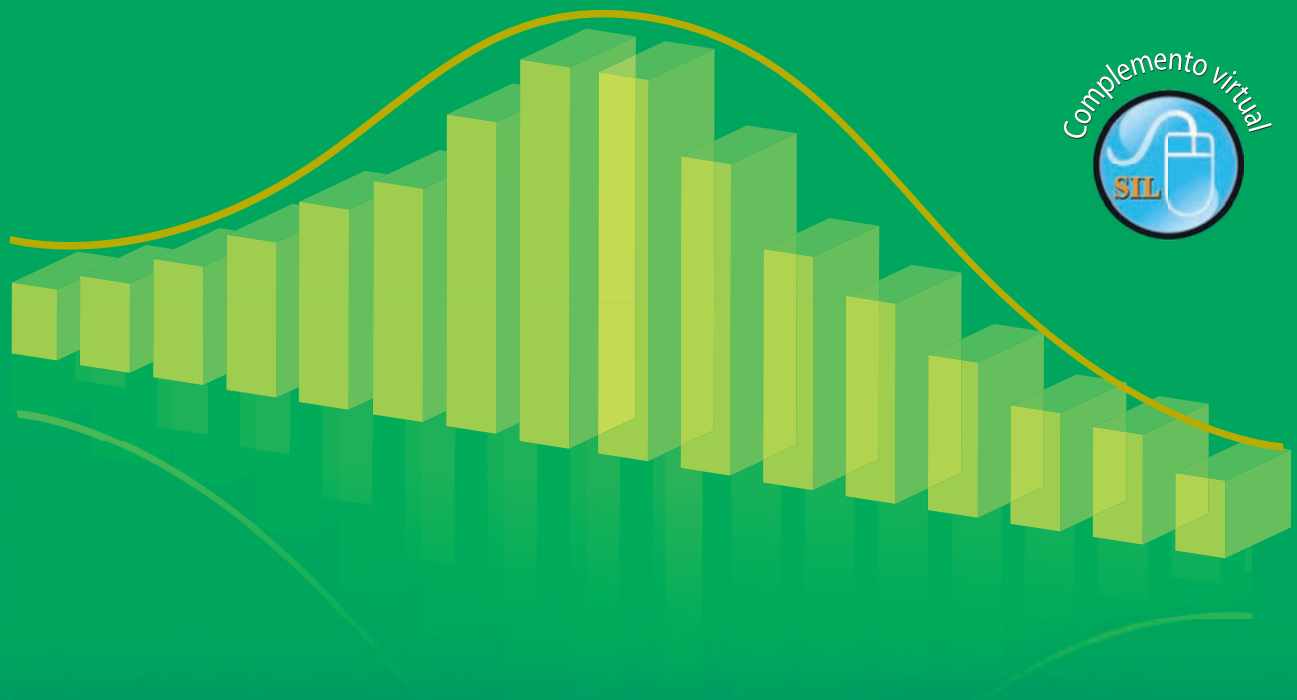
Décima tercera edición

Estadística y muestreo



Ciro Martínez Bencardino

ECOE EDICIONES

CIRO MARTINEZ BENCARDINO

Nacido en Convención (Norte de Santander - Colombia). Economista de la Universidad Jorge Tadeo Lozano de Bogotá, D.C. Bio-estadística (Universidad de los Andes, Bogotá, D.C.). Técnicas Estadísticas (CIENES-Santiago de Chile) y Estadística Laboral (Universidad de Río Piedras y Negociado Laboral de Puerto Rico).

Vinculado a la enseñanza de la Estadística en un gran número de instituciones universitarias de Bogotá. Durante muchos años ha trabajado en el campo de la estadística, ocupando diferentes cargos gubernamentales.

Entre sus publicaciones destacamos: *Estadística comercial*; *Muestreo, algunos métodos y sus aplicaciones prácticas*; *Estadística Básica Aplicada*; libros editados por Ecoe Ediciones.

Décima tercera edición

Estadística y muestreo

Ciro Martínez Bencardino

Catalogación en la publicación – Biblioteca Nacional de Colombia

Martínez Bencardino, Ciro
Estadística y muestreo / Ciro Martínez Bencardino. -- 13ª. ed. -- Bogotá : Ecoe Ediciones, 2012.
900 p. – (Ciencias exactas. Matemáticas)

ISBN 978-958-648-702-3

1. Estadística matemática 2. Muestreo (Estadística) I. Título II. Serie

CDD: 519.52 ed. 20 CO-BoBN– a746986

Colección: Textos universitarios
Área: Matemáticas
Primera edición: Bogotá, D.E., noviembre de 1978
Segunda edición: Bogotá, D.E., enero de 1982
Tercera edición: Bogotá, D.E., enero de 1984
Cuarta edición: Bogotá, D.E., julio de 1987
Quinta edición: Bogotá, D.E., agosto de 1990
Sexta edición: Santa Fe de Bogotá, D.C., julio de 1992
Séptima edición: Santa Fe de Bogotá, D.C., enero de 1995
Octava edición: Santa Fe de Bogotá, D.C., febrero de 1997
Novena edición: Santa Fe de Bogotá, D.C., febrero de 1998
Reimpresión: Santa Fe de Bogotá, D.C., abril de 1999
Décima edición: Bogotá, D.C., octubre de 2000
Onceava edición: Bogotá, D.C., enero de 2002
Reimpresión: Bogotá, D.C., septiembre de 2003
Décimo segunda edición: Bogotá, D.C., septiembre de 2005
Reimpresión: Bogotá, D.C., enero de 2007
Décima tercera edición: Bogotá, D.C., 2012

ISBN:978-958-648-702-3

© Ciro Martínez Bencardino e-mail: cimarben@yahoo.es
© Del contenido del SIL, Ciro Martínez Bencardino
© ECOE ediciones Ltda.
E-mail: correo@ecoeediciones.com
www.ecoeediciones.com
Carrera 19 No. 63C-32, Tel. 2481449

Coordinación editorial: Alexander Acosta Quintero
Diagramación: Raúl Enrique Rodríguez
Portada: Edwin Nelson Penagos
Impresión: Imagen Editorial Impresores
imagenimvega@yahoo.com

Impreso y hecho en Colombia

A mis padres, hermanos, esposa, hijos y nietos, por ellos vivo, investigo y trabajo.

Como un homenaje póstumo a la memoria del doctor Carlos Alzate Giraldo, fundador de ECOE.

ÍNDICE GENERAL

PRESENTACIÓN **XIX**

CAPÍTULO 1

CONCEPTOS GENERALES **1**

ASPECTOS GENERALES 2
Reseña histórica 2
Aplicación de la estadística 2
Finalidad de la estadística 4
Colectivos investigados por la estadística 5
Algunos términos y conceptos que se deben conocer y manejar 7
LA PROYECCIÓN Y PREPARACIÓN DE INVESTIGACIONES ESTADÍSTICAS 10
Planeamiento y preparación de una investigación de tipo estadístico 10
El objeto de la investigación 11
Las fuentes de información 15
Los procedimientos de investigación 17
El material estadístico 23
El presupuesto de la investigación 26
Recolección 28
Crítica y codificación 29
Tabulaciones o procesamiento 30
Análisis e interpretación 30
Publicación 32
MONOGRAFÍAS Y ENCUESTAS 33
Ejercicios para resolver 35
SÍNTESIS DE CAPÍTULO 39

CAPÍTULO 2

DISTRIBUCIONES DE FRECUENCIAS **43**

GENERALIDADES 44
ELABORACIÓN DE TABLA DE FRECUENCIAS 45
Variables discretas, continuas y de atributos 45
Propiedades de las frecuencias 51
Ejercicios para resolver 54
Recomendaciones en la elaboración de cuadros y tablas 58
ELABORACIÓN DE GRÁFICAS 59

Gráficas aplicadas al desarrollo de la Teoría Estadística 60
Ejercicios para resolver 63
APLICACIONES DE EXCEL 64
Gráficas utilizadas en la presentación de informes 68
Ejercicios para resolver 85
SÍNTESIS DE CAPÍTULO 88

CAPÍTULO 3

MEDIDAS DE POSICIÓN Y DE TENDENCIA 91

ASPECTOS GENERALES 92
Características, uso, ventajas y desventajas de promedios 92
Media aritmética 93
Ventajas y desventajas 93
Simple y ponderada 94
Desviaciones y propiedades 97
Ejercicios para resolver 99
Mediana 103
Ventajas y desventajas 103
Simple y ponderada 104
Propiedades 104
Modo 107
Ventajas y desventajas 107
Simple y ponderada 107
Datos sin agrupar u originales 107
Relación entre la Media, Mediana y el Modo. 109
Ejercicios para resolver 110
Media Geométrica 112
Ventajas y desventajas 112
Simple y ponderada 112
Aplicaciones 113
Media Armónica 114
Ventajas y desventajas 115
Simple y ponderada 115
Aplicaciones 116
Ejercicios para resolver 117
Media Cuadrática 118
Ventajas y desventajas 119
Simple y ponderada 119
Media Cúbica 119
Simple y ponderada 120
Relación entre promedios 121
Centro Recorrido 121
Cuartiles, Deciles y Percentiles 122
Ejercicios para resolver 124
SÍNTESIS DE CAPÍTULO 128

ANEXOS DE LA UNIDAD 3 130
Aplicaciones con calculadora 130
APLICACIONES DE EXCEL 131

CAPÍTULO 4

MEDIDAS DE DISPERSIÓN, DE DEFORMACIÓN Y APUNTAMIENTO 143

CONCEPTOS GENERALES 144
MEDIDAS DE DISPERSIÓN 144
Varianza 144
Simple y ponderada 145
Propiedades 147
Aplicaciones 147
Desviación Típica 149
APLICACIONES DE EXCEL Y LA CALCULADORA 150
Ejercicios para resolver 152
Variación Absoluta 154
Coeficiente de Variación 154
Puntaje Típico o Estandarizado 155
Desviación Media 157
Desviación Mediana 158
Ejercicios para resolver 159
Recorrido Intercuartílico, Desviación Cuartil y Coeficiente de Desviación Cuartil 163
Recorrido u Oscilación 166
MEDIDAS DE ASIMETRÍA O DE DEFORMACIÓN 167
Ejercicios para resolver 169
MEDIDAS DE APUNTAMIENTO 170
Ejercicios para resolver 172
SÍNTESIS DE CAPÍTULO 176

CAPÍTULO 5

NOCIONES ELEMENTALES DE PROBABILIDAD 177

GENERALIDADES 178
Definición, elaboración de espacios muestrales 180
Ejercicios para resolver 184
PERMUTACIONES Y COMBINACIONES 188
Uso de la calculadora 192
Ejercicios para resolver 193
ALGUNAS REGLAS BÁSICAS DE PROBABILIDAD 196
Regla de adición 196
Sucesos mutuamente excluyentes 196
Regla de la multiplicación 201
Sucesos independientes y dependientes 201
Ejercicios para resolver 204
Probabilidad condicional 207

TEOREMA DE BAYES 209
Ejercicios para resolver 211
SÍNTESIS DE CAPÍTULO 215

CAPÍTULO 6

DISTRIBUCIONES DE PROBABILIDAD 217

CONCEPTOS GENERALES 218
VARIABLE ALEATORIA DISCRETA 219
Distribución Binomial 219
Definición, fórmulas y aplicaciones 222
APLICACIÓN CON LA CALCULADORA Y EXCEL 225
Ejercicios para resolver 230
Distribución de Poisson 235
Definición, fórmulas y aplicaciones 236
APLICACIÓN CON LA CALCULADORA Y EXCEL 237
Ejercicios para resolver 241
Distribución Hipergeométrica 243
Definición, fórmulas y aplicaciones 244
Ejercicios para resolver 246
VARIABLE ALEATORIA CONTINUA 247
Distribución normal 247
Representación gráfica 248
Condiciones que debe tener en cuenta 249
Definición, usos y aplicaciones 249
APLICACIÓN DE EXCEL 253
Ejercicios para resolver 257
SÍNTESIS DE CAPÍTULO 271

CAPÍTULO 7

DISTRIBUCIONES MUESTRALES. MUESTREO ALEATORIO 273

CONCEPTOS GENERALES 274
DISTRIBUCIONES MUESTRALES 277
Definición, fórmulas y aplicaciones 277
Distribución de medias muestrales 278
Ejercicios para resolver 284
Distribución de diferencia entre dos medias muestrales 293
Ejercicios para resolver 297
Distribución de diferencias entre dos proporciones 299
Ejercicios para resolver 302
TAMAÑO DE LA MUESTRA 303
Ejercicios para resolver 309
SÍNTESIS DE CAPÍTULO 320

CAPÍTULO 8

PRUEBAS DE HIPÓTESIS Y LÍMITES DE CONFIANZA 323

CONCEPTOS GENERALES 324
PRUEBA DE HIPÓTESIS 324
Tipo de error 326
Hipótesis nula y alternativa 327
Prueba unilateral y bilateral 328
Nivel de significación y puntos críticos 328
Etapas de la realización de la prueba 329
Distribución de medias muestrales 332
Ejercicios para resolver 336
Distribución en una proporción 339
Ejercicios para resolver 341
Distribución de diferencias entre dos medias 343
Ejercicios para resolver 345
Distribución de diferencias entre dos proporciones 347
Ejercicios para resolver 349
TEORÍA SOBRE MUESTRAS PEQUEÑAS 351
Distribución "t" de student 351
Distribución de medias muestrales 352
Ejercicios para resolver 357
Distribución en una proporción muestral 358
Ejercicios para resolver 359
Distribución de diferencias entre dos medias 360
Ejercicios para resolver 365
Distribución de diferencias entre dos proporciones 368
Ejercicios para resolver 369
APLICACIÓN CON EXCEL 369
LÍMITES DE CONFIANZA 380
Aplicación en las diferentes distribuciones 381
Ejercicios para resolver 392
APLICACIÓN CON EXCEL 395
SÍNTESIS DE CAPÍTULO 401

CAPÍTULO 9

OTRAS PRUEBAS DE HIPÓTESIS 403

ASPECTOS GENERALES 404
PRUEBA DE HIPÓTESIS DE UNA VARIANZA 404
Teoría, aplicaciones y procedimientos 404
Procedimiento más utilizado en la prueba con una varianza 408
Ejercicios para resolver 410
COMPARACIÓN ENTRE VARIANZAS DE DOS POBLACIONES 410
Distribución F 410
Límites de confianza 413

Ejercicios para resolver 414
PRUEBA DEL COEFICIENTE DE CORRELACIÓN DE PEARSON 414
Ejercicios para resolver 418
PRUEBAS CON OBSERVACIONES APAREADAS 419
Ejercicios para resolver 423
PRUEBAS NO PARAMÉTRICAS 427
Prueba de chi-cuadrado 427
Tablas de contingencia 431
Dócimas de homogeneidad y de independencia 434
APLICACIONES CON EXCEL 436
Ejercicios para resolver 441
Dócima o pruebas del signo 450
Ejercicios para resolver 455
Correlación por rangos 460
Prueba T de Wilcoxon 462
Muestras Pequeñas 462
Muestra Grande 464
Prueba U de Mann-Whitney 466
Muestras Grandes 466
Muestras Pequeñas 468
Muestras Muy Pequeñas 469
Prueba H de Kruskal y Wallis 470
Teoría y aplicaciones 471
Ejercicios para resolver 472
SÍNTESIS DE CAPÍTULO 477

CAPÍTULO 10

NÚMEROS ÍNDICES 479

GENERALIDADES 480
ÍNDICES SIMPLES 480
Encadenamiento de índices 482
ÍNDICES AGREGATIVOS SIMPLES 483
ÍNDICES PONDERADOS DE PRECIOS Y CANTIDAD 484
Índice de Laspeyres, Paasche, Fisher 484
Índice de Sidgwick-Drobisch, Marshall-Edgeworth 485
Índice de Walsh, Keynes 486
ÍNDICES DE PROMEDIOS PONDERADOS 487
Ejercicios para resolver 488
ÍNDICE DE VALOR 493
Ejercicios para resolver 493
EMPALME DE DOS O MÁS SERIES 494
Ejercicios para resolver 495
ALGUNAS APLICACIONES DE LOS NÚMEROS ÍNDICES 496
Índice de precios al consumidor (IPC) 496
Tasa de cambio (TC) 500

Ejercicios para resolver ... 500
Índices de comercio exterior ... 504
Ejercicios para resolver ... 505
Índices de precios implícitos ... 505
Índice de productividad (IP) ... 508
Ejercicios para resolver ... 509
Índice precios del productor (IPP) ... 510
Índice bursátil mundial Dow Jones ... 511
Indicadores de desempleo ... 511
Otros indicadores (OI) ... 512
Ejercicios para resolver ... 513
SÍNTESIS DE CAPÍTULO ... 517

CAPÍTULO 11

SERIES CRONOLÓGICAS ... 519

GENERALIDADES ... 520
COMPONENTES DE UNA SERIE TEMPORAL ... 520
TENDENCIA ... 521
Método analítico ... 521
Método de mínimos cuadrados ... 521
Cálculo, aplicaciones, estimativos ... 521
Límites de confianza, coeficiente de correlación ... 524
Representación gráfica ... 526
APLICACIÓN DE EXCEL ... 527
Ejercicios para resolver ... 529
Estimaciones mensuales con base en datos anuales ... 533
Ejercicios para resolver ... 536
Método de los semipromedios ... 536
Ejercicios para resolver ... 537
Método de los promedios móviles ... 538
APLICACIÓN DE EXCEL ... 541
Tendencia parabólica ... 543
Cálculo, aplicaciones, estimativos ... 543
Límites de confianza, coeficiente de correlación ... 546
Ejercicios para resolver ... 547
Tendencia exponencial ... 549
Cálculo, aplicaciones, estimativos ... 549
Límites de confianza, coeficiente de correlación ... 551
Ejercicios para resolver ... 553
OTROS MÉTODOS ... 555
Curva exponencial, curva logística y de Gompertz ... 555
Método de los componentes y gráfico ... 555
Ejercicios para resolver ... 556
Variaciones estacionales ... 558
ÍNDICE ESTACIONAL ... 559

MÉTODO DE RAZÓN 559
MÉTODO DE LAS RAZONES AL PROMEDIO MÓVIL 562
Ejercicios para resolver 566
SÍNTESIS DE CAPÍTULO 567

CAPÍTULO 12

REGRESIÓN Y CORRELACIÓN 569

ASPECTOS GENERALES 570
REGRESIÓN Y CORRELACIÓN 570
APLICACIÓN DE EXCEL 573
Diagrama de dispersión 575
REGRESIÓN Y CORRELACIÓN LÍNEAL 578
Método de los mínimos cuadrados 580
Ecuación de la recta, cálculo de los parámetros, estimaciones 580
Covarianza, varianza residual y explicada. Graficación 581
Coeficiente de correlación. Límites de confianza. Gráficas 581
APLICACIÓN DE EXCEL 582
Prueba del coeficiente de regresión y correlación 593
Estimativos de x en función de y viceversa 595
Uso de la calculadora 597
Ejercicios para resolver 602
REGRESIÓN PARABÓLICA SIMPLE 622
Método de los mínimos cuadrado. Cálculo de los coeficientes. 622
Varianza residual y explicada. Estimativos. Coeficiente de Correlación. Límites de confianza. Gráficas. 627
REGRESIÓN PARABÓLICA PONDERADA 630
Ejercicios para resolver 633
REGRESIÓN EXPONENCIAL Y LOGARÍTMICA 634
Método de los mínimos cuadrado. Cálculo de los coeficientes. 634
Varianza residual y explicada. Estimativos. Coeficiente de correlación. Límites de confianza. Gráficas 636
REGRESIÓN Y CORRELACIÓN MÚLTIPLE 637
Método de los mínimos cuadrados 640
Método de Jordan-Gauss 643
Ejercicios para resolver 645
COEFICIENTE DE CORRELACIÓN POR RANGOS 647
Ejercicios para resolver 649
SÍNTESIS DE CAPÍTULO 655

CAPÍTULO 13

ALGUNOS MÉTODOS DE MUESTREO 657

GENERALIDADES 658
Población o universo 658

Unidad y elemento 658
Población finita e infinita 660
Características 661
Investigación total y parcial 661
MUESTREO NO ALEATORIO 663
Muestreo opinático, circunstancial, por cuotas, a juicio 663
MUESTREO ALEATORIO 663
Marco 666
Falseamiento del esquema de selección 666
Sustitución de unidades 667
Dominio de estudio 668
Métodos de selección 668
APLICACIÓN DE EXCEL Y CALCULADORA EN NÚMEROS 670
Objeto del muestreo aleatorio 677
Ejercicios para resolver 680
APLICACIÓN DE MÉTODOS Y TÉCNICAS DE MUESTREO 687
Muestreo aleatorio simple (M.A.S) 687
Tamaño de la muestra 687
Cálculo de algunos estimativos 691
Estimación de promedios y totales 693
Estimación de proporciones y totales 696
Estimación de proporciones y totales en conglomerados 697
Estimativos de razones como método indirecto para estimar promedios y totales 698
Estimación de promedios y totales mediante la regresión 699
Estimación de promedios, proporciones y totales en dominio de estudio 703
Ejercicios para resolver 708
MUESTREO ALEATORIO ESTRATIFICADO MAE 718
Asignación igual 721
Cálculo de los tamaños muestrales 722
Cálculo de estimativos y fijación de límites 725
Muestra asignación igual 726
Estimación de promedios y totales 727
Estimación de una proporción y total 730
Estimación para la razón, promedio y el total 731
Estimación de promedios y totales mediante la regresión lineal simple 735
Estimación de una proporción y el total en conglomerados 738
Ejercicios para resolver 740
Asignación proporcional 744
Tamaño de la muestra 745
Encuesta preliminar – Asignación proporcional 746
Muestra de asignación proporcional 748
Estimación de promedios y totales 749
Estimación de proporciones y totales 750
Estimación de proporciones y totales en conglomerados 750
Estimación indirecta del promedio y el total a través del método de la razón 751
Estimación indirecta del promedio y total mediante la regresión lineal 753

Asignación óptima 757
Tamaño de la Muestra 757
MÉTODO DE NEYMAN 763
Ejercicios para resolver 764
MUESTREO SISTEMÁTICO 773
MUESTREO POR CONGLOMERADOS 777
MUESTREO BIETÁPICO 783
MUESTREO POR CONGLOMERADOS DE DOS ETAPAS 784
Ejercicios para resolver 789
MUESTREO POR FASES MÚLTIPLES 791
MÉTODOS MIXTOS 792

APÉNDICE 793

GLOSARIO 794
TABLAS 817
RESUMEN DE FÓRMULAS 865

BIBLIOGRAFÍA 873

ÍNDICE DE TABLAS

TABLAS

I. Áreas bajo la curva normal de probabilidad 817
II. Distribución "t" de student 820
III. Distribución binomial 822
IV. Distribución de j_i cuadrado. Áreas bajo la curva normal de probabilidad 828
V. Transformación de r a z 830
VI. Números al azar 831
VII. Distribución de poisson 837
VIII. Intervalo de confianza del 95% (2 sigma) 838
IX. Seguridad del 95% (2 sigma) 839
X. Seguridad del 99,7% (3 sigma) 840
XI. Tabla para la determinación de una muestra (n) sacada de una población finita para márgenes de error de 1%, 2%, 3%, 4%, 5%, 10%, en la hipótesis de p = 50%. Margen de confianza del 95% 841
XII. Tabla para la determinación de una muestra (n) sacada de una población finita para márgenes de error de 1%, 2%, 3%, 4%, 5%, 10%, en la hipótesis de p = 50%. Margen de confianza del 99,7% 842
XIII. Probabilidades asociadas con valores tan pequeños como los valores observados de U en la prueba de Mann-Whitney 843
XIV. Tabla de los valores críticos de U en la prueba de Mann-Whitney. Valores críticos de para una prueba de una cola en a = 0,001 o para una prueba de dos colas en a = 0,002. 846
XV. Percentillas de las distribuciones (x^2/g.L.) 850
XVI. Distribución F 851
XVII. Coeficientes de Spearman para la correlación de rango 859
XVIII. Prueba de Wilcoxon 860
XIX. Distribución binomial 861

FÓRMULAS

XX. Fórmulas aplicadas en las medidas de posición, dispersión, asimetría y apuntamiento 865
XXI. Fórmulas aplicadas a la regresión y correlación 867
XXII. Fórmulas para series de tiempo 869
XXIII. Fórmulas aplicadas en los números índices 870
XXIV. Fórmulas para inferencia estadística 871

CONTENIDO DEL SIL

1. Solucionario de los ejercicios para resolver en el libro - Capítulos 1 al 13
2. Aplicación de EXCEL en los procesos Estadística
3. Resumen de fórmulas de aplicación en la Estadística Descriptiva y en la Inferencia
4. Teoría y aplicaciones, sobre: Desarrollo de un sistema de ecuaciones; Sumatoria y Productoria; Porcentajes, Tasas y Razones

PRESENTACIÓN

Gracias a la generosidad de profesores y estudiantes que lo han preferido como texto o como fuente de consulta, se ofrece esta nueva Edición, con nueva presentación mucho más cómoda y fácil de manejar, con una revisión y actualización de los contenidos, así como de estilo, que esperamos sea del agrado de todos aquellos que de alguna manera tendrán la oportunidad de consultarla.

En esta décima tercera edición se suprimieron algunos temas que han perdido vigencia en los programas, en contraprestación se incluyeron y mejoraron otros, manteniendo las exigencias y la actualización en la enseñanza de esta materia de acuerdo a los diferentes niveles, así como en los objetivos que se exigen en las carreras intermedias y profesionales. En esta ocasión se ha suprimido el CD que siempre lo acompañaba, ahora se ofrece la consulta a la Página web w**ww.ecoeediciones.com** donde encontrará entre otros, a) Proceso continuado de la aplicación del Excel en todos los temas de esta publicación. b) Solucionario con todo el proceso que conlleva el desarrollo de los problemas propuestos, c) Además incluye un tema que fue suprimido pero que lo seguimos considerándolo importante en proceso integral del conocimiento de esta materia, como son la Sumatoria, Productoria, y el uso de los Porcentajes, Tasas y Razones

El objetivo principal y la razón de esta publicación sigue siendo el estudiante, a fin de facilitarle el aprendizaje de esta valiosa herramienta, de ahí la equivocación de algunos que la consideran "elemental", comentarios que nos alegra, pues no está dirigido a profesionales de la Estadística con grandes conocimientos matemáticos, por el contrario se pretende ayudar a aquellos que apenas se inician en esta disciplina, eliminando procesos que no solo dificultan su aprendizaje, que de alguna manera hacen perder el interés sobre aquellos temas que tienen gran importancia en su aplicación en tantos campos del conocimiento.

Creo firmemente que los contenidos de esta publicación van a ser un valioso apoyo, para quienes creen que los programas de computador la han desplazado, este último solo permite observar la base de datos en su conjunto y los resultados obtenidos que son los requeridos, pero elimina los procedimientos que fueron aplicados para su obtención.

Sea la oportunidad para manifestar nuevamente, mis más sinceros agradecimientos a profesores y estudiantes que me han hecho llegar comentarios, críticos algunos, que han sido de gran ayuda al mejoramiento de los contenidos en cada edición, como aquellos que reconocen esta labor, lo que me da aliento para seguir en esta tarea de divulgación.

El Autor.

CAPÍTULO 1

Conceptos generales

TÉRMINOS Y DEFINICIONES

El ser humano aprende en la medida en que participa en el descubrimiento y la invención. Debe tener la libertad para opinar, para equivocarse, para rectificarse, para ensayar métodos y caminos para explorar

Ernesto Sábato

CONTENIDO

- Términos, definiciones, conceptos generales.
- Finalidad de la Estadística,
- Etapas en una investigación de carácter estadístico.
- Monografías y Encuestas Diferencias.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- ❑ Entender y manejar correctamente cada uno de los términos que deberán ser utilizados.
- ❑ Conocer la importancia que tiene la Estadística no sólo como asignatura, sino en el ejercicio profesional.
- ❑ Poder realizar pequeñas investigaciones estadísticas, aplicando los conocimientos recientemente adquiridos, en cada una de las etapas del proceso.
- ❑ Distinguir con claridad la diferencia entre una Monografía y una Encuesta, de carácter estadístico.

ASPECTOS GENERALES

RESEÑA HISTÓRICA

La *estadística* que muchos la consideran como algo nuevo, dado que hoy en día todo se fundamenta en el manejo de grandes cantidades de información o datos de carácter estadísticos, pero su aplicación o uso se remonta a la antigüedad, pues hay vestigios de su inicio como en China donde el emperador Yao (2.239 A de C.) dispuso la realización de un Censo; algo similar ocurrió en la época del Rey Herodes, relacionado con un empadronamiento cuando se habló del nacimiento de Jesús.

La *estadística* como disciplina, tuvo su inicio en Alemania a mediados del Siglo XVII, en el reinado de Godofredo de Achewall (1.719-1.772) se utilizó la palabra estadística y se separó de la Sociología. Hoy en día ha alcanzado un alto desarrollo, con el apoyo de tantos investigadores dedicados a perfeccionar métodos estadísticos y a través del avance tecnológico, permitiendo la selección de aquellos procedimientos que buscan mejorar los resultados, al mismo tiempo disminuyendo los márgenes de error.

La palabra *estadística,* para algunos, proviene de la palabra status, que significa estado; para otros se deriva del vocablo griego statera que significa balanza, otra más confiable proviene de la palabra alemana Stara que significa estado, la encargada de hacer estadísticas.

APLICACIÓN DE LA ESTADÍSTICA

En la vida diaria, los diversos fenómenos de orden económico, social, político, educacional e incluso biológico, aparecen, se transforman y finalmente desaparecen.

Para tan abundante y complejo material informativo recogido, se hace necesario contar con un registro ordenado y continuo, a fin de determinar o seleccionar los datos requeridos por el estudio o investigación, en cuanto a lo sucedido, sucede o puede suceder. Para ello se requiere contar con un método o conjunto de reglas o principios, que nos permita la observación, el ordenamiento, la cuantificación y el análisis de dichos fenómenos. Este método se le denomina *Estadística.*

> La palabra *estadística* se refiere a un sistema o método usado en la recolección, organización, análisis y descripción numérica de la información. También se puede decir que la estadística estudia el comportamiento de hechos o fenómenos de grupo.

Se consideran dos fases en el campo de la estadística. En primer lugar, está la fase que sólo se limita a la descripción de un conjunto de datos sin llegar a conclusiones o a generalizar con respecto a un grupo mayor. A esta fase se le da el nombre de *Estadística Descriptiva o Deductiva*. En la segunda, ella implica el análisis mediante la cual trata de llegar a conclusiones acerca de un grupo más grande o población, basado en la información de un grupo menor o muestra, procedimiento o técnica denominada como *Estadística Inductiva* o *Inferencia*.

En un principio se consideraba que la función primordial de la *Estadística*, era la descripción de las características de un grupo, actividad que la hacía confundir con el papel que cumple la historia, consistente en observar y describir el hecho. En su origen, las estadísticas eran históricas; hoy en día, la estadística, además de ser descriptiva, es analítica, considerándose ésta cómo la función más importante que

ella realiza, ya que además, permite obtener conclusiones para un grupo mayor, denominado población, partiendo de una investigación realizada en un grupo menor, conocido como muestra, cuyos elementos o unidades, en la mayoría de los casos, se seleccionan aleatoriamente o al azar.

La *estadística descriptiva o deductiva*, tiene como finalidad, colocar en evidencia aspectos característicos del grupo (promedios, variabilidad de los datos, etc.) que sirven para realizar comparaciones, sin pretender sacar conclusiones de tipo más general. Esta descripción, por lo general se realiza a través de la elaboración de cuadros, gráficos, cálculo de promedios, proporciones, varianzas de una o más variables relacionadas.

La *estadística analítica o inductiva*, busca dar explicaciones al comportamiento de un conjunto de observaciones, probar la significación o validez de los resultados, además, intenta descubrir las causas que lo originan, siendo de gran aplicación en el muestreo, logrando de esta manera, obtener conclusiones que se extienden mas allá de las muestras estadísticas mismas.

Uno de los objetivos principales de la *estadística*, consiste en realizar inferencias acerca de los valores estadísticos de la población, denominados parámetros, a través de la información obtenida mediante una muestra, que nos permite el cálculo de estimadores.

> La *estadística descriptiva-analítica*, se podría definir: como un conjunto sistemático de procedimientos utilizados para observar y describir numéricamente un fenómeno, además, descubrir las leyes que regulan la aparición, transformación y desaparición de los mismos.

También se puede considerar la *estadística*, como aquel método que permite no sólo describir el hecho o fenómeno, sino también, que permite deducir, evaluar y sacar conclusiones acerca de una población, con los resultados obtenidos a través de una muestra.

Generalmente asociamos la *estadística* con cifras sobre un campo en particular. Como por ejemplo: el número de nacimientos, defunciones, transacciones comerciales, valor de las acciones en el mercado de valores, volumen físico y monetario de las importaciones y exportaciones, beneficios y utilidades de las empresas, demanda presente o potencial de algún producto.

> *Estadísticas* (en plural). Se refiere a un ordenamiento sistemático de datos, presentados en forma de tablas o cuadros y gráficas. En otras palabras, las *estadísticas* son datos agrupados en forma metódica y presentados en publicaciones, elaboradas por diversas empresas o entidades privadas o del estado, buscando ser consultadas por los interesados.

Se clasifican en:

Estadísticas o fuentes primarias. Son datos obtenidos mediante encuestas directas, a través de la utilización de cuestionarios o como resultado de la observación directa, técnica muy utilizada en estudios de carácter científico o en investigaciones de mercados. Se podrá decir, que son datos publicados por quien recoge directamente la información primaria.

Éstas podrían clasificarse en*: personales* (entrevistas, correo, experimentos, etc.); ***unipersonales*** (auditoría, análisis de rastreo o de contenido, simulación); mixta (observaciones).

Estadísticas o fuentes secundarias. En éstas, los datos se obtienen de publicaciones, las cuales podrán ser reproducidas en forma total o parcial. Se les considera de gran valor para cualquier tipo de investigación. Son consideradas como *fuentes secundarias*: bibliotecas, centros de documentación, folletos, revistas. archivos, etc.

Estadísticas o series temporales. Generalmente denominadas como series de tiempo o cronológicas. La información es obtenida y ordenada en forma cronológica, a medida que se va produciendo a través del tiempo, ya sean días, meses, semestres, años. Cuando la investigaciones se realizan aisladamente, no periódicas, se les denomina: investigaciones *atemporales o aisladas*.

Las estadísticas, también se pueden clasificar como *internas o externas.* Las primeras se originan y se elaboran en la misma fuente de información, por ejemplo en una empresa, son registros internos, tales como: la producción y ventas diarias, semanales, etc. salarios, horas trabajadas, etc. Las *externas*, son registros de información originados fuera de la empresa: opinión de los consumidores, precios de la competencia, etc.

Clases de estadísticas. Se clasifican de acuerdo con el tema, sector o materia de estudios, así por ejemplo: estadísticas industriales, agropecuarias, de transporte, comunicaciones, salud, educación, etc.

Estadístico. Nos referimos a la persona que trabaja en la elaboración y análisis de información de carácter estadístico. En nuestro medio utilizamos el término estadígrafo, para referirnos a esa persona, cuando en realidad su significado es el de una medida estadística.

FINALIDAD DE LA ESTADÍSTICA

En términos generales, se considera que su finalidad es suministrar información y su utilidad dependerá en gran parte del fin que se proponga y de la forma como se obtengan los datos. La creciente especialización lograda por las empresas, ha obligado a la utilización continua de métodos estadísticos, como en la producción masiva, en el control de calidad, en éste último, evitando la inspección de cada artículo y reemplazado por selección de muestras. Por ejemplo, en el control realizado sobre dureza, resistencia y duración, que por lo general implica la destrucción del artículo, lo cual hace necesario la aplicación de las técnicas de muestreo, a fin de reducir su número, como forma de obtener conclusiones que puedan ser extensivas al total de la producción de la cual se extrajo la muestra.

Otras aplicaciones que se pueden hacer, son las investigaciones realizadas con el objeto de conocer los cambios en los gastos o gustos de los consumidores, situaciones en la demanda: exceso en las existencias; capacidad de compra en un grupo familiar. Precisando sobre la finalidad de la *estadística*, podríamos mencionar:

- *Conocer la realidad en una observación o fenómeno.* Al cuantificarlo, estamos conociendo su situación actual, es así como al tener información por horas, semanalmente, etc. sobre la producción de un artículo, se sabrá si se está cumpliendo con lo programado o, si por el contrario, estamos por debajo o por encima de lo establecido, pues ello afectará de alguna manera las existencias del producto o de la materia prima, ventas, precios, costos, etc.

- *Determinar lo típico o normal de esa observación*. Cuando se cuantifica la característica de un fenómeno, se obtiene un valor denominado ***promedio***, es ésta la forma de referirnos al grupo cuantificando así su comportamiento. Si decimos que el promedio de unidades vendidas de un artículo A, en 20 almacenes, en el mes de marzo fue de 1.200 unidades, es una forma de determinar el comportamiento del grupo; pero con este valor típico o normal del conjunto no se quiere decir que los 20 almacenes vendieron la misma cantidad; posiblemente algunos más, otros menos, pero el promedio fue de 1.200.
- *Determinar los cambios que presenta el fenómeno.* Cualquier fenómeno de carácter estadístico presenta variaciones a través del tiempo, pero requiere una observación continua para poder determinar la magnitud del cambio. La demanda, la producción, las ventas, los precios, son fenómenos que constantemente registran cambios que deben ser tenidos en cuenta por la administración de una empresa.
- *Relacionar dos o más fenómenos*. Desde el punto de vista de correlación, se puede determinar si existe una relación válida entre dos o más características de una misma observación, o entre dos o más fenómenos. Tal es el caso de la relación de ingresos y gastos para un grupo de familias o la relación entre las unidades producidas, vendidas y el precio.
- *Determinar las causas que originan el fenómeno.* Los cambios que se observan en un fenómeno pueden tener una o varias causas de origen. El aumento de las exportaciones de un artículo puede ser causado por un mejoramiento del precio en el mercado externo, por un volumen de producción superior a la demanda interna, o por otras causas que producen cambios en el comportamiento del fenómeno.
- *Hacer estimativos sobre el comportamiento futuro del fenómeno*. En numerosas ocasiones se requiere proyectar, para estimar el comportamiento futuro de un fenómeno. La proyección a dos o tres años de la población de una región, sirve para determinar el mercado potencial del producto, es decir, el conocimiento, aunque sea estimado, de la demanda en los años próximos.
- *Obtener conclusiones de un grupo menor (muestra), para hacerlas extensivas a un grupo mayor (población).* Al realizar mediciones en una muestra, cuyo tamaño debe ser el óptimo deseado, los resultados obtenidos pueden ser considerados como el comportamiento que se obtendrá si se hubiera trabajado con todos los elementos que constituyen la población, de la cual se extrajo la muestra.

Además de lo anterior, la *estadística* facilita una serie de instrumentos o técnicas que, al ser utilizadas correctamente, permiten determinar el grado de validez y confiabilidad, ya sea de las predicciones o de las conclusiones obtenidas a partir de muestras.

COLECTIVOS INVESTIGADOS POR LA ESTADÍSTICA

La conformación de un grupo o colectivo alrededor de cierta característica, o características, que permiten ser investigadas en cuanto a su comportamiento, se denomina fenómeno. Los fenómenos pueden ser de carácter económico, social, politico, u otros, considerando como tales: exportaciones, importaciones, ventas, producción, defunciones, nacimientos, accidentes laborales, etc.

Los fenómenos de carácter estadístico deben reunir ciertos requisitos, como ***manifestarse al exterior para poder ser observados***, lo cual se logra mediante registros. Se consideran como ejemplos: las nóminas de pago, donde quedan consignados el número de empleados, las categorias, sueldos y otras tantas características que podrian ser observadas; los manifiestos de aduana, en donde se registran, tanto

las importaciones como las exportaciones; las licencias de construcción concedidas por las Secretarías de Obras Públicas; los registros de nacimientos, defunciones, escrituras de venta, hipotecas, todas ellas realizadas en las Notarías, etc. Además, las características de estos fenómenos *deben ser cuantificables* en tal forma que permitan determinar la intensidad con que se producen. Un ejemplo clarifica mejor lo anterior: si consideramos las exportaciones de café realizadas por las diferentes firmas exportadoras en un período determinado, encontramos que se trata de un fenómeno que se manifiesta al exterior y que se puede observar en los registros de exportación, y por otra parte, se cuantifica el número y valor de las remesas realizadas, además se podrá obtener la intensidad o sea la cantidad de sacos o kilos correspondientes a cada remesa o al total exportado en dicho período.

Fuera de las anteriores condiciones implícitas en un fenómeno estadístico, podemos anotar otras que se consideran dentro del campo de estudio de la estadística:

- *Colectivos.* La estadística estudia el comportamiento de los fenómenos de grupo, prescindiendo de aquellos fenómenos individuales que pueden ser considerados como resultados de casos aislados. Si bien es cierto que para estudiar el colectivo se requiere información individualizada, las conclusiones que se obtienen de la investigación no se refieren a cada elemento individualmente, sino al conjunto de los elementos considerados como grupo.

- *Frecuente repetición.* Aquellos fenómenos que se presentan una vez y no vuelven a repetirse en las mismas condiciones, no son estudiados por la estadística. Por el contrario, se requiere que el fenómeno sea de frecuente repetición. Tal es el caso de las transacciones comerciales, inventarios, accidentes de trabajo, nacimientos, etc.

- *Distinta frecuencia.* Si observamos las ventas realizadas por un almacén, encontramos que se trata de un fenómeno de distinta frecuencia, porque de lunes a sábado se repite dicha operación, pero el valor de esas ventas varían de un día a otro, de semana a semana o de mes a mes, dependiendo del período examinado. Si el fenómeno tiene la misma frecuencia, no requerirá de la estadística, pues basta su registro una sola vez y se tendrá igual información en cada período.

- *Manifestarse al exterior.* El contrabando, el cultivo y comercio de la marihuana, las transacciones en dólares en el mercado negro, son fenómenos que diariamente se suceden pero no son registrados. Por lo tanto, se dice que no se manifiestan al exterior para ser observados estadísticamente.

- *Distantes en el tiempo.* Las ventas, en cierta cantidad, de determinados artículos estacionales, tales como pólvora, juguetes, artículos para celebrar el dia de la secretaria, la primera comunión o la recolección de un determinado producto agrícola, son fenómenos que no se suceden todos los días. Por el contrario, se producen en una determinada época o fecha y hay que esperar cierto tiempo para que se presente nuevamente.

- *Distantes en el espacio.* El pago de cheques por el *Banco Comercial,* es un fenómeno que podemos observar en diferentes sitios, en las sucursales que tiene en una ciudad, en todo el departamento o en todo el pais. Las exportaciones e importaciones que se realizan por los diferentes puertos del país. Son ejemplos que muestran al fenómeno originado en diferentes sitios, sin ser indispensable que se presente en un mismo lugar.

- *Cualitativos que pueden cuantificarse.* Algunas características cualitativas investigadas, requieren ser cuantificadas para que sean consideradas dentro del campo estadístico. Sin embargo no todas son cuantificables, tales como el grado de religiosidad, la moral y tantas otras y por tal razón quedan fuera de la acción estadística. Son ejemplos de hechos cualitativos,

los elementos clasificados por: sexo, estado civil, profesión, ocupación, primaria, secundaria, universitaria, por nacionalidad, mes de nacimiento, colores, etc.

- *Aquellos fenómenos que son accidentales,* tanto en el tiempo como en el espacio, no son propios de la investigación estadística, se presentan una sola vez y no vuelven a ocurrir.

ALGUNOS TÉRMINOS Y CONCEPTOS QUE SE DEBEN CONOCER Y MANEJAR

A continuación se definen en forma sencilla y fácil de entender, algunos términos que serán utilizados en el desarrollo de los diferentes capítulos, en los cuales se ha dividido este libro.

Vale la pena recalcar la importancia del conocimiento y el manejo de los términos que a continuación se expondrán, ya que nos permitirá hablar un solo lenguaje, precisar lo que se va a hacer y entender su uso en los diversos aspectos que conlleva el desarrollo de los temas del presente libro. Se recomienda leer detenidamente las primeras páginas de los capítulos 7 y 13, donde encontrará más información que le ayude a comprender mejor estos términos.

Población. Es un conjunto de medidas o el recuento de todos los elementos que presentan una característica común. El término población se usa para denotar el conjunto de elementos del cual se extrae la muestra.

Lo ideal sería, que el número de elementos o unidades de observación que constituyen la muestra, denominada también población por muestrear o *población* muestreada, fuera igual al contenido total en la *población* o *población objetivo.* Pero como no ocurre así, las conclusiones consideradas válidas para la muestra pueden ser extendidas a la población, garantizándolas estadísticamente mediante la indicación de la validez del proceso.

Los *elementos* o *unidades* que integran la población o la muestra pueden corresponder a personas objetos o cosas. Además, el *elemento* puede ser una entidad simple (una persona) o una entidad compleja (una familia) que se denomina *unidad*, la que, a su vez, puede estar compuesta de elementos o de unidades, que en algunos casos se les define como *conglomerados*.

Según sea el número de elementos, la población puede considerarse como *finita* o *infinita.* Es una población *finita* cuando el número de elementos que la compone es limitado; *infinita* cuando consta de ilimitado número de elementos. Sin embargo, algunas veces, una población finita con un número grande de elementos, el comportamiento en el cálculo del tamaño de muestra resulta igual al de una población infinita, ya que existe un tamaño muestral óptimo tal que, a partir de un determinado valor calculado, éste no se altera por más que aumente el tamaño poblacional.

Marco: se denomina *marco, marco muestral o marco de referencia*, a la lista, mapa o cualquier otro material aceptable, que contenga todas las *unidades* o *elementos* perfectamente identificados y actualizados, donde se selecciona la muestra.

Características: es importante resaltar el hecho de que a pesar de encontrarse una población constituida por un grupo de elementos o unidades, a la estadística no le interesa el elemento o la unidad en si, sino sus características.

Éstos son algunos ejemplos de lo que podrÍan denominarse elementos de una población y características de la misma:

ELEMENTOS	CARACTERÍSTICAS
Personas	Salarios
Personas	Horas trabajadas
Personas	Cargos
Departamentos	Ventas
Hogar	Consumo

Las *características* (o caracteres). Corresponden a ciertos rasgos, cualidades o propiedades que poseen los elementos que constituyen la población o la muestra.

Algunos caracteres son mensurables y se describen numéricamente; por tal motivo se denominan caracteres *cuantitativos* o *variables*, tales como: estatura, peso, ingreso, valor, producción, ventas, etc. Otros se expresan mediante *palabras* por no ser mensurables pero si cuantificables, tales como: profesión, cargo, marcas, calidad, etc. y se denominan *caracteres cualitativos o atributos*.

Las variables o aspectos cuantitativos se clasifican en: *variables discretas* y *continuas*. La primera se refiere a aquélla que sólo puede tomar unos determinados valores, siendo imposible que llegue a tomar valores intermedios entre dos consecutivos, es decir, toma únicamente valores enteros: 1, 2, 3 etc., tal es el caso del número de hijos por familia, nunca se tendrá 2 hijos y medio o dos hijos y cuarto, etc. En el segundo, se toman todos los valores infinitos posibles en un intervalo, es decir, se admiten valores fraccionarios, como el número de años de una persona: 20 años, tres meses, cinco días etc. La distinción que generalmente se hace entre variables *discretas* y *continuas* es más teórica que práctica, pues en la práctica se podrá decir que todas las variables son discretas, pero dada la facilidad de agruparlos en intervalos, como por ejemplo: edades entre 6 y 10 años, más de 10 y 20 años y así sucesivamente hacen realidad esa distinción.

Si analizamos o describimos una característica (atributo o variable) en forma independiente nos referiremos a *distribuciones unidimensionales* o *univariante*. Si relacionamos dos características entre si (variables; variable y atributo; o dos atributos) nos referimos a *distribuciones bidimensionales* o *bivariante*. En el caso de relacionar más de dos características se dirá que la *distribución* es *pluridimensional* o *multivariante*.

Investigación total. Denominada también como *censo* o *investigación exhaustiva*, es aquélla en la cual se toma la totalidad de los elementos o unidades que conforman la población objeto de estudio.

No siempre se puede realizar una investigación total, y debemos observar una parte de ellas, debido a circunstancias como las que se detallan a continuación:

- *Poblaciones* muy grandes o infinitas.
- *Tiempo* requerido demasiado grande.
- *Costos* tan elevados que sería imposible con los recursos disponibles.
- *Recursos humanos* con los que no se cuenta para el desarrollo de la investigación.
- *Destrucción* del elemento como ocurre en el control de calidad de un determinado producto, como por ejemplo, probar su resistencia o dureza.

- *Homogeneidad* de la característica, como sucede en la preparación de alimentos o en los exámenes de laboratorio.

Muestra o investigación parcial. Se define como un conjunto de medidas o el recuento de una parte de los elementos pertenecientes a la población. Los elementos se seleccionan aleatoriamente, es decir, todos los elementos que componen la población tienen la misma posibilidad de ser seleccionados.

Para que una muestra sea representativa de la población, se requiere que las unidades sean *seleccionadas al azar,* ya sea utilizando métodos tales como: el sorteo, *las tablas de números aleatorios, la selección sistemática o cualquier otro método que sea al azar.*

Una muestra *no probabilística* corresponde a la selección de las unidades o elementos, según el criterio o conveniencia del investigador. En estos casos, algunas unidades tienen mayor posibilidad que otras de ser seleccionadas; por tal razón no se puede determinar la validez, ni la confianza que merecen dichos resultados. Casi por lo general, las unidades se seleccionan en forma *caprichosa, voluntaria*, por cuotas, tal como sucede en las encuestas de opinión, que en la mayoría de los casos no tienen credibilidad, de ahi que se le conozca como *Muestra circunstancial o errática.*

Podemos concluir, en forma resumida, mediante un esquema, que la observación de una población se puede hacer de diferentes maneras, tal como se presenta a continuación. Este tema se amplia en los capítulos 7 y 13, por lo tanto se recomienda consultarlo.

Parámetros. Son todas aquellas medidas que describen numéricamente la característica de una *población.* También se les denomina *valor verdadero*, ya que una característica poblacional tendrá un solo parámetro (media, varianza, etc.). Sin embargo, una población puede tener varias características y, por tanto, varios parámetros. Algunos lo denominan como *valor estadístico de la población.*

Estimador o *Estadígrafo*. La descripción numérica de una característica correspondiente a los elementos de una *muestra*, se denomina *estadígrafo*. De una población se pueden obtener (M) número de muestras posibles y en cada una de ellas se puede cuantificar la característica, obteniéndose por lo general, valores diferentes para cada muestra, a pesar de ser utilizado el mismo estadígrafo o medida.

También se le conoce como *Estimador Puntual* si se trata de un *promedio, varianza, proporció*n, etc. Como por lo general, existe una diferencia entre el estimador y el parámetro, denominado *error*, es aconsejable utilizar el *estimador por intervalos,* dentro del cual deberá estar contenido el parámetro con cierto margen de seguridad, establecido por el investigador.

La diferencia entre el *estimador* y el *parámetro*, como ya se dijo, se le denomina como *error de muestreo,* y lo fija el investigador para determinar el tamaño de la muestra, con base en la experiencia y el conocimiento que tenga sobre el comportamiento y la variación que tenga la característica o características principales relacionados con el objetivo mismo de la investigación.

Errores ajenos al muestreo: no son cuantificables (como el anterior) y son resultados de errores sistemáticos que se producen a lo largo de la investigación, como: en la preparación, formulación, recopilación presentación, entre otras.

LA PROYECCIÓN Y PREPARACIÓN DE INVESTIGACIONES ESTADÍSTICAS

En nuestro diario quehacer profesional, en la mayoría de las casos, nos corresponde analizar información estadística para la toma de decisión en el desarrollo de nuestro trabajo y gran parte del libro trata de colaborar con el lector, presentando técnicas que les permitan hacerlo, en gran manera acertada; pero otras veces, no pocas, hay necesidad de realizar la investigación a fin de obtener los datos necesarios.

El método estadístico permite recolectar u observar mediante registros, que se ordenan, clasifican, cuantifican y presentan mediante cuadros, gráficas, permitiendo la depuración de datos en dos aspectos, en su presentación clara y en su simplificación, ya que de esta manera facilita manejar grandes cantidades de información presentadas en forma resumida, tarea que realizamos a través de la estadística descriptiva; luego la inferencia estadística permite el análisis y la búsqueda de relaciones que puede existir entre ellas ya que trabajamos sobre abstracciones, situación muy diferente comparada con el *método experimental* que descansa sobre realidades físicas.

Algunos problemas que nos planteamos en la vida diaria, tendrán contestación a través del análisis de los datos que se han registrado, a medida que se producen en un estricto orden cronológico: precios, ventas, accidentes, exportaciones, inventarios al finalizar el año, etc. Cuando no existen esos registros es necesario la realización de encuestas, ya sea tomando toda la población (censo) o una parte de ella (muestra) para deducir el comportamiento de las características de la totalidad de la población. En este caso podemos considerar algunas etapas, no necesariamente todas, que deben ser tenidas en cuenta para su desarrollo.

El proceso de elaboración estadística lo dividimos en seis (6) fases:

A. Planeamiento o preparación
B. Recolección
C. Crítica y codificación
D. Procesamiento
E. Análisis e interpretación (numérico y gráfico)
F. Publicación

PLANEAMIENTO Y PREPARACIÓN DE UNA INVESTIGACIÓN DE TIPO ESTADÍSTICO

Toda recolección de datos estadísticos de un hecho, requiere el establecimiento previo de un plan que detalle los aspectos que la investigación va a abarcar, que fije los procedimientos a seguir, y que resuelva de antemano las posibles dificultades que se presume hayan de presentarse.

La implantación de una nueva información estadística en una empresa privada u oficial o en un caso particular, exige el estudio y la redacción de un proyecto al respecto.

En general, un proyecto o plan de investigación estadística debe contemplar los siguientes aspectos:

1) El objeto de la investigación
2) La finalidad
3) La fuente de información
4) Los procedimientos
5) El material
6) El costo y su financiación

El objeto de la investigación

Con el término de objeto de la investigación, nos referimos al hecho o fenómeno que vamos a observar y registrar numéricamente. En una investigación de salarios: el salario; en un censo de población: los habitantes; en una estadística industrial: la producción, en una información de créditos: las operaciones crediticias.

En el proyecto, al analizar qué es o qué vamos a investigar, nos proponemos:

a) Definir el objeto o finalidad de la investigación
b) Determinar su naturaleza cuantitativa y cualitativa
c) Determinar la posibilidad de su investigación, y
d) Limitar el objeto investigable

El proyecto de una investigación estadística tiene que atender los fines que se persiguen con la investigación, porque así como sean éstos, habrán de ser los demás puntos del proyecto: la fuente de información, los procedimientos, el material y en consecuencia última, el presupuesto.

Dos ejemplos, el uno bioestadístico y el otro económico, aclaran la necesidad de establecer, a priori, el punto de mira que se persigue con la recopilación estadística.

Se organiza una estadística de natalidad. Se puede aspirar a muy diferentes fines. Por ejemplo, se trata de conocer la proporción de los sexos con un propósito biológico, o si no (y este es otro fin), con un propósito demográfico de la población que entra a la vida. En el supuesto, el fin nos obliga a preguntar por el sexo de los hijos nacidos, muertos o vivos; en segundo caso, nos bastará con indagar los nacidos vivos.

Se prospecta una estadística de salarios. El objetivo de esta investigación, entre otros, puede ser, calcular el salario medio general simplemente, o bien obtener el salario más frecuente en cada uno de los oficios de cierta rama industrial.

a) Definir el objeto o finalidad de la investigación: vamos a emprender un censo de edificios: ¿qué es un edificio para efectos de la recolección de los datos? En una investigación de salarios: ¿a qué salarios nos referimos? En una investigación industrial, ¿qué hemos de entender como "industria", como "empresa", como "fábrica"?

   ¿Indagaremos cualquier industria, toda clase de salarios, toda especie de edificios? o por el contrario, exigiremos determinados requisitos particulares? Lo anterior nos permite identificar *A quién* se investiga o sea el elemento o la unidad. Veamos algunas formas de definirlas:

   Unidad de vivienda: "Es la unidad final considerada como una estructura separada e independiente destinada para alojar una o más personas, con acceso directo a la calle, o a un pasillo o escalera, de tal forma que sus ocupantes puedan entrar o salir sin pasar por los cuartos de otras unidades de vivienda".

Hogar: "Constituye la unidad final de observación y se define como una o más personas bajo la autoridad de un sólo jefe que ocupan en común la totalidad o parte de una vivienda y alimentos, a menos que circunstancias especiales lo impidan (por ejemplo el horario de trabajo) y no se aprovisionen conjuntamente de otros bienes y servicios de primera necesidad, sin interesar el parentesco entre ellos".

La primera cuestión por tanto, al tratar el objeto de la investigación en una prospectación estadística, es la fijación precisa del concepto de lo que se aspira a indagar. Decir con claridad y exactitud lo que la Estadística va a recoger. La unidad o elemento de investigación debe ser: *clara, adecuada, mensurable y comparable.*

b) Determinar la naturaleza cuantitativa y cualitativa del objeto de la investigación: o sea, establecer los aspectos cuantitativos del fenómeno que se pretende investigar y consecuencialmente pueden registrarse por medio de números, y otros que por ser cualitativos se han de recoger mediante anotaciones literarias, bien de su existencia o de su ausencia (si, no), bien de su clase o categoría (1a, 2a, 3a), su origen (importación de Estados Unidos, Inglaterra, Argentina), su destino (créditos a la agricultura, a la ganadería, a la industria manufacturera), su marca (café brasilero, colombiano, mexicano), su grado de bondad (bueno, regular, malo), etc.

c) *Determinar la posibilidad de investigación*: de los aspectos establecidos en el considerando anterior, no basta conocer el objeto a investigar, y sus modalidades cuantitativas y cualitativas; se necesita examinar si una y otras pueden ser conocidas con precisión, si se exteriorizan, si pueden contarse, si admiten su existencia y su intensidad.

En un proyecto de estadística sobre la morbilidad de una población, ,cómo comprobaremos la presencia o la falta de la enfermedad?; el ahorro popular no depositado en cuentas de ahorros sino atesorado personalmente, ¿cómo podremos conocerlo para registrarlo estadísticamente?; el alcoholismo en un barrio obrero, divisas extranjeras en las transacciones en la bolsa negra, la declaración de las edades, los consumos familiares de víveres, los precios de las mercancías de importación, se exteriorizan de alguna manera, constan en algún sitio, como para poder anotar el hecho en su cantidad o en su calidad o en ambos aspectos?

d) *Limitar el objeto investigable:* conocemos lo que vamos a investigar, sus modalidades cualitativas y cuantitativas y la posibilidad de registrarlas. A continuación, el proyecto tiene que delimitar la investigación. Todo objeto a investigar en si, en el tiempo y en el espacio y a veces en el número, debe tener sus méritos.

Por imposibilidad práctica o por ser innecesaria la observación completa, la estadística reduce sus trabajos a un doble aspecto:

Primero: Limitando el objeto mismo de la investigación y

Segundo: Limitando el campo de la investigación

Limita el propio hecho o fenómeno que investiga, cuando se concreta a ciertos aspectos o modalidades, prescindiendo de todos los demás. Y ello por no serle posible abarcar todos los caracteres del fenómeno.

Estadística de salarios. Limita a las tasas de salarios en dinero, en los hombres de 18 años o más, en la industria de transformación. Se prescinde de los salarios percibidos por el obrero en trabajo extra, por trabajos reducidos, primas de días feriados, bonificaciones, etc.; se prescinde de las tasas de salarios en la agricultura y ganaderas; en la minería y en los servicios se hace caso omiso de los salarios femeninos y de los masculinos, en los trabajadores menores de 18 años.

Estadística de créditos bancarios. Se limita a los préstamos y descuentos comerciales e industriales que conceden los bancos. No se recogen los préstamos hipotecarios o de "amortización gradual", se pregunta la clase de actividad económica, a que se destina, sobre si son personales o con garantía prendaria, cuantía, interés o descuento y sus plazos de devolución. Pero se dejan de indagar muchísimos otros aspectos de la empresa a la que se le concede el préstamo, del particular prestatario, del destino geográfico, del capital prestado, etc. Al admitir como necesarias unas características del hecho que vamos a investigar y rechazar otras, el proyecto está limitando al objeto mismo de la investigación: los créditos bancarios, los salarios, en los ejemplos propuestos.

La *Estadística* puede interesarse por conocer la producción algodonera del país, pero no necesita ni le es posible llegar al detalle de los kilogramos recogidos en cada hacienda o región, le bastará con estimar aproximadamente las toneladas o incluso los miles de toneladas. La *Estadística* nos proporciona el dato sobre la cotización de las monedas extranjeras hasta las milésimas o diezmilésimas del peso; aquí la aproximación no puede ser mayor del tercero o cuarto decimal de la unidad monetaria.

Según la clase de investigación de que se trate, la exactitud podrá ser mayor o menor, y ésta deberá fijarse en el Proyecto de Investigación. Es una forma de limitar el objeto de la recolección estadística.

Estudiamos ahora la limitación del campo de la investigación. Comprende cuatro aspectos: el *espacio*, el *tiempo*, el *número de casos* y la *combinación de los tres anteriores*.

Limitación del espacio o cobertura geográfica. Toda estadística nacional presupone una limitación geográfica de orden internacional. Una limitación internacional, e incluso más de una estadística municipal, está circunscrita a determinados barrios del área urbana del municipio. Los censos nacionales, por su propia naturaleza, no están limitados en el espacio nacional. Pero, por ejemplo, la estadística sobre los índices de precios al consumidor, sólo se realiza en un reducido número de ciudades.

Como siempre, la limitación de este género se efectúa por falta de recursos para recoger los datos en todos los lugares, o por ser innecesaria la recopilación completa para efectos de la observación y conocimiento suficiente del fenómeno que investiga la estadística. El proyecto ha de concretarse expresamente los lugares que va a contemplar la investigación.

Limitación del tiempo. También el tiempo se debe establecer; no todas las investigaciones pueden ser continuas; algunas se repiten, son periódicas; otras se celebran una vez en forma accidental y no vuelven a realizarse. En el proyecto se fija la continuidad, la periodicidad o la accidentalidad de la investigación. En los dos últimos casos se trata de una limitación del tiempo.

El censo de población, como los censos industriales o agrícolas no se celebran todos los días, ni todos los meses ni siquiera es necesario que sean anuales, basta y así está convenido internacionalmente, que sean decenales.

La criminalidad, por el contrario, tiene lugar en cualquier momento con características muy variadas y su registro policial por otra parte es imprescindible. La estadística de delitos, muy bien, podrá ser continua. La estadística de mortalidad, como otras demográficas, deberá ser permanente.

Esta limitación cronológica atemporal, se refiere a lo que en técnica estadística se denomina *"el momento estadístico";* la época, la fecha de una investigación.

Esta limitación, en la mayor o menor periodicidad, depende de la variabilidad del fenómeno que se investiga. A mayores modificaciones en el fenómeno, más breve el tiempo transcurrido entre una y otra investigación; los hechos de naturaleza estática se investigan muy de tiempo (estado de las edificaciones, higiénico, sanitario de la población, características de la industria, de la ganadería, de las explotaciones agrícolas, mineras, de la enseñanza); los hechos de grandes fluctuaciones se investigan en cortos intervalos de tiempo (cotizaciones de bolsa, precios de víveres, nacimientos, producción industrial, comercio exterior, costo de vida, ingresos y egresos del Estado, etc.).

El proyecto dirá el período o el carácter continuo o accidental de la investigación.

Limitación del número de casos. La observación, a veces, puede abarcar todos los casos que se presentan; en otras ocasiones no es posible, y por lo tanto, hay que limitar su número. Los cálculos y estimaciones estadísticas parten generalmente del estudio de un número de observaciones. Así se precede comúnmente en las investigaciones científicas de tipo médico, antropológico, económico, sanitario, social, etc.

En la recolección de precios, no se hace de todos los artículos que se ofrecen y demandan en el mercado, sino de un número parcial de ellos; en la estadística de los salarios se investiga a un grupo de obreros o empleados.

El proyecto también tiene que fijar el número exacto o aproximado de casos que deben indagarse.

Limitación coordinada del espacio, del tiempo y del número de casos: por último y es lo que más frecuentemente sucede, en un proyecto de investigación estadística hay que limitar a la vez, el espacio y el tiempo, aquél y el número de casos, éstos y el tiempo, o los tres juntos.

Cuando en un proyecto se recomienda cualquiera de esas reducciones de la investigación, se plantea el problema de la elección de la época, del lugar y de las propias unidades que sometemos al examen estadístico.

Sólo existe una norma general, para efectuar esas elecciones: los casos, los lugares y las épocas designadas deberán ser normales. No deberán encerrar características, aspectos excesivamente intensos o excesivamente débiles del hecho o fenómeno de que se trate.

Los procedimientos prácticos son dos:

Elegir al azar sin buscar intencionadamente los casos extremos, los lugares o las fechas en que el fenómeno presente una situación anormal.

Si se pretende realizar un análisis sobre las condiciones del trabajo en la industria manufacturera de un país, y decidimos la investigación, por ejemplo, a doscientas veinte fábricas y talleres, diez de cada departamento, efectuaremos una elección al azar; asi del Directorio de la Industria Manufacturera entresacamos de cada departamento las diez primeras empresas fabriles según el orden alfabético de sus razones sociales; probablemente, entrarán pequeños talleres y grandes instalaciones, producciones muy diversas, condiciones de trabajo distintas, formas de asociación de los propietarios, en fin, las más variadas circunstancias y no empresas de igual naturaleza.

Elegir por casos extremos, en los cuales el hecho social económico tiene intensidades máximas y mínimas con el objeto de buscar lo normal en el punto medio de los límites extremos.

En el ejemplo anterior se pueden recoger las informaciones de la gran industria por un lado, y de la pequeña por otro. De esta forma conoceremos los casos y situaciones extremas, y podremos de ellos deducir los valores y situaciones intermedias. Es una manera indirecta de hallar lo normal o típico partiendo de lo anormal o atípico.

Las fuentes de información

Hasta ahora hemos analizado lo que se desea investigar y el por qué de esa investigación.

A continuación, es necesario preguntarse: ¿dónde es posible recoger alguna manifestación externa del fenómeno económico?, ¿en dónde podremos registrar datos sobre los saldos de moneda depositados en los bancos?, ¿en dónde cifras sobre importaciones?, ¿en dónde información sobre precios de víveres, etc.?

El lugar, el documento, la oficina, la persona que constituye la fuente informativa podrá ser:

❑ Directa e Indirecta

Investigaciones directas e indirectas. Directamente recogemos datos de un acontecimiento de cualquier índole, cuando acudimos a él, lo observamos y anotamos su presencia o su ausencia y su intensidad mediante un número. Nos proponemos conocer el canje de cheques entre los bancos, acumulamos los cheques, los contamos en su número, sumamos su valor y lo anotamos. Hemos observado y anotado directamente el mismo hecho cuyo conocimiento deseábamos.

Indirectamente investigamos en estadística, cuando se recurre a un hecho distinto del que nos interesa, para después deducir de éste el valor del que en definitiva deseábamos conocer. En estadística agrícola frecuentemente se emplea este procedimiento. Del rendimiento de determinada hectárea se deduce la producción total. De la cosecha probable se deduce la siembra.

De la transmisión de herencias durante la vida media de un hombre en un país, el capital y la renta nacional; de los consumos, la población; de los precios del café hemos visto deducir estimativamente la importación y el consumo de determinados artículos suntuarios. Son inducciones lógicas, cálculos aproximados, estimaciones que constantemente se realizan en los negocios.

Se da el caso de que las cifras obtenidas indirectamente son más necesarias y frecuentes en los países de pequeño desarrollo estadístico por la misma falta de muchos renglones desconocidos. Y su proyección y aproximación suficiente depende precisamente de la abundancia, extensión y calidad de las estadísticas directas. Puesto que la garantía de esta clase de cálculos se funda en varias pruebas por varias estadísticas directas.

Se refieren sobre todo, a la investigación de elevado costo que directamente no se puede realizar, las que por su naturaleza reservada, consumos de bebidas alcohólicas, vicios sociales, la pregunta directa se hace difícil.

Muchas veces, se estima y otras se calcula. Se estima subjetivamente cuando por ejemplo, un agrónomo tasa la cosecha, valora las tierras. Se calcula objetivamente cuando, por ejemplo, el demógrafo nos da la cifra de habitantes en una fecha intermedia a dos censos directamente obtenidos; cuando los consumos de cierta población se deducen de los consumos directamente investigados en un grupo parcial

de consumidores; cuando de las necesidades de materias primas de las grandes empresas se calcula la cantidad necesaria para todo un pais; por generalización de una parte al todo.

También, en caso de pasar a la intensidad de un hecho partiendo de la intensidad de otro; el consumo de carne y leche a la población; de la importación de autos, a las necesidades de llantas; de la importación y exportación de libros, a la frecuencia de la lectura; del impuesto a la cerveza al consumo; de la pluviosidad a los precios, etc. Para que este cálculo dé resultados suficientemente buenos, se requiere imprescindiblemente que entre el hecho conocido y el desconocido exista una simple relación, cierta proporcionalidad conocida o una dependencia de causa a efecto, de ésta a aquélla.

Llamamos fuente de información estadística ***directa*** allí donde el hecho económico se produce. Son ejemplos, el hogar de la familia obrera, para investigar sus consumos y el costo de su vivir; los peajes, para obtener los datos de la circulación por carreteras; en las notarías para determinar el número de nacimientos, defunciones, matrimonios y otras informaciones estadísticas; los balances comerciales, para conocer los resultados de los ejercicios semestrales; el taller, para registrar el tiempo empleado en cada operación productora.

Son fuentes de información estadística ***indirecta,*** aquéllas donde el hecho se manifiesta indirectamente, en donde se refleja. Las rentas departamentales para determinar el consumo de bebidas alcohólicas; la policía para indagar datos sobre criminalidad; las listas o nóminas de salarios para obtener información de las cesantías de éstos; el degüello del ganado para estimar la población consumidora; los manifiestos de importación y exportación para averiguar el consumo interno.

En el proyecto hay que fijar concretamente la ***fuente*** o ***fuentes*** de información que conviene o se necesita adoptar.

Las mejores son las ***fuentes directas***. Pero no siempre son posibles. Una estadística de los delitos contra la propiedad no podría organizarse de forma que el registro estadístico se efectúe en el lugar del robo. Necesariamente la fuente informativa ha de ser la comisaría o inspección de policía, donde el hecho se registra.

De aquí nace la desconfianza sobre la estadística, por la diferencia en los registros: una persona que ha sido atracada, violada, atropellada, etc., cree que denunciando el hecho no le representa ningún beneficio, sólo perderá tiempo, de ahí que los resultados publicados no muestran la magnitud del problema, sino parte del mismo, lo cual origina la idea de que las estadísticas son mentirosas, en realidad se debe decir, que son deficientes, debido al no registro de lo que sucede.

Cuando sea posible, debe emplearse una fuente directa. Cuando no, se empleará una indirecta. Frecuentemente éstas se usan como complementarias de las primeras, y a efecto de establecer un control sobre las directas.

Algunos autores clasifican la fuente de información como ***Primaria***, cuando se obtiene directamente de la investigación, realizada a través de una encuesta, y ***Secundaria***, cuando se trata de información complementaria, publicada por la misma institución o cualquier otra.

También la clasifican en ***Interna*** cuando la información se produce dentro de la institución, ***Externa*** en el caso de ser obtenida por fuera.

La anterior clasificación puede dar lugar a confusiones en cuanto a la denominación de directa o indirecta.

Los procedimientos de investigación

¿Qué características debe reunir la investigación? ¿bajo qué formas haremos la recolección de los datos?

En este aspecto del proyecto, comúnmente los autores incluyen varios considerandos que se ajusta a estos interrogantes: ¿cómo investigaremos?, ¿quién debe hacer la investigación y por qué sistema?

¿Cómo debe realizarse una investigación? Es decir, ¿cuáles deben ser las normas generales y particulares de la estadística que proyectamos?

Suelen resumirse en estos puntos:

- *Claridad y publicidad*. Toda la investigación, en todos sus pormenores, debe ser clara y conocida por todos los que vayan a intervenir: observadores y observados.

  La claridad se hermana con la comprensión. La estadística, primero en el período de investigación, luego en el de elaboración y más tarde en presentación y el análisis, debe ser entendida por todo el mundo: investigador, investigado, empleado que elabora los datos y el público que va a leerlos y quiera utilizarlos.

- *Sencillez*. Sencillez en todo. En los formularios, en las instrucciones, en el proyecto, en la finalidad, en los cuadros de exposición, en los comentarios literarios de análisis, en las operaciones de cálculo, en los gráficos. La sencillez es conducción esencial de la claridad; difícilmente se dará aquélla sin ésta.

- *Utilidad*. Toda estadística que se inicie debe tener alguna aplicación práctica de interés. De otro modo nos hallaremos ante una acumulación molesta de datos que perjudicarán, por muchos motivos, a la información realmente aprovechable, a los negocios y a la política económica de un país.

¿Cuándo deberá efectuarse la recopilación estadística? En el denominado "Momento Estadístico" Una estadística podrá ser *continua*, *periódica* o *aislada*. En el primer caso, nada tenemos que decir respecto del momento estadístico apropiado, porque son todos. Es el caso de la estadística de criminalidad, constantemente funcionan las comisarías de policía recogiendo los datos sobre los delitos que vayan denunciándose.

Pero la mayoría de las estadísticas económicas y demográficas, son más o menos periódicas, se producen en intervalos grandes o pequeños, según el hecho de que se trate; es la cuestión del tiempo limitado. Se le denomina como *series cronológicas o series de tiempo*.

Ahora se nos presenta el problema de elegir ese tiempo; el período, la duración de la investigación aislada y la fecha de ésta o fechas de aquéllos.

Las investigaciones pueden ser, según la clasificación de Conrado Gini y de otros autores: *ocasional, periódica* y *continua*.

- Es *ocasional* una recolección de datos en circunstancias extraordinarias, cuando eventualmente se presenta un problema, o se agita su solución. Una investigación del *costo de la vida* o de los

salarios cuando se plantea una huelga; cuando se denuncia el incumplimiento de la legislación laboral; al discutirse reforma de la tributación; del proteccionismo aduanero, etc. Se ejecuta la indagación numérica en esa ocasión y luego no se vuelve a efectuar. Son investigaciones *intencionales* o *aisladas*.

- Es *periódica* aquella investigación que se repite de tiempo en tiempo, en lapsos regulares. Los censos en períodos decenales; la estadística industrial, con periodicidad anual; la del comercio exterior, mensual; los precios de los víveres, semanal; semestral, la estadística de los dividendos de las sociedades anónimas; diaria, la información de los dividendos de las sociedades anónimas, la información de los precios del café, del ganado sacrificado en el matadero de Bogotá, D.C. y de las cotizaciones de las divisas extranjeras; dos veces al día, las estadísticas de cotizaciones y operaciones en la Bolsa.

- Es *continua* la investigación automática, es decir, la que se produce sin interrupción. Las informaciones demográficas (la natalidad, la mortalidad, la nupcialidad); las de criminalidad, las del tráfico por carreteras y ríos, las de los transportes férreos, de ciertos espectáculos (cines), de artículos de consumo.

- Es registro permanente, a medida que el hecho tiene lugar.

Son normas generales, puesto que particulares no se pueden señalar, sino en la simple discusión concreta de cada estadística, son:

- Periodicidad y duración suficiente para observar el hecho o fenómeno.

- Fechas normales, cuando el hecho o fenómeno presente características habituales, corrientes o cuando no sea ni muy intenso ni muy débil.

Por ejemplo, una investigación sobre los ingresos y gastos familiares en Europa tiene que durar todo el año porque las cuatro estaciones climáticas afectan el presupuesto del hogar, cuatro características consumidoras y cuatro gastos distintos; en el trópico se considera suficiente un mes de duración, porque la variabilidad de las compras es pequeña en los cuatro trimestres del año.

Los fenómenos que cambian mucho y en cortos lapsos, exigen una periodicidad de investigación grande, las cotizaciones en la bolsa se registran mañana y tarde. Las temperaturas y demás condiciones meteorológicas, tres y cuatro veces al día, cuando no continuamente. Pero los precios de las mercancías al por mayor suelen investigarse semanalmente. La estadística bancaria se publica de un mes a otro, como la de la producción minera y del comercio exterior. Las modificaciones entre dos semanas o entre dos días no son tan grandes como para anotarlas.

Los datos sobre el estado de desarrollo de la educación se conocen anualmente. En períodos semestrales los balances de las sociedades anónimas. Decenales, los censos de población.

Las estadísticas *discontinuas*, pueden tener una mayor o menor periodicidad en función de su mayor o menor variabilidad de fenómenos que se alteran mucho, frecuentes períodos; en los que se alcanzan a notar cambios muy de tarde en tarde, períodos bastante separados. Pero además, el "momento estadístico" encierra la elección concreta de la fecha o fechas. Y decimos que ésta debe ser un período normal. Investigar la ocupación obrera, por ejemplo, cuando la actividad del trabajo se halle en su grado habitual y no cuando por un motivo extraordinario y pasajero haya un número excesivo de desocupados o una actividad también excesiva.

El costo de vida no podría investigarse únicamente durante el mes de diciembre o el de enero, porque el último mes del año por exceso de gastos y el primero del año por defecto de gastos, serían notablemente anormales.

Ni el censo de población podría levantarse el 24 o 25 de diciembre, porque un núcleo no despreciable de gentes, se halla en esas fechas en un lugar fuera de su habitual o legal de residencia o domicilio.

Por último, no hay que olvidar en el proyecto de investigación el período preliminar de preparación, que proceda a la recopilación de los datos. Su iniciación y duración deben determinarse de antemano. Por tal motivo se recomienda la elaboración de un calendario de trabajo que incluya a cada una de las diferentes etapas de investigación con sus fechas respectivas de iniciación y terminación.

Todos estos elementos intervienen en los puntos siguientes: el personal, el material, la propaganda y el costo.

¿Quién debe hacer la investigación? La pregunta de quién debe hacer la investigación comprende un doble aspecto: primero la entidad u organismo, segundo el personal. Ambas hay que fijarlas en el proyecto.

Entidad: pueden ser instituciones u organismos públicos oficiales o privados, particulares, de los gobiernos o de las personas o asociaciones particulares.

Ambos tienen ventajas y ambos, unos frente a otros, tienen inconvenientes, veamos:

Principales desventajas de las entidades públicas:

- Infunden recelo y suspicacias sobre la finalidad que los gobiernos persiguen con la estadística: temor fiscal, temor policial, etc.
- Frecuentemente el costo de las investigaciones es más elevado que en las entidades privadas.

Principales ventajas de las entidades privadas:

- Son más indicadas para las encuestas ocasionales de tipo científico, problemas sociales, médicos, higiénicos, etc.
- Ofrecen al público más confianza que los gobiernos: los datos se conceden con mayor sinceridad.

Los inconvenientes de las entidades privadas corresponden a las ventajas de las instituciones públicas: menos recursos económicos, sobre todo en los países de escaso desarrollo de las investigaciones científicas; más reducida organización administrativo-territorial y falta de poder coactivo.

En las naciones de gran proceso estadístico, como Estados Unidos, Francia, Italia, Inglaterra y Alemania, son numerosas las entidades privadas que realizan investigaciones de esta clase. En América del Centro y del Sur, todavía es deficiente el desarrollo de la investigación no oficial.

Desde el punto de vista de la Administración Pública no es discutible la conveniencia y la necesidad de que los gobiernos dispongan de un servicio de Estadística Nacional y Seccional; las estadísticas generales o parciales de los países sólo pueden llevarse por los gobiernos, máxime teniendo en cuenta que para el público las informaciones deben ser gratuitas y generalmente conocidas.

En la actualidad se vienen realizando encuestas a través de los medios escritos (periódicos, revistas), radiales y de TV., algunos mediante entrevistas, otros en forma voluntaria con llamadas al medio de información o escritas, siendo de mayor uso del Internet, telefónicas o del celular, sin embargo los resultados son poco confiables, dado que no hay selección al azar del informante, procedimiento válido para una muestra aleatoria.

Personal. También en el proyecto tiene que tratarse el personal que ha de ser reclutado para la investigación. Se fijará el número, las atribuciones y las categorías de los distintos puestos. Alrededor del tema del personal, plantea el problema delicadísimo de su elección, de su instrucción y de su remuneración.

Nos parece conveniente hacer estas advertencias al respecto:

- La instrucción, requiere cursos de preparación, sobre todo en donde la cultura y formación estadística no sean grandes.

- La elección debe atender a la instrucción, puede ser mediante examen previo, de aptitud y vocación y a la imparcialidad de los aspirantes al puesto de estadística, es decir, que el empleado sea ajeno a los intereses particulares que puedan entrar en juego en las informaciones estadísticas.

Aparte de eso, todo el personal que trabaje en estadística debe ser constantemente controlado en sus tareas, sobre todo cuando éstas se refieren a cálculos, a clasificaciones, recuentos y construcción de tablas. El control también debe existir para el personal de recolectores, con el fin de comprobar si los datos que ellos recogen son ciertos o no.

Desgraciadamente el personal que se vincula, lo hace en forma transitoria, sólo quiere resolver su situación laboral por un tiempo menor, mientras consiga mejores condiciones salariales y de estabilidad.

Sistemas de investigación: se distinguen varios procedimientos de investigación:

- Las recopilaciones *automáticas* de datos por declaración espontánea del sujeto de la investigación. Se refiere generalmente a las investigaciones por medio de registros o inscripciones obligatorias; como en el caso de las estadísticas demográficas (natalidad, nupcialidad, mortalidad, migración). Su carácter automático estriba en el hecho de producirse en cumplimiento de una disposición legal o de una costumbre, de una tradición. La estadística del movimiento de divisas extranjeras, por ejemplo, es automática, porque se forma en virtud de la obligación de solicitar la compra de monedas extranjeras y acudir a su cambio o venta a la Oficina de Control de Cambios del Banco del Estado. La estadística del comercio exterior, la de las edificaciones, la estadística de los recaudos gubernamentales por concepto de los impuestos, son otros ejemplos de estadística automática, que automáticamente y sin necesidad de buscarlo, se producen mediante la anotación de los hechos en oficinas administrativas: Recaudación de impuestos, Aduanas, Banco del Estado, Secretarías de Obras Públicas Departamentales, etc.

  Es un buen sistema para recoger informaciones de esta clase, puesto que la obligatoriedad legal o la fuerza de la costumbre influyen en la exactitud sinceridad y generalización de las informaciones.

- Las recopilaciones *intencionales* de datos, obtenidas mediante el empleo de un agente que ex profeso vaya a la fuente de información para registrar los datos: el censo de la población, las encuestas de hogares sobre ingresos y gastos, las encuestas sobre sistemas de trabajo, sobre condiciones de una determinada industria, etc. En ellas se utilizan los cuestionarios, boletines, formularios, cédulas o patrones de inscripción de los datos.

La organización y establecimiento de este sistema de compilar datos se efectúa con una finalidad propiamente estadística, a diferencia de las recopilaciones automáticas en las cuales el motivo de la declaración no es propiamente estadístico. El registro civil, demográfico, la inscripción catastral, de la propiedad, de las sociedades comerciales, etc. La estadística, en estos casos, lo que hace es adjudicar a los registros de otra naturaleza la obtención de estadísticas, aprovechando los servicios del Estado.

Las investigaciones sistemáticas, por ser exclusivamente estadísticas, ofrecen las ventajas de poder organizarse conforme a una técnica completa y abarcar todos los aspectos que requieren las necesidades del servicio de estadística.

- Las recopilaciones *combinadas automáticas e intencionales*. En ellas se añaden a las informaciones del servicio ajeno al de la estadística, las preguntas que exige una buena investigación estadística. En el registro eclesiástico se anotan automáticamente los matrimonios y los nacimientos, se aprovecha esta circunstancia para poder indagar algunas cuestiones que al registro de la Iglesia no le interesarían.

- Algunos estadísticos, clasifican las investigaciones, en *extensivas* e *intensivas* y éstas en *monográficas y encuestas*. La diferencia se basa en el número de datos que abarcan y en los detalles a que llegan.

Cuando la recopilación se limita a las preguntas esenciales, a los datos muy generales, tal como ocurre en los censos, en la estadística bancaria, en la del comercio exterior y en la fiscal; y su extensión es completa por proceder de todo el campo de la investigación: se denomina *extensiva*. La mayoría de las recolecciones de datos periódicos y continuos son de esta naturaleza.

Cuando, por el contrario, la investigación se detiene en aspectos muy particulares del hecho, se hacen muchas preguntas y muy minuciosas y cuando se reduce a un sector del campo de la investigación, se le llama *intensiva*. La mayoría de las investigaciones ocasionales son intensivas.

Ejemplos reales de las primeras son los censos, la estadística de precios y salarios, las monetarias, entre muchas otras. Ejemplos de la segunda, intensivas, son las encuestas campesinas, de costo de vida, de alimentación, etc.

- *Investigaciones completas, incompletas y más que completas,* continuamos la división didáctica del profesor Conrado Gini, división necesaria a efectos de la enseñanza del método estadístico, y más que nada para poder efectuar una crítica y corrección técnica.

Son *completas*, las investigaciones que recogen todos los casos. Indagan todo el campo de la observación; cuantas unidades de las operaciones bancarias, (es otro ejemplo de información completa), entran en todos los balances de la banca. La producción de sal, la de cemento, los transportes aéreos, la estadística fiscal, son otros tantos ejemplos de estadísticas completas; se tomarían todas las fábricas de cemento, todos los aeropuertos y todas las oficinas que tenga el Ministerio de Hacienda.

Son *incompletas*, las investigaciones que sólo atienden a una parte de las unidades estadísticas, bien por no ser posible recoger la totalidad de los datos, o bien por no ser necesario para el fin que se persigue. Si la estadística incompleta no es representativa del conjunto, no es típica como para generalizar los resultados parciales al conjunto de los casos; su valor es nulo, o bien pequeño. En caso contrario, cuando el circulo estudiado numéricamente puede sustituir al total, la estadística incompleta es de extraordinaria utilidad.

Son *más que completas*, las investigaciones que repiten varias veces una misma anotación, a veces intencionalmente como control, otras inadvertidamente, dando origen a un error.

Pero no faltan ocasiones en las cuales el estadístico voluntariamente hace una misma pregunta en un formulario de dos formas distintas para verificar la una con la otra: salario medio de una empresa por un lado, y número de jornadas y valor total de los salarios, por otro; el primer dato se deduce de los segundos. Edad y fecha de nacimiento, etc.

- *Investigaciones representativas y no representativas*. En la crítica de este problema hemos de ocuparnos de la responsabilidad de un grupo, de una muestra estadística. Bástenos advertir ahora que la condición esencial para que un grupo limitado represente el total de casos, es que el grupo parcial contenga elementos de la misma característica del total y con iguales frecuencias.

La encuesta de costo de vida obrera hecha sólo entre los obreros de una fábrica de cervezas, no serviría como muestra de toda la clase obrera de una ciudad, porque los salarios de aquella empresa no reúnen las mismas características, en sus ingresos, gastos, prestaciones sociales, clase de trabajos, oficios, residencia, que se observan en toda la masa de trabajadores de la ciudad. Incluso, tal vez, la composición demográfica de las familias de aquel sector de obreros sea distinta de la composición familiar de todo el obrerismo del país. Nos hallaríamos ante una estadística poco representativa.

- *Investigaciones obligatorias voluntarias*. Son obligatorias en cuanto a la rendición de datos; por lo general, las que organizan los gobiernos, porque valiéndose de la atribución legal para exigir las aclaraciones, así se dispone. El registro civil, base de la estadística demográfica; la declaración de las importaciones y exportaciones, son obligatorias, como la mayoría de las estadísticas que llevan las entidades que el Estado ha creado para ello.

Sin perjuicio de esto, los organismos del Estado también pueden y suelen organizar investigaciones voluntarias, ante las cuales el público puede contestar o dejar de hacerlo.

Las recopilaciones *voluntarias* de datos, frecuentemente se llevan a cabo por las instituciones privadas y se refieren comúnmente a las monografías y encuestas científicas. La radio, prensa y las revistas suelen invitar a sus lectores a opinar sobre algunos problemas candentes o a declarar un dato de su vida o negocio particular.

Pues bien, de estos sistemas, el proyecto, para el caso particular, tendrá que decir cuál interesa más y cuál debe emplearse.

Algunos autores consideran tres sistemas importantes de recolección: *correo*, *entrega personal del cuestionario* y *la entrevista*; otros sistemas de menor importancia corresponden a: *internet, teléfono*, *fax* y *panel*.

Todos ellos presentan ventajas y desventajas. Se considera que la *entrevista* presenta más ventajas que los otros sistemas, siendo la más importante: mayor número de cuestionarios recolectados, mayor número de respuestas, permite aclarar el objetivo de la investigación y las dudas del informante. Entre las desventajas se menciona: mayor costo, más tiempo para la recolección y una cantidad mayor de personas, además, la influencia negativa que puede presentar el entrevistador.

La organización. Es éste otro de los problemas que debe estudiarse en un proyecto; qué forma se va a dar a la organización de la estadística de que se trate. Existen dos tendencias: *centralista* y *descentralista*.

La centralización. Es incuestionablemente necesaria en el orden técnico. La técnica, las normas generales, la dirección de toda investigación debe ser central y general para todo el territorio, para el campo de cada investigación y para todos los organismos inferiores y el personal que intervenga.

Es el único derecho y medio de conseguir unidad en los sistemas y de poder totalizar los datos de todo un municipio, de todo un departamento, o del conjunto del país. Si se deja a cada sección que realice las estadísticas que quiera y por los medios que se le ocurran, las diferencias regionales impedirán reunir en totales nacionales los datos e incluso harán imposible las comparaciones interregionales.

La descentralización. Debe emplearse en el orden administrativo, en el orden ejecutivo, para seleccionar personal, para cooperar en el presupuesto de la estadística, para aportar el material, para recoger los datos, para controlar la bondad de las investigaciones, para, en fin, servir de intermediaria entre las diversas categorías de dependencias.

Claro está que a la centralización, en ningún caso, puede corresponder uniformidad absoluta en lo esencial y en lo accidental. La dirección central debe tener en cuenta las particularidades regionales del hecho o fenómeno que se investiga estadísticamente, con el fin de ajustar el proyecto a las distintas características geográficas.

El material estadístico

Está constituido por todos los útiles, documentos o instrumentos necesarios para llevar a buen fin la investigación. El proyecto tiene que enumerarlo en sus clases y en su número. Podemos dividirlo en: material *impreso* e *instrumental*.

Material general para la investigación y material para la investigación de que se trate.

Material impreso. Se refiere a los formularios o cuestionarios, boletines, hojas de inscripción, registros, circulares, pliegos de instrucciones.

El formulario, boletín, boleta o cuestionario, que de todas las formas suele denominársele, es el instrumento que sirve para anotar las respuestas o el producto de la observación personal del hecho que se investiga.

Los formularios pueden ser:

- *Colectivos.* Destinados a la inscripción de los datos de varias unidades estadísticas, principalmente cuando esas unidades son personales. El formulario del censo de población es colectivo, puesto que en uno se reúne la información de todos los habitantes de una vivienda u hogar.

Tiene la ventaja de facilitar la recepción de los mismos en la oficina que investiga, puesto que los colectivos son menos en cuanto al número y supone por esto mismo, una economia en la impresión. Exigen, generalmente, un investigador que haga las anotaciones y que lleve personalmente el formulario.

- *Individuales*. Destinados a la inscripción de los datos de un solo elemento o unidad estadística: una persona, una fábrica, un almacén, una sociedad industrial, una nueva construcción, un hospital, una escuela, etc.

Ofrecen la gran ventaja de facilitar el agrupamiento, el recuento o escrutinio, con sólo amontonar los formularios de una misma naturaleza: por departamentos, por clases, por industrias, por países de origen, por países de destino, por sexo, por importancia de un carácter cuantitativo como capital, producción, consumo, importe del salario, etc.

Contenido de un formulario. En general todo formulario debe contener:

- El *encabezamiento* que comprende: el nombre de la institución u organismo que realiza la investigación: Universidad Nacional, Investigaciones Económico-Sociales, Federación Nacional de Cafeteros, Departamento de Investigación de Mercadeo, etc. Un número que lo personifique de los demás. Esta enumeración puede ser doble: 1. Numeración de orden igual para todos los ejemplares de un mismo formulario. 2. Numeración de orden, dentro de un mismo formulario, para cada unidad estadística.

El título del formulario, que nos diga de qué investigación se trata: "Investigación sobre la propiedad". "Investigación sobre precios". Además cómo, cuándo y dónde se realizó la investigación.

- El *cuerpo* abarca la indicación de la clase, dirección y nombre de la unidad o grupo estadístico que se investiga. Además las preguntas generales relativas a toda la unidad. No siempre es necesario dedicar la primera parte del cuerpo a generalidades, que tiene por objeto recoger información sobre el nombre del informante, dirección, sexo, ocupación, cuando se trata de personas; o nombre de la entidad, razón social, dirección, actividad económica, etc. De acuerdo al objetivo, se puede prescindir de este tipo de preguntas.

También comprende las *preguntas particulares* relativas a la unidad o unidades examinadas. El cuerpo de un formulario puede disponerse en serie o en cuadro. En forma de serie o cuestionario, una pregunta después de la otra, se emplea frecuentemente cuando el formulario es individual; en forma de cuadro, cuando es colectivo. No es extraña tampoco la combinación de ambos sistemas en un mismo formulario. Incluye preguntas de *control*, *filtro*, *abiertas*, *cerradas* que pueden ser de selección *múltiple* o *dicotómicas*.

Las preguntas pueden ser *cerradas*, cuando se responde si o no; *abiertas*, en el caso de que se pida una opinión; *selección múltiple*, si se da una serie de alternativas para que escoja una o varias; *control*, para determinar la veracidad del informante; y preguntas *filtro*, que nos determinan si se debe dar por terminada la entrevista o si debe pasar a otra parte del formulario.

Las preguntas deben ser: claras y sencillas; utilizando el mismo vocablo del informante; se inicia con las más fáciles y se termina con las más complejas; no se deben hacer preguntas que no vayan a tener respuesta.

Las *observaciones*: en todo formulario conviene destinar un espacio libre al final, en la última columna del cuadro, para que el investigador o el investigado puedan anotar las aclaraciones y observaciones que crean convenientes, como complemento de las preguntas impresas, de por si rígidas.

- Las *instrucciones*: casi siempre, la brevedad y concisión de las preguntas de un formulario exigen explicaciones de su sentido o sobre la manera de dar respuesta. Para facilitar la labor del encargado de llenar el documento, se incluyen algunas instrucciones, bien en el mismo formulario (al pie, en la cabeza o al dorso), con sus correspondientes llamadas entre paréntesis (1), (2), etc., bien en pliego aparte cuando son extensas. En este último caso, ninguna pregunta puede ser omitida, aún cuando no haya necesidad, por la claridad y precisión de la redacción, deberá acompañarse una enumeración metódica en forma de código y con ejemplos: 10, 20, 30, etc.

No hay criterio unificado, sobre el lugar del formulario donde deben colocarse las instrucciones. Algunos consideran que debe ser después del encabezamiento, para obligar al informante a leerlas antes de iniciar la contestación, pero otros, aconsejan ubicarlas ya sea al final o en cartilla separada del formulario aduciendo una mejor presentación.

- La fecha de envío del formulario a la entidad que recoge los datos.
- La firma responsable del investigador o agente cuando lo hay, o del investigado cuando éste informe directamente: propietario, administrador, gerente de una empresa.

También suele incluirse al pie, un renglón para la firma o visto bueno del jefe o coordinador de la investigación, el nombre o firma del entrevistador.

Trazado y redacción de los formularios. Al proyectar una nueva investigación, se debe diseñar el material que vaya a repartirse y someterlo a discusión y aprobación como el resto del proyecto.

Las normas del diseño y redacción son las siguientes:

- Debe ser *sucinto*, es decir, limitado a las preguntas esenciales, justamente las que se necesitan para los fines de la investigación y que en realidad puedan obtenerse de la fuente informativa.
- Debe prescindirse de toda pregunta *indiscreta* que levante suspicacias y temores o que moleste al investigado.
- Debe *ser claro*, fácilmente comprensible por cuantos, de una u otra manera, tengan que manejarlo. No ofrecer dudas sobre la forma de contestar cada pregunta; que no admita sino una sola interpretación.
- *Categoría*, generalmente las respuestas deben limitarse a un número o a una afirmación (sí) o una negación (no), por lo tanto las preguntas deben ser cerradas.
- Debe procurarse evitar los *juicios personales* del investigador o del investigado, como cuando se deja a criterio del calificador juzgar la *importancia* o la *bondad* de un hecho (grande, mediano, o pequeño); (bueno, regular, malo).

Otros aspectos: la cuestión de los formularios trae consigo otros considerandos que deben resolverse en el proyecto. Estos son: *clase de papel, forma de impresión* (mimeografiados, en imprenta), tamaño general y espacios intensos; anulación de casillas en los cuadros por no tener que contestar, inclusión de

preguntas de control, diversidad de colores en las tintas, rayados fuertes y débiles, etc. Todos los problemas relativos a las labores de imprenta.

Pero lo esencial es, en la etapa de prospección, calcular el número de formularios que se han de necesitar para la investigación previa y para la definitiva.

Todo formulario antes de imprimirlo, debe ser ensayado con una recolección real de datos. Por muy perfecto que se suponga el modelo en una oficina, la práctica enseña a perfeccionarlo, a corregir la redacción, a suprimir lo imposible e innecesario, a añadir lo que se dejó en olvido. Por otra parte permite establecer el tiempo necesario para su diligenciamiento, la reacción del informante y la claridad de las instrucciones. Lo anterior implica la realización de una *encuesta preliminar o piloto*. El material impreso lo podemos dividir en dos: *material de recolección* y *material* de *elaboración.*

Equipos. En la recolección de datos, a veces, y en la elaboración posterior casi siempre se requiere de muchos instrumentos, aparatos, máquinas y útiles, que quien proyecta debe tener en cuenta, en su número y clase.

En cualquier investigación, generalmente se necesitan máquinas calculadoras, sumadoras, de escribir, tablas de cálculo, computadores con programas estadísticos (Microstat, SPSS, SAS, Estagraf, TSP, etc.) material de dibujo y los útiles propios de toda oficina.

En las grandes investigaciones (censos, comercio exterior, estadística fiscal, de criminalidad, etc.), se necesitan equipos que permitan el procesamiento de grandes cantidades de información.

En ciertas investigaciones especiales, se requiere también de un instrumento especial, sin el cual no se podrían recoger los datos. Si es una investigación antropométrica, requiere escalas cromáticas de la piel, del pelo y de los ojos, cinta métrica, balanza, etc. Si es de una investigación meteorológica, termómetros, barómetros, pluviómetros, etc. Si se trata de llevar en una empresa un registro estadístico del horario de entradas y salidas del personal, se recurrirá a un reloj de marcar. Lo mismo en un aparcadero sobre entrada y salida. En un almacén, la estadística de ventas e ingresos se lleva en la caja registradora y por último, los datos del movimiento de viajeros en los buses se conocen por el control mecánico de los usuarios.

El presupuesto de la investigación

El presupuesto comprende el cálculo del costo y su financiación; las partidas del costo, en general, suelen ser las siguientes:

a. Organización

Estudios preliminares	Asesorías
Trabajos geográficos	Formulación del plan
Plan de propaganda	Impresión del formulario
Adiestramiento del personal	Contratación de servicios auxiliares

Materiales y equipos.

b. Trabajos de campo (recolección)

c . Sistematización

d. Publicación.

Ambos aspectos, la forma de allegar los recursos y la estimación previa de los gastos, constituyen el último punto de un proyecto de Investigación Estadística.

La *preparación*. Una vez redactado el plan de la Investigación Estadística y antes de llevarlo a la práctica, requiere ser sometido a un ensayo denominado preparación.

Todo proyecto de esta clase, conviene ser discutido y aprobado, en último término, por un grupo de entendidos y técnicos en estadística, puesto que, por regla general, los proyectos enteramente individuales no suelen ser perfectos, adolecen de algún defecto que el encargado de realizarlo no alcanzó a ver, la critica de un plan estadístico no sólo debe estar hecha por los peritos en procedimientos estadísticos, sino también por los técnicos en las materias que van a investigarse.

En materia de estadística agrícola, se necesita el parecer de los agrónomos y de los agricultores. En una investigación industrial el de los administradores y empresarios. En una información comercial el del gremio de comerciantes y para una estadística bancaria, o sindical, no puede prescindirse de la opinión de los banqueros o de los sindicatos, respectivamente.

Aprobado el plan con las modificaciones del consejo de técnicos, se inicia la preparación del mismo. Ésta comprende la práctica y el ensayo o prueba en los siguientes aspectos:

- *Adiestramiento y elección* del personal por medio de cursos más o menos breves y entrevistas.
- *Ensayo de investigación sobre el terreno*. Esto supone el ensayo correspondiente a la época y fecha del material, del propio personal, utilizado de la fuente de información que genera, de la organización y de los programas a los sistemas que se utilizan.
- *Elección y preparación de la fuente informativa*. No basta con adiestrar personal de recolección e investigadores. Se requiere, al mismo tiempo, instruir a los sujetos o a las entidades que van a ser investigados, sobre los registros contables o previamente estadísticos que deben llevar, sobre la manera de contestar los cuestionarios, sobre la fecha de visita de los recolectores, sobre la obligación de rendir los datos, etc.
- *Preparación del material de la encuesta*. La impresión, conexión, el reparto de material, su adquisición cuando se trata de instrumentos, son gestiones que como deben estar terminadas antes de dar comienzo a la recolección propiamente dicha de los datos, entran dentro del período preparatorio.
- *Propaganda*. La labor de anunciar que se va a proceder a una investigación estadística, las disposiciones del ánimo del público afectado por la investigación, el aislarla de todo

intervencionismo perjudicial del Estado, de todo recelo contra la solicitud de datos o de todo temor por las consecuencias que puedan presumirse; el dar a conocer el interés general de los datos estadísticos, incluso y principalmente para aquél que los rinde, llena esta función de propaganda de las estadísticas, cuya realización encaja necesariamente en la fase preparatoria.

- La preparación y actualización del *listado de informantes* o *marco de referencia* que son objeto de investigación; esta labor debe desarrollarse en su totalidad.

- En algunas investigaciones se requiere del uso de la *cartografía*, por lo tanto es necesario haberla preparado o actualizado.

- *Financiación*. La consecución del dinero necesario para cubrir todos los gastos, incluso la elaboración posterior, no debe dejarse para más tarde de esta etapa de preparación. Desgraciadamente, son frecuentes las investigaciones que no se terminan, aun después de haberse hecho el esfuerzo y el gasto parcial, necesarios en la recopilación de los datos.

- Hay aspectos, además de los anteriores que merecen especial atención por ejemplo: el *calendario de trabajo* que consiste en una descripción de las diferentes etapas de la investigación, indicando la fecha de inicio y de término de cada etapa; en algunos casos se utilizan las barras Gantt que se colorean a medida que se realiza el trabajo. La *formulación de hipótesis* en aquellas investigaciones por muestreo en las que se desea hacer comprobaciones; el *examen de la documentación* requerida antes de comenzar la investigación, siendo posible que la investigación haya sido realizada y no se requiera de un nuevo estudio; en otros casos el proceso será únicamente de actualización, y en situaciones extremas se podrían utilizar las experiencias, corrigiendo los errores cometidos; finalmente la *unidad de investigación*, es decir, a quién va dirigido: una persona, familia, hogar, manzana, barrio, empresa, etc.

La duración del período preparatorio depende de la clase de investigación. Hay estadísticas que requieren un corto plazo para realizarlas (precios, salarios, producción, transportes) y otras exigen meses (censos de población, industriales, ganaderos y agrícolas, encuestas de hogares).

RECOLECCIÓN

Preparada la investigación, comienza la recolección definitiva de los datos.

De esta tarea depende todo el resultado posterior de la estadística. Si está mal hecha, la elaboración resultará incorrecta e incluso imposible de efectuar, y si se realiza, dará origen a un análisis erróneo y a interpretaciones falsas.

La investigación externa de los hechos, recopilación del material básico, puede efectuarse por muy diversos sistemas y con muy distintos fines.

Puede referirse a los datos preliminares o a los datos definitivos. La investigación preliminar coincide con las labores de preparación que hemos visto antes. Su necesidad se deja sentir más cuanto más extensa y complicada sea la investigación. Presta un triple servicio: enseña, controla los resultados y sirve para la organización.

Su valor como medio de control, porque los datos previstos, su distribución y sus peculiaridades indican la calidad del material (formularios), el personal indispensable, la propaganda conveniente, la

elección de las fuentes informativas, la manera de repartir y recoger el material, la fecha apropiada, la duración, y los procedimientos a regir en la sistematización de los datos.

La etapa de recolección comprende varios aspectos:

- ❑ Distribución del material o instrumento de recolección
- ❑ La recolección propiamente dicha
- ❑ Control del número de formularios recolectados
- ❑ Control sobre la calidad de las informaciones recogidas

CRÍTICA Y CODIFICACIÓN

De la investigación se obtienen los datos individuales, colectivos o aislados. Cada cifra presenta la intensidad de un hecho particular; el salario de un obrero, la producción de una fábrica o los préstamos otorgados por una entidad de crédito.

> **Es un conjunto de operaciones de revisión y corrección de la información recolectada, que nos permita agruparla y procesarla, de tal manera que nos facilite la elaboración de cuadros, gráficas y análisis, necesarios en la publicación.**

Todo ese conjunto de material recogido en los formularios antes de ser totalizado y utilizado, requiere un examen crítico, severo, con el objeto de comprobar si cumple con estas condiciones indispensables; ¿son exactos o erróneos los datos?, ¿son completos, precisos o imprecisos, representativos o no representativos?

El objeto de la crítica, es clasificar el material primario que precede de la misma investigación, en tres grupos: material bueno, material incorrecto pero corregible y material incorregible o desechable.

La clase y la importancia del error cometido, determinan la admisión o no del dato primitivo resultante de la observación o declaración. Además, se critican los datos numéricos directos para establecer las causas de los errores y modificar o perfeccionar las fuentes y los métodos de observación.

Difícilmente se consigue, por primera vez, informaciones perfectas, en cuanto a la exactitud, a la extensión y consecuencialmente en cuanto a la responsabilidad de las investigaciones parciales.

La necesidad de procesar las informaciones, recogidas en los cuestionarios, ha obligado a traducir las respuestas en códigos. Por ejemplo el código para la pregunta estado civil:

Soltero	1
Casado	2
Viudo	3

Divorciado	4
Separado	5
Otro	6

Cuando el número de respuestas pasa de 9, es preciso utilizar cifras de dos dígitos.

EJEMPLO A

CLASIFICACIÓN POR PROFESIÓN	
Abogados	01
Arquitectos	02
...	
...	
Zootecnistas	20

EJEMPLO B

CLASIFICACIÓN POR CATEGORÍAS SOCIO-PROFESIONALES
0 Agricultores
00 Agricultores (sin ninguna indicación)
01 Propietarios explotadores
02 Colonos
03 Aparceros

TABULACIONES O PROCESAMIENTO

Puede ser manual, mecánica o computarizada y su elección dependerá:

- De la cantidad de formularios que se van a utilizar.
- Del número de preguntas que tenga el formulario.
- Del tiempo y de los recursos, ya sean financieros o de equipo disponible.

Cuando la tabulación se acuerda desde el principio, como parte integrante de la planeación general de la investigación, es de suponer que todo el proceso será totalmente satisfactorio, lo cual ha sido demostrado por la experiencia.

Del proceso que origina una base de datos, de la cual se obtienen *listados*, son revisados a fin de detectar las inconsistencias que se presenten en los procesos anteriores. Una vez hechas las correcciones, se procede a elaborar los *cuadros*, con el fin de facilitar el cruce y *análisis* de la información, la elaboración de *gráficas*, las conclusiones y las recomendaciones, si las hay.

Hoy en día, el procesamiento de datos, la obtención de resultados, la aplicación de ciertas medidas estadísticas, se realizan con mayor rapidez y confianza si se utilizan paquetes estadísticos elaborados para tal fin. Entre otros, como ya se dijo: el SAS, el SPSS, el Microstat, Minitab, el Stagraf, TSP, etc. Por otra parte, el uso de la calculadora ayuda a agilizar los cálculos en la aplicación de las diferentes medidas, cuando se están desarrollando ejercicios en una clase de estadística o no se dispone del computador.

ANÁLISIS E INTERPRETACIÓN

Esta etapa, se puede considerar como la más importante del informe, ya que el análisis de los datos tendrá que ver con la formulación del objetivo mismo de la investigación y de las hipótesis establecidas; sin embargo, este proceso de análisis tendrá menos dificultad, si el investigador tiene pleno conocimiento de los problemas que son inherentes al planeamiento de una investigación.

En este proceso, se debe considerar la elaboración de distribuciones o tablas de frecuencias obtenidas a través de una sistematización de la información para poder ser presentada en forma de cuadros y gráficos. Con los anteriores resultados, se procede luego a hacer un resumen y a aplicar las diferentes medidas que hemos denominado *estadígrafos o estimadores, puntuales*, cuando son aplicados a las características de las unidades en la muestra o como *parámetros* aplicados en las características de la población, entre

los cuales figuran las medidas de *dispersión y los promedios, incluyendo en éstos los porcentajes y proporciones*.

Con las cifras resultantes, se pueden hacer comparaciones con otros estudios, para poder llegar a mejores conclusiones.

De esta última fase de la metodología se puede decir que encierra dos aspectos:

a) Análisis y evaluación estadística de los resultados.
b) Análisis y evaluación técnica de acuerdo con la naturaleza de la investigación.

Estos dos aspectos permitirán determinar el grado de consistencia y confiabilidad de los resultados obtenidos de la investigación.

El profesor Jhon W. Best en su libro *Cómo investigar en educación* nos da una posible guía del análisis, sugiriendo los siguientes puntos:

1. Título
   a) ¿Es claro y conciso?
   b) ¿No promete más de lo que el estudio puede proporcionar?
2. El problema
   a) ¿Se halla establecido con claridad?
   b) ¿Está bien delimitado?
   c) ¿Se reconoce su significado?
   d) ¿Las preguntas son específicas o se han establecido las hipótesis con claridad?
   e) ¿Se establecen supuestos y limitaciones?
   f) ¿Se definen los términos importantes?
3. Revisión de la bibliografía relacionada:
   a) ¿Es de amplitud adecuada?
   b) ¿Se destacan los hallazgos importantes?
   c) ¿Está bien organizada?
   d) ¿Se procura un resumen efectivo?
4. Procedimientos utilizados:
   a) ¿Se describe detalladamente el diseño experimental?
   b) ¿Es adecuado este diseño?
   c) ¿Se describen las muestras?

d) ¿Se reconocen las variables relevantes?
e) ¿Se procuran controles adecuados?
f) ¿Son idóneos los instrumentos de recolección de datos?
g) ¿Se establecen la validez y la fiabilidad?
h) ¿Es adecuado el tratamiento estadístico?

5. Análisis de los datos:
   a) ¿Es adecuado el uso de tablas y figuras?
   b) ¿Es concisa y clara la exposición del texto?
   c) ¿Es lógico y perceptible el análisis de las relaciones de datos?
   d) ¿Se interpreta con precisión el análisis estadístico?

6. Resumen y conclusiones:
   a) ¿Se replantea el problema?
   b) ¿Se describen en detalle los procedimientos?
   c) ¿Se presentan concisamente los hallazgos?
   d) ¿Es objetivo el análisis?
   e) ¿Los datos presentados y analizados justifican los hallazgos y conclusiones?

PUBLICACIÓN

Corresponde a la fase final de la investigación, y con ella se propone hacer llegar a las personas interesadas, el resultado total del estudio, teniendo en cuenta todos los aspectos considerados en el proceso, de tal forma que los datos sean comprensibles, con la correspondiente validez que merezcan las conclusiones.

En términos generales se puede decir que un informe deberá contener:

- Planteamiento del problema
- Objetivo de la investigación.
- Hipótesis que se quieren probar.
- Breve exposición de la metodología adoptada, diseño y tamaño de la muestra. Proceso de selección de las unidades de información y de recolección.
- Se podrá incluir en el informe, copia del formulario utilizado en la recolección, aún relacionando y justificando, en forma muy sucinta, las preguntas que se consideran más importantes dentro de la investigación.

- Descripción de los resultados en forma de cuadros y gráficas, acompañados del análisis y comparaciones obtenidas a través de los datos.
- Conclusiones y recomendaciones. Estas últimas cuando así lo exija la investigación.
- En algunos casos, el informe tiene una parte final, denominada apéndice, en donde se incluyen cuadros más generales, que permiten aclarar o comprobar rápidamente cualquier información más detallada. También se puede incluir documentación complementaria al informe.

MONOGRAFÍAS Y ENCUESTAS

Al dividir las investigaciones en *extensivas* e *intensivas,* indicamos que estas últimas se refieren, ante todo a estos dos métodos de recolección de datos tan en boga hoy y que denominamos *Monografías* y *Encuestas*.

Las monografías se iniciaron en los sistemas de observación social y económica por el Ingeniero y Sociólogo francés Federico Le Play (1806-1882) a mediados del siglo *XIX*. Se aplicó la monografía por el propio Le Play, por sus continuadores en las investigaciones de la situación física, económica, sanitaria, moral, social, etc., de las familias obreras. Luego se extendieron estos mismos estudios a otras clases sociales y más tarde a monografías de los municipios, regiones y talleres industriales.

Como podemos deducir etimológicamente de la palabra monografía, se trata de la descripción de un caso, para nosotros, de una unidad estadística. Desde este punto de vista, estrictamente, la monografía no es investigación estadística, puesto que esta última, por su misma naturaleza, tiene que referirse a un grupo grande, el mayor número posible de casos, salvo que se apliquen métodos de muestreo aleatorio, con lo cual se logra validez de los resultados obtenidos.

Al analizar un solo caso, la monografía se detiene en multitud de aspectos y cuestiones que directa e indirectamente se relacionan o afectan al sujeto u objeto investigado. Intensifica el examen, recoge muchos detalles, se hacen muchas preguntas, que sólo en un estudio profundo interesan. El fundamento científico de este método arranca de la razón de "muestra", "sampling" en inglés, de la posibilidad de representar un conjunto o masa por un solo caso; es la tipicidad individual. Sus defensores nos dicen que así como el químico toma una cantidad (un litro) de agua para analizar la composición química de un embalse y el ingeniero de minas se conforma con granos de mineral para calificar la riqueza de un filón, el sociólogo, el economista y el hombre de negocios pueden, igualmente, limitarse a una muestra social o económica, a un ejemplo de lo sucedido en la vida del comercio y de las finanzas para luego generalizar al conjunto de lo observado en su caso.

El *estudio de un sujeto elegido como tipo*. Un individuo, una familia, una empresa, una fábrica, un almacén, una hacienda, etc. En realidad el procedimiento monográfico se diferencia, además, de la estadística porque indaga aspectos cualitativos que no se pueden reducir a cifras, sin prescindir de preguntas cuantitativas de números e intensidades, pero también incluye inscripciones cualitativas y descripciones literarias, subjetivas, impresiones de observador o del observado.

Es un sistema de observación directa y de generalización indirecta: de un caso a todos.

Dificultades. Aún cuando se discute el hecho de que un caso, una familia, un edificio, un taller, etc., pueda representar bien todas las familias, oficios, talleres, etc., es evidente que existe un caso más *representativo* que todos los demás; la muestra mejor, el que al sustituir al conjunto comete menor error, compensa exactamente o tiende a comprender los errores, por ser unos por defecto y otros por exceso. El análisis monográfico de ese caso precisamente y no de otro, no deja de tener interés.

Pero la dificultad estriba en determinar de antemano en su conjunto, cuál es ese caso más representativo de la unidad a investigar. El tropiezo del método monográfico, es la elección a priori de la unidad a investigar: ¿Cómo podemos saber que la familia, el taller, el municipio, o la empresa elegida al azar o con orientaciones parciales, representa bien a la colectividad de familias, talleres, municipios o empresas?

Esta crítica, con razón ha hecho que el sistema monográfico puro, apenas o incorrectamente se emplee en la actualidad.

A pesar de lo expuesto anteriormente, es un gran auxiliar de la estadística propiamente dicha. La investigación estadística afecta a un gran número de casos, en sus manifestaciones cuantitativas o cualitativas traducibles a número, y siempre indagando los aspectos esenciales del problema, fenómeno o hecho del mundo económico en nuestro caso. La monografía, si es representativa, confirma las conclusiones de la estadística y las reviste de detalles minuciosos, cualitativos, de descripciones sagaces y de interpretaciones, que ella –la Estadística– por su naturaleza mecánica y estrictamente objetiva, no estaba en condiciones de captar.

DIFERENCIAS DE LA ESTADÍSTICA CON LA MONOGRAFÍA

- La estadística observa y registra fenómenos colectivos, en su conjunto o en su mayor parte posible.
- La monografía, estudia intencionalmente sólo un hecho. La estadística, es exclusivamente numérica. La monografía es descriptiva o literal sin perjuicio de poder emplear el número.
- El análisis matemático y gráfico que emplea la estadística, no caben en la monografía.

La Encuesta. Para subsanar el defecto de la monografía, se comenzó a aplicar un procedimiento llamado por algunos autores franceses “Polomonográfico”. Si un caso es insuficiente, estudiemos varios, si una monografía no basta, recojamos varias; de éstas saquemos la más representativa, a quien le atribuiremos, a su vez, la representación de todo el grupo. De esta forma, además de obtener una sola muestra bien elegida, de un grupo, el grupo parcial nos permite estudiar las variaciones que entre las unidades se presentan. El grupo parcial nos permite conocer lo normal, lo típico y lo anormal o atípico. Luego generalizaremos ambos conocimientos: el normal y el anormal. A este sistema se le llama hoy *encuesta*.

La Encuesta, es pues, una investigación estadística parcial o incompleta cuando recoge datos numéricos y deja de ser una investigación estadística propiamente dicha, aún cuando luego pueda someterse a un recuento o elaboración estadística, cuando se limite a anotaciones literarias. Ejemplo de estas últimas son las que suelen organizar los periódicos y revistas, y muchas instituciones profesionales y técnicas; preguntan el parecer, la opinión, por ejemplo, sobre la situación económica, sobre temas de la política, de los deportes, aficiones y gustos artísticos, etc.

A las encuestas se les suele dividir, en cuanto al organismo que las realiza, en: *privadas y públicas*, y a estas últimas en *parlamentarias y estadísticas.*

Las encuestas *privadas* son muy corrientes en los países anglosajones en donde existe una inclinación grande entre el público a dar respuesta a las preguntas de una encuesta. Son menos frecuentes en los países latinos. La prensa, con el objeto de pulsar la opinión y dar la publicidad y las entidades científicas privadas, con el propósito de alcanzar avances en las ciencias, organizan esta clase de investigaciones.

La distinción de las encuestas ***públicas*** y *oficiales*, en ***parlamentarias*** y ***estadísticas***, se funda, principalmente, en el hábito de los sistemas democráticos de estudiar la realidad de los problemas sobre los cuales tiene que legislar. Así, el Congreso, Asambleas, los Concejos, encargan a sus miembros realizar encuestas alrededor de los tópicos de discusión. Por lo general los senadores, representantes, diputados, concejales, buscan interrogar a las personas interesadas, a observar visualmente la realidad y luego informan.

La Estadística, por el contrario, recoge datos numéricos, realiza anotaciones. Son ejemplos constantemente repetidos, las encuestas sobre ingresos y gastos que llevan a cabo en casi todos los países, así como otras sobre problemas sociales.

La sociología emplea de ordinario el sistema de observación por encuesta.

En cuanto al procedimiento de recoger los datos y hacer las preguntas de una encuesta, se distinguen dos: ***orales*** y ***escritos***.

Las encuestas parlamentarias, por lo regular son orales. Las estadísticas tienen que ser escritas.

Las escritas, unas veces se organizan repartiendo cuestionarios entre el público en general, para que quien desee las responda.

Otras veces se remiten los formularios ex profeso a determinadas fuentes de información, determinadas personas, entidades, empresas, que se consideran de interés o representativas de su sector. También se utiliza el método de agentes investigadores, (reportajes) que acuden personalmente al campo de la observación con el cuestionario a recoger los datos y opiniones.

Las encuestas suelen ser muy frecuentes en los períodos difíciles, en las crisis económicas o cuando se agudiza un problema social o económico, huelgas, epidemias, plagas en la agricultura, alzas de precios, de empleo, pliegos de peticiones obreras, alzas de impuestos, etc. Por eso resultan investigaciones ocasionales.

Las diferencias con la investigación estadística propiamente dicha, son las mismas que hemos señalado para las monografías, con la particularidad de que en la encuesta se observa un mayor número de casos y suelen emplearse para conocer opiniones de personas o entidades en una materia.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. ¿Cuáles son los nuevos términos aprendidos en este capítulo?

2. ¿Qué significado tiene cada uno de los términos recordados?

3. Ilustrar con ejemplos lo que se entiende por población, población finita, población infinita, muestra, característica, variable.

4. Mediante ejemplos, explicar la diferencia entre la estadística descriptiva y la estadística inductiva.

5. ¿Por qué es útil la estadística en el campo para el cual se está preparando?

6. Dar como mínimo tres ejemplos de fenómenos económicos, considerados dentro del campo de la investigación estadística.

7. Dar tres ejemplos de fenómenos no considerados dentro del campo de la estadística.

8. "La estadística estudia el comportamiento de fenómenos colectivos y nunca de una observación individual". Comentar este principio.

9. Señalar el literal que más concuerde con cada una de las siguientes aseveraciones.

   9.1 Una buena muestra industrial debe ser:

   a) Pequeña. b) Grande.
   c) Representativa de la población objetivo. d) Incluir sólo los establecimientos grandes.
   e) Incluir sólo los establecimientos pequeños.

   9.2 La clasificación de los establecimientos dedicados al comercio al detal se hace teniendo en cuenta:

   a) Su función. b) El valor de las ventas.
   c) El valor de la producción. d) El valor del trabajo realizado.

   9.3 Una población infinita es aquélla en donde:

   a) El número de elementos está completamente delimitado.
   b) El número de elementos pasa de 1.000.
   c) El número de elementos está escondido.
   d) El número de elementos no está delimitado.

10. Contestar verdadero o falso, según el caso:

   a) La palabra población, significa en la metodología estadística lo mismo que en cualquier otra disciplina.

   b) La teoría estadística es de naturaleza general y puede aplicarse a cualquier campo del conocimiento.

   c) A las medidas que se obtienen de una muestra, se les da el nombre de parámetros, mientras que a las obtenidas de una población se les denomina estadígrafos.

   d) Se conoce como fuente primaria, aquella información que se obtuvo inicialmente, es decir directamente de la persona, empresas o entidad investigada.

   e) Una muestra aleatoria, es aquélla en la cual ciertos elementos tienen mayor posibilidad que otros de ser seleccionados.

11. Señalar el literal más adecuado para las siguientes observaciones:

11.1 La investigación preliminar permite:

a) Establecer la hipótesis. b) Determinar la muestra.
c) Coordinar el personal de campo. d) Ninguna de los anteriores.

11.2 Antes que nada, la investigación estadística requiere

a) Que exista un objetivo. b) Que se hayan trazado planes.
c) Que se tenga un problema. d) Ninguna de los anteriores.

11.3 El costo de una encuesta, por correo, es generalmente:

a) Igual al de una encuesta por medio de entrevistas personales.
b) Mayor al de una encuesta por medio de entrevistas personales.
c) Menor al de una encuesta por medio de entrevistas personales.
d) Imposible de medir en relación con el costo de una encuesta por medio de entrevistas personales.

11.4 En el diseño del cuestionario, las preguntas más difíciles deben colocarse:

a) Al principio, para salir inmediatamente de la parte más difícil.
b) En el centro para que sean precedidas y seguidas por preguntas fáciles.
c) Al final, luego que se haya establecido un clima de confianza, al comenzar por las más fáciles hasta llegar a las difíciles.
d) Ninguna de los anteriores.

12. Contestar verdadero o falso, según el caso:

a) Código es la representación cualitativa de un hecho cuantitativo.

b) Después de elaborar el formulario se define el objetivo de la investigación.

c) Al recolectar información por medio de entrevistadores se tiene la ventaja de que estos pueden observar el sitio de la operación que se está llevando a cabo.

d) Se conoce como fuente primaria, aquélla que obtuvo inicialmente la información directamente de la persona o entidad.

e) Al diseñar un cuestionario no es de gran importancia la forma de hacer la pregunta, siempre que ésta sea clara.

f) No hay posibilidad alguna de que en una encuesta, por correo, se interpreten mal las preguntas de un cuestionario, siempre y cuando la persona que la reciba sepa leer.

g) El examen de la documentación y metodología se efectúa después de tabulada la información.

13. Se ha dicho que en una investigación se consideran tres etapas; las que a su vez se subdividen en otras fases. ¿Cuáles son? ¿Podría usted reagrupar los titulares de este capítulo en un índice de temas de acuerdo con estas etapas?

14. Mencionar algunos aspectos técnicos y materiales que deben tenerse en cuenta en el diseño de un formulario.

15. Un ejemplo de característica cualitativa pueden ser datos sobre:

a) Salarios b) Pulsaciones por minuto c) Gasto mensual en alimentación.
d) Ocupación e) Temperatura.

16. Una muestra es aleatoria cuando las unidades se seleccionan:
    a) En forma caprichosa. b) Por conveniencia.
    c) A través de un censo. d) En forma repetitiva.
    e) De tal manera que todas tengan la misma posibilidad.

17. Por población o universo se entiende:
    a) Un recuento de unidades.
    b) Un conjunto de seres humanos.
    c) Un conjunto de datos.
    d) Un conjunto de medidas o el recuento de todas las unidades que tienen una característica común.
    e) Ninguna de las anteriores.

18. Cualquier medida aplicada a la característica de las unidades en la población se denomina:
    a) Parámetro. b) Estimador.
    c) Estadístico. d) Variable.
    e) Población.

19. Dentro de los hechos o fenómenos que no caen dentro del campo de la estadística están:
    a) Los de frecuente repetición. b) Los de distinta frecuencia.
    c) Los colectivos. d) Los individuales.
    e) Los cualitativos que pueden cuantificarse.

20. La estadística descriptiva tiene como objetivo:
    a) Probar la significación de los resultados.
    b) Ser herramienta indispensable en el muestreo.
    c) Descubrir las causas que originan el hecho.
    d) Lograr conclusiones más allá de las muestras.
    e) Efectuar comparaciones sin sacar conclusiones de tipo más general.

21. Se debe responder *verdadero* si el enunciado es siempre válido. En caso contrario, se deberá sustituir la palabra *subrayada* por otra, con la cual el enunciado tenga validez.

    a) Parámetro, es el resultado al aplicar una medida a las características de las unidades de una *población.*

    b) El recuento de los empleados de una empresa, de acuerdo al cargo, es un ejemplo de características *cuantitativas.*

    c) La estadística *descriptiva,* es el "estudio" de una muestra para hacer estimaciones acerca de los valores estadísticos de la población, de la cual se tomó la muestra.

    d) Una muestra aleatoria significa que, cada elemento tiene una probabilidad *diferente* al ser seleccionado.

    e) La *inferencia,* es un ordenamiento sistemático de la información en cuadros y gráficas que observamos en las diferentes publicaciones e informes.

22. Explique brevemente los siguientes puntos.
    a) Factores o impedimentos que obligan a realizar una investigación parcial, en vez de una investigación total.

b) Marco muestral.
c) Hechos que abarca y no abarca la estadística.

23. En los tres (3) ejemplos siguientes, determinar en cada uno de ellos ¿cuál es la población?; ¿cuál es la muestra?; ¿cuál es la unidad?; ¿cuál es la característica?; ¿cuál es cualitativa o cuantitativa?; ¿cuál de las variables es discreta o continua?
   a) Se realiza un estudio en 350 hogares de la clase media de la ciudad Bellavista para conocer el tipo de aceite o grasa usada en la cocina. Los resultados fueron: 130 hogares utilizan el ajonjolí; maíz en 90 hogares; girasol en 75 hogares; etc.
   b) El laboratorio de control de calidad de una empresa realiza un test de rapidez de acción de un pesticida de jardín, en 50 plantas infestadas. Los resultados fueron observados cada hora, habiéndose obtenido algunos datos del número de plantas totalmente libres de plaga, después de los períodos de tiempo que se indican: 6 horas, 6 plantas; 7 horas, 9 plantas; 8 horas. 5 plantas; etc.
   c) En un plantel de 800 niños de ambos sexos, de 5 a 12 años, se realizó un test de aceptación a 20 niños, utilizando una escala de 10 puntos, para medir el grado de aceptación de un nuevo producto que fabrica la compañía Chocolatera la Avispa S.A.

24. Diga brevemente que diferencia hay entre:
   a) Parámetro y estimador
   b) Población y muestra
   c) Atributo y variable
   d) Muestreo aleatorio y no aleatorio

25. Qué entiende usted por:
   a) Estadística
   b) Estadísticas
   c) Fuentes o estadísticas primarias
   d) Fuentes o estadísticas secundarias
   e) Estadísticas internas y externas
   f) Unidad o elemento
   g) Objetivo de la investigación
   h) Proceso de recolección
   i) Muestreo aleatorio y no aleatorio
   j) Encuestas y monografías.

Síntesis de capítulo

Existe una diferencia entre los términos *estadística* y *estadísticas*; la primera se refiere a la técnica de recolección, procesamiento y análisis del dato, la segunda, corresponde al ordenamiento sistemático de la información en forma de cuadros y gráficas que aparecen en publicaciones.

La estadística cumple dos funciones: a) la primera, de análisis *descriptivo* en forma cuantitativa de las características observadas en el fenómeno, y b) la segunda, *inferencia* estadística o inducción, lográndose generalizaciones para un grupo mayor denominado población, partiendo de un grupo menor o muestra.

Se dan una serie de términos estadísticos que hay que conocer para un buen desarrollo del curso de estadística, entre otros: población, muestra, variables y atributos.

Se considera que la finalidad de la estadística, por ejemplo, en una empresa es suministrar información que permita: a) tener una visión general de ella en su conjunto, para que la

administración pueda formular directrices con pleno conocimiento de causa; b) descubrir las relaciones de causa a efecto en las diversas manifestaciones económicas de la empresa; c) reconocer y separar, en vista del control, lo normal y anormal, observando las fluctuaciones internas ligadas estrechamente con las condiciones externas, para obtener una mayor orientación a la actividad de la empresa.

También, se establecen las condiciones que deben presentar los fenómenos para que sean considerados dentro del campo de la estadística

El campo de acción de la Estadística es muy amplio; sin embargo, no todos los fenómenos son abarcados, únicamente aquellos que reúnen ciertas condiciones, a saber: fenómenos colectivos o de grupos; de frecuente repetición; de distinta frecuencia; distantes en el espacio; distantes de tiempo; cualitativos que pueden cuantificarse.

En cambio, quedan fuera del campo de acción de la estadística, los enumerados a continuación: Fenómenos individuales; que no se exteriorizan; accidentales en el tiempo y en el espacio; cualitativos que no pueden cuantificarse.

Son numerosas las finalidades de la estadística; entre las más importantes podemos anotar:

Conocer la realidad acerca de un fenómeno; determinar lo que sea normal o típico; observar los cambios que presentan; relacionar dos o más hechos; determinar las causas que originan; hacer estimaciones sobre el comportamiento futuro; partiendo de un grupo menor, obtener conclusiones para un grupo mayor (inferencia estadística).

En esta unidad se quiere resaltar, que no todas las investigaciones son indispensables de ser llevadas a cabo, debido al hecho de que en, algunos casos, se dispone de información obtenida de investigaciones realizadas con anterioridad, por la misma entidad o por otras entidades. En otras ocasiones, dicha información puede ser considerada como complementaria a la nueva investigación.

Una investigación requiere en la primera etapa, ante todo, de un planeamiento que comprenda:

- Fijar los objetivos de la investigación, determinando si el fenómeno puede ser observado mediante la aplicación de métodos estadísticos.
- Establecer la unidad o unidades de observación
- Determinar, si se trata de un censo o de una muestra.
- Si la encuesta se va a hacer mediante entrevista, entrega personal del cuestionario, por correo, teléfono, observación, panel, internet, fax, etc. Elaborar el presupuesto.
- Diseñar el calendario de trabajo. Diseñar el formulario e instrucciones.
- Seleccionar y capacitar a las personas que van a trabajar en la investigación. Hacer una preencuesta.
- En la segunda etapa, le corresponde como actividad la recolección.

- Distribución y recolección de formularios.
- Verificación del número de formularios enviados y recogidos. Efectuar un control sobre la calidad de los datos.

La tercera y última etapa comprende: la sistematización, análisis y presentación de los resultados

- Elaboración de códigos.
- Crítica y codificación
- Elaboración de cuadros de tabulación.
- Selección del proceso de tabulación, si es manual, mecánico o computarizado.
- Elaboración de cuadros y gráficas.
- Análisis y comparación de los datos.
- Publicación.

Las estadísticas y la estadística, que para muchos significan lo mismo, son términos empleados frecuentemente, siempre vinculados a cifras que miden algo. Así oímos hablar de la inflación mediante el crecimiento del índice de precios al consumidor, lo mismo de la tasa de desempleo o de criminalidad, llegando al extremo de desafiar y criticar duramente esas cifras, sin tener medianamente conocimiento de cómo fueron obtenidas ¿Cuál fue el procedimiento utilizado? ¿Qué significado tienen? y algo más, no pensamos si la interpretación dada es la correcta. Todo esto nos muestra lo poco que sabemos de esta herramienta para la interpretación y análisis de una información.

Por la anterior razón, se comenzó por definir aquellos términos que son indispensables para entender la teoría estadística, tales como población, unidad, muestra, característica, etc., que a primera vista parecen sencillos, en algunos casos resultan complejos. Veamos, en el salón de clase se dice que se va a realizar una encuesta a un grupo de familias de un barrio determinado de la ciudad, preguntémosle al estudiante: ¿Qué se entiende por familia, como unidad de investigación? resultando, un término definido de diferentes maneras. El investigador seguramente ha tomado alguna de ellas, desgraciadamente algunos de los que proporcionan información estadística no definen los conceptos que supuestamente miden, prestándose a muchas interpretaciones con los resultados presentados.

La parte metodológica, en la enseñanza de la estadística, parece que ha perdido su importancia si observamos el desarrollo y contenido de los de programas, lo cual podemos constatar en los últimos programas académicos. Se ha despertado un gran interés por la aplicación de modelos matemáticos que sólo son parte del proceso, algunos se centran en deducciones y comprobaciones, en vez de ayudar a comprender la estadística, sólo se ha logrado aumentar el grado de dificultad en su aprendizaje, dejando a un lado el conocimiento de la importancia que tiene la etapa de planeación y de recolección para producir una información consistente, que permita la aplicación de los métodos estadísticos tanto descriptivos como los de inferencia, a fin de realizar un buen análisis de los resultados y cumplir con los objetivos propuestos.

2 CAPÍTULO

Distribuciones de frecuencias

El fin primordial de la educación es formar los hombres capaces de hacer cosas nuevas y no de repetir simplemente lo que otras generaciones han hecho.

Jean Piaget

CONTENIDO

- Elaboración de tablas o cuadros de frecuencias.
- Elaboración de gráficos aplicados a la teoría estadística.
- Elaboración de gráficas generalmente utilizadas en informes.
- Recomendaciones para la elaboración de tablas y gráficos.
- Síntesis del Capítulo.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Seleccionar y elaborar gráficas que visualicen mejor la información.
- Aplicar las técnicas y recomendaciones en la elaboración de tablas o cuadros.
- Poder analizar y sacar conclusiones sobre las informaciones que se tienen, tanto en tablas o en gráficas
- Distinguir la importancia y diferencias entre el texto, cuadros y gráficas.

GENERALIDADES

La *distribución de frecuencias*, es un método utilizado para organizar y resumir información. Bajo este método, los datos recolectados se ordenan y clasifican, indicándonos la frecuencia o sea el número de veces que se repiten.

Se podrá decir también, que nos permite manejar grandes cantidades de información en espacios reducidos, en formas de cuadros o tablas, complementadas con gráficas.

Por *población o universo,* se entiende como un conjunto de medidas para ser aplicadas a una característica cuantitativa, o como el recuento de todas las unidades que presentan una característica común, siendo ésta cualitativa. También se puede definir a la población como un conjunto de elementos o unidades. Lo que se estudia en una *unidad o elemento* son sus *características*.

Cuando se toman todas las unidades o elementos de la población, se habla de una *investigación exhaustiva* o censo. Si sólo se investiga una parte, se le considera como *investigación parcial o muestra.*

La *muestra,* para que sea representativa de la población, requiere que las unidades o elementos sean seleccionadas al *azar*, en tal forma que cada una de ellas tenga la misma posibilidad de ser seleccionada.

Se usan letras mayúsculas o letras del alfabeto griego como símbolos en poblaciones, en cambio, en la muestras se emplean letras minúsculas.

Los caracteres de los *elementos* de una población pueden ser *cualitativos o cuantitativos*, Los datos cualitativos, denominados también *atributos*, son todos aquellos elementos que pueden ser descritos cualitativamente, es decir mediante palabras; son ejemplos de atributos: la clasificación de los alumnos de una universidad por lugar de origen; clasificación de un grupo de personas por ocupación, por cargo, por sexo, etc.

Los caracteres *cuantitativos* denominados *variables*, son todas aquellas características susceptibles de ser expresadas cuantitativamente, es decir, mediante números. Ejemplo: peso, estatura, edad, número de hijos, salarios, etc.

Las variables se dividen en *discretas y continuas.* Es de tener en cuenta que esta clasificación tiene más valor teórico que práctico.

Las *variables discretas* son aquéllas que admiten solamente valores enteros, es decir, no tienen valores intermedios. Ejemplo: el número de hijos por familia será discreta, ya que no se podrá decir que una familia tiene dos hijos y medio; el número de empleados por departamento en una empresa, etc.

Las *variables continuas* son aquéllas que admiten valores fraccionarios, pudiéndose establecer intervalos. Ejemplo: la estatura de una persona que mide un metro con setenta centímetros; que pesa sesenta kilos, una libra y cuatro onzas, etc.

> *Nota*: Se hará un pequeño resumen sobre aquellos términos que más serán utilizados en los capítulos posteriores y que fueron presentados en forma amplia en el anterior capítulo.

ELABORACIÓN DE TABLA DE FRECUENCIAS

VARIABLE DISCRETA

Ante todo es conveniente familiarizarnos con ciertos símbolos que utilizaremos tanto para la variable discreta como para la variable continua.

$\overline{n}$	=	n	=	Tamaño de la muestra.
N	=	N	=	Tamaño de la población o universo.
X_i	=	x_i	=	Identificación para cada valor observado. (minúscula en la muestra)
f_i	=	n_i	=	Frecuencias absolutas.
f_i/n	=	h_i	=	Frecuencias relativas.
F_i	=	N_i	=	Frecuencias absolutas acumuladas.
H_i	=	H_i	=	Frecuencias relativas acumuladas.
X_i	=	y_i	=	Identifica la variable discreta o las marcas de clase en la continua.
X'_{i-1}-X'_i	=	y'_{i-1}- y'_i	=	Identifica a la variable continua con sus intervalos.
i	=	c	=	Amplitud del intervalo
$\underline{m}$		m	=	Número de valores de la variable o de intervalos.

Se ha considerado conveniente presentar las dos simbologías más utilizadas en los textos de Estadística, de tal manera que el estudiante pueda consultar o aplicar indistintamente cualquiera de ellas.

Ejemplo 1. Supongamos que se tiene una población constituida por 2.000 cajas y deseamos examinarlas con el fin de determinar el número de piezas o elementos defectuosos que contiene cada caja. Por diferentes razones, se desea que la investigación no sea exhaustiva, es decir, no examinar la totalidad de las 2.000 cajas o universo sino, por el contrario, seleccionar una muestra de tamaño 30, correspondiente a una investigación parcial. $N = 2.000$ $n = 30$

Los resultados obtenidos en esta encuesta, se anotan a continuación. Siendo x_1, la primera caja examinada y 3, el número de piezas defectuosas encontradas en ella y así sucesivamente:

$x_1 = 3$	$x_2 = 2$	$x_3 = 0$	$x_4 = 2$	$x_5 = 3$	$x_6 = 3$	$x_7 = 1$	$x_8 = 1$	$x_9 = 0$
$x_{10} = 1$	$x_{11} = 3$	$x_{12} = 3$	$x_{13} = 4$	$x_{14} = 4$	$x_{15} = 3$	$x_{16} = 2$	$x_{17} = 4$	$x_{18} = 2$
$x_{19} = 4$	$x_{20} = 2$	$x_{21} = 4$	$x_{22} = 0$	$x_{23} = 0$	$x_{24} = 1$	$x_{25} = 3$	$x_{26} = 4$	$x_{27} = 2$
$x_{28} = 3$	$x_{29} = 1$	$x_{30} = 2$						

En la anterior información o conjunto de datos, observamos que no están ordenados y los denominamos d*atos originales* o *no ordenados*. Quizás, dado que la muestra es muy pequeña, podemos darnos cuenta, en forma muy ligera del comportamiento de la característica. Pero si en vez de tener 30 observaciones, fueran 200 o 2.000 es muy poco lo que podemos decir o conocer de la variable; por lo tanto se hace necesario el ordenamiento o la clasificación de esos datos, mediante un proceso que llamaremos *tabulación* o *procesamiento de datos.*

En primer lugar, se determinan los valores que toma la variable, en cuanto al número de piezas defectuosas, siendo en este caso de 0 1 2 3 4. También se hubiera podido establecer los valores extremos (0 y 4) y luego los valores intermedios (1, 2, 3); procedemos en seguida al conteo, a fin de determinar cuántas veces se repite cada valor de la variable, resultados que denominaremos *frecuencias o frecuencias absolutas.*

Tabla 2.1 Tabulación

NÚMERO DE PIEZAS	NÚMERO DE CAJAS	TABULACIÓN	
		A	B
0	4	IIII	□
1	5	卌	⊿
2	7	卌 II	⊿ ∟
3	8	卌 III	⊿ ⊔
4	6	卌 I	⊿ I
Σ	30	30	30

En el ejemplo 1 de la Tabla 2.1 se puede interpretar, diciendo que en la muestra de 30 cajas encontramos cuatro de ellas sin piezas defectuosas; 5 cajas con una; 7 cajas con dos; 8 cajas con tres; y 6 con cuatro piezas defectuosas. Sin embargo, a veces, es mejor expresarlo en términos porcentuales, diciendo que el 13% de las cajas examinadas no tenían piezas defectuosas, que el 17% tenían una; que el 23%, dos; el 27%, tres; y el 20% con cuatro piezas o unidades defectuosas, como se puede observar en el cuadro siguiente:

Se recomienda formar grupos de cinco, 卌 卌 卌 III que nos representará 18 observaciones, siendo más fácil contar de cinco en cinco, que hacerlo de uno en uno, IIIIIIIIIIIIIIIIII. También se pueden formar cuadros de cinco, lo cual nos quedaría así: ⊿ ⊿ ⊿ ⊔ para un total de 18.

En el caso en que la variable no emplee valores en forma continua, solamente se debe considerar en la distribución aquellos que aparecen en la base de datos, como se puede observar en la Tabla 2.2, donde se omiten los valores 2 4 6 7 y 9.

Tabla 2.2 Distribución de frecuencias

NÚMERO DE PIEZAS DEFECTUOSAS
0
1
3
5
8
10

Tabla 2.3

Distribución de frecuencias - variable discreta

y_i	n_i	h_i	N_i	H_i
0	4	0,13	4	0,13
1	5	0,17	9	0,30
2	7	0,23	16	0,53
3	8	0,27	24	0,80
4	6	0,20	30	1,00
Σ	30	1,00	-	-
X_i	f_i	f_i/n	F_i	H_i

y_i	n_i	h_i	N_i	H_i
y_1	n_1	h_1	N_1	H_1
y_2	n_2	h_2	N_2	H_2
y_3	n_3	h_3	N_3	H_3
y_4	n_4	h_4	N_4	H_4
y_5	n_5	h_5	N_5	H_5
Σ	n	1,00	-	-
X_i	f_i	f_i/n	F_i	H_i

Las frecuencias relativas se obtienen de la siguiente manera:

$$h_1 = \frac{n_1}{n} = \left[\frac{f_1}{n}\right] \qquad h_2 = \frac{n_2}{n} = \left[\frac{f_2}{n}\right] \qquad h_3 = \frac{n_3}{n} = \left[\frac{f_3}{n}\right] \qquad h_4 = \frac{n_4}{n} = \left[\frac{f_4}{n}\right] \qquad h_5 = \frac{n_5}{n} = \left[\frac{f_5}{n}\right]$$

$$\left[f_1/n\right] = h_1 = \frac{4}{30} = 0{,}13 \qquad \left[f_2/n\right] = h_2 = \frac{5}{30} = 0{,}17 \quad \text{... etc.}$$

Las *frecuencias absolutas acumuladas* las utilizaremos más adelante para el cálculo de algunas medidas. Se obtienen así:

$$\begin{aligned}
N_1 &= n_1 && = N_1 = n_1 = 4\\
N_2 &= n_1 + n_2 && = N_1 + n_2 = 4 + 5 = 9\\
N_3 &= n_1 + n_2 + n_3 && = N_2 + n_3 = 9 + 7 = 16\\
N_4 &= n_1 + n_2 + n_3 + n_4 && = N_3 + n_4 = 16 + 8 = 24\\
N_5 &= n_1 + n_2 + n_3 + n_4 + n_5 && = N_4 + n_5 = 24 + 6 = 30
\end{aligned}$$

$$\begin{aligned}
F_1 &= f_1 && = f_1\\
F_2 &= f_1 + f_2 && = F_1 + f_2\\
F_3 &= f_1 + f_2 + f_3 && = F_2 + f_3\\
F_4 &= f_1 + f_2 + f_3 + f_4 && = F_3 + f_4\\
F_5 &= f_1 + f_2 + f_3 + f_4 + f_5 && = F_4 + f_5
\end{aligned}$$

Las frecuencias relativas acumuladas resultan asi.

$$\begin{aligned}
H_1 &= h_1 && = H_1 = h_1 = 0{,}13 && = F_1/n\\
H_2 &= h_1 + h_2 && = H_1 + h_2 = 0{,}13 + 0{,}17 = 0{,}30 && = F_2/n\\
H_3 &= h_1 + h_2 + h_3 && = H_2 + h_3 = 0{,}30 + 0{,}23 = 0{,}53 && = F_3/n\\
H_4 &= h_1 + h_2 + h_3 + h_4 && = H_3 + h_4 = 0{,}53 + 0{,}27 = 0{,}80 && = F_4/n\\
H_5 &= h_1 + h_2 + h_3 + h_4 + h_5 && = H_4 + h_5 = 0{,}80 + 0{,}20 = 1{,}00 && = F_5/n
\end{aligned}$$

Ejemplo 2. Con la información de la tabla No. 2.3 determinar los valores que toma:

Solución

$$y_3 = \left[X_3\right] = 2 \qquad N_2 = \left[F_2\right] = 9 \qquad n_5 = \left[f_5\right] = 6$$

$$H_3 = \left[\frac{F_3}{n}\right] = 0{,}53 \qquad h_4 = \left[\frac{f_4}{n}\right] = 0{,}27 \qquad x_{16} = 2$$

VARIABLE CONTINUA

Ejemplo 3. Consideremos que se seleccionó una muestra de 30 alumnos, a fin de conocer su peso en kilos; para facilitar el trabajo redondeamos las cifras.

$x_1 = 74$	$x_2 = 67$	$\boxed{x_3 = 94}$	$x_4 = 70$	$x_5 = 69$	$x_6 = 61$	$x_7 = 71$	$x_8 = 79$	$x_9 = 47$
$x_{10} = 85$	$x_{11} = 82$	$x_{12} = 55$	$x_{13} = 65$	$x_{14} = 88$	$x_{15} = 52$	$x_{16} = 58$	$x_{17} = 76$	$x_{18} = 57$
$x_{19} = 72$	$x_{20} = 66$	$x_{21} = 48$	$x_{22} = 56$	$x_{23} = 63$	$x_{24} = 71$	$x_{25} = 60$	$x_{26} = 64$	$x_{27} = 68$
$x_{28} = 83$	$x_{29} = 74$	$x_{30} = 92$						

El primer paso a seguir consiste en determinar el valor máximo y el mínimo que toma la variable. En este ejemplo tenemos: $x_{max} = 94$ y $x_{min} = 47$. La diferencia entre el valor máximo y el mínimo se denomina recorrido. En el presente caso el recorrido es de 47.

$$x_{máx} - x_{mín} = \text{Recorrido o rango}$$

$$94 - 47 = 47$$

Introducimos dos nuevos símbolos que son: *m = número de intervalos y*

C = amplitud del intervalo = $\lceil i \rfloor$

El valor de *m* se puede determinar de varias maneras:

- Un número arbitrario que sea mayor o igual a cinco, y menor o igual a 16
  $5 \leq m \leq 16$ ó $5 \leq m \leq 18$
- Métodos muy utilizados en clase, consiste en la aplicación de la fórmula Sturges
  $m = 1+3{,}3 \log n$
- De acuerdo a la clasificación que requiera la variable o la forma en que se ha venido presentando la información.
- Algunos utilizan un procedimiento, no muy recomendado, calculando *m* mediante la fórmula $m = \sqrt{\log n}$

En la práctica, se determina *m* atendiendo a varios factores, tales como: finalidad del estudio, grado de variabilidad de los datos, necesidad de efectuar comparaciones con otros estudios que obligue a mantener los intervalos y su número. Sin embargo, se recomienda que el número de intervalos, hasta donde sea posible, no sea menor de 5 ni mayor de 16. Algunos lo admiten hasta 18.

En este ejercicio se tendrá que $m = 1 + 3{,}3 \log 30 = 5{,}29 \simeq 6$, lo aproximamos a 6, pero se hubiera podido considerar que $m = 5$. La aproximación no tiene importancia alguna en estos casos. Esta fórmula recibe el nombre de Sturges, su descubridor.

En cuanto la amplitud $\lceil C \rfloor$ o $\lceil i \rfloor$ que debe tomar cada intervalo de la distribución, dependerá del criterio establecido para presentar la información. Puede variar el valor de la amplitud para cada intervalo, sin embargo en clase y con fines de explicación es mejor trabajar con una amplitud constante, para ello se tendrá que:

$$i = C = \frac{Rango}{m}$$

Reemplazando se tiene que:
C = 47 ÷ 6 = 7,83. Siempre aproximamos al número inmediatamente superior, por pequeña que sea la fracción. Si *C* hubiese resultado 7,06 o 7,10, se aproxima a 8. Si usted no aproxima al número superior por pequeña que sea la fracción al calcular amplitud, es muy posible que le queden valores observados por fuera del rango.

Para nuestro ejercicio fijamos m igual a 6, con lo cual podemos determinar el valor de $\boxed{C} = \lceil i \rceil$

$$C = \frac{X_{máx} - X_{mín}}{m} \qquad C = \frac{97 - 47}{6} = \frac{47}{6} = 7{,}83 \qquad C = \lceil i \rceil = 7{,}83 \simeq 8$$

Siempre que el resultado de *C* sea un valor entero con alguna fracción, por pequeña que ésta sea, debe ser aproximada siempre al valor inmediatamente superior; según esto, el valor de *C* será de 8, con el cual el recorrido se aumenta en una unidad, puesto que 6 x 8 = 48. La unidad de incremento se le puede restar al límite inferior o aumentarse al límite superior. En este caso se lo restamos a 47, quedando el límite inferior en 46, a partir de este valor comenzamos a agregar el valor de *C*, para formar los distintos intervalos.

En la variable continua. Se designa con $\boxed{y'_{i-1} - y'_i}$ ó $\boxed{X'_{i-1} - X'_i}$, la columna correspondiente a los distintos valores que toma la variable. Al primer valor $\boxed{y'_{i-1}}$ ó $\boxed{X'_{i-1}}$ se denomina límite inferior del intervalo y $\boxed{y_i'}$ ó $\boxed{X_i'}$ al límite superior; la diferencia entre estos límites corresponde al valor de la amplitud del intervalo $\boxed{C}$ ó $\boxed{i}$. Es preferible con fines de trabajo, que la amplitud sea constante, sin embargo, en la práctica no es muy frecuente que esto se presente.

La tabla de frecuencias para este ejercicio nos quedará de la siguiente forma:

Tabla 2.4

Distribución de frecuencias - variable continua

$y'_{i-1} - y'_i$	n_i	h_i	N_i	H_i	y_i	$y'_{i-1} - y'_i$	n_i
46,1 – 54	3	0,10	3	0,10	50	$y'_0 - y'_1$	y_1
54,1 – 62	6	0,20	9	0,30	58	$y'_1 - y'_2$	y_2
62,1 – 70	8	0,27	17	0,57	66	$y'_2 - y'_3$	y_3
70,1 – 78	6	0,20	23	0,77	74	$y'_3 - y'_4$	y_4
78,1 – 86	4	0,13	27	0,90	82	$y'_4 - y'_5$	y_5
86,1 – 94	3	0,10	30	1,00	90	$y'_5 - y'_6$	y_6
Σ	30	1,00			-	Σ	-
$X'_{i-1} - X'_i$	f_i	f_i/n_i	F_i	H_i	X_1	$X'_{i-1} - X'_i$	f_i

Es de anotar que a cada límite inferior del intervalo se le ha agregado 0,1, lo que permite una fácil tabulación y así se sabe a cuál intervalo corresponde el valor de x_i, obtenido. **El 0,1 se omitirá para el cálculo de las marcas de clase y de los diferentes estadígrafos o medidas que se requiera calcular.**

El cuadro de tabulación con el cual se elaboró la tabla de frecuencias, se estableció de la siguiente forma:

Tabla 2.5

Tabulaciones de datos - variable continua

$y'_{i-1} - y'_i$	n_i	TABULACIÓN	
		A	B
46,1 – 54	3	𝍸	𝍩
54,1 – 62	6	𝍸 𝍷	𝍩 𝍷
62,1 – 70	8	𝍸 𝍸	𝍩 𝍩 𝍷
70,1 – 78	6	𝍸 𝍷	𝍩 𝍷
78,1 – 86	4	𝍸	𝍷𝍷
86,1 – 94	3	𝍸	𝍩
Σ	30	30	30

Las *marcas de clase* se pueden obtener de 3 maneras:

a) Se promedia cada intervalo, de la siguiente forma:

$$\left[X_1\right] = y_1 = \frac{y'_0 + y'_1}{2} = \frac{46+54}{2} = 50 = \left[\frac{X'_0 + X'_1}{2}\right]$$

$$\left[X_2\right] = y_2 = \frac{y'_1 + y'_2}{2} = \frac{54+62}{2} = 58 = \left[\frac{X'_1 + X'_2}{2}\right]$$

$$\left[X_3\right] = y_3 = \frac{y'_2 + y'_3}{2} = \frac{62+70}{2} = 66 = \left[\frac{X'_2 + X'_3}{2}\right]$$

$$\left[X_4\right] = y_4 = \frac{y'_3 + y'_4}{2} = \frac{70+78}{2} = 74 = \left[\frac{X'_3 + X'_4}{2}\right], \text{ etc.}$$

b) Se obtiene la primera marca de clase por el método anterior y si la amplitud C es constante, se le suma a la primera marca de clase obtenida y así sucesivamente.

$$y_1 = \frac{y'_0 + y'_1}{2} = \frac{46+54}{2} = 50 = \left[\frac{X'_0 + X'_1}{2}\right] \qquad \left[X_2\right] = X_2 = 50 + 8 = 58 \qquad \left[X_3\right] = X_3 = 58 + 8 = 66$$

c) Se divide la amplitud de cada intervalo por dos y se le suma al límite inferior del intervalo o se le resta al límite superior del intervalo.

$$\left[\frac{i}{2}\right] = \frac{C}{2} = \frac{8}{2} = 4$$

$$y'_0 + \frac{C}{2} = y_1 \qquad \left[X_1\right] = y_1 = 46 + 4 = 50 \qquad \left[X_2\right] = y_2 = 54 + 4 = 58, etc.$$

$$y'_1 - \frac{C}{2} = y_1 \qquad \left[X_1\right] = y_1 = 54 - 4 = 50 \qquad \left[X_2\right] = y_2 = 62 - 4 = 58, etc.$$

Nota: una distribución es simétrica cuando las frecuencias absolutas o relativas, equidistantes a un punto central son iguales, además, la gráfica que la representa tiene forma acampanada o convexa como se observa con los datos que se presentan a continuación:

Tabla 2.6

A continuación se presenta dos ejemplos de Distribuciones simétricas:

a)

$y'_{i-1} - y'_i$	n_i	h_i
46,1 – 54	2	0,05
54,1 – 62	6	0,15
62,1 – 70	12	0,30
70,1 – 78	12	0,30
78,1 – 86	6	0,15
86,1 – 94	2	0,05
Σ	40	1,00
$X'_{i-1} - X'_i$	f_i	f_i/n

b)

y_i	n_i	h_i
2	2	0,10
4	5	0,25
6	6	0,30
8	5	0,25
10	2	0,10
Σ	20	1,00
X_i	f_i	f_i/n

Propiedades de las frecuencias

- Las frecuencias absolutas son siempre valores enteros.
- La suma de las frecuencias absolutas es igual a n.

$$\sum_{i=1}^{n} n = n \qquad \sum_{i=1}^{5} n = n = n_1 + n_2 + n_3 + n_4 + n_5 = n$$

$$\sum_{i=1}^{5} n_i = 30 = 4 + 5 + 7 + 8 + 6 = 30$$

$$\sum_{i=1}^{n} f_i = n \qquad \sum_{i=1}^{5} f_i = f_1 + f_2 + f_3 + f_4 + f_5 = n \qquad \sum_{i=1}^{5} f_i = 30$$

- Las frecuencias relativas son siempre valores fraccionarios.

$$0 < h_i < 1 \qquad 0 < f_i/n < 1$$

- La suma de las frecuencias relativas es igual a 1

$$\Sigma h_i = 1 \qquad \sum_{i=1}^{5} h_i = h_1 + h_2 + h_3 + h_4 + h_5 = 1$$

$$\sum_{i=1}^{5} h_i = 0{,}13 + 0{,}17 + 0{,}23 + 0{,}27 + 0{,}20 = 1{,}00$$

$$\Sigma f_i/n = 1 \qquad \sum_{i=1}^{5} f_i/n = \frac{f_1}{n} + \frac{f_2}{n} + \frac{f_3}{n} + \frac{f_4}{n} + \frac{f_5}{n} = 1$$

- El último valor de las frecuencias absolutas acumuladas es igual a n.

$N_m = n$ $\quad N_5 = n \quad 30 = 30$ $\quad F_m = n$ $\quad F_5 = n = 30$

- El último valor correspondiente a las frecuencias relativas acumuladas debe ser igual a 1.

$H_m = 1$ $\quad H_5 = 1$ $\quad F_m/n = 1$ $\quad F_6/n = 1$

Ejemplo 3.

a. Indique dos ejemplos de distribuciones unidimensionales, dos de bidimensionales y dos de multidimensionales, relacionadas con la economía del país. b.Señale cuatro distribuciones de variable discreta, tres de variable continua y tres de atributos.

Solución:

a.	Unidimensional (univariante)	1. Producción
		2. No. de explotaciones agrícolas
		3. Ingresos
	Bidimensional	1. Producción e inversión (bivariante)
		2. No. de explotaciones agrícolas y hectáreas cultivadas.
		3. Ingresos y gastos.
	Pluridimensional (polivariante)	1. Producción; inversión; horas laborales
		2. No. de explotaciones agrícolas, hectáreas cultivadas número de obreros
		3. Ingreso, gasto, No. de familias
b.	Variable discreta	1. No. de hijos por familia
		2. No. de hermanos por alumno.
		3. No. de billetes por numeración
		4. No. de piezas defectuosas.
	Variable continua	1. Ingresos
		2. Peso en Kg.
		3. Longitud.
c.	Atributos	1. Exportaciones por puertos colombianos.
		2. Número de alumnos por Universidad en Caracas.
		3. Consumo de cigarrillos por marcas.

Ejemplo 4.

En los ejemplos siguientes señalar con una **A** las series constituidas por datos cualitativos (atributos) y con una **V** los datos cuantitativos (variables).

Solución:

a. Distribución de alumnos por mes de nacimiento (**A**)
b. Distribución de alumnos por nacionalidad (**A**)
c. Distribución de profesionales por estatura y peso. (**V**)
d. Distribución de oficiales por sueldo. (**V**)
e. Distribución de accidentes por causa. (**A**)

Ejemplo 5.

Señale con **C** las series de variable continua y con **D** las de variable discreta.

Solución:

a. Distribución de obreros por salarios (**C)**

b. Distribución de fallecimientos por edades (**C)**

c. Distribución de alumnos por número de hermanos (**D)**

d. Distribución de alumnos por estatura. (**C)**

Ejemplo 6.

Leer el siguiente texto: "*Una vez recolectados los datos en forma ordenada, es necesario presentarlos en forma tal que se facilite su comprensión y su posterior análisis y utilización. Para ello se ordenan en cuadros numéricos y luego se representan en gráficos, para variable discreta mediante diagramas de frecuencias, tanto para absolutas o relativas*".

a. Formar una tabla de frecuencias absolutas (n_i), según el número de letras que componen cada palabra (*rr* y *ll*, debe considerarse una sola letra).

b. Considerando el número de letras por palabra. ¿Qué valor tienen $x_4 \quad x_{27} \quad x_{12} \quad x_{34} \quad x_2$?

c. En la tabla de frecuencias del punto a. ¿Qué valores poseen $y_2 \quad y_4 \quad y_7$?

d. Análogamente ¿Qué valores poseen $n_2 \quad n_4 \quad n_7$?

e. ¿Qué valores poseen $h_2 \quad h_4 \quad h_7$?

f. Calcular las frecuencias relativas acumuladas (H_i) y las frecuencias absolutas acumuladas (N_i).

g. Puede ocurrir que $h_4 > h_7$ (en general). ¿Qué $H_4 < H_7$?

Solución:

x_1 = una = 3 letras $\qquad$ x_2 = vez = 3 letras, etc.

a)

$x_1 = 3$	$x_{11} = 12$	$x_{21} = 2$	$x_{31} = 7$	$x_{41} = 8$
$x_2 = 3$	$x_{12} = 2$	$x_{22} = 9$	$x_{32} = 9$	$x_{42} = 8$
$x_3 = 12$	$x_{13} = 5$	$x_{23} = 8$	$x_{33} = 1$	$x_{43} = 9$
$x_4 = 3$	$x_{14} = 3$	$x_{24} = 1$	$x_{34} = 5$	$x_{44} = 2$
$x_5 = 5$	$x_{15} = 3$	$x_{25} = 11$	$x_{35} = 2$	$x_{45} = 11$
$x_6 = 2$	$x_{16} = 2$	$x_{26} = 4$	$x_{36} = 11$	$x_{46} = 5$
$x_7 = 5$	$x_{17} = 8$	$x_{27} = 3$	$x_{37} = 2$	$x_{47} = 4$
$x_8 = 8$	$x_{18} = 2$	$x_{28} = 2$	$x_{38} = 8$	$x_{48} = 9$
$x_9 = 2$	$x_{19} = 11$	$x_{29} = 7$	$x_{39} = 4$	$x_{49} = 1$
$x_{10} = 9$	$x_{20} = 1$	$x_{30} = 2$	$x_{40} = 8$	$x_{50} = 9$

y_i	n_i	h_i	N_i	H_i
1	4	0,08	4	0,08
2	11	0,22	15	0,30
3	6	0,12	21	0,42
4	3	0,06	24	0,48
5	5	0,10	29	0,58
7	2	0,04	31	0,62
8	7	0,14	38	0,76
9	6	0,12	44	0,88
11	4	0,08	48	0,96
12	2	0,04	50	1,00
Σ	50	1,00	-	-
X_i	f_i	f_i/n	F_i	N_i/n

b) $x_4 = 3 \qquad x_{27} = 3 \qquad x_{12} = 2 \qquad x_{34} = 5 \qquad x_2 = 3$

c) $y_2 = 2 = \boxed{X_2} \qquad y_4 = 4 = \boxed{X_4} \qquad y_7 = 8 = \boxed{X_7}$

d) $n_2 = 11 = \boxed{f_2} \qquad n_4 = 3 = \boxed{f_4} \qquad n_7 = 7 = \boxed{f_7}$

e) $h_2 = 0{,}22 \qquad h_4 = 0{,}06 \qquad h_7 = 0{,}14$

$\boxed{f_2/n = 0{,}22} \qquad \boxed{f_4/n = 0{,}06} \qquad \boxed{f_7/n = 0{,}14}$

f) En general puede ocurrir que $h_4 > h_7$ ó $h_7 > h_4$
Pero siempre se tendrá $H_4 < H_7$ y nunca se dará que $H_4 > H_7$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. En un grupo de familias considerando el número de hijos, se han obtenido los siguientes valores:

2	0	2	4	4	6	6	4	6	7	4	4	7	4	2	0	4	6	7	7

Construir una tabla con base en los anteriores datos. Calcular las columnas correspondientes a: n_i h_i N_i y H_i.

2. Dé contestación a los siguientes puntos, señalando con una *X* si es cierto o falso.

	Cierto	Falso
a. $H_5 = 0{,}36$ $N_4 = 30$ $n_5 = 6$ $n = 50$	( )	( )
b. Para calcular las marcas de clase, se suma el límite inferior al superior del intervalo y se divide entre dos.	( )	( )
c. El número de accidentes según sus causas, es una variable discreta.	( )	( )
d. Si $H_6 = 0{,}7$ y $H_4 = 0{,}3$ un 40% de los valores de la variable es menor que Y_6 y mayor que Y_4.	( )	( )
e. $H_5 = 1{,}10$ $H_4 = 0{,}80$ $h_5 = 0{,}30$	( )	( )
f. La suma de las frecuencias absolutas es igual a 1.	( )	( )

3. Reconstruya la siguiente distribución simétrica.

y_i	n_i	H_i
10	6	0,12
20	-	0,32
30	-	-
40	-	-
50	-	-
X_i	f_i	F_i/n

4. Se realiza un estudio en la ciudad de Guayaquil a 150 familias de clase media, para conocer el tipo de aceite o manteca usados en la cocina. Los resultados son los siguientes: Maíz, 14 hogares; Soya, 65 hogares; Ajonjolí, 21 hogares; compran aceite al detal sin especificar tipo, 17 hogares; manteca de cerdo, 21 hogares; grasas de origen vegetal, 6 hogares; Oliva, 13 hogares.
   a) ¿Cuál es la población?
   b) ¿Cuál es el tamaño de la muestra?
   c) ¿Qué carácter tiene la población?
   d) ¿Cuál es la característica?
   e) ¿Cuántas clases de características tiene la distribución?
   f) Construya una distribución de frecuencias en el mismo orden indicado.
   g) ¿Cuál es la 5ª clase?
   h) ¿Cómo se explica que la suma de frecuencias sea superior al número de hogares?

5. Una fábrica de gaseosas proyecta lanzar al mercado un nuevo sabor. Se realiza un test de aceptación de dicho sabor en una muestra de 30 niños, utilizando una escala de 10 puntos, para medir el grado de aceptación. Los puntos obtenidos en los 30 niños fueron los siguientes:

2 6 8 7 4 5 10 6 6 7 6 7 3 8 7 6 8 6 5 4 7 8 5 7 6 7 2 7 2 7

La muestra estuvo compuesta por igual número de niños de ambos sexos, de 5 a 12 años de edad, residentes en el Barrio San Eduardo de la ciudad de Maracaibo.

a) ¿Cuál es la población?
b) ¿Cuál es la muestra?
c) ¿Cuál es la característica?
d) ¿La característica es cualitativa o cuantitativa?
e) ¿Cuántas clases tiene el atributo?
f) ¿Qué clase de escala se ha utilizado en la medición de la variable?
g) Construya una distribución de frecuencias
h) ¿Por qué la suma de la frecuencias es diferente a 150?

6.Contestar si son ciertas o falsas las siguientes relaciones y por qué.

a. $m = 6$ $h_1 = 0{,}2$ $h_4 = 0{,}2$ $H_2 = 0{,}6$ $H_3 + H_4 = 1{,}9$

$m = 6$ $f_1 = 0{,}2$ $f_4 = 0{,}2$ $F_2/n = 0{,}6$ $F_3/n + F_4/n = 1{,}9$

b. $h_2 = 0{,}20$ $n = 80$ $n_2 = 16$

$f_2/n = 0{,}20$ $n = 80$ $f_2 = 16$

c. $m = 6$ $n_2 + n_4 = 15$ $n_1 + n_3 = 25$ $n = 50$ $n_5 + n_6 = 20$

$m = 6$ $f_2 + f_4 = 15$ $f_i + f_3 = 25$ $N = 50$ $f_5 + f_6 = 20$

7. Determine si las siguientes características son cualitativas o cuantitativas. En el caso de ser cuantitativas, determine si son discretas o continuas.
   a) Almacenes de la localidad de acuerdo a la especialidad (ropa, zapatos, etc.).
   b) La cantidad de artículos para el hogar comprado en el mes, por 30 hogares.
   c) El número de unidades por marca, comprado por las amas de casa en un barrio de la ciudad.
   d) Ocupación de los familiares y amigos.
   e) Pagos mensuales en servicios por los propietarios en un conjunto residencial.
   f) La marca de microcomputadora personal que usan los compañeros de la universidad.
   g) La cantidad de tiempo que utilizan por semana los compañeros de la universidad en la microcomputadora personal.

8. Qué entiende usted por:
   a) Distribución de frecuencias
   b) Variable discreta
   c) Frecuencia absoluta
   d) Frecuencia relativa

9. Una encuesta realizada a 30 fumadores para determinar el número de cigarrillos que encienden (fuman) en un día corriente arrojó los siguientes resultados:

3 7 5 10 8 4 5 8 10 8 8 4 5 3 10 5 7 10 8 5 5 12 8 4 4 3 5 8 12 10

Con los datos anteriores elabore una tabla de frecuencias.

10. Elabore una distribución de frecuencias para atributos; tome para ello cualquier característica que vea en el salón o en el instituto.

11. Diga si son ciertas o no las siguientes afirmaciones. ¿Por qué?
    a. El número de accidentes de trabajo por días de la semana es un ejemplo de variable discreta.
    b. La suma de las frecuencias absolutas acumuladas, debe ser el doble al valor del tamaño de la muestra.
    c.

$H_5 = 0,8$	$h_4 = 0,2$	$H_3 = 0,3$	$h_5 = 0,3$
$F_1/n = 0,8$	$f_4 = 0,2$	$F_3/n = 0,3$	$F_5 = 0,3$

12. Con los datos siguientes, elaborar una tabla de frecuencias sabiendo que la distribución es simétrica.

$m = 7$	$c = 10$	$n_1 = 8$	$n_2 + n_5 = 62$	$y_3 n_3 = 1.260$	$h_3 = 0,21$	$H_6 = 0,96$
$m = 7$	$i = 10$	$f_i = 8$	$f_2 + f_5 = 62$	$X_3 f_3 = 1.260$	$f_3/n = 0,21$	$F_6/n = 0,96$

13. Las estaturas (en centímetros) de los socios de un club juvenil de Panamá, son las siguientes:

153	123	129	132	147	138	137	134	131	147
138	128	134	148	125	139	146	145	148	135
152	128	146	143	138	138	122	146	137	151
145	124	132	138	144	141	137	146	138	146
152	136	160	159	157	150	160	142	148	130

    Se pide agrupar los datos en una tabla de frecuencias con 6 intervalos.

14. Los siguientes datos corresponden a la distribución de frecuencias de los gastos en publicidad (millones de pesos) de 50 empresas comerciales, durante el último trimestre de 2011. Dichos gastos se agruparon en cuatro clases de amplitud constante, de la cual se sabe:

$y_1 = 3,5$	$y_4' = 8,75$	$n_1 = 4$	$N_2 = 20$	$n_3 = 25$
$X_1 = 3,5$	$X_4' = 8,75$	$f_1 = 4$	$F_2 = 20$	$f_3 = 25$

    Se pide elaborar una tabla de frecuencias.

15. Una empresa dedicada a la producción de ambientadores en forma de atomizadores, realiza una investigación entre amas de casa del barrio El Recuerdo de esta ciudad, para lo cual selecciona 50 de ellas, en cuanto al tiempo de permanencia del aroma en su casa, con los siguientes resultados: 3 horas, 3 amas; 4 horas, 7 amas; 5 horas, 10 amas; 6 horas, 16 amas; 7 horas, 9 amas; 8 horas, 5 amas.
    a) ¿Cuál es la población?
    b) ¿Cuál es la muestra?
    c) ¿Cuál es la característica?
    d) Diga si la característica es cualitativa o cuantitativa.
    e) Diga si la variable es discreta o continua.

16. En una empresa, se hace un estudio sobre el número de horas que el personal deja de laborar durante la semana. Los resultados son: 10 horas, 3 obreros; 3 horas, 7 obreros; 9 horas, 1 obrero; 4 horas, 2 obreros; 8 horas, 5 obreros; 5 horas, 8 obreros; 6 horas, 6 obreros.
    a) ¿Cuál es la población?
    b) ¿Cuál es la variable?

c) ¿Que clase de variable es?
d) ¿Cuántos valores toma la variable?
e) ¿Cuales son los valores de:

y_3	n_2	H_5	h_4

X_3	f_2	F_5/n	f_4/n

17. Con los siguientes datos, correspondientes a los saldos (miles de $) de cuentas pendientes:

77	70	65	62	53	78	41	48	74	63	34	38	69	
79	76	55	59	69	78	75	34	68	56	61	41	83	
54	49	68	48	64	84	74	68	73	69	31	69	78	64

Elabore una tabla de frecuencias, calculando m, mediante la fórmula: $m = 1 + 3{,}3 \log n$
Determine los valores de:

y_3	y_5	H_4	N_5	C
X_3	X_5	F_4/n	F_5	i

18. Con los siguientes datos:

$y_2 = 22$	$y'_5 = 50$	c = constante	$m = 6$
$n = 150$	$n_1 = n_6$	$n_2 = n_5 = n_1 + 5$	$n_3 = n_4 = 30$

$X_2 = 22$	$X'_5 = 50$	i = constante	$m = 6$
$n = 150$	$f_1 = f_6$	$f_2 = f_5 = f_1 + 5$	$f_3 = f_4 = 30$

Elaborar una tabla de frecuencias.

19. Según la regla de Sturges $m = 1 + 3{,}3 \log n$ ¿Cuántas marcas de clase calcularía usted cuando:
a) $n = 50$
b) $n = 200$
c) $n = 1.000$
d) La amplitud del intervalo es la diferencia que hay entre el límite superior e inferior de cada intervalo.

20. ¿Qué opinión le merece a usted, si elaboramos una tabla de frecuencias y utilizamos:
a) Dos marcas o número de clases
b) 28 marcas de clase
c) 12 marcas de clase

21. Contestar si son ciertos o falsos los siguientes puntos:
a) Las frecuencias absolutas son valores que admiten fracciones
b) La suma de las frecuencias relativas puede ser mayor a 1
c) Se dice que $0 > h_1 > 1$
d) El último término de la frecuencia absoluta acumulada es igual a n.

22. Decir si son ciertas o no las siguientes relaciones:
a) $m = 6$ $h_1 = 0{,}2$ $h_4 = 0{,}2$ $H_2 = 0{,}6$ $H_3 + H_4 = 1{,}9$
b) $H_4 = 0{,}20$ $H_5 = 0{,}12$ $h_5 = -0{,}08$
c) $m = 4$ $H_5 + h_6 = H_6$

23. Con el ejercicio 18 ¿Qué porcentaje de los valores:
    a) Son superiores a 42?
    b) Inferiores a 34?
    c) Entre 18 y 50?

RECOMENDACIONES EN LA ELABORACIÓN DE CUADROS O TABLAS PARA INFORMES

Las anteriores tablas de frecuencias nos permiten desarrollar con más facilidad la aplicación de *medidas* de *posición, dispersión, asimetría* y *apuntamiento*, nociones básicas en la estadística descriptiva, sin embargo en los informes puede haber cambios en la forma de presentación de los datos.

Aunque no se pueden fijar normas rígidas en la presentación de los cuadros de informes, si es posible hacer algunas recomendaciones.

- Todo cuadro debe estar numerado, en especial, cuando el informe presenta dos o más cuadros.
- El título debe indicar: ¿qué?, ¿cómo?, ¿dónde? y ¿cuándo?.
- El cuadro consta de un encabezamiento, en el cual van los títulos que deben escribirse en mayúsculas y los subtítulos con minúsculas.
- En el cuerpo del cuadro va la información.
- En un cuadro, ninguna casilla debe quedar en blanco, siendo necesario emplear símbolos.
- El pie, se ubica abajo de la línea inferior del cuadro y se coloca en él, la fuente, nota o llamada que se crean necesarias.
- Por presentación, el cuadro no debe estar cerrado a los lados, sin embargo no se hace.
- El cuadro deberá estar cerrado en la parte superior e inferior con líneas más gruesas.
- Se repetirá el título del cuadro en todas las páginas por él ocupadas, utilizando la palabra continúa, continuación o conclusión, según el caso.
- El cuadro podrá tener subtotales y totales

Cuadro N° Título

MAYÚSCULAS	MAYÚSCULAS			MAYÚSCULAS
	minúsculas	minúsculas	minúsculas	

Encabezamiento

Cuerpo

Pie

Veamos ahora, las clases y la estructura que se debe tener en cuenta en la elaboración de cuadros o tablas de informe.

Los cuadros estadísticos pueden ser clasificados como:

1. De simple entrada
   - Simple
     - Homogéneos
     - Heterogéneos
   - Complejos
2. De doble entrada
   - Corriente
   - Correlación

La presentación de las informaciones obtenidas en encuestas, se puede realizar de varias formas, ya sean aisladas o combinadas, manteniendo el siguiente orden:

a) Textual,
b) Cuadros o tablas,
c) Gráficas.

a) La presentación en *texto o textual*, tiene aplicación bastante limitada; se le utiliza principalmente para informes de empresas, artículos, reportajes y otras publicaciones de igual naturaleza. En este caso, los datos aparecen intercalados con los comentarios o la simple interpretación de ellos.
b) Los *cuadros* o *tablas* corresponden a arreglos sistemáticos de los datos en filas o columnas. Los cuadros son un buen complemento del texto en los informes.

ELABORACIÓN DE GRÁFICAS

Hemos visto que la estadística descriptiva cumple con algunas funciones, ya sea suministrando técnicas para recolectar, ordenar, clasificar mediante cuadros o tablas, de las informaciones obtenidas como las que acabamos de ver, pero por si solas no es suficiente, de ahí la necesidad de complementarlas por medio de *gráficas*, que nos permitan ver con mayor ligereza la información que se ha querido presentar mediante cuadros.

A continuación en términos generales, se dan algunas recomendaciones para la elaboración de gráficas.

- La mejor gráfica es la más simple.
- Las gráficas deben ser tan sencillas y claras, de tal manera que sean comprensibles sin la ayuda de las descripciones del texto.
- Las gráficas deben seguir y nunca preceder a la exposición del texto.
- Las gráficas no sustituyen al cuadro, al contrario deben complementarse.
- Las gráficas más comunes se elaboran teniendo como base los ejes de coordenadas cartesianas.
- La finalidad de las gráficas es visualizar mejor la información.
- La gráfica es considerada como el medio de expresión de la estadística, más llamativa y sugestiva, a la vez que presenta la ventaja de dejar en la memoria una expresión más duradera que los cuadros o el texto, en un menor tiempo de lectura.
- Si el informe tiene dos o más gráficas deberán numerarse.

- Toda gráfica debe tener título que indique con claridad el contenido de la misma. Las líneas y símbolos usados, deben ser los estrictamente indispensables.
- Las líneas que representan los datos o las variables, deben ser más gruesas que las escalas.
- La línea vertical (ordenada) representa las frecuencias, y se debe comenzar de cero.
- Las características cualitativas y cuantitativas, por lo general, van en la línea horizontal (abscisa).
- La lectura de la escala del eje horizontal se hace de izquierda a derecha. La del eje vertical debe hacerse de abajo hacia arriba.
- Cuando la gráfica presenta más de una característica o variable, deberá diferenciarse por medio de leyendas, notas o signos convencionales.
- En toda gráfica se debe explicar la fuente de donde fueron obtenidos los datos; además aclarar las escalas, leyendas, notas y convenciones que ayuden a identificar las características presentadas.

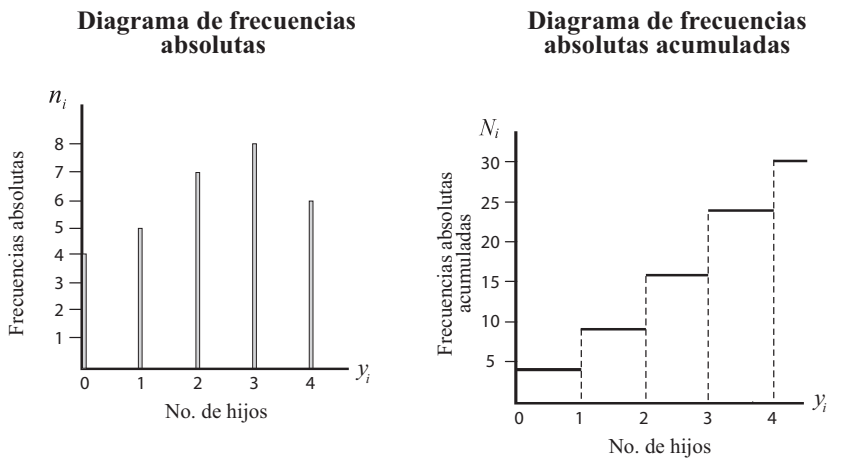
Hemos dividido las gráficas en dos grandes grupos: aquéllas que generalmente se utilizan para el desarrollo de la teoría estadística, tales como los *diagramas de frecuencias y de dispersión, los histogramas, polígonos y ojivas*. El segundo grupo lo conforman las gráficas que frecuentemente vemos en las publicaciones, llámense informes, periódicos, revistas, publicidad, etc., cómo son las *gráficas: circulares,* de *barras,* de *líneas, pictogramas, cartogramas* y *muchas más.*

GRÁFICAS APLICADAS AL DESARROLLO DE LA *TEORÍA ESTADÍSTICA*

Diagramas de frecuencias. Son muy utilizados para representar las frecuencias absolutas y relativas, incluyendo las acumuladas que ocurren con respecto a una variable aleatoria discreta. Se representan por líneas delgadas ya sean verticales u horizontales, colocando las frecuencias.

Ejemplo 1. Con los datos de la tabla 2.1 elaboraremos los respectivos *diagramas* de *frecuencias*.

Figura 2.1



Histograma. Son diagramas de frecuencias unidimensionales, en los cuales se levantan rectángulos de áreas, proporcionales a las frecuencias de clases sobre los intervalos del eje horizontal, por lo tanto es utilizado para representar a la variable continua, cuando la amplitud es constante.

La representación gráfica, para las frecuencias relativas y absolutas en variables continuas, se hace mediante *histogramas de frecuencias*. En el eje de las abscisas se señalan los intervalos que constituyen la base del rectángulo, y en el eje vertical u ordenada corresponde a las frecuencias absolutas o número proporcional de ellas, formando la altura del rectángulo (Figura No. 2.2). Sin embargo, en la confección de este gráfico se pueden presentar grupos irregulares o sea, cuando la amplitud es diferente para cada intervalo, presentando unos espacios pequeños y otros demasiados grandes. Este gráfico aparentemente correcto (Figura No. 2.3), es en realidad falso y engañoso. Para obtener un gráfico correcto es preciso dividir la frecuencia absoluta, por la amplitud del intervalo, es decir:

$$\text{Altura} = \frac{n}{C_i}$$ (Ver figura 2.4 y tabla 2.10)

El histograma de frecuencias se representa por medio de rectángulo o áreas:

Tabla 2.8

Tabla de frecuencias

$y'_{i-1} - y'_i$	n_i	h_i	N_i	H_i
46,1- 54	3	0,10	3	0,10
54,1- 62	6	0,20	9	0,30
62,1- 70	8	0,27	17	0,57
70,1- 78	6	0,20	23	0,77
78,1- 86	4	0,13	27	0,90
86,1-94	3	0,10	30	1,00
Σ	30	-	1,00	-
$X'_{i-1} - X'_i$	f_i	f_i/n	F_i	F_i/n

Figura 2.2

Histograma de frecuencias

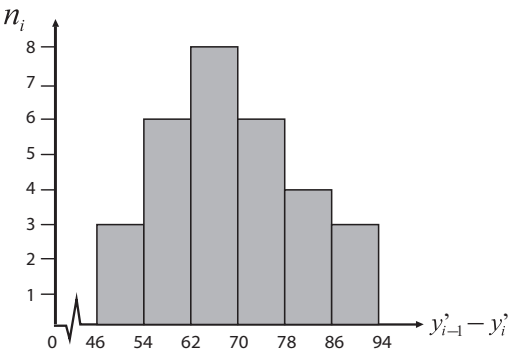


Tabla 2.9

Obreros clasificados por edades en una gráfica

EDAD (AÑOS)	No. OBREROS
16,1 - 25	55
25,1 - 28	47
28,1 - 31	32
31,1 - 40	26
Σ	**160**

Figura 2.3

Histograma elaborado incorrecto

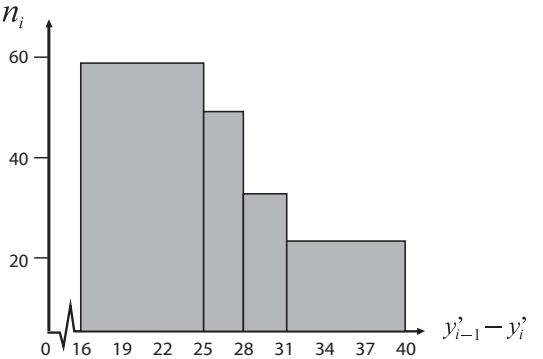


Tabla 2.10

Obreros clasificados por edades en una gráfica

AMPLITUD	FRECUENCIAS	ALTURA
C_i	n_i	n_i / C_i
9	55	6,11
3	47	15,66
3	32	10,66
9	26	2,88
i	f_i	f_i/i

Figura 2.4

Histograma correcto

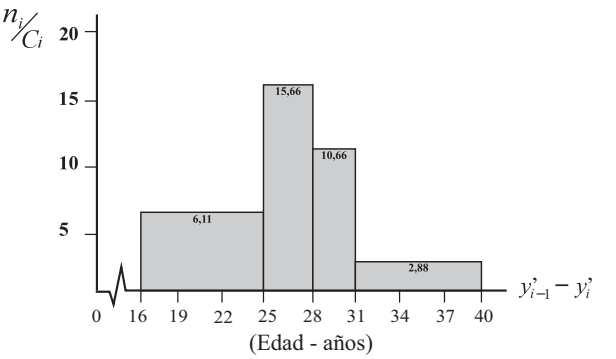
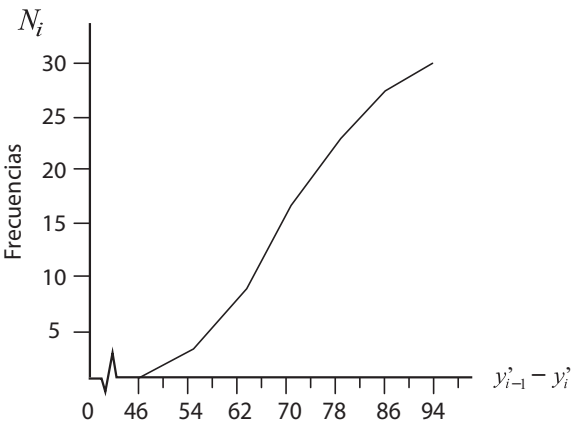


Figura 2.5

Ojiva



Ojiva. La representación gráfica para las frecuencias absolutas y relativas acumuladas en una variable continua se hace a través de una *ojiva ascendente*. Para ello se determinan los puntos de intersección entre cada valor de la variable y su respectiva frecuencia; luego se unen con trazos rectilíneos. (Figura No 2.5). Prácticamente es un *polígono de frecuencias*, con la diferencia de ser aplicado a una tabla de frecuencias absolutas acumuladas.

Polígono de frecuencias. En la variable continua es bastante utilizado este diagrama, fijando puntos, utilizando las marcas de clase y las frecuencias, luego se unen dando una línea quebrada. Si en el histograma de frecuencias unimos los puntos medios en la parte superior de cada rectángulo, obtenemos el *polígono de frecuencias* (figura No. 2.6 y 2.7).

Figura 2.6

Polígono de frecuencias

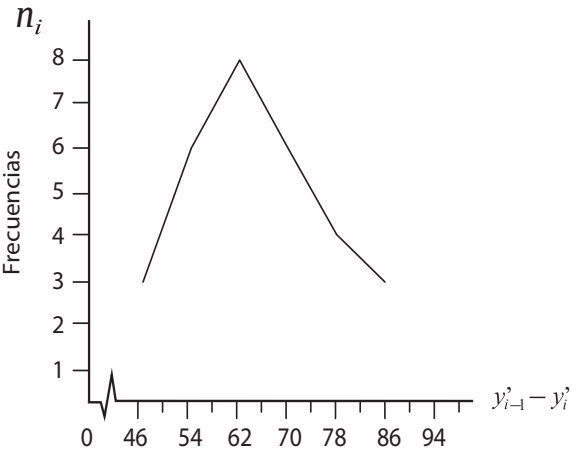
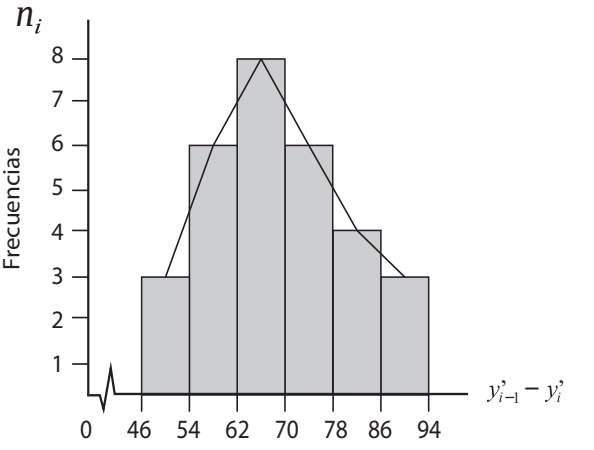


Figura 2.7

Polígono e Histograma



EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

24. Responda a los siguientes puntos, diciendo si son ciertas o falsas las siguientes afirmaciones:
    a) Al hacer una gráfica, la variable a estudiar va en el eje de las abscisas y las frecuencias en la ordenada.
    b) El polígono de frecuencias se emplea para representar a la variable continua.
    c) Entre más características tenga la gráfica, se le considera como la mejor.
    d) Si los intervalos no son iguales, nos facilita la elaboración del histograma por lo tanto será mucho más representativo.

25. Con los datos del ejercicio No. 9 construya los diagramas de frecuencias tanto con las frecuencias absolutas, como con las acumuladas.

26. Con los datos del ejercicio No. 13, se pide elaborar el histograma de frecuencias.

27. Con los datos del ejercicio No. 15, se pide graficar el conjunto de observaciones.

28. Con los datos del ejercicio No. 18, se pide elaborar el histograma de frecuencias.

29. Elaborar un histograma y polígono utilizando la siguiente distribución:

$y'_{i-1} - y'$	40,1 – 52	4,1 – 20	20,1 – 24	24,1 – 32	32,1 – 40		$X'_{i-1} - X'_i$
n_i	30	16	20	10	24	100	f_i

30. Explique brevemente qué entiende usted por:
    a) Marcas de clase
    b) Variable continua
    c) Intervalo de clase
    d) Histograma de frecuencias
    e) Polígonos de frecuencias
    f) Amplitud del intervalo.

31. En una investigación realizada a 800 empresas sobre valor de las ventas, en una quincena, se obtuvo la siguiente distribución.

VENTAS	PORCENTAJE DE EMPRESAS
8 – 18	0,30
18 – 48	0,25
48 – 98	0,18
98 – 148	0,14
148 – 200	0,13

Se pide:
    a) Completar la distribución de frecuencias
    b) Cuántas empresas venden menos de 98 mill $
    c) Elaborar el histograma y polígono de frecuencias.

32. Los ingresos semanales (miles de $) de 50 trabajadores de la construcción fueron los siguientes:

170	172	183	164	166	164	160	168	173	176
184	190	163	151	158	166	166	172	170	186
170	166	169	160	158	155	160	157	163	165
170	172	178	183	170	166	165	172	168	158
158	168	166	159	154	166	172	168	163	172

a) Se pide elaborar una tabla de frecuencias.
b) Construir un histograma, el polígono y la ojiva.

33. Con los siguientes datos sin agrupar, correspondientes a saldos (en miles de pesos) de cuentas pendientes:

77	70	65	62	53	78	41	48	74	63
34	38	69	64	79	76	55	59	69	78
75	34	68	56	61	41	83	54	49	68
48	64	84	74	68	73	69	31	69	78

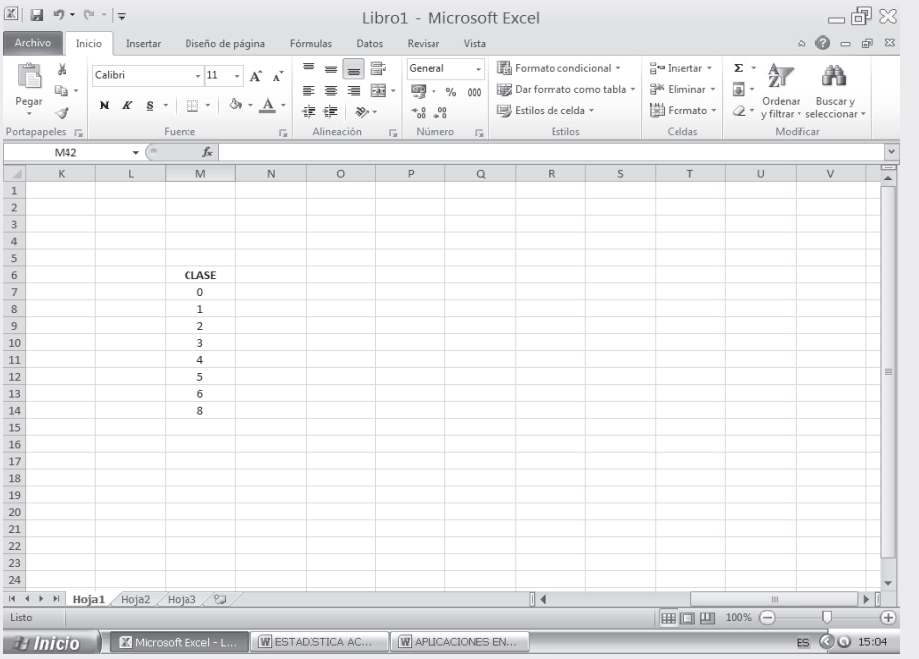
Elaborar una tabla de frecuencias y dibujar la ojiva.

APLICACIONES DE LA FUNCIÓN EXCEL EN LA ELABORACIÓN DE TABLAS O CUADROS

Complemento virtual SIL

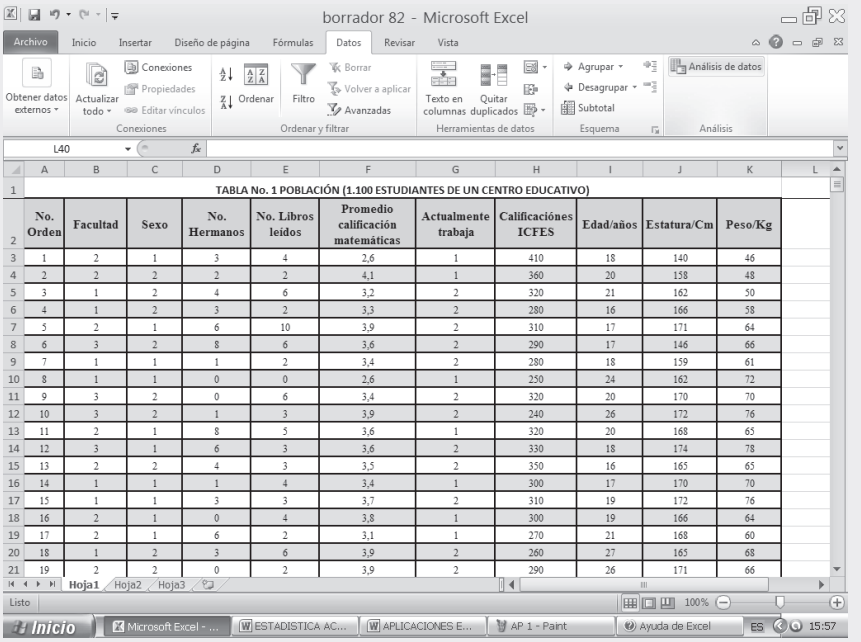
Acudimos a la información que aparece en la tabla No. 1 (ver SIL: Aplicaciones estadísticas en Excel). Para elaborar nuestra tabla, seleccionamos otras celdas donde aparezcan los datos ya organizados, por ejemplo, ubicamos el cursor desde M6 hasta M15, tecleamos el título CLASE y a continuación (debajo) en orden ascen-dente se teclean los valores que toma nuestra variable, siendo 0 1 2 3 4 5 6 8 como lo podemos observar en la aplicación siguiente:

Figura No. 1. Ubicación de Datos



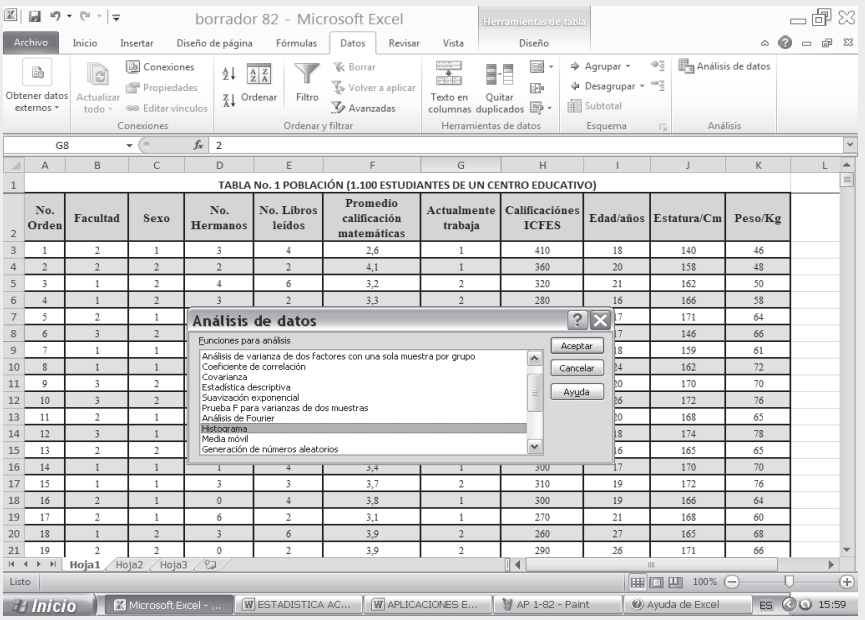
- Guiándonos por la Tabla No. 1 del SIL, realizamos la selección de la información que encontramos en esta tabla y la llevamos a una hoja de Excel.
- Hacemos clic en ANÁLISIS DE DATOS en el grupo ANÁLISIS de la ficha DATOS. Si el comando ANÁLISIS DE DATOS no está disponible, debemos hacer CLIC en el botón y en OPCIONES EN EXCEL seleccionamos COMPLEMENTOS y damos CLIC en IR. A continuación, aparece un cuadro de diálogo en el que activamos la opción HERRAMIENTAS PARA ANÁLISIS y finalmente hacemos CLIC en ACEPTAR. El proceso anterior nos permite ver el ícono ANÁLISIS DE DATOS en la tabla de cálculo, tal como se presenta a continuación:

Figura No. 2. Selección Comando Análisis de Datos



No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
3	1	2	4	6	3,2	2	320	21	162	50
4	1	2	3	2	3,3	2	280	16	166	58
5	2	1	6	10	3,9	2	310	17	171	64
6	3	2	8	6	3,6	2	290	17	146	66
7	1	1	1	2	3,4	2	280	18	159	61
8	1	1	0	0	2,6	1	250	24	162	72
9	3	2	0	6	3,4	2	320	20	170	70
10	3	2	1	3	3,9	2	240	26	172	76
11	2	1	8	5	3,6	1	320	20	168	65
12	3	1	6	3	3,6	2	330	18	174	78
13	2	2	4	3	3,5	2	350	16	165	65
14	1	1	1	4	3,4	1	300	17	170	70
15	1	1	3	3	3,7	2	310	19	172	76
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
18	1	2	3	6	3,9	2	260	27	165	68
19	2	2	0	2	3,9	2	290	26	171	66

Figura No. 3. Opción Histograma

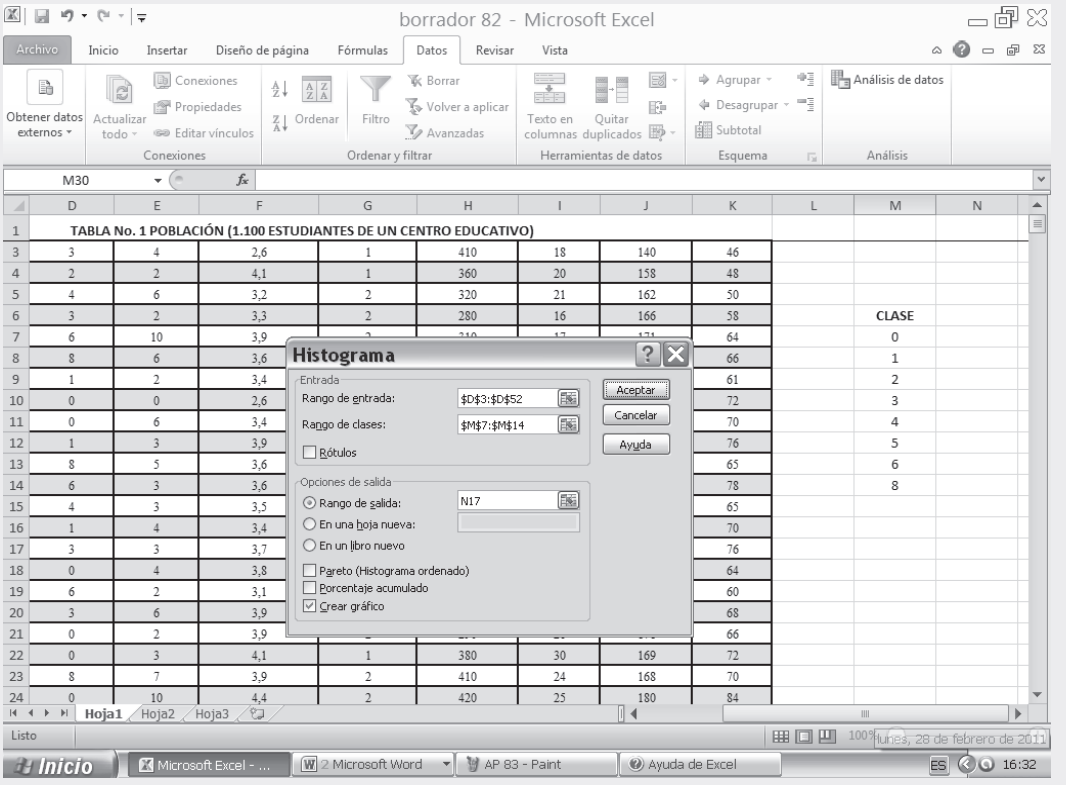


No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
3	1	2	4	6	3,2	2	320	21	162	50
4	1	2	3	2	3,3	2	280	16	166	58
5	2	1						17	171	64
6	3	2						17	146	66
7	1	1						18	159	61
8	1	1						24	162	72
9	3	2						20	170	70
10	3	2						26	172	76
11	2	1						20	168	65
12	3	1						18	174	78
13	2	2						16	165	65
14	1	1	1	4	3,4	1	300	17	170	70
15	1	1	3	3	3,7	2	310	19	172	76
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
18	1	2	3	6	3,9	2	260	27	165	68
19	2	2	0	2	3,9	2	290	26	171	66

Con base en la cuarta columna denominada "Número de hermanos", ubicada en la celda D, que aparece en la Tabla No.1 del ejemplo que seguimos, realizamos el HISTOGRAMA. Nos aparece un cuadro de diálogo denominado HISTOGRAMA, el cual diligenciamos tal como se especifica a continuación:

- En el RANGO DE ENTRADA seleccionamos el rango correspondiente a D3:D52
- Con el cursor nos ubicamos en el RANGO DE CLASES, que lo hemos establecido como M7:M14 (ver figura No. 1)
- Nos situamos en la casilla RANGO DE SALIDA y la activamos haciendo CLIC en el círculo que se antepone. Tecleamos en esta casilla la celda en la cual queremos que se presenten los resultados, en nuestro caso seleccionamos la celda N17, columna que presentará las distribuciones de frecuencias correspondientes.
- Si además del cuadro nos interesa crear el gráfico, EXCEL lo realiza como HISTOGRAMA, cuando en realidad es un gráfico de BARRAS. Para obtener la representación gráfica, se debe seleccionar la opción CREAR GRÁFICO.
- Finalmente hacemos CLIC en ACEPTAR.

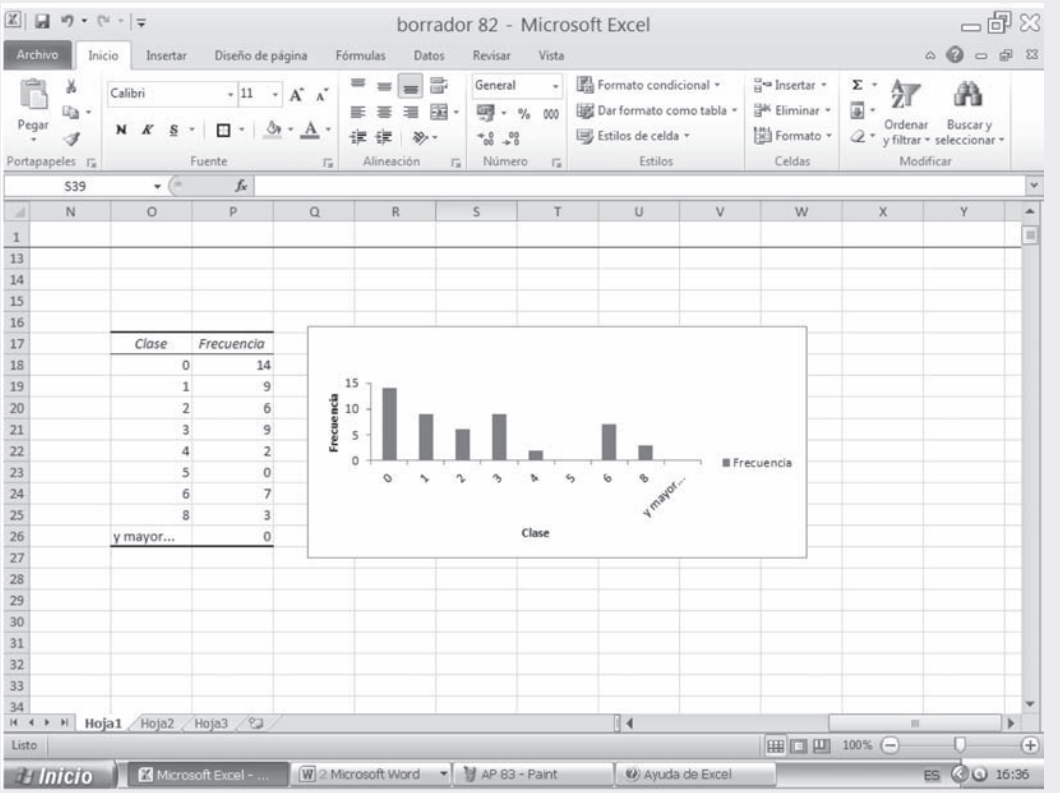
Figura No. 4. Datos de Histograma.



Una vez hecho el CLIC en ACEPTAR; obtenemos en pantalla los resultados esperados, es decir, se tendrá el cuadro de frecuencias con sus respectivos valores de clases, desde N17 hasta N27, a su lado debe aparecer la gráfica por haberse seleccionado la opción CREAR GRÁFICO.

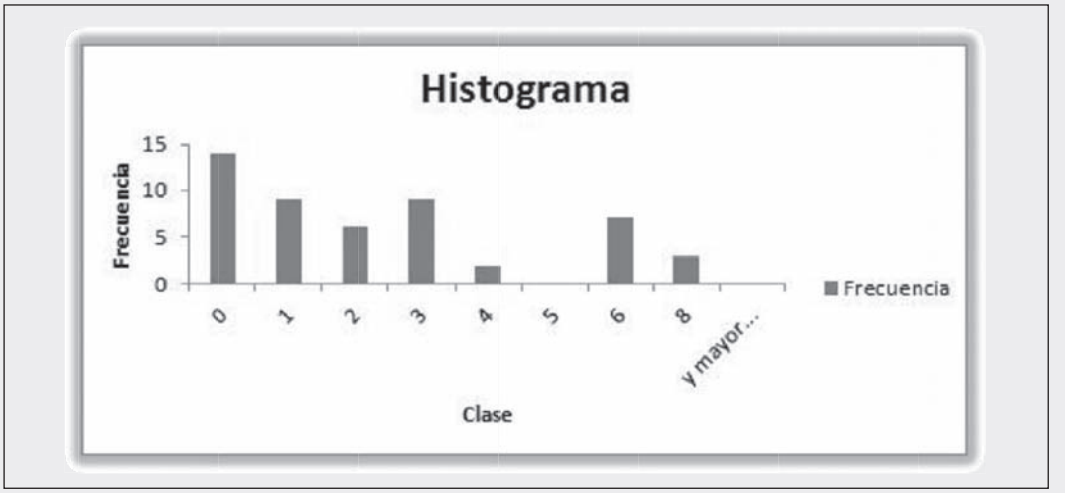
A continuación se presentan los resultados, en el cual aparece la TABLA DE FRECUENCIAS acompañada de un HISTOGRAMA similar a la GRÁFICA DE BARRAS.

Figura No. 5. Resultados Histograma.



Luego de finalizado el procedimiento anterior, se procede a trasladar la figura a un procesador de texto para efectos de presentación.

Gráfica 1. Histograma



Nota:

(1) La línea que puede aparecer en este tipo de gráfico, corresponderá a un DIAGRAMA DE FRECUENCIAS PORCENTUALES, cuando señalamos o hacemos CLIC en PORCENTAJE ACUMULADO.

(2) Observe que Excel incluye una última clase como mayor por si existen valores superiores.

(3) Como se puede observar, en vez de HISTOGRAMA, aparece una gráfica de barras. Se recomienda hacer el histograma en forma correcta utilizando la función gráficas de barras, al principio aparecen barras separadas y utilizando área SERIES ofrece dos alternativas (FILAS) (COLUMNAS) con una de ellas toma la forma adecuada.

GRÁFICAS UTILIZADAS EN LA PRESENTACIÓN DE INFORMES

Los gráficos, son utilizados con el fin de visualizar mejor la información, no sustituyen al CUADRO, antes por el contrario lo complementan. Algunos lo consideran como algo decorativo del informe, pero esto no es cierto, ya que ayuda a mostrar la información mucho más rápida y agradable, que la obtenida a través del TE*X*TO o del mismo CUADRO.

El E*X*CEL es una buena herramienta para la elaboración de GRÁFICOS, ofreciéndonos variadas formas de presentación, tales como:

COLUMNAS; BARRAS; LINEAL; CIRCULAR; DISPERSIÓN; ÁREAS; ANILLOS RADIAL; SUPERFICIE; BURBUJAS; COTIZACIONES

De las anteriores opciones, las primeras cuatro (4) son las más utilizadas, las que a su vez son muy sencillas, para visualizar mejor la información.

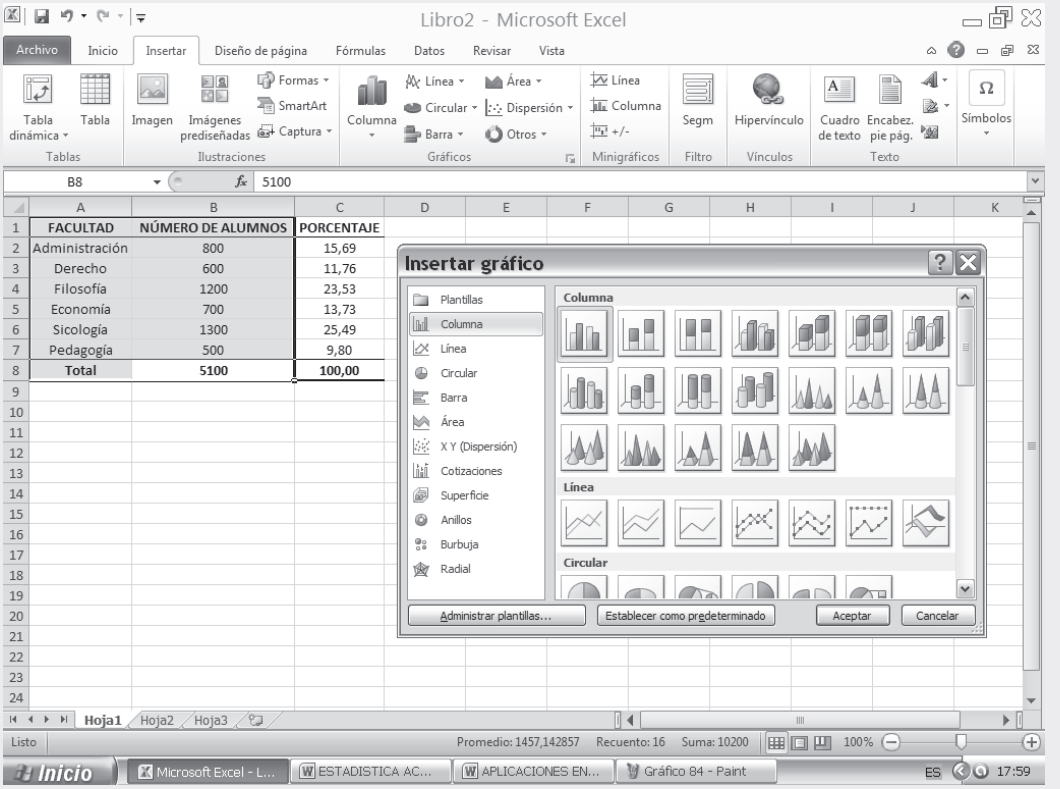
Ejemplo 1. Para indicar el proceso que se sigue en la elaboración de una GRÁFICA consideramos como información (cualitativa) la matrícula de 5.100 alumnos, distribuidos por facultades, en una institución universitaria de la capital.

Cuadro 2.1. Alumnos matriculados por facultad en el instituto X (primer semestre de 2011)

FACULTAD	NÚMERO DE ALUMNOS	%
Administración	800	15,69
Contaduría	600	11,76
Derecho	1.200	23,53
Economía	700	13,73
Ingeniería	1.300	25,49
Psicología	500	9,80
Total	**5.100**	**100,00**

- ❑ Digitamos la anterior información en una hoja de cálculo EXCEL, recomendando aumentar el ancho de las columnas (la primera) a fin de lograr una mejor presentación.
- ❑ Activamos las CELDAS que contienen la información o datos que deseamos graficar, incluidos los TÍTULOS y RÓTULOS. En este caso consideramos las columnas: **Facultades y Número de alumnos.**
- ❑ Procedemos a hacer CLIC sobre la pestaña □ del grupo GRÁFICOS en la ficha INSERTAR. Una forma más rápida consiste en seleccionar directamente sobre el grupo GRÁFICOS el que más se adapte según el tipo de información que tengamos

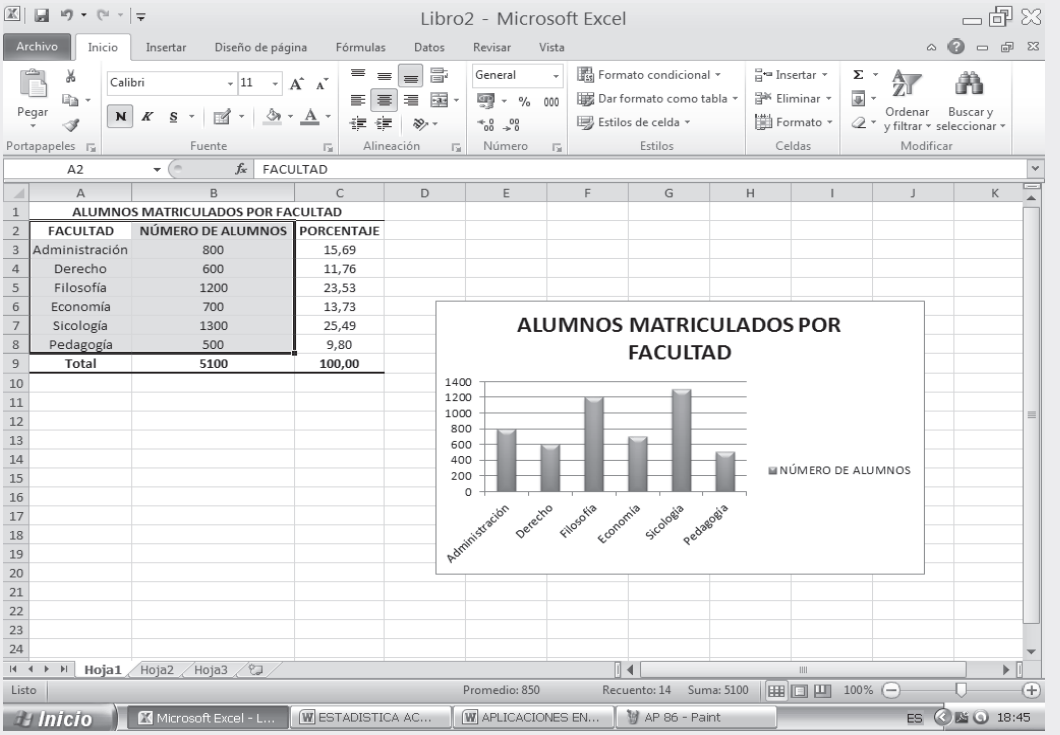
Figura No. 6. Insertar Gráfico.



	FACULTAD	NÚMERO DE ALUMNOS	PORCENTAJE
2	Administración	800	15,69
3	Derecho	600	11,76
4	Filosofía	1200	23,53
5	Economía	700	13,73
6	Sicología	1300	25,49
7	Pedagogía	500	9,80
8	Total	5100	100,00

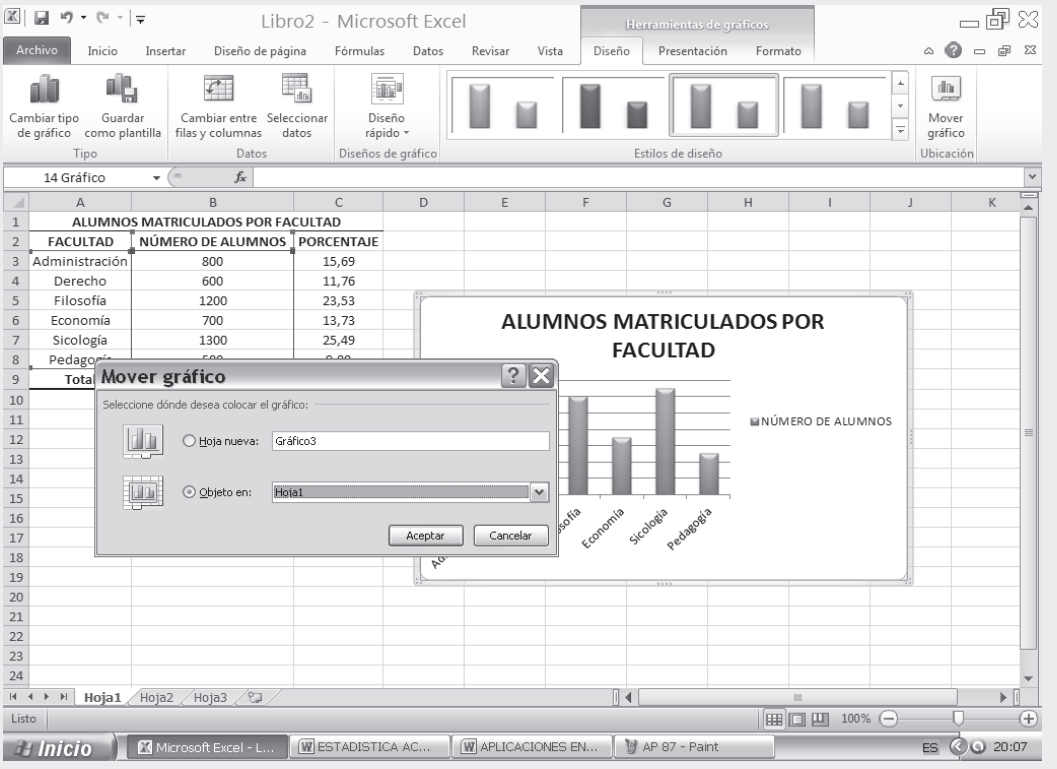
- Los datos pueden estar organizados en FILAS o COLUMNAS. Debemos tener en cuenta el rango de datos que vayamos a seleccionar; en este caso, seleccionamos el rango de A2:B8.
- Seleccionamos el TIPO DE GRÁFICO que deseemos crear, haciendo CLIC en él. En nuestro caso consideramos el gráfico de COLUMNAS, optando por uno de los DISEÑOS de las diecinueve (19) posibilidades que se presentan.
- Después de crear un gráfico, es posible modificar cualquiera de sus elementos en la ficha HERRAMIENTAS DE GRÁFICO, de acuerdo a las pestañas de DISEÑO, PRESENTACIÓN y FORMATO. De este modo, podemos cambiar la forma en que se presentan los ejes, el título del gráfico, mover u ocultar la leyenda así como incluir elementos de gráfico adicionales. Esto nos permite AGREGAR o QUITAR aquello que deseamos que aparezca o desaparezca.
- El gráfico se puede ver de dos formas, como objeto y como una hoja nueva.
- Como una HOJA NUEVA: Mostrando el gráfico en la misma hoja, denominada HOJA DE GRÁFICO, hacemos CLIC sobre el gráfico ya insertado y en la pestaña de DISEÑO, encontramos el fichero UBICACIÓN; Hacemos CLIC en MOVER GRÁFICO y elegimos la opción HOJA NUEVA.
- Como OBJETO, seguimos el procedimiento anterior sólo que esta vez elegimos la opción OBJETO EN, especificando la hoja en la que deseamos insertar el gráfico. En este caso el gráfico no genera una hoja nueva sino un cambio de ubicación, por ejemplo de hoja 1 a hoja 2. Lo anterior se presenta en la figura siguiente:

Figura No. 7. Gráfico de Columnas



ALUMNOS MATRICULADOS POR FACULTAD		
FACULTAD	NÚMERO DE ALUMNOS	PORCENTAJE
Administración	800	15,69
Derecho	600	11,76
Filosofía	1200	23,53
Economía	700	13,73
Sicología	1300	25,49
Pedagogía	500	9,80
Total	5100	100,00

Figura No. 8. Opciones de pegado de Gráfico



- Finalmente, mejoramos la presentación del gráfico, haciendo CLIC en el área del gráfico, le damos opción COPIAR, y luego lo pegamos donde quiera que estemos realizando el informe o trabajo tal como se presenta a continuación:

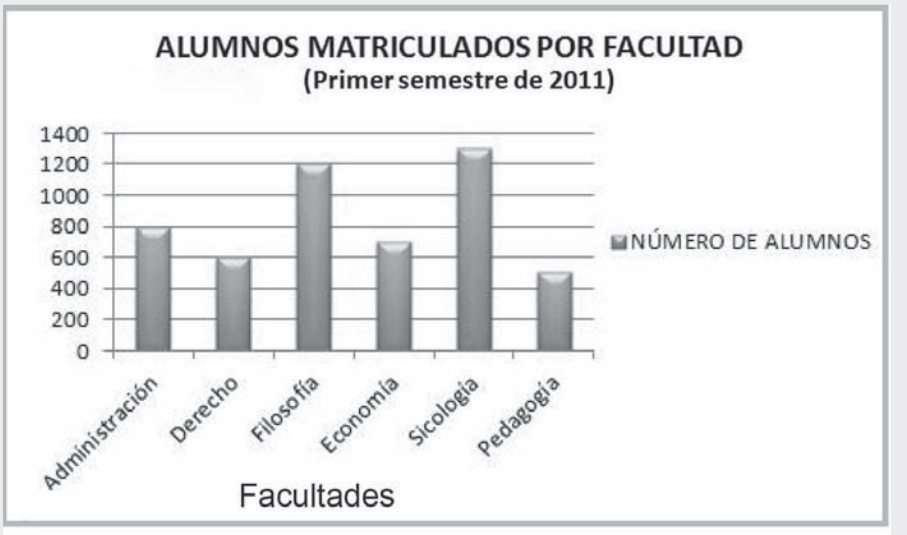


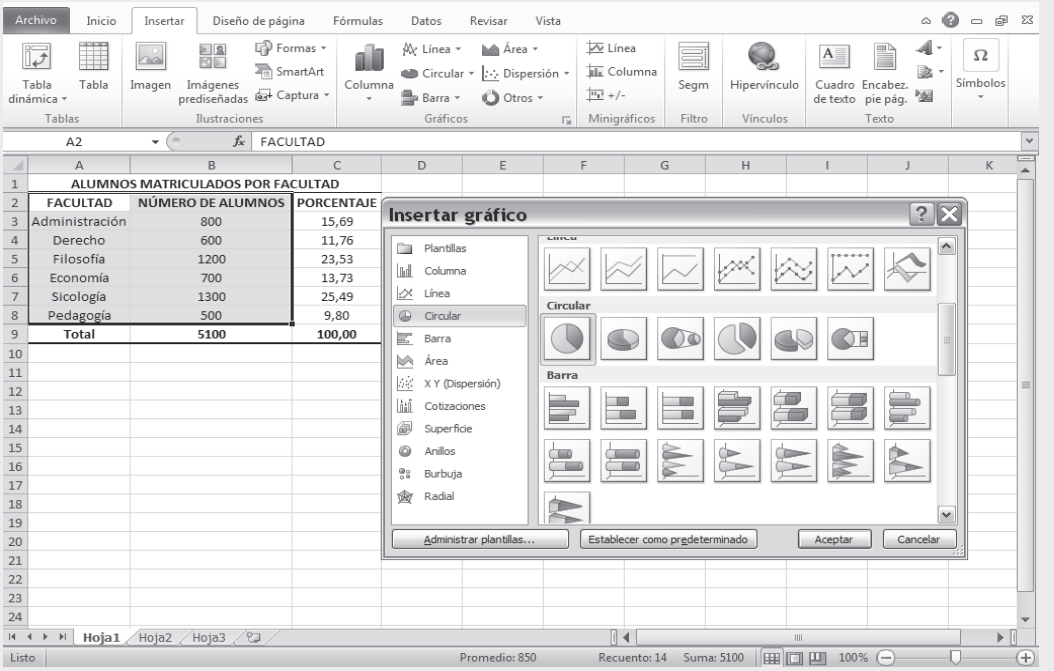
GRÁFICO CIRCULAR O PASTEL

El uso de este gráfico CIRCULAR, se hace muy conveniente cuando la(s) característica(s) que se quieren visualizar son ATRIBUTOS o CUALITATIVOS, tal como en el ejercicio que se realizó anteriormente al aplicar la gráfica de BARRAS, para lo cual consideramos la matrícula de 5100 alumnos en una institución universitaria de la capital.

Vale la pena tener en cuenta que este tipo de gráfico no es muy indicado en aquellos casos donde la característica presenta numerosos aspectos, igual como en el ejercicio que nos ocupa, si la institución tuviera más de 10 facultades, esto haría un gráfico muy recargado y no se lograría una buena visualización de la información. El proceso que se sigue en la graficación es muy similar al de las barras. Veámoslo:

- Tomamos la misma información para la gráfica de barra.
- Activamos las CELDAS respectivas, incluyendo los TÍTULOS, en este caso FACULTADES y NÚMERO DE ALUMNOS.
- Nos ubicamos en la ficha INSERTAR haciendo CLIC en la pestaña GRÁFICOS para cambiar el tipo de gráfico.
- Seleccionamos CIRCULAR y uno de los SUBTIPO DE GRÁFICO. En nuestro ejemplo, optamos por el primero de las seis (6) opciones que nos ofrece y hacemos CLIC en ACEPTAR
- De ahí en adelante seguimos los pasos similares a los mostrados en el GRÁFICO DE BARRAS, de acuerdo con la siguiente presentación, tal como se puede observar en las aplicaciones que aparecen a continuación:

Figura No. 9. Insertar Gráfico.



- Seleccionando la opción HOJA NUEVA, nos resulta la siguiente GRÁFICA en una hoja completa.

Figura No. 10. Resultados de Salida

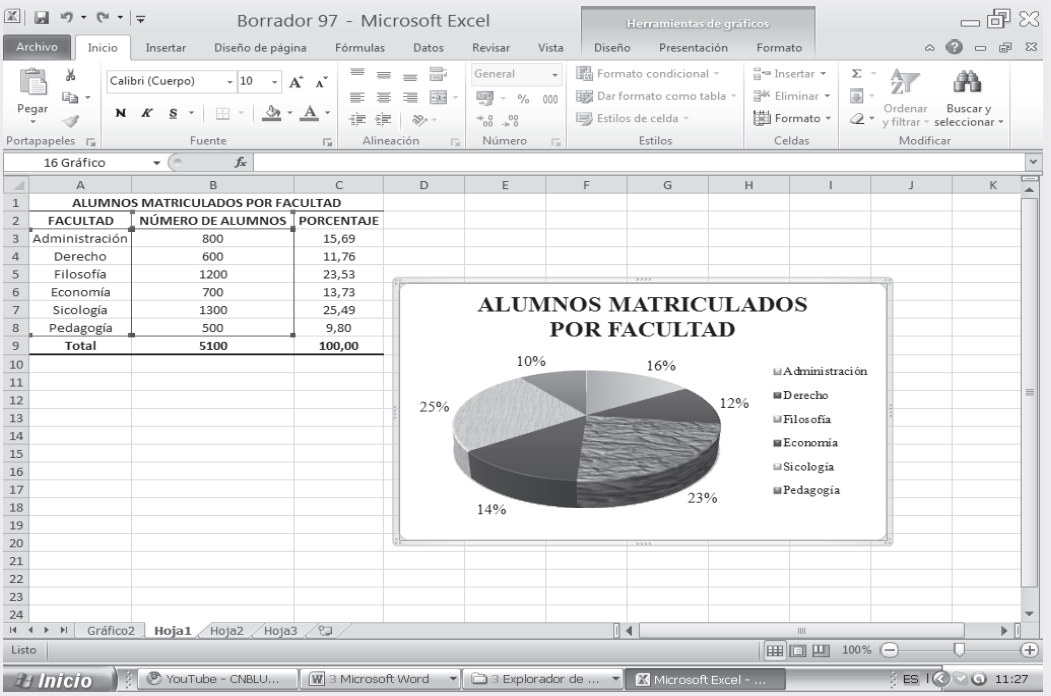
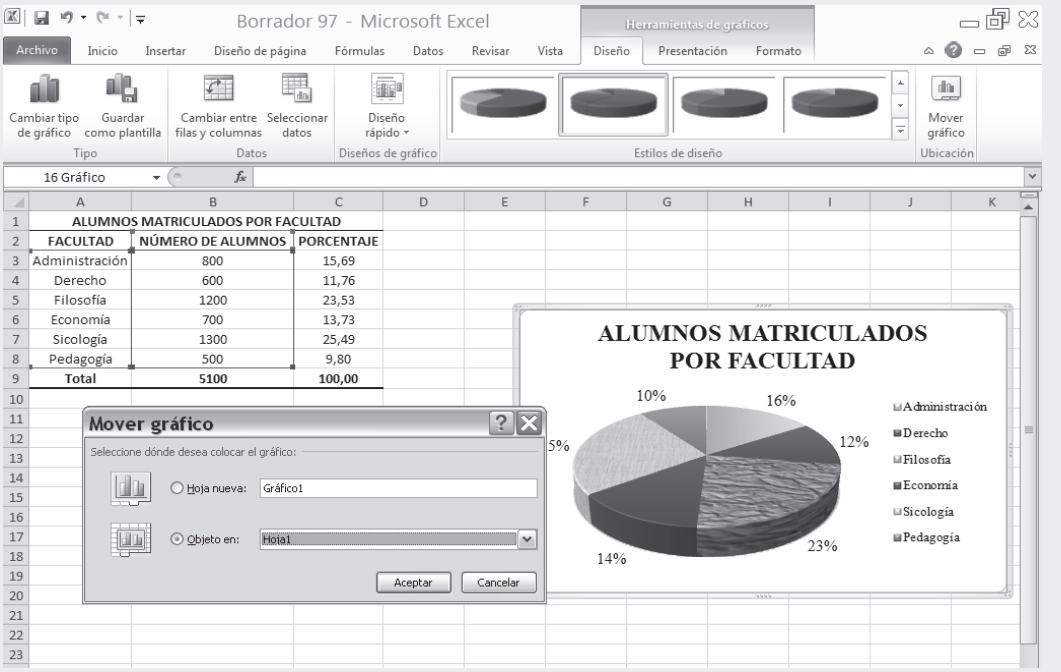


Figura No. 11. Opciones de Pegado



- Si tecleamos la opción de ETIQUETA DE DATOS, es posible mostrar la distribución que presentan las diferentes categorías. Esto es, nos indica la cantidad de ALUMNOS que hay por cada facultad.
- Finalmente, se puntea la gráfica y se pega en el informe o tarea que se esté realizando.

Figura No. 12. Gráfico Circular en una Hoja Nueva.

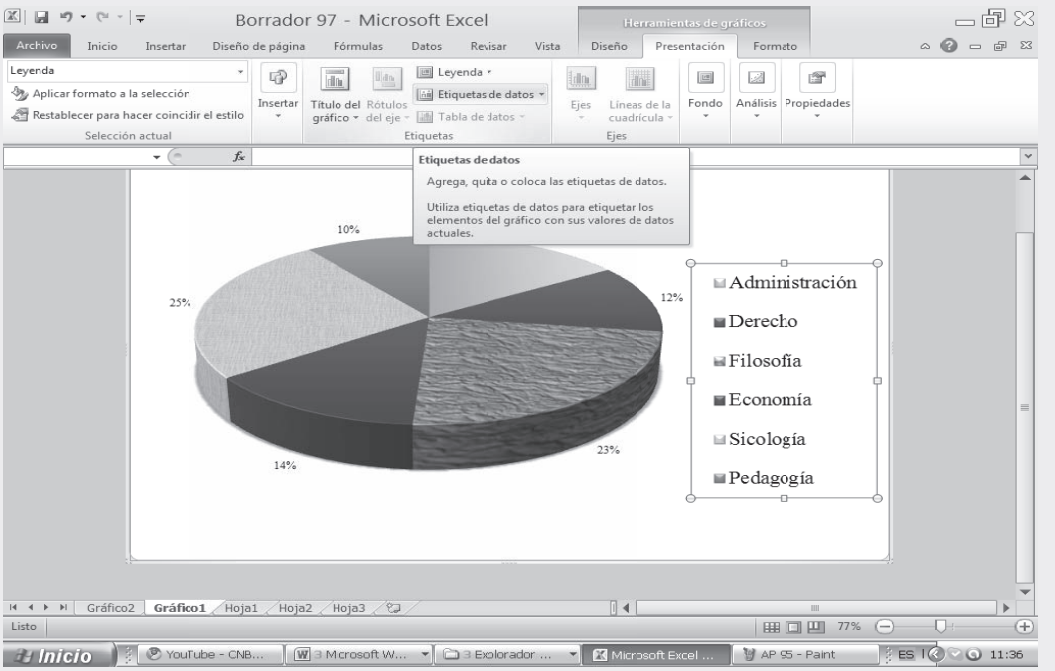


Gráfico No. 3. Gráfico Circular.

GRÁFICO DE LÍNEAS

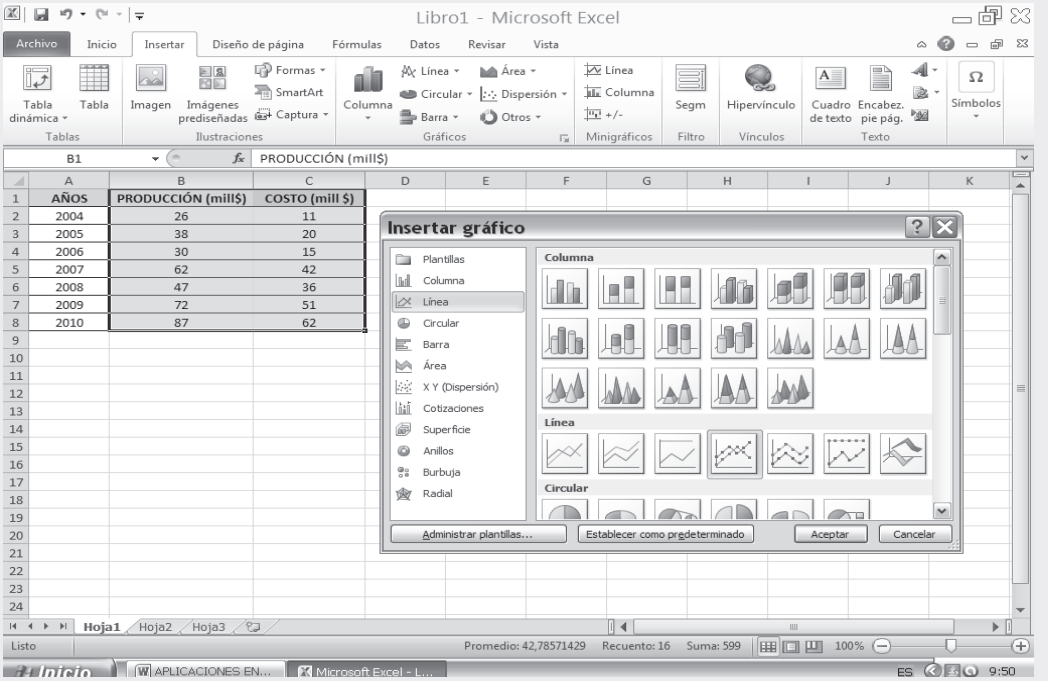
Son de gran aplicación en las denominadas SERIES DE TIEMPO O SERIES CRONOLÓGICAS, donde una de las variables corresponde al TIEMPO (*X*) (años, meses, días, etc.) y la segunda la variable investigada (Y), en nuestro caso PRODUCCIÓN (ciento mill $) y/o COSTO (ciento mill $). Recordemos que en el eje *horizontal o abscisa* van los valores correspondientes a la *variable tiempo* y en el eje vertical u ordenada los valores de la variable (Y) *producción y/o costo.*

Cuadro 2.2 Producción y Costo (ciento mill $) en la empresa X

AÑOS	PRODUCCIÓN (ciento mill $)	COSTO (ciento mill $)
2004	260	110
2005	380	200
2006	300	150
2007	620	420
2008	470	360
2009	720	510
2010	870	620

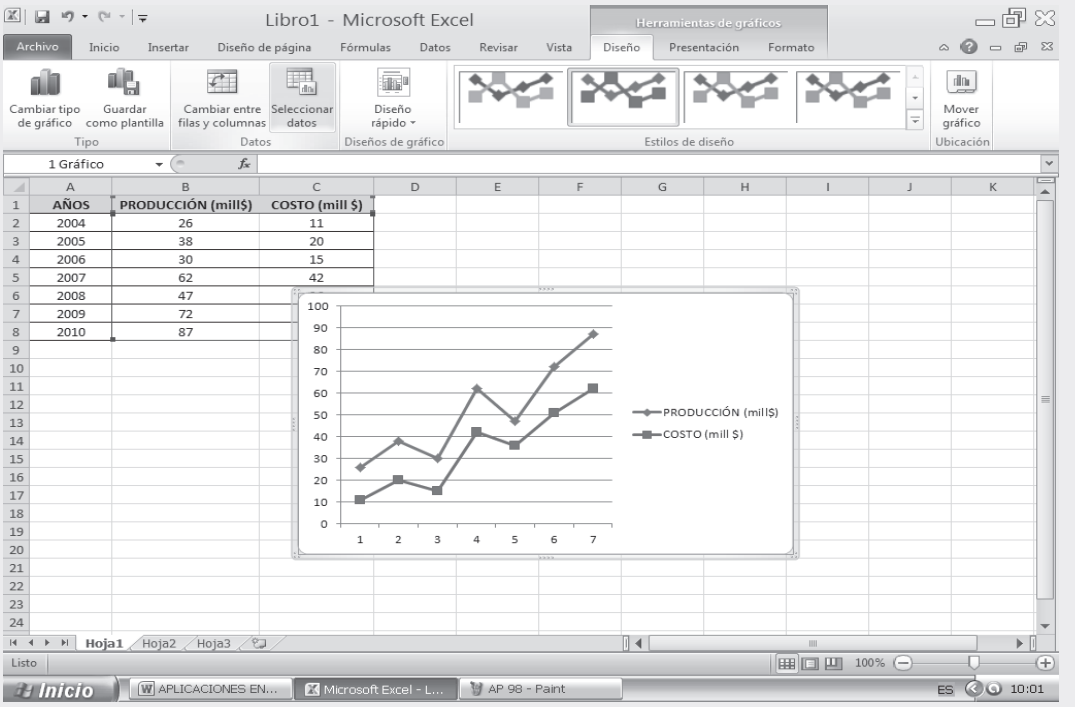
- Una vez construida la tabla en una hoja electrónica EXCEL, se deben activar las celdas correspondientes a PRODUCCIÓN y COSTOS, al momento de hacer la selección de los datos de gráfico.
- En la ficha INSERTAR hacemos CLIC en el grupo Gráficos y seleccionamos el tipo de gráfica LÍNEAS.
- Activamos el subtipo de gráfico cuatro (4) de las siete (7) opciones que nos muestran.

Figura No. 13. Insertar Gráfico.



Nota: Importante para agregar los AÑOS en el eje vertical, luego de incrustada la gráfica, vamos al menú HERRAMIENTAS DE GRÁFICO y en la pestaña DISEÑO seleccionamos el ÍCONO SELECCIONAR DATOS

Figura No. 14. Resultados de Salida



En la opción ETIQUETAS DEL EJE HORIZONTAL que se encuentra dentro del cuadro SELECCIONAR ORIGEN DE DATOS, elegimos la opción EDITAR y seleccionamos el rango de datos correspondiente a los AÑOS.

Figura No. 15. Selección Origen de Datos.

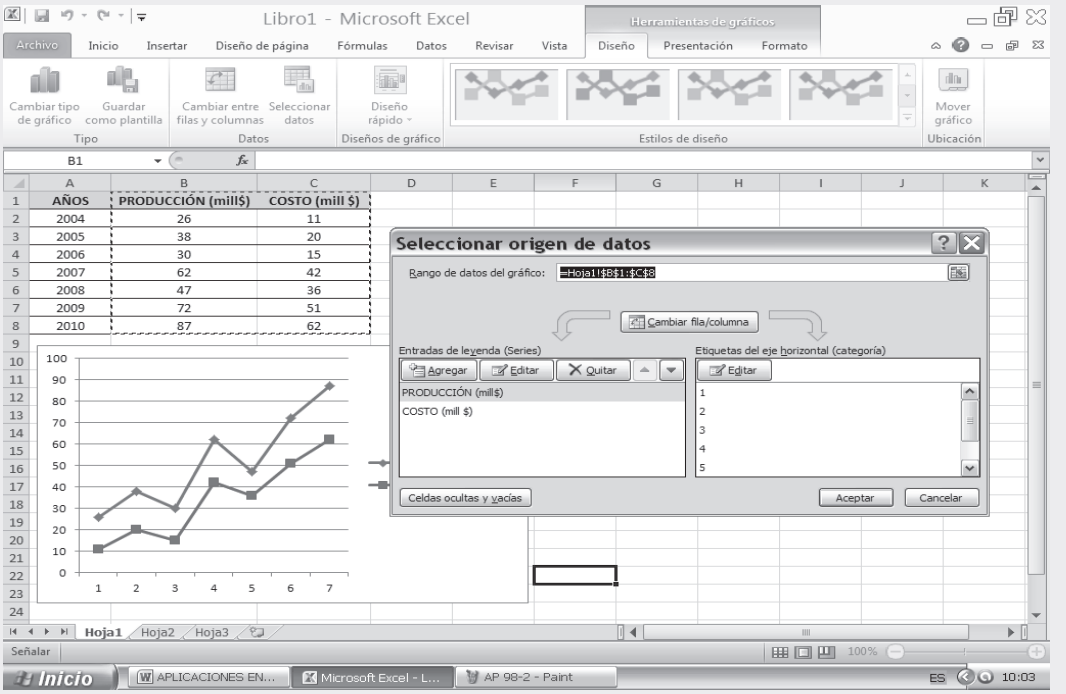


Figura No. 16. Selección Años de la producción y costo de la Empresa X. (Pantalla Adic.)

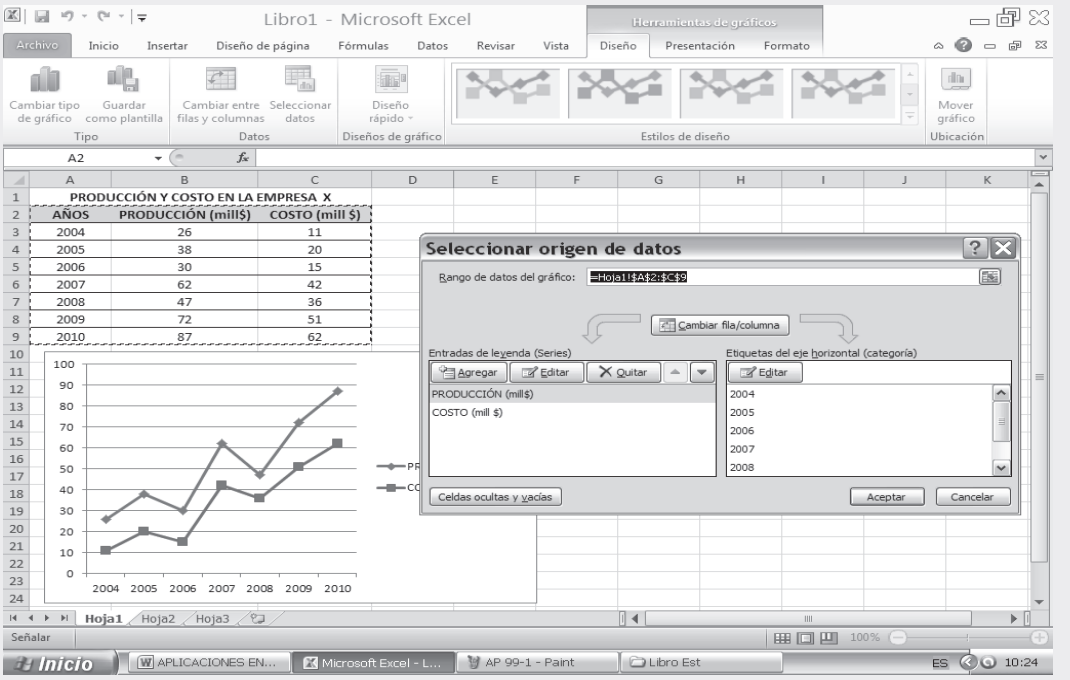
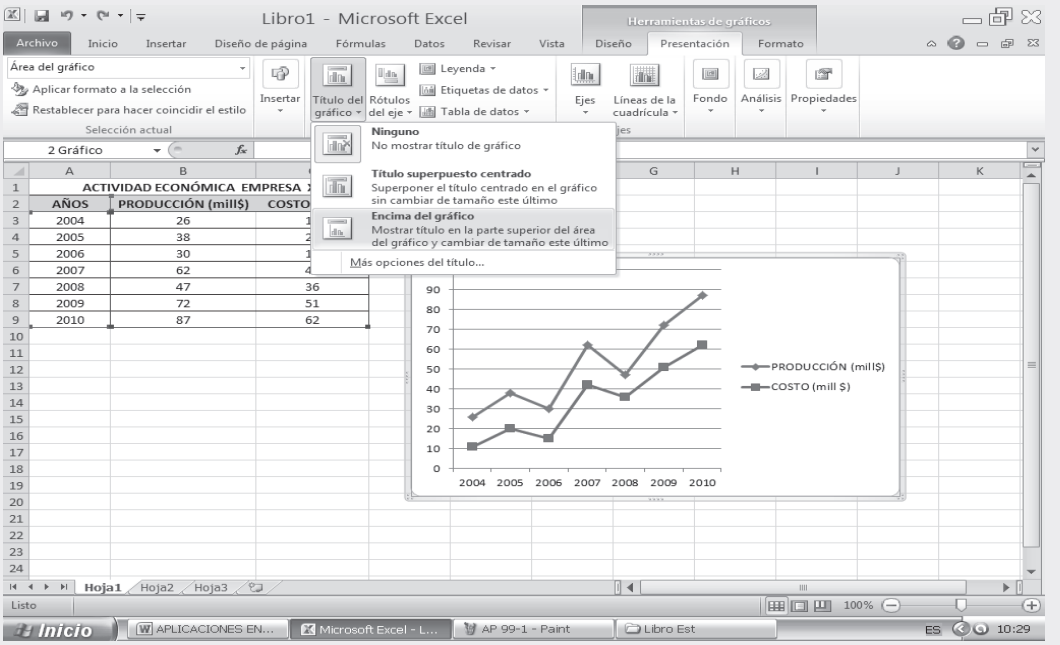
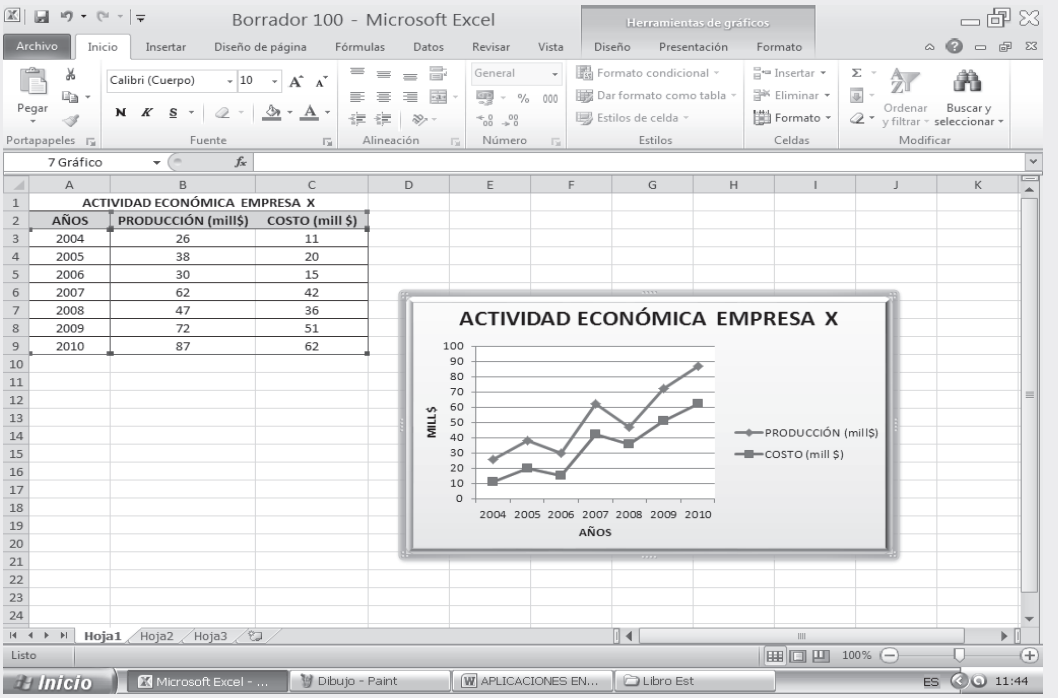


Figura No. 17. Título del Gráfico. (Pantallazo Adicional).



- Finalmente, agregamos el titulo y hacemos los ajustes de diseño que consideremos pertinentes.

Figura No.18. Resultados Gráfica de Líneas.



- Luego de este paso se selecciona el gráfico, y se copia en el informe o tarea que se esté realizando.

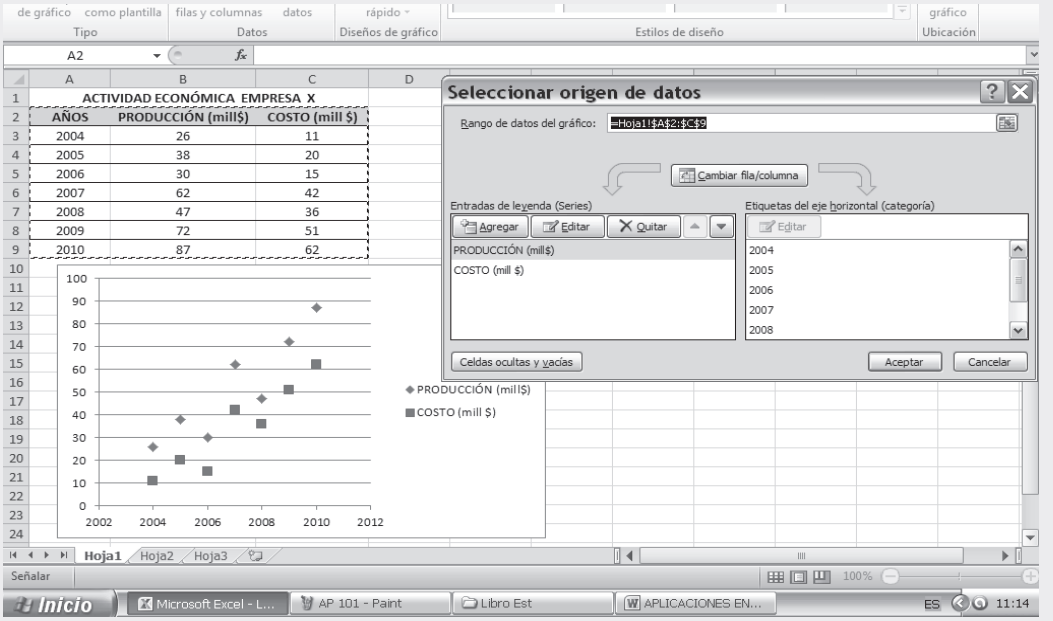
Gráfica 3. Gráfica de Líneas.

NUBE DE PUNTOS O DIAGRAMAS DE DISPERSIÓN

Esta gráfica es utilizada en el ANÁLISIS DE REGRESIÓN (ver capítulo 10), es decir cuando se trabaja simultáneamente con dos variables, a fin de visualizar mejor la información y decidir cuál es la función (lineal, parabólica, exponencial, logarítmica) que más se ajusta a ese conjunto de puntos

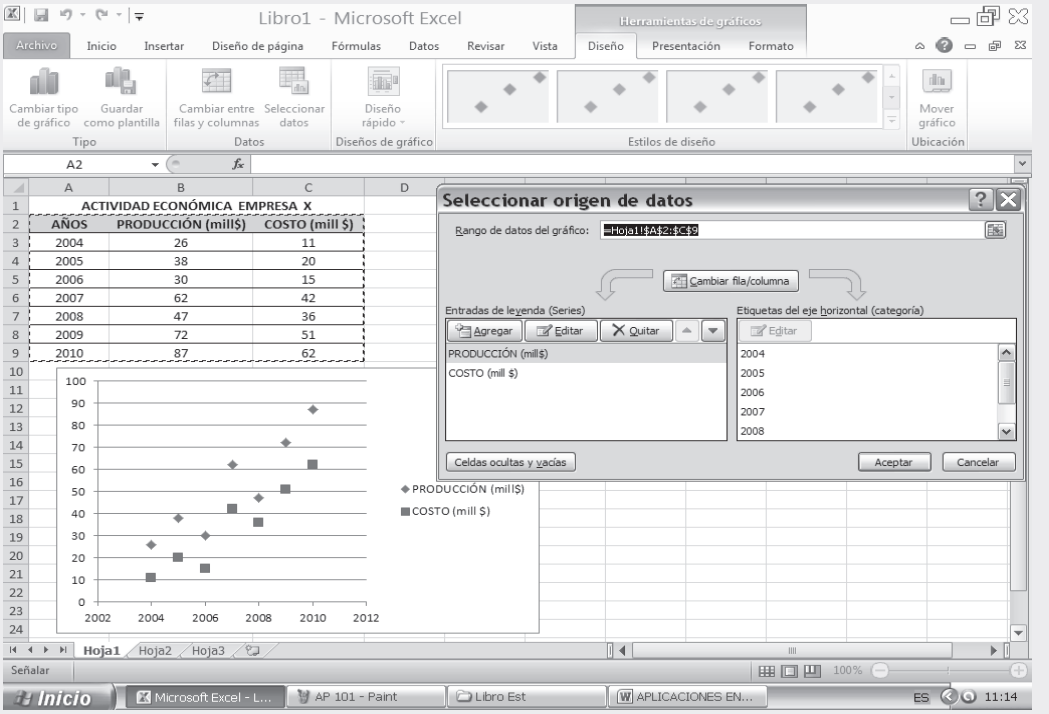
- Con los datos utilizados en el cuadro No. 2.2 (ver página 97), seleccionamos las variables PRODUCCIÓN y COSTO para elaborar la representación gráfica.
- En el cuadro de diálogo INSERTAR GRÁFICO que nos aparece, seleccionamos la opción XY DISPERSIÓN y hacemos CLIC en el primer (1) SUBTIPO DE GRÁFICO de las 5 opciones posibles.

Figura No. 19. Insertar Gráfico.



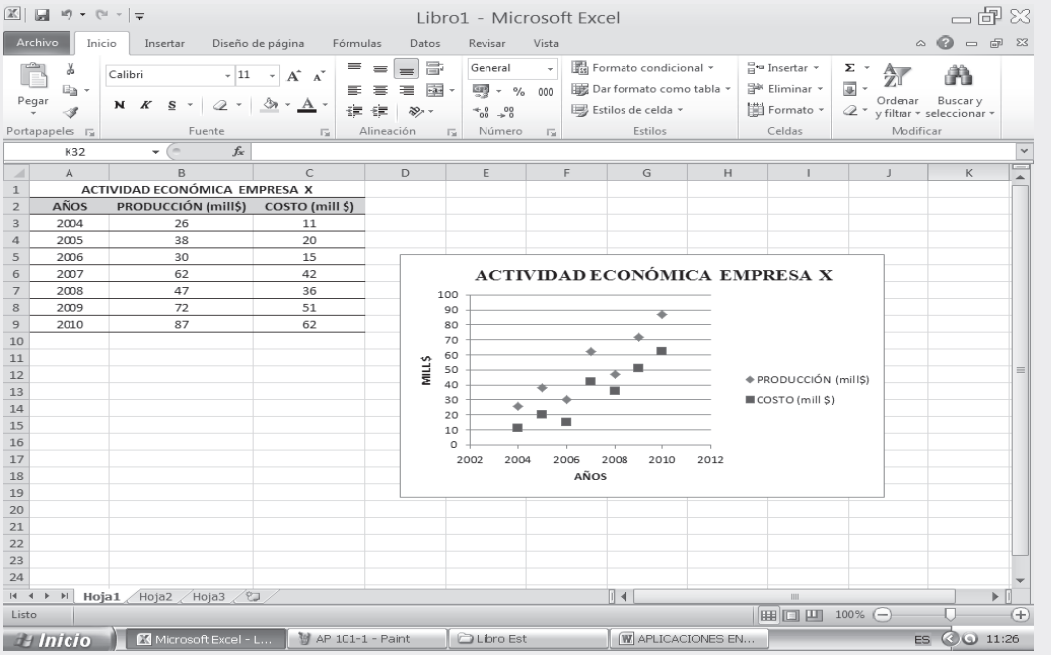
- De los datos de origen se obtienen los siguientes resultados:

Figura No.20. Salida de Resultados.



- Ahora procedemos a colocar el título del gráfico y el de los ejes, como se puede observar en la aplicación siguiente:

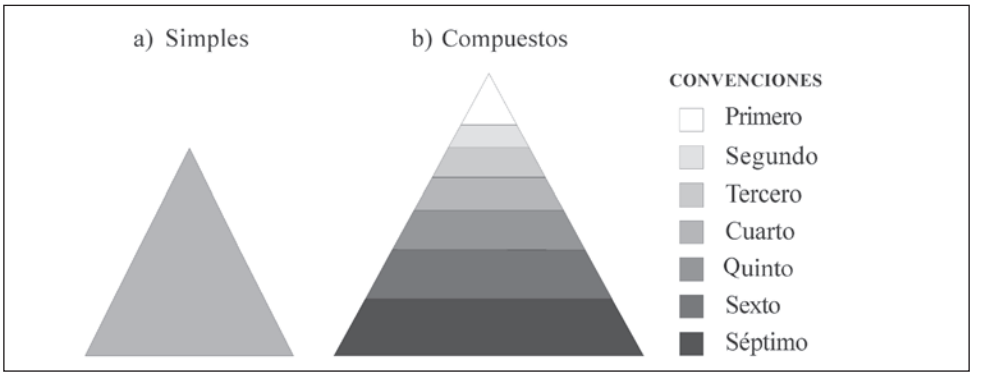
Figura No. 21. Inserción Título del Gráfico y de los Ejes.



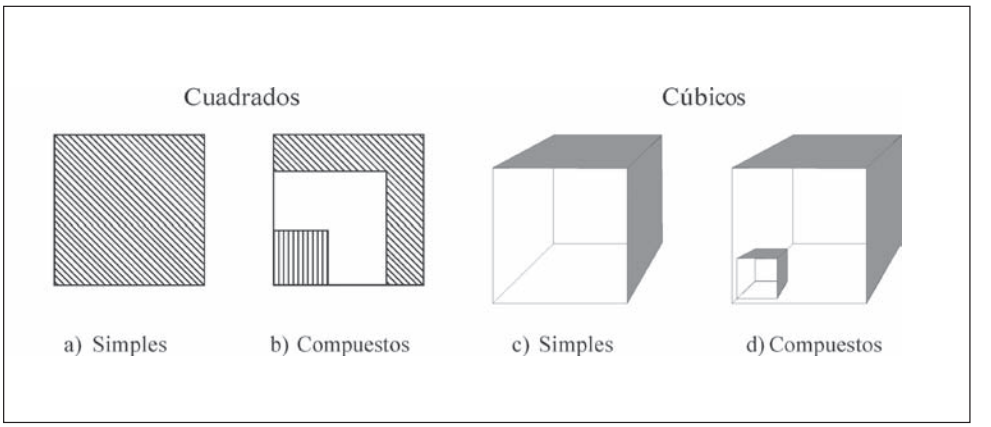
Gráfica No. 4. Gráfico de Dispersión XY

OTRAS GRÁFICAS (NO SE ENCUENTRAN EN EL PROGRAMA EXCEL)

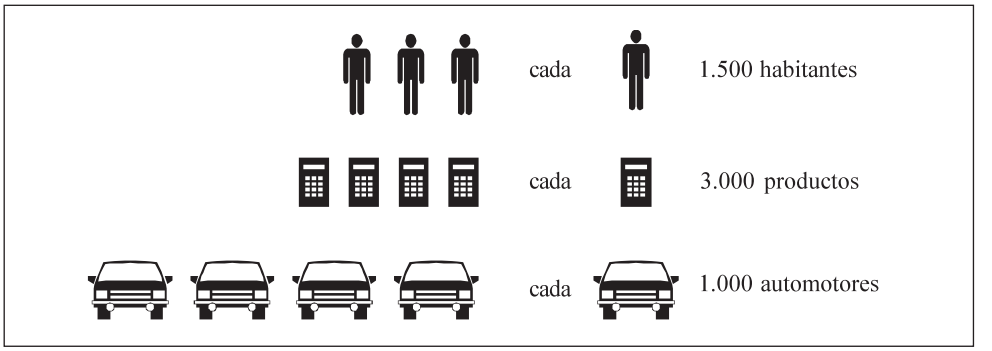
Gráficas triangulares



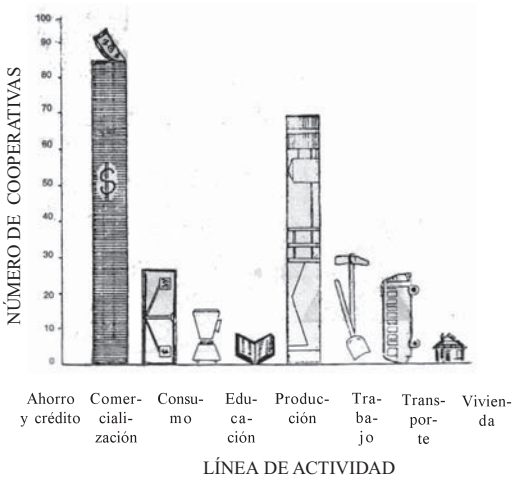
Gráficos cuadrados



Pictogramas



Número de entidades Cooperativas creadas según línea de actividad
de julio 1º de 2006 a 30 junio de 2011



Cartogramas

Gráficas semilogarítmicas

Estas gráficas se emplean en casos concretos, cuando se requiere destacar las variaciones relativas más que las absolutas; tradicionalmente utilizamos la *escala aritmética* espaciando los valores de la escala vertical (ordenada) en proporción a las diferencias absolutas; en cambio en la *escala logarítmica* los espacios corresponden proporcionalmente a las diferencias logarítmicas, siendo menores estos espacios. Consideremos la representación de un determinado índice para un período, siendo inicialmente de 400 y al finalizar el período, de 460; el incremento en valor absoluto es de 60; pero si tenemos otro índice que representa otra variable con la cual estamos comparando para el mismo período, iniciando con 800 y finalizando en 920, podemos observar que el incremento es de 120, es decir el doble del anterior, pero porcentualmente el incremento en ambos casos es igual, es decir, del 1 %. Estos dos indicadores representados gráficamente en una *escala aritmética* (valor absoluto) muestran la variación, en cambio en la *logarítmica* representan la misma magnitud siendo esta última la que concuerda con la realidad. Es necesario anotar en la gráfica, que solamente la escala vertical es de tipo logarítmico y la horizontal es aritmética.

Para la confección de un gráfico *semilogarítmico* se requiere desarrollar los siguientes pasos:

a) Se debe establecer el valor máximo y mínimo que tomará la escala vertical, en este caso hacemos referencia al índice (no hay problema, si los valores tomados en la escala son ligeramente superiores e inferiores a los extremos obtenidos en la información a representar).

b) Se calcula la diferencia de los logaritmos de los valores máximo y mínimo obtenidos en el paso anterior.

c) Se establece cuántos centímetros tendrá la escala del gráfico correspondiente.

d) Se divide el valor dado en centímetros, establecido en el punto c, por el resultado obtenido en el punto b, es decir, (c/ b).

e) Multiplicamos el logaritmo correspondiente al valor máximo dado en la escala, en el punto a, por el resultado obtenido en el punto d y se le resta el valor del punto c.

f) Luego, para cada valor (en nuestro caso el índice) de la tabla se obtiene su respectivo logaritmo y se le multiplica por el valor del punto d y se le resta el valor obtenido en el punto.

g) Los valores resultantes son dados en centímetros y corresponden a la ordenada por cada valor que tome la abscisa (eje horizontal).

Gráficas logarítmicas

El gráfico logarítmico, se emplea cuando las dos variables muestran un crecimiento geométrico, por lo tanto en ambos ejes (ordenada y abscisa) se utilizará la escala logarítmica y sirve para comparar los cambios relativos de una variable, conforme a los cambios relativos de otra variable. Esta representación es muy usual en economía cuando se trabaja con oferta y demanda, precios relativos y cantidades, bien sea de oferta o de demanda; en estos dos ejemplos, estamos comparando los valores relativos de una con respecto a valores relativos de otra.

Se dice, que si las dos variables relacionadas presentan un crecimiento geométrico se deberá utilizar el papel logarítmico, pero si una de ellas presenta un crecimiento aritmético mientras que la otra variable lo hace en forma geométrica se empleará el papel semilogarítmo, con lo cual se logra una representación gráfica correcta.

Gráficas Gantt

Tienen como fin presentar las tareas o etapas de un proceso, relacionándolas con el tiempo de su ejecución, de gran interés en el planeamiento o actividad empresarial, siendo su aplicación tan amplia y de tanto éxito que puede ser desarrollada en una empresa que tenga 2 ó 3 obreros, como en aquellos establecimientos comerciales o industriales de gran magnitud.

- Los elementos que conlleva a la elaboración del gráfico Gantt es tan sencillo, que es fácil de comprenderlo, representando a la vez el espacio, tiempo y la tarea a ejecutar. Podríamos clasificar estos gráficos en tres formas:
- Los que tratan de resaltar la relación existente entre lo realizado y lo que podría haberse hecho, tal es el caso de un obrero o máquina al realizar una determinada labor, podríamos estar determinando el tiempo inactivo.
- Cuando se desea establecer un reparto equitativo de la tarea con el fin de evitar la inactividad (obrero o máquina) y de esta manera lograr su ejecución según el orden establecido de antemano.
- Para comparar de una manera continua la ejecución de una tarea de acuerdo al plan trazado con anterioridad y dar a conocer las causas que han impedido su cumplimiento.

Gráficas de volumen

Constantemente observamos en los diferentes medios de comunicación (periódicos, revistas, T.V.) la utilización de gráficas de volúmenes y pictogramas, en estos últimos, como ya lo vimos se emplean figuras tales como personas, animales, fábricas, aviones, casas, etc., debiendo en ambos casos ser proporcionales a las cantidades para que sean representativos.

Veamos un pequeño ejemplo en el cual se quieren representar los gastos de una pequeña empresa, tomando tres aspectos: salarios, publicidad y alquiler, utilizando cubos. En este caso si se conoce o se establece la longitud de un lado (arista), se podrá calcular el volumen siendo a^3, supongamos que la arista para el cubo que representa los salarios es de 4 cms., se tendrá que el volumen es $4^3 = 64\ cm^3$, que representa a un gasto en salarios de 40 millones de pesos; ahora bien, si el gasto en publicidad es de 8 millones de pesos se tendrá que $40/8 = 5$ veces más pequeño proporcionalmente y su volumen será $64/5 = 12{,}4\ cm^3$; finalmente consideramos alquileres por un valor de $5 millones siendo $40/5 = 8$ veces más pequeño y su volumen será $64/8 = 8\ cm^3$. De acuerdo a los resultados anteriores la arista para cada caso será igual:

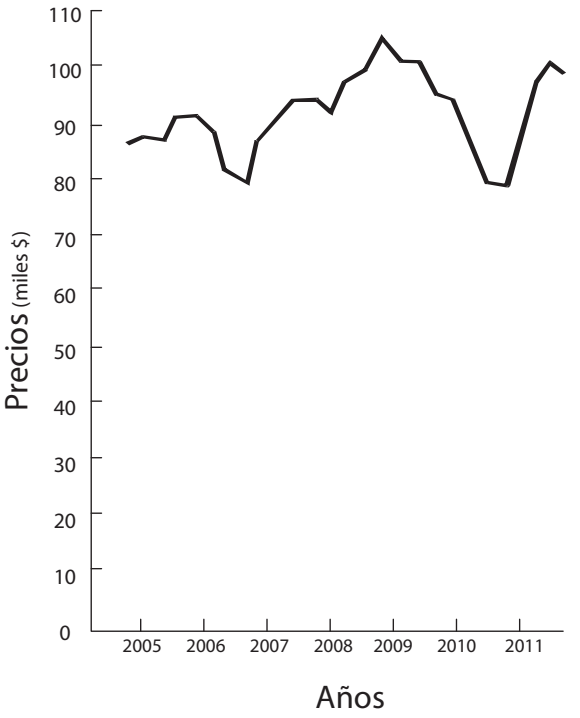
$$\text{Salarios} = \sqrt[3]{64} = 4\text{cms.} \qquad \text{Publicidad} : \sqrt[3]{12{,}4} = 2{,}31\text{cms.}$$

$$\text{Alquileres} = \sqrt[3]{8} = 2{,}20\,\text{cms.}$$

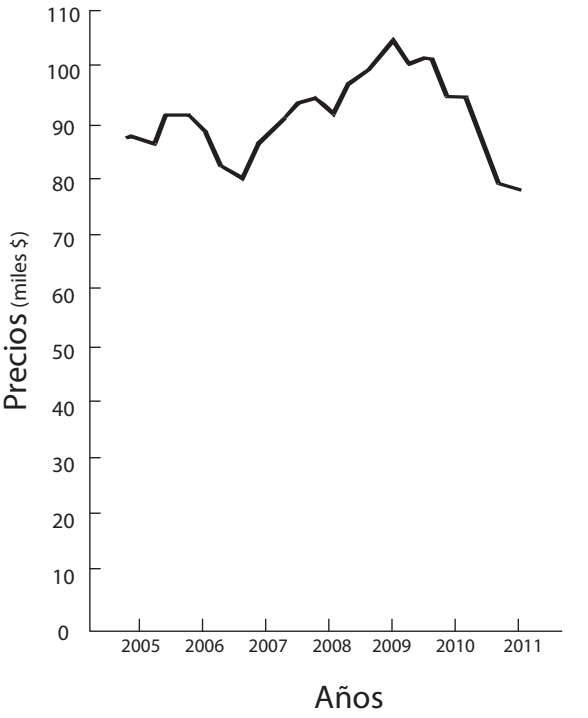
Gráficas engañosas

Con frecuencia se presentan gráficas engañosas al falsear una de las escalas, veamos estas dos representaciones gráficas lineales presentadas en forma desfavorable, casi con fluctuaciones suaves tratando de disimular las bajas súbitas.

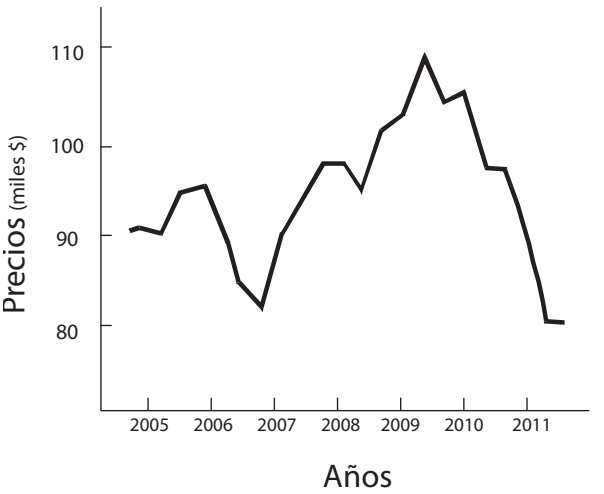
Gráfica A



Gráfica B



Gráfica C



Se puede observar en la GRÁFICA B, como al omitirse el último año hay cambio total con respecto a la anterior, desaparece la recuperación que ofrece la variable para el año 2011, dejando la sensación de que la tendencia de esa serie es a la baja. En ambos casos las escalas están mal trazadas y las gráficas mal presentadas.

Veamos ahora otras situaciones que se podrían presentar con los mismos datos, distorsionando totalmente la información que se desea presentar en forma clara y objetiva. La gráfica C es la forma correcta de presentación de la serie.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

34. Contestar verdadero si el enunciado es verídico. Si no lo es, la palabra en negrilla debe sustituirse por una expresión con la cual el enunciado sea válido.
    a) Un cuadro o tabla es la mejor forma de visualizar la información.
    b) El recuento de los empleados en una empresa, de acuerdo al cargo, es un ejemplo de variable discreta.
    c) El pictograma es la representación estadística por medio de mapas, dentro de los cuales se ubican símbolos.
    d) El histograma es un tipo de representación en gráfica de rectángulos de una distribución continua.
    e) El grosor de la lámina de metal que la compañía A utiliza en sus procesos de manufactura es un ejemplo de atributos.
    f) La población o universo es un subconjunto de unidades extraídas al azar en tal forma que nos permita tener información sobre un grupo mayor.

35. El costo estimado, en miles de pesos ($) por kilómetro, para la operación de un automóvil (Renault, Mazda, Chevrolet) modelo 2002, se muestra en la siguiente tabla:

Depreciación	$236
Mantenimiento	212
Gasolina	85

Seguros	$1060
Parqueadero	38
Impuestos	37

Construya una gráfica circular y otra de barras

36. ¿Qué ventajas tienen las gráficas en un resumen de datos?

37. Represente los siguientes datos en millones de $ en ventas:

AÑOS	2006	2007	2008	2009	2010	2011
y_i	2.000	1.800	4.600	1.200	4.600	8.230

38. Representar los siguientes datos en millones de $, en costos y ventas:

AÑOS	2007	2008	2009	2010	2011
y_i	120	180	75	240	200
x_i	150	170	100	200	250

39. Opinión de los electores sobre si van a votar para Congreso y Presidente en las próximas elecciones. Se pide elaborar el gráfico correspondiente:

OPINIÓN	CONGRESO %	PRESIDENTE %
Afirmativo	22,24	48,32
Negativo	77,76	51,68
Total	100,00	100,00

40. Elabore la gráfica correspondiente.

AÑOS	2007	2008	2009	2010	2011
Producción (miles ton.)	380	450	700	740	720
Ventas (miles ton.)	260	350	600	520	508

41. Con los siguientes datos presentados en una tabla de frecuencias:

$y'_{i-1} - y_i'$	n_i
4,1 – 24	36
24,1 – 32	20
32,1 – 36	18
36,1 – 48	22
48,1 – 52	14
Σ	110

Se pide elaborar:
a) Un histograma
b) Un polígono de frecuencias

42. Suponga que los precios de 5 artículos para dos períodos han sido:

ARTÍCULOS	2010	2011	VARIACIÓN % 2010 – 2011
Vehículos (usados)	$12.500.000	$11.300.000	–9,60
Computadores	1.700.000	1.100.000	–35,29
Betamax	250.000	350.000	40,00
Celular	1.200.000	1.500.000	25,00
Antenas parabólicas	300.000	410.000	36,67

Elaborar una gráfica de barras en dos direcciones (positivo y negativo) o en dos sentidos:

43. El consejo de administración de una gran cooperativa de vivienda desea investigar la posibilidad de contratar un supervisor para el parque de juegos infantiles. Se hizo una encuesta en la totalidad de las 616 familias de la cooperativa y cada familia tuvo un solo voto, cualquiera que fuera el tamaño del apartamento. Los resultados fueron:

¿Debe la cooperativa contratar un supervisor?

Si	146
No	91
No está seguro	158
No Respondió	221

Convierta los datos en porcentajes y construya:
a) Una gráfica de barras
b) Una gráfica circular
c) Una gráfica de componentes de porcentajes

44. Los siguientes datos corresponden al tiempo (minutos) que han necesitado 30 clientes de un Banco para llevar a cabo una transacción bancaria.

20	42	62	32	28	20	38	26	30	18	42	54	16	32	42
36	18	56	41	16	14	42	34	14	24	51	49	24	18	56

Se pide elaborar:
a) Una tabla de frecuencias
b) Histograma y el polígono en una sola gráfica.

45. Una empresa petrolera "Compañía X" en su informe anual mencionó las siguientes ventas netas y el costo de ventas desde 2006 (en miles de millones de pesos).

VARIABLE	2006	2007	2008	2009	2010	2011
Ventas netas	19.116	15.586	13.534	21.344	27.342	30.620
Costo de ventas	15.776	12.895	18.287	18.476	20.698	25.382

Se pide:
a) Elaborar un cuadro que represente esta información para ser publicada en una revista de la facultad.
b) Haga como mínimo dos gráficas que permitan representar la información anterior.

46. Con los siguientes datos correspondientes a una variable discreta, dibujar el diagrama de frecuencias, tanto para frecuencias absolutas como acumuladas.

y_i :	2	4	6	8	10	12	
n_i:	7	15	8	10	16	4	=60

47. Un almacén al cierre de sus ventas encuentra que el porcentaje de ellas por articulo se distribuye así: 42% camisas; 8% corbatas; 34% pantalones; 5% calcetines; 11% otros artículos. Se pide e1aborar:
a) Un gráfico de barras
b) Una gráfica circular.

48. Los siguientes datos corresponden a1 número de retardos (en la mañana) por parte de los empleados de una empresa, durante el mes:

2	5	0	0	6	3	5	0	0	2	7	4	4	2	3	8	4	3	3	1
3	2	3	2	4	2	2	3	1	3	2	0	0	1	4	2	2	0	3	5

Se pide elaborar una tabla de frecuencias y los diagramas correspondientes.

49. Las calificaciones de 30 alumnos que presentaron examen de admisión a la facultad, utilizando una escala de 0 a 100, fueron:

83	64	51	46	82	91	73	65	82	61	74	64	75	81	94
65	42	81	56	61	72	65	39	54	70	93	42	46	54	72

Elaborar un cuadro de frecuencias y dibujar los gráficos correspondientes (histograma, polígono y ojiva).

50. Una oficina de publicidad fue contratada con el fin de presentar en una revista de ventas, la distribución porcentual de ventas (valor) por jornadas (mañana, tarde, noche), realizada por 4 grandes almacenes de la capital, cuyos resultados fueron:

ALMACENES	MAÑANA (9 – 1)	TARDE (1 – 5)	NOCHE (5 – 10)
Carulla	22,5	42,6	34,9
Carrefour	38,6	45,4	16,0
Éxito	30,3	36,8	32,9
Vivero	46,4	32,8	20,8

Con la información del cuadro adjunto, se pide elaborar dos (2) gráficas, con el fin de seleccionar una, para que aparezca en la revista ¿Cuáles elaboraría usted?

51. Con los siguientes datos (miles mill $), el gerente de una compañía, le propone hacer dos gráficos, con el fin de escoger uno de ellos. ¿Cuales haría usted?

AÑOS	2007	2008	2009	2010	2011
Ventas	260	350	600	520	510
Producción	380	450	700	740	630

52. Una investigación realizada a 120 estudiantes en un instituto universitario requiere elaborar una gráfica.

DEPORTES	Ajedrez	Baloncesto	Balompié	Natación	Ciclismo	Tenis
ALUMNOS	13	47	62	34	20	8

53. Dibuja un histograma y al polígono de frecuencias, partiendo de la siguiente información obtenida de una tabla de frecuencias.

$m = 6$	$h_1 = 0{,}10$	$H_2 = 0{,}25$	$y_3 = 60$	$H_3 + H_4 = 0{,}95$
	$h_6 = 2(h_1)$	$h_4 = 0{,}20$	$n_1 = 4$	$c = 10$
$m = 6$	$f_1/n = 0{,}10$	$F_2/n = 0{,}25$	$X_3 = 60$	$F_3/n + F_4/n = 0{,}95$
	$f_6/n = 2(f_1/n)$	$f_4/n = 0{,}20$	$f_1 = 4$	$i = 10$

Síntesis de capítulo

Dá la sensación en numerosos informes, que el objetivo final es la presentación de *cuadros* o *tablas*, para mostrar lo dispendioso que ha sido al proceso de recolección y la abundancia de resultados obtenidos, sin preocuparse por lo que ellos digan. E1 procesamiento de la información es una de las etapas importantes en la elaboración de estadísticas, pues mediante ellas organizamos, relacionamos y en parte describimos el comportamiento de un conjunto de datos.

Un *cuadro* por si solo, no es explicativo, su elaboración se constituye en un instrumento que se emplea como complemento del texto explicativo; sin embargo esto no basta, pues a su vez debe estar acompañado de una buena *gráfica*. Esta

última tiene la gran ventaja de visualizar mejor la información pero nunca sustituye al *cuadro*, ni al texto.

La *gráfica* por fallas en confección, generalmente por desconocimiento de las técnicas necesarias para una buena representación, se convierte en una herramienta que a menudo se utiliza para engañar incautos, produciendo gráficas aparentemente llamativas pero desorientadoras.

En este capítulo vimos unas gráficas generalmente utilizadas para exp1icar la teoría estadística, ellas son: los *diagramas* en la variable discreta; los *histogramas*, *polígonos* y *ojivas* en la variable continua. Otras gráficas, tales como las circulares, *pictogramas, barras, cartogramas*, etc., que se emplean en los informes y que son las que frecuentemente observamos en los diarios o revistas que presentan estadísticas.

3
CAPÍTULO

Medidas de posición y de tendencia central

Los sabios son los que buscan sabiduría: los necios piensan ya haberla encontrado.
Napoleón Bonaparte

CONTENIDO

- Términos, definiciones, conceptos generales.
- Finalidad, usos, aplicaciones, ventajas y desventajas de cada uno
- Promedios: Media, Mediana, Moda, Media geométrica, Armónica, Cuadrática y Cúbica.
- Cuartiles, Deciles y Percentiles. Centro recorrido
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- ❑ Conocer, manejar y aplicar en cada caso particular el respectivo promedio.
- ❑ Entender el uso, las ventajas y desventajas que ofrece ese promedio.
- ❑ Saber aprovechar las propiedades que ellas tienen.

ASPECTOS GENERALES

En el capítulo anterior, se dijo que la *estadística* cumplía una función *descriptiva* haciéndolo mediante cuadros y gráficas, ahora lo complementamos con la aplicación de una serie de *medidas de posición* y de *dispersión*, que nos permiten la descripción de un hecho o de un conjunto de observaciones.

Las medidas de *posición* o de *tendencia central,* denominadas también como *promedios*, nos permiten determinar la posición de un valor respecto a un conjunto de datos, el cual lo consideramos como representativo o típico, para el total de las observaciones.

Estas medidas aplicadas a las características de las unidades en una muestra, se les denomina estimadores o *estadígrafos*. En cambio, aplicadas a las características de los elementos de una población, se les conoce como *parámetros* o *valores estadísticos* de la población.

Si con el resultado obtenido, en una encuesta aplicada en una zona o barrio de la ciudad, afirmamos que el consumo promedio de leche por familia es de dos litros por semana, estamos representando una gama o variedad de consumos, que van desde familias que no consumen, hasta un consumo superior a dos litros. Con esta información hacemos referencia al comportamiento del consumo de leche en una zona de la ciudad; también, el resultado puede ser comparado con los consumos promedios de otros barrios, o el consumo promedio por persona, o establecer la relación que puede haber entre el consumo y los niveles de ingresos.

Dependiendo de la naturaleza de los datos, la misma necesidad que se tiene y/o la característica que ellos representan, requieren de la aplicación de un promedio especial dentro de los diferentes tipos que se expondrán a continuación, pero para ello deberán tenerse en cuenta algunas reglas que serán básicas en la selección de estas herramientas de descripción.

CARACTERÍSTICAS, USO, VENTAJAS Y DESVENTAJAS DE LAS MEDIDAS DE TENDENCIA CENTRAL Y DE POSICIÓN.

Un *promedio* es un valor en la escala de las X correspondiente a una distribución de frecuencias, cuyo valor se calcula para representarla en calidad de medida típica del grupo.

Algunas condiciones esenciales para que este valor pueda cumplir su cometido:

- Debe estar representada por una sola cifra.
- Debe ser objetiva y definida por una fórmula algebraica, de tal manera que cualquiera que sea la persona que trabaje con la distribución, se obtenga siempre el mismo valor.
- Debe ser descriptiva de los datos, de manera que su significado sea fácil de entender.
- No debe ser una abstracción matemática, que únicamente pueda ser entendida por personas muy versadas en esta ciencia, ya que una de las condiciones de la estadística, es simplificar los datos y no hacerlos más complejos.
- Debe ser fácil de calcular, pero esta condición no debe preferirse en perjuicio de otras ventajas.
- Debe depender de cada uno de los elementos del grupo, de tal manera que si se altera alguno de éstos, consecuentemente se altera el valor del promedio ya que

por definición, ésta es representante típica de todos los miembros del grupo y no solamente del alguno de ellos.

- A pesar de que cada uno de los valores de los elementos del grupo debe tener influencia en el valor de la medida, ésta debe ser de tal calidad que no se deje afectar demasiado por alguno o algunos pocos valores.
- Debe tener lo que los estadísticos llaman estabilidad en el muestreo.
- Debe ser de uso fácil para cálculos matemáticos posteriores.

El *promedio* es un término muy utilizado por todos, aunque indeterminado. Se dice generalmente que un valor promedio, intenta representar o resumir las características de un conjunto de valores. La expresión corriente de promedio, suele en la mayoría de los casos referirse a la *media aritmética.*

MEDIA ARITMÉTICA: $\overline{X} = \mu$

Es la *medida de posición* o *promedio* más conocida, la más utilizada y entendida por todos, por su gran estabilidad es la preferida en el muestreo, sus fórmulas admiten tratamiento algebraico. Su desventaja principal, es ser muy sensible a cambios en sus valores u observaciones, también, cuando alguno de sus valores extremos es demasiado grande o pequeño

Ventajas

- El *promedio aritmético* es, en sí, la medida más entendida y la más utilizada. Es un valor tal, que si se calculara para un grupo en el cual todos los elementos fueran iguales, cada uno de ellos sería igual a su promedio aritmético, por lo cual podemos deducir que el promedio de *n* elementos, es un nuevo elemento formado, correspondiernte a una parte igual a 1/*n* de cada uno de los elementos originales.
- Esta medida se define mediante una ecuación matemática muy fácil de entender y algunas veces se puede obtener cuando no es posible calcular otros tipos de medidas de *tendencia central* o de *posición* y, aún en el casos de no conocerse los valores individuales de la serie, por ejemplo: si 10 familias consumen 20 litros de leche, el promedio aritmético será de 2 litros por familia.
- El promedio aritmético es extraordinariamente estable en el muestreo.
- Es altamente sensible a cualquier cambio en los datos de la distribución.
- Como ventaja final, se podrá anotar que es excepcionalmente adaptable cuando se trata de hacer cálculos matemáticos posteriores con él.

Desventajas

- Es muy sensible a valores muy pequeños o grandes, especialmente cuando estos últimos se encuentren incluidos en la distribución que se está estudiando, puede resultar en un promedio que no represente lo típico para el total del grupo.
- Cuando la distribución es marcadamente asimétrica, de tal forma que el *promedio aritmético*, la *mediana* y el *modo* difieran en forma apreciable, debe considerarse siempre la posibilidad de que el promedio aritmético pueda no ser el valor único representativo de la serie.

- Otro inconveniente o desventaja, es cuando la distribución tiene la forma de U, es decir, parabólica; éste corresponde a los valores menos comunes en la serie y por tanto, puede darnos una idea irreal de la distribución.

Esta medida, por lo general, se representa mediante una equis con una raya o trazo colocado en su parte superior, sin embargo se utilizan otros símbolos como se pueden observar a continuación.

$$a \qquad a_y \qquad a_x \qquad M_{[x]} \qquad M_{[y]} \qquad \overline{x} \qquad \overline{y} \qquad M_{[x]} \qquad \mu$$

Se define como la "*suma de todos los valores observados, divididos por el número total de observaciones*" De esta forma definida, sólo se aplica en datos sin agrupar, tambien denominados como datos originales.

Media aritmética simple

Se trabaja con datos sin agrupar u originales, siendo:

$$\overline{x} = \frac{\Sigma x_i}{n} \qquad \text{Donde} \quad \overline{x} = \frac{x_1 + x_2 + x_3 ..... x_n}{n} \quad \text{(Minúsculas en las muestras)}$$

$$\mu = \overline{X} = \frac{\Sigma X_i}{N} \; ; \; \text{Siendo} \quad \mu = \frac{X_1 + X_2 + X_3 ..... X_N}{N} \quad \text{(Mayúsculas y letras griegas en la población)}$$

Ejemplo 1. Aplicando la fórmula dada para calcular la media aritmética, en datos no agrupados, de acuerdo a los datos del Ejemplo 1, de la Unidad 2, correspondiente a una muestra realizada en 30 cajas, para determinar el número de unidades con piezas defectuosas, se tendrá que:

$$\overline{x} = \frac{\Sigma x_i}{n} \qquad \overline{x} = \frac{3 + 2 + 0 + ... + 2}{30} = \frac{67}{30} = 2{,}23$$ Aproximadamente 2 unidades defectuosas por caja

Ejemplo 2. Apliquemos la anterior fórmula para datos no agrupados con los datos dados, en el caso de una variable continua (Ejemplo 2 de la Unidad 2), correspondiente a una muestra de 30 alumnos, con la cual se quiere investigar su peso promedio. El resultado es.

$$\overline{x} = \frac{\Sigma x_i}{n} \qquad \overline{x} = \frac{74 + 67 + 94 + ... + 92}{30} = \frac{2.067}{30} = 68{,}9$$

Media aritmética ponderada.

Se aplica en datos agrupados, es decir, aquellos que se encuentran organizados en una *tabla de frecuencias,* siendo las frecuencias absolutas sus ponderaciones..

Ejemplo 3. Supongamos que dispone de información para 10 0bservaciones: 8, 2, 8, 6, 2, 2, 6, 8, 2, 4. La media aritmética será:

$$\overline{x} = \frac{8 + 2 + 8 + ... + 4}{10} = \frac{48}{10} = 4{,}8$$ Ahora si calculamos la media aritmética de las mismas 10 observaciones, pero ordenadas, el resultado obtenido será el mismo.

$$\bar{x} = \frac{2+2+2+2+4+6+6+8+8+8}{10} = \frac{48}{10} = 4{,}8$$ Resultado que sería exactamente igual a:

$$\bar{x} = \frac{2(4)+4(1)+6(2)+8(3)}{10} = \frac{8+4+12+24}{10} = 4{,}8$$ ¿Qué se ha hecho? Simplemente se han organizado los datos en una tabla de frecuencias

y_i	n_i	$y_i\, n_i$
2	4	8
4	1	4
6	2	12
8	3	24
Σ	10	48
X_i	f_i	$X_i\, f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} \quad \text{o} \quad \bar{X} = \frac{\Sigma X_i f_i}{N}$$

$$\bar{y} = \frac{y_1 n_1 + y_2 n_2 + \ldots\ldots}{n_1 + n_2 + \ldots\ldots\ldots} = \frac{2(4)+4(1)\ \ldots\ldots\ldots}{10} = \frac{48}{10} = 4{,}8$$

Ejemplo 4. En el ejemplo de las 30 cajas, seleccionadas con el fin de establecer el promedio de unidades defectuosas por caja, se podrá obtener también aplicando la fórmula de cálculo para datos agrupados, como se observa a continuación:

y_i	n_i	$y_i\, n_i$
0	4	0
1	5	5
2	7	14
3	8	24
4	6	24
Σ	30	67
X_i	f_i	$X_i\, f_i$

Con la información de la tabla 2.1- Unidad 2
La *Media aritmética* trabajando con datos agrupados será:

$$\bar{y} = \frac{\Sigma y_i n_i}{n} \quad \text{o} \quad \bar{X} = \frac{\Sigma X_i f_i}{N}$$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} = 2{,}23 \quad \text{Unidades defectuosas}$$

Ejemplo 5. Algo similar lo podemos hacer, para el cálculo de la media aritmética en la variable continua. Se obtiene en primer lugar las marcas de clase y luego se trabaja exactamente, como se hizo en la variable discreta.

$y'_{i-1} - y'_i$	n_i	y_i	$y_i n_i$
46,1–54	3	50	150
54,1–62	6	58	348
62,1–70	8	66	528
70,1–78	6	74	444
78,1–86	4	82	328
86,1–94	3	90	270
Σ	30	--	2.068
$X'_{i-1} - x'_i$	f_i	X_i	$X_i\, f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} \quad \text{o} \quad \bar{X} = \frac{\Sigma X_i f_i}{N}$$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{2.068}{30} \equiv 68{,}93 \ \text{Kilos}$$

Cálculo de la Media, utilizando Frecuencias Relativas

Esta medida también se puede calcular, utilizando como ponderaciones a las frecuencias relativas, con las cuales se obtendrá el mismo resultado:

$\bar{y} = \frac{\Sigma y_i n_i}{n}$ $\quad \bar{y} = \frac{y_1 n_1 + y_2 n_2 + ... + y_m n_m}{n_1 + n_2 + ...}$ Se sabe que $h_i = \frac{n_i}{n}$ por lo tanto se tiene que

$\bar{y} = y_i h_1 + y_2 h_2 + ...y_m h_m \Rightarrow$ $\bar{y} = \Sigma y_i h_i$ ó $\bar{x} = \Sigma x_i (f_i / n)$

Ejemplo 6. Considerando las tablas de frecuencias (2.1 y 2.4) calculamos las medias aritméticas, en cada una de ellas, trabajando con las frecuencias relativas, cuyos resultados deberían ser iguales, pero en algunos casos se observan algunas diferencias, debido a la aproximación a dos dígitos, que se hizo para las frecuencias relativas.

y_i	h_i	$y_i h_i$
0	0,13	0
1	0,17	0,17
2	0,23	0,46
3	0,27	0,81
4	0,20	0,80
Σ	1,00	2,24
X_i	f_i	$X_i\left(f_i/n\right)$

(2.1) $\bar{y} = \Sigma y_i h_i = 2,24$

$\bar{x} = \Sigma x_i (f_i / n)$

(2.2) $\bar{y} = \Sigma y_i h_i = 68,88$

y_i	h_i	$y_i h_i$
50	0,10	5,00
58	0,20	11,60
66	0,27	17,82
74	0,20	14,80
82	0,13	10,66
90	0,10	9,00
Σ	1,00	68,88
X_i	f_i	$X_i\left(f_i/n\right)$

DESVIACIONES

Se considera necesario mencionar las *Desviaciones con respecto a la media aritmética*, las cuales se definen como las diferencias que hay entre los distintos valores que toman la variable y la media aritmética, tanto en *datos sin agrupar* como en *datos agrupados*, por lo general simbolizada por Z_i o d_i

$Z_i = x_i - \bar{x}$ (En datos sin agrupar) $\qquad$ $Z_i = y_i - \bar{y}$ (Para datos agrupados

Ejemplo 7. Procedamos a considerar una nueva fuente de información con 8 datos, sin agrupar: 6, 10, 4, 10, 8, 2, 6, 2. En primer lugar calculamos la media y luego sus desviaciones.

$\bar{x} = \frac{\Sigma x_i}{n} = \frac{48}{8} = 6$:

$z_1 = x_1 - \bar{x} = 6 - 6 = 0 \Leftrightarrow d_1$

$z_2 = x_2 - \bar{x} = 10 - 6 = 4 \Leftrightarrow d_2$

$z_3 = x_3 - \bar{x} = 4 - 6 = -2 \Leftrightarrow d_3$

Ejemplo 8. Si calculamos las desviaciones con respecto a la media, en una tabla con datos correspondiente a una variable discreta (2.1) en la cual la media fue de 2,23, se tendrá:

y_i	n_i	$y_i - \bar{y}$
0	4	–2,23
1	5	–1,23
2	7	–0,23
3	8	0,77
4	6	1,77
Σ	30	--
X_i	f_i	$X_i - \bar{X}$

$z_1 = y_1 - \bar{y} = 0 - 2,23 = -2,23$

$z_2 = y_2 - \bar{y} = 1 - 2,23 = -1,23$

$z_3 = y_3 - \bar{y} = 2 - 2,23 = -0,23$

$z_4 = y_4 - \bar{y} = 3 - 2,23 = 0,77$

$z_5 = y_5 - \bar{y} = 4 - 2,23 = 1,77$

$d_i = x_i - \bar{x}$

Ejemplo 9. Para el cálculo de estas desviaciones en la variable continua, se debe trabajar con las marcas de clase. La media aritmética de los datos de la tabla 2.2 fue de 68,93.

$y'_{i-1} - y'_i$	y_i	$y_i - \bar{y}$
46,1 – 54	50	–18,93
54,1 – 62	58	–10,93
62,1 – 70	66	– 2,93
70,1 – 78	74	5,07
78,1 – 86	82	13,07
86,1 – 94	90	21,07
Σ	--	--
$X'_{i-1} - X'_i$	X_i	$X_i - \bar{X}$

$$z_1 = y_1 - \bar{y} = 50 - 68{,}93 = -18{,}93$$
$$z_2 = y_2 - \bar{y} = 58 - 68{,}93 = -10{,}93$$
$$z_3 = y_3 - \bar{y} = 66 - 68{,}93 = -\ 2{,}93$$
$$z_4 = y_4 - \bar{y} = 74 - 68{,}93 = +\ 5{,}07$$
$$z_5 = y_5 - \bar{y} = 82 - 68{,}93 = +13{,}07$$
$$z_6 = y_6 - \bar{y} = 90 - 68{,}93 = +21{,}07$$

PROPIEDADES DE LA MEDIA ARITMÉTICA

Se hace necesario conocer y manejar correctamente las propiedades que presenta la Media aritmética, pues facilita la realización de ciertas operaciones, algunas de ellas necesarias para el desarrollo de la teoría estadística y otras, como simplificación de cálculos.

a) La suma de las desviaciones respecto a la media, siempre serán iguales a cero.

$\Sigma Z_i = \Sigma(X_i - \bar{X}) = 0$ ó $\Sigma d_i = \Sigma(X_i - \bar{X}) = 0$ En datos sin agrupar

$\Sigma Z_i n_i = \Sigma(y_i - \bar{y}) n_i$ ó $\Sigma d_i f_i = \Sigma(X_i - \bar{X}) f_i = 0$ En datos agrupados

Aplicación

y_i	n_i	$y_i - \bar{y}$	$(y_i - \bar{y}) n_i$
0	4	-2,23	-8,92
1	5	-1,23	-6,15
2	7	-0,23	-1,61
3	8	0,77	6,16
4	6	1,77	10,62
Σ	30	--	0
X_i	f_i	$X_i - \bar{X}$	$(X_i - \bar{X}) f_i$

$\Sigma Z_i n_i = \Sigma(y_i - \bar{y}) n_i = 0$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} \equiv 2{,}23$$

Nota: Si sumamos los valores de la última columna de la tabla, debería haber dado cero, pero resulta que nos da 0,10, en vez de 0, ello se debe que el valor de la media es 2,23333 que usted puede comprobarlo trabajando con todos estos decimales.

b) La media aritmética de una variable por una constante, es igual a la constante por la media aritmética de la variable.

Aplicación

y_i	n_i	$y_i n_i$
0	4	0
1	5	5
2	7	14
3	8	24
4	6	24
Σ	30	67
X_i	f_i	$X_i f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} \equiv 2{,}23$$

K = 4

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{268}{30} \equiv 8{,}93$$

$\bar{y} = k\bar{y} = 4(2{,}23) = 8{,}92$

Por las aproximaciones en la media inicial.

y_i	n_i	$y_i n_i$
0	4	0
4	5	20
8	7	56
12	8	96
16	6	96
Σ	30	268
X_i	f_i	$X_i f_i$

La variable de la primera tabla, está dada en Kilos, y en la segunda tabla convertida en libras y se desea calcular la media aritmética en libras.

$$\bar{X} = \frac{\Sigma X_i f_i}{N}$$

c) La media aritmética de una constante es igual a la constante $\Rightarrow M_{(K)} = K$. Supongamos que un estudiante durante el semestre le hicieron 6 evaluaciones y en cada una de ellas la calificación obtenida fue 4, su promedio será exactamente cuatro (4)

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{4+4+4+4+4+4}{6} = \frac{24}{6} = 4 \qquad \bar{x} = \frac{\Sigma k}{n} = \frac{nk}{n} = k$$

(Propiedad vista en la sumatoria)

d) La media aritmética de una variable más una constante, es igual a la media aritmética de la variable. Esta propiedad de la suma, es válida para la diferencia

Aplicación

y_i	n_i	$y_i n_i$
0	4	0
1	5	5
2	7	14
3	8	24
4	6	24
Σ	30	67
X_i	f_i	$X_i f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} \equiv 2{,}23$$

K = 4

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{187}{30} \equiv 6{,}23$$

$$\bar{y} = \bar{y} + k = 2{,}23 + 4 = 6{,}23$$

Por las aproximaciones en la media inicial.

$$M_{[x+k]} = x + k \qquad M_{[x-k]} = x - k$$

y_i	n_i	$y_i n_i$
4	4	16
5	5	25
6	7	42
7	8	56
8	6	48
Σ	30	187
X_i	f_i	$X_i f_i$

e) La media aritmética de la suma de dos variables, es igual a la suma de las dos medias correspondiente a las dos variables. Algo similar sucede con la diferencia o cuando se toma más de dos de dos variables.

$$M_{[x+y]} = M_{[x]} + M_{[y]} = \bar{x} + \bar{y} \qquad M_{[x-y]} = M_{[x]} - M_{[y]} = \bar{x} - \bar{y}$$

f) La media aritmética de dos muestras, es igual, a la media ponderada de las submuestras, siendo sus ponderaciones los tamaños de esas submuestras.

Aplicación

Muestra

y_i	n_i	$y_i n_i$
50	3	150
58	6	348
66	8	528
74	6	444
82	4	328
90	3	270
Σ	30	2.068
X_i	f_i	$X_i f_i$

Submuestra 1

y_i	n_i	$y_i n_i$
50	3	150
58	6	348
66	8	528
Σ	17	1.026
X_i	f_i	$X_i f_i$

Submuestra 2

y_i	n_i	$y_i n_i$
74	6	444
82	4	328
90	3	270
Σ	13	1.942
X_i	f_i	$X_i f_i$

Muestra **Submuestra 1** **Submuestra 2**

$n = 30;\quad \bar{y} = \frac{2.068}{30} = 68{,}93;$ $n_1 = 17;\ \bar{x}_1 = \frac{1.026}{17} = 60{,}35;$ $n_2 = 13;\ \bar{x}_1 = \frac{1.042}{13} = 80{,}15$

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + ...}{n_1 + n_2 + ...} \qquad \bar{x} = \frac{60{,}35(17) + 80{,}15(13)}{17 + 13} = \frac{1.026 + 1.042}{30} = \frac{2.068}{30} = 68{,}93$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en **Sistema de Información en Línea SIL.**

1. Con los siguientes datos referentes a una distribución simétrica, se pide calcular la media aritmética aplicando uno de los métodos que usted conoce.

X_i	y_i:	10	20	30	40	5	
f_i	n_i:	6	10	18	10	6	= 50

2. Con los siguientes datos

153	123	129	132	147	138	137	134	131	147	158	128	134	148	125	139	146
145	148	135	152	128	146	143	138	138	122	146	137	151	145	124	132	138
144	141	137	146	138	146	152	136	160	159	157	150	160	142	148	130	

Se pide:

a) Calcular la media aritmética para datos no agrupados u originales.

b) Elaborar una tabla de frecuencias.

c) Calcular la media aritmética para datos agrupados, primero utilizando las frecuencias absolutas y luego las relativas.

d) Comprobar que la suma de las desviaciones con respecto a la media es cero.

3. Con la tabla anterior, se pide:

a) Dividir la tabla en dos, cada una con tres intervalos y calcular la media para cada una. Luego con esas medias aplicar una de sus propiedades, para calcular una sola media.

b) Utilizando la tabla completa, multiplique por dos a la variable, calculando la nueva Media; luego aplique la propiedad para abreviar el procedimiento. ¿Cuál?

4. Con la siguiente distribución de frecuencias, calcular la media ponderada.

$y'_{i-1} - y'_i$	6,1-12	12,1-18	18,1-24	24,1-30	30,1-36	36,1-42	$X'_{i-1} - X'_i$
N_i	8	12	40	70	90	100	F_i

5. Si la media de 75 artículos es 52,6 galones y, la de 25 artículos es de 48,4 galones, encuentre la media de los 100 artículos.

6. De 500 estudiantes de secundaria, cuya estatura media es de 1,57 metros, 150 de ellos son mujeres. Si la estatura media de las mujeres es de 1,52 metros ¿Cuál es la estatura media de los varones?

7. Un grupo de 200 estudiantes, cuya estatura media es de 160,96 centímetros, se dividen en dos grupos, uno con una estura media de 163,4 centímetros y otro con 157,3 centímetros. ¿Cuántos estudiantes hay en cada grupo?

8. Se realiza una prueba a dos secciones de un curso de matemáticas, con un promedio general de 60,98. La sección I tiene una media de 57,30 y la sección II de 65,30. Si hay 27 estudiantes en la sección I.¿Cuántos hay en la sección II?

9. En una competencia atlética con 45 participantes, 20 compiten el primer día y los restantes al día siguiente. Si en el primer día, los 20 obtienen un promedio de 48,4 puntos. ¿Cuál es el menor número de puntos que deben obtener en promedio los restantes atletas, de modo que el puntaje medio de todo el equipo sea por lo menos igual a 55,0?

10. Un grupo de 100 atletas viaja en dos aviones. El primero lleva 40 atletas y el segundo los restantes. Se sabe que el peso medio de todos los atletas es de 186,3 libras y que el de los del segundo avión es tres libras menos que el de los atletas del primer avión ¿Cuál es el peso medio de los atletas en cada avión?

11. En un examen tomado a tres secciones en un curso de estadística de 91 alumnos, el puntaje medio general fue de 69,3. Los puntajes medios de las secciones 1 y 2 fueron 70,4 y 64,2 respectivamente. Se perdieron los archivos con las notas de la sección 3, pero los monitores recuerdan que las secciones 1 y 2 tenían exactamente el mismo número de alumnos, mientras que el monitor de la sección 3, afirma que su sección tenía 5 estudiantes menos que la 1.¿Cuál fue el promedio de la sección 3?

12. La media aritmética de los salarios pagados en un mes a los empleados de una empresa, ascendió a $920.000. La media aritmética de los salarios pagados a los hombres y a las mujeres, fueron respectivamente $970.000 y $840.000. Determinar el porcentaje de hombres y mujeres empleados en dicha empresa?.

13. El salario medio mensual por obrero en la empresa A fue de $938.000 durante el 2010. Para el presente año la empresa aumentó a cada uno de sus empleados la suma $78.000. Admitiendo que los salarios no se han modificado desde entonces. ¿Cuál es el promedio mensual?

14. En cierto examen el puntaje medio de los estudiantes del curso A es 68,4 y los del B 71,2. Si el puntaje medio del grupo es 70,0 ¿Cuál es la relación, entre el número de los alumnos, en los cursos A y B?

15. En un curso hay 35 hombres con una media de 17,5 años y 15 mujeres, las que en promedio son 22% más jóvenes. ¿Cuál es la edad promedio del curso?.

16. El precio promedio de un centenar de artículos es $18.750, los artículos se dividen en dos grupos, con medias $17.580 y $19.780 ¿Cuántos artículos hay en cada grupo?

17. Se sabe que ninguna de las sucursales de una empresa comercial, tiene más de 19 empleados o menos de 17. La mayoría tiene 18 empleados, pero el 25% tiene 19 empleados y cada una de las 10 sucursales tiene 17 empleados. ¿Cuál es el promedio de empleados por sucursal?

18. De 10 familias investigadas con teléfono, auto y antena parabólica, se obtuvo la siguiente información, en miles de pesos, sobre los gastos mensuales en:
    a) Costo del teléfono por familia $56 64 38 60 42 28 55 70 42 y 63 (Miles pesos).
    b) Costo promedio mensual en mantenimiento $83
    c) La distribución de costos de servicio mensual de la parabólica, es

COSTOS (Miles de pesos)	80,1-120	120,1-160	160,1-200	200,1-240	240,1-280
Nº DE FAMILIAS	1	3	2	3	1

Se pide calcular el costo total promedio mensual por familia, de los 3 servicios.

19. Los siguientes datos corresponden a los salarios mensuales, en miles de pesos, pagados por una empresa a su personal.

SALARIOS (Miles pesos)	500	600	700	800	900
Nº EMPLEADOS	10	16	35	26	13

En la empresa se presenta un conflicto laboral. El gerente propone un aumento del 7% mensual, para cada uno de los empleados. La Junta directiva propone un aumento de $49.000, para cada empleado. Se solicita su asistencia técnica, para que se discuta que es más ventajoso para la empresa y qué para cada grupo de empleados, según su clasificación por nivel de salarios. ¿Cómo actuaría usted?

20.

a)	$H_5 = 0{,}36$	$N_4 = 30$	$n_5 = 6$	$n = 5$
b)	$y_1 = 17$	$m = 5$	$c = 6$	$\bar{y} = 125$
c)	$y_i n_i = 240$	$\bar{y} = 8$	$n = 40$	
d)	$H_5 = 1{,}10$	$H_4 = 0{,}80$	$h_3 = 0{,}30$	

21. Señale y justifique la respuesta correcta a las siguientes afirmaciones: a) El promedio de las notas de un curso fue de 5,8. Las 25 mujeres obtuvieron un promedio de 3,8 y los hombres de 6,4. Luego el curso tiene más de 90 alumnos. b) Si $H_6 = 0{,}7$ y $H_3 = 0{,}3$ un 40% de las observaciones de la variable es menor que y_6 y mayor que y_4 c) El ingreso mensual medio de los obreros agrícolas es $635.000, el de los mineros es $754.000 y el de los industriales es $864.000, es decir, el ingreso medio del conjunto de obreros es de $754.000. d) Para calcular la media geométrica, es preciso que todos los intervalos sean iguales. e) El 75% de los empleados públicos son hombres, mientras que en el sector privado es el 81%. Esto es, hay un 22% de mujeres empleadas en ambos sectores. f) Si se multiplica por 2 todas las frecuencias relativas, el valor de la media aritmética también se duplica.

22. Contestar si es *cierto o falso* a los siguientes puntos: a) La suma de las desviaciones es igual a cero, cuando se toma con respecto a la media. b) Un grupo de valores puede tener más de una media aritmética. c) Para calcular las marcas de clase, se suma el límite inferior al superior del intervalo y se divide por dos.

23. Un laboratorio de control de calidad de una empresa, quiere lanzar al mercado un nuevo spray ambientador, para determinar el tiempo de permanencia del aroma; los resultados fueron observados cada hora, habiéndose obtenido:10 horas, 5 apartamentos; 3 horas, 3 apartamentos; 5 horas, 10 apartamentos; 4 horas, 7 apartamentos; 6 horas, 16 apartamentos; y al final, 8 horas, 9 apartamentos. a) Cuál es la población? b) Cuál es la muestra? c) Cuál es la variable d) Qué tipo de variable es? e) Cuál es el promedio de duración del aroma? f) Elabore una gráfica que la represente.

24 En una empresa constructora de vivienda, los jornales mensuales de los obreros tienen una media de $655.000. Como solución a un conflicto laboral se proponen dos alternativas: a) un aumento del 8%; b) un aumento del 5%, más una bonificación mensual de $25.000 a cada obrero. ¿Cuál de las dos alternativas le conviene aceptar al sindicato, si se quiere que la distribución de los salarios sea lo mejor posible?

25. El promedio de salario mensual para un grupo de obreros es de $662.000. El gerente de la firma resuelve reconocerles un aumento diario por cada obrero de $3.000; si consideramos los meses de 30 días. ¿Cuál será el nuevo promedio de salario mensual?

26. Un almacén de artículos electrónicos, se dedica a la venta de cuatro líneas de producto. Se sabe que el año anterior el total de ventas (Millones de $) y margen de utilidad por línea fueron los siguientes:

LÍNEA DE PRODUCTOS	VENTAS	MARGEN DE UTILIDAD
1 - Televisores……	214.000	12,6 %
2 - Neveras……	90.000	5,8%
3 - Equipos de sonido….	183.000	9,3%
4 - Lavadores……..	75.000	4,6%
TOTAL……..……..	562.000	-------

Se pide calcular: a) La media aritmética simple del margen de utilidad. b) La media ponderada del margen de utilidad. c) ¿Cuál de las dos medias es más representativa? ¿Por qué?

27. Una firma de producción múltiple, tiene cuatro líneas de productos. Durante el mes de octubre del presente año, los resultados obtenidos al realizar las operaciones de rutina de calidad, fueron las siguientes:

LÍNEAS DE PRODUCCIÓN	UNIDADES PRODUCIDAS	PORCENTAJES DEFECTUOSOS
A	8.300	0,8
B	12.600	1,1
C	24.300	2,6
D	15.800	1,4
TOTAL	61.000	-------

Determinar el promedio porcentual de artículos defectuosos producidos por dicha firma en el mes de octubre del presente año.

28. Dos fábricas tienen 80 y 120 empleados, respectivamente. El salario promedio mensual para el total de empleados es de $620.000. Sabiendo, además, que los empleados de la primera de ellas, ganan en promedio $58.500 más que los empleados de la segunda.¿Cuál es el salario promedio mensual de los empleados en cada una de las fábricas?

29. Un grupo de 400 empleados se divide en operarios y técnicos, con un salario promedio de $980.600. Los salarios promedios para los dos grupos son $725.730 y $1.076.500. ¿Cuántos operarios y cuántos técnicos hay?

30. Tres almacenes tienen un total de 80 vendedores. Los dos primeros pagan mensualmente $925.000 y $870.000 respectivamente y tienen además, el mismo número de vendedores. Se quiere saber, cuál es el salario promedio de los vendedores del tercer almacén, si se sabe, que el salario promedio de los vendedores de los tres almacenes es $890.000 y además, tiene 10 vendedores menos que el primer almacén.

31. Un grupo de empleados tienen los siguientes salarios mensuales (Miles de $).

$ 730	510	640	680	480	640	710	600	780

Los empleados piden un reajuste del 15%, pero el empresario ofrece un aumento del 10%, más una bonificación mensual de $20.000. ¿Cuál cree usted que debe ser la decisión más ventajosa para el empleado?

32. Conteste cada una de las siguientes preguntas, investigando.
a) ¿Qué significa el término “promedio”?

b) ¿Qué ventajas y desventajas tiene la aplicación de la media aritmética?
c) ¿Se pondera la media aritmética? ¿Por qué o por qué no?

33. Un inversionista compró acciones de varias empresas en la Bolsa de Bogotá, así: 500 acciones a $2.500 cada una; 1.200 a $5.800 cada una; 600 a $10.000 c/u y 2.500 a $800 c/u. a) ¿Cuál es la inversión total? b) ¿Cuál es el precio promedio por acción?

34. A un profesor de estadística le asignaron 5 cursos con igual número de alumnos, al finalizar el semestre, los promedios de calificación de cada curso fueron: 3,6; 4,1; 3,2; 3,8; y 4,6. a) ¿Cuál es la calificación promedio con ese profesor? b) Ahora suponga, que el número de alumnos por curso fue; 32; 40; 26; 34; y 15 ¿Cuál es el promedio de calificación por curso?

MEDIANA M_E

Es considerada también, al igual que la *Media*, como una *medida de tendencia central*. Su importancia es menor, sus fórmulas son rígidas ya que no admiten tratamiento algebraico, es por eso que tendremos que aplicar seis procedimientos de cálculo diferente, que lo hace engorroso, como lo veremos a continuación:

Se define como "*aquel valor de la variable que supera a no más de la mitad de las observaciones, al mismo tiempo, es superado por no más de la mitad de las observaciones"* en otras palabras, se puede definir como el "valor central". Se simboliza por M_e

Ventajas

- Esta medida tiene una definición bastante rígida y el concepto de ella es tan claro, que cualquier persona lo entiende, aún sin ser familiar su término.
- Los datos deben estar ordenados, pero los valores extremos no tienen ninguna incidencia en su cálculo.
- Tiene menor estabilidad en el *muestreo* que el promedio aritmético, pero mayor estabilidad que otras medidas.
- Hay situaciones en que la única medida de *tendencia central* que puede utilizarse es la *mediana*, especialmente cuando los valores extremos de una distribución de variable continua no están definidos

Desventajas

- No es tan conocida como la *media aritmética*.
- Se hace necesario ordenar los datos para poderla calcular.
- Esta medida no se adapta a cálculos posteriores aritméticos, por cuanto que si obtenemos las medianas de diferentes grupos, no se podrá calcular una mediana para el conjunto.
- Tampoco es sensible a cambios en los valores de los elementos, manteniéndose el valor central y los valores ordenados.

Cálculo en datos sin agrupar u originales

a) *Número impar de observaciones.* Cuando ésta medida la aplicamos en los datos originales o sin agrupar, lo primero que se debe hacer, es ordenarlos de menor a mayor o de mayor a menor.

Ejemplo 1. Supongamos se tienen los siguientes datos: y $x_1=2$ $x_2=18$ $x_3=4$ $x_4=12$ Observemos que la serie es *impar*, ya que $n=5$ por lo tanto ordenamos los datos, por ejemplo, de menor a mayor:

2	4	6	12	18

La *mediana* será igual al valor central $M_e=6$

b) *Número par de observaciones.* Ahora calcularemos la mediana, cuando se tenga un número par de observaciones. En estos casos, encontramos dos valores en el centro de la serie, por tal razón la mediana deberá ser el promedio de ellos.

Ejemplo 2. Supongamos que en vez de 5 datos se tengan 8 datos así:

$$x_1=8 \quad x_2=16 \quad x_3=4 \quad x_4=2 \quad x_5=20 \quad x_6=3 \quad x_7=12 \quad x_8=20$$

Ordenamos los datos de menor a mayor o viceversa;

2	3	4	8	12	16	20	20

$$\Rightarrow M_e=\frac{x_{j-1}+x_j}{2}=\frac{8+12}{2}=10$$

En los dos casos anteriores, correspondiente al cálculo de la mediana para datos sin agrupar, la localización de ella, luego de ser ordenados, se facilita aplicando el siguiente procedimiento:

$\frac{n+1}{2}=\frac{5+1}{2}=3^a$ *posición*, la *mediana* en el primer caso o ejemplo 1, será $M_e=x_j=6$ y en el ejemplo 2, será:

$\frac{n+1}{2}=\frac{8+1}{2}=4{,}5^a$ *posición*, es decir, se encuentra entre la 4ª y la 5ª observación y la *mediana* será igual a $M_e=x_j=10$, el valor promedio de los dos valores centrales.

Cálculo en datos agrupados o en tablas de frecuencias

Veamos las fórmulas que se aplican cuando la variable es discreta y cuando es continua. En cada caso, tiene dos procedimientos

Variable discreta

Cuando $N_{j-1}=n/2$ La *mediana* será igual a: $M_e=\frac{y_{j-1}+y_j}{2}$

Cuando $N_{j-1}<n/2$ La *mediana* será igual a: $M_e=y_j$

Variable continua

Cuando $N_{j-1}=n/2$ La *mediana* será igual a: $M_e=y'_{j-1}$

Cuando $N_{j-1}<n/2$ La *mediana* será igual a: $M_e=y'_{j-1}+c\left[\frac{n/2-N_{j-1}}{n_j}\right]$

Pasos a seguir en el cálculo de la Mediana Variable discreta

Ejemplo 3

Caso A $\frac{n}{2} = \frac{30}{2} = 15$ Se localiza en la columna N_j

Cuando $N_{j-1} = n/2$ o sea $15 = 15$

Se dice que $M_e = \frac{y_{j-1} + y_j}{2}$

$M_e = \frac{2+3}{2} = \frac{5}{2} = 2{,}5$

	y_i	n_i	N_j	
	0	4	4	
	1	5	9	
y_{j-1}	→2	6	15→	N_{j-1}
y_j	→3	7	22→	N_j
	4	8	30	
	Σ	30	---	
	X_i	f_i	N_i	

Es decir que la *mediana* se encuentra localizada entre los valores 2 y 3, siendo igual a 2,5. También se puede comprobar el anterior resultado, transformando la distribución de frecuencias en una variable que nos muestre los datos originales o que éstos se encuentren sin agrupar.

0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4

La *mediana* se localiza entre los valores ubicados entre la posición 15ª y 16ª

Caso B $\frac{n}{2} = \frac{30}{2} = 15$ No se localiza en la columna

Cuando $N_{j-1} < n/2$ o sea que $9 < 15$

Se dice que $M_e = y_j$ $M_e = 2$

	y_i	n_i	N_j	
	0	4	4	
	1	5	9→	N_{j-1}
y_j	→2	7	16→	N_j
	3	8	24	
	4	6	30	
	Σ	30	---	
	X_i	f_i	N_j	

La mitad de las observaciones es 15, por lo tanto la 15ª y la 16ª corresponden al valor de 2, el cual se toma como el valor de la *Mediana*, como se puede ver a continuación:

Tal como se hizo en el ejercicio anterior, en datos sin agrupar, el valor central lo ocupa el 2

0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4

Mediana Datos agrupados: Variable discreta

a) $M_e = \frac{y_{j-1} + y_j}{2}$ Cuando $N_{j-1} = n/2$

b) $M_e = y_j$ Cuando $N_{j-1} = n/2$

Pasos a seguir en el cálculo de la mediana en la Variable continua

Datos agrupados: Variable continua

a) $M_e = y'_{j-1}$ Cuando $N_{j-1} = n/2$

b) $M_e = y'_{j-1} + c\left[\frac{n/2 - N_{j-1}}{n_j}\right]$ Sí $N_{j-1} < n/2$

Ejemplo 4 .

Caso A $\frac{n}{2} = \frac{40}{2} = 20$

Cuando $N_{j-1} = n/2$ en este caso $N_{j-1} = 20$

y $N_j = 30$ así que el valor de la *mediana*

$M_e = y'_{j-1} = 70$

Es decir corresponde al valor central que toma la variable

$y'_{i-1} - y_i$	y_i	n_j	N_J	
46,1 – 54	50	3	3	
54,1 – 62	58	8	11	
62,1 – 70	66	9	20 →	N_{j-1}
y'_{j-1} → 70,1 – 78	74	10	30 →	N_j
78,1 – 86	82	8	38	
86,1 – 94	90	2	40	
Σ	--	40	---	
$X'_{i-1} - X'_i$	x_i	f_i	N_J	

Caso B $\frac{n}{2} = \frac{40}{2} = 20$ No se localiza en la columna

En este caso $N_{j-1} < n/2$ por lo tanto se tendrá que $N_{j-1} = 11$ y $N_j = 22$

La *mediana* será igual a:

$$M_e = y'_{j-1} + c\left[\frac{n/2 - N_{j-1}}{n_j}\right]$$

$y'_{i-1} - y_i$	y_i	n_j	N_J	
46,1 – 54	50	3	3	
54,1 – 62	58	8	11 →	N_{j-1}
y'_{j-1} → 62,1 – 70	66	11	22 →	N_j
70,1 – 78	74	7	29	
78,1 – 86	82	9	38	
86,1 – 94	90	2	40	
Σ	--	40		
$X'_{i-1} - X'_i$	x_i	f_i	N_j	

$$M_e = 62 + 8\left[\frac{20 - 11}{11}\right] = 62 + \left[\frac{72}{11}\right] = 62 + 6{,}55 = 68{,}55$$

Algunos autores, utilizan otras simbologías para realizar el anterior cálculo, pero el procedimiento y resultado son siempre iguales

$$M_e = L_i + i\left[\frac{n/2 - F_{j-1}}{f_i}\right]$$

Pasos a seguir en el cálculo de la *Mediana*

a) Obtener en primer lugar las frecuencias absolutas.

b) Localizar la mitad de las observaciones, por medio de la expresión $n/2$.

c) El resultado anterior lo buscamos en la columna de las frecuencias absolutas acumuladas. Si no aparece, se toma el valor inmediatamente anterior y lo simbolizamos por N_{j-1} y al inmediatamente superior o posterior por N_j. En caso contrario, cuando dicho valor aparece registrado en esa columna, será simbolizado N_{j-1} y el inmediatamente superior posterior por N_j.

d) Cuando aparece $n/2$, se dice que es igual a N_{j-1} ($N_{j-1} = n/2$); en caso contrario, que no aparezca, se dirá que $N_{j-1} < n/2$. En ambos casos, dependiendo de si la variable es discreta o continua, se aplica la fórmula respectiva, tal como aparece en el resumen anterior.

MODO – MODA O VALOR MODAL M_D

Se define como *"el valor de la variable que más se repite"* o *"aquel valor que presenta la máxima frecuencia"*. Puede suceder que una distribución tenga *dos Modas*, en este caso se dice que la distribución *es Bimodal*, en el caso que haya más de dos modas, se dice que es *plurimodal* o *multimodal*. Es la única medida de posición que puede ser utilizada en atributos, es decir, cuando la característica es cualitativa. *Modo, Moda o valor Modal*, medida originalmente concebida, como aquel valor de la variable, que presenta el mayor número de observaciones, es decir, el valor de la variable que más veces se repite. Se empleó formalmente en 1.984 por Karl Pearson, ya que no era utilizado anteriormente.

Desventajas

- El hecho que el *modo* nos indique el punto de mayor concentración, lo hace tal vez, la mejor medida de *tendencia central*. Cuando una distribución es muy *asimétrica*, claramente se ve que el *Modo* es el más representativo del grupo, y en algunos casos, si el *modo* y el promedio *aritmético* son significativamente diferentes al valor, es preferible usar el *modo*.
- En series *plurimodales*, el modo permite dividir la distribución con fines de estratificación.
- Se considera de uso riguroso en variables continuas, aunque puede extenderse a casos discontinuos.

Ventajas

- Se le considera con cierto grado de dificultad de cálculo en datos agrupados, siendo su resultado de poca confianza, especialmente cuando la amplitud no es constante, en este caso, si se hace necesario se debe seguir un determinado proceso.
- Es muy inestable en el muestreo.
- No puede ser utilizado en procesos algebraicos posteriores.
- No es sensible, especialmente cuando hay cambios en los valores de la variable, salvo que no afecten a su propio valor.

Datos sin agrupar u originales

Ejemplo 1. Consideremos los siguientes datos: 5, 10, 8, 5, 10, 18, 5, 12, 5, 12 calculemos la media, mediana y la moda. $n = 10$

$$Media \rightarrow \bar{x} = \frac{\sum x_i}{n} = \frac{90}{10} = 9$$

Mediana $\rightarrow$ Ordenamos los datos de mayor a menor o de menor a mayor

$$5 \quad 5 \quad 5 \quad 5 \quad \underset{M_e}{\overset{\overset{8}{\frown}\quad\overset{10}{\frown}}{\lfloor\quad\rfloor}} \quad 10 \quad 12 \quad 12 \quad 18$$

Posición: $\frac{n+1}{2} = \frac{10+1}{2} = 5,5^a$ La mediana se encuentra entre la 5^a y la 6^a observación

Siendo: $M_e = \frac{8+10}{2} = 9$ $\qquad M_e = \frac{x_{j-1} + x_j}{2}$

Moda $\rightarrow$ Corresponderá a 5, siendo el valor de la variable que más se repite $\rightarrow$ $M_d = x_j = 5$

Ejemplo 2. Con dos casos nuevos y con diferentes valores, calculemos otra vez las tres (3) medidas que se aplicaron en el ejemplo anterior.

Caso 1 Con los siguientes datos 5 10 8 12 18 5 12 5 12

$Media \rightarrow \bar{x} = \frac{8}{9} \cong 9,6$

Mediana → Ordenamos los datos de mayor a menor o de menor a mayor

Posición: $\frac{n+1}{2} = \frac{9+1}{2} = 5^a$ 5 5 5 8 [10] 12 12 12 18

$M_e = x_j = 10$

Moda → ncontramos en esta serie de 9 datos, que el 5 y el 12 se repiten veces n mero mayor de repetición por lo tanto ay dos modas y en este caso se dice que es *bimodal.* Si ay mas de dos modas se abla de *Plurimodal* o *Multimodal.*

$M_d = 5$ y $M_d = 12$

Caso 2 Con los datos 13 4 8 12 10 12 16 6 11

Media $\rightarrow \bar{x} = \frac{80}{10} = 8$

Mediana → Ordenamos los datos de mayor a menor o de menor a mayor

$M_e = \frac{10+11}{2} = 10,5$

Moda → No hay, dado que ninguno de los valores de la variable se repite más de una vez.

Datos agrupados

En datos agrupados, tanto en la variable discreta como en la continua, la *Moda* corresponderá a aquel valor que representa la mayor frecuencia. Hay que observar en la variable continua que la amplitud del intervalo sea constante.

a) *Variable discreta*

y_i	n_i	
0	4	
1	5	
2	7	
3	8	n_j
4	6	
Σ	30	
X_i	f_i	

(y_j → 3)

$M_d = y_j = 3$

b) *Variable continua*

$y'_{i-1} - y'_i$	y_i	n_j	
46,1 – 54	50	3	
54,1 – 62	58	8	n_{j-1}
62,1 – 70	66	11	n_j
70,1 – 78	74	7	n_{j+1}
78,1 – 86	82	9	
86,1 _ 94	90	2	
Σ	--	40	
$X'_{i-1} - X'_i$	X_i	f_i	

Amplitud constante

$M_d = y_j = 66$

Como se observa en el ejercicio anterior, primero localizamos el valor de la frecuencia más alta y luego el valor al frente de la frecuencia, en la columna de la variable, corresponderá al valor del Modo. La frecuencia la simbolizaremos por n_j, la frecuencia inmediatamente anterior a ese por n_{j-1} y la frecuencia inmediatamente superior por n_{j+1}.

El requisito primordial al calcular el *Modo*, es que la frecuencia absoluta debe ser un valor mayor que n_{j-1} y además, menor a: n_{j+1}. $\boxed{n_{j-1} < n_j < n_{j+1}}$

Existen otras fórmulas, menos importantes, para calcular el *Modo*, diseñadas especialmente para ser aplicadas cuando la variable es continua y cuando la amplitud del intervalo es constante.

a) $M_d = y'_{j-1} + c\left[\dfrac{n_{j+1}}{n_{j-1} + n_{j+1}}\right]$ b) $M_d = y'_{j-1} + c\left[\dfrac{n_j - n_{j-1}}{(n_j - n_{j-1}) + (n_j - n_{j+1})}\right]$

c) $M_d = y'_{j-1} + c\left[\dfrac{n_{j+1} - n_{j-1}}{2(2n_j - n_{j-1} - n_{j+1})}\right]$ d) $M_d = y'_{j-1} + c\left[\dfrac{n_{j-1} - n_{j+1}}{n_j + n_{j+1} + n_{j-1}}\right]$

Con los datos correspondientes al último ejercicio, de una variable continua, que fue utilizado para calcular la M oda, con datos agrupados, será utilizado en la aplicación de las fórmulas anteriores.

a) $M_d = 62 + 8\left[\dfrac{7}{8+7}\right] = 62 + \dfrac{56}{15} = 62 + 3{,}73 = 65{,}73$

b) $M_d = 62 + 8\left[\dfrac{11-8}{(11-8)+(11-7)}\right] = 62 + 8\left[\dfrac{3}{7}\right] = 62 + 3{,}43 = 65{,}43$

c) $M_d = 62 + 8\left[\dfrac{7-8}{2(2(11)-8-7)}\right] = 62 + 8\left[\dfrac{-1}{14}\right] = 62 - 0{,}57 = 61{,}43$

d) $M_d = 62 + 8\left[\dfrac{8-7}{11+7+8}\right] = 62 + 8\left[\dfrac{1}{26}\right] = 62 + 0{,}31 = 62{,}31$

Hemos calculado la Moda en datos agrupados, correspondiente a una variable continua, con una amplitud constante. Ahora veamos el procedimiento a seguir cuando no lo es. Por ejemplo:

$y'_{i-1} - y'_i$	y_i	n_j	c_i	n_j/c_i
46,1 – 50	48	3	4	0,75
50,1 – 62	56	8	12	0,67
62,1 – 77	69,5	11	15	0,73
77,1 – 82	79,5	7	5	1,40
82,1 – 84	83	9	2	4,50
84,1 – 90	87	12	6	2,00
Σ	--	50	--	--
$X'_{i-1} - X'_i$	X_i	f_i	i	f_i / c_i

Observemos en la tabla adjunta, que el mayor valor registrado en la columna de n_i/c_i es 4,50, por lo tanto el valor correspondiente al Modo, estará al frente, en la columna de las marcas de clase (y_i). En este caso, se tendrá que:

$$M_d = y_J = 83$$

Relación entre la *Media, Mediana* y el *Modo*.

Son tres situaciones que se pueden presentar entre estas tres medidas de posición.

$\boxed{M_x = M_e = M_d}$ $\boxed{M_x < M_e < M_d}$ $\boxed{M_x > M_e > M_d}$

En un polígono de frecuencias, la Moda se ubica en la cima, es decir en la parte superior de la curva o línea poligonal. Como término parroquial, corresponde al llamado horas pico, siendo el tiempo de mayor congestión vehicular dentro de la ciudad o de mayor consumo de energía, durante el día.

Algunos calculan la moda mediante la fórmula: $M_d = M_x - 3(M_x - M_d)$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

35. Con el siguiente ejercicio, referente a los datos obtenidos sobre el número de hijos por familia, se pide calcular la media, mediana y modo. (n = 20)

y_i	0	2	4	6	7	X_i
n_i	2	3	7	4	4	f_i

36. Con el siguiente ejercicio, referente a una distribución asimétrica, se pide calcular: la media, mediana y modo. (n = 50)

$y'_{i-1} - y'_i$	2,75 - 4,25	4,25 - 5,75	5,75 - 7,25	7,25 - 8,75	$x'_{i-1} - x'_i$
n_i	4	16	25	5	f_i

37. Con la siguiente serie de números, calcule: 2.000 2.500 3.000 3.250 4.000 5.000 6.500 7.500 50.000 a) La media b) Mediana c) Moda
d) ¿Qué factor de la variable afecta la media en esta serie?
e) ¿Cuál de las medidas calculadas estima usted más representativa?

38. Los sueldos mensuales de 70 empleados de una oficina, son los siguientes:

OCUPACIÓN	Nº DESEMPLEADOS	SUELDOS
Recepcionista.....	2	642.000
Mecanógrafa.....	12	751.000
Secretaria.....	8	758.000
Aux. Contabilidad.....	10	794.000
Técnicos Eléctricos....	24	801.000
Téc. Mecánicos.....	14	911.000

Se pide calcular la media, mediana y modo.

39. Utilizando la siguiente distribución sobre resistencia a la tensión de láminas de acero (en kgrs/) obtenga la: media, mediana y modo.

RESISTENCIA	10 a 20	20 a 30	30 a 40	40 a 50
% DE LÁMINAS	0,20	?	?	0,12

Nota: Se sabe que la resistencia promedio es de 29,4 kgrs/

40. El presidente de un sindicato, clasificó a los compañeros de trabajo según el salario mensual en la siguiente forma:

SALARIO MENSUAL (Miles $)	Nº DE TRABAJADORES
Menor o igual a 600	30
600,1 - 800	60
800,1 - 1.000	45
1.000,1 - 1.200	15
1.200,1 y más	50

41. Calcular el Modo utilizando las diferentes fórmulas presentadas aquí.

SALARIOS ($)	Nº EMPLEADOS
648.000,1 – 668.000	120
668.000,1 – 688.000	38
688.000,1 – 708.000	22
708.000,1 – 728.000	10
728.000,1 – 748.000	6
748.000,1 – 768.000	4

42. Con base en los siguientes datos y medidas de tendencia central, determinar si hay simetría o asimetría y cuál es su dirección.

a) $\overline{x} = 68$ $M_e = 74$ $M_d = 80$ c) $\overline{x} = 74$ $M_e = 74$ $M_d = 74$

b) $\overline{x} = 80$ $M_e = 74$ $M_d = 68$ d) $\overline{x} = 74$ $M_e = 74$ $M_d = 60$

43. Consideremos que un vendedor de cigarrillos, vende de lunes a sábado el siguiente número de cajetillas. ¿Cuál es la media, la mediana y la moda?

a) 55 60 65 60 65 65 b) 55 60 64 70 65 50 c) 40 36 48 35 100 35

44. Con los siguientes datos, correspondientes a dos submuestras, se pide calcular:

x_i	4	6	7	6	10	6	4	15	6
y_i	10	18	12	20	8	10			

a) La mediana, moda y media aritmética para cada una de las variables.
b) La media conjunta de las dos submuestras.

45. Con los siguientes datos calcular la media, mediana y el modo.

$y'_{i-1} - y'_i$	3,1-8	5,1-9	10,1-16	16,1-24	24,1-30	30,1-32	$X'_{i-1} - X'_i$
n_i	8	8	14	16	18	10	f_i

46 Una muestra realizada a 12 profesores de tiempo completo, en Universidades privadas de la capital, se encontró que su sueldo quincenal en miles de $, es de: 1.800, 720, 750, 1.680, 900, 720, 840. 810, 720, 810, 720, 840. Calcular la media, mediana y la moda de estos salarios ¿Cuál cree usted que sea el mejor promedio? ¿Por qué?

47. Si al realizar los cálculos en una distribución, se obtienen los siguientes resultados: $\overline{x} = 36$ $M_e = 32$; $M_d = 23$ y se desea corregirlos ¿Cuáles serían los nuevos promedios en los casos siguientes: a) Cada uno de los valores se incrementó en 5 unidades? b) La variable está dada en años y se desea que esté dada en meses?

48. Una corporación financiera, en los últimos 9 meses ha hecho captaciones en dinero con interés anual de: 7,9% 8,3% 6,1% 6,8% 7,4% 8,6% 8,2% 7,6% y 8% ¿Cuál es la media, mediana y moda de estos valores?

49. Se tiene el control de llegada (en minutos) de un empleado a su oficina, cuando llega temprano se tendrá tiempo a su favor (+) y sus retardos podrán ser compensados (-) Se le realizó un control durante 8 días hábiles, así: 12 -8 15 -12 -6 10 -8 y -18. Calcular la media, mediana y la moda.

50. Con los siguientes datos, calcular la media, mediana y moda.

$y'_{i-1} - y'_i$	2,1-5	5,1-9	9,1-16	16,1-20	20,1-28	28,1-36	$X'_{i-1} - X'_i$
n_i	120	15	8	6	6	4	f_i

MEDIA GEOMÉTRICA M_G M_o G

Se considera como una medida de posición o promedio, pudiendo utilizarse cualquiera de estos símbolos y el resultado obtenido deberá ser menor al de la *Media aritmética.* Se aplica en todos aquellos casos en que la variable muestra un crecimiento geométrico, como sucede con la población o que se encuentre relacionada con ella, además, como en el caso de un capital colocado a una tasa de interés compuesto, durante un determinado período de tiempo.

La *Media geométrica* se define como "*la raíz enésima del producto de los valores que toma la variable"*

$$M_0 = \sqrt[n]{\pi x_i} = \sqrt[n]{x_1 . x_2 . x_3 ..... x_n}$$

Observemos que el signo π se lee *productoria*, siendo este símbolo la forma de expresar que los valores correspondientes a la variable deben estar multiplicados entre sí.

Ventajas

- Se define rígidamente mediante una fórmula matemática.
- Se utiliza cuando se quiere dar importancia a pequeños valores que toma la variable.
- Es sensible a cualquier cambio que se haga en los valores de la distribución.
- Su valor no es muy influenciado por valores extremos grandes, tal como sucede con los otros promedios.
- Se hace indispensable, cuando se desea calcular el promedio en una serie de valores que están dados en progresión geométrica o aproximadamente.
- Su resultado puede ser utilizado en procesos estadísticos posteriores, puesto que los promedios geométricos en diferentes muestras, pueden ser combinados para el total de la muestra.

Desventajas

- Es bastante complicado el proceso de cálculo.
- Si cualquier valor de la serie es cero, el promedio será cero, dependiendo de la fórmula empleada.
- Si un valor de la serie es negativo, su valor será negativo o imaginario.

Cálculo en datos sin agrupar.

Ejemplo 1. Supongamos que se tienen 5 observaciones ($n = 5$) cuyos valores son: x_i: 6 4 8 2 5. Se pide calcular la media geométrica

Solución:

$$M_0 = \sqrt[5]{6.4.8.2.5} = \sqrt[5]{1.920} = 4{,}54$$

> Esta fórmula presenta algunos problemas en su aplicación: a) Si un valor de la variable es cero el producto será cero y su raíz cero. b) Si uno de los valores es negativo, el producto será negativo, por lo tanto se tendrá una raíz imaginaria. c) Sí el valor de n es grande, será tedioso su cálculo.

Frente a la anterior situación es preferible utilizar los logaritmos. Veamos el procedimiento:

$$M_0 = \sqrt[n]{\pi x_i} = \pi x_i^{1/n} \quad \Leftrightarrow \quad \log M_0 = \frac{\Sigma \log x_i}{n}$$

Reemplacemos y se tendrá que $\log M_0 = \dfrac{\Sigma \log x_i}{n} = \dfrac{\log 6 + \log 4 + \log 8 + \log 2 + \log 5}{5}$

$$\log M_0 = \frac{0{,}77815 + 0{,}60206 + 0{,}90309 + 0{,}30103 + 0{,}69897}{5} = \frac{3{,}28330}{5} = 0{,}65666$$

M_0 = *anti* log de 0,65666 = 4,54 El mismo resultado obtenido anteriormente.

Cálculo en datos agrupados.

Ejemplo 2.

y_i	n_i	$\log y_i$	$n_i \log y_i$
1	4	0	0
2	5	0,30103	1,50515
3	7	0,47712	3,33985
4	8	0,60206	4,81648
6	6	0,77815	4,66891
Σ	30	--	14,33039
X_i	f_i	$\log X_i$	$f_i \log X_i$

Solución:

$$\log M_0 = \frac{\Sigma n_i \log y_i}{n} = \frac{14{,}33039}{30} \equiv 0{,}477680$$

$$M_0 = \textit{anti}\log 0{,}477680 = 3$$

$$\log M_0 = \frac{\Sigma f_i \log x_i}{n}$$

Ejemplo 3.

y_i	n_i	$y_i^{n_i}$
1	4	1
2	5	32
3	7	2.187
4	8	65.536
6	6	46.656
Σ	30	--
X_i	f_i	$X_i^{f_i}$

Solución:

$$M_0 = \sqrt[n]{\pi y_i^{n_i}}$$

$$M_0 = \sqrt[30]{1x32x2.187x65.536x46.656} = 3{,}00$$

Como lo podemos observar, en los dos ejercicios que se acaban de realizar los resultados son exactamente iguales. Este último cálculo hizo ver que el anterior procedimiento fuera se considerara más engorroso.

Ejemplo 4. La media aritmética de 3 números es 8, la mediana es 8 y la media geométrica es $M_0 = \sqrt[3]{480}$. Se pide calcular los valores de esos tres números.

Solución:

$M_1 = 8 \quad M_e = 8$

$24 = x_1 + x_2 + x_3$

$M_0 = \sqrt[3]{480}$

$24 = x_1 + 8 + x_3$

$24 - 8 = x_1 + x_3$

$16 = x_1 + x_3$

$\sqrt[3]{480} = \sqrt[3]{x_1 x_2 x_3}$

$480 = x_1.(8).x_3$

$60 = x_1.x_3$

$x_1 = \dfrac{60}{x_3}$

Reemplazando $16 = x_1 + x_3 \Leftrightarrow 16 = \dfrac{60}{x_3} + x_3 \Leftrightarrow 16x_3 = 60 + x_3^2$

Despejando se tiene que $x_3^2 - 16x_3 + 60 = 0$, de donde: $x_1 = 6 \quad x_2 = 8 \quad x_3 = 10$

(Teniendo en cuenta dos números que sumados den 16 y multiplicados 60)

Ejemplo 5. Supongamos que un país tuvo en 1997 una población de 8 millones y subió a 12 millones para el 2011. ¿Cuál será el promedio (media) de poblacion para ese período?

$M_0 = \sqrt{x_1.x_2} = \sqrt{8(12)} = \sqrt{96} = 9{,}8$ Millones de personas

Ejemplo 6. Supongamos que un capital de $50.000 sea colocado a una tasa anual del 24% el primero de enero del 2.008. Si se capitaliza los días primero de enero, calcule el promedio del dinero invertido entre el 31 de diciembre del 2.008 y el 31 de diciembre del 2011.

Solución:

AÑOS	x_i	$\log x_i$
2008	50.000	4,698970
2009	62.000	4,792392
2010	76.880	4,885813
2011	93.331,2	4,979235
Σ		19,356410

$$\log M_0 = \frac{\Sigma \log x_i}{n}$$

$$\log M_0 = \frac{19{,}356410}{4} = 4{,}8391025$$

$$M_0 = \textit{anti}\log\ 4{,}8391025 = 69.040{,}27$$

(Promedio de dinero invertido durante el período)

MEDIA ARMÓNICA: $M_H \quad M_{-1}$

Este promedio, menos importante que los tres anteriores mencionados, se aplica en especial cuando la variable está dada en forma de tasas o cuando se trata calcular la velocidad media. Vale la pena decir, que esta medida le da gran importancia a aquellos valores pequeños de la variable, contrario a la media aritmética.

Se define, diciendo que: *dada una serie de datos, el inverso de la media armónica, es igual a la media aritmética del inverso de los valores de la variable.*

$$\frac{1}{M_H} = M\left[\frac{1}{x_i}\right] \qquad \text{Por lo tanto se tendrá que} \quad M_H = \frac{n}{\sum\left[1/x_i\right]} \quad \text{ó} \quad M_H = \frac{N}{\sum\left[1/x_i\right]}$$

Ventajas

- Se usa con preferencia en el cálculo de la velocidad media.
- De gran utilidad cuando la variable está dada en forma de tasas.

Desventajas

- Que un valor de la variable sea cero.
- El promedio armónico está rígidamente definido por una fórmula matemática, su valor depende de cada uno de los datos de la distribución y su resultado no puede ser utilizado en cálculos posteriores.

Datos no agrupados

Ejemplo 1. Supongamos que se dispone de 10 observaciones, cuyos valores son:

x_i = 8 2 8 6 2 2 6 8 2 4 Con ellos se desea calcular la media armónica, aplicando la fórmula anterior.

$$M_{-1} = \frac{10}{1/8 + 1/2 + 1/8 + 1/6 + 1/2 + 1/2 + 1/6 + 1/8 + 1/2 + 1/4} = 3{,}38$$

Si comparamos este resultado, con el de la media aritmética aplicada a los mismos datos, siendo igual a 4,8, notaremos que es inferior, hecho que se dará en todos los ejercicios que realicemos, siendo: $M_{-1} < M_x$

Datos agrupados

Ejemplo 2. Con los datos correspondientes a una tabla de frecuencias, tal como aparece a continuación, tanto para la *variable discreta* como la continua, la Media armónica será:

1- VARIABLE DISCRETA

y_i	n_i	n_i/y_i
1	4	4,00
2	5	2,50
3	7	2,33
4	8	2,00
6	6	1,00
Σ	30	11,83
X_i	f_i	f_i/X_i

(1) $M_H = 30/11{,}83 =$

$M_H = 2{,}54$

(2) $M_H = 50/0{,}70 =$

$M_H = 71{,}43$

2 - VARIABLE CONTINUA

$y'_{i-1} - y'_i$	y_i	n_i	n_i/y_i
46,1 – 50	50	3	0,06
50,1 – 62	58	8	0,14
62,1 – 77	66	11	0,17
77,1 – 82	74	7	0,09
82,1 – 84	82	9	0,11
84,1 – 90	90	12	0,13
Σ	--	50	0,70
$X'_{i-1} - X'_i$	X_i	f_i	f_i/X_i

Lo mismo que en datos sin agrupar, los anteriores resultados en datos agrupados, deben ser inferiores a los obtenidos aplicando la media aritmética.

Fórmula aplicada en datos agrupados

$$M_H = \frac{n}{\sum\left[n_i / y_i\right]}$$

$$M_H = \frac{N}{\sum\left[f_i / x_i\right]}$$

Ya se mencionó al definir la *media armónica*, que una de sus aplicaciones se realizaba al calcular la *velocidad media*, es decir, cuando hay necesidad de relacionar tiempo con velocidad. Supongamos disponer de la siguiente información:

$S_i = V_i\, t_i$ $\quad S_i$= *espacio* $\quad V_i$= *velocidad* $\quad t_i$= *tiempo* $\quad V_m$= *velocidad media*

$$V_m = \frac{S}{\sum\left[S_i / V_i\right]} \qquad \text{Siendo} \quad S = S_1 + S_2 + \ldots\ldots\ldots \quad \text{Total de espacio}$$

Ejemplo 3. Consideremos, un auto recorre 60 km a una velocidad promedio de 100 km. por hora y luego, los siguientes 40 km. a una velocidad media de 80 km. por hora. ¿Cuál será la velocidad media del recorrido total?

Solución:

$V_1 = 100 \quad V_2 = 80 \quad S_1 = 60 \quad S_2 = 40$

$$V_m = \frac{100}{60/100 + 40/80} = \frac{100}{0{,}6 + 0{,}5} = \frac{100}{1{,}1} = 90{,}9 \text{ Velocidad media}$$

A continuación veremos una forma errónea en la aplicación de la media aritmética. En este caso en particular, debe aplicarse la media armónica

Ejemplo 4. Un grupo de trabajadores producen 120 papeleras para escritorio, en un promedio diario de 12 papeleras. Una vez terminado ese contrato, se dedican a producir otras 120 papeleras, a razón de 8 por día. Se desea determinar la productividad diaria en la elaboración de las 240 papeleras.

Solución:

Si calculamos la media aritmética obtendríamos un resultado erróneo

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{12 + 18}{2} = 15 \quad \text{Papeleras diarias}$$

Se observa que las primeras papeleras se realizan en 10 días (120/12 = 10) y las siguientes papeleras en 6,67 días (120/8 = 6,67); las 240 papeleras se harían en 16,67 días; si son 16,67 y el rendimiento por día es de 15 se tendrían un total de 250 papeleras, cantidad diferente a las 240 papeleras del problema

Mediante la media armónica se tiene: $\quad M_{-1} = \dfrac{2}{1/12 + 1/18} = \dfrac{2}{0{,}13888} = 14{,}4$

La aplicación de la media no es correcta, debido a que existe una relación inversamente proporcional. Con la media armónica, el número de papeleras producidas diariamente es de 14,4, que a la vez multiplicadas por 16,67 días nos daría un total de producción de 240 papeleras durante ese período, diferente a las 250 como resultado al aplicar la media.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

51. Calcule la media geométrica y la media armónica en la siguiente distribución de seis intervalos de amplitud constante, de la cual se sabe:

$n=150 \quad y_5 = 60 \quad n_1 = n_6 \quad n_3 = n_4 = 30 \quad n_2 = n_5 = n_1 + 5 \quad \overline{y} = 45$

$N=150 \quad X_5 = 60 \quad f_1 = f_6 \quad f_3 = f_4 = 30 \quad f_2 = f_5 = f_1 + 5 \quad \overline{X} = 45$

52 Halle la media aritmética, media geométrica, modo y mediana con los siguientes datos sin agrupar: 3 6 13 22 46 89 y conteste ¿Cuál de estas medidas es mejor y por qué?

53. Con los siguientes datos: 2 5 8 12 y 20
    a) Calcular la media aritmética, media geométrica y media armónica.
    Observar que $M_1 > M_0 > M_{-1}$
    b) Calcular la media, la mediana y la moda. Observar en cuál de las dos situaciones se presenta: $M_1 > M_e > M_d$ o $M_1 < M_e < M_d$

54. Los gastos (miles de millones de $) en publicidad de 50 empresas comerciales, durante el mes de diciembre de 2011 agruparon en 4 clases de amplitud constante, de la cual se sabe:

$y_1 = 3{,}5 \quad y'_4 = 8{,}75 \quad n_1 = 4 \quad N_2 = 20 \quad n_3 = 25$

$X_1 = 3{,}5 \quad X'_4 = 8{,}75 \quad f_1 = 4 \quad F_2 = 20 \quad f_3 = 25$

Se pide calcular la media aritmética y la media geométrica.

55. Una persona maneja su automóvil durante 400 kilómetros. Los primeros 120 km viaja a razón de 60 km por hora; los siguientes 120 a 100 km por hora; el 25% del total lo hace a razón de 80 km por hora. ¿A qué velocidad debe viajar el resto, para tener en total una velocidad promedio de 70 km por hora?

56. Un avión vuela una distancia de 900 km. Si cubre el primero y el último tercio del viaje a 250 km por hora y el tercio medio a 300 km por hora, se pide calcular su velocidad media.

57. Una persona viaja durante 4 días. Diariamente recorre 200 km, pero maneja el primero y el último día a 50 km por hora, el segundo a 55 y el tercero a 60 km por hora. ¿Cuál es la velocidad media durante el viaje?

58. Una persona viaja en coche de la ciudad "A" a la "B" con una velocidad media de 40 km por hora y vuelve de la ciudad "B" a la "A" con una velocidad media de 60 km por hora, hallar la velocidad media del viaje completo.

59. Las ciudades A, B y C, son equidistantes entre sí. Un motorista viaja de A a B a 30 km por hora, de B a C a 40 km por hora y de C a A a 50 km por hora. Determinar el promedio de velocidad para el viaje completo.

60. Durante un mes se construyeron 134 kilómetros de carretera en la siguiente forma: la primera semana 3,6% del total, la segunda semana 7,6% del total, la tercera semana 15,3%, la cuarta semana,

24,5% y la última 49%. Halle la medida de tendencia central que represente mejor el promedio de la distribución, en kilómetros por semana.

61. Se sabe que dos obreros gastan en la ejecución de un trabajo 50 y 40 minutos respectivamente, ¿cuál es el tiempo medio requerido para hacerlo en conjunto?

62. Un grupo de trabajadores produce 140 papeleras para piso con una productividad de 14 papeleras diarias; una vez terminado ese contrato se dedica a producir otras 140 papeleras razón de 10 por día. Se desea determinar la productividad diaria en la elaboración de las 280 papeleras.

63. Una persona va a tres tiendas del barrio a comprar azucar; los precios son como siguen:

TIENDA	PRECIO
A	$2.250
B	$2.830
C	$2.570

Si la persona compra azúcar en dos formas diferentes. La primera forma consiste en que compra 3 paquetes en cada tienda. La segunda forma corresponde a comprar en cada tienda el equivalente de $10.000 en azúcar.

a) ¿Cuál es el precio promedio por paquete que la persona paga cuando compra el azúcar de la primera forma?
b) ¿Cuál es el precio promedio por paquete cuando lo compra de la segunda forma?

64. En una tabla de frecuencias, se clasifican 50 datos utilizando 4 intervalos de igual magnitud. Se pide calcular la mediana y la media armónica sabiendo que:

$y_2 = 50$ $n_1 = 4$ $N_2 = 20$ $n_3 = 25$ $\overline{y} = 62{,}4$

65. Se sabe que dos obreros gastan en la ejecución de un trabajo 150 y 140 minutos respectivamente ¿Cuál es el tiempo medio requerido para hacerlo en conjunto?
66. Un fabricante dispone de $100.000, para la compra de cierta materia prima. Durante tres años invierte la misma cantidad de dinero. Si el precio promedio por kilo en los tres años sucesivos sube de $1.200 a $1.800 y a $3.600. ¿Cuál es el precio promedio que ha pagado el fabricante en dicho período?

67. Con la siguiente distribución:

$y'_{i-1} - y'_i$:	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	$\dfrac{X'_{i-1} - X'_i}{f_i}$
n_i :	3	7	15	25	10	

Se pide calcular la mediana y la media armónica.

MEDIA CUADRÁTICA: M_2

Se usa con muy poca frecuencia esta medida de tendencia central. Es un promedio que involucra a los cuadrados de las cantidades promediadas. La desviación estándar o típica de la cual hablaremos en la próxima unidad, es una media cuadrática. Es utilizada a menudo con el fin de reducir la media al mismo grado de los valores promediados, siendo la razón por la cual se calcula la raíz cuadrada de la misma, después de haber realizado el procedo de su promedación.

Se define como *la raíz cuadrada de la media aritmética, de los cuadrados de los valores que toma la variable".*

Fórmulas: Datos sin agrupar $\Leftrightarrow M_2 = \sqrt{\frac{\Sigma x_i^2}{n}}$

Datos agrupados: $M_2 = \sqrt{\frac{\Sigma y_i^2 n_i}{n}}$ $\quad M_2 = \sqrt{\frac{\Sigma X_i^2 f_i}{N}}$

Ventajas

- Es de importancia en algunos problemas de probabilidad, siendo teóricamente recomendable trabajar con el cuadrado de los valores.

Desventajas

- Su fórmula está definida rígidamente por una fórmula matemática.
- Se deja influenciar mucho por valores extremos, especialmente cuando éstos son demasiado grandes.

Datos sin agrupar

Ejemplo 1. Supongamos que se tienen 6 observaciones, cuyos valores son: $x_i =$ 2 6 5 4 6 8 y se desea calcular la media cuadrática de esos seis números.

$$M_2 = \sqrt{\frac{2^2 + 6^2 + 5^2 + 4^2 + 6^2 + 8^2}{6}} = 5{,}49$$ La media hubiese sido $\bar{x} = \frac{31}{6} = 5{,}17$

Ejemplo 2. Consideremos los datos de la tabla siguiente para calcular la media aritmética y la media cuadrática.

y_i	n_i	$y_i n_i$	$y_i^2 n_i$
1	4	4	4
2	5	10	20
3	7	21	63
4	8	32	128
6	6	36	216
Σ	30	103	431
X_i	f_i	$X_i f_i$	$X_i^2 f_i$

Media: $\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{103}{30} = 3{,}43$ $\quad M_2 = \sqrt{\frac{\Sigma X_i^2 f_i}{N}}$

Media cuadrática: $M_2 = \sqrt{\frac{\Sigma y_i^2 n_i}{n}} = \sqrt{\frac{431}{30}} = 3{,}79$

En los dos casos, donde hemos calculado las dos medias, permite observar que $M_1 < M_2$ (La Media aritmética es menor que la media cuadrática)

$3{,}43 < 3{,}79$

MEDIA CÚBICA: M_3

Este promedio, al igual que el anterior es poco conocido y por lo tanto de uso limitado. Se define como "*la raíz cúbica de la media aritmética de los cubos de los valores de la variable*"

$M_3 = \sqrt[3]{\frac{\Sigma x_i^3}{n}}$ $\quad M_3 = \sqrt[3]{\frac{\Sigma X_i^3}{N}}$ Datos sin agrupar

$M_3 = \sqrt[3]{\frac{\Sigma y_i^3 n_i}{n}}$ $\quad M_3 = \sqrt[3]{\frac{\Sigma X_i^3 f_i}{N}}$ Datos agrupados

Datos sin agrupar

Ejemplo 1. Consideremos nuevamente los datos no agrupados, que fueron utilizados para calcular la Media cuadrática $x_i = 2 \quad 6 \quad 5 \quad 4 \quad 6 \quad 8$.

$$M_3 = \sqrt[3]{\frac{\Sigma x_i^3}{n}} = \sqrt[3]{\frac{2^3+6^3+5^3+4^3+6^3+8^3}{6}} = 5{,}75 \qquad \boxed{M_1 < M_2 < M_3}$$

Veamos: 3,43 < 3,79 < 5,75 por lo tanto se cumple la anterior relación. Si relacionamos estas tres medidas, con la media geométrica y la media armónica, observamos que: $M_{-1} < M_0 < M_1 < M_2 < M_3$ que el lector podrá comprobar desarrollando un ejercicio que involucre el cálculo de estas medidas.

Datos agrupados

Ejemplo 2. Con los siguientes datos, correspondiente a una tabla de frecuencias, se podrá calcular la Media cúbica. Veamos como es el proceso:

y_i	n_i	$y_i\, n_i$	$y_i^2\, n_i$	$y_i^3 n_i$
1	4	4	4	4
2	5	10	20	40
3	7	21	63	189
4	8	32	128	512
6	6	36	216	864
Σ	30	103	431	1.609
X_i	f_i	$X_i\, f_i$	$X_i^2\, f_i$	$X_i^3 f_i$

$$M_3 = \sqrt[3]{\frac{1^3(4)+2^3(5)+.....6^3(6)}{30}} = \sqrt[3]{\frac{1.609}{30}} = 3{,}77$$

$$M_3 = \sqrt[3]{\frac{\Sigma y_i^3 n_i}{n}} \qquad \boxed{M_3 = \sqrt[3]{\frac{\Sigma X_i^3 f_i}{N}}}$$

Ejemplo 3. Consideremos que la media aritmética de tres números es 7, su mediana es 6 y su media geométrica es $\sqrt{216}$. Se pide calcular la media cuadrática y la cúbica.

Solución:

$$\bar{x} = 7 \qquad M_e = 6 \qquad M_0 = \sqrt{216}$$

$$\bar{x} = \frac{x_1 + 6 + x_3}{3} = 7 \Rightarrow 21 = x_1 + 6 + x_3 \Rightarrow 15 = x_1 + x_3$$

$$M_0 = \sqrt[3]{216} = \sqrt[3]{x_1.x_2.x_3} = \sqrt[3]{x_1(6)x_3} \Leftrightarrow 216 = x_1(6)x_3$$

$$\frac{216}{6} = x_1.x_3 \Rightarrow 36 = x_1.x_3 \Rightarrow x_1 = \frac{36}{x_3} \qquad \text{Reemplazando } 15 = \frac{36}{x_3} + x_3 \text{ se tiene que}$$

$$15x_3 = 36 + x_3^2 \Leftrightarrow x_3^2 - 15x_3 + 36 = 0 \qquad \text{Siendo: } x_1 = 3 \quad x_2 = 6 \quad \text{y} \quad x_3 = 12$$

La media cuadrática será: $M_2 = \sqrt{\frac{3^2+6^2+12^2}{3}} = \sqrt{\frac{9+36+144}{3}} \cong 7{,}94$

La media cúbica será: $M_3 = \sqrt[3]{\frac{3^3+6^3+12^3}{3}} = \sqrt[3]{\frac{27+216+1.728}{3}} = 8{,}69$

RELACIÓN ENTRE LAS MEDIDAS DE TENDENCIA CENTRAL O PROMEDIOS.

a) Cuando tratamos la media aritmética, la mediana y la moda, mencionamos algunas relaciones existentes entre ellas

$\bar{x} = M_e = M_d$ En este caso la Distribución se considera *Simétrica*

$\bar{x} > M_e > M_d$ Se dice que la Distribución es *Asimétrica positiva*

$\bar{x} < M_e < M_d$ Es un caso de Distribución *Asimétrica negativa*

Los tres casos anteriores serán nuevamente mencionados en el Capítulo siguiente.

b) Si la distribución es convexa y moderada asimétrica, la *Mediana* se ubica entre la *Media* y el *Modo*, quedando aproximadamente, dos veces más lejos de este último que del primero. $\bar{x} - M_d = 3(\bar{x} - M_e)$ y $M_d = 3M_d - 2\bar{x}$

c) En cualquier distribución, donde los elementos originales difieren en tamaño, las siguientes medidas diferirán en valor, de acuerdo al siguiente orden.

$$M_{-1} < M_0 < M_1 < M_2 < M_3$$

CENTRO RECORRIDO C_R

Esta medida poca utilizada, se define como la media aritmética de los valores extremos que toma la variable.

$$C_r = \frac{x_{máx} + x_{mín}}{2} \qquad C_r = \frac{y_m + y_1}{2} \qquad C_r = \frac{y'_m + y'_0}{2}$$

(Datos sin agrupar) (Variable discreta) (Variable continua)

Ejemplo 1. Supongamos que se tiene la siguiente información, correspondiente a datos sin agrupar:
x_i = 2 6 5 4 6 8 12 7 6 10 Su centro recorrido será:

$$C_r = \frac{x_{máx} + x_{mín}}{2} = \frac{12+2}{2} = 7$$

Ejemplo 2. Su aplicación tanto en la variable discreta como en la continua, se puede observar a continuación.

a)

y_i	n_i
1	4
2	5
3	7
4	8
6	6
Σ	30
X_i	f_i

$$C_r = \frac{y_m + y_1}{2}$$

$$C_r = \frac{6+1}{2} = 3{,}5$$

b) $C_r = \frac{y'_m + y'_0}{2}$

$$C_r = \frac{94+46}{2} = 70$$

En ninguno de los dos casos se tiene en cuenta, el valor que tomen las frecuencias absolutas.

$y'_{i-1} - y'_i$	y_i	n_i
46,1 – 54	50	3
54,1 – 62	58	8
62,1 – 70	66	11
70,1 – 78	74	7
78,1 – 86	82	9
86,1 – 94	90	2
Σ	--	40
$X'_{i-1} - X'_i$	X_i	f_i

CUARTILES, DECILES Y PERCENTILES

Cuando la distribución está constituida por un número grande de intervalos o de marcas de clase, haciéndose necesario calcular un promedio sobre una parte de ella, en estos casos, la distribución puede ser distribuida en cuatro, en diez o en cien partes. En el primer caso nos referiremos a *Cuartiles*, en el segundo a *Deciles* y en el tercero a *Percentiles* o *Centiles.*

En el primero o sea los cuartiles, se podrá obtener: Q_1 como *primer cuartil*, Q_2 *segundo cuartil* y Q_3 correspondiente al *tercer cuartil*.

El *primer cuartil,* es aquel valor de la variable que supera al 25% de las observaciones y a la vez es superado por el 75%.

El *segundo cuartil,* es aquel valor de la variable que supera al 50% de las observaciones y a la vez es superado por el 50%. Es decir, corresponde a la Mediana

En los *deciles, el séptimo decil,* es aquel valor de la variable que supera al 70% de las observaciones y a la vez es superado por el 30%.

Y en los percentiles, el *percentil cuarenta*, es aquel valor de la variable que supera al 40% de las observaciones y a la vez es superado por el 60%.

Tal como lo mencionamos, el segundo cuartil, el quinto decil y el percentil cincuenta, en una misma distribución presenta el mismo resultado, siendo en los tres casos igual al de la mediana.

Datos no agrupados

Ejemplo 1. Con los siguientes datos: 16 10 4 8 12 10 8 20 4 13 12 22 16 26 20, se pide calcular a) Primer y tercer cuartil. b) Cuarto y sexto decil y c) el 30 y 90 percentil.

a) Lo primero que se hace, es ordenar los datos de menor a mayor o viceversa.

4 4 8 8 10 10 12 12 13 16 16 20 20 22 26 $\quad n = 15$

a) Para el *primer cuartil*, aplicamos el siguiente procedimiento, muy parecido al calcular la mediana, es decir: $\frac{1(n+1)}{4} = \frac{1(15+1)}{4} = 4^a$ *posición* $\quad Q_1 = x_j = 8$

Para el *tercer cuartil*: $\frac{3(n+1)}{4} = \frac{3(15+1)}{4} = 12^a$ *posición* $\quad Q_3 = x_j = 20$

b) *Cuarto decil* : $\frac{4(n+1)}{10} = \frac{4(15+1)}{10} = 6{,}4^a$ *posición.* Observemos que el proceso correspondiente al resultado obtenido es diferente a los dos anteriores, pues aparece con la posición de 6,4, es decir, que debe estar entre la sexta y la séptima posición, siendo el valor de 10 la sexta posición y la diferencia entre la 7ª y 6ª es 12-10=2, este resultado se multiplica por la fracción 0,4, así 2(0,4)=0,8 y se lo agregamos al valor anterior de 10, siendo: $D_4 = x_j = 10{,}8$

El *sexto decil*, será: $\frac{6(n+1)}{10} = \frac{6(15+1)}{10} = 9{,}6^a$ *posición*. El valor correspondiente a la novena posición es 13 y se tiene además, 0,60(16-13) = 0,6(3) = 1,8. Por lo tanto $D_6 = 14{,}8$

c) *Percentil 30* será: $\frac{30(n+1)}{100} = \frac{30(15+1)}{100} = 4{,}8^a$ *posición*. El valor de la cuarta posición es 8, además se tiene que, 0,8(10 – 8) = 1,6. Por lo tanto $P_{30} = 9{,}6$

El *Percentil 90* será: $\frac{90(n+1)}{100} = \frac{90(15+1)}{100} = 14{,}4^{a}$ *posición*.El valor en la posición 14 es de 22, además, 0,4(26 – 22) = 1,6, se tendrá que P_{90}= 23,6

Datos agrupados

El cálculo de los *cuartiles, deciles* y *percentiles* es bastante parecido al procedimiento que se había utilizado para calcular la *Mediana*, salvo una pequeña variación, al establecer la posición donde se debe ubicar la respectiva medida.

Los pasos a seguir en los cálculos son:
- Se determinan las *frecuencias absolutas acumuladas.*
- Determinamos la *posición* donde se localiza.
- Observamos si el valor obtenido en el paso anterior, aparece o no en dicha columna, correspondiente a las frecuencias absolutas acumuladas.
- Se aplica la fórmula respectiva, dependiendo de la variable trabajada.

$y'_{i-1} - y'_i$	n_i	N_i	y_i
2,1 – 6	3	3	4
6,1 – 10	12	15	8
10,1 – 14	10	25	12
14,1 – 18	18	43	16
18,1 – 22	10	53	20
22,1 – 26	6	59	24
26,1 – 30	20	79	28
30,1 – 34	8	87	32
34,1 – 38	3	90	36
Σ	90	--	-----
$X'_{i-1} - X'_i$	f_i		X_i

Se pide calcular:

a) *Cuartil tres*. La posición se da así: $\frac{3(n)}{4} = \frac{3(90)}{4} = 67{,}5^{a}$ *posición*. Observemos que este valor no aparece en la columna de las frecuencias absolutas acumuladas, por tal razón, seleccionamos el valor inmediatamente anterior (59) como N_{j-1} y al valor inmediatamente superior teniendo a 79 como N_j. Ahora veamos las fórmulas a aplicar dependiendo de la variable.

Variable Discreta $\rightarrow Q_3 = y_j = 28$. Como se observa, se toma el valor de y_j al frente de N_j si consideramos a las marcas de clase como *variable discreta*.

Variable continua. En este caso nos ubicamos al frente de los intervalos y aplicamos la fórmula parecida a la utilizada al calcular la *Mediana,* con una ligera modificación:

$$Q_3 = y'_{j-1} + c\left[\frac{3n/4 - N_{j-1}}{n_j}\right] = 26 + 4\left[\frac{67{,}5 - 59}{20}\right] = 26 + 1{,}7 = 27{,}7$$

Observemos que el valor de 26 fue seleccionado en el intervalo correspondiente a N_j

b) *Decil seis* Determinamos la posición así: $\frac{6(n)}{10} = \frac{6(90)}{10} = 54^{a}$ Cómo este valor no aparece en la columna de las frecuencias absolutas acumuladas, se toma a N_{j-1}=53 y a N_j =59.

Ahora observemos el procedimiento a seguir en ambas variables.

Variable Discreta $\rightarrow D_6 = y_j = 24$ *(Marcas de clase).*

Variable Continua $\rightarrow D_6 = 22 + 4\left[\frac{54-53}{6}\right] = 22 + 0{,}66 = 22{,}66$

c) *Percentilil* 80. La posición es: $\frac{80(n)}{100} = \frac{80(90)}{100} = 72^{a}$*posición.* Al igual que en los dos casos anteriores, se deberán aplicar los mismos procedimientos para esta tabla. N_{j-1}= 59 y para N_j= 79

$$\textit{Variable Discreta} \rightarrow P_{80} = y_j = 28 \ (\textit{Marcas de clase}).$$

$$\textit{Variable Continua} \rightarrow P_{80} = 26 + 4\left[\frac{72-59}{20}\right] = 26 + 2{,}6 = 28{,}6$$

EJERCICIOS MISCELÁNEOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **SIL.**

68. Una compañía de aviación que ofrece un vuelo diario a una determinada región del país no cumplió con el horario de llegada en los últimos 10 días de abril, con los siguientes minutos de retardo o de anticipación (número negativo):

-3	6	4	10	-4	124	2	-1	4	1

a) Si la compañía contratara un especialista en estadística para mostrar el cumplimiento, ¿cuáles serían las medidas que utilizaría?
b) Si el objetivo fuese mostrar que ofrece un buen servicio (cumplimiento), ¿qué medida utilizaría?
Si el objetivo es mostrar un mal servicio de una compañía con la que él desea competir ¿qué medida utilizaría para lograrlo?

69. Explique brevemente:
a) Tres ventajas de la media
b) Dos desventajas de la mediana
c) Dos ventajas de la moda
d) ¿En qué condiciones serán iguales la media, mediana y moda?

70. Con los siguientes datos, se pide calcular:

$y'_{i-1} - y'_i$	n_i
6,1 — 12	2
12,1 — 15	14
15,1 — 20	5
20,1 — 28	3
28,1 — 36	7
36,1 — 40	16
40,1 — 50	3
Σ	50

a) Media aritmética
b) Mediana
c) Moda
d) Tercer decil
e) Segundo cuartil
f) Percentil sesenta

71. Un grupo de 400 empleados, que tiene una compañía, se dividen en operarios y técnicos con un salario promedio de $1.260.960. Los salarios promedio para cada uno de los grupos son de $857.300 y $1.320.856 respectivamente.
a) ¿Cuántos operarios y cuántos técnicos tiene la compañía?
b) Si el gerente establece una bonificación de $30.000 para los operarios y del 8% para los técnicos, ¿cuál será el salario promedio para los 400 empleados de la compañía?

72. ¿Qué es un promedio?. ¿cuáles podría mencionar?; ¿qué condiciones debe tener el promedio para cumplir su cometido, mencione tres?

73. Con los datos del ejercicio N° 45, se pide calcular: la media cuadrática, la cúbica y el séptimo decil.

74. Con los siguientes datos:

$x_1 = 4$	$x_2 = 10$	$x_3 = 6$	$x_4 = 4$	$x_5 = 4$	$x_6 = 12$

a) Comprobar que $M_{-1} < M_O < M_1 < M_2 < M_3$

b) Suponga que al calcular la mediana, media aritmética y moda, estos resultados provienen de una distribución de frecuencias. ¿Cuál sería su relación, es decir, simétrica o asimétrica?

75. Suponga que los salarios pagados a los empleados de una compañía, son como se representa a continuación:

CARGOS	NÚMERO	SALARIO MENSUAL
Directores	2	1.930.000
Supervisores	4	1.510.000
Economistas	6	1.370.000
Contadores	4	1.350.000
Auxiliares	26	646.000
Obreros	110	590.000

Se pide:

a) Calcular la media, mediana y moda.

b) En su opinión. ¿Cuál de estos promedios considera más representativo?

76. En una fábrica de tres secciones se sabe que en la sección A, con 120 obreros la asistencia promedio es de 240 días al año; en la sección B, que tiene 180 operarios, la asistencia media es de 216 días al año. Si la asistencia media en toda la fábrica es de 226,70 días. ¿cuántos obreros hay en la sección C, donde la asistencia promedio es de 230 días al año?

77. Durante diez días se observó en un almacén de autos, el número de estos vendidos por su empleado estrella.

DÍA:	1	2	3	4	5	6	7	8	9	10
AUTOS:	3	4	2	1	3	2	4	6	5	4

Cada auto vale $18.500.000 y al empleado se le paga un sueldo de $270.000 básico mensual más el 0,5% del valor de cada auto vendido. ¿Cuál será el sueldo promedio para el vendedor estrella en este lapso de 10 días?

78. Se sabe que el salario mensual promedio de los empleados de una empresa es de $1.052.000. El total de los empleados se aumenta en 20% con relación a los que había hasta la fecha. Los nuevos empleados tienen salario mensual promedio igual al 36,8% de salario promedio de los antiguos. Un año después se hace un aumento de $57.600 a todos los empleados de la empresa. ¿Cuál será el salario promedio de la empresa actualmente?

79. La siguiente es la población de 3 zonas de un Departamento en que se encuentra dividido el país según los censos de 2001 y de 2011

ZONA	2001	2011
1	322.867	240.776
2	204.903	267.358
3	345.604	575.864

Calcule la población total de las tres zonas del Departamento para una fecha equidistante entre los dos censos, suponiendo que:

a) El porcentaje (tasa) de aumento poblacional es constante en todo el Departamento.
b) El porcentaje (tasa) de aumento es constante en cada zona del Departamento.

80.

$y'_{i-1} - y'_i$	n_i
800 – 1.000	5
1.000 – 1.200	13
1.200 – 1.400	17
1.400 – 1.600	8
1.600 – 1.800	7
$X'_{i-1} - X'_i$	f_i

Con la siguiente distribución, calcular:

a) La media, mediana y modo.
b) La media cúbica, cuadrática, armónica y geométrica.

81. Contestar verdadero si el enunciado es verdadero. En caso contrario, la palabra subrayada debe sustituirse por una expresión con la cual el enunciado sí sea válido.
   a. La media de una muestra divide los datos en dos mitades iguales: la mitad mayores y la mitad menores que su propio valor.
   b. La media armónica es la raíz enésima del producto de los valores que toma la variable.
   c. El decil ocho representa aquel valor de la variable que supera el 80% de las observaciones y es superado por el resto 20%.
   d. Una distribución o un conjunto de datos, permite el cálculo de varias medidas de tendencia central.
   e. Un promedio aplicado o calculado en un conjunto de datos provenientes de una muestra, se denomina parámetro.

82. Explique brevemente los siguientes puntos:
   a. Cuatro condiciones para que los promedios puedan cumplir su cometido.
   b. Cuatro reglas que deben ser tenidas en cuenta para el uso de los promedios.

83. Se sabe que la media aritmética de dos números es igual a 5 y la media geométrica de los mismos es igual a 4. ¿cuál es la media cúbica?

84. En una empresa constructora de vivienda, los jornales semanales tienen una media de $169.000. Como una solución al conflicto laboral surgido, se proponen dos soluciones al conflicto: (a) un aumento del 6% en el salario semanal; (b) un aumento del 4%, más una bonificación semanal de $5.800 a cada obrero. ¿cuál de las dos alternativas mejora la situación del obrero?

85. Dos empresas tienen ocupados 600 obreros, distribuidos así, el 30% en A y el resto en la empresa B. Se sabe que el promedio de salario en esta última es de $860.000 y en A es del 20% menos que de B. ¿Cuál es el promedio de salarios para el total de obreros?

86. Se tienen dos cursos A y B, donde el primero tiene un promedio de calificación de 3,4, mientras que en el B es de 4,2. Diga qué porcentaje de alumnos tiene A y B, si el promedio de calificación para la suma de los dos cursos es de 3,7.

87. Se tiene que

$y_2 = 22$	$y_5' = 50$	$m = 6$	C = constante	$n_2 + n_4 = 15$	
$n_1 + n_3 = 25$	$n = 50$	$n_5 + n_6 = 10$	$n_1 = 8$	$n_5 = 2$	$n_2 = 7$

Se pide calcular la media armónica y cúbica.

88. Decir si son ciertos o falsos y por qué:

a) $\bar{x} = 68$	$M_e = 62$	$M_d = 68$			
b) $\bar{x} = 70$	$M_o = 65$	$M_2 = 72{,}50$	$M_{-1} = 63$	$M_3 = 76{,}8$	

89. Dados los números 8, 12, 16, 20, encontrar la media armónica, la media geométrica, la media cúbica y la media cuadrática.

90. Contestar ***verdadero*** si el enunciado siempre es verídico. Si no lo es, la palabra en negrilla deberá sustituirse por una expresión con la cual el enunciado si sea válido.

a) A un grupo de valores se le puede calcular **más de un** promedio.

b) Un grupo de valores puede tener **más de una media** aritmética.

c) La **media cúbica** es la raíz enésima del producto de los valores de la variable.

d) La suma de las desviaciones con respecto al **origen de trabajo** es igual a cero.

e) En una serie sencilla (datos sin agrupar) cuando el valor de n **es impar**, la mediana es igual al valor central.

f) La **media** divide a los datos de una distribución en dos mitades: la mitad mayores y la mitad menores que su propio valor.

g) La **moda** es el valor de la variable que más se repite.

91. Un industrial tiene dos fábricas, la primera con 200 trabajadores y la segunda con 300, además se sabe que en promedio el salario mensual es de $1.026.328 y que los trabajadores de la segunda fábrica ganan $86.400 menos que los de la primera fábrica. ¿Cuál es el salario promedio en cada fábrica?

92. Con los siguientes datos, $n = 200$, con frecuencias relativas y acumuladas diferentes, ya que se trata de dos problemas (A y B), se pide:

	(A)	(B)
$y_{i-1}' - y_i'$	h_i	N_i
15,1 – 25	0,04	6
25,1 – 35	0,10	23
35,1 – 45	0,21	57
45,1 – 55	0,30	110
55,1 – 65	0,21	152
65,1 – 75	0,10	190
75,1 – 85	0 04	200
Σ	1,00	–
$X_{i-1}' - X_i'$	f_i/n_i	F_i

Primero trabajar con la columna de las frecuencias relativas (A) y luego con las frecuencias absolutas acumuladas. (B)

Nota: Son dos problemas diferentes, donde las frecuencias relativas no corresponden a las frecuencias absolutas *acumuladas*.

a) Calcular la media aritmética utilizando las fórmulas ya conocidas.
b) Calcular la mediana (variable continua). Obtener las marcas de clase y calcular la mediana.
c) Calcular el modo (aplicando la fórmula general y las establecidas para variables continuas).

93. Con los datos del punto anterior se pide calcular:
a) La media geométrica
b) El modo mediante la aplicación de la siguiente fórmula: $M_d = 3M_e - 2M_1$
c) Las medias: armónica, cuadrática y cúbica.

94. De tres números se sabe que su media cúbica es $\sqrt[3]{657}$ su media aritmética es 7 y su mediana es 6. Se pide calcular los valores de esos tres números.

95. Se sabe que la media aritmética de dos números es igual a 5 y la media geométrica de los mismos es igual a 4. ¿Cuál es su media armónica?

96. Con los siguientes datos sin agrupar, calcular la media cuadrática.

x_i :	8	2	6	10	2	6	8	10	2	6

97. Hallar la media cuadrática de la siguiente distribución

$y'_{i-1} - y'_i$	2,1 – 6	6,1 – 10	10,1 – 14	14,1 – 18	18,1 – 22	Σ	$X'_{i-1} - X'_i$
n_i	3	7	10	6	4	30	f_i

98. Con los siguientes datos, calcular el primer cuartil, tercer cuartil, sexto decil y el percentil 80

a) *Variable discreta*

y_i	n_i
2	3
4	6
6	15
8	8
10	2
12	2
14	10
Σ	50
X_i	f_i

b) *Variable continua*

$y'_{i-1} - y'$	n_i
3,1 — 8	14
8,1 — 13	15
13,1 — 18	8
18,1 — 23	6
23,1 — 28	7
28,1 — 33	10
Σ	60
$X'_{i-1} - X'_i$	f_i

Síntesis de capítulo

La posición, en una distribución de frecuencias, lo da el valor de la variable denominada promedio, alrededor de la cual se agrupan las frecuencias.

A los promedios se les conoce como medidas de posición, valor promedio o valor central.

Para cada caso en particular se deben aplicar algunas de las medidas de posición o de tendencia central, de tal manera que el valor represente al conjunto de datos, los cuales tendrán algunas ventajas y desventajas en su aplicación.

Hay que tener en cuenta que estas medidas o promedios, proporcionan información acerca de un conjunto, siendo una forma de describir el comportamiento de esos datos, pero no proporcionan ninguna idea del grado de variabilidad, es decir, su verdadera o aproximada representación de ese conjunto observado.

Se ha dicho en repetidas ocasiones, que los promedios aplicados a datos correspondientes a una muestra se les denomina estimadores y aplicados a datos de una población, se les conoce como parámetros.

Al tener varias clases de promedios: media, mediana, modo, etc., la cifra resultante al aplicar una de esas medidas puede traer confusión, ya que algunas de ellas se ajustan más a determinadas situaciones que otras. Hoy por hoy, la gente utiliza más la media aritmética, algunos pocos la mediana y la moda, sin tener el cuidado de seleccionar el promedio que mejor represente el resumen de los datos mediante una cifra que no distorsione.

Estos promedios utilizados sólo pueden ocultar información, cuando no representan correctamente el conjunto de datos. En estos casos se hace necesaria la aplicación de otras medidas como las de dispersión, que ayudan a una mejor interpretación de ese valor central y que se verán en el capítulo siguiente.

Reglas para el uso de los promedios

- ❑ Cuando la serie tenga forma de progresión geométrica, debe usarse el promedio geométrico.
- ❑ Para calcular la velocidad media debe usarse la media armónica.
- ❑ Cuando la distribución sea muy asimétrica, debe considerarse la posibilidad de usar le mediana o el modo.
- ❑ Cuando la distribución tenga forma de U o sea que pueda representarse por una cóncava de extremos iguales, debe usarse el modo.
- ❑ Cuando quiera dársele importancia a valores pequeños de la variable, es aconsejable la media geométrica.
- ❑ En una distribución cuyos valores extremos no están definidos, es aconsejable la mediana o el modo.
- ❑ Cuando haya alguna razón para pensar que el promedio aritmético no representa muy bien a la distribución, debido a que valores extremos lo afectan, o por otras razones, debe considerarse la posibilidad de usar la mediana o el modo.
- ❑ Cuando la amplitud de la distribución no es constante, no debe usarse el modo.
- ❑ Cuando se requiere promediar relaciones (tasas), se debe usar la media armónica.
- ❑ En los demás casos debe usarse la media aritmética.

ANEXOS A LA UNIDAD 3

Se explicará brevemente el uso de la calculadora, así como la aplicación del programa EXCEL, con lo cual el lector podrá agilizar los cálculos correspondientes, tal como se verá a continuación.

APLICACIONES CON LA CALCULADORA

- **Calculadora CASIO** f_x – 5000*F*
  - ***Primero.*** Se oprimen las teclas [MODE] [X], con lo cual debe aparecer en la parte superior de la pantalla, el símbolo [SD]
  - ***Segundo.*** Se hace necesario, antes de iniciar los cálculos, borrar la información que hubiese con anterioridad, mediante la ejecución [SHIFT] [DEL] [EXE].
  - ***Tercero.*** Se comienza con la entrada o captura de la información (datos).

Ejemplo 1. Consideremos los siguientes datos sin agrupar: x_i: 8 10 20 3 7 Iniciamos digitando el [8] y luego la tecla $\frac{\sqrt{}}{\text{DT}}$ identificada por DT, correspondiente a DATA, a continuación se repite el proceso, pero ahora con 10, luego con 20 y así sucesivamente.

Ejemplo 2. Procedamos ahora a la entrada de los datos correspondientes a una tabla de frecuencias o datos agrupados, siendo $n = 40$

x_i: 2 4 6 8 10

n_i: 3 6 10 12 9

La entrada de estos datos se hará de la siguiente forma: [2] [SHIFT] [;] [3] $\frac{\sqrt{}}{\text{DT}}$ Luego se hará [4] [SHIFT] [;] [6] [DATA] y así sucesivamente.

Ya digitados los datos, la obtención de los resultados al aplicar las fórmulas respectivas, tanto para el primero como en el segundo ejemplo, se hará de la siguiente forma:

[SHIFT] [PROG]

- [1] → Σx_i^2 ó $\Sigma y_i^2 n_i$
- [2] → Σx_i ó $\Sigma y_i n_i$
- [3] → n

Ahora con [SHIFT]

- [1] → $\bar{x}$ ó $\bar{y}$
- [2] → $_x\sigma_n$
- [3] → $_x\sigma_{n-1}$

- **Calculadora CASIO** f_x – 30*MS* **G**

❑ ***Primero.*** Al oprimir la tecla [MODE] aparecen en pantalla tres operadores [COMP] [SD] [REG]
[1] [2] [3]
oprimimos en el [2] para obtener [SD] con el cual iniciaremos los cálculos. Vale la pena indicar que se oprime el [1] para salir, una vez realizadas todas las operaciones.

❑ ***Segundo***. Borramos toda la información que se tenga con [SHIFT] [MODE] (CLR) con el cual aparecen otros operadores, seleccionando el [1] correspondiente [SCL].

❑ ***Tercero***. La entrada o captura de todos los datos, se hace igual que con la anterior calculadora.

❑ ***Cuarto***. La salida u obtención de los resultados se hace de la siguiente forma:

Con SHIFT 1 se tienen tres opciones:
- 1 → Σx_i^2 ó $\Sigma y_i^2 n_i$
- 2 → Σx_i ó $\Sigma y_i n_i$
- 3 → n

Con SHIFT 2 hay tres opciones:
- 1 → $\bar{x}$ ó $\bar{y}$
- 2 → $_x\sigma_n$
- 3 → $_x\sigma_{n-1}$

■ **Calculadora CASIO** f_x – 570*MS*

❑ ***Primero.*** Al oprimir la tecla MODE aparecen en pantalla dos opciones 1 COMP 2 CMPLY

Oprimimos nuevamente MODE y nuevamente se nos dan otras tres opciones SD REG BASE

1 2 3

Procedemos a seleccionar el 1 correspondiente a SD

❑ ***Segundo***. Procedemos a capturar información, tal como se realizó en la primera calculadora. Datos sin agrupar: 8 DT luego 10 DT y así sucesivamente.

En datos agrupados, se procede así: 2 ; 3 luego 4 ; 6 y así sucesivamente.

❑ ***Tercero***. Los resultados aparecen, mediante el siguiente proceso:

Con SHIFT 1 se seleccionan así: 1 Σx_i^2 2 Σx_i 3 n

Con SHIFT 2 se seleccionan así: 1 $\bar{x}$ 2 $_x\sigma_n$ 3 $_x\sigma_{n-1}$

■ **Calculadora CASIO** f_x – 350*TL*

Con MODE nos aparecen 1 COMP 2 SD 3 REG, seleccionamos la opción 2 y procedemos a capturar o digitar la información, exactamente igual como se hizo en los procesos anteriores. La salida de los resultados se hará así:

Con SHIFT 1 se obtiene $\bar{x}$ → Media aritmética

Con SHIFT 2 se obtiene $_x\sigma_n$ → Desviación típica

Con SHIFT 3 se obtiene $_x\sigma_{n-1}$ → Desviación típica corregida

Ahora con la tecla u opción RCL y las letras A B C se obtienen los resultados para RCL A Σx_i^2 RCL B Σx_i y RCL C n respectivamente.

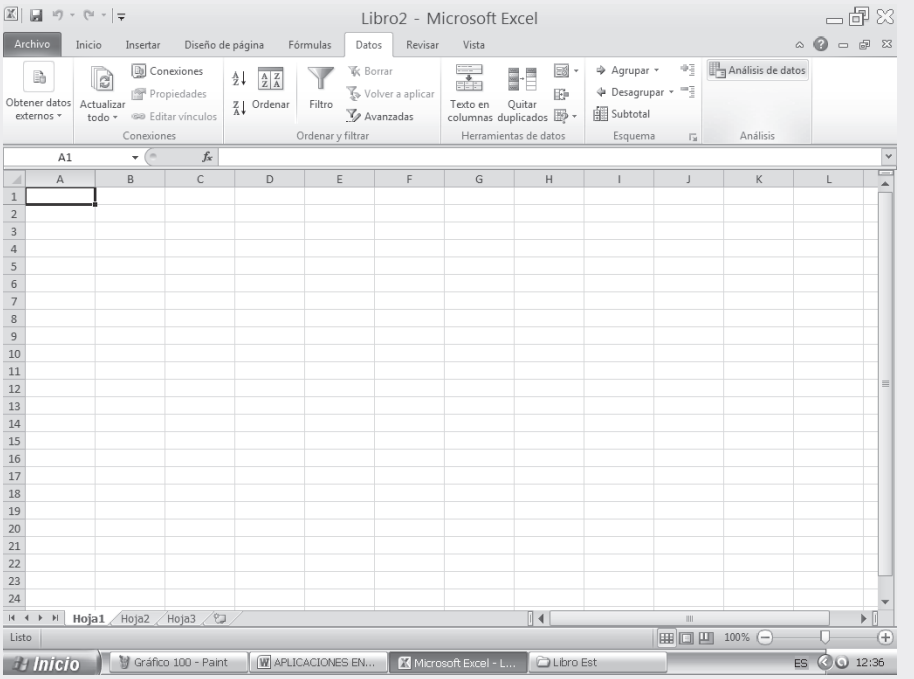
TRABAJO EN EXCEL PARA EL CÁLCULO DE LA MEDIA

Complemento virtual SIL

Para el cálculo de la media se realiza el siguiente procedimiento:

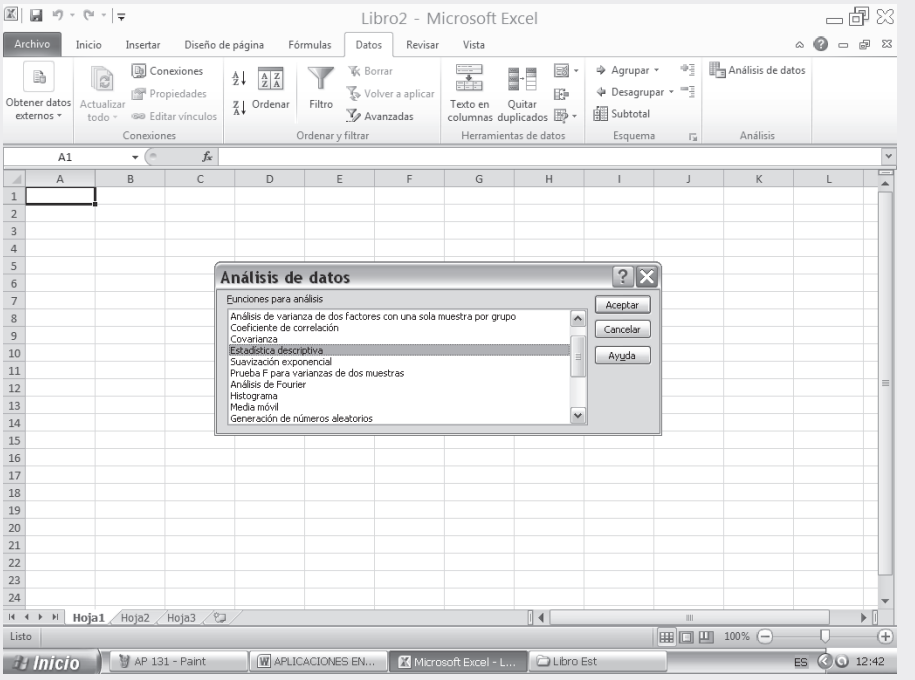
❑ Hacemos CLIC en ANÁLISIS DE DATOS. Para tener acceso a estas herramientas, hacemos clic en ANÁLISIS DE DATOS en el grupo ANÁLISIS de la ficha DATOS (Ver aplicativos en Excel del Capítulo 2)

Figura No. 1. Microsoft Excel



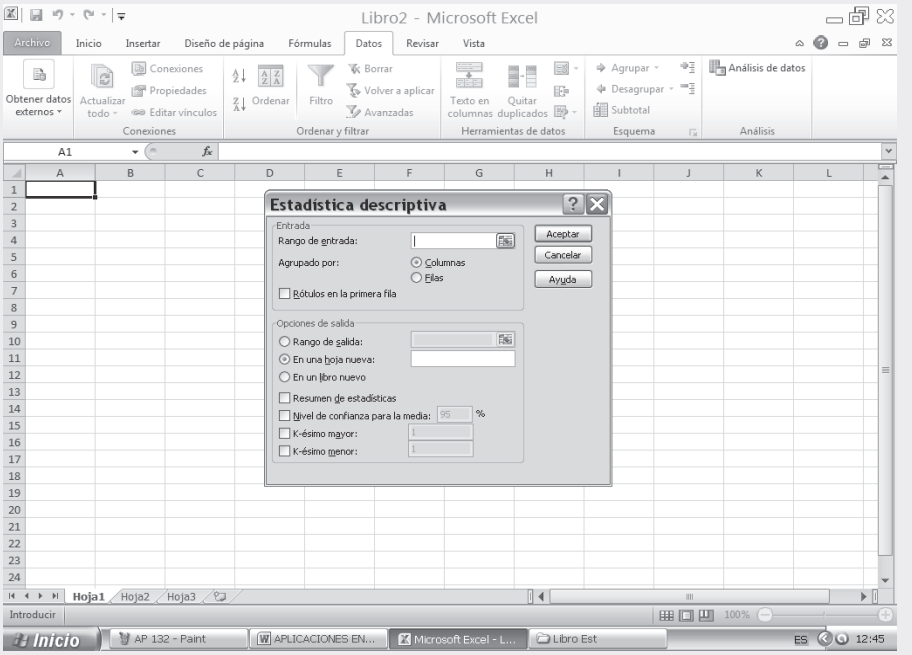
- Hacemos CLIC en el submenú ANÁLISIS DE DATOS. Nos aparece un cuadro de diálogo, en el cual seleccionamos una de las funciones, en nuestro caso la opción identificada como ESTADÍSTICA DESCRIPTIVA.

Figura No. 2. Funciones para análisis



- Luego al hacer CLIC en ACEPTAR, deberá aparecer la figura siguiente:

Figura No. 3. Estadística Descriptiva



- Tabla No. 1- Los datos que se presentan a continuación se encuentran en la segunda parte del SIL en internet. La tabla siguiente sólo muestra una parte de los datos que contiene el SIL, por tanto es conveniente para el desarrollo del siguiente ejercicio tomar la totalidad de los datos, es decir, los 1100 datos que aparecen allí.

Tabla No. 1. Población Total de una Institución Educativa

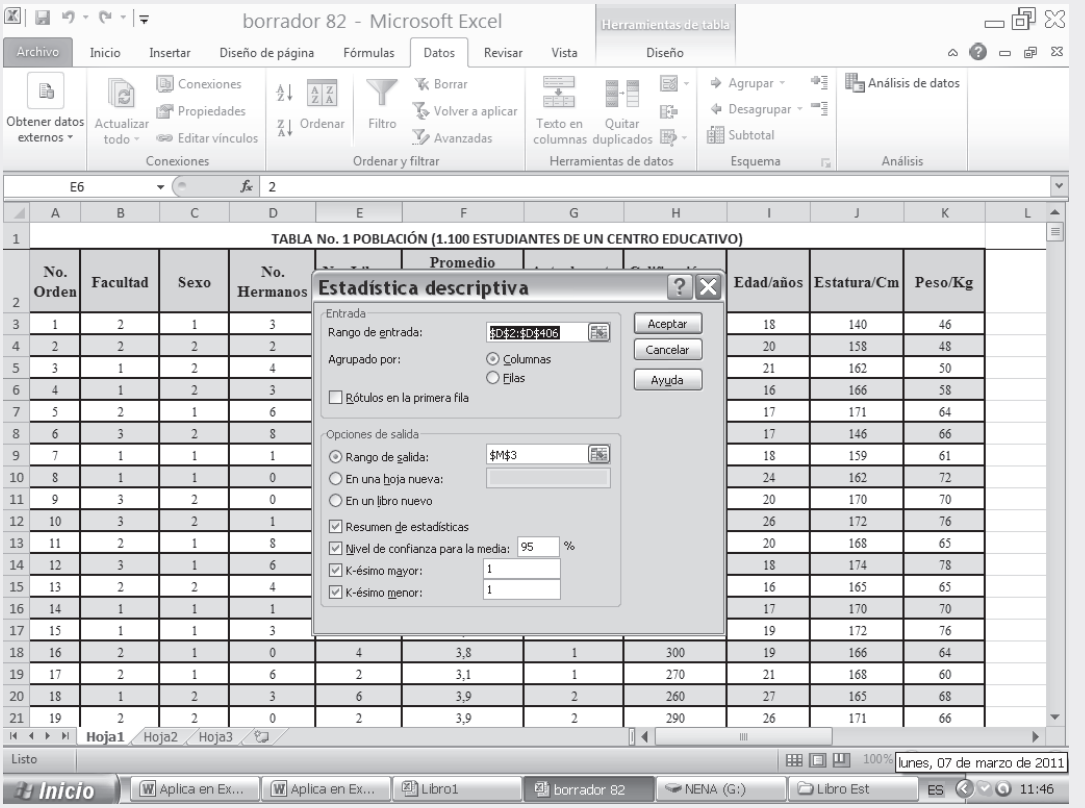
No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciones ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
11	2	1	8	5	3,6	1	320	20	168	65
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
75	3	2	3	3	3,1	1	280	19	154	50
147	3	1	1	8	5	1	310	17	174	83
148	2	2	4	2	4,2	2	310	16	182	70
149	1	1	6	3	2,6	1	320	20	178	88
150	3	1	2	4	2,8	1	400	24	165	70
151	1	2	0	6	3,4	2	380	20	165	58
152	1	2	0	3	2,8	1	310	20	171	59
153	1	2	3	5	4,2	2	350	22	169	64
154	2	2	4	8	4,1	2	400	23	172	68
155	2	1	2	2	4	2	400	24	163	72
156	3	2	6	2	4	2	380	20	165	58
157	1	1	5	2	3,6	2	420	26	178	88
158	1	1	6	2	3,2	1	300	20	182	84
159	3	2	2	6	3,8	2	315	19	180	70
160	1	1	4	4	4,2	1	310	18	174	81
276	1	1	0	2	4	2	350	19	162	69
277	1	1	0	6	3,8	2	320	20	168	68

276	1	1	0	2	4	2	350	19	162	69
277	1	1	0	6	3,8	2	320	20	168	68
376	2	1	0	4	3,8	1	300	19	166	64
377	2	1	6	2	3,1	1	270	21	168	60
378	1	2	3	6	3,9	2	260	27	165	68
379	2	2	0	2	3,9	2	290	26	171	66
380	3	1	0	3	4,1	1	380	30	169	72
381	2	2	8	7	3,9	2	410	24	168	70
382	3	1	0	10	4,4	2	420	25	180	84
383	1	1	1	3	4,2	1	390	22	174	80
384	2	2	6	6	4	2	370	23	166	68
385	3	1	0	11	3,9	2	290	24	162	60
386	3	2	0	8	3,9	2	300	19	151	66
387	2	2	2	2	4	1	330	18	158	60
388	1	2	3	2	3,3	2	310	17	159	58
399	1	1	2	2	4,5	1	300	28	150	56
400	3	2	0	6	3,6	2	280	17	148	46
408	1	1	0	3	3,5	2	270	17	163	65
440	1	1	4	3	2,7	2	320	15	168	76
441	3	1	0	3	3	2	290	17	171	82
442	3	1	0	6	3	2	410	18	174	86
443	1	1	4	2	3,6	1	400	19	176	84
444	2	2	7	4	3,8	1	380	20	180	70
445	3	2	2	2	4,5	2	360	17	178	69
814	2	2	3	10	2	1	280	18	165	60
815	2	2	2	8	4,3	2	270	18	160	61
816	1	1	0	2	4	2	350	19	162	69
817	1	1	0	6	3,8	2	320	20	168	68
818	1	2	2	2	3,6	1	320	17	174	68
819	3	1	3	2	4,6	2	410	16	176	69
820	2	2	1	6	1,8	1	240	17	172	69
821	3	1	3	5	3,1	2	280	17	169	76
822	3	1	8	5	1,9	2	250	16	164	70
828	2	2	0	6	4	2	370	17	174	69
869	1	1	6	3	2,6	1	320	20	178	88
870	3	1	2	4	2,8	1	400	24	165	70
871	1	2	0	6	3,4	2	380	20	165	58
997	1	1	0	6	3,8	2	320	20	168	68
998	1	2	2	2	3,6	1	320	17	174	68
999	3	1	3	2	4,6	2	410	16	176	69
1000	2	2	1	6	1,8	1	240	17	172	69
1068	3	2	2	2	3,4	2	300	26	172	70
1069	2	1	6	6	3,6	1	310	21	164	68
1075	1	2	3	2	2,6	1	270	15	165	60
1082	3	2	3	8	3,6	1	320	32	165	64
1083	1	1	6	10	2,9	2	370	30	162	70
1084	1	2	0	8	3,4	2	400	28	163	58
1085	1	2	3	6	2,6	1	270	15	168	56
1096	1	2	2	2	3,6	1	320	17	174	68
1097	3	1	3	2	4,6	2	410	16	176	69

El cuadro de diálogo que se presenta en la figura No. 3. ESTADÍSTICA DESCRIPTIVA será tecleado con la siguiente información:

- La casilla RANGO DE ENTRADA, corresponde a la variable NÚMERO DE HERMANOS, es decir, seleccionamos una muestra de rango D2:D406. Debemos tener en cuenta que esta selección incluye el nombre de la fila.
- Nuestra información esta agrupada por Columnas, en este caso activamos el campo correspondiente.

Figura No. 4. Estadística Descriptiva (Pantallazo Adicional)



- Hacemos CLIC en el campo RÓTULOS EN LA PRIMERA FILA, con el fin de que los resultados salgan clasificados.
- En la categoría OPCIONES DE SALIDA, podemos establecer dónde queremos que nos muestre el cuadro de resultados. En nuestro ejemplo, activamos la opción RANGO DE SALIDA, y seleccionamos la celda M3.
- Activamos las opciones RESUMEN DE ESTADÍSTICAS; NIVEL DE CONFIANZA PARA LA MEDIA, que para nuestro ejemplo, lo consideramos en un 95%, así que tecleamos este valor.
- También seleccionamos las opciones de K-ÉSIMO MAYOR Y K-ÉSIMO MENOR, haciendo CLIC en estos campos. En nuestro caso, tecleamos el número 1 en cada campo, esto es, en caso de que se presente un número mayor de resultados para el análisis.
- Finalmente hacemos CLIC en ACEPTAR. Al hacer CLIC en ACEPTAR, se obtiene la información, tal como puede observarse en la figura No. 4, que será profundizada mediante la aplicación de las fórmulas correspondientes en los capítulos 3 y 4 de este libro.
- Los resultados de la figura No. 5, nos muestran un cuadro resumen con los valores de la Media, Error Típico, Mediana, Asimetría, Mínimo, Máximo, Suma, así como el Conteo para la variable NÚMERO DE HERMANOS.

Figura No. 5 Cuadro de Resultados

No. Hermanos	
Media	2,68069307
Error típico	0,12714053
Mediana	2
Moda	0
Desviación estándar	2,55549299
Varianza de la muestra	6,53054443
Curtosis	0,58527109
Coeficiente de asimetría	1,04257772
Rango	12
Mínimo	0
Máximo	12
Suma	1083
Cuenta	404
Mayor (1)	12
Menor(1)	0
Nivel de confianza(95,0%)	0,24994148

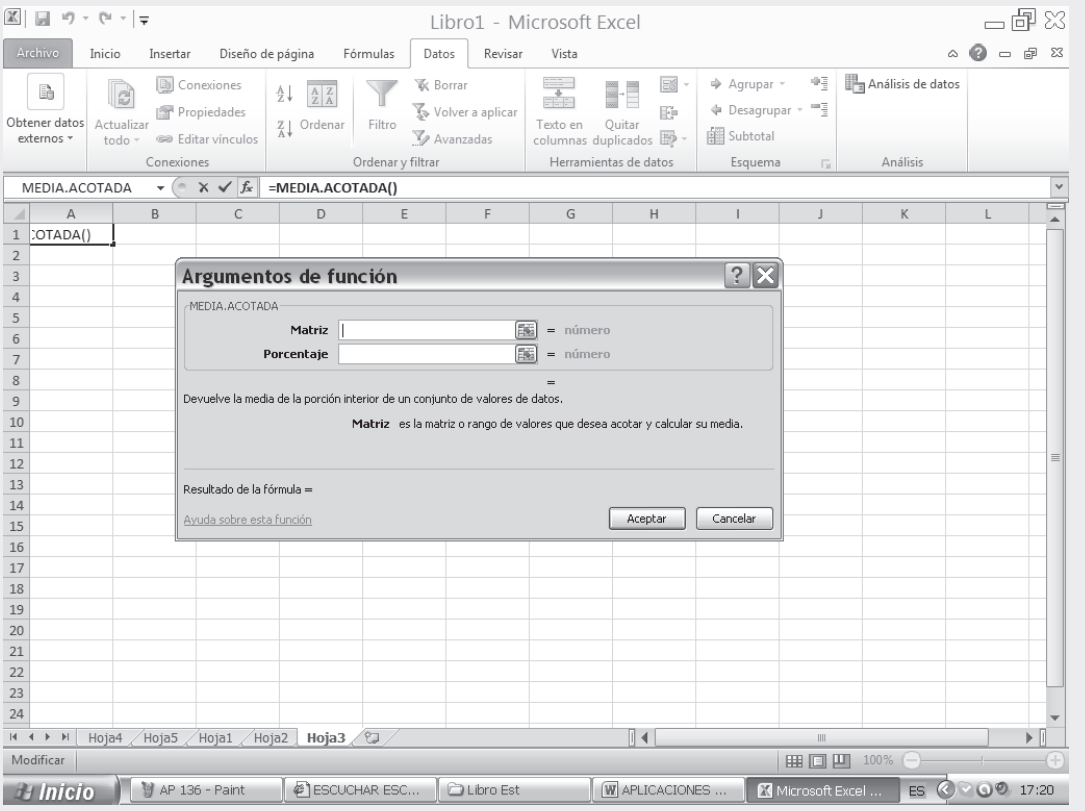
- Nota: Para lograr los resultados en todas y cada una de las opciones (Resumen de estadísticas; nivel de confianza para la media, K-ésimo mayor y K-ésimo menor), deben activarse.

Se podrá calcular en forma independiente, cada una de las medidas denominadas de TENDENCIA CENTRAL y de DISPERSIÓN, para ello sólo se va a tomar una de las variables, en este caso será el NÚMERO DE HERMANOS, de acuerdo con la información de la Tabla No. 1. El cálculo de estas medidas se presenta a continuación:

- Seleccionamos la CELDA, es decir, la activamos en el lugar donde deseamos que aparezca el resultado.

- Con el MOUSE nos movemos al ÍCONO fx, y hacemos CLIC dos veces. De este modo aparecerá la figura correspondiente a INSERTAR FUNCIÓN.

Figura No. 5. Insertar Función

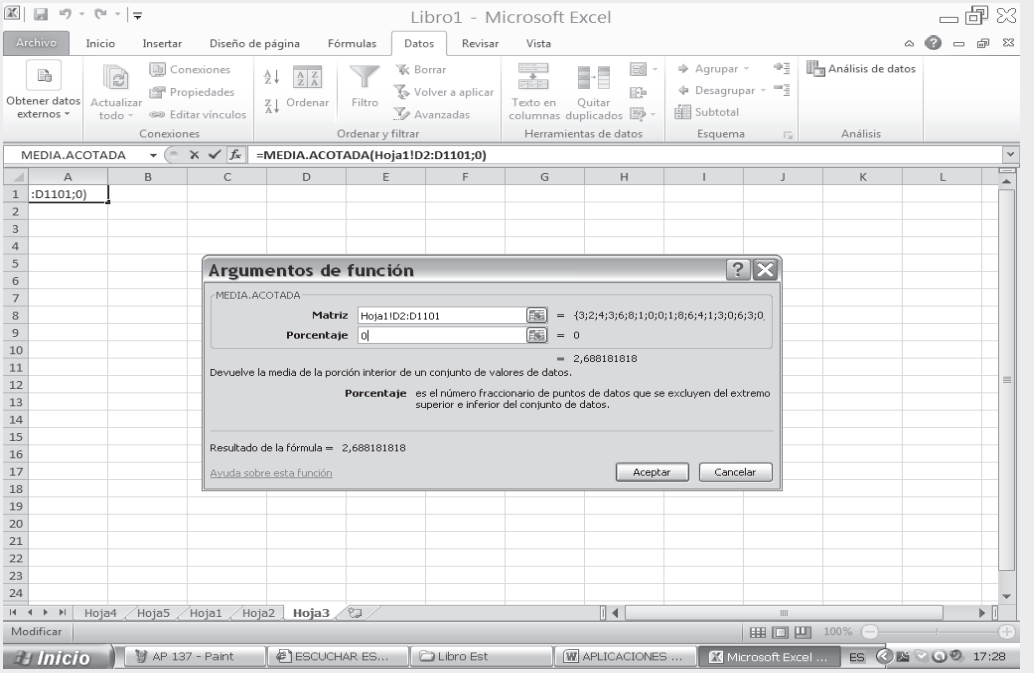


Cuadro No. 1. Resultados para cada una de las variables cuando los datos están sin agrupar

Medidas	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciones ICFES	Edad/años	Estatura/Cm	Peso/Kg
Media	2,687898089	4,66515014	3,489626934	1,651501365	316,5696087	20,3963636	166,9363636	67,1172727
Error típico	0,077286891	0,08489441	0,019766201	0,014379939	1,334706027	0,12466909	0,222495241	0,27102907
Mediana	2	4	3,6	2	310	20	168	66
Moda	0	2	3,6	2	310	20	168	60
Desviación estándar	2,562150779	2,81434901	0,655272679	0,476711778	44,24706493	4,13480593	7,379332318	8,98901725
Varianza de la muestra	6,564616616	7,92056034	0,429382283	0,227254119	1957,802755	17,0966201	54,45454545	80,8024311
Curtosis	0,67110538	-0,30739702	0,063374502	-1,597435984	-0,223624021	0,48917427	1,349355724	-0,02761855
Coeficiente de asimetría	1,066995481	0,75017149	-0,27308826	-0,636768724	0,653420663	1,04230999	-0,484056991	0,3668056
Rango	12	12	3,4	1	180	18	50	42
Mínimo	0	0	1,6	1	240	15	140	46
Máximo	12	12	5	2	420	33	190	88
Suma	2954	5127	3835,1	1815	347910	22436	183630	73829
Cuenta	1099	1099	1099	1099	1099	1100	1100	1100
Mayor (1)	12	12	5	2	420	33	190	88
Menor(1)	0	0	1,6	1	240	15	140	46
Nivel de confianza(95,0%)	0,151646685	0,16657361	0,038783795	0,028215264	2,618862553	0,24461632	0,436563451	0,53179288

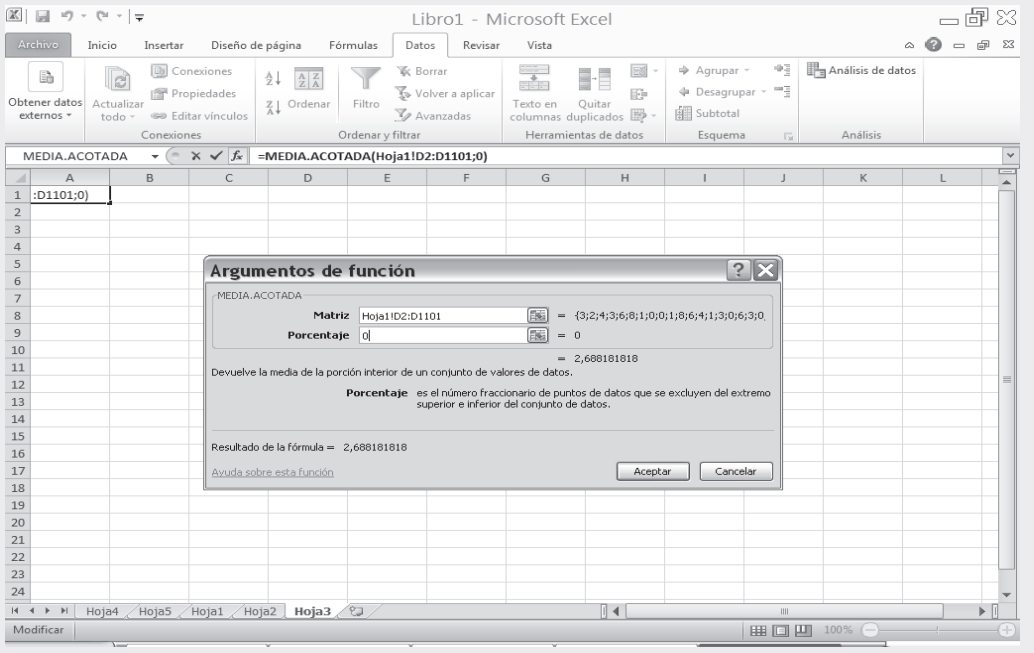
Nota. Observe cuidadosamente los errores que se pueden cometer si consideramos aquellas características que son cualitativas, tales como FACULTAD, SEXO, TRABAJA. En estos casos el resultado no nos dice absolutamente nada, pues las medidas a aplicar, entre otras, sería la proporción, para luego presentarlo como porcentajes.

Figura No.6. Media



- En la figura No. 6, se nos pide en la primera celda la MATRIZ. Se debe ir a la Tabla No. 1 y sombrear todos los datos sin soltar el botón izquierdo.
- En la casilla que dice PORCENTAJE se le da el valor de cero (0) y se observará que el resultado de la MEDIA es de 2.68, siendo el mismo resultado que se obtuvo por el procedimiento anterior, tal como lo muestra la figura No. 7.

Figura No. 7 Cálculo de la media (número de hermanos)



Nota: el procedimiento anterior es aplicado igualmente para las diferentes medidas de POSICIÓN y de DISPERSIÓN, sólo cambia cuando nos ubicamos en el nombre de la función, es decir, la medida que se requiere calcular.

En las siguientes páginas se expondrá de forma general la TEORÍA de ciertas medidas de posición y de Tendencia Central, tales como LA MEDIANA, MODA Y MEDIA GEOMÉTRICA, especificando en cada caso a través de un ejemplo el procedimiento a seguir para obtener los resultados correspondientes en la hoja de cálculo. De forma complementaria, al final del capítulo, se encuentra la unidad de ejercicios para resolver a fin de profundizar sobre estos temas.

MEDIANA- M_E

En datos sin agrupar hay dos maneras de calcularla:

a) Número de Observaciones IMPAR. En este caso se ordenan los datos de menor a mayor o viceversa, luego seleccionamos como MEDIANA al valor central.

Ejemplo: 3 5 12 16 30

Me

b) Número PAR de observaciones. También se ordenan de mayor a menor o de menor a mayor; sólo que en este caso se tienen dos valores en el centro, así que debemos promediar.

Ejemplo: 3 5 12 16 30 46

Me = (12 + 16) / 2 = 14

Figura No. 8. Mediana (Pantallazo Adicional)

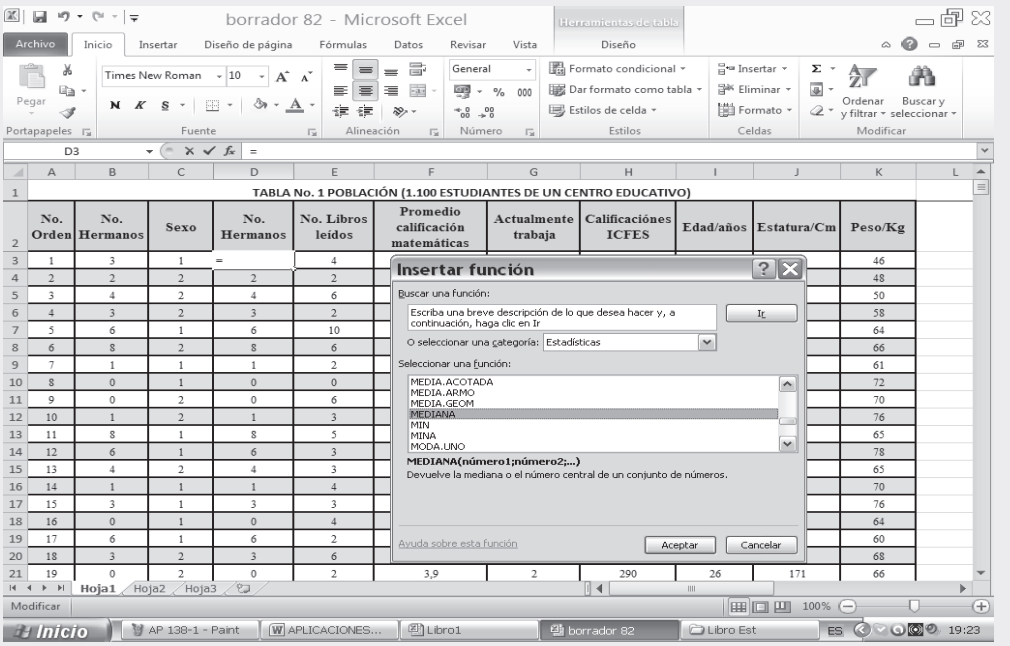
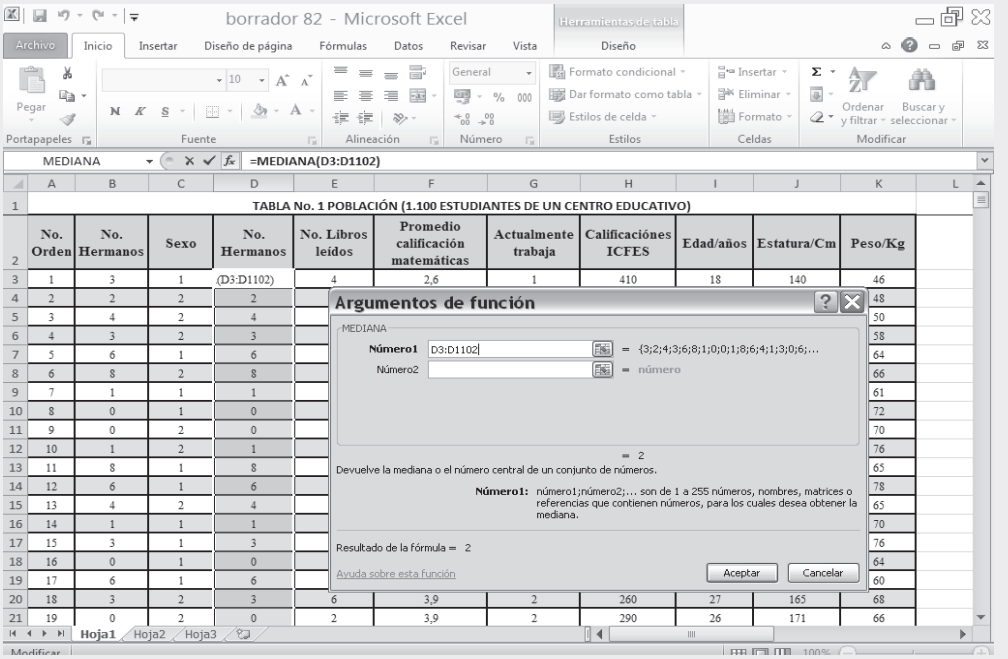


Figura No. 8. Mediana (Pantallazo Adicional)



- Observe, que para el cálculo de la MODA y demás MEDIDAS, aparece en las primeras casillas la palabra NÚMERO 1, es en ésta donde se debe copiar el rango de los valores que aparecen en la Tabla No. 1 correspondiente a la columna NÚMERO DE HERMANOS. Luego de establecer este rango es posible observar, que al compararlo con el obtenido en el anterior método, la MEDIANA sigue siendo igual a 2.0.

Nota: Este resultado se localiza en la celda que se escoja y así sucesivamente. No obstante lo anterior y para efectos de profundizar sobre estos temas, le invitamos a realizar el mismo procedimiento para las demás columnas de la Tabla No. 1.

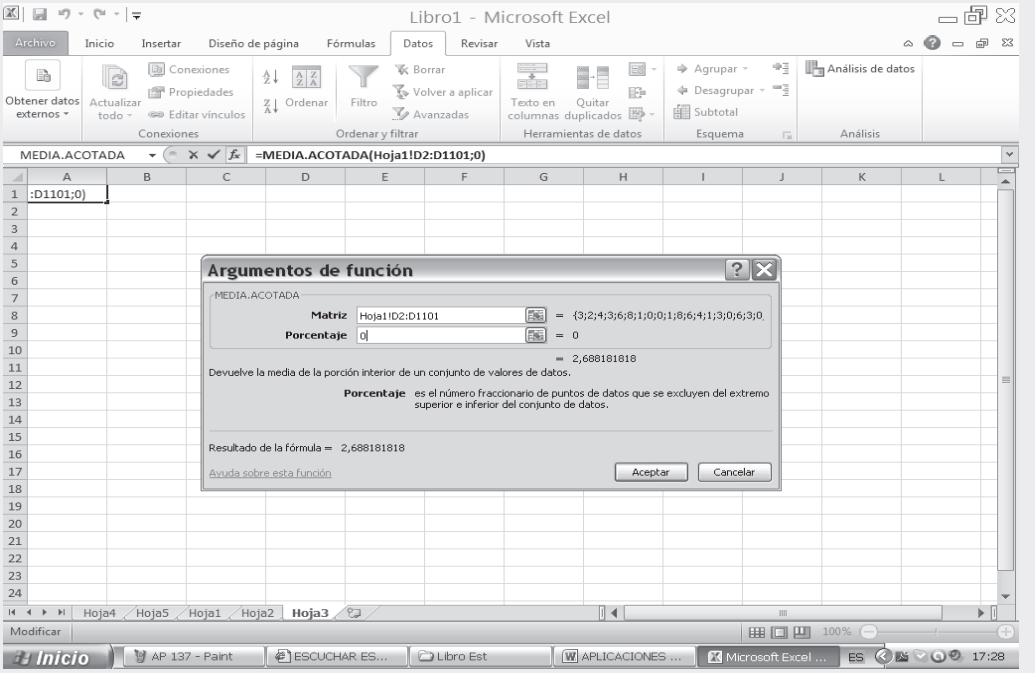
MODO = MODA = VALOR MODAL: M_D

Se define como aquel valor de la variable que más se repite. Para algunos no es aconsejable calcularla mediante la utilización de Excel en caso de que haya más de una MODA, es decir, que sea BIMODAL o PLURIMODAL. Sólo se debe aplicar si es UNIMODAL, es decir, cuando hay una sola moda. En caso de ser BIMODAL O PLURIMODAL, sólo se reconoce la primera MODA de la lista.

Los pasos son los mismos utilizados para el cálculo de la MEDIA y la MEDIANA, esto es: (Figura No. 9).

Nota: El resultado de una MODA = 0 nos indica que cero es el valor que más se repite en la muestra. Obsérvese que el resultado anterior, es el mismo al obtenido en el cuadro No. 5.

Figura No.9. Moda

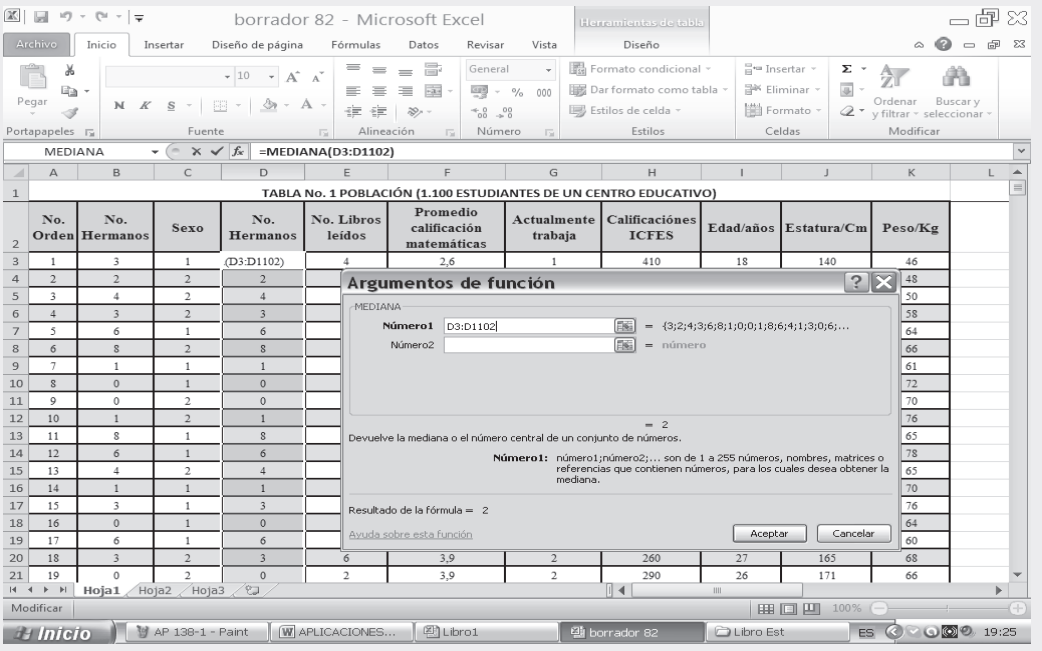


MEDIA GEOMÉTRICA : $M_G = M_O$

Se define como la "raíz enésima del producto de los valores de la variable.

Fórmula $M_o = \sqrt[n]{X_1.X_2.....X_n}$ $\qquad M_o = \sqrt[n]{\prod_{i=1}^{n} x_i}$

Figura No. 10. Estimación Moda (número de hermanos)



- Nuevamente se siguen los mismos pasos aplicados para calcular la MEDIA, MEDIANA Y MODA. Hacemos CLIC en *fx* y seleccionamos CATEGORÍA ESTADÍSTICA, luego, a la derecha elegimos la función MEDIA GEOMÉTRICA, finalmente hacemos CLIC en ACEPTAR.

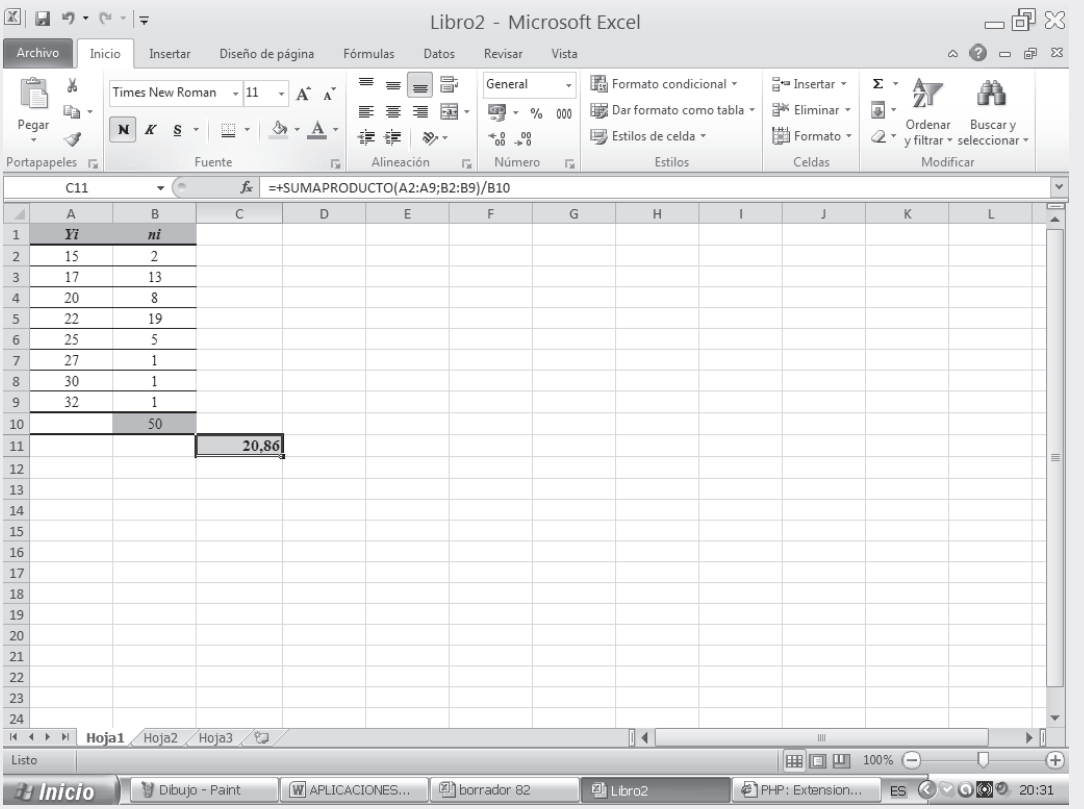
MEDIA ARITMÉTICA PONDERADA

Para utilizar Excel en el cálculo de la MEDIA PONDERADA, se realiza el siguiente procedimiento:

- Tomemos ahora la tabla que aparece a continuación, correspondiente a la variable (y_i): Edad (años); y su frecuencia absoluta, siendo (n_i) el número de veces que se repite cada valor de la variable. Para ello vamos a tener presente una tabla de datos agrupados, aplicando la fórmula que deberá ser digitada:

=SUMA PRODUCTO (A2:A9;B2:B9)/SUMA(B2:B9)

Figura No. 11. Media Aritmética Ponderada



$$\overline{y} = \frac{\sum y_i n_i}{n} = 20{,}86$$

4

CAPÍTULO

Medidas de dispersión, de deformación y apuntamiento

Estudia no para saber algo más sino para saberlo mejor. Ya que el saber más que los otros es fácil. Lo difícil es saber algo mejor que los otros

a

CONTENIDO

Aplicaciones prácticas de las medidas de Dispersión o de deformación, Asimetría y de Apuntamiento.

- Definiciones, fórmulas, propiedades y aplicaciones.
- Medidas de dispersión: Varianza, Desviación típica, Coeficiente de variación, Puntaje típico, Rango o recorrido.
- Medidas de asimetría.
- Medidas de apuntamiento.
- Síntesis de la unidad.
- Ejercicios para Resolver, resueltos en el Sistema de Información en Línea SIL.

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Seleccionar, comprender y aplicar cada una de las medidas, en casos que lo ameriten.
- Manejar los conceptos, fórmulas, usos e importancia que cada una de estas medidas lo exige, no sólo en los talleres, sino en casos reales.
- Comprender sus aplicaciones y agilizar sus cálculos.
- Entender los conceptos, las ventajas y desventajas en su aplicación.

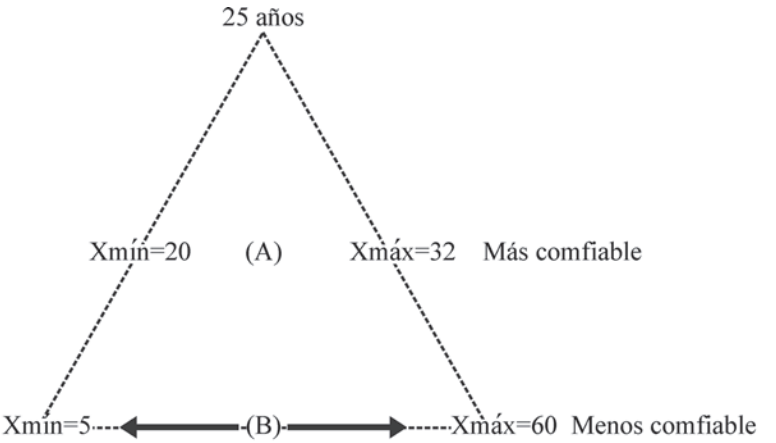
CONCEPTOS GENERALES

Estos primeros cuatro capítulos han sido dedicados a la *estadística descriptiva*. Comenzamos con el manejo de términos y conceptos, luego, el uso de la simbología para hablar un mismo lenguaje, lo cual nos permitió la elaboración de cuadros y gráficas; luego la aplicación de *promedios* y ahora, el cálculo de las *medidas de dispersión*, que será el tema de este capítulo.

Al calcular un promedio, por ejemplo la media aritmética, no sabemos hasta donde se dá la representatividad para ese conjunto de datos, sin embargo es posible determinar si hay concentración de datos alrededor del promedio, el cual nos indicaría una buena aplicación, por el contrario una gran dispersión estaría indicando poca representatividad, por lo tanto no sería confiable o adecuada para el conjunto de datos. Aún más, las medidas de dispersión son indicadas cuando queremos evaluar dos o más promedios. Para este capítulo el estudiante tiene la necesidad de repasar los contenidos del capítulo 3.

MEDIDAS DE DISPERSIÓN

Supongamos que dos distribuciones de edades tienen igual media, pero diferentes grados de dispersión, por lo tanto hace que una de ellas sea más representativa o más confiable.



Observe que 32 - 20 = 12, y 60 - 5 = 55, por lo tanto 12 < 55 permite decir, que hay menos variabilidad en el primer caso; lo cual hace que el promedio de 25 años de edad sea mucho más representativo.

VARIANZA : S^2 ó σ^2

Las medidas de dispersión más conocidas y utilizadas son la *varianza y la desviación típica o estándar*. Esta última, es la raíz cuadrada de aquélla.

La *varianza* se define como: *la media aritmética de los cuadrados de las diferencias (desviaciones) entre los valores que toma la variable y su media aritmética*. Su símbolo es $S2$ en la muestra, $\sigma 2$ (sigma al cuadrado) en la población.

ato a a

1) $S^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}$

2) $S^2 = \dfrac{\Sigma x_i^2 - n\bar{x}^2}{n}$

3) $S^2 = \dfrac{\Sigma x_i^2}{n} - \bar{x}^2$

(muestra);

$\sigma^2 = \dfrac{\Sigma(X_i - \quad)^2}{N}$

$\sigma^2 = \dfrac{\Sigma X_i^2 - N \quad ^2}{N}$

$\sigma^2 = \dfrac{\Sigma X_i^2}{N} - \quad ^2$

Veamos con un ejercicio cómo se aplican las fórmulas anteriormente presentadas, obteniéndose un resultado igual a la aplicación en cada uno de ellos.

Ejemplo 1. Con los siguientes datos (sin agrupar) calcule la varianza por los diferentes métodos

$x_1 = 5 \qquad x_2 = 3 \qquad x_3 = 1 \qquad x_4 = 6 \qquad x_5 = 10$

1) $S^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}$ $\qquad S^2 = \dfrac{\Sigma d_i^2}{n}$

$$\bar{x} = \frac{5+3+1+6+10}{5} = \frac{25}{5} = 5$$

$$S^2 = \frac{(5-5)^2 + (3-5)^2 + (1-5)^2 + (6-5)^2 + (10-5)^2}{5}$$

$$S^2 = \frac{0+4+16+1+25}{5} = \frac{46}{5} = 9{,}2$$

También se podrá calcular asi:

2) $S^2 = \dfrac{\Sigma x_i^2 - n\bar{x}^2}{n}$ $\qquad$ 3) $S^2 = \dfrac{\Sigma x_i^2}{n} - \bar{x}^2$

$$\Sigma x_i^2 = 5^2 + 3^2 + 1^2 + 6^2 + 10^2 = 25 + 9 + 1 + 36 + 100 = 171$$

$$S^2 = \frac{171 - 5(5)^2}{5} = \frac{171 - 125}{5} = \frac{46}{5} = 9{,}2$$

$$S^2 = \frac{171}{5} - 5^2 = 34{,}2 - 25 = 9{,}2$$

Procedimiento de cálculo

- Se calcula la media aritmética, por cualquiera de los procedimientos que han visto.
- En la primera fórmula, se establecen las diferencias entre los valores de la variable con respecto a su media y luego se elevan al cuadrado: $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2$ y así sucesivamente.
- Para la segunda y tercera fórmula se calcula la Σx_i^2, se debe tener presente que corresponde a la suma de $x_1^2 + x_2^2 + x_3^2$ y así sucesivamente.

ato a a o

Las fórmulas son casi iguales a las anteriores, salvo que los valores están ponderados, es decir, multiplicados por las frecuencias absolutas o las relativas.

1) $s^2 = \frac{\Sigma(y_i - \bar{y})^2 n_i}{n}$ $\quad S^2 = \frac{\Sigma(X_i - \bar{X}) f_i}{n}$

2) $s^2 = \frac{\Sigma y_i^2 n_i - n\bar{y}^2}{n}$ $\quad S^2 = \frac{\Sigma X_i^2 f_i - n\bar{X}^2}{n}$

3) $s^2 = \frac{\Sigma y_i^2 n_i}{n} - \bar{y}^2$ $\quad S^2 = \frac{\Sigma X_i^2 f_i}{n} - \bar{X}^2$

Ejemplo 2. Calculemos la varianza para la siguiente distribución, aplicando las tres fórmulas anteriores. Siendo el procedimiento igual tanto al trabajar con variable discreta o continua.

y_i	n_i	$y_i n_i$	$y_i - \bar{y}$	$(y_i - \bar{y}) n_i$	$(y_i - \bar{y})^2 n_i$	y_i^2	$y_i^2 n_i$
5	2	10	−9,75	−19,50	190,1250	25	50
10	4	40	−4,75	−19,00	90,2500	100	400
15	8	120	0,25	2,00	0,5000	225	1.800
20	5	100	5,25	26,25	137,8125	400	2.000
25	1	25	10,25	10,25	105,0625	625	625
Σ	20	295	—	0	523,7500	—	4.875
X_i	f_i	$X_i f_i$	d_i	$d_i f_i$	$d_i^2 f_i$	X_i^2	$X_i^2 f_i$

$\bar{y} = \frac{\Sigma y_i n_i}{n}$ $\quad \bar{y} = \frac{295}{20} = 14{,}75$ $\quad \bar{X} = \frac{\Sigma X_i f_i}{n}$

1) $s^2 = \frac{\Sigma(y_i - \bar{y})^2 n_i}{n} = \frac{523{,}75}{2} = 26{,}1875$ $\quad S^2 = \frac{\Sigma d_i^2 f_i}{n}$

2) $s^2 = \frac{\Sigma y_i^2 n_i - n\bar{y}^2}{n} = \frac{4.875 - 20\,(14{,}75)^2}{20} = 26{,}1875$ $\quad S^2 = \frac{\Sigma X_i^2 f_i - n\bar{X}^2}{n}$

3) $s^2 = \frac{\Sigma y_i^2 n_i}{n} - \bar{y}^2 = \frac{4.875}{20} - 14{,}75^2 = 26{,}1875$ $\quad S^2 = \frac{\Sigma X_i^2 f_i}{n} - \bar{X}^2$

Hay dos fórmulas más, conocidas como *métodos abreviados*, utilizando las desviaciones con respecto a un valor arbitrario (cualquier valor que usted escoja) que no serán presentados y por lo tanto desarrollados.

Se pudo observar en el ejercicio anterior que en el cálculo de la *varianza*, utilizando las diferentes fórmulas, producen el mismo resultado cuando son aplicadas en un mismo conjunto de observaciones.

Propiedades de la varianza

- ***a:*** La varianza debe ser siempre un valor positivo. $S^2 > 0$.
- ***a***: La varianza de una constante es igual a 0.

$$V_{[\ ]} = M_{[\ -M(k)]^2} = M_{[\ -\ ]^2} = 0$$

- ***a:*** La varianza de una constante más una variable, es igual a la varianza de la variable.

$$V_{[y+\ ]} = V_{[y\ ]} + V_{[\ ]} = V_{[y]} \qquad V_{[x+\ ]} = V_{[x]} + V_{[\ ]} = V_{[x]}$$

Válido también para la diferencia: $V_{[x-\ ]} = V_{[x]} - V_{[\ ]} = V_{[x]}$

- ***a ta:*** La varianza de una constante por una variable, es igual al producto de la constante al cuadrado por la varianza de la variable.

$$V_{[\ y]} = K^2 V_{[y]} = K^2 S_y^2$$

- ***ta:*** Para el cálculo de la varianza en una muestra, cuando se está trabajando con submuestras, se aplica la siguiente fórmula:

$$S^2 = \frac{S_1^2 n_1 + S_2^2 n_2 + \ldots}{n_1 + n_2 + \ldots} + \frac{(\overline{x}_1 - \overline{x})^2 n_1 + (\overline{x}_2 - \overline{x})^2 n_2 + \ldots}{n_1 + n_2 + \ldots}$$

sabiendo que: $\overline{x} = \dfrac{\overline{x}_1 n_1 + \overline{x}_2 n_2 + \ldots}{n_1 + n_2 + \ldots}$

Ejercicios de aplicación de las propiedades

Ejemplo 4. *Primera propiedad*: supongamos que se tienen dos distribuciones A y B y se desea aplicar la fórmula de la varianza en ambos casos.

a)

x_i :	8	2	10	6	4	$\overline{x} = 6$

b)

y_i :	0	1	2	3	4	$\overline{y} = 1{,}65$
n_i :	4	6	5	3	2	$= 20$

Las varianzas serán:

a) $S^2 = \dfrac{\Sigma x_i^2 - n\overline{x}^2}{n} = \dfrac{220 - 5(6)^2}{5} = 8$ $\qquad \left[\ S^2 = \dfrac{\Sigma x^2 - n\overline{x}^2}{n}\ \right]$

b) $S^2 = \dfrac{\Sigma y_i^2 n_i - n\overline{y}^2}{n} = \dfrac{85 - 20(1{,}65)^2}{20} = 8 = 1{,}527 \cong 1{,}53$

Observemos que en ambos casos se cumple la *primera propiedad, la varianza de una variable siempre será mayor a cero* ($S^2 > 0$). Por lo tanto siempre será positiva.

Ejemplo 5. *La segunda propiedad*, no requiere que se haga un ejercicio de aplicación, pues si consideramos que en un curso el profesor califica un trabajo o taller con 4 para todos los alumnos, el promedio será de cuatro y la varianza será igual a cero, pues ninguna nota será superior o inferior a cuatro es decir, no hay variación.

Ejemplo 6. *Tercera propiedad:* consideremos que las dos variables del ejemplo 4 están dadas en la misma unidad de medida, por ejemplo en kilogramos, y se requiere que el resultado se dé en libras (aproximadamente 2 libras por kg); para ello se tendrá que:

x_i :	16	4	20	12	8

$\overline{x} = 12$ libras y $S^2 = 32$ libras^2

$$S^2 = \frac{880 - 5(12)^2}{5} = 32 \text{ libras}^2 \qquad V_{[+y]} = {}^2S^2 = 2^2(8) = 32 \text{ libras}^2$$

Algo similar ocurre al trabajar con datos agrupados. Observe que el resultado obtenido, siempre será expresado en las mismas unidades, *pero al cuadrado*.

Con los datos del ejemplo 4–b se harán los calculos

y_i	n_i	$y_i n_i$	$y_i^2 n_i$
0	4	0	0
2	6	12	24
4	5	20	80
6	3	18	108
8	2	16	128
Σ	20	66	340
X_i	f_i	$X_i f_i$	$X_i^2 f_i$

$$S^2 = \frac{\Sigma y_i^2 n_i - n\bar{y}^2}{n}$$

$$S^2 = \frac{340 - 20(3,3)^2}{20} = 6,11$$

$$\bar{y} = \frac{66}{20} = 3,3 \qquad \left[\bar{X} = \frac{\Sigma x_i f_i}{n}\right]$$

$$V_{[KX]} = K^2 S^2 = 2^2(1,527) = 6,11$$

Ejemplo 7. *Cuarta propiedad:* si procedemos a sumarle una determinada cantidad a cada valor observado, por ejemplo 5, se tendrá que x_i : 13 7 15 11 9 (ver ejemplo 4–a)

La media será: $\bar{x} = 11$ y la varianza = 8

$V_{[X+K]} = S^2 + V_{[K]} = 8 + 0 = 8$

Se podrá comprobar aplicando la fórmula: $S^2 = \frac{645 - 5(11)^2}{5} = 8 \qquad V_{[K]} = 0$

En datos agrupados al aplicar la *cuarta propiedad*, se tendrá que:

y_i	n_i
0	4
1	6
2	5
3	3
4	2
Σ	20

y_i	n_i	$y_i n_i$	$y_i^2 n_i$
5	4	20	100
6	6	36	216
7	5	35	245
8	3	24	192
9	2	18	162
Σ	20	133	915
X_i	f_i	$X_i f_i$	$X_i^2 f_i$

$$\left[\bar{x}\right] = \bar{y} = \frac{133}{20} = 6,65$$

$$M_{[y+K]} = 1,65 + 5 = 6,65$$

$$S^2 = \frac{915 - 20(6,65)^2}{20} = 1,527 \cong 1,53$$

Ejemplo 8. *Quinta propiedad*: finalmente consideramos que las dos variables o las dos distribuciones anteriores (a y b) corresponden a dos submuestras, siendo los resultados para cada submuestras los siguientes:

$n_1 = 5 \qquad n_2 = 20 \qquad \bar{x}_1 = 6 \qquad \bar{x}_2 = 1,65 \qquad S_1^2 = 8 \qquad S_2^2 = 1,53$

Se quiere calcular la media y la varianza para el conjunto de las 25 observaciones, aplicando la *quinta propiedad*.

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n} \qquad \bar{x} = \frac{6(5) + 1,65(20)}{25} = 2,52$$

$$S^2 = \frac{S_1^2 n_1 + S_2^2 n_2}{n} + \frac{(\bar{x}_1 - \bar{x})^2 n_1 + (\bar{x}_2 - \bar{x})^2 n_2}{n}$$

$$S^2 = \frac{(5)8 + 1{,}53(20)}{25} + \frac{(6 - 2{,}52)^2 5 + (1{,}65 - 2{,}52)^2 20}{25}$$

$$S^2 = 2{,}824 + \frac{60{,}552 + 15{,}138}{25} = 2{,}824 + 3{,}027 = 5{,}851$$

DESVIACIÓN TÍPICA : S y σ

La *desviación típica* es la raíz cuadrada de la varianza, considerada siempre con signo positivo. Es la medida de dispersión más extensamente aplicada.

$$S = +\sqrt{S^2} \qquad S = +\sqrt{26{,}19} = 5{,}12$$

La desviación típica o estándar será la medida que más se va a utilizar de aquí en adelante, no sólo en este capítulo, sino en todos los siguientes de este libro.

La varianza se expresa siempre en unidades diferentes a las originales, es decir, si la variable se refiere al peso en kg, al calcular la varianza el resultado obtenido estará dado en kg al cuadrado, sucediendo exactamente igual con cualquiera de las unidades de medidas que utilicemos. Esta es la razón por la cual siempre estamos utilizando *la desviación típica o desviación estándar*, pues el resultado se expresa en las mismas unidades de la variable.

Supongamos que se tiene una población normalmente distribuida, con media:

$$\mu = 30 \qquad \sigma^2 = 15 \qquad \text{y} \qquad \sigma = 3{,}87$$

De acuerdo con la desigualdad de Tchebycheff, puede deducirse para cualquier distribución normal o simétrica que la media aritmética es más o menos....

Si se toma una sola vez la desviación típica al lado y lado de la media, el intervalo obtenido, incluye el 68,3% de las observaciones.

$$\mu \pm 1\sigma \Rightarrow 30 \pm 1(3{,}87) \begin{cases} \rightarrow 33{,}87 \\ \rightarrow 26{,}13 \end{cases} \quad 68{,}3$$

$$\mu \pm 1{,}5\sigma \Rightarrow 30 \pm 1{,}5(3{,}87) \begin{cases} \rightarrow 35{,}805 \\ \rightarrow 24{,}1\ 5 \end{cases} \quad 86{,}6$$

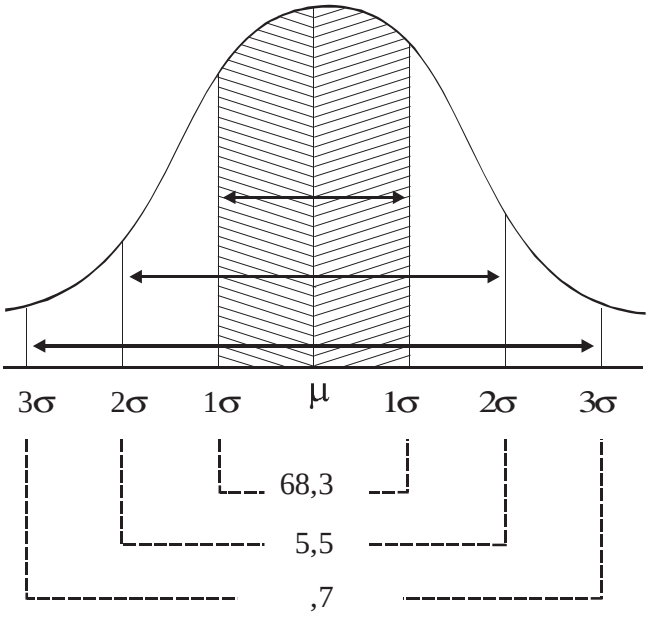
$$\mu \pm 2\sigma \Rightarrow 30 \pm 2(3{,}87) \begin{cases} \rightarrow 37{,}74 \\ \rightarrow 22{,}26 \end{cases} \quad 5{,}5$$

$$\mu \pm 3\sigma \Rightarrow 30 \pm 3(3{,}87) \begin{cases} \rightarrow 41{,}46 \\ \rightarrow 18{,}3 \end{cases} \quad ,7$$

En los próximos capítulos 7 y 8 se ampliarán estos conceptos de *limites de confianza* y, cual es su aplicación y significado.

Gráficamente para una distribución normal, teniendo en cuenta que la ***media aritmética***, la ***mediana*** y el ***modo*** son iguales, corresponderán al valor central de la distribución

$$= M_e = M_d$$



APLICACIONES DEL EXCEL

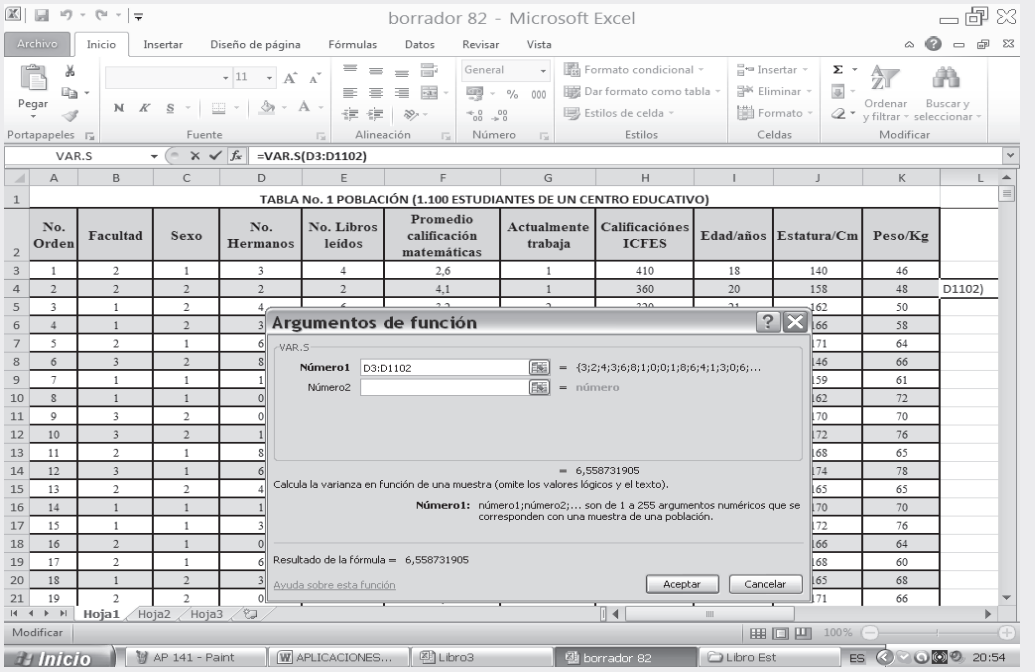
Varianza (aplicada a la muestra) $S^2 = \sigma^2$

El procedimiento es exactamente igual a los aplicados para calcular: Media, ***Mediana, Moda y Media Geométrica***. Seleccionamos el renglón correspondiente a la función VAR, de acuerdo a la figura No. 1. Luego establecemos el rango separado por dos puntos pertenecientes a la variable NÚMERO DE HERMANOS y hacemos CLIC en ACEPTAR

- Observemos que el valor de la varianza por este método es el mismo obtenido por el anterior método de cálculo.

  Nota: En Excel 2010 es posible encontrar nuevas funciones de varianza. Si los datos representan la población total se utiliza la función VARP y si por el contrario sólo se toma una muestra de la población total, se utiliza la función VARM. En este caso aplicamos VARM.

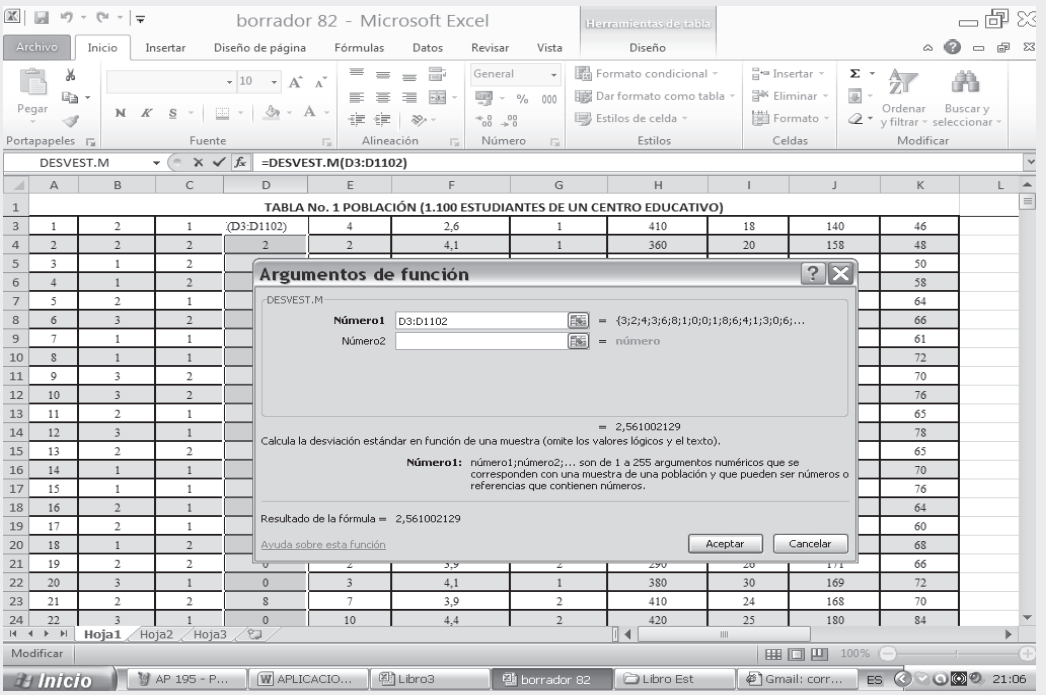
Figura No. 1. Argumentos para una Varianza



Desviación típica estándar $= \sigma$

Se define como la Raíz cuadrada de la Varianza: $\overline{S} = \sqrt{S^2}$, siguiendo el mismo procedimiento que hemos venido utilizando, pero esta vez, buscamos la función DESVEST y de esta manera se obtienen los resultados siguientes:

Figura No. 2. Cálculo Desviación Típica



Al observar este resultado se ve claramente que es la raíz cuadrada de la varianza, es decir que

$$\mathbf{S = \sqrt{6{,}558} = 2{,}56}$$

Nota: Todas las medidas que aparecen en la columna NOMBRE DE LA FUNCIÓN podrán calcularse de la mism que las anteriores.

En Excel 2010 es posible encontrar nuevas funciones de desviación estándar. Si los datos representan la población total se utiliza la función DESVESTP y si por el contrario sólo se toma una muestra de la población total se utiliza la función DESVESTM. En este caso aplicamos DESVESTM.

USO DE LA CALCULADORA

El cálculo de la desviación típica por medio del uso de la calculadora, se sigue el mismo procedimiento que se indicó para el cálculo de ΣX_i ΣX_i^2 n y $\overline{X}$ (ver página 130). Se debe calcular la desviación típica para las poblaciones o en muestras ($n > 30$) ver la tecla 2, donde generalmente aparece indicado ${}_x\sigma_n$ que corresponde a desviación típica sin corregir.

En algunas, el valor de σ se obtiene oprimiendo SHIFT 2; en otras calculadoras será: SHIFT 2 OPCIONES y seleccione la opción deseada y ese resultado obtenido al ser elevado al cuadrado corresponde al valor de la varianza ${}_x\sigma_n^2$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

1. En una distribución unidimensional se sabe, que la media aritmética es igual a 3,04 y que la media cuadrática es igual a 4,26. Se pide calcular la varianza y la desviación típica de dicha distribución.

2. En la sección financiera de un diario se publicó recientemente la (distribución unidimensional de variable discreta) tabla adjunta y decía el texto del artículo, que la media aritmética era igual a 120 y la varianza igual a 25.

   Desgraciadamente la publicación apareció con dos manchas de tinta caídas en las columnas de las frecuencias absolutas, lo cual impidió comprobar directamente lo que en el artículo se decía. Sin embargo, se le pregunta ahora a usted, ¿pueden aceptarse dichos valores de la media y la varianza, teniendo en cuenta lo que puede verse en la tabla? Conteste si puede ser posible o no y porqué.

y_i :	105	110	115	120	125	130	135	140
n_i:		37	90	95	85	60		

3. Con la siguiente ecuación:
   $V_{[1/8X\,-\,8]}$ sabiendo que $S_x = 8$. Calcular la varianza de y_i.

4. Con los siguientes datos:
   $y = 1/4\,(8x - 2)$ $\quad x = 4$ $\quad S_x = 2$ Se pide calcular: $V_{[Y]}$ $\quad y \quad M_{[X]}$

5. Una muestra de 70 datos da para una cierta variable una media de 120 y una desviación típica de 6, otra semejante, pero de 30 observaciones, da para la misma variable una media de 125 y una desviación típica de 5. Si se reúnen las dos muestras formando una sola de 100 datos, ¿cuál será su media y su desviación típica?

6. Durante un período de 3 años, los precios de un producto fueron en promedio de $8.000, con una desviación típica de $120. En el período anterior de 3 años, el promedio fue de $5.000 con una varianza de $3.600. ¿En qué período hubo mayor estabilidad?

7. La varianza de dos números es 1 y su media aritmética 8. Calcular esos dos números.

8. Si un conjunto de n valores de x_i se sabe que: $\sum x_i = 10$ $\Sigma x_i^2 = 260$ $S^2 = 25$. Se pide encontrar el valor de n.

9. La media de 10 observaciones es 3 y la suma de sus cuadrados es 100. Encontrar la desviación estándar del conjunto.

10. Sumando 4 a cada uno de los números de la serie 2 6 5 9 1, se obtiene la serie 6 10 9 13 5. Comprobar que ambas series tienen la misma varianza y distintas medias.

11. Hallar la desviación típica de dos números cuya media aritmética es 9,0 y su media geométrica es 7,2.

12. Tres profesores de estadística registraron una calificación media en sus exámenes de 71 78 89 con desviaciones típicas de 9 8 7; sus clases estaban formadas por 30 25 y 15 estudiantes respectivamente. Determinar la calificación media y la desviación típica para el conjunto de los 70 estudiantes. ¿Cuáles serían estas medidas, si todos los cursos tienen el mismo número de estudiantes?

13. La distribución por edades de los inmigrantes extranjeros que en el mes de Abril de 2011 arribaron a un país X por vía marítima fueron:

EDAD (AÑOS):	0 - 10	10,1 - 20	20,1 - 30	30,1 - 40	40,1 - 50	50,1 - 60	60,1 - 70	70,1 - 80	80,1 - 90
PASAJEROS:	640	684	863	876	753	663	414	154	13

a) Calcule el coeficiente de variación de la distribución.

b) Aumente en 15 años todas las edades y calcule el coeficiente de variación ¿Qué opina al respecto?

c) Fijar los límites de confianza para la $\overline{y}$ obtenida en el punto (a), siendo las posibilidades del 86,64% y 98,76%.

14. En un conjunto de n valores de X, se sabe $\sum x_i^2 = 1.360$ $\sum x_i = 40$ y $ns^2 = 1.280$. Se pide encontrar n.

15. Se toman las medidas de 80 personas las que tienen una estatura media de 170 cmts y una varianza de 7,4 cmts. Posteriormente se verificó que el metro usado en la medición tenía 9 cmts menos. Se pide rectificar los estadígrafos mencionados.

16. Dada una serie de valores de x: 3, 2, 8, 1, comprobar que siendo y = 2x + 5, la media de la serie ( ) es $\overline{y} = 2\overline{x} + 5$ y su varianza: $s_y^2 = 4\, s_x^2$

17. Responda a los siguientes puntos señalando si es cierto o falso.
a) En una distribución simétrica la desviación media es igual a 0,7979 S.
b) Si se multiplica la distribución por una constante, el coeficiente de variación se altera.
c) La desviación mediana es menor o igual a la desviación media, a su vez, es menor que la desviación típica

18. Un profesor de educación física informa, que el peso promedio de un grupo de alumnos que conforman el equipo de fútbol es de 56 kilos y que su desviación típica es de 7 cmts. ¿Es correcto? ¿cual debería ser el coeficiente de variación?

19. Si se tienen dos distribuciones A y B en el primero $S^2 = 16$ y en la segunda $S^2 = 25$, ambos se dan en la misma unidad de medida. ¿En cuál de ellas hay una mayor variabilidad absoluta?

20. Si las dos varianzas del ejercicio anterior se dieran en unidades de medidas diferentes, ¿cuál de ellas tienen mayor variabilidad absoluta? Si no se puede responder ¿qué medida de dispersión utilizaría usted?

21. Si en una distribución se tiene que la varianza calculada es de 4 horas ¿cuál sería la varianza de esa misma distribución, pero en minutos?

22. Con los siguientes datos obtenidos en dos distribuciones, con datos en las mismas unidades de medida, ¿cuál será la desviación típica para el conjunto A + B?
$s_A^2 = 28{,}2 \qquad s_B^2 = 14{,}6 \qquad \overline{y}_A = 12 \qquad \overline{y}_B = 9 \qquad n_A = 20 \qquad n_B = 30$

23. ¿Es posible que una varianza sea negativa? ¿Por qué?

VARIACIÓN ABSOLUTA

Dos o más varianzas y dos o más desviaciones típicas se pueden comparar entre sí, si las variables están dadas en las mismas unidades de medidas, en caso contrario, se deberá utilizar el coeficiente de variación.

COEFICIENTE DE VARIACIÓN CV

Más conocido como *variación relativa*. En ocasiones nos interesa comparar la variabilidad de dos series de datos, sin embargo podemos encontrar, al hacerlo, que ambas series están expresadas en diferentes unidades de medidas, por lo tanto no se podrán comparar sus varianzas o sus desviaciones típicas. Puede darse el caso de que estén expresadas en la misma unidad, pero nos interesa determinar la variación respecto a una base. Para resolver los anteriores problemas se usa el *coeficiente de variación*. Si es multiplicado por 100, el resultado se dará en términos porcentuales.

$$CV = \frac{S}{\overline{x}} \quad \text{(relativo)} \qquad\qquad CV = d = \frac{S}{\overline{x}}\ 100 \quad \text{(porcentual)}$$

De acuerdo a la fórmula anterior, es *el cociente entre la desviación típica estándar y su media aritmética*, como en ocasiones se expresa en porcentajes, por esa razón aparece multiplicado por 100. Fue propuesto por Karl Pearson (1895) con el propósito de comparar la variabilidad entre varias distribuciones de frecuencias.

$S = 5{,}12$
$\overline{x} = 14{,}75$

$$CV = d = \frac{5{,}12}{14{,}75}\ 100 = 34{,}71\%$$

La aplicación de esta fórmula presenta el inconveniente de que varía para distribuciones que tienen diferentes medias aritméticas e igual varianza, es decir, que con igual dispersión tienen diferentes coeficientes de variación.

Ejemplo 1. Si consideramos dos distribuciones cuyas medias aritméticas son:

$\overline{x}_1 = 24{,}5$ y $\overline{x}_2 = 30$, además sus desviaciones típicas o estándar son idénticas, $S_1 = 2$ y $S_2 = 2$, esto nos indica que tienen la misma variación absoluta, sin embargo al calcular los coeficientes de variación, encontramos que presentan diferente variación relativa a pesar de tener la misma desviación.

o :

$$CV_1 = d_1 = \frac{S_1}{\overline{x}_1} = \frac{2}{24{,}5} = 0{,}0816 = 8{,}16\% \qquad CV_2 = d_2 = \frac{S_2}{\overline{x}_2} = \frac{2}{30} = 0{,}0666 = 6{,}66\%$$

Ejemplo 2. Para la media y la varianza de un conjunto se han hallado, respectivamente, los valores 4 y 25. ¿Qué opinión merece la media aritmética?

o :

$$CV = d = \frac{S}{\overline{x}}\ 100$$

$$CV = d = \frac{5}{4} = 1{,}25 = 125\%$$

Esta media aritmética no es lo suficientemente representativa de la distribución, debido a la variabilidad tan alta que presenta.

Ejemplo 3. Multiplicando por 4 cada uno de los valores de la variable x_i : 3 2 0 5 se obtiene la serie x_i: 12 8 0 20. Comprobar que ambas series tienen el mismo coeficiente de variación.

o :

x_i	$4(x_i)$	x_i^2	$16(x_i^2)$
3	12	9	144
2	8	4	64
0	0	0	0
5	20	25	400
10	40	38	608

$$\overline{x}_1 = \frac{10}{4} = 2{,}5$$

$$\overline{x}_2 = \frac{40}{4} = 10$$

$$M_{KX} = 4(2{,}5) = 10 = \overline{x}_2$$

$$S_1^2 = \frac{\Sigma x_i^2}{n} - \overline{x}^2 = \frac{38}{4} - 2{,}5^2 \qquad S_1^2 = 9{,}5 - 6{,}25 = 3{,}25 \qquad S_1 = \sqrt{3{,}25} = 1{,}80$$

$$V_{[KX]} = K^2 V_{[x]} \qquad V_{[4X]} = 16\ (3{,}25) = 52$$

$$V_{[4X]} = \frac{608}{4} - 10^2 = 152 - 100 = 52 \qquad V_{[4X]} = 52 \qquad S_2 = 7{,}21$$

$$CV_1 = \frac{1{,}8}{2{,}5} = 0{,}72 = 72\% \qquad CV_2 = \frac{7{,}21}{10} = 0{,}72 = 72\% \qquad CV_1 = CV_2$$

PUNTAJE TÍPICO O ESTANDARIZADO: z ó "t

Este estadígrafo de dispersión, muy utilizado en la distribución normal y en el análisis del coeficiente de correlación, mide la *desviación de una observación con respecto a la media aritmética en unidades de desviación típica*, determinando la posición de una observación dada, dentro de un conjunto de observaciones.

$$Z = \frac{x - \bar{x}}{S} \qquad \text{ó} \qquad t = \frac{x - \bar{x}}{S}$$

El *puntaje típico* sirve para comparar dos o más datos individuales, aunque pertenezcan a distribuciones diferentes, aún en casos en que la media y/o varianza no coincidan.

Tanto el puntaje "*t*" o "*z*" es muy utilizado en las pruebas educacionales y psicológicas, aunque hoy en día es muy frecuente en otros tipos de experimentos.

Ejemplo 1 Supongamos que una persona presenta exámenes de admisión para la Facultad de Economía, en dos universidades de San José. En la primera obtuvo una calificación de 85 y en la segunda de 60, además se sabe que el promedio de calificación para el ingreso de la universidad A es de 90, con una desviación estándar de 10, en cambio en la universidad B es de 55, con una desviación estándar de 12. Se desea saber en cuál de las universidades esta persona se desempeñó relativamente mejor?

$$Z_A = \frac{85 - 90}{10} = \frac{-5}{10} = -0{,}50 \qquad Z_B = \frac{60 - 55}{12} = \frac{5}{12} = 0{,}42$$

Lo anterior indica, que el estudiante se desempeñó mejor en la universidad B, donde logró 0,42 de desviación estándar por encima del promedio, en cambio, en B fue de – 0,50 por debajo del promedio.

En una *distribución normal la media de la variable Z vale* 0 *y su varianza es igual a* 1. El puntaje típico estandarizado va a ser utilizado en gran medida en el Capitulo 6, correspondiente a la Distribución Normal.

Ejemplo 2. En un examen final de estadística, la puntuación media de un grupo de 150 estudiantes fue de 78 y la varianza 64. En álgebra, sin embargo, la media final del grupo fue 73 y la desviación típica 7,6. En qué asignatura hubo mayor.

a) Dispersión absoluta b) Dispersión relativa

c) Si el estudiante consiguió una nota de 75 en estadística y 70 en álgebra ¿En qué asignatura fue su puntuación relativa superior?

o :

a) Dispersión absoluta:

$S_1^2 = 64$ $S_1 = 8$ $S_2^2 = 57{,}76$ $S_2 = 7{,}6$

En estadística hubo una mayor dispersión absoluta: $S_1 > S_2$ $8 > 7{,}6$

b) Dispersión relativa

$$CV_1 = \frac{8}{78} = 0{,}1025 \qquad CV_1 = 10{,}25\%$$

$$CV_2 = \frac{7{,}6}{73} = 0{,}1041 \qquad CV_2 = 10{,}41\%$$

En álgebra hubo una mayor dispersión relativa: $CV_2 > CV_1$ → 10,41% > 10,25%

c) Puntuación relativa: $Z = \frac{x - \bar{x}}{S}$

$$Z_1 = \frac{75 - 78}{8} = \frac{-3}{8} = -0{,}37$$

$$Z_2 = \frac{71 - 73}{7{,}6} = \frac{-2}{7{,}6} = -0{,}26 \qquad Z_2 > Z_1$$

Su puntuación relativa fue superior en álgebra, dado que – 0,26 > – 0,37

DESVIACIÓN MEDIA : $_a$

Es otro estadígrafo de dispersión de menor importancia, en relación a la varianza y la desviación típica. Se define como la *media de las desviaciones respecto a la media aritmética*, *tomadas en valor absoluto.*

$$D_a = \frac{\Sigma | x_i - \bar{x} |}{n} \quad \text{(Datos no agrupados)} \qquad D_a = \frac{\Sigma | X_i - \bar{X} |}{N} = \frac{\Sigma | d_i |}{N}$$

$$D_a = \frac{\Sigma | y_i - \bar{y} | n_i}{n} \quad \text{(Datos agrupados)} \qquad D_a = \frac{\Sigma | X_i - \bar{X} | f_i}{N} = \frac{\Sigma | d_i | f_i}{N}$$

Ejemplo 1. Con los siguientes datos no agrupados, calcular la desviación media:

$x_1 = 5 \qquad x_2 = 3 \qquad x_3 = 1 \qquad x_4 = 6 \qquad x_5 = 10$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
5	0	0
3	−2	2
1	−4	4
6	1	1
10	5	5
25	0	12

$$\bar{x} = \frac{25}{5} = 5$$

$$D_a = \frac{\Sigma | x_i - \bar{x} |}{n} \qquad D_a = \frac{\Sigma | d_i |}{N}$$

$$D_a = \frac{12}{5} = 2{,}4$$

Ejemplo 2. Con los datos agrupados, que se presentan a continuación, calcular la Desviación Media.

El procedimiento de cálculo para datos agrupados es el siguiente:

y_i	n_i	$y_i - \bar{y}$	$\lvert y_i - \bar{y} \rvert$	$\lvert y_i - \bar{y} \rvert n_i$
5	2	−9,75	9,75	19,50
10	4	−4,75	4,75	19,00
15	8	0,25	0,25	2,00
20	5	5,25	5,25	26,25
25	0	10,25	10,25	10,25
Σ	20	—	—	77,00
X_i	f_i	d_i	$\lvert d_i \rvert$	$\lvert d_i \rvert f_i$

$$\bar{y} = 14{,}75 = \bar{X}$$

$$D_a = \frac{\Sigma | y_i - \bar{y} | n_i}{n}$$

$$D_a = \frac{77}{20} = 3{,}85$$

$$D_a = \frac{\Sigma | d_i | f_i}{n}$$

En una distribución Simétrica o Normal, la Desviación Media es igual al producto de 0,7979 por la desviación típica.

$$D_a = 0{,}7979\, S$$

En el ejercicio anterior tenemos una desviación típica de 5,12, por lo tanto la Desviación media será igual a: $D_a = 0{,}7979\ (5{,}12) = 4{,}09$

Dado que la distribución anterior no es *simétrica,* presenta el resultado obtenido alguna diferencia; sin embargo, vemos que la relación se cumple aproximadamente, por ser la distribución ligeramente *asimétrica*.

La *desviación media* se usa muy poco; en el caso de ser calculada en una *distribución de frecuencias* será necesario suponer que la *media aritmética* representa adecuadamente a los valores de la variable.

El valor obtenido al calcular la *desviación media* siempre será menor al de la *desviación típica*

$$D_a < S$$

COEFICIENTE DE DESVIACIÓN MEDIA : $_a$

Esta medida de dispersión es muy parecida en su presentación y cálculo al *Coeficiente de variación*, con la diferencia de considerar a la *Desviación media* (*Da*) en vez de la *Desviación típica* (*S*). También se puede expresar en términos porcentuales.

Ejemplo 1. De acuerdo a los datos de dos ejemplos anteriores (desviación media) se pide aplicar esta medida de dispersión:

o **:**

Ejemplo1: $CD_a = \frac{D_a}{\bar{x}} 100 = \frac{2{,}4}{5} = 0{,}48$ o 48% (en datos no agrupados)

Ejemplo 2: $CD_a = \frac{D_a}{\bar{y}} 100 = \frac{3{,}85}{20} = 0{,}1925 = 19{,}25\%$ (en datos agrupados)

DESVIACIÓN MEDIANA : $_e$

Corresponde a los denominados estadígrafos o medidas de dispersión, siendo su uso más bien limitado. Se define como la *Media de los valores absolutos de las diferencias entre los valores que toma la variable y su mediana.*

Sus fórmulas son:

$$D_e = \frac{\Sigma \left| x_i - M_e \right|}{n} \quad \text{(Datos no agrupados)}$$

$$D_e = \frac{\Sigma \left| y_i - M_e \right| n_i}{n} \quad \text{(Datos agrupados)} \qquad \left| D_e = \frac{\Sigma \left| X_i - M_e \right| f_i}{N} \right|$$

La desviación mediana, será menor o igual a la *Desviación media* y ésta a la vez, será menor que la *Desviación típica*: $D \leq D_a < S$

Ejemplos 1. Con los siguientes datos calcular la desviación mediana.

o :

a) $x_1 = 5$ $x_2 = 3$ $x_3 = 1$ $x_4 = 6$ $x_5 = 10$

Se calcula primero la mediana para ello ordenamos los valores de menor a mayor.
1 – 3 – 5 – 6 – 10 Luego establecemos las diferencias entre la variable y el promedio M_e

x_i	$x_i - M_e$	$\lvert x_i - M_e \rvert$
1	– 4	4
3	– 2	2
5	0	0
6	1	1
10	5	5
Σ	0	12

$$D_e = \frac{\Sigma \lvert x_i - M_e \rvert}{n}$$

$$D_e = \frac{12}{5} = 2{,}4$$

Ejemplo 2. Con datos agrupados, se pide calcular la desviación media

	y_i	n_i	N_i	
	5	2	2	
	10	4	6	$\leftarrow N_{j-1}$
$y_j \rightarrow$	15	8	14	$\leftarrow N_j$
	20	5	19	
	25	1	20	
	Σ	20	–	
	X_i	f_i	F_i	

$$N_{j-1} < \frac{n}{2}$$

$$6 < 10$$

$$M_e = y_j = 15$$

$$M_e = X_j = 15$$

Continuación:

$y_i - M_e$	$\lvert y_i - M_e \rvert$	$\lvert y_i - M_e \rvert n_i$
–10	10	20
–5	5	20
0	0	0
5	5	25
10	10	10
0	–	75
$X_i - M_e$	$\lvert X_i - M_e \rvert$	$\lvert X_i - M_e \rvert f_i$

$$D_e = \frac{\Sigma \lvert y_i - M_e \rvert n_i}{n}$$

$$D_e = \frac{75}{20} = 3{,}75$$

$$D_e = \frac{\Sigma \lvert X_i - M_e \rvert f_i}{N}$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

24. Un conjunto de 20 valores tiene una media igual a 50; otro conjunto de 20 valores tiene una media igual a 30; la desviación estándar de los 40 valores considerados conjuntamente es igual a 10. Calcular el coeficiente de variación del conjunto de los 40 números.

25. Reconstruya y calcule el coeficiente de variación de la siguiente distribución simétrica. (Para el cálculo de la media y la varianza, hacerlo mediante el método que usted conoce).

X_i:	$\Rightarrow$	y_i	10	20	30	40	50
f_i:	$\Rightarrow$	n_i	6				
F_i/n	$\Rightarrow$	H_i	0,12	0,32			

26. Hallar el coeficiente de variación de dos números, cuya media aritmética es 5 y la media geométrica es 4.

27. Se tienen los siguientes datos correspondientes a dos submuestras:
$\bar{x}_1 = 120 \quad \bar{x}_2 = 125 \quad n_1 = 70 \quad n_2 = 30 \quad S_1^2 = 36 \quad S_2 = 3$
Se pide calcular el coeficiente de variación para el conjunto.

28. Se supone que $x_1 \quad x_2 \quad x_3 ... \; x_n$ son elementos que poseen una media aritmética 10 y una desviación típica de 3. Si se considera una nueva media y_i, de cada uno de los elementos relacionados con la media x_i mediante la ecuación.
$$y_i = 4\,x_i + 2$$
¿Cuál es la media aritmética, la varianza y el coeficiente de variación de los nuevos valores y_i obtenidos?

29. Las notas de 50 alumnos se clasifican en una tabla de frecuencias con cuatro intervalos de igual magnitud. Se pide calcular el coeficiente de variación, sabiendo que:
$$y_2 = 50 \quad n_1 = 4 \quad N_2 = 20 \quad n_3 = 25 \quad \bar{y} = 62{,}4$$
$$X_2 = 50 \quad f_1 = 4 \quad F_2 = 20 \quad f_3 = 25 \quad \bar{X} = 62{,}4$$

30. El cuadro de pagos en dos fábricas para el año 2011 fue así:

MEDIDAS	FÁBRICA A	FÁBRICA B
Salario medio semanal	$930.000	$1.020.000
Desviación típica del salario semanal	$96.000	$97.000

a) ¿En cuál de las fábricas los jornales fueron absolutamente más variables?
b) ¿Relativamente más variables?

31. Se analizaron en el primer semestre del año 1 los gastos de una empresa de la construcción y se obtuvo un promedio de $1.740 millones y una desviación típica de $90 millones. Se determinó luego que los contadores de esta empresa habían prescindido de $13 millones en los gastos por un error de apreciación. Corrigiendo las medidas enunciadas, obtenga el coeficiente de variación de los gastos de esta empresa.

32. Para cierta variable se sabe que $M_{[4+X]} = 11{,}2$ y $M_{[X+2]^2} = 30{,}25$
Calcule el coeficiente de variación.

33. En cierta región la distribución de predios por extensión tiene una media de 35,4 hectáreas y una desviación típica de 19,33 hectáreas, mientras que la distribución por canon de arrendamiento tiene una media de $945.750 y una desviación de $74.708. ¿Cuál de las dos distribuciones tiene mayor variabilidad?

34. Una empresa fabrica bombillas eléctricas de dos tipos (A y B), con base en muestras de la producción, se sabe que las distribuciones de la duración en horas de esas bombillas son tales que tienen las siguientes medias y varianzas.

TIPO	MEDIA	VARIANZA
A	800 horas	7.800
B	650 horas	5.400

Se pide:

a) Comparar ambas distribuciones en cuanto a su variabilidad absoluta y relativa.

b) Si se extrajo una bombilla de cada tipo y su duración fue de 700 y 630 horas respectivamente, se quiere saber cuál tipo de bombilla tiene mejor posición relativa.

35. Un almacén vende un promedio mensual de \$3.000.000 (cientos \$) en telas con una desviación típica de \$30.000 (cientos de \$).

a) El almacén paga un impuesto igual al 10% sobre las ventas. ¿Cuál será la varianza de las ventas una vez pagado el impuesto?

b) Las utilidades del almacén se calculan teniendo en cuenta que mensualmente se pagan \$2.450.000 (cientos de \$) por salarios, gastos, etc., además del impuesto sobre las ventas. Calcule el coeficiente de variación de las utilidades.

36. Los siguientes datos muestran el número de cuartos habitables en 60 viviendas localizadas en la ciudad de Cúcuta.

NO. DE CUARTOS:	2	4	6	8	10
NO. DE VIVIENDAS:	6	18	16	12	8

Obtenga e interprete para estos datos la desviación típica, la desviación mediana y el coeficiente de variación.

37. Los trabajadores de una empresa solicitan, en una convención colectiva, que cada salario semanal de sus afiliados sea aumentado según la ecuación:

$$y_i = 1{,}13x + 10.750$$

La empresa tiene 1.000 trabajadores y antes de solicitar el reajuste salarial, devengaban un promedio de \$120.000 semanal.

a) ¿Cuál será el nuevo promedio de ingreso mensual de los trabajadores si la empresa acepta la petición?

b) El coeficiente de variación antes de solicitar el reajuste salarial es del 38%. ¿Cuál será el nuevo coeficiente de variación?

38. Con los datos siguientes:

$\Sigma y_i n_i = 1.500 \rightarrow \Sigma x_i f_i$

$\Sigma y_i^2 n_i = 40.000 \rightarrow \Sigma x_i^2 f_i$

$n = 100 \rightarrow N$

Se pide calcular:

a) La desviación estándar.

b) El coeficiente de variación.

c) El puntaje típico semanal, si $x = 24$

39. Al trabajar con cierta información se obtienen los siguientes datos: $\bar{x} = 36 \quad S = 8 \quad CV = 0{,}22$.
Si a cada uno de los valores de la variable se les aumenta en 5 ¿Cuáles serían las nuevas medidas: media, desviación estándar y coeficiente de variación?

40. Con los datos correspondientes a dos submuestras.

a)

x_i :	4	6	7	6	10	6	4	15	6	X_1
y_i :	10	18	12	20	8	10				X_2

Se pide:

b) Determinar en cada una de las submuestras la mejor posición de los siguientes valores $x = 8$ y $y = 15$.

c) Transformando las dos submuestras en una sola muestra, calcular el coeficiente de variación.

41. Con la siguiente distribución de frecuencias:

a)

$y'_{i-1} - y'_i$	n_i
8,1 – 16	3
16,1 – 24	6
24,1 – 32	10
32,1 – 40	15
40,1 – 48	4
48,1 – 56	2
Σ	40
$X'_{i-1} - X'_i$	f_i

b) $n = 8$

x_i : 8 10 7 30 25 38 47 70

Se pide determinar:

1) En cuál de las dos distribuciones (a y b) hay una mayor variabilidad absoluta.
2) En cuál hay una menor variabilidad relativa
3) Si se tiene un valor $x = 33$, en cuál de las dos distribuciones, dicho valor tiene una mejor posición relativa.

42. En un grupo de 70 personas, se sabe que 30 de ellas tienen un salario medio diario de $27.000 y los restantes de $32.000.

a) Se pide calcular el salario medio diario de las 70 personas.

b) Si se sabe que el coeficiente de variación es del 35%, ¿cuál es la varianza de los salarios diarios de las 70 personas?

43. Si la media de 10 observaciones es 5,7 y la suma de los cuadrados es 490. Encontrar el coeficiente de variación para el conjunto.

44. Los siguientes datos corresponden a la distribución de frecuencias de los gastos de publicidad (miles de millones de pesos) de 50 empresas comerciales, durante el último trimestre de 2011. Dichos gastos se agruparon en cuatro clases de amplitud constante, de la cual se sabe:

$y_1 = 3{,}5$ $y'_4 = 8{,}75$ $n_1 = 4$ $N_2 = 20$ $n_3 = 25$

Se pide calcular el coeficiente de variación, la desviación media y la desviación mediana.

45. Con la siguiente distribución calcule:

$y'_{i-1} - y'_i$	N_i
10,1 – 16	4
16,1 – 22	12
22,1 – 28	25
28,1 – 34	35
34,1 – 40	44
40,1 – 46	50
Σ	–
$X_{i-1} - X_i$	F_i

a) La desviación típica o estándar.
b) La desviación media.
c) La desviación mediana.
d) Coeficiente de variación.

46. En la siguiente distribución de resistencia a la tensión (en kgrs./mm^2) de láminas de acero. Obtenga la desviación mediana:

RESISTENCIA:	10 a 20	20 a 30	30 a 40	40 a 50
% DE LÁMINAS:	0,20	?	?	0,12

Nota: se sabe que la resistencia promedio es de 29,4 kgrs./mm^2.

47. La siguiente distribución de frecuencias se refiere al número de compradores de un determinado producto, en 30 barrios de clase media en Bogotá.

No. DE COMPRADORES	No. DE BARRIOS
130	3
148	6
160	5
220	3
280	2
320	4
400	7

a) Calcule la desviación mediana.
b) Calcule el coeficiente de variación mediana.
c) Si la empresa considera que el número de compradores por barrio es el doble. ¿Cuál será la nueva varianza y el coeficiente de variación?

48. Con los siguientes datos se pide:

1)

$y_{i-1} - y_i$	n_i
8,1 – 16	3
16,1 – 24	6
24,1 – 32	10
32,1 – 40	15
40,1 – 48	4
48,1 – 56	2
Σ	40
$X_{i-1} - X_i$	f_i

2) x_i: 8 10 7 30 38 25 47 70 .

b) Si se considera (1) y (2) como submuestras, ¿cuál será el coeficiente de variación para el conjunto o sea cuando $n = 48 = 40 + 8$.
c) Con la distribución (1) determine la desviación media.
c) Con la distribución (2) determine la desviación mediana.
d) Compruebe con la distribución (a) que

$$De \leq Da < S$$

49. Con los siguientes datos: 2 5 8 12 y 20. Calcular la desviación típica, desviación media y mediana. Observar que $S > Da \geq De$

50. Calcular el coeficiente de la desviación media a los ejercicios: 47 48 y 49

51. Calcular el coeficiente de la desviación mediana a los ejercicios: 47 48 y 49

RECORRIDO INTERCUARTÍLICO, DESVIACIÓN CUARTIL Y COEFICIENTE DE DESVIACIÓN CUARTIL

Se considera una serie de medidas de dispersión cuando se ha trabajado con *cuartiles, deciles* o *percentiles.* Entre los más importantes se tienen:

El *recorrido intercuartílico*, se define como la diferencia entre el tercer cuartil y el primero.

$$Q_R = Q_3 - Q_1$$

El *recorrido interdecil*, corresponde a la diferencia entre el noveno y el primer decil.

$$D_R = D_9 - D_1$$

La *desviación cuartil o semi-recorrido intercuartílico*, se obtiene, mediante el cálculo del recorrido intercuartílico divido entre 2.

$$Q_D = \frac{Q_3 - Q_1}{2}$$

Este estadígrafo, en una distribución normal, equivale a 0,6745 veces la desviación estándar.

El *coeficiente de desviación cuartil*, es aplicado, especialmente, en aquellas distribuciones cuyos valores extremos no se encuentran definidos o cuando se desea obtener una rápida estimación de la dispersión, no necesaria para posteriores cálculos.

$$C_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

El *recorrido interdecil* se calcula estableciendo la diferencia entre el noveno y el primer decil, lo mismo, para obtener el recorrido interpercentil será la diferencia entre el percentil 99 y el percentil uno.

$$D_D = D_3 - D_1 \qquad P_D = P_{99} - P_1$$

Ejemplo 1. Con la siguiente distribución calcular el recorrido intercuartílico, la desviación cuartil y el coeficiente de desviación cuartil.

o **:**

$y'_{i-1} - y'_i$	n_i	N_i	
30,1 – 38	6	6	
38,1 – 46	14	20	$\leftarrow N_{j-1}$
46,1 – 54	$n_j \rightarrow 36$	56	$\leftarrow N_j$
54,1 – 62	50	106	
62,1 – 70	43	149	$\leftarrow N_{j-1}$
70,1 – 78	$n_j \rightarrow 32$	181	$\leftarrow N_j$
78,1 – 86	18	199	
86,1 – 94	9	208	
$\sum$	208	-	
$X'_{i-1} - X'_i$	f_i	F_i	

Primer cuartil

$\frac{n}{4} = \frac{208}{4} = 52$ no está, en la columna de N_i, por lo tanto se toma a $N_{i-1} = 20$ y $N_i = 56$.

Siendo que $N_{J-1} < \frac{n}{4}$ se tendrá:

$$Q_1 = y'_{J-1} + C\left[\frac{\frac{n}{4} - N_{J-1}}{n_J}\right] \qquad Q_1 = 46 + 8\left[\frac{52 - 20}{36}\right] = 53{,}11$$

Tercer cuartil

$\frac{3n}{4} = \frac{3(208)}{4} = 156$ como no se encuentra en la columna de las frecuencias absolutas acumuladas, se toma a $N_{j-1} = 149 < 156$ y $N_j = 181 > 156$. Ahora, cuando se presenta $N_{j-1} < \frac{3n}{4}$ se tendrá:

$$Q_3 = y'_{j-1} + C\left[\frac{\frac{3n}{4} - N_{J-1}}{n_j}\right]$$ Reemplazando $Q_3 = 70 + 8\left[\frac{156-149}{32}\right] = 71{,}75$

Recorrido intercuartílico: $Q_D = Q_3 - Q_1$ $Q_D = 71{,}75 - 53{,}11 = 18{,}64$

Desviación cuartil será igual a: $Q_{D_2} = \frac{Q_3 - Q_1}{2}$ $Q_{D_2} = \frac{71{,}75 - 53{,}11}{2} = 9{,}32$

Coeficiente de desviación cuartil será igual a:

$$C_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$ $C_Q = \frac{71{,}75 - 53{,}11}{71{,}75 + 53{,}11} = \frac{18{,}64}{124{,}86} = 0{,}1439$ *ó* $14{,}39\%$

Ejemplo 2. Con la misma distribución anterior calculemos el recorrido interdecil, la desviación interdecil y el coeficiente de desviación decil.

Primer decil:

$\frac{n}{10} = \frac{208}{10} = 20{,}8$ Este valor no se encuentra en la columna de las frecuencias absolutas acumuladas, por lo tanto $N_{j-1} = 20 < 20{,}8$ y $N_j = 56 > 20{,}8$, la fórmula de cálculo que debe ser aplicada será:

$$D_1 = y'_{j-1} + C\left[\frac{\frac{n}{10} - N_{J-1}}{n_j}\right]$$

$$D_1 = L_i + i\left[\frac{\frac{n}{10} - F_{J-1}}{f_J}\right]$$

Reemplazando en la fórmula anterior se tendrá: $D_1 = 46 + 8\left[\frac{20{,}8 - 20}{36}\right] = 46{,}18$

Noveno decil:

$\frac{9n}{10} = \frac{9(208)}{10} = 187{,}2$ Siendo que $N_{J-1} = 181 < 187{,}2$ y $N_J = 199 > 187{,}2$ se tendrá que:

$$D_9 = y'_{j-1} + C\left[\frac{\frac{9n}{10} - N_{J-1}}{n_j}\right]$$ $D_9 = 78 + 8\left[\frac{187{,}2 - 181}{18}\right] = 80{,}76$

$$D_9 = L_i + i\left[\frac{\frac{9n}{10} - F_{J-1}}{f_J}\right]$$

El *recorrido interdecil*: $D_R = D_9 - D_1$ $D_R = 80{,}76 - 46{,}18 = 34{,}58$

La *desviación interdecil*: $D_{D_2} = \dfrac{D_9 - D_1}{2}$ $\quad D_{D_2} = \dfrac{80{,}76 - 46{,}18}{2} = 17{,}29$

El *coeficiente de desviación decil*:

$$C_D = \frac{D_9 - D_1}{D_9 + D_1} \qquad C_D = \frac{80{,}76 - 46{,}18}{80{,}76 + 46{,}18} = \frac{17{,}29}{126{,}94} = 13{,}62\% \qquad 13{,}62\%$$

RECORRIDO U OSCILACIÓN

El recorrido, denominado también como oscilación o fluctuación, es un estadígrafo fácil de calcular y su significado muy sencillo. Se calcula tomando la diferencia entre los valores extremos, sin tener en cuenta la repetición o frecuencia de los valores que toma la variable, de ahí su gran inestabilidad, especialmente cuando se añaden o suprimen cifras extremas.

Recorrido = x_{max} x_{min} (para datos no agrupados)

En datos agrupados, para variable discreta el recorrido se obtendrá así: recorrido = $y_m - y_l$.

Para la variable continua se calcula aplicando recorrido = $y'_m - y'_o$

Ejemplo 1. Supongamos que se tienen 5 observaciones: (datos no agrupados)

$x_1 = 8$ $\quad x_2 = 4$ $\quad x_3 = 6$ $\quad x_4 = 6$ $\quad x_5 = 12$

Recorrido = $x_{max} - x_{min}$ $\qquad$ Rango = $12 - 4 = 8$

Ejemplo 2. Tomemos una distribución de frecuencias para una *variable discreta*

y_i	n_i
2	3
4	5
6	6
8	4
10	2
Σ	20
X_i	f_i

Recorrido = $y_m - y_1$

Recorrido = 10 – 2 = 8

Recorrido = $X_m - X_1$

Ejemplo 3. En el caso de una *variable continua* el procedimiento es casi igual al anterior.

$y'_{i-1} - y'_i$	n_i
3,1 – 7	6
7,1 – 11	4
11,1 – 15	12
15,1 – 20	8
20,1 – 23	5
Σ	35
$X'_{i-1} - X'_i$	f_i

Recorrido = $y'_m - y'_0$

Recorrido = 23 – 3 = 20

Recorrido = $X_m - X_0$

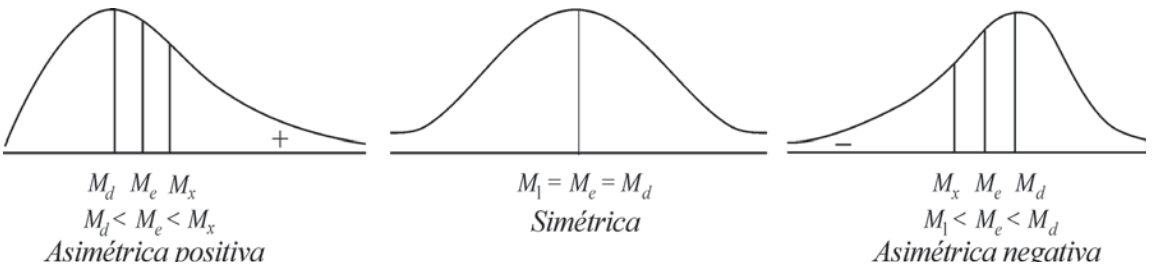
MEDIDAS DE ASIMETRÍA O DE DEFORMACIÓN

Fueron propuestas, algunas de ellas, por Karl Pearson. Una distribución simétrica no tiene riesgo; recordemos que en este caso $M_x = M_e = M_d$, estas medidas son iguales, por lo tanto consideramos que la distribución tiene la forma de una campana, denominada de Gauss o normal, ya que el promedio se ubica en todo el centro de ella.

Si las frecuencias son considerablemente altas, la distribución deja de ser *simétrica* y hablamos de distribuciones *asimétricas positivas* o *negativas* y las tres medidas (media, moda y mediana) no tienen igual valor.

En las gráficas que se presentan a continuación se observa que en la parte más alta de la distribución se ubica el Modo.

Si $M_l > M_e > M_d$ se dice que la distribución es *asimétrica negativa*, ya que la curva presenta un alargamiento hacia la derecha. Si por el contrario $M_x = M_e = M_d$, el alargamiento es hacia la izquierda y se dice que es asimétrica positiva.



Las fórmulas utilizadas para calcular el grado de asimetría son:

1) $A_S = \dfrac{M_1 - M_d}{S}$ 2) $A_S = \dfrac{3\,(M_1 - M_e)}{S}$ 3) $A_S = \dfrac{m_3}{S^3}$

En la última se utiliza el *momento de orden 3 con respecto a la media*

$$m_3 = \frac{\Sigma d_i^3 f_i}{n} \qquad m_3 = \frac{\Sigma (y_i - \bar{y})^3 n_i}{n}$$

El grado de asimetría, también se puede calcular aplicando la fórmula establecida por Bowley, siendo: $A_s = \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 + Q_1}$, para ello es necesario, trabajar con *cuartiles*.

Si $A_s = 0$ Simétrico Si $A_s > 0$ Asimétrico positivo Si $A_s < 0$ Asimétrico negativo, ya que el signo nos indicará hacia que lado se presenta la deformación o alargamiento de la distribución y el valor será el grado de asimetría, entre más grande sea este valor, más grande será la asimetría.

Ejemplo 1. Con los siguientes datos de una distribución de frecuencias, se puede calcular la asimetría y su grado, por las fórmulas anteriormente establecidas.

y_i	n_i	$y_i n_i$	N_i	$y_i^2 n_i$	$(y_i - \bar{y})$	$(y_i - \bar{y})^3$	$(y_i - \bar{y})^3 n_i$
2	4	8	4	16	–3,3	–35,94	–143,76
4	6	24	10 $\rightarrow N_{J-1}$	96	–1,3	–2,20	–13,20
6	5	30	15 $\rightarrow N_J$	180	0,7	0,34	1,70
8	3	24	18	192	2,7	19,68	59,04
10	2	20	20	200	4,7	103,82	207,64
Σ	20	106	–	684	–	–	111,42
X_i	f_i	$X_i f_i$	F_i	$X_i^2 f_i$	d_i	d_i^3	$d_i^3 f_i$

$$\bar{y} = M_1 = \frac{106}{20} = 5{,}3 \qquad \boxed{\bar{X} = 5{,}3} \qquad M_d = y_J = 4 \text{ (modo)}$$

$$M_e = \frac{y_{J-1} + y_J}{2} \qquad \boxed{M_e = \frac{X_{J-1} + X_J}{2}} \qquad M_e = \frac{4+6}{2} = 5$$

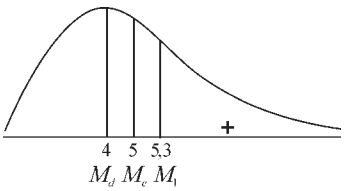
a) $$A_S = \frac{M_1 - M_d}{S} \qquad A_S = \frac{5{,}3 - 4}{2{,}47} = \frac{1{,}3}{2{,}47} = 0{,}53 \qquad \boxed{S^2 = \frac{\Sigma X_i^2 f_i - n\bar{X}^2}{n}}$$

$$S^2 = \frac{\Sigma y_i^2 n_i}{n} - \bar{y}^2 = \frac{684}{20} - 5{,}3^2 = 34{,}2 - 28{,}09 = 6{,}11\,; \qquad S = \sqrt{6{,}11} = 2{,}47$$

b) $$A_S = \frac{3(M_1 - M_e)}{S} \qquad A_S = \frac{3(5{,}3 - 5)}{2{,}47} = \frac{3(0{,}3)}{2{,}47} = \frac{0{,}9}{2{,}47} = 0{,}36$$

c) $$A_S = \frac{m_3}{S^3} = \frac{5{,}57}{15{,}07} = 0{,}37 \qquad m_3 = \frac{\Sigma(y_i - \bar{y})^3 n_i}{n} = \frac{111{,}42}{20} = 5{,}57 \qquad \boxed{m_3 = \frac{\Sigma d_i^3 f_i}{N}}$$

Siendo: $M_1 > M_e > M_d$
$5{,}3 > 5 > 4$



La distribución es asimétrica positiva

La *asimetria*, también es calculada mediante la aplicación de la fórmula de Bowley, donde

$$A_S = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 + Q_1}$$

Veamos cómo se calcula con los datos de la distribución anterior:

$Q_1 = y_5 = 4 \quad \rightarrow \quad$ siendo: $\frac{20}{4} = 5 \qquad N_{J-1} = 4$ y $N_J = 10$

$Q_3 = \frac{6+8}{2} = 7 \quad \rightarrow \quad$ siendo: $\frac{3(20)}{4} = 15 \qquad N_{J-1} = 15$ y $N_J = 18$

$Q_2 =$ mediana $= 5 \quad \rightarrow \quad A_S = \frac{(7+4) - 2(5)}{(7-4)} = \frac{1}{3} = 0{,}33$

Nota: la aplicación de las anteriores medidas es preferible hacerla con la variable continua, usando los intervalos de clase.

Nota: Los resultados obtenidos en la aplicación de las fórmulas anteriores, arrojan resultados diferentes sin que esto tenga importancia alguna.

Ejemplo 2. Determinar si es grande o pequeño el grado de asimetría, en una distribución cuyos estadígrafos de posición son:

$\bar{x} = 189{,}87 \qquad M_e = 189{,}16 \qquad M_d = 187{,}60$

o :

Aplicando la fórmula empírica: $\bar{x} - M_d = 3\,(\bar{x} - M_e)$

$189{,}87 - 187{,}60 = 2{,}27 \neq 2{,}13$

$3\,(189{,}87) - 189{,}16 = 2{,}03$ La diferencia nos indica que existe una pequeña asimetría.

Ejemplo 3. Tomando una distribución ligeramente asimétrica. Calcular el modo sabiendo que su media es igual a 3 y que la diferencia entre la media y mediana es igual a –2

o :

$\bar{x} - M_d = 3\,(\bar{x} - M_e) \qquad 3 - M_d = -6$

$3 - M_d = 3\,(-2) \qquad M_d = \bar{x} - 3\,(\bar{x} - M_e) \qquad M_d = 9$

Ejemplo 4. Calcular el coeficiente de asimetría en la siguiente distribución, utilizando la fórmula de Pearson.

y_i :	1	2	3	4	5	6
n_i :	2	8	3	5	7	5

o :

y_i	n_i	$y_i\,n_i$	$x_i^2\,n_i$
1	2	2	2
2	8	16	32
3	3	9	27
4	5	20	80
5	7	35	175
6	5	30	180
Σ	30	112	496
X_i	f_i	$X_i\,f_i$	$X_i^2\,f_i$

$\bar{X} = \dfrac{\Sigma X_i\,f_i}{N}$ $\qquad S^2 = \dfrac{\Sigma X_i^2\,f_i}{N} - \bar{X}^2$

$\bar{y} = \dfrac{\Sigma y_i\,n_i}{n}$ $\qquad \bar{y} = \dfrac{112}{30} = 3{,}73 \qquad M_d = 2$

$S^2 = \dfrac{\Sigma y_i^2\,n_i}{n} - \bar{y}^2$ $\qquad S^2 = \dfrac{496}{30} - 3{,}73^2 = 2{,}62 \Rightarrow S = 1{,}62$

$A_S = \dfrac{\bar{y} - M_d}{S}$ $\qquad A_S = \dfrac{3{,}73 - 2}{1{,}62} = 1{,}07$ $\qquad A_S = \dfrac{\bar{x} - M_d}{S}$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

52. Conteste si es cierto o falso y por qué:

a) En toda serie de datos $\dfrac{\Sigma x_i^2}{n}$ es mayor o igual a $\left(\dfrac{\Sigma x_i}{n}\right)^2$

b) Si para dos distribuciones $\bar{x}_1 = \bar{x}_2$ y $S_1^2 > S_2^2$ entonces $CV_1 < CV_2$

c) Si el coeficiente de asimetría es negativo, la curva es achatada.
d) Si las frecuencias de una distribución se multiplican por 8, la varianza queda multiplicada por 64.
e) La varianza está expresada en las mismas unidades de la variable.

53. Conteste los siguientes puntos diciendo si es cierto o falso y por qué:
a) El coeficiente de variación no puede ir expresado en unidades de la variable.
b) El coeficiente de variación, es la desviación estándar expresada en términos de porcentajes de $\bar{x}$.
c) Si una distribución es simétrica o moderadamente asimétrica, entre $\overline{x} \pm 2S$ está el 95,5% de las observaciones.
d) La desviación típica no se mide en las mismas unidades de la variable.

54. Con con la información sobre el número de retardos anuales, de los 500 empleados de una empresa calcular el coeficiente de asimetría de Pearson.
$\overline{x} = 9.725 \qquad M_e = 9.672 \qquad S = 1.217{,}50$

55. Con los siguientes datos determine la simetría o asimetría de la distribución aplicando las diferentes fórmulas vistas.

a)

y_i	n_i
5	3
7	39
9	10
11	8
13	7
15	3
Σ	70
X_i	f_i

b)

y_i	n_i
5	3
7	7
9	8
11	9
13	30
15	3
Σ	60
X_i	f_i

c)

y_i	n_i
5	5
7	10
9	20
11	20
13	10
15	5
Σ	70
X_i	f_i

Haga los respectivos histogramas y dibuje los polígonos de frecuencias.

MEDIDAS DE APUNTAMIENTO : A_p

urtosis o estadígrafos de apuntamiento. Una característica importante de la variación en algunas distribuciones, es su grado de agudeza en la cima de la curva que las representa. Esta agudeza que se observa en la región del *Modo*, comparada con las condiciones halladas para el mismo sitio en la curva normal, es lo que se llama *urtosis o curtosis*.

Si la curva es más *plana* que la *normal*, la distribución se le considera *achatada* o *planticúrtica* y si es más aguda se le denomina *apuntada* o *leptocúrtica*. Si la curva es *normal* se dice que es *mesocúrtica.*

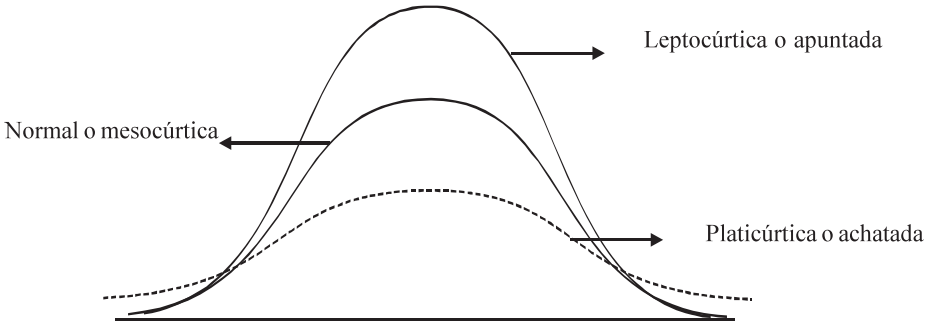
La *curtosis* es una medida de altura de la curva y por tanto está representada por el cuarto momento de la media. En la misma forma que para la asimetría, su cálculo se efectúa en función de la desviación típica y de los momentos unidimensionales de orden cuatro con respecto a la media aritmética:

$$A_p = \frac{m_4}{S^4}$$ Siendo 3 el valor de $\frac{m_4}{S^4}$ en la curva normal,

$$m_4 = \frac{\Sigma(y_i - \bar{y})^4 n_i}{n}$$

$$m_4 = \frac{\Sigma(X_i - \bar{X})^4 f_i}{N} = \frac{\Sigma d_i^4 f_i}{N}$$

Si $A_p = 3$ → la distribución es normal o mesocúrtica
$A_p > 3$ → la curva es leptocúrtica o apuntada
$A_p < 3$ → la curva es platicúrtica o achatada



Ejemplo 1. Calcular el grado de apuntamiento en la siguiente distribución:

y_i	n_i	$(y_i - \bar{y}_i)^3$	$(y_i - \bar{y}_i)^4 n_i$
2	4	–35,94	474,41
4	6	– 2,20	17,16
6	5	0,34	1,19
8	3	19,68	159,41
10	2	103,82	975,91
Σ	20	—	1.628,08
X_i	f_i	d_i^3	$d_i^4 f_i$

$$m_4 = \frac{\Sigma (y_i - \bar{y}_i)^4 n_i}{n} = \frac{1.628,08}{20} = 81,404$$

$$S^2 = \frac{\Sigma (y_i - \bar{y}_i)^2 n_i}{n} = 6,11; \quad S^4 = 37,33$$

$$m_4 = \frac{\Sigma d_i^4 f_i}{N} \qquad S^2 = \frac{\Sigma d_i^2 f_i}{N}$$

$$A_p = \frac{m_4}{S^4} = \frac{81,404}{37,33} = 2,18$$

2,18 < 3 la curva es achatada o platicúrtica.

Ejemplo 2. Con los siguientes datos:

$y'_{i-1} - y'_i$	n_i
10,1 – 16	4
16,1 – 22	8
22,1 – 28	13
28,1 – 34	10
34,1 – 40	9
40,1 – 46	6
Σ	50
$X'_{i-1} - X'_i$	f_i

Se pide calcular:

a) Los coeficientes de asimetría.
b) Los coeficientes de apuntamiento.

o :

a) $A_s = \dfrac{3(\bar{y} - M_e)}{s}$ $\quad A = \dfrac{3(28{,}6 - 28{,}1)}{8{,}736} = 0{,}1717$ (asimetría positiva $A_s > 0$)

$A_s = \dfrac{3(\bar{X} - M_e)}{s}$

b) $A_s = \dfrac{m_4}{s^4}$ $\quad m_4 = \dfrac{\Sigma(y_i - \bar{y})^4 n_i}{n}$ $\quad m_4 = \dfrac{\Sigma d_i^4 f_i}{n}$

y_i	n_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^3$	$(y_i - \bar{y})^4$	$(y_i - \bar{y})^4 n_i$
13	4	−15,6	243,36	−3.796,41	59.223,99	236.895,96
19	8	−9,6	92,16	−884,74	8.493,50	67.948,00
25	13	−3,6	12,96	− 46,66	167,98	2.183,74
31	10	2,4	5,76	13,82	33,17	331,70
37	9	8,4	70,56	592,70	4.978,68	44.808,12
43	6	14,4	207,36	2.985,98	42.998,11	257.988,66
Σ	50	—	—	—	—	610.156,18
X_i	f_i	d_i	d_i^2	d_i^3	d_i^4	$d_i^4 f_i$

$Z_i = y_i - \bar{y}$ $\quad d_i = X_i - X_i$ $\quad \bar{y} = \dfrac{1.430}{50} = 28{,}6 = \bar{X}$

$m_4 = \dfrac{\Sigma(y_i - \bar{y})^4 n_i}{n}$ $\quad m_4 = \dfrac{610.156{,}18}{50} = 12.203{,}12$ $\quad s^2 = 76{,}32$ $\quad s^4 = 5.824{,}74$

$A_p = \dfrac{m_4}{s^4}$ $\quad A_p = \dfrac{12.203{,}12}{5.824{,}74} = 2{,}09$ $\quad A_p < 3 \rightarrow 2{,}09 < 3$ achatada ó platicúrtica

EJERCICIOS MISCELÁNEOS

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

56 Con los siguientes datos:

28	42	64	32	68	25	34	40	51	62
36	56	76	40	70	28	38	46	58	64

Se pide:

a) Trabajando con los datos originales calcular la desviación típica, desviación mediana y el coeficiente de variación.

b) Agrupar los datos en una tabla de frecuencias, teniendo en cuenta que la amplitud es constante (la variable es continua) y $m = 6$, calcular el coeficiente de variación.

57. Suponga que se tiene otra distribución que presenta las siguientes medidas: $\bar{y} = 38{,}5$ y $S^2 = 36$ Comparando con los resultados del ejercicio anterior (No. 56), determinar:

a) ¿Cuál de las dos distribuciones presenta una menor variabilidad absoluta?
b) ¿Cuál de ellas presenta una mayor variabilidad relativa?
c) Si consideramos que la primera distribución presenta un valor de x = 48 y en la segunda (o sea la del punto c) el valor de x = 50, ¿cuál de estos valores presenta una mejor posición relativa?

58. a) Calcular el grado y la dirección de la asimetría con los datos de la tabla de frecuencias del punto 56 (b)
b) Calcular el grado de apuntamiento en esa distribución, determinando si es normal, achatada o apuntada.

59. Se tiene una distribución de 6 intervalos de amplitud constante, de la cual se sabe que:

$$n = 150 \quad n_3 = n_4 = 30 \quad n_2 = n_5 = n_1 + 5 \quad n_1 = n_6 \quad y_5 = 50 \quad \bar{y} = 41$$
$$N = 150 \quad f_3 = f_4 = 30 \quad f_2 = f_5 = f_1 + 5 \quad f_1 = f_6 \quad X_5 = 50 \quad \bar{X} = 41$$

a) Calcular el coeficiente de variación
b) Calcular el coeficiente de asimetría aplicando las diferentes fórmulas expuestas c) Calcular el coeficiente de apuntamiento.

60. Con los siguientes datos, se pide calcular:

$y'_{i-1} - y'_i$	n_i
6,1 – 12	3
12,1 – 15	10
15,1 – 20	6
20,1 – 28	4
28,1 – 36	8
36,1 – 40	16
40,1 – 50	3
Σ	50

a) Varianza
b) Desviación típica
c) Coeficiente de variación
d) Desviación media
e) Desviación mediana

61. Con los datos del punto 51 calcular: a) el grado de asimetría; b) grado de apuntamiento.

62. Se tiene dos distribuciones:
l) Corresponde a datos agrupados

$y'_{i-1} - y'_i$	n_i
2,1 – 6	22
6,1 – 10	14
10,1 – 14	10
14,1 – 18	8
18,1 – 22	4
22,1 – 26	2
$X'_{i-1} - X'_i$	F_i

2) Corresponde a datos sin agrupar

$x_1 = 12$	$x_3 = 6$	$x_5 = 23$	$x_7 = 4$	$x_9 = 6$
$x_2 = 18$	$x_4 = 20$	$x_6 = 17$	$x_8 = 10$	$x_{10} = 4$

Se pide determinar:
a) En cuál distribución se observa una mayor variabilidad absoluta?
b) En cuál distribución se observa una menor variabilidad relativa?
c) Si consideramos un valor de 18, en cuál de las distribuciones se tiene una mayor posición relativa?

63. Se pide calcular el coeficiente de variación para el total de las 70 observaciones (A+B) del ejercicio 62.
    a) ¿Qué pasaría con este coeficiente de variación, si multiplicamos a cada una de las 70 observaciones por 5?
    b) ¿Qué pasaría con este coeficiente de variación, si le sumamos 10 a cada una de las 70 observaciones?
64. Considerando únicamente la distribución (A) para datos agrupados: (Ejercicio 61)
    a) ¿La distribución es simétrica o asimétrica? ¿cuál es su grado de asimetría (si fuera la respuesta afirmativa) a qué lado se presenta?
    b) Calcule el grado de apuntamiento.

65. Los salarios mensuales que paga una fábrica a los operarios que trabajan en dos turnos, tiene las siguientes estadísticas:

MEDIDAS	TURNO I	TURNO II
Número de trabajadores	38	62
Salario medio mensual	978.000	1.203.500
Varianza	851.600	962.600

    a) Comparar los salarios en los dos turnos en cuanto a variabilidad absoluta y relativa.
    b) Obtener el coeficiente de variación para el salario de los 100 operarios.
    c) Cual será el coeficiente de variación para el año entrante si el gerente ofrece un aumento de $22.000 para los del primer turno y un 7% para los del segundo turno. El coeficiente de variación a calcular, se refiere al total de los 100 operarios.

66. Con los siguientes datos, correspondientes a una distribución de frecuencias, se pide calcular el grado de simetría o asimetría, indicando hacia qué lado, lo mismo que el grado de apuntamiento. Además calcular:

$y'_{i-1} - y'_i$	n_i
2,1 – 6	3
6,1 – 10	12
10,1 – 14	25
14,1 – 18	11
18,1 – 22	7
22,1 – 26	2
Σ	60
$X'_{i-1} - X'_i$	F_i

    a) El coeficiente de variación
    b) El puntaje típico, si se tiene queun valor de $y = 12$
    c) La desviación media y mediana

67. A los trabajadores de una empresa, el próximo año, el salario mensual les será aumentado en un 4,2% más $8.000. La empresa tiene 600 trabajadores y actualmente devengan un salario medio mensual de $810.000, con un coeficiente de variación de 0,36.
    a) ¿Cuál será la varianza de los salarios el próximo año?
    b) El próximo año, ¿cuál será el valor total de la nómina mensual?

68. Con los siguientes datos no agrupados: x_i: 6 4 8 2 10, calcular:
    a) Desviación media
    b) Desviación mediana

c) Desviación típica
d) Comprobar que $D_e \leq D_a < S$
e) Calcular el coeficiente de variación

69. Con la siguiente información: $x = \frac{6}{8}$ [16 y +16] $CV_y = 0{,}40$ $\bar{y} = 5$
Hallar el coeficiente de variación de x

70. Sabiendo que $y = 6 - 10x$ $V_{[x]} = 8$. Hallar la varianza de $4x - y$

71. ¿Por qué es cero el coeficiente de asimetría para una distribución simétrica?

72. Dos obreros del mismo y trabajo muestran los siguientes resultados en un período determinado

MEDIDAS	OBREROS	
	A	B
Tiempo promedio para el desarrollo de su trabajo	42	35 minutos
Desviación típica	8	6 minutos

a) ¿Cuál grupo es el más regular en el desarrollo de su trabajo?
b) ¿Cuál grupo es el más rápido en terminar el trabajo?

73. Un estudiante obtuvo las siguientes notas en 4 de sus materias así:

Economía = 4 Matemática = 3 ,3 Inglés = 3 ,6 y Derecho = 4 ,2

a) ¿Qué clase de opinión podríamos dar con respecto a esas notas?

Recurrimos a los profesores que en un análisis han podido reunir más información sobre el rendimiento del curso donde se encuentra el estudiante.

Promedio del curso: Economía = 4,3 Matemática = 2,8 Inglés = 3,2 y Derecho = 4,6. Además, conocen las desviaciones típicas, siendo: 0,6 0,75 0,8 0,6 respectivamente.

b) ¿Su opinión se mantiene igual respecto al rendimiento de este estudiante?

74. Supongamos que un acucioso empleado de la empresa de Acueducto de la ciudad, realiza una muestra de 60 usuarios del servicio, sobre los reclamos en los 2 últimos años por esas personas, con los siguientes resultados:

NÚMERO DE RECLAMACIONES	NÚMERO DE USUARIOS
0	26
1	10
2	8
3	6
4	4
5	3
6	2
7	1

Se pide hallar:
a) El promedio de reclamos
b) La varianza y su desviación típica
c) El coeficiente de variación

Síntesis de capítulo

En el resumen del capítulo anterior, se decía que no basta con calcular el *promedio* para describir el comportamiento de un conjunto de datos, sino que además, en numerosos casos, requiere de la aplicación de alguna de las medidas de dispersión, las que nos determinan la extensión que ocupan los valores de la variable respecto al valor promedio elegido.

Supongamos que dos personas conocedoras de estadística se dedican a lanzar dardos y en el primer lanzamiento ambos lo colocan el tercer círculo uno arriba y el otro abajo, es de imaginar que ambos se sientan satisfechos del lanzamiento, pues lograron la misma distancia con respecto al centro. Se puede sacar como conclusión que el punto central fue para ellos menos importante y la atención se centró más en la distancia o dispersión que a dar en el blanco o punto central.

El cálculo de un promedio para dos grupos puede ser igual, como por ejemplo el promedio de calificación de dos cursos de estadística con diferentes profesores, en ambos se obtuvo 3,7, pero se observa que en el primero de ellos la calificación mínima fue de 1,1 y la máxima de 4,3, en cambio con el segundo profesor la mínima fue de 3,4 y la máxima de 4,2 ¿qué profesor le recomendaría a su hermano si hubiese necesidad de elegir uno de ellos? En este ejemplo se le da importancia a la dispersión; como se puede observar, no siempre el promedio por sí mismo nos proporciona una suficiente información.

Son varias las medidas o varios los procedimientos que nos indican el mayor o menor grado de dispersión, siendo la más conocida la *desviación estándar* o *desviación típica*. La varianza es el procedimiento que nos permite llegar a la desviación típica, con el inconveniente que no se expresa en la misma unidad de medida de la variable, ya que siempre corresponde al cuadrado de ella, así se tendrá: metros, kilos, horas al cuadrado, en cambio la desviación estándar y el recorrido, se expresan en las mismas unidades. Cuando el coeficiente de variación se expresa en porcentajes, permite realizar mejor la comparación de la dispersión entre dos o más distribuciones, cuando las variables están expresadas en unidades de medidas diferentes.

Decimos que una *distribución es simétrica*, cuando la *media*, *mediana* y la *moda* calculada en una variable son iguales. El grado de *simetría* o *asimetría* en una distribución se obtiene aplicando la medida denominada coeficiente de asimetría o simplemente observando el histograma de frecuencias. El problema se presenta cuando la distribución tiene forma de U, siendo considerada no normal, ya que no tiene la forma de campana de Gauss que siempre utilizamos para indicar una distribución simétrica.

Finalmente recordemos que los valores observados, si bien es cierto se repiten, en conjunto son diferentes si las comparamos entre sí, por lo tanto esta dispersión no puede pasar desapercibida y su información radica cuando tiende a agruparse alrededor del promedio, de ahí que no solo calculamos estos últimos, pero además, debemos acompañarlo con técnicas que nos permitan estudiar la variación, cuando encontramos diferentes distribuciones con diversos grados de dispersión. Vale la pena resaltar cómo en la fabricación de artículos de precisión, una pequeña variación puede influir en su mal funcionamiento.

5

CAPÍTULO

Nociones elementales de probabilidad

No a o a ta
t o a o
a o
o o

CONTENIDO

- Conceptos, usos y aplicaciones de las probabilidades.
- Probabilidad a priori, clásica, subjetiva y de frecuencia relativa.
- Espacio muestral, Diagrama de árbol, Técnicas de conteo, Combinaciones y Permutaciones.
- Sucesos mutuamente excluyentes, independientes y dependientes. Probabilidad condicional. Teorema de Bayes.
- Síntesis de la Unidad
- Ejercicios para Resolver, resueltos en el Sistema de Información en Línea SIL.

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Conocer, entender y manejar los conceptos básicos de probabilidad.
- Calcular probabilidades, aplicando las reglas de adición y multiplicación.
- Manejar y aplicar los conceptos de la Permutaciones y Variaciones.
- Entender y aplicar el Teorema de Bayes.
- Entender la importancia que ella tiene en la inferencia, para realizar aseveraciones sobre un entorno incierto o de incertidumbre.

GENERALIDADES

El concepto de *probabilidad* puede ser interpretado como algo indefinible, pero utilizado para expresar, de algún modo, un grado de creencia que uno tiene de la ocurrencia de un suceso; nos referimos a algo que puede suceder con base en la experiencia que se tenga. Con frecuencia observamos o escuchamos el estado del tiempo, o sea los pronósticos meteorológicos sobre la posibilidad de buen tiempo o la presencia de lluvias fuertes o ligeras, gran nubosidad, vientos fuertes o en calma, etc.; los hinchas de los diferentes equipos de fútbol discuten frecuentemente sobre la posibilidad de clasificación o de ganar el campeonato; algo similar ocurre con los que juegan lotería o apuestan en las carreras de caballos; el mismo alumno, cuando se refiere a la posibilidad que tiene de ganar o de perder una asignatura; todos ellos son pronósticos que hacemos con la esperanza de que sucedan.

El origen de las probabilidades en los juegos de azar se remonta al siglo XVII, cuando Antoine Gombauld, más conocido como el caballero de Meré, jugador profesional quien pensó haber descubierto una táctica infalible de jugar a los dados, con muy buenos resultados durante un determinado tiempo, después del cual comenzó a perder y por ende disminuyó su fortuna, situación que lo obligó a consultar a Blas Pascal y a Pierre de Fermat, iniciándose así poco a poco una ciencia bien fundamentada.

En la actualidad las probabilidades guardan una estrecha relación con la Teoría de Conjuntos, de gran importancia en el campo de la *Inferencia Estadística* debido a la incertidumbre que siempre se tiene en la toma de decisiones, permitiendo el análisis de los riesgos que se corren y la forma de minimizar el azar inherente. En estadística, el uso de las predicciones es de gran utilidad cuando se realizan investigaciones por muestreo, en la mayoría de los casos obligado por el costo y tiempo que con llevaría la realización de una investigación total, lo cual nos limita a un reducido número de elementos; y con base en esa información disponible, procedemos a la realización de predicciones o estimaciones, asignando límites de confianza a esos resultados.

Las probabilidades, conjuntamente con la estadística tienen infinidad de aplicaciones a problemas de economía y ciencias sociales, de la misma manera a las ciencias físicas, industria, comercio y gobierno, con la observación de que en cada uno de ellos tendrá sus requisitos particulares.

Se puede hablar de *Posibilidades* y de *Probabilidades,* el primero hace referencia a la comparación entre el número de resultados favorables con los desfavorables: $\frac{\text{N°. de defectuosos}}{\text{N°. de no defectuosos}}$ En el segundo es el cuociente entre el número favorable sobre el total de casos posibles: $P = \frac{\text{N°. de defectuosos}}{\text{Total de casos posibles}}$

Todos en esencia somos jugadores. En los negocios, en nuestra vida y siempre que tomamos una decisión, siempre va haber incertidumbre por la dificultad de predecir con exactitud: ¿Ganaré el parcial? ¿El semestre? ¿Me ganaré el Baloto si lo compro? ¿Si le hablo a esa persona me corresponderá? Todas estas preguntas y muchas más, tendrían en nuestra mente una posible respuesta ya que nos dejamos guiar por la experiencia y la intuición.

Definiciones. Es difícil dar una definición exacta de las probabilidades, sin embargo trataremos de obtener alguna que nos aproxime a ella.

M todo axiomático. El cual concibe la probabilidad de ocurrencia de un suceso, *como un número comprendido entre y 1*. Este concepto tiene que ver directamente con la noción de frecuencias relativas, donde $0 < h_i < 1$.

Supongamos que se lanza 100 veces una moneda, anotamos el número de veces que sale cara y las veces que sale sello; los resultados fueron los siguientes:

Frecuencia absoluta:	cara 56 veces	sello 44 veces
Frecuencia relativa:	56/100	44/100
Probabilidad:	p = 56% (éxito)	q = 44% (fracaso)

Se puede establecer una escala de valores comprendidos entre 0 y 1. La probabilidad igual a uno corresponde al límite superior, el cual se considera como la *certeza absoluta*; por ejemplo, la probabilidad de que algún día moriremos, esto es un hecho que tiene que cumplirse. Para el otro extremo igual a 0, consideraremos por ejemplo, "que se atravesará el Océano Atlántico a nado", lo cual es imposible, así que la probabilidad será 0. En este caso se habla de *imposibilidad absoluta*.

Si la probabilidad fuera 0 ó 1 no habría problema, pero existen una serie de fenómenos cuyos valores están comprendidos entre esos límites que dificultan un poco el cálculo.

Se llamará *suceso* a cada caso posible, es decir, a la realización de un acontecimiento y este puede ser:

- Se dirá que un *hecho es cierto*, cuando son favorables todos los casos posibles. Un ejemplo podría ser, el de comprar todos los billetes de un sorteo, por lo tanto será un hecho cierto que ganará el sorteo.
- Llamaremos un *hecho verosímil* a un suceso susceptible de realizarse, pero su probabilidad favorable es menor que la unidad y mayor que 0,5.
- Si la probabilidad es igual a 0,5 será un *hecho dudoso*, ya que las probabilidades ventajosas y desventajosas son iguales. Tal es el caso del lanzamiento de una moneda, en la aparición de cara o sello.
- *Hecho inverosímil*, se presenta cuando la probabilidad es menor que 0,5 y mayor que cero.
- *Hecho imposible*, es cuando no existe posibilidad alguna de salir favorecido; por ejemplo el individuo que no compra lotería, la probabilidad que tiene de ganar será cero.

M todo empírico o práctico. Considera la probabilidad de un suceso, como aquel número al cual se aproxima cada vez más la frecuencia relativa de la ocurrencia de un suceso, cuando las veces que se repite el experimento que origina ese suceso es lo bastante grande. Este concepto tiene algo que ver con el *experimento de Quetelet*, en donde la probabilidad de un suceso tiende a estabilizarse en un punto, cuando el número de experimentos se va haciendo cada vez más grande.

Probabilidad empírica, se determina mediante una serie de experimentos, en el caso de determinar la probabilidad de éxito de una operación quirúrgica practicada por determinado médico, será: número de casos favorables, dividido por el total de operaciones practicadas.

$$P = \frac{\text{Número de casos favorables}}{\text{Número de casos posibles}}$$

Si lanzamos una moneda 10 veces, es posible que 8 sean caras y dos sellos, pero aquí hablamos de una moneda teórica, perfectamente equilibrada caerá el mismo número de caras y sellos, en nuestro ejemplo 5 caras y 5 sellos. Con un dado teórico, se tendrá que la probabilidad de aparición de cada cara será 1/6. La probabilidad teórica se aplica a algo que no existe en la práctica, pues en la vida diaria veremos que cuanto mayor sea el número de lanzamientos de la moneda más nos acercaremos al ideal. El número de observaciones debe ser lo suficientemente grande, si se quiere una inferencia válida para ella.

M todo clásico. Considera la probabilidad como el cuociente de dividir al número de casos favorables o éxitos que pueden ocurrir en una prueba, por el total de casos posibles y que anteriormente lo consolidamos como empírico.

$$P = \frac{\text{Número de éxitos}}{\text{Total de casos posibles}}$$

Probabilidad a priori. Es aquélla que se puede determinar de antemano, sin necesidad de realizar el experimento. Ejemplo: la probabilidad de que aparezca cara en el lanzamiento de una moneda, se sabe de antemano que es igual a 0,5.

Probabilidad objetiva y subjetiva. Las probabilidades obtenidas mediante el método empírico y método clásico, comúnmente se les denomina como *probabilidades objetivas*, ya que se toma de la experiencia, es decir, de las repeticiones del hecho. Sin embargo vale la pena aclarar que también existen las llamadas *probabilidades subjetivas* correspondientes a una evaluación muy personal de la ocurrencia del suceso, por ejemplo ¿Perderá la selección del fútbol en el próximo partido? ¿Sacaré más de 4 en el próximo examen? etc. En ellos la persona fijará la probabilidad correspondiente dependiendo de su criterio personal.

Probabilidad con base en las frecuencias relativas. Es otro concepto probabilístico, que consiste en una observación que nos determina en qué momento ocurrieron eventos semejantes en el pasado, que permitan establecer la probabilidad de que vuelva a ocurrir en el futuro.

$$\textit{Probabilidad de ocurrencia} = \frac{\text{Número de veces de ocurrencia del evento en el pasado}}{\text{Número total de observaciones}}$$

Prueba es la realización de un acto. El conjunto de pruebas realizadas en las mismas condiciones se denomina *experimento*. La respuesta de una prueba se llama resultado, *punto muestral* o *suceso*. El conjunto de todos los resultados posibles constituye un *espacio muestral*. Un *evento* es un conjunto de uno o más puntos muestrales, por ejemplo, al lanzar un dado, se tendrá, que 1, 2, 3, 4, 5, 6 son 6 elementos; si se pide la aparición de un número impar el evento estará constituido por: 1 3 5.

ELABORACIÓN DE ESPACIOS MUESTRALES

Hemos visto que los resultados posibles de un experimento se denominan *puntos muestrales* y una lista de todos los resultados posibles de un experimento se denomina *espacio muestral*. Veamos varios ejemplos.

Ejemplo 1. *Experimento*: elegir un alumno del curso de estadística en la facultad.

o ***:***
(conjunto) ⇒ U { Acero, Alvarez, Cuellar... Serrano}

Ejemplo 2. *Experimento*: lanzamiento de una moneda teórica

o ***:*** U = {C S}
$\frac{1}{2}$ $\frac{1}{2}$

Ejemplo 3. *Experimento*: lanzamiento de dos monedas, asignando la probabilidad a cada suceso.

o ***:*** U = {CC CS SC SS}
¼ ¼ ¼ ¼

Ejemplo 4 *Experimento*: lanzamiento de tres monedas

o ***:*** U = {CCC CCS CSC SCC CSS SCS SSC SSS}
$\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$ $\frac{1}{8}$

Ejemplo 5. *Experimento*: lanzamiento de 4 monedas

o ***:***

$$U = \left\{ \begin{array}{ccccc} CCCC & CCCS & CCSS & CSSS & SSSS \\ & CCSC & CSCS & SCSS & \\ & CSCC & CSSC & SSCS & \\ & SCCC & SCSC & SSSC & \\ & & SSCC & & \\ & & SCCS & & \\ 1/16 & 4/16 & 6/16 & 4/16 & 1/16 \end{array} \right\}$$

Observemos en el ejemplo 5, cómo se ordenan, cambiando la posición de las apariciones de cara (C) y sello (S). La probabilidad de obtener exactamente tres caras, vemos que hay 4 resultados {CCCS, CCSC, CSCC, SCCC} sobre un total de 16 sucesos, por lo tanto su resultado será de 4/16 = ¼.

El número de casos posibles será fácil de determinar, si consideramos el lanzamiento de una moneda, se obtienen dos sucesos y tendrá como exponente el número de lanzamientos, así: 2^n. Observemos en los ejemplos anteriores: 2 3 4 y 5 el número de sucesos posibles:

2. 2^1 = 2 sucesos posibles
3. 2^2 = 4 sucesos posibles
4. 2^3 = 8 sucesos posibles
5. 2^5 = 16 y asi sucesivamente.

Ejemplo 6. *Experimento*: lanzamiento de un dado teórico
(conjunto) U = { 1 2 3 4 5 6} $6^n \rightarrow 6^1 = 6$ sucesos posibles
1/6 1/6 1/6 1/6 1/6 1/6

Ejemplo 7. *Experimento*: lanzamiento de dos dados

o **:**

$$U=\begin{Bmatrix} 11 & 12 & 13 & 14 & 15 & 16 \\ 21 & 22 & 23 & 24 & 25 & 26 \\ 31 & 32 & 33 & 34 & 35 & 36 \\ 41 & 42 & 43 & 44 & 45 & 46 \\ 51 & 52 & 53 & 54 & 55 & 56 \\ 61 & 62 & 63 & 64 & 65 & 66 \end{Bmatrix}$$

$6^2 = 36$ sucesos posibles
La probabilidad de presentación de cada suceso = 1/36

$\rightarrow \frac{1}{36}$ = parición de (66)

Ejemplo 8. *Experimento*: lanzamiento de tres dados

o **:**

$6^3 = 216$ sucesos posibles; la probabilidad de cada suceso = 1 / 216

Ejemplo . *Experimento*: extraer una carta de una baraja de 40 cartas

o **:**

Copas	⇒ AS	2	3	4	5	6	7	Zota	Caballo	Rey
Oros	⇒ AS	2	3	4	5	6	7	Zota	Caballo	Rey
Espadas	⇒ AS	2	3	4	5	6	7	Zota	Caballo	Rey
Bastos	⇒ AS	2	3	4	5	6	7	Zota	Caballo	Rey suceso 1/40

Ejemplo 10. *Experimento*: extraer una carta de la baraja española (52 cartas)

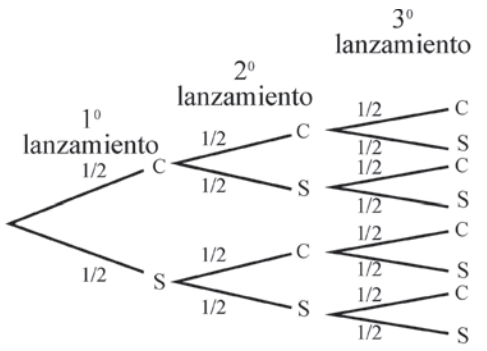
o

Diamantes	⇒ AS	2	3	4	5	6	7	8	9	10	J	Q	K	
Corazón	⇒ AS	2	3	4	5	6	7	8	9	10	J	Q	K	
Trébol	⇒ AS	2	3	4	5	6	7	8	9	10	J	Q	K	
Picas	⇒ AS	2	3	4	5	6	7	8	9	10	J	Q	K	→ 1/52

Diagrama del árbol. Es una de las maneras que permite determinar diversos eventos posibles, al contar los puntos muestrales. Veamos dos ejemplos.

Ejemplo 11. En el lanzamiento de tres monedas, se tendrá:

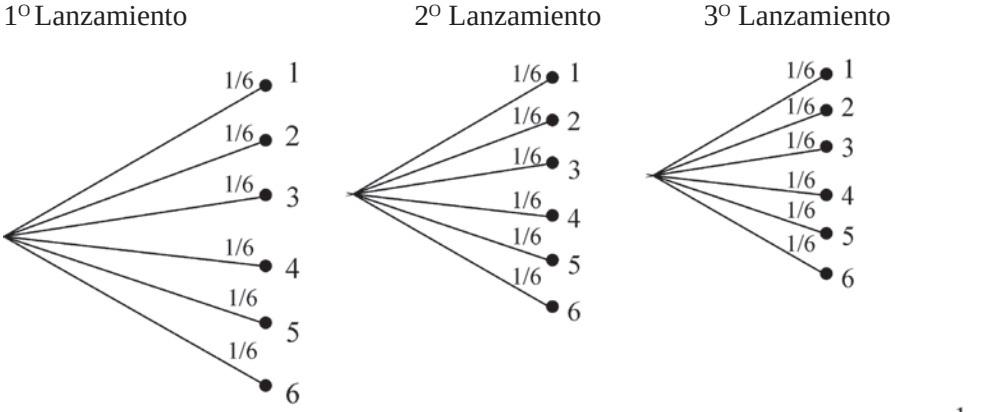
o **:**



CCC ⇒ (1/2)(1/2)(1/2) = 1/8
CCS ⇒ (1/2)(1/2)(1/2) = 1/8
CSC ⇒ (1/2)(1/2)(1/2) = 1/8
CSS ⇒ (1/2)(1/2)(1/2) = 1/8
SCC ⇒ (1/2)(1/2)(1/2) = 1/8
SCS ⇒ (1/2)(1/2)(1/2) = 1/8
SSC ⇒ (1/2)(1/2)(1/2) = 1/8
SSS ⇒ (1/2)(1/2)(1/2) = 1/8

Ejemplo 12. Lanzamiento de tres dados

o :



La probabilidad de obtener las caras (3 4 2) $\rightarrow (1/6)(1/6)(1/6) = \frac{1}{216}$

Nota: Cada uno de los puntos muestrales se ramifica tal y como lo hicimos con el suceso 3.

Esperanza matemática. Si *p* es la probabilidad de éxito de un suceso en un solo ensayo, el número de sucesos o la esperanza de ese suceso en n ensayos, estará dado por el producto de *n* y la probabilidad de éxito *p*.

$E = np$

Ejemplo 13. En el lanzamiento de dos dados 900 veces, ¿cuál es la esperanza de que en la suma de sus caras se obtenga un valor menor a 6?

o :

Primero obtenemos la probabilidad de éxito que el suceso sea un valor menor a 6 en un solo ensayo:

(1,1) (1,2) (2,1) (2,2) (2,3) (3,2) (1,3) (3,1) (4,1) (1,4) $\rightarrow \quad p = \frac{10}{36}$

Como se lanza el dado 900 veces, se obtiene que: $E = np = 900\left(\frac{10}{36}\right) = \frac{9.000}{36} = 250$

Se espera que 250 de los 900 lanzamientos, la suma de sus caras resulte un valor menor a 6.

Ejemplo 14. Lanzamiento de tres monedas, ¿Cuál es la esperanza de la aparición de caras por lanzamiento?

X_i	$P_{(x)}$	$X_iP_{(x)}$
0	1/8	0
1	3/8	3/8
2	3/8	6/8
3	1/8	3/8
Σ	8/8	12/8

E = np = 3 (1/2) = 1,5

$\mu = \Sigma X_i P_{(x)} = \frac{12}{8} = 1{,}5$

El resultado de 1,5 es el número esperado de caras por lanzamiento.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

1. Las caras de un dado común se hallan numeradas de 1 a 6.
   a) ¿Cuál es la probabilidad de que habiéndose lanzado el dado, aparezca en la cara superior un valor par?
   b) ¿Cuál es la probabilidad de obtener un número mayor a dos?

2. ¿Cuál es la probabilidad de que al lanzar dos dados comunes, se presenten dos valores tales que la suma sea. a) 3? b) 4?

3. ¿Cuál es la probabilidad de que al lanzar tres monedas:
   a) ¿Todos sean caras? b) ¿Qué dos sean caras? c) ¿Qué dos sean sellos?

4. ¿Cuál es la probabilidad que sean varones, los tres hijos de una familia?

5. ¿Cuál es la probabilidad, en la experiencia de lanzar dos dados, uno blanco y otro rojo, de obtener?
   a) Primero, construir un espacio muestral
   b) ¿Que en uno de ellos se presente el 4 y en el otro un valor menor a 4?
   c) Obtener en el dado blanco un número menor de tres y en el dado rojo, un valor mayor a tres?
   d) La suma de las dos caras resulte un valor de: 6; 8; 7; más de 9.

6. Si se tienen 2 lápices uno rojo y otro verde, cuyas caras están numeradas 1, 2, 3, 4 y se hacen rodar sobre el piso, leyendo los números correspondientes a sus caras superiores. Con lo anterior:
   a) Establezca el espacio muestral de los acontecimientos.
   b) Determine la probabilidad de que la cara superior del lápiz rojo sea 1 ó 3, mientras que la de verde sea 2 ó 4.
   c) ¿Cuál es la probabilidad de que la suma de sus caras sea 4?
   d) ¿Que la suma de sus caras, sea un número par?

7. Suponga que agrega un lápiz azul al ejercicio anterior. ¿Puede establecer un espacio muestral? Si es así, determínelo:
   a) ¿Cual es la probabilidad de que exactamente una de las caras expuestas presente el número dos?
   b) Hallar la probabilidad correspondiente a que exactamente dos de las caras expuestas presenten el número dos.
   c) Que las tres caras presenten el número dos.

8. Tres corredores A, B y C compiten entre ellos frecuentemente; han ganado el 60, el 30 y el 10 por 100 de las competiciones respectivamente. En la próxima carrera:
   a) ¿Cuál será el espacio muestral?
   b) ¿Qué valores podríamos asignar a los puntos muestrales?
   c) ¿Cuál es la probabilidad de que A pierda?

9. Después de un extenso estudio, los archivos de una compañía de seguros revelan que la población de un país cualquiera puede clasificarse, según sus edades, como sigue: un 35 por ciento menores de 20 años, un 25 por ciento entre 21 y 35 años, un 20 por ciento entre 36 y 50 años, un 15 por ciento

entre 51 y 65 años y un 5 por ciento mayores de 65 años.
Suponga que se puede elegir un individuo de tal manera que cualquier habitante del país supuesto tenga la misma posibilidad de ser elegido. Empleando la anterior información, describir un espacio muestral para la edad del individuo elegido y asignar valores a los puntos muestrales. ¿Cuál es la probabilidad de que el individuo elegido sea mayor a 35 años?

10. Un embarque de pintura contiene 2.000 latas de 5 kilos, de las cuales 800 son de pintura blanca, 500 de amarilla, 300 de roja, 300 de verde y 100 de azul. Durante el viaje, las latas se han sumergido accidentalmente en agua y se han borrado todos los rótulos. A la llegada, las latas se colocan sobre una plataforma, se coge una y se abre. Respecto del color de la lata elegida:
    a) ¿Cuál es el espacio muestral?
    b) ¿Qué valores podrían asignarse a los diversos puntos muestrales?
    c) ¿Cuál es la probabilidad de que la lata elegida contenga pintura roja, blanca o azul?

11. Suponga que el observatorio meteorológico clasifica cada día según las condiciones del viento como ventoso o en calma; según la cantidad de lluvia caída, en húmedo o seco y según la temperatura como caluroso, normal o frío. ¿Qué espacio muestral es necesario para caracterizar un día? ¿Qué valores se pueden asignar a los puntos muestrales?

12. En una urna que contiene cuatro bolas blancas y dos rojas.
    a) ¿Cuál es el espacio muestral?
    b) ¿Qué valores parecen razonables para los puntos muestrales?
    c) ¿Cuál es la probabilidad de sacar una bola roja?

13. Si se lanzan 3 dados, encontrar la probabilidad de que:
    a) Los 3 presenten un cuatro.
    b) Los 3 presenten, el mismo número.
    c) Dos dados presenten el cuatro y el tercero cualquier otro número.
    d) Sólo dos dados tengan el mismo resultado.

14. Tenemos en una caja 3 bolas azules, 2 blancas, 6 negras y 5 verdes. ¿Qué probabilidad hay de ganar o perder, si las premiadas son las blancas y azules?

15. Un experimento consiste en seleccionar una ficha de una caja que contiene 6 de ellas numeradas del 1 al 6. Dadas las siguientes descripciones, establecer cuáles son espacios muestrales adecuados al experimento considerado.
    a) U = [1, 2, 3, 4, 5, 6]
    b) U = [1, 2, 3, 4, 5]
    c) U = [N° par, N° impar]
    d) U = [1, 3, 5, N° par]
    e) U = [1, 2, N° menor que 6, 6]
    f) U = [N° menor que 3, 3, N° mayor que 3]

16. Se elige un comité de 3 miembros entre 6 candidatos A B C D E y F.
    a) Especificar el espacio apropiado y asignar adecuadamente probabilidades a los sucesos elementales del espacio muestral.
    b) Hallar la probabilidad de que se elija A
    c) Hallar la probabilidad de que A y B sean elegidos.
    d) Hallar la probabilidad de que A o B sean elegidos.
    e) Hallar la probabilidad de que A no sea elegido
    f) Hallar la probabilidad de que ni A ni B sean elegidos

17. ¿Cuál es la probabilidad que una bola, extraída al azar, de una urna que contiene tres bolas rojas, cuatro blancas y cinco azules, sea blanca?

18. En una encuesta a familias con dos niños, se registra el sexo de cada niño según el orden de sus nacimientos. Por ejemplo, si el primer niño es varón y el segundo es mujer se anotará: HM. Esto es un punto del espacio muestral. Listar los demás puntos muestrales.

19. Supongamos que la encuesta del ejercicio anterior, se realice a familias con 3 hijos.
    a) Listar un espacio muestral apropiado
    b) ¿Cuántos puntos tendrá dicho espacio?
    c) ¿Cuántos de tales puntos corresponden a familias con dos hijos hombres y una mujer?
    d) ¿Cuántas familias en las que el primer hijo nacido sea mujer?

20. Asignando la misma probabilidad a cada uno de los puntos del espacio muestral del ejercicio anterior.
    a) ¿Cuál es la probabilidad de que en una familia con 3 hijos los dos primeros sean mujeres y el tercero hombre?
    b) ¿Cuál es la probabilidad de que dos sean mujeres y uno sea hombre?

21. Una persona guarda en su bolsillo una moneda de $50, una de $100, otra de $200 e imaginemos una cuarta de $500. Si esa persona toma sucesivamente (una después de la otra) dos monedas, describir un espacio muestral apropiado.

22. Se lanzan dos dados correctos. Calcular la probabilidad del suceso, “los números obtenidos no suman 4”.

23. A continuación se mencionan algunos experimentos aleatorios. En cada uno de los casos se pide describir un espacio muestral.
    a) Se extrae una carta de una baraja española (40 cartas) y se anota la carta escogida.
    b) Se lanzan 6 monedas y se observa la aparición de caras y sellos.
    c) Un joven tiene en su bolsillo 4 billetes: de $100, de $200, de $1.000 y de $10.000 pesos. Saca, uno tras otro, dos billetes.
    d) Los socios de un club eligen un comité de tres miembros entre los seis candidatos A, B, C, D, E y F.

24. Suponga tres figuras de 4 caras, numeradas del uno al cuatro. Al ser lanzadas al aire, y luego observados los números que aparecen en la cara superior, se pide:
    a) Elaborar un espacio muestral.
    b) Utilizando la regla del exponente, hallar el número de eventos posibles.
    c) Determinar la probabilidad de obtener exactamente un dos; exactamente dos dos; exactamente tres dos.

25. La probabilidad de que en un día cualquiera llueva, es del 20%. ¿Cuál es la probabilidad de que llueva? ¿Que no llueva?

26. Al lanzar un par de dados correctos, ¿cuál es la probabilidad de que:
    a) Ambos dados presenten el número tres?
    b) Ambos presenten números impares?

c) La suma de sus caras sea un número impar?
d) En uno de ellos aparezca el 3 y en el otro el 6?
e) En el primero aparezca el 3 y en el segundo el 6?

27. Si se lanza un dado correcto, ¿cuál es la probabilidad de obtener un número?
a) Par b) Impar c) Mayor que 0 d) Menor que cinco

28. Si se lanzan tres monedas, ¿cuál es la probabilidad de obtener:
a) exactamente una cara? b) por lo menos dos caras?
c) exactamente dos caras? d) como máximo tres caras?

29. a) ¿Qué es un evento?
b) Si se tiene una baraja de 52 cartas al sacar una K, ¿cuál será el evento de interés? y ¿cuáles son los puntos muestrales de ese evento?
c) ¿Cuál es la probabilidad, al extraer una carta de una baraja de 52, de obtener una figura (K, Q, J)?

30. Suponga que en un grupo (curso) de alumnos, se van a seleccionar 3 de ellos con el fin de observar, si trabajan (A) o no trabajan (B)
a) Enumérense los elementos del espacio muestral, con los tres alumnos seleccionados.
b) Enumérense los elementos en el suceso de que el número de alumnos que trabajan sea cero.
c) ¿Cómo podría definir usted el suceso {AAB, ABA, BAA}?

31. Explique brevemente qué es:
a) Probabilidad a priori b) Probabilidad empírica
c) Espacio muestral d) Diferencia entre posibilidad y probabilidad
e) Probabilidad subjetiva f) Experimento
g) Prueba h) Evento

32. Estime la probabilidad de que su equipo favorito sea uno de los 4 primeros en el campeonato del año entrante ¿En qué factores se basa para hacer dicha estimación? ¿Cómo llamaría esta probabilidad?

33. De un ejemplo de probabilidad subjetiva y otra de probabilidad objetiva.

34. Hay dos reglas que se aplican a los espacios muestrales: (a) Cada resultado posible del experimento debe corresponder a un punto muestral. (b) Analizando el espacio muestral correspondiente al lanzar un dado: aparece un número par, aparece un número impar.
En ambos casos a) y b), conteste si es cierto o falso, la aplicación de esas reglas.

Esperanza matemática

35. Se propone un juego de dados, en las siguientes condiciones: si sale el "uno" gano $5.000, pero si sale cualquier otro número pierdo $1.000. ¿Debo aceptar esa propuesta?

36. En una urna hay 50 sobres, de los cuales, 10 contienen $5.000, 10 contienen $1.000 cada uno y el resto está vacío, ¿Cuál es la esperanza obtenida al sacar un solo sobre?

37. Una persona saca sucesivamente tres bolas, sin reposición, de una urna que contiene 8 bolas negras, 8 blancas y 8 rojas. Si recibe $5.000 si no saca ninguna bola negra ¿Cuál es su esperanza?

38. Una compañía dispone de 7 obreros calificados para operar una máquina que requiere 3 obreros en cada turno.
    a) ¿Cuántos turnos son disponibles?
    b) ¿En cuántos de estos turnos aparecerá uno cualquiera de tales obreros?

39. El número de accidentes que ocurren en una determinada esquina los días viernes entre las 5:30 y las 6:30 p.m. pueden ser: 0, 1, 2, 3 ó 4; con probabilidades asociadas iguales a 0,90, 0,04, 0,03, 0,02 y 0,01.
    a) Halle el número esperado de accidentes, durante dicho período.
    b) Durante 200 períodos como el que se consideró.

40. Aseguro mi automóvil contra el riesgo de robo en la suma de $850.000. Si la probabilidad de que sea robado en el curso de un año es 0,04 ¿cuál es el precio justo de la prima anual que debo pagar?

REGLA DE MULTIPLICACIÓN, PERMUTACIONES Y COMBINACIONES

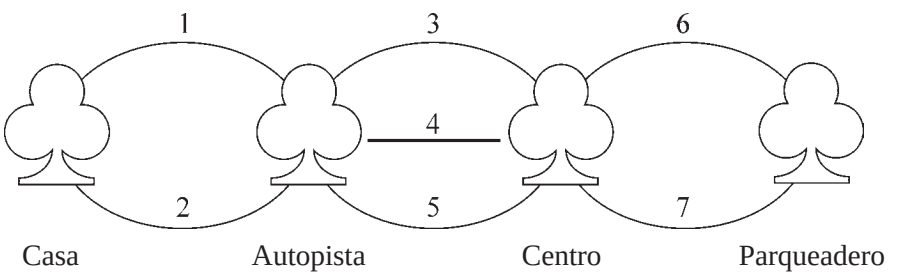
Hemos visto cómo el *diagrama de árbol* nos ayuda a establecer los puntos muestrales, y también puede ser utilizado en aquellos *experimentos compuestos*, que puede resultar tedioso, sobretodo cuando el número de resultados posibles o el número de etapas es bastante grande. La regla de multiplicación, la aplicación de las permutaciones y combinaciones, evitan en muchos casos trazar un diagrama de árbol. Veámoslo:

Regla de multiplicación

Algunos de los problemas de probabilidad tienen solución a través de la aplicación de la regla de multiplicación.

Ejemplo 1. Una persona vive en el extremo norte de la ciudad y sólo cuenta con dos rutas para poder llegar a la llamada Autopista Norte. Una vez alcanzada la autopista tiene tres rutas de menor congestión para llegar al centro de la ciudad. Ya en el centro puede seleccionar dos rutas para llegar al parqueadero más cercano a su oficina. La pregunta que se haría dicha persona es: ¿de cuántas maneras o rutas podría conducir su automóvil de la casa al parqueadero más próximo a la oficina?

o :



Las alternativas que tendrá son:

{1,3,6} {1,3,7}	{1,4,6} {1,4,7}	{1,5,6} {1,5,7}
{2,3,6} {2,3,7}	{2,4,6} {2,4,7}	{2,5,6} {2,5,7}

En total se tienen 12 maneras o rutas para elegir. Aplicando la técnica de la multiplicación, se tendrá: 2 x 3 x 2 = 12.

Para el total de arreglos que se pueden hacer, la fórmula general es: Total arreglos posibles = $m \cdot n$
Si tenemos por ejemplo cuatro eventos será igual : m x n x o x p

Ejemplo 2. En el experimento de lanzar una moneda y a la vez un dado ¿cuál es el número de puntos muestrales?

o :

En el primer caso (moneda) son 2 y en el dado son 6 las posibilidades por lo tanto el número de puntos muestrales del experimento compuesto es 2 x 6 = 12

Ejemplo 3. Supongamos que se va a extraer una carta en una baraja de 52 cartas ¿Cuántos puntos muestrales tendrá un experimento compuesto (a) si después de extraer una carta, se vuelve al mazo y luego se extrae otra carta? (b) si luego de extraer una carta, ésta se deja por fuera y luego se extrae otra segunda carta?

o :

a) En la primera se tienen 52 cartas y como inmediatamente se devolvió, la que se extrajo en la segunda se tendrán otra vez las 52 cartas, por lo tanto el total de puntos muestrales será: 52 x 52 = 2.704
b) En la primera extracción se tendrán 52 puntos, pero en la segunda no se devolvió la carta, sólo hay 51 puntos, en este caso será: 52 x 51 = 2.652 puntos muestrales.

Ejemplo 4. Se lanzan tres dados (a) ¿Cuántos puntos muestrales se tiene? (b) ¿Cuál es la probabilidad de obtener un 2 en uno de los dados? (c) Que aparezca un 2 en el primer dado (d) De obtener exactamente dosd 2.

o :

a) Se tendrá 216 puntos muestrales $6^3 = 216$.
b) En total se obtiene 77 resultados, la probabilidad será: 77/216
c) Si tiene 36 casos posibles, la probabilidad será: 36/216
d) Son 15 los casos posibles, la probabilidad será: 15/216

Permutaciones y combinaciones

La regla de la multiplicación es la base de dos fórmulas que nos permiten simplificar en forma considerable el conteo de puntos muestrales, siendo ellas las *permutaciones* y las *combinaciones*.

Se puede decir que las *permutaciones* es una forma de ordenar o arreglar la totalidad de los elementos de un conjunto; también se puede considerar como un conjunto de cosas extraídas en un orden *específico* y sin reemplazo de un conjunto igual o mayor.

Generalmente simbolizado por $P_n = n!$ o como ${}_nP_n = n!$
Lo anterior se lee como “permutaciones de n elementos tomados de n en n”
El símbolo $n!$ se lee como n factorial, consideremos el ejemplo de 8!

o **:**

8! = 8 × 7 × 6 × 5 × 4 × 3 × 2 × 1 También se puede escribir a: 8! = 8 × 7 × 6 × 5!

Ejemplo 1. Supongamos que se tienen los siguientes números naturales 1, 2, 3, 4 y se quiere formar cifras de 4 dígitos. Según la fórmula anterior se tendrá que ${}_4P_4 = P_4 = 4! = 4 \times 3 \times 2 \times 1 = 24$. Veamos cuales serían esas cifras:

1234	2134	3142	4132
1243	2143	3124	4123
1324	2314	3214	4213
1342	2341	3241	4231
1432	2413	3412	4312
1423	2431	3421	4321

$P_n = n!$

$P_4 = 4! = 24$ cifras

Ejemplo 2. En la primera línea del salón de clases se tienen colocados 8 pupitres y se quiere sentar a 8 alumnos; ¿de cuántas maneras se podrán colocar?

o : ${}_8P_8 = P_8 = 8! = 40.320$ maneras

Ejemplo 3. Con las letras de la palabra PALO, ¿cuántas palabras se pueden formar?

PALO	APLO	LPAO	OPAL
PAOL	APOL	LPOA	OPLA
PLAO	AOPL	LOPA	OLAP
PLOA	AOLP	LOAP	OLPA
POLA	ALOP	LAPO	OALP
POAL	ALPO	LAOP	OAPL

$P_4 = 4! = 4 \times 3 \times 2 \times 1 = 24$ palabras

$P_n = n!$

Cuando uno o varios elementos están repetidos, el cálculo de las permutaciones varía: en este caso nos referimos a *Permutaciones con repetición*; por ejemplo con la palabra CASA tendríamos un número de palabras inferior a 24, en el caso de que no se haga distinción de la A.

CASA	ACSA	SCAA	CSAA	AACS	SAAC
CAAS	ACAS	SACA	ASAC	AASC	ASCA

$$P_{n\,(r)} = \frac{n!}{r!}$$

La fórmula que se debe aplicar es: $P_{n\,(r=2)} = \frac{n!}{r!} = \frac{4!}{2!} = \frac{4 \times 3 \times \cancel{2!}}{\cancel{2!}} = 12$ palabras

Las *permutaciones con repetición, r* son un caso particular de las variaciones y no existe una ley sencilla para su formación. Dado lo complicado del sistema, solamente se presenta la fórmula con la cual se logra obtener el número de esta clase de permutaciones. Sean los elementos aa – bbb – cc – d para permutar con repetición. Tendremos 8 elementos repartidos así: dos del primero, tres del segundo, dos del tercero y uno del cuarto, entonces las permutaciones se presentan así: $P_{8\,(r:2,3,2,1)}$ y la fórmula será:

$$P_{n(r:2,3,2,1)} = \frac{8!}{2!\ 3!\ 2!} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times \cancel{3!}}{2 \times 1 \times \cancel{3!} \times 2 \times 1} = 1.680;$$

$$P_{n(r:r_1 \,.\, r_2)} = \frac{n!}{r_1!\ r_2!}$$

Ejemplo 4. ¿De cuántas maneras distribuiríamos 3 monedas de $50 y 4 monedas de $100 en una misma linea?

o **:** $P_{7(r:3,\,4)} = \dfrac{7!}{3!\ 4!} = 35$ maneras

Ejemplo 5. ¿Cuántos grupos de 6 letras se pueden formar con las letras de la palabra Amigas?

o **:** $P_{6(r:2)} = \dfrac{6!}{2!} = 360$ grupos de letras

Ejemplo 6. ¿Cuántas palabras se pueden formar con las letras de la palabra Barranquilla? Primero contamos el número de letras, para determinar el valor de *n*, en este caso *n* = 12; luego las letras repetidas; A se repite 3 veces (r_1= 3); R dos veces (r_2= 2) y la L dos veces (r_3= 2). Aplicando la fórmula se tendrá:

$$P_{12(r:3,\,2,\,2)} = \frac{12!}{3!\ 2!\ 2!} = \frac{12\times11\times10\times9\times8\times7\times6\times5\times4\times\cancel{3!}}{\cancel{3!}\times2\times1\times2\times1} = 19.958.400 \text{ palabras}$$

En las *permutaciones* cuando no se utilizan todos los elementos, sino una parte de ellos, algunos lo denominan como *variaciones*, cuya fórmula está dada así:

$$_nP_r = {}_nV_r = \frac{n!}{(n-r)!}$$

Ejemplo 7. Volvamos nuevamente a los cuatro números naturales: 1, 2, 3, 4 del ejemplo 1 y formemos cifras de tres dígitos, con los siguientes resultados:

o **:**

123	213	312	413
132	231	321	431
124	214	314	412
142	241	341	421
134	234	342	423
143	243	324	432

$$_4P_3 = \frac{4}{4-3} = \frac{4}{1} = 24 \text{ cifras}$$

Ejemplo 8. Si ahora se quieren formar cifras de dos digitos, se tendrá:

o **:**

12	21	31	41
13	23	32	42
14	24	34	43

$$_4P_2 = \frac{4}{4-2} = \frac{4}{2} = \frac{24}{2} = 12 \text{ cifras}$$

Ejemplo . Si en el ejemplo 2 de los 8 pupitres colocados en la primera línea del salón se quiere determinar de cuántas maneras podemos ordenar a 5 alumnos, se tendrá:

$$_8\mathrm{P}_5 = \frac{}{-5} = \frac{}{3} = \frac{\times7\times6\times5\times4\times3}{3} = 6.720 \text{ maneras}$$

Combinaciones. Son un arreglo de los elementos sin importar el orden en que se dispongan.

La fórmula que se utiliza en el cálculo de las combinaciones es:

$$ {}_nC_r = \frac{n!}{(n-r)!\ r!} $$

Ejemplo 10. Cambiemos el ejercicio de los 4 números naturales por las primeras cuatro letras del alfabeto A, B, C, D. Si se desea combinarlos, ¿cuántas combinaciones se podrán hacer?

o :

Una sola combinación, ya que al no importar el orden de colocación da lo mismo
ABCD = ADBC = ACBD = CBAD = DACB = etc.

$$ {}_4C_4 = \frac{4}{4-4\quad 4} = \frac{4}{0\quad 4} = \frac{4}{1\quad 4} = 1 \text{ combinación} $$

Ejemplo 11. Si se fueran a combinar esas 4 letras de dos en dos, se tendría: AB = BA AC = CA
BC = CB BD = DB CD = DC AD = DA

Se pueden combinar de 6 maneras: $ {}_4C_2 = \frac{4}{4-2\quad 2} = 6 $

Ejemplo 12. En este mismo ejercicio nos pidieran que combináramos las letras de 3 en 3, el resultado sería:

ABC	=	BCA	=	CBA	=	ACB	=	BAC	=	CAB
ABD	=	ADB	=	BDA	=	BAD	=	DAB	=	DBA
ACD	=	ADC	=	CAD	=	CDA	=	CDA	=	CAD
BCD	=	BDC	=	CBD	=	CDB	=	DBC	=	DCB

En total se puede combinar de 4 maneras. Mediante la aplicación de la fórmula se tendrá:

$$ {}_nC_r = \frac{n!}{(n-r)!\ r!} $$

$$ {}_4C_3 = \frac{4}{4-3\quad 3} = \frac{4}{1\quad 3} = 4 $$

APLICACIÓN DE LA CALCULADORA

Los cálculos anteriores, para determinar el total de puntos muestrales que tenga el experimento, se pueden agilizar mediante el uso de la calculadora. La mayoría de ellas tienen en sus teclas la indicación correspondiente, si se trata de permutaciones aparece ${}_nP_r$ y en las combinaciones ${}_nC_r$ algunas implican una vez digitado el valor de n, oprimir la tecla SHIFT y luego el valor de r, de esta forma se obtiene el resultado respectivo:

$$ {}_nC_r = \frac{n!}{(n-r)!\ r!} \qquad {}_{10}C_6 = \frac{10}{4\quad 6} = 210 \qquad {}_{10}P_6 = \frac{10}{4} = 151\,200 $$

En la calculadora CASIO fx 5.000F, las fórmulas están incorporadas así: con 12 y luego la tecla FMLA aparece la fórmula para el cálculo de las permutaciones y con la cifra 13 y FMLA se obtiene la correspondiente a las combinaciones.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

41. En una Universidad de Bogotá a 5 estudiantes se les califica con las letras A, B, C, D. E. ¿De cuántas maneras se les puede calificar, si los estudiantes obtienen todos calificaciones diferentes?

42. Si un futbolista conoce 7 jugadas diferentes y si el entrenador le instruye para que juegue las 7 sin que ninguna se repita, ¿qué libertad le queda a ese jugador?

43. Una señora invita a cenar a 8 amigos y después de sentarse ella, ¿De cuántas maneras se pueden sentar sus invitados?

44. ¿Cuántas permutaciones se pueden hacer con las letras: A_1 A_2 B_1 B_2?

45. ¿Cuántas cifras de 9 dígitos se pueden formar con los dígitos del 1 al 9?

46. Un examen consta de cuatro preguntas y se deja libertad para contestarlas en el orden que se desee. ¿De cuántas maneras se podrá contestar?

47. ¿Cuántas palabras de 5 letras, con o sin sentido idiomático pueden fomarse a partir de las letras de la palabra COSER?

48. ¿Cuántas permutaciones pueden efectuarse con las letras A, B, C?

49. ¿Cuántas permutaciones se pueden hacer con las letras de la palabra MISSISSIPPI?

50. a) ¿Cuántas permutaciones se pueden hacer con las letras de la palabra COOPERADOR?
    b) ¿Ahora, si se considera que las O deben estar juntas?

51. ¿Cuántas permutaciones se pueden hacer con las letras: A A B_1 B_2 B_3?

52. ¿Cuántas permutaciones se pueden hacer con las letras: A A B B B?

53. ¿De cuántas maneras se pueden ordenar en un estante con 5 litros de whisky y tres botellas de aguardiente, a condición de que dos litros de whisky estén siempre juntos y dos botellas aguardiente siempre juntas?

Permutaciones (Variaciones)

54. Si un estudiante tiene 9 libros y desea ordenar 5 de ellos sobre un estante. ¿De cuántas maneras distintas puede hacerlo?

55. ¿Cuántos números de 4 dígitos pueden formarse con los dígitos: 1, 3, 5, 7, 8 y 9. si ninguno puede aparecer más de una vez en cada número?

56. ¿Cuántas palabras de cinco letras diferentes se pueden formar con las 27 letras del alfabeto?

57. ¿Cuántos números de tres cifras se pueden formar con los dIgitos: 2, 3, 4, 5, 6?

58. ¿Cuántas señales diferentes se pueden formar con 10 banderas distintas, levantando al menos 3 y no más de 6 banderas en una driza de un mástil?

59. ¿De cuántas maneras diferentes se puede contestar un examen de 5 preguntas, si sólo hay que dar respuesta a 3 de ellas?

60. ¿Cuántas ordenaciones diferentes de diez letras se pueden hacer utilizando la palabra CONVENCIÓN?

61. Determine el valor para cada uno de los siguientes casos: a) 6! b) 10! c) 3! d) 0!

62. a) ¿Cuántos números de tres dígitos pueden formarse con 1, 2, 3, 4, 5, 6, 7, 8?
    b) ¿Cuántas cifras de 4 dígitos?
    c) ¿Cuántas cifras de 5 dígitos?

63. ¿Cuántas permutaciones se pueden obtener con las letras de la palabra CARRASQUILLA?

64. Si consideramos un alfabeto con 26 letras ¿cuántas placas de tres letras se pueden elaborar, si una letra no debe aparecer más de una vez?; b) ¿Cuántas placas se tendrán, si las letras pueden aparecer más de una vez?

65. Un mecánico contratado para arreglar una máquina, cree que hay posiblemente 4 causas y se le propone dar solución a cada una de ellas, señalándolas por A, B, C, D
    a) Enumérese el orden que podría dar a las soluciones.
    b) Aplicando la fórmula correspondiente ¿de cuántas maneras se pueden permutar?

66. Si consideramos el actual campeonato de balompié conformado por 16 equipos:
    a) ¿De cuántas maneras se podrán clasificar, suponiendo que no hay empates?
    b) Si sólo consideramos los 4 primeros puestos, ¿cuántas clasificaciones diferentes se pueden hacer?

67. ¿De cuántas maneras posibles se pueden ordenar en una biblioteca 3 libros de matemática y dos de estadística, se fija como condición que los libros de matemática estén siempre juntos, lo mismo que las estadísticas?

68. Un joven ha invitado a 6 amigos a comer. Después de sentarse él ¿de cuántas maneras diferentes pueden sentarse los amigos?

69. Obtener los resultados para los siguientes: a) ${}_6P_2$ b) ${}_4P_1$ c) ${}_{12}P_5$ d) P_6 e) P_8

Combinaciones

70. Desarrolle las siguientes combinaciones:

a) $\binom{8}{6}$ b) $\binom{5}{3}$ c) $\binom{5}{2}$ d) $\binom{8}{2}$ e) $\binom{10}{4}$ f) $\binom{10}{6}$

¿Observa algo especial en el desarrollo de los ejercicios anteriores? ¿Por qué se dan esos resultados iguales?

71. ¿De cuántas maneras puede formarse un equipo de balompié (bajo el supuesto que pueden jugar en cualquier puesto) entre un plantel de 30 jugadores?

72. ¿De cuántas maneras sepuede organizar una familia de 5 hijos, si desea que dos sean niñas y tres niños?

73. ¿Cuántas comisiones de 6 personas pueden formarse con un grupo de 10 personas?

74. Supongamos que Pedro, María, Grisel, Juan y Jorge son los candidatos para conformar un comité, compuesto de tres personas. a) ¿Cuántos comités de tres personas se pueden conformar?; b) Grisel y Juan, por ser hermanos, no deben estar juntos en los comités.

75. Al desarrollar las siguientes combinaciones, ¿qué observa usted en los resultados?

a) $\binom{10}{6}=\binom{10}{4}$ b) $\binom{8}{3}=\binom{8}{5}$ c) $\binom{7}{2}=\binom{7}{5}$

76. ¿Cuántos comités diferentes de 4 personas se pueden formar a partir de un grupo de 12 personas?

77. a) Suponga que en el ejercicio anterior el comité de 4 personas tiene que estar conformado por una mujer y 3 hombres.
b) Si el grupo está conformado por 4 mujeres y 8 hombres,
¿En los dos casos, de cuántas maneras diferentes los podemos organizar?

78. ¿Cuántos grupos de 7 cartas, pueden obtenerse de una baraja de 40 cartas?

79. ¿Cuántos comités diferentes pueden seleccionarse entre 7 hombres y 4 mujeres si deben constituirse de:
a) 3 hombres y 2 mujeres
b) 5 personas de las cuales por lo menos tres deben ser hombres.

80. Es necesario elegir un comité de 10 personas entre 6 abogados, 8 economistas y 5 ingenieros. Si el comité debe estar integrado por 4 abogados, 3 economistas y 3 ingenieros.

81. ¿Cuántos comités compuestos de 3 diputados y 5 senadores pueden formarse tomando como base un grupo de 5 diputados y 8 senadores?

82. a) ¿Cuántas comisiones de 3 personas se pueden formar seleccionándolas de entre 10 personas?
b) ¿Comisiones de 7 personas entre 10?

83. ¿Cuántos grupos de cinco cartas se pueden obtener de una baraja de 52 cartas?

84. ¿De cuántas maneras se pueden extraer dos manzanas de una caja que contiene 8 manzanas?

85. Una caja contiene 7 fichas rojas, 6 fichas blancas y 4 fichas azules. ¿Cuántas selecciones de 3 fichas se pueden formar si:
a) las 3 deben ser rojas?
b) ninguna puede ser roja?

86. Un examen consta de 4 preguntas, hay que dar respuesta a solo 3 de las 4 preguntas, ¿Cuántos exámenes de diferente contenido habrá que corregir como máximo?

ALGUNAS REGLAS BÁSICAS DE PROBABILIDAD

Clases de sucesos. Es necesario primeramente considerar varias clases de sucesos:

- *Sucesos igualmente probable*: lanzar una moneda, aparición de cara o sello.
- *Sucesos opuesto o contrario*: siendo aquellos que se complementan básicamente.
- *Sucesos ciertos*: una moneda con dos caras.
- *Sucesos imposibles*: lanzar un dado y que aparezca en la cara superior un 8.
- *Sucesos compatibles*: que puede suceder en una baraja, aparezca simultáneamente un seis y que sea oros.
- *Sucesos mutuamente excluyentes*: al lanzar un dado, aparece un dos o un seis.
- *Sucesos independientes*: al lanzar dos dados, obtener en el primero un dos y en el segundo, un seis.
- *Sucesos dependientes*: la ocurrencia de uno afecta la ocurrencia del otro.

Para facilitar el cálculo de las probabilidades se emplean algunas reglas como se verán a continuación.

REGLA DE ADICIÓN

o t a t t

Si dos o más sucesos son tales, que solamente uno de ellos puede ocurrir en un solo ensayo, se dice que son mutuamente excluyentes. Se denomina *probabilidad aditiva* y será igual a la suma de las probabilidades de cada suceso.

$$P = p_1 + p_2 + p_3 + ... + p_n$$

Consideremos que p_1, p_2, p_3, ... p_n son las distintas probabilidades de n sucesos mutuamente excluyentes, la probabilidad (P) de que uno de estos sucesos se presente en un *solo ensayo*, estará dada por la suma de las probabilidades para cada suceso $P = p_1 + p_2 + p_3 + ... + p_n$

De acuerdo a lo anterior, *mutuamente excluyente* significa que solamente un solo suceso o evento puede ocurrir, o sea que los demás no se pueden presentar al mismo tiempo. La fórmula anterior la podemos expresar de una manera más fácil y entendible:

$$P\,(A \text{ ó } B) = P_{(A)} + P_{(B)}$$

$$P_{(A \text{ o } B \text{ o } C)} = P_{(A)} + P_{(B)} + P_{(C)}$$

También se presenta como: $P_{(A \cup B)} = P_{(A)} + P_{(B)}$ donde $\cup$ significa unión.

En el tema sobre *probabilidad condicional* existen agunos autores que utilizan simbología diferentes a las presentadas en esta publicación.

Ejemplo 1. La probabilidad de obtener un As o un Rey, sacando una sola carta en una baraja española de 40 cartas. Si uno de los casos aparece, queda excluido el otro.

o :

$P_A = \frac{4}{40} = \frac{1}{10}$ *AS* $\qquad P_B = \frac{4}{40} = \frac{1}{10}$ *REY*

$P_{A\,o\,B} = P_A + P_B = \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = 1\ 5$

Ejemplo 2. La probabilidad de obtener un 2 o un 5, en el lanzamiento de un dado.

o :

$P_{(A)} = 1/6$ (aparición del 2) $\qquad P_{(A)} = 1/6$ (aparición del 5)

$P_{(A\,o\,B)} = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 1\ 3$

Ejemplo 3. Tenemos una caja o recipiente con 16 bolas de 4 colores diferentes:

3 bolas azules 6 bolas negras 2 bolas blancas 5 bolas verdes.

¿Qué probabilidad tenemos de ganar o perder si las premiadas son las blancas y azules?

o :

Posibilidades: $3 + 2 + 6 + 5 = 16$

Probabilidad adversa: $q = \frac{6}{16} + \frac{5}{16} = 11\ 16$

Probabilidad: $p = \frac{3}{16} + \frac{2}{16} = 5\ 16$

Probabilidad total: $p + q = \frac{5}{16} + \frac{11}{16} = 1$

Se ha dicho que un *espacio muestral* es el conjunto de resultados posibles de un experimento o ensayo. Un *ensayo* es considerado como un experimento cuyos resultados no necesariamente tienen que ser los mismos cada vez que se repita.

Cualquier subconjunto del espacio muestral es considerado como evento; un subconjunto que contiene un solo punto muestral se denomina un evento elemental o punto muestral. De ahí que los resultados individuales que constituyen el espacio muestral reciban la denominación de sucesos o puntos muestrales.

Ejemplo 4. En una empresa el personal se distribuye así:

	Personal	N
A	Vigilancia y aseo	30
B	Auxiliar	58
C	Operarios	180
D	Supervisores y técnicos	70
E	Profesionales	40

Se desea saber Cuál es la probabilidad

a) Al elegir una persona, corresponda al personal auxiliar?
b) Al elegir una persona sea alguien de vigilancia y aseo o profesional?
c) Que sea un profesional o un operario o un auxiliar.

o :

a) El número total de empleados es de 378, por lo tanto la probabilidad de que sea personal auxiliar, será de:

$$P_{(B)} = \frac{58}{378} = 0{,}1534 = 15{,}34\% \Rightarrow P_{(B)} = 0{,}1534$$

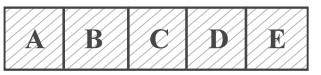
b) $P_{(A\ o\ E)} = P_{(A)} + P_{(E)} = \frac{30}{378} + \frac{40}{378} = \frac{70}{378} = 0{,}1852 \Rightarrow P_{(A\ o\ E)} = 0{,}1852$ Probabilidad que sea de vigilancia y aseo o profesional.

c) $P_{(E\ o\ C\ o\ B)} = P_{(E)} + P_{(C)} + P_{(B)} = \frac{40}{378} + \frac{180}{378} + \frac{58}{378} = \frac{278}{378} = 0{,}7354$ Probabilidad que sea profesional u operario o auxiliar.

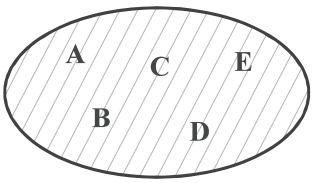
Diagramas

El investigador inglés J. Venn (1834 — 1888) elaboró diagramas que permitieran en forma gráfica visualizar mejor el resultado de un experimento. El *espacio muestral*, estará delimitado por un espacio en forma de cuadrado, rectángulo, circular o figura, que no corresponda a ninguna de las anteriores, que represente todos los posibles resultados.

En el Ejemplo 4 el diagrama de Venn, cuando los sucesos o eventos son mutuamente excluyentes será:



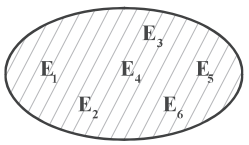
o también



Ejemplo 5. a) En el lanzamiento de un dado, trace un diagrama de Venn.
b) ¿Cuál es la probabilidad de que aparezca un 3?
c) ¿Que aparezca 3 o un 6?

o :

a)

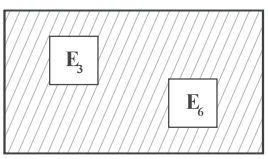


b) Probabilidad de que aparezca un 3 = 1/6 = 0,167

c) $P_{(E_3\ o\ E_6)} = P_{(E_3)} + P_{(E_6)} = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3} = 0{,}3333$

Probabilidad que aparezca el 3 o el 6.

La representación para el punto (C) será:



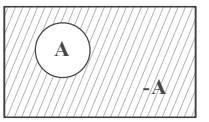
Recordemos que la probabilidad representada por el espacio muestral es 100% y la probabilidad de cualquier evento A, corresponderá a un valor que puede variar entre 0 y 1.

$$0 \leq P_{(A)} \leq 1$$

Además la probabilidad de que un evento no ocurra, es igual a uno menos la probabilidad de que si ocurra.

$P(A') = 1 - P(A)$ $\quad P\ A\ = 1 - \frac{1}{6} = \frac{6}{6} - \frac{1}{6} = \frac{5}{6} = 0,\ 33 = \ 3{,}33$

En el *diagrama de Venn* se observará así:



Ejemplo 5. Volvamos al experimento de extraer una carta (óigase bien, una sola carta, de una sola baraja): ¿Cuál es la probabilidad de obtener un as o un seis de copas o un rey?

o **:**

Probabilidad de que sea un as: $P_A = \frac{4}{40}$

Probabilidad de que sea un seis de copas: $P_B = \frac{1}{40}$

Probabilidad de que sea un rey: $P_C = \frac{4}{40}$

Como sólo debe ocurrir uno de estos eventos, dado que se extraerá una carta, la probabilidad de que sólo uno de ellos ocurra estará dada por:

$$P(A \text{ o } B \text{ o } C) = P(A) + P(B) + P(C) =$$

$$P_{(A\,o\,B\,o\,C)} = \frac{4}{40} + \frac{1}{40} + \frac{4}{40} = \frac{}{40} = 0{,}225 = 22{,}5$$

Sucesos compatibles

Ejemplo 6. Ahora, el experimento con la baraja de 40 cartas consiste en extraer una carta, y se desea saber cuál es la probabilidad de que la carta extraida sea as o copas.

Observamos que al extraer una carta puede ser *as*, pero también puede ser *as de copas*, cumpliéndose la realización de las dos pruebas en forma simultánea; por tal razón, se dice que los sucesos son *compatibles*, o también nos podemos referir a una *probabilidad conjunta*.

Se recomienda en el tema de *probabilidad condicional* ver o consultar otras simbologías que se han utilizado en otros libros.

Se dice que dos sucesos son compatibles, o que no son mutuamente excluyentes, cuando la posibilidad de que ocurra un suceso no impide la ocurrencia del otro. En este caso la probabilidad de uno de los dos sucesos se halla así:

$$P(A \text{ o } B) = P_{(A)} + P_{(B)} - P_{(A \text{ y } B)}$$

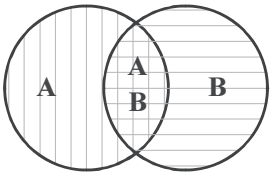
$$P(A \text{ o } B) = \frac{4}{40} + \frac{10}{40} - \frac{1}{40} = \frac{13}{40} = 0{,}325 = 32{,}5\%$$ Probabilidad que sea as o copas.

La probabilidad de que aparezca un as: $P_{(A)} = \frac{14}{40}$

La probabilidad de que aparezca copas $P_{(B)} = \frac{10}{40}$

La probabilidad de que sea el as de copas: $P_{(A\,y\,B)} = \frac{1}{40}$

En un diagrama de Venn se pueden representar *eventos compatibles*.



Los eventos A y B no son mutuamente excluyentes, ya que tienen algunos elementos en común.

Es corriente utilizar símbolos, tales como $A \cup B$ para indicar la unión de A con B, y $A \cap B$ la intersección de A con B.

La anterior expresión de:

$P_{(A\,o\,B)} = P_{(A)} + P_{(B)} - P_{(A\,y\,B)}$ Se podrá reemplazar por $P_{(A \cup B)} = P_{(A)} + P_{(B)} - P_{(A \cap B)}$

Ejemplo 7. Al lanzar un dado, usted apuesta $1.000 a que el número obtenido debe ser par o divisible por 3. ¿Cuál es la probabilidad de que usted gane en este lanzamiento?

o **:**

Que aparezca un número par {2, 4, 6} $P_{(A)} = \frac{3}{6}$

Ahora, que sea divisible por 3 será {3, 6} $P_{(A)} = \frac{2}{6}$. Pero observamos que el 6 aparece tanto en el suceso A como en el B por tanto, se tendrá que {6} $P_{(A \cap B)} = \frac{1}{6}$

$$P_{(A \cup B)} = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3} = 0{,}6667 = 66{,}67\%$$ La probabilidad de ganar.

Ejemplo 8. La probabilidad de que un alumno del instituto tenga un libro de matemáticas en su biblioteca es 0,7; un libro de estadística es 0,4 y de que tenga ambos es 0,30. ¿Cuál es la probabilidad de que un estudiante del instituto tenga un libro de matemáticas o un libro de estadística o ambos libros?

$P(A \cup B) = 0{,}7 + 0{,}4 - 0{,}3 = 0{,}8$ Probabilidad de obtener un libro de matemática o de estadiística.

Colectivos ex austivos

Se dice también, que los eventos son *colectivamente exhaustivos* si no es posible obtener otro resultado para un experimento dado, es decir, por lo menos uno de ellos debe ocurrir. Así por ejemplo, al extraer una carta de una baraja de 52, ¿cuál es la probabilidad de que sea trébol o diamantes o corazones o espadas? Se puede observar que se agotan todas las posibilidades, ya que se considera la totalidad de los elementos; la probabilidad será igual a 52/ 52 = 1,0.

REGLA DE LA MULTIPLICACIÓN

Sucesos independientes

Se dice que dos o más sucesos son *independientes*, si la probabilidad de presentación de ninguno de ellos queda influenciada por la presentación del otro. En caso contrario se dice que son *dependientes*.

En otras palabras, si el resultado de un suceso no afecta al otro, se dice que son *independientes*, por lo tanto se efectuará la multiplicación de las probabilidades para cada suceso.

> Si p_1, p_2, p_3,..... p_n, son las distintas probabilidades de presentación de n sucesos *independientes*, la probabilidad (P) de que ocurran todos estos sucesos en un solo ensayo, estará dada por el producto de las probabilidades para cada suceso.
> $P = p_1 \times p_2 \times p_3 .... \times p_n$

Ejemplo 8. ¿Qué probabilidad tendremos de obtener dos Reyes sacando una carta de una baraja y la otra de una segunda baraja?

o ***:***

$$P = \frac{4}{40} \times \frac{4}{40} = \frac{16}{1.600} = 1/100$$

$$P_{(A\,o\,B)} = P_{(A)}\,P_{(B)}$$

Ejemplo 10. Al lanzar dos dados, ¿cuál es la probabilidad de sacar dos cincos?

o ***:***

$P_1 = \frac{1}{6}$ (5 en el primer dado); $p_2 = \frac{1}{6}$ (5 en el segundo dado); $P = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$

> Recordemos cuando se dice que dos o más eventos son *independientes* entre si, la ocurrencia de un evento no está relacionada con la ocurrencia de los otros. Si hay tres eventos independientes A, B y C, la probabilidad de que ocurran A, B y C se obtiene al multiplicar las tres probabilidades.
>
> $$P(A \text{ y } B \text{ y } C) = P(A) \,.\, P(B) \,.\, P(C)$$

Ejemplo 11. Supongamos que se dispone de tres barajas de 40 cartas cada una. Se desea extraer tres cartas, una de cada baraja; ¿cuál es la probabilidad de obtener un as y un rey de oros y un seis de copas?

o ***:***

En la primera baraja se tienen 4 ases, siendo: P(A) = 4/40
En la segunda baraja se tiene un rey de oros: P(B) = 1/40
Y en la tercera baraja hay un seis de copas: P(C) = 1/40

Observamos que los resultados son independientes, pues ninguno de ellos se ve afectado por la aparición del otro; en estos casos aplicamos la regla especial de multiplicación, siendo:

$$P(A \text{ y } B \text{ y } C) = \frac{4}{40} \times \frac{1}{40} \times \frac{1}{40} = \frac{4}{64.000} = \frac{16}{16.000} = 0{,}0000625$$

Es necesario clarificar un poco más la diferencia entre sucesos mutuamente excluyentes y sucesos independientes.

a) En el primero se tiene un solo dado, una baraja; en el segundo son dos o más dados o barajas.
b) En el primero se extrae una sola carta, o se obtiene una sola cara, es decir, se espera la presentación de un suceso; en el segundo se espera la presentación de dos o más sucesos.
c) En el primero utilizamos la conjunción "o" y en el segundo la conjunción "y".

Ejemplo 12. Una fábrica de calzado produce, independientemente costura (toda la parte superior del calzado relacionada con el cuero), suela y tacón, siendo estas partes armadas aleatoriamente en cada zapato. Se sabe que en este proceso, el cinco por ciento de las costuras, el cuatro por ciento de las suelas y el uno por ciento de los tacones tiene fallas; ¿qué porcentaje de pares de zapatos resulta: a) con fallas en sus tres componentes?; b) sin fallas en su tres componentes?

o :

a) $P(C\ y\ C\ y\ S) = (0{,}05)\ (0{,}04)\ (0{,}01) = 0{,}00002 = 0{,}002\%$

b) $P_{(C)} = 1 - 0{,}05 = 0{,}95 \qquad P_{(s)} = 1 - 0{,}04 = 0{,}96 \qquad P_{(t)} = 1 - 0{,}01 = 0{,}99$

$P(C\ y\ C\ y\ S) = (0{,}95)\ (0{,}96)\ (0{,}99) = 0{,}903 = 90{,}3\%$

Ejemplo 13. Una máquina en buenas condiciones de trabajo, produce un artículo defectuoso por cada mil. Los resultados correspondientes a artículos producidos sucesivamente son independientes. ¿Cuál es la probabilidad para que los próximos dos artículos producidos por esta máquina no tengan fallas?

La probabilidad de que el artículo no sea defectuoso es igual a:

$$P = 1 - \frac{1}{1.000} = \frac{1.000}{1.000} - \frac{1}{1.000} = \frac{999}{1.000} = 0{,}999$$

Ahora la probabilidad para A y B será: P(A y B) = (0,999) (0,9998001) = 99,8%

Sucesos dependientes

Se dice que los sucesos son *dependientes* o *eventos compuestos*, si la ocurrencia o no ocurrencia de un evento en cualquier prueba afecta la probabilidad de otros eventos en otras pruebas, es decir que la probabilidad del segundo suceso depende del primer suceso, el del tercero de lo que haya sucedido en el primero y segundo y así sucesivamente.

Si se van a extraer tres cartas de una baraja, se debe hacer sin reposición, es decir, al extraer una carta, ella no vuelve a formar parte del total y en vez de tener 40 cartas, para la segunda se tendrán 39. Recordemos, que dos o más eventos son *dependientes*, cuando la ocurrencia de uno afecta la ocurrencia de los otros, en un orden determinado. En caso contrario los sucesos son independientes. La fórmula general será: $P = p_1 \times p_2 \times p_3 \times \ldots\ldots\ldots \times p_n$

Ejemplo 14. Probabilidad de obtener 3 ases, sacando sucesivamente tres cartas de una baraja española, sin volverlas a incluir (sin reposición) en el montón.

o :

$$p_1 = \frac{4}{40} \qquad p_2 = \frac{3}{3} \qquad p_3 = \frac{2}{3} \qquad P = \frac{4}{40} \times \frac{3}{3} \times \frac{2}{3} = \frac{1}{2\,470}$$

Ejemplo 15. Probabilidad de obtener un As, un Rey y una Zota, sacando sucesivamente tres cartas, sin reposición, de una baraja de 40 cartas.

o :

$$p_1 = \frac{4}{40}\ (As) \qquad p_2 = \frac{4}{3}\ (\ \ y) \qquad p_3 = \frac{4}{38}\ (Zota) \qquad P = \frac{4}{40} \cdot \frac{4}{3} \cdot \frac{4}{38} = \frac{64}{5\ .280}$$

Ejemplo 16. De una baraja de 40 cartas se desea extraer tres cartas en forma sucesiva sin reposición, es decir, la carta que se extrae no regresa a la baraja; ¿cuál es la probabilidad de que en la primera extracción aparezca un *as* y en la segunda un *rey de oros* y en la tercera un *seis de copas*?

o :

Al extraer la primera carta as, se tiene que $P_{(A)}$ = 4/40; luego al extraer la segunda rey de oros, se hará sobre un total de 39 cartas, por tanto $P_{(B)}$ = 1/39; luego, la tercera carta seis de copas, se tendrá que $P_{(C)}$ = 1/38. La probabilidad de que todos estos sucesos dependientes ocurran será igual a:

$$P(\,A\ y\ B\ y\ C\,) = P_{(A)} \times P_{(B)} \times P_{(C)} = \frac{4}{40} \times \frac{1}{39} \times \frac{1}{38} = \frac{4}{59.280} = \frac{1}{14.820} := 0{,}00006747$$

Generalmente se expresa lo anterior, de la siguiente manera, con el mismo resultado:

$$P(\,A\ y\ B\ y\ C\,) = P(\,A) \cdot P(B \mid A) \cdot P(C \mid A\ y\ B)$$

$$P(\,A\ y\ B\ y\ C\,) = \frac{4}{40} \times \frac{1}{39} \times \frac{1}{38} = \frac{4}{59.280}$$

Ejemplo 17. Supongamos que se tiene una caja con 10 monedas de $100, y dos de ellas son falsas. Se van a extraer dos monedas, una después de la otra sin reposición; ¿cuál es la probabilidad de seleccionar una moneda falsa seguida por otra también falsa?

o :

$P(A\ y\ B) = P(\,A) \cdot P(B \mid A)$ También se puede expresar así. $P(\,A \cap B) = P(\,A) \cdot P(B \mid A)$

$$P(A\ y\ B) = \frac{2}{10} \times \frac{1}{9} = \frac{2}{90} = \frac{1}{45} = 0{,}0222 = 2{,}22\%$$

> P ($A \cap B$) se denomina probabilidad conjunta de A y B; P(A) se denomina probabilidad marginal de A; P(B/A) es la probabilidad condicional de B respecto a A.

Ejemplo 18. Se extraen tres cartas sin reposición de una baraja de 40 cartas; ¿cuál es la probabilidad de que las tres sean zotas?

$$P(\,A_1 \cap A_2 \cap A_3\,) = P(\,A_1\,) \cdot P(\,A_2 / A_1\,) \cdot P(\,A_3 / A_1\ y\ A_2\,)$$

$$P\ A_1\ A_2\ A_3\ = \frac{4}{40} \times \frac{3}{3} \times \frac{2}{3} = \frac{24}{5\ \ 2\ \ 0} = \frac{1}{2\,470} = 0{,}000404 = 0{,}04$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

Sucesos mutuamente excluyentes y sucesos compatibles

87. Se extrae una carta al azar de una baraja de 40 cartas. ¿Cuál es la probabilidad de que sea As o figura?

88. Se encuentran reunidas 4 personas con diferentes profesiones: Abogado, Economista, Ingeniero, Administrador. Se elige una persona al azar. ¿Cuál es la probabilidad de que sea Abogado o Economista o Administrador?

89. Considere una baraja de 52 cartas y se desea extraer una carta. ¿Cuál es la probabilidad de obtener una que sea J o corazón?

90. Si en el ejercicio anterior se dijera, al extraer una carta, ¿cuál es la probabilidad de obtener una que sea diamante o trébol?

91. En un grupo de estudiantes la probabilidad de que tengan computador es de 0,60: auto de 0,30 y que tengan ambos, 0,25. ¿Cuál es la probabilidad de que un estudiante tenga computador o auto o ambas cosas?

92. Suponga que se tienen 30 fichas de tres colores así: amarillo, 15 fichas; negro, 10 fichas y azul, 5 fichas. Al mezclarlas, ¿Cuál es la probabilidad, al sacar una de ellas, de que sea: a) azul; b) azul o negra; c) amarilla o negra?

93. En una baraja de 40 cartas, al extraer una carta, ¿Cuál es la probabilidad de obtener una carta que sea figura o copas?

94. En un grupo de estudiantes la probabilidad de obtener un puntaje bajo es 0,20; que se haya graduado en la universidad es 0,5 y que se den ambos es 0,05. ¿Cuál es la probabilidad que un estudiante obtenga un puntaje bajo o se haya graduado en la universidad?

95. Si de un naipe bien barajado, de 40 cartas, se extrae una carta, ¿Cuál es la probabilidad de obtener:
    a) Un caballo o un rey b) Una zota de copas o un rey c) Una figura o copas
    d) Oros o un 6 e) Seis de espadas o figura f) Un as o figura

96. Se tiene una urna con 20 bolas de plástico distribuidas en los siguientes colores: 5 amarillas, 8 negras y 7 rojas. Extraiga una bola, teniendo el cuidado de revolverlas antes de extraerlas. ¿Cuál es la probabilidad de que la bola seleccionada?
    a) Sea negra b) No sea amarilla c) Sea roja d) Sea amarilla o negra

97. Suponga que: $P(A) = 0,20$ $P(B) = 0,70$ y $P(A \text{ y } B) = 0,10$
    a) ¿A y B son mutuamente excluyentes?
    b) Obtenga la probabilidad para P(A o B)
    c) ¿Son A y B colectivamente exhaustivos?

98. Supongamos una baraja de 52 cartas de la que debemos extraer una carta. Nos dan un premio si la carta extraída es trébol o K. ¿Cuál es la probabilidad de ganar?

99. a) Consideremos el lanzamiento de un dado. Usted gana, si el resultado es impar o divisible por dos. ¿Cuál es la probabilidad de ganar?
    b) Si planteamos el ejercicio de ganar obteniendo un resultado par o divisible por 3, ¿cuál sería la probabilidad de ganar?

100. La mamá lleva a su hijo a una tienda y le ofrece una de tres galguerías. La probabilidad de que escoja un helado es del 70%; kumis, 0,40 y helado y kumis, 0,30. ¿Cuál es la probabilidad de que compre helado o kumis?

101. En un día programado para realizar un paseo por el parque, la probabilidad de que haga sol es de 0,60; de que llueva, 0,20 y de que haga sol y llueva, es de 0,03. ¿Cuál es la probabilidad de que llueva o haga sol?

102. Si el Banco de la República exige que se rebaje la tasa de interés al 32%, existirá una probabilidad del 80% de que la inflación para ese año sea superior al 25%. ¿Qué interpretación le daría usted al 80%?

103. Se compraron 30 lápices de diferentes colores: 12 azules, 8 amarillos y 10 verdes. ¿Cuál es la probabilidad al extraer un lápiz de que sea: a) azul; b) azul o amarillo; c) amarillo o verde?

104. A un cargo se presentan 16 candidatos de diferentes profesiones; 6 economistas, 4 administradores 2 contadores y 4 ingenieros industriales. ¿Cuál es la probabilidad de que el cargo sea ocupado por un economista o un administrador?

Sucesos independientes

105. Qué probabilidad tendremos de obtener 2 reyes, sacando una carta de una baraja y la otra de una segunda baraja? (Ambas barajas son españolas).

106. En un solo lanzamiento de dos dados, ¿qué probabilidad tenemos de sacar dos cincos?

107. Juan y Grisel estudian en un mismo curso. La probabilidad de que Juan no pierda ninguna materia es de 0,8 y que Grisel obtenga el mismo resultado es de 0,90. ¿Cuál es la probabilidad (a) de que los dos no pierdan ninguna materia; (b) que Juan pierda por lo menos una y Grisel ninguna; (c) que ambos pierdan?

108. De una baraja de 40 cartas se van a extraer 3 cartas con reposición. ¿Cuál es la probabilidad de que la primera carta sea Rey y la segunda As y la tercera 6 de copas?

109. Supongamos que un taller dispone de dos máquinas. En la primera se produce el 1,5% de unidades defectuosas y en la segunda el 3%. ¿Cuál es la probabilidad de que al extraer una de cada máquina, las dos sean defectuosas?

110. ¿Cuál es la probabilidad, al seleccionar 5 personas en un grupo, de que las 5 sean mujeres?

111. Se lanzan 3 monedas, ¿cuál es la probabilidad de que las tres sean caras?

112. Cuatro amigos que se dirigen a un lugar, toman 4 rutas diferentes de acuerdo al riesgo que se corre de tener algún accidente. Si se le asignan las probabilidades de riesgo para cada ruta: 0,2 0,15 0,25 0,10. Encuentre la probabilidad:

    a) De que ninguno sufra dificultades
    b) Que los cuatro sufran accidentes
    c) Los dos primeros sufran accidentes y los restantes no

113. Al sacar dos cartas con reposición de una baraja de 52 cartas, cuál es la probabilidad de que:
a) Ambas sean diamantes; b) Ambas sean figuras (J, K); c) Ambas sean corazón y diamante

114. ¿Cuáles serían sus respuestas al ejercicio anterior, si las dos cartas se extraen sin reposición?

115. Un hombre posee un negocio y es, además, propietario de su casa. En un año cualquiera la probabilidad de que la casa sea robada es 0,08, y la probabilidad de que su negocio sea robado es 0,14. Suponiendo que estos eventos sean independientes, cuál es la probabilidad de que:
a) Sufra robos en ambos lugares en este año
b) No se presenten robos en ninguno de los dos

116. En forma independiente se lanza una moneda y se extrae una carta de una baraja de 52 cartas y se lanza un dado. ¿Cuál es la probabilidad de observar cara en la moneda, as en la carta y el tres en el dado?

117. De una baraja de 40 cartas se van a extraer 3 cartas con reposición. ¿Cuál es la probabilidad de que la primera carta sea un Rey, la segunda un As y la tercera un 6 de Copas?

118. Una fábrica tiene cuatro máquinas de diferentes modelos, la primera del año 2005 y una probabilidad del 12% de dañarse en un día de trabajo, otra de 2007, con el 7%; la tercera de 2009, con el 2% y la última de 2011 con el 1%, en un día de producción. Calcule la probabilidad de que:
a) Ninguna se descomponga o dañe b) Todas se descompongan.

119. Una máquina que produce un determinado artículo fue adquirida bajo la condición de que el 3% de los artículos producidos son defectuosos. Si el proceso se realiza bajo control, es decir independiente, ¿Cuál es la probabilidad de que:
a) dos artículos seguidos sean defectuosos?
b) dos artículos seguidos no sean defectuosos?
c) un artículo defectuoso y el otro bueno, en cualquier orden?
d) tres artículos seguidos sean buenos?

120. Tengo en el bolsillo del saco dos bolas de plástico una roja y otra verde. ¿Cuál es la probabilidad de sacar 3 veces sucesivas (con reposición) la bola roja?

121. En un recipiente se tienen 10 bolas azules y 5 rojas y en un segundo recipiente se tienen 8 bolas blancas y 12 rojas. ¿Cuál es la probabilidad de sacar bolas rojas, si extraemos una de cada recipiente?

122. Suponga que pintamos dos caras de un dado de rojo, tres de verde y una de azul. ¿Cuál es la probabilidad, al lanzar cuatro veces el dado, de obtener:
a) las tres primeras veces verde y la última rojo?
b) sólo las tres primeras rojo?
c) que en los tres primeros lanzamientos se obtenga azul?

Sucesos dependientes

123. Una persona saca de una baraja de 52 cartas sin reposición; simultáneamente otra persona lanza una moneda. ¿Cuál es la probabilidad de que la primera persona saque en compañía de la segunda tres cartas de la misma pinta y una cara?

124. Cuatro personas sacan sucesivamente una carta de un mazo de 40 sin reposición. ¿Cuál es la probabilidad de que la primera persona saque el As de espadas, la segunda el Rey de copas, la tercera un Rey y la cuarta un As?

125. En una bolsa hay seis bolitas blancas y cinco amarillas. Se sacan de una en una sin reposición. ¿Cuál es la probabilidad de que la primera sea blanca, la segunda amarilla, la tercera blanca y así sucesivamente?

126. Se extraen 3 cartas sin reposición de una baraja de 52 cartas. ¿Cuál es la probabilidad de que:
   a) en la primera carta aparezca una K y en las siguiente no aparezca?
   b) en la tercera carta aparezca la primera K?

127. ¿Cuál es la probabilidad de obtener 3 ases, sacando sucesivamente 3 cartas de una baraja de 40 cartas, sin volverlas a incluir en el mazo (montón)?

128. Suponga que se tiene una urna con veinte bolas, de las cuales, cinco son amarillas, ocho negras y siete rojas.
   a) Extraiga tres bolas sin reposición, ¿cuál es la probabilidad de que la primera sea amarilla, la segunda negra y la tercera roja?
   b) Si se hubiese pedido con reposición, ¿cuál sería la probabilidad anterior?

129. Se extraen tres cartas sin reposición de una baraja de 40 cartas.
   a) ¿Cuál es la probabilidad de que las tres sean ases
   b) ¿De que las tres sean oros?

130. Se extraen cinco cartas sin reposición de una baraja de 40 cartas, ¿Cuál es la probabilidad de que las cinco sean copas?

131. Dos equipos de balompié revuelven sus camisetas, 11 son de color verde y las restantes 11 azules. Se van a extraer 2 camisetas. ¿Cuál es la probabilidad de que las dos sean rojas?

132. Después de inspeccionar un lote de 10 unidades, se encontró que 2 son defectuosas y las restantes buenas. Si de los 10 extraemos dos unidades, ¿cuál es la probabilidad de que las dos sean defectuosas?

PROBABILIDAD CONDICIONAL

Los símbolos que se presentan a continuación son algunos de los más comunes en textos de estadística.

$P_{(A)}$	⇒	Probabilidad de que *ocurra* el suceso A
$P_{(A')} = P_{(\overline{A})}$	⇒	Probabilidad de que no *ocurra* el suceso A. También simboliza la probabilidad del complemento A
$P_{(A \mid B)}$	⇒	Probabilidad de que ocurra A *dado que* ha ocurrido B. También simboliza la *probabilidad condicional* de A dado B.
$P_{(A \cap B)}$	⇒	Probabilidad de que ocurra t*anto A como B* en un experimento. También simboliza la probabilidad de la intersección de A y B o la probabilidad *conjunta* de A y B.
$P_{(A \cup B)}$	⇒	Es la probabilidad de que ocurra A, *o bien* B, *o ambos*, en un experimento. También simboliza la probabilidad de la *unión* de A y B.

En nuestro tema, de la *probabilidad condicional* de A, dado B, se tendrá que: $P_{(A \mid B)} = \dfrac{P_{(A \cap B)}}{P_{(B)}}$.

En el caso contrario, es decir, la probabilidad condicional de $P_{(A\ B)} = \frac{P_{(A \cap B)}}{P_{(A)}}$

La raya vertical significa *dado que*.

- Recordemos que en la probabilidad de *sucesos mutuamente excluyentes* simbolizamos así: $P_{(A\,o\,B)} = P_{(A)} + P_{(B)}$ De acuerdo a las anotaciones se podrá decir también que $P_{(A \cup B)} = P_{(A)} + P_{(B)}$
- En los *sucesos compatibles* lo simbolizamos así: $P_{(A\,o\,B)} = P_{(A)} + P_{(B)} - P_{(A\,y\,B)}$
  Ahora lo podemos presentar: $P_{(A \cup B)} = P_{(A)} + P_{(B)} - P_{(A \cap B)}$
- En los sucesos independientes, utilizamos $P_{(A\,y\,B)} = P_{(A)}\,P_{(B)}$ Ahora se puede presentar que $P_{(A \cap B)} = P_{(A)}\,P_{(B\ A)} = P_{(B)}\,P_{(A\ B)}$

La *probabilidad condicional* es aquella que se presenta en un evento o suceso, dado que otro evento ya ha ocurrido.

Existe también la *probabilidad conjunta* cuando se presentan dos o más eventos en forma simultánea.

Todos se presentan bajo condiciones de dependencia estadística. No hay que olvidar que existen las probabilidades marginales, correspondientes a una probabilidad incondicional de que se presente un evento, se refiere a la probabilidad de un solo evento.

En la regla de la multiplicación, la probabilidad conjunta de A y B se calculaba mediante la aplicación de la fórmula $P_{(A\,y\,B)} = P_{(A)}\,P_{(A\ B)} = P_{(A \cap B)}$ de donde podemos despejar la fórmula para la probabilidad condicional de un evento.

$$P_{(A\ B)} = \frac{P_{(A\,y\,B)}}{P_{(B)}} = \frac{P_{(A \cap B)}}{P_{(B)}}$$

Ejemplo 1. Se encuentra en una facultad que del 70% de los alumnos matriculados, el 70% son mujeres y el 18%, mujeres estudiantes de economía. Si elegimos un estudiante al azar y resulta que es mujer, ¿cuál es la probabilidad de que esté estudiando economía?

o **:** $P_{(A\ B)} = \frac{P_{(A \cap B)}}{P_{(B)}} = \frac{0{,}18}{0{,}70} = 0{,}2571 = 25{,}71\%$ Probabilidad que este estudiando economia.

Ejemplo 2. Por una investigación reciente se encontró que el 10% de los conductores de taxi en la ciudad son hombres con estudios universitarios. También se sabe que el 80% de los conductores de taxi son hombres. ¿Cuál es la probabilidad, al tomar un conductor de taxi al azar, que resulte ser hombre, y que tenga además estudios universitarios?

o **:** $P_{(A\ B)} = \frac{P_{(A \cap B)}}{P_{(B)}} = \frac{0{,}10}{0{,}80} = 0{,}125 = 12{,}5\%$ Probabilidad que conduzca taxi y tenga estudios superiores.

Ejemplo 3. El 18% de las familias de un barrio tienen vehículo propio, el 20% tiene vivienda de su propiedad y el 12%, vivienda y vehículo. ¿Cuál es la probabilidad de tener vivienda si se tiene vehículo?

A = Propietario de vehículo

A' = No propietario de vehículo

B = Propietario de vivienda

B' = No propietario de vivienda

	B	B'	TOTAL
A	0,12	0,06	0,18
A'	0,08	0,74	0,82
TOTAL	0,20	0,80	1,00

$$P_{(B\ A)} = \frac{0,12}{0,18} = 0,66$$

$$P_{(B\ A)} = \frac{P_{(A \cap B)}}{P_{(B)}}$$

TEOREMA DE BAYES

El matemático y reverendo Thomas Bayes, (1763) en el siglo XVIII intentó desarrollar una fórmula para evaluar la probabilidad de la existencia de Dios con base en evidencias terrenales. Más tarde fue Laplace quien terminó su desarrollo denominándolo "Teorema de Bayes".

El *teorema de Bayes* es un método aplicado generalmente para calcular probabilidades posteriores.

Este teorema se aplica cuando se formulan hipótesis a posteriori sobre la probabilidad a priori de eventos ya ocurridos.

La fórmula general aplicable es:

$$P_{(A_i\ B)} = \frac{P_{(A_i)} P_{(B\,A_i)}}{P_{(A_1)} P_{(B\,A_1)} + P_{(A_2)} P_{(B\,A_2)} + \ldots\ldots\ldots + P_{(A_n)} P_{(B\,A_n)}}$$

Este teorema establece, que si sucede cierto evento, que depende de la ocurrencia de los eventos A o B o C correspondientes a un conjunto de sucesos mutuamente excluyentes, la probabilidad de que B haya ocurrido a consecuencia de A, lo cual lo expresamos: $P_{(A_i\ B)}$ corresponda al producto de las probabilidades individuales del evento A y del evento B, dividido por la probabilidad alternativa del evento B con respecto a cada uno de los eventos independientes de A, B y C.

La fórmula siguiente está dada en el caso que utilicemos las letras A, B y C.

$$P_{(A_i\ E)} = \frac{P_{(A)} P_{(E\,A_i)}}{P_{(A)} P_{(E_1\ A)} + P_{(B)} P_{(E_2\ B)} + P_{(C)} P_{(E_3\ C)}}$$

Ejemplo 1. Volvamos a nuestro ejemplo de las 4 máquinas A, B, C y D. Por especificaciones y control conocemos la capacidad de producción de cada máquina, durante un determinado período (1 hora) así: A. una producción de 600; B, de 400; C, de 300, y D, de 700 unidades, es decir, en términos porcentuales A produce el 30%, B el 20%, C el 15%, y D el 35%. Mediante un proceso de observación se ha detectado que el porcentaje de unidades defectuosas producidas por cada una de las máquinas es del 4%, 3%, 6% y 5%, respectivamente. Si procedemos a sacar un elemento del total del lote examinado. ¿Cuál es la probabilidad de que haya sido producida por la máquina A, o por la máquina B, o por la máquina C o por la máquina D?

o :

Aplicaremos la fórmula general, dada inicialmente y no la segunda expresión. Sabemos que la $P_{(A_1)} = 0{,}30$ $P_{(A_2)} = 0{,}20$ $P_{(A_3)} = 0{,}15$ $P_{(A_4)} = 0{,}35$

son las probabilidades independientes de A_1 A_2 A_3 y A_4 respectivamente.

Siendo: $P_{(BA_1)} = 0{,}04$ $P_{(BA_2)} = 0{,}03$ $P_{(BA_3)} = 0{,}06$ $P_{(BA_4)} = 0{,}05$

Podemos calcular:

$P_{(A_1)}P_{(BA_1)} = 0{,}30\ (0{,}04) = 0{,}012$ $P_{(A_2)}P_{(BA_2)} = 0{,}20\ (0{,}03) = 0{,}006$

$P_{(A_3)}P_{(BA_3)} = 0{,}15\ (0{,}06) = 0{,}009$ $P_{(A_4)}P_{(BA_4)} = 0{,}35\ (0{,}05) = 0{,}0175$

La suma de las probabilidades será = 0,012 + 0,006 + 0,009 + 0,0175 = 0,0445

Para calcular la probabilidad de que la unidad defectuosa haya sido producida por cada una de las 4 máquinas, utilizamos la fórmula de Bayes, así:

$$P_{(A_1|B)} = \frac{0{,}30\ (0{,}04)}{0{,}30\ (0{,}04) + 0{,}2\ (0{,}03) + 0{,}15\ (0{,}06) + 0{,}35\ (0{,}05)}$$

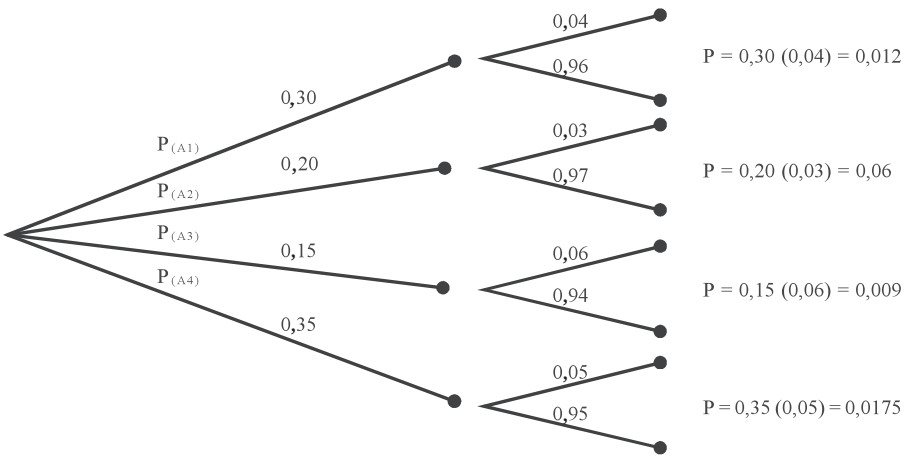
$$P_{(A_1|B)} = \frac{0{,}3\ (\ 0{,}04)}{0{,}0445} = 0{,}2\ \ 67 = 2\ \ ,67$$

$$P_{(A_2|B)} = \frac{0{,}20\ (\ 0{,}03)}{0{,}0445} = 0{,}1348 = 13{,}48$$

$$P_{(A_3|B)} = \frac{0{,}15\ (0{,}06)}{0{,}0445} = 0{,}2022 = 20{,}22$$

$$P_{(A_4|B)} = \frac{0{,}35\ (0{,}05)}{0{,}0445} = 0{,}3\ \ 33 = 3\ \ ,33$$

Podemos utilizar el *diagrama de árbol* en dos etapas



Ejemplo 2. Se tienen 3 recipientes; el primero contiene 6 bolas azules y 2 rojas; el segundo 4 azules y 4 rojas y el tercero 6 azules. Se selecciona una de las tres urnas al azar y de ella se extrae una bola que resulta ser azul. Con la anterior información. ¿Cuál es la probabilidad de que el recipiente escogido sea el primero?, Sea el tercero?

o :

$P_{(A_1)} \quad \frac{1}{3} \qquad P_{(A_2)} \quad \frac{1}{3} \qquad P_{(A_3)} \quad \frac{1}{3}$

$P_{(B|A_1)} \quad \frac{6}{8} \quad \frac{3}{4} \qquad P_{(B|A_2)} \quad \frac{4}{8} \quad \frac{1}{2} \qquad P_{(B|A_3)} \quad \frac{6}{6} \quad 1$

La probabilidad de que la bola azul provenga del primer recipiente será:

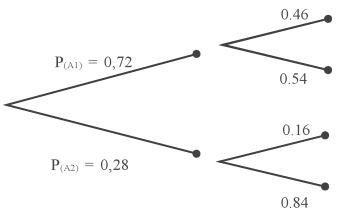
$$P_{A_1|B} \quad \frac{P_{(A_1)} P_{B|A_1}}{P_{(A_1)} P_{B|A_1} \quad P_{(A_2)} P_{B|A_2} \quad P_{(A_3)} P_{B|A_3}}$$

$$P_{A_1|B} \quad \frac{(1/3)(3/4)}{(1/3)(3/4) \quad (1/3)(1/2) \quad (1/3)(1)} \quad \frac{1}{3} \quad 0{,}33$$

La probabilidad de que provenga del tercer recipiente

$$P_{A_3|B} \quad \frac{(1/3)(1)}{(1/3)(3/4) \quad (1/3)(1/2) \quad (1/3)(1)} \quad \frac{4}{9} \quad 0{,}44$$

Ejemplo 3. Un autor, por intermedio de la editorial envía folletos promocionando su libro de estadística al 72% de los profesores que enseñan la asignatura en las universidades que fueron seleccionadas para la promoción. Un mes después se constató que el 46% que recibieron el folleto adoptaron el libro y un 16% que no lo recibieron, también lo adoptaron. ¿Cuál es la probabilidad de que un profesor que adopta el libro, fue el resultado del folleto de promoción?



$$P_{A_1|B} \quad \frac{0{,}72\ (0{,}46)}{0{,}72(0{,}46) \quad (0{,}28)(0{,}16)} \quad 0{,}8809$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

133. Un analista financiero descubrió que el 40% de las acciones experimentaron un comportamiento superior al promedio, el 18% inferior y el 42% se mantuvieron alrededor del promedio. El 40% del primer grupo fue considerado como buenas adquisiciones, lo mismo que 30% del segundo grupo y un 10% del último grupo. ¿Cuál es la probabilidad de que un valor correspondiente al primer grupo crezca por encima del promedio?

134. La probabilidad de que se presenten 3 eventos A, B, C son $P_{(A)}$= 0,15; $P_{(B)}$= 0,50; $P_{(C)}$ = 0,35. Supongamos que se han presentado A, B o C, la probabilidad que se presente otro evento D, es: $P_{(D/A)}$= 0,72; $P_{(D/B)}$= 0,4 y $P_{(D/C)}$= 0,60. Encuentre la probabilidad de $P_{(B/D)}$.

135. Una empresa utiliza tres métodos para recuperar la cartera morosa. El 50% es requerido telefónicamente, el 30% es visitado por el cobrador y el 10% por correo. Las probabilidades de que haya cancelación de la deuda o por lo menos abono a la misma, de acuerdo a los tres sistemas anteriores son: 0,62; 0,80 y 0,54 respectivamente ¿Cuál es la probabilidad de que la petición de pago haya sido efectiva, mediante el correo?

136. Con la siguiente información $P_{(Al)}$ = 0,6; $P_{(A2)}$ = 0,25; $P_{(A3)}$ = 0,15 y las probabilidades conjuntas de $P_{(Ai\ y\ B)}$ son: 0,09; 0,12 y 0,18 respectivamente. ¿Cuál será la probabilidad de $P_{(A2/B)}$?

137. En un curso de estadística se sabe que el 75 % realiza el taller en forma personal. También se sabe que el 92%, de los que hacen el taller en forma individual, gana el curso y el 40% lo pierde cuando el taller lo copian o no lo hacen en forma individual. ¿Cuál es la probabilidad que él haya hecho sus tareas individualmente?

138. Use el teorema de Bayes para determinar la probabilidad de $P_{(A3/B)}$ si se tiene la siguiente información $P_{(Al)}$ = 0,18; $P_{(A2)}$ = 0,46 y $P_{(A3)}$ = 0,36, además, las probabilidades condicionales son $P_{(B/Al)}$= 0,21; $P_{(B/A2)}$= 0,08 y $P_{(B/A3)}$= 0,14

139. Supongamos que se tienen dos recipientes A_1, y A_2, en el primero se tienen 26 bolitas, de las cuales 10 son azules y 16 verdes; y en el segundo son 26, distribuidas así: 6 azules y 20 verdes. Si se elige al azar una urna y de ella se extrae una bolita, ¿cuál es la probabilidad de que sea verde?

140. De acuerdo con el punto anterior. ¿Cuál es la probabilidad de que sea verde proveniente del recipiente A_1?

141. De acuerdo al ejercicio 139. ¿Que sea la bolita verde proveniente del recipiente A_2?

EJERCICIOS MISCELÁNEOS

142. Se lanzan dos dados correctos. Calcular la probabilidad del suceso: “la suma de sus caras sea mayor a 8”.

143. Se extrae al azar una carta de una baraja de 40 cartas. Sea A el suceso de que la carta elegida es un Rey y B el suceso de que la carta elegida sea Copas. Calcular la probabilidad de que uno de los sucesos ocurra.

144. ¿Cuántas combinaciones posibles pueden hacerse con 52 cartas de una baraja, tomadas de 4 en 4?

145. En una carrera de caballos participan 10 caballos. La apuesta consiste en escoger tres caballos que pasen en los primeros lugares a la meta.
    a) ¿Cómo se daría un caso de permutaciones? b) ¿Cómo se daría un caso de combinaciones?

146. Se lanza un dado rojo y otro verde.
   a) ¿Cuál es la probabilidad de que habiendo salido el 5, en el dado Rojo, la suma sea mayor que 10?
   b) ¿Cuál es la probabilidad de que habiendo salido el dos, en el dado Verde, la suma sea menor que 7?
   c) ¿Cuál es la probabilidad de obtener 7 como suma, sabiendo que en el dado Rojo aparece un número menor de 4?

147. Se tiene una baraja de 40 cartas y se sacan cinco de ellas sin reposición. ¿Cuál es la probabilidad que
   a) las 3 primeras cartas sean ases y las dos últimas zotas?
   b) las 3 primeras ases, la cuarta 4 de oros y la quinta un rey?

148. Se lanza tres veces una moneda correcta:
   a) ¿Cuál es la probabilidad de que únicamente en el tercer lanzamiento resulte cara?
   b) ¿Cuál es la probabilidad de que en el tercer lanzamiento resulte cara, sabiendo que en uno de los dos anteriores se obtuvo cara?

149. ¿De cuántas maneras puede formarse un comité de 5 personas teniendo un total de 9 personas?

150. ¿Cuántos números de 5 cifras se pueden formar con los dígitos: 2, 4, 6, 8 y 9 sin repetirlos?

151. ¿Cuántos números de tres dígitos existen?

152. ¿Cuántas placas para vehículos se pueden obtener empleando tres letras seguidas de 3 dígitos?

153. a) ¿Cuántas palabras de 5 letras se pueden formar con las letras de la palabra LOBOS?
   b) ¿En el caso, sin usar ninguna O?

154. ¿De cuántas maneras se pueden sacar dos manzanas de una caja que contiene 8?

155. Diga cierto o falso en las siguientes aseveraciones:
   a) Una permutación es un caso particular de las combinaciones.
   b) ABC = CBA = BCA es un caso de combinación.
   c) $V_2^4 = 8$
   d) $C_3^5 = C_2^5$
   e) Una permutación se diferencia de otra por la posición de los elementos.
   f) Combinaciones de elementos, tomados de *n* en *n*, es igual a 1.

156. Si la sexta válida del 5 y 6 corresponde al juego del marcador de 5 caballos, ¿cuántos marcadores diferentes resultarán con once caballos, teniendo en cuenta el orden de llegada?

157. La probabilidad de que Grisel estudie para un examen de estadística es 0,30. Si estudia, la probabilidad de aprobar el examen es 0,75, en tanto que si no estudia, la probabilidad es de 0,40.
   a) ¿Cuál es la probabilidad de que Grisel apruebe su examen final?
   b) Dado que Grisel aprobó el examen ¿Cuál es la probabilidad de que haya estudiado?

158. Se piensa elegir una Junta Directiva compuesta por un Presidente, un Tesorero y un Secretario. ¿Cuántas planchas diferentes se podrán postular, si la asamblea está compuesta por 20 miembros y todos tienen igualdad de derecho para pertenecer a la Junta?

159. Un equipo capitalino juega el 60% de sus partidos de día. El equipo gana el 40% de sus partidos diurnos y el 80% de los nocturnos. De acuerdo a las noticias que aparecen en un diario de la capital, perdió ayer. ¿Cuál es la probabilidad de que el partido se haya desarrollado de día?

160. Un fabricante produce 6.000 unidades por semana. Cada unidad pasa por tres puestos de inspección, antes de ser despachada a sus distribuidores. Por lo general, en el primer puesto se rechaza el 5%; de las que fueron aceptadas, el 3% se rechaza en la segunda inspección y finalmente en la tercera inspección el 2% aproximadamente. ¿Cuál es la probabilidad que una unidad seleccionada al azar pase las tres inspecciones?

161. Supongamos que un proyecto de ley, debe ser aprobado en primera instancia por la Cámara de Representantes, luego por el Senado y finalmente recibe la sanción Presidencial, para convertirse en Ley. Un politólogo asevera que las probabilidades son: 0,8; 0,6 y 0,5 respectivamente. ¿Cuál es la probabilidad que el proyecto finalmente se convierta en Ley?

162. Se tienen tres ciudades localizadas en el Eje cafetero y se quiere seleccionar una de ellas para el desarrollo de un evento deportivo. Por estudios anteriores se sabe que Pereira tiene una probabilidad del 42%, Armenia del 34% y Manizales del 24%. El comité organizador del evento supone que al ser consultado a un grupo mayor de personas, las probabilidades de favorabilidad en cada región será 55%; 60%; 62% respectivamente ¿Qué ciudad tendrá la mayor probabilidad de ser escogida?

163. En la última temporada de vacaciones, se seleccionó una muestra de 400 turistas de los cuales 240 visitaron Santa Marta y 100 a Cartagena ¿Cuál es la probabilidad de que una persona seleccionada haya visitado durante la temporada las dos ciudades de la Costa Atlántica? Una verificación de la encuesta encuentra que 70 de ellos habían visitado las dos ciudades.

164. En una Facultad de una Universidad se sabe que el 35% están matriculados en el diurno y el 65% son del nocturno. La probabilidad de que el estudiante diurno esté trabajando es del 15%, en tanto, para el estudiante nocturno es del 70%. Se elige al azar un estudiante de esa Facultad.
    a) ¿Cuál es la probabilidad de que esté trabajando?
    b) Dado que el estudiante elegido esté trabajando, determinar la probabilidad que sea del diurno.

165. Tres distribuidores de gas se reparten el mercado de una ciudad, al distribuidor A le corresponde el 50%, al B, el 30% y al C únicamente el 20%. Las autoridades locales hacen una inspección en cada una de las distribuidoras y encuentran que en A el 5% de las válvulas de los tanques están defectuosos, en B el 3% y en C es del 8%. Suponiendo que la distribución no está demarcada por zonas, se presenta un escape con las consecuencias de una explosión que produce daños. ¿Cuál es la probabilidad de que el tanque causante del daño haya sido suministrado por el distribuidor A, B o C?

166. Defina brevemente los siguientes puntos:
    a) Probabilidades previas b) Probabilidades posteriori
    c) Diagrama de árbol d) Teorema de Bayes.

Síntesis de capítulo

Hablamos de *probabilidad* cuando hay *incertidumbre* o no se tiene la seguridad total de que pueda suceder. Es un concepto básico que se interpreta como algo indefinible, expresado de alguna manera como el grado de creencia de que algo puede suceder.

A menudo, sin fundamento matemático, se usa como una forma conveniente de representar la opinión de una persona con respecto a la posibilidad de que se obtenga un resultado determinado, lo que llamamos como *probabilidad subjetiva*. En otras oportunidades hablamos de *probabilidades a priori* donde no se requiere la realización del experimento, ya que se puede determinar de antemano. Se tiene la *probabilidad empírica* que implica la realización del experimento para ser calculada.

La *probabilidad relativa*, se refiere en los casos cuando observamos que ocurrieron sucesos semejantes en el pasado y queremos determinar la probabilidad de que ocurra ese suceso a largo plazo, es decir, el número de veces que el evento ocurrió en el pasado, dividido, por el número total de observaciones.

Existen unas reglas que pueden ayudar a su cálculo, tales como la de *adición* cuando los sucesos son *mutuamente excluyente*s; la *multiplicación* cuando los *sucesos son independientes* o *dependientes*; se usan también las *reglas de conteo*, con el fin de determinar el número de formas de posible ocurrencia, en especial cuando el número de sucesos posibles es grande.

Se dice que la probabilidad de ocurrencia de un evento es un número comprendido entre 0 y 1. En la mayoría de los casos, esa cifra se obtiene al dividir el número total de éxitos o casos favorables por el total de casos posibles.

El investigador matemático Thomas Bayes estableció una fórmula, conocida como el Teorema de Bayes, con el fin de determinar la probabilidad de que ocurra un evento A, dado que ha ocurrido un evento B.

El estudio de la probabilidad en su aplicación, se inició en los juegos de azar, marcando el camino para trabajos más serios; como se puede observar en los últimos años, por el gran auge que ha tomado una de sus ramas conocida como Teoría de los Juegos, de gran aplicación donde se requiere del planeamiento para afrontar situaciones donde no se tiene información o se dispone de alguna, por lo general, de poca confiabilidad.

6 CAPÍTULO

Distribuciones de probabilidad

Tener conocimiento de la propia ignorancia es un gran paso acia el conocimiento.

a a

CONTENIDO

- Conceptos y aplicaciones teórico-prácticos de las distribuciones de probabilidad.
- Variable aleatoria discreta: Distribución Binomial, de Poisson e Hipergeométrica.
- Variable aleatoria continua: Distribución normal.
- Manejo de Tablas, Calculadora y EXCEL para la realización de cálculos.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el Sistema de Información en Línea SIL.

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Explicar y diferenciar las distribuciones de frecuencias de las distribuciones de probabilidad.
- Definir e identificar las variables aleatorias discretas y continuas.
- Identificar las condiciones que se debe reunir, en la aplicación de cada una de las distribuciones de probabilidad.
- Construir, aplicar y calcular probabilidades asociadas a estas distribuciones.
- Manejar tablas, calculadora, Excel y fórmulas en la solución de ejercicios.

CONCEPTOS GENERALES

Los capítulos del dos al cuatro, fueron dedicados a explicar algunos temas referentes a la *Estadística Descriptiva* en cuanto al proceso de recolección de datos, elaboración de tablas o cuadros, la representación de ellos por medio de gráficas y la aplicación de algunas medidas, tales como las de: *posición, dispersión, asimetría* y *apuntamiento*.

Entre las medidas vistas, las más importantes a considerar fueron la *Media aritmética*, la *Varianza* y la *Desviación estándar* o *típica*. La Primera de ellas tenía como fin primordial, calcular un solo valor para representar a ese conjunto de observaciones y la Desviación típica para permitir determinar la posible dispersión o el grado de variabilidad que ocurre entre esos datos. Recordemos, que a menor variabilidad se presente, mayor será la representatividad del promedio.

Veamos algunas definiciones de términos que se deben entender y manejar.

- Distribución de probabilidad. Son todos los posibles valores que resultan de un experimento aleatorio, junto con la probabilidad asociada a cada valor.

  En el capítulo 3 se vio el procedimiento a seguir en la elaboración de las *Distribuciones de frecuencias*, con base en informaciones de hechos que ya habían sucedido, recolectados a través de una encuesta. En este capítulo, consideraremos la *Distribución de probabilidades* la cual se basa en hechos que probablemente pueden suceder, es decir, estamos hablando de dos situaciones correspondiente a hechos que suceden: *antes y después*.

Tabla 6.1 Distribución de frecuencias

y_i	n_i	$y_i n_i$	$(y_i - \bar{y}_i)^2 n_i$
1	3	3	27,5427
2	5	10	0,6045
4	10	40	0,0090
5	8	40	7,5272
7	4	28	35,2836
Σ	30	121	90,9670

Tabla 6.2 Distribución de probabilidad

X_i	$p_{(x)}$	$X_i P_{(x)}$	$(X_i - \mu)^2 P_X$
0	1/16=0,0625	0	0,2500
1	4/16=0,2500	0,25	0,2500
2	6/16=0,3750	0,75	0
3	4/16=0,2500	0,75	0,2500
4	1/16=0,0625	0,25	0,2500
Σ	16/16=1,000	2,00	1,0000

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{121}{30} = 4{,}03 \qquad \mu = \Sigma X_i P_x = 2$$

$$s^2 = \frac{\Sigma (y_i - \bar{y})^2 n_i}{n} = \frac{90{,}967}{30} = 3{,}03 \qquad \sigma^2 = \Sigma (X_i - \mu)^2 P_X = 1$$

$$s = \sqrt{3{,}03} = 1{,}74 \qquad \sigma = \sqrt{1{,}0} = 1$$

Recordemos que la *Distribución de frecuencias* permitía el manejo de cantidades de información en forma resumida, describiendo un hecho transcurrido (pasado), en cambio la *distribución de probabilidades* es similar a una distribución de frecuencias, con la diferencia de que presenta valores probables de un experimento en un evento futuro.

- Variable aleatoria. Corresponde a una caracterización cualitativa de los resultados que constituyen un espacio muestral. Cada cantidad o valor es el resultado de un experimento aleatorio y, como tal, puede tomar distintos valores.

  Las variables aleatorias se clasifican en d*iscretas* y *continuas.*

- Variable aleatoria discreta. Se considera así, cuando los valores que asume se puedan contar, y éstos pueden organizarse en una secuencia al igual que los números enteros positivos. Sólo podrá asumir un número finito de valores.

- Variable aleatoria continua. Se da, cuando puede asumir cualquier valor dentro de un intervalo o en una unión de intervalos. Como ejemplo se podría considerar cualquier resultado en cuanto de medición al ancho, longitud de una cosa, así como a tiempo en la realización de una tarea; en estos casos las variables admiten fracciones.

> Dentro de los modelos de probabilidad, correspondiente a variables aleatorias discretas, con mayor aplicación se tienen: *Bernoulli, Binomial, Poisson, Exponencial, Multinomial e Hipergeométrico* y en cuanto a la variable aleatoria continua se considera el modelo *normal estandarizado*.

A continuación se verán algunas de estas distribuciones: B*inomial, Poisson, Hipergeométrica y Normal.*

VARIABLE ALEATORIA DISCRETA

DISTRIBUCIÓN BINOMIAL

Algunos la denominan como *método exacto* y como ya se mencionó, corresponde a una distribución de *variable aleatoria discreta*. Pero antes de comenzar a explicar en qué consiste, cómo se calcula y en qué casos se debe aplicar, vale la pena recordar algunos conocimientos, que nos puedan ser útiles en la distribución binomial: el *Binomio de Ne ton y el Triángulo de Pascal. (Ver SIL., capítulo 6).*

Se puede considerar "que si durante repetidos ensayos, siendo p la probabilidad de éxito en un solo ensayo, la cual debe permanecer fija, y q la probabilidad de fracaso, entonces la probabilidad P de que se obtenga *X* éxitos en *n* ensayos, es el término del desarrollo binomial de $(q+p)^n$ ". La fórmula general para cada término es:

$$P_{(x)} = \binom{n}{x} p^x \, q^{n-x}$$

n = Número de ensayos
X = Número de éxitos
p = probabilidad de éxito en cada ensayo
q = probabilidad de fracaso en cada ensayo

Los criterios que debe satisfacer una experiencia binomial son:

a) Debe existir un número fijo de pruebas repetidas (n)
b) Cada una de las n pruebas debe tener dos resultados, favorable o desfavorable. En el caso de la moneda, será: cara o sello, por lo tanto son sucesos mutuamente excluyentes.
c) La probabilidad de éxito de un acontecimiento es fijo, algo similar sucede con la probabilidad de fracaso.
d) Las pruebas son independientes, ya que el resultado de un ensayo no afecta el resultado de otro.
e) Nos interesa el número de éxitos en n pruebas.

Ahora procedemos a elaborar el espacio muestral para el experimento consistente en lanzar cuatro monedas (n = 4), con el consiguiente resultado.

Tabla 6.3 Espacio muestral en el lanzamiento de 4 monedas

0 caras	1 cara	2 caras	3 caras	4 caras
ssss	*csss*	*ccss*	*cccs*	*cccc*
	scss	*cscs*	*ccsc*	
	sscs	*sccs*	*cscc*	
	sssc	*scsc*	*sccc*	
		sscc		
		cssc		
$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Consideremos como éxito la aparición de cara (C) y como fracaso la obtención de sello (S). Supondremos que las probabilidades de C y S son p y q respectivamente, siendo p + q = 1. Así tenemos que $P_{(c)} = p$ y $P_{(s)} = q$. Por otra parte, se debe considerar que C y S se presentan independientemente; por tal razón, cualquiera de las puntos muestrales, la probabilidad de que ocurran todos estos sucesos en un solo ensayo, se obtiene multiplicando las probabilidades para cada suceso. Supongamos que en el lanzamiento de 4 monedas, se quiera obtener éxito (cara) en la primera y tercera moneda y fracaso (sello) en la segunda y cuarta; se tendrá que:

$$P_{(cscs)} = P_{(c)}\ P_{(s)}\ P_{(c)}\ P_{(s)} = p\ q\ p\ q = p^2\ q^2$$

Veamos las probabilidades asignadas a los 16 puntos muestrales del experimento anterior al lanzar cuatro monedas.

Tabla 6.4 Espacio muestral con asignación de probabilidades para cada suceso

$$
\begin{aligned}
P_{(ssss)} &= qqqq = q^4 = q^4 \\
\left.\begin{aligned} P_{(sssc)} &= qqqp = q^3p \\ P_{(sscs)} &= qqpq = q^3p \\ P_{(scss)} &= qpqq = q^3p \\ P_{(csss)} &= pqqq = q^3p \end{aligned}\right\} &= 4q^3p \\
\left.\begin{aligned} P_{(sscc)} &= qqpp = q^2p^2 \\ P_{(scsc)} &= qpqp = q^2p^2 \\ P_{(sccs)} &= qppq = q^2p^2 \\ P_{(cscs)} &= pqpq = q^2p^2 \\ P_{(cssc)} &= pqqp = q^2p^2 \\ P_{(ccss)} &= ppqq = q^2p^2 \end{aligned}\right\} &= 6q^2p^2 \\
\left.\begin{aligned} P_{(sccc)} &= qppp = q\,p^3 \\ P_{(cscc)} &= pqpp = q\,p^3 \\ P_{(ccsc)} &= ppqp = q\,p^3 \\ P_{(cccs)} &= pppq = q\,p^3 \end{aligned}\right\} &= 4qp^3 \\
P_{(cccc)} &= ppqq = p^4 = p^4
\end{aligned}
$$

Si X representa al número de éxitos (caras) que se desea ocurran, se podrá elaborar la siguiente tabla de probabilidades. Siendo: $p = 1/2$ y $q = 1/2$.

Tabla 6.5 Distribución de probabilidades

X	0	1	2	3	4 ← *No. de caras*
$P_{(x)}$	q^4	$4q^3p$	$6q^2p^2$	$4qp^3$	p^4
$P_{(x)}$	$(1/2)^4$	$4(1/2)^3(1/2)$	$6(1/2)^2(1/2)^2$	$4(1/2)(1/2)^3$	$(1/2)^4$
$P_{(x)}$	$\frac{1}{6}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$ $= \frac{16}{16} = 1$

Recordemos el Binomio de Newton y observemos lo que pasa, si reemplazamos las notaciones clásicas de *a* y *b* por *q* y *p*, que representan las probabilidades de sello y cara.

$$(a+b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{1\times 2}a^{n-2}b^2 + \frac{n\ n-1\ \ n-2}{1\times 2\times 3}a^{n-3}b^3 \quad + b^n$$

La expansión binomial estará expresada de la siguiente forma:

$$q+p\ ^n = \binom{n}{0}q^n + \binom{n}{1}q^{n-1}p + \binom{n}{2}q^{n-2}p^2 + \quad + p^n$$

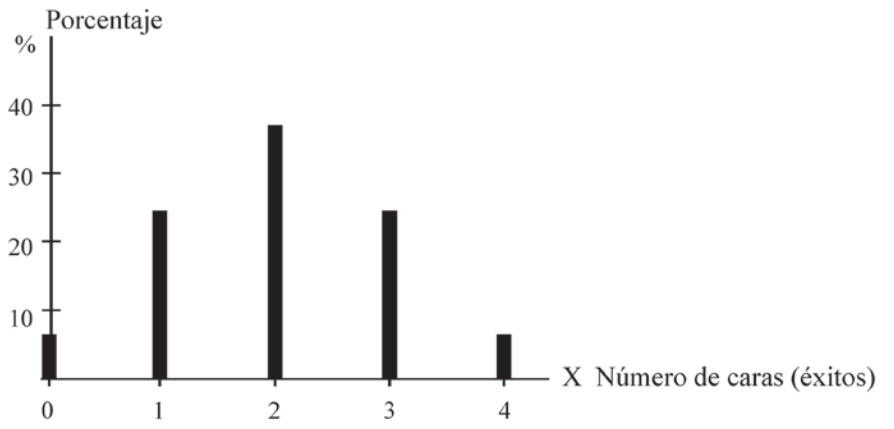
Distribución de probabilidad binomial y su representación gráfica

Elaboremos la distribución de probabilidad, con el experimento consiste en el lanzamiento de cuatro (4) monedas, para la cual la variable aleatoria discreta esté dada por la obtención de exactitud: cero, uno, dos, tres y cuatro caras (X = 0, 1, 2, 3 y 4)

Tabla 6.6 Distribución de probabilidades

X	PROBABILIDADES		
	Fracción	Decimal	%
0	1/16	0,0625	6,25
1	4/16	0,2500	25,00
2	6/16	0,3750	37,50
3	4/16	0,2500	25,00
4	1/16	0,0625	6,25
Σ	16/16	1,0000	100,00

Figura 6.1 Diagrama de frecuencias



Ejemplo 1. Supongamos el lanzamiento de 12 monedas o una moneda lanzada 12 veces. ¿Cuál es la probabilidad de obtener: (a) exactamente 4 caras; (b) exactamente 10 caras?; (c) por lo menos dos caras; (d) como máximo 9 caras; (e) más de 10 caras y menos de 3 caras?

o :

a) Exactamente 4 caras

$p=\frac{1}{2}$ $q=1-\frac{1}{2}=\frac{1}{2}$ $n=12$ X = 4 éxitos $P_{(x)}=\binom{n}{x}p^x\,q^{n-x}$

$$P_{(x=4)}=\binom{12}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^8=\left[\frac{12!}{(12-4)!\,4!}\right]\left[\frac{1}{16}\right]\left[\frac{1}{256}\right]$$

$$P_{(x=4)}=\left[\frac{12!}{8!\,4!}\right]\left[\frac{1}{4.096}\right]=\left[\frac{12\times11\times10\times9\times\cancel{8!}}{\cancel{8!}\times4\times3\times2\times1}\right]\left[\frac{1}{4.096}\right]=495\left(\frac{1}{4.096}\right)$$

$$P_{(x=4)}=\frac{495}{4.096}=0{,}1208=12{,}08\%$$

$P_{(x=4)}=12{,}08\%$

Nota: Con calculadora, se hubiese hecho más rápido el procedimiento anterior, algo similar se logra con el uso de tablas o la aplicación de Excel, como se verá después de desarrollar estos ejercicios.

b) Exactamente10 caras

$p=0{,}5$ $q=1-0{,}5=0{,}5$ $n=12$ X = 10 $P_{(x)}=\binom{n}{x}p^x\,q^{n-x}$

$$P_{(x=10)}=\binom{12}{10}(0{,}5)^{10}(0{,}5)^2$$

$$P_{(x=10)}=\left[\frac{12!}{(12-10)!\;10!}\right]\left[(0{,}0009766)(0{,}25)\right]$$

$P_{(x=10)}=66\ (0{,}00024415)=0{,}0161=1{,}61\%$ $P_{(x=10)}=1{,}61\%$

C) Por lo menos dos caras = Como mínimo dos caras

X= 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 y 12 p = 0,5 q = 0,5

$P_{(x>1)}=P_{(x\geq2)}=P_{(x=2)}+P_{(x=3)}+P_{(x=4)}+\ldots\ldots\ldots P_{(x=12)}$

Recordemos que $\sum_{i=1}^{n}\binom{n}{x}p^x q^{n-x}=1$ o sea 100%, por lo tanto se tendrá que:

$P_{(x)}=\binom{n}{x}p^x q^{n-x}$

$P_{(x\geq2)}=1-\left[P_{(x=0)}+P_{(x=1)}\right]$

$$P_{(x\geq2)}=1-\left[\binom{12}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{12}+\binom{12}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^{11}\right]$$

$P_{(x\geq2)}=1-[0{,}0002+0{,}0029]=1-0{,}0031=0{,}9969=99{,}69\%$ $P_{(x\geq2)}=99{,}69\%$

d) Como máximo 9 caras = No más de 9 caras

X = 0, 1, 2, 3, 4, 5, 6, 7, 8 y 9

$$P_{(x<10)} = P_{(x\leq 9)} = P_{(x=0)} + P_{(x=1)} + P_{(x=2)} + \ldots\ldots\ldots + P_{(x=9)} \qquad P_{(x)} = \binom{n}{x} p^x q^{n-x}$$

Desarrollándolo con el complemento se tendrá que:

$$P_{(X\leq 9)} = 1 - [P_{(X=10)} + P_{(X=11)} + P_{(X=12)}]$$

$$P_{(x\leq 9)} = 1 - \left[P_{(x=10)} + P_{(x=11)} + P_{(x=12)}\right]$$

$$P_{(x\leq 9)} = 1 - \left[\binom{12}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{2} + \binom{12}{11}\left(\frac{1}{2}\right)^{11}\left(\frac{1}{2}\right)^{1} + \binom{12}{12}\left(\frac{1}{2}\right)^{12}\left(\frac{1}{2}\right)^{0}\right]$$

$$P_{(x\leq 9)} = 1 - [0{,}0161 + 0{,}0029 + 0{,}0002]$$

$$P_{(x\leq 9)} = 1 - 0{,}0192 = 0{,}9808 = 98{,}08\%$$

$$P_{(x\leq 9)} = 98{,}08\%$$

e) Más de 10 caras y menos de 3 caras

X = 11 12 0 1 2

$$P_{(10<x<3)} = P_{(11\leq x\leq 2)} = P_{(11)} + P_{(12)} + P_{(0)} + P_{(1)} + P_{(2)}$$

$$P_{(11\leq x\leq 2)} = \binom{12}{11}(0{,}5)^{11}(0{,}5)^{1} + \binom{12}{12}(0{,}5)^{12}(0{,}5)^{0} + \binom{12}{0}(0{,}5)^{0}(0{,}5)^{12}$$
$$+ \binom{12}{1}(0{,}5)^{1}(0{,}5)^{11} + \binom{12}{2}(0{,}5)^{2}(0{,}5)^{10}$$

$$P_{(11\leq x\leq 2)} = 0{,}0029 + 0{,}0002 + 0{,}0002 + 0{,}0029 + 0{,}0161 = 0{,}0223 = 2{,}23\%$$

$$P_{(11\leq x\leq 2)} = 2{,}23\%$$

Ejemplo 2. Una institución universitaria establece nuevos métodos de aprendizaje y de evaluación, con el resultado donde el 85% de sus alumnos aprueban todas las asignaturas. Supongamos que se seleccionan 8 estudiantes de dicho plantel ¿Cuál es la probabilidad: (a) exactamente 3 aprueben todas las asignaturas; (b) exactamente 3 pierdan alguna asignatura; (c) por lo menos dos pierdan alguna asignatura?

o ***:*** (a) Exactamente 3 aprueben

p = 0,85 (aprobar) n = 8 X = 3

$$P_{(=3)} = \binom{8}{3}(0{,}85)^{3}(0{,}15)^{5} = 0{,}0026 = 0{,}26$$

Se pueden utilizar las tablas de binomial que aparecen al final del libro correspondiente a la tabla III

- Se busca en la primera columna de la tabla que aparezca el valor de n = 8 (Ver tabla 6.5).
- Luego en la segunda columna el valor de X.
- Se localiza 0,85 en el renglón de p, observemos que se llega a 0,50, por lo tanto se toma el complemento de 0,85 siendo 0,15.
- Observemos en la tabla al frente de 5 y en la columna 0,15 aparece 0,0026.

Como nos piden que 3 aprueben, es lo mismo que decir, 5 pierdan alguna asignatura, siendo: n = 8 p = 0,15 X = 5

$$P_{(x=5)} = \binom{5}{5} (0,15)^3 (0,85)^5 = 0,0026 = 0,26\%$$

Como observamos el resultado es igual al anterior

b) Exactamente 3 pierdan alguna asignatura

$n = 8$ $p = 0,15$ $q = 0,85$ X = 3

$$P_{(x=3)} = \binom{8}{3} (0,15)^3 (0,85)^5 = 0,0839 = 8,39\% .$$

De acuerdo a la Tabla 6.5, al frente de 3 y debajo de la columna 0,15 aparece el valor de 0,0839

Tabla 6.5 Distribución binomial. Cuando n = 8

n	x	*p*									
		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3825	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0312
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2000	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039

c) Por lo menos dos pierdan alguna asignatura

$$P_{(x\geq 2)} = 1 - \left[P_{(x=0)} + P_{(x=1)}\right] \quad \text{(Observar Tabla 6.5)}$$

$$P_{(x\geq 2)} = 1 - \left[0,2725 + 0,3847\right]$$

$$P_{(x\geq 2)} = 1 - 0,6572 = 0,3428 = 34,28\%$$

Al final del libro se presenta una tabla donde p va desde .05 hasta .95 lo cual ayuda a simplificar el proceso.

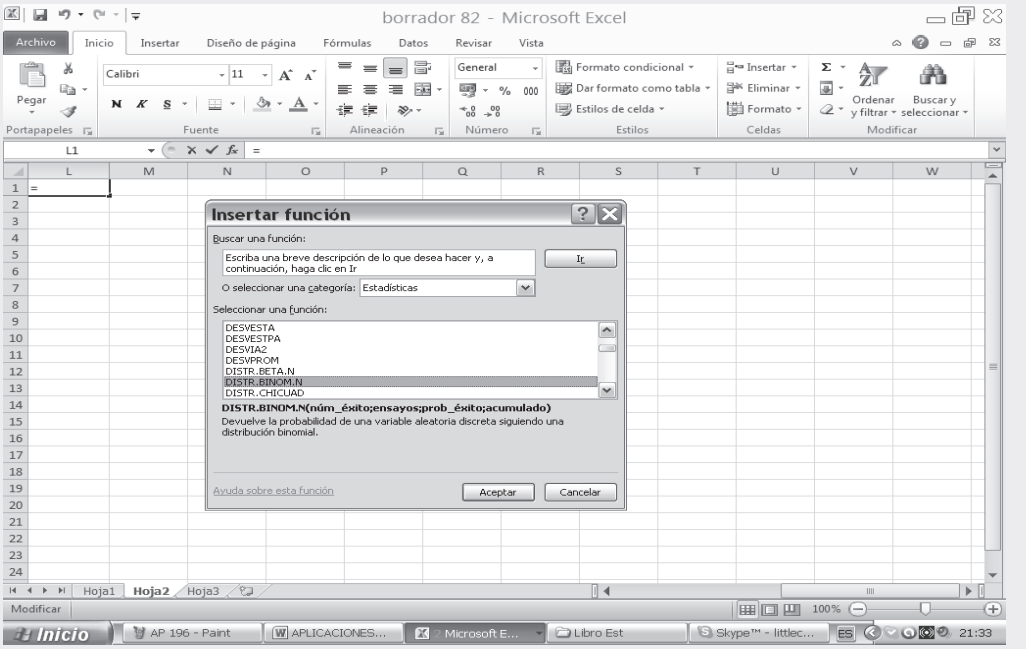
APLICACIÓN DEL EXCEL

Complemento virtual SIL

DISTRIBUCIÓN BINOMIAL

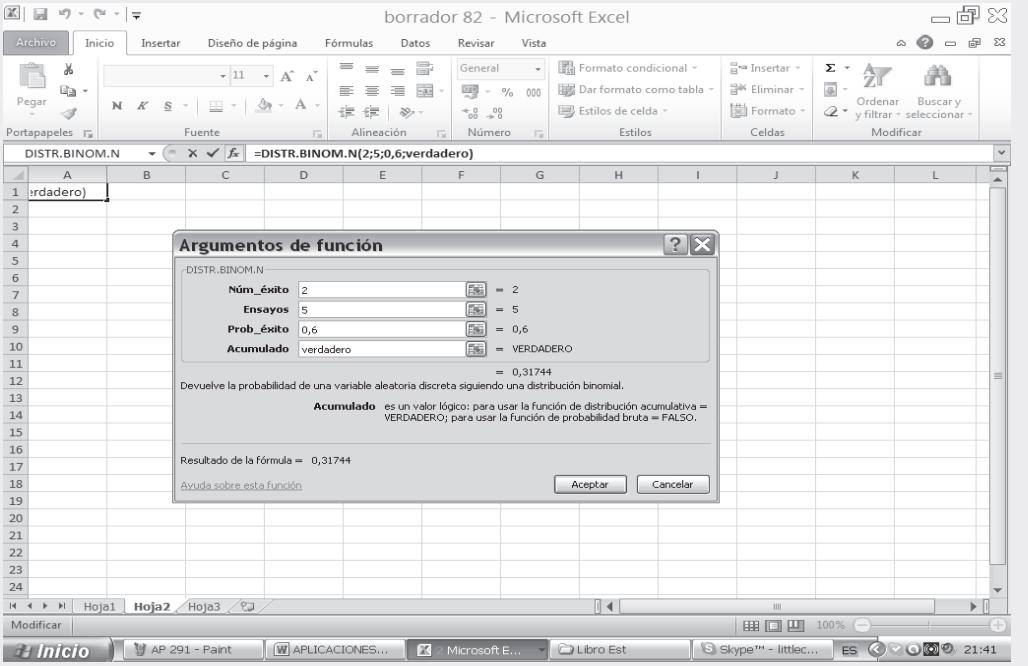
- Nos ubicamos en *f*x haciendo CLIC en DIST.BINOM y nuevamente CLIC en ACEPTAR.
- Aparece la figura correspondiente a DIST.BINOM. N, allí nos ubicamos en NUM-ÉXITO en esta celda digitamos el valor de X=2 (el número de éxitos es 2).
- Luego en la celda correspondiente a ENSAYOS, digitamos el valor n siendo en este caso igual a 5 (número de repeticiones).
- Se procede luego a digitar el valor correspondiente a PROB_ÉXITO, siendo igual a 0.6 (p=0.6).

Figura No. 1. Insertar Función



- En la casilla de ACUMULADO (es un valor lógico) en este caso, si digitamos la palabra VERDADERO obtendremos la probabilidad de 0.31744, correspondiente a la obtención de máximo 2, en este caso X toma valores de 0, 1 y 2 siendo P ($x \leq 2$).

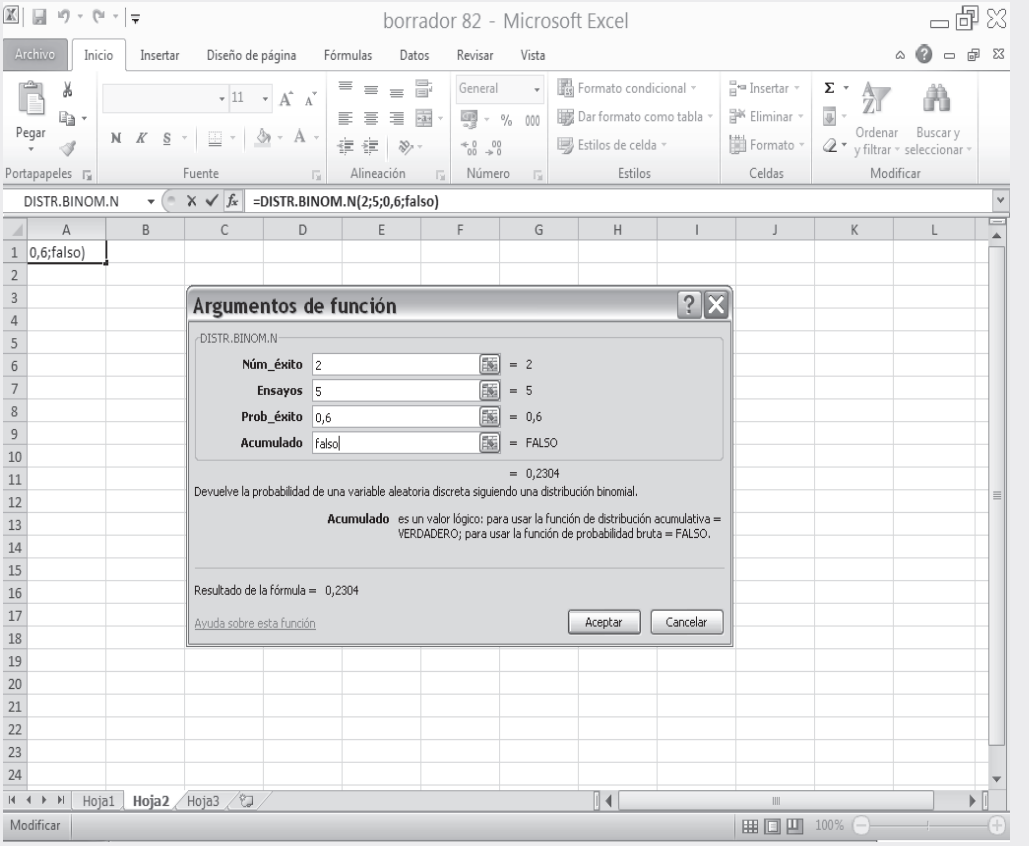
Figura No. 2. Argumentos función Distribución Binomial. Parámetro Verdadero



En la SEGUNDA PARTE DEL SIL que acompa a al libro encontrará en conjunto toda la aplicación de EXCEL a la ESTADÍSTICA, al igual en el capítulo 13, se ofrece una BASE DE DATOS para trabajar.

- Si por el contrario digitamos FALSO, en vez de VERDADERO, se calculará la probabilidad de obtener EXACTAMENTE DOS.

Figura No. 3. Argumentos función Distribución Binomial. Parámetro Falso.



- Observemos que en la parte inferior, aparece el valor de 0.2304 cuando escribimos FALSO, proceso utilizado cuando se quiere calcular cada término en forma independiente. Si por el contrario se escribe VERDADERO será el resultado igual a 0.31744, este caso implicará el acumulado de varios valores de X.

Ejemplo 4. Se encuentran 5 amigos cierto día ¿Cuál es la probabilidad (a) por lo menos dos de ellos hayan nacido un martes?; (b) ¿Si el grupo lo conforman 40 amigos, cuántos esperaríamos hubiesen nacido el día lunes?

USO DE CALCULADORA

Para hallar la probabilidad de un suceso en la binomial se presentan algunas marcas y modelos que más se utilizan en el salón de clases.

o :

a) $P_{(x\geq 2)} = P_{(x=2)} + P_{(x=3)} + P_{(x=4)} + P_{(x=5)}$

$P_{(x\geq 2)} = 1 - \left[P_{(x=0)} + P_{(x=1)}\right] = 1 - [0{,}4627 + 0{,}3856]$

$P_{(x\geq 2)} = 1 - 0{,}8483$

$P_{(x\geq 2)} = 0{,}1517 \;=\; 15{,}17\ \%$

b) $E = np$

$n = 40 \qquad p = 1/7$

$E = 40(1/7) = 5{,}7 \cong 6$

Aproximadamente 6 personas de las 40 se espera haber nacido el día lunes.

- Casio fx - 5.000F
  - Se digita el código de la fórmula binomial (103 FMLA)
  - Observe en pantalla dónde nos pide n? (digite en este caso 5 y luego EXE)
  - Aparece p? es la probabilidad de éxito (podemos digitar 1 ÷ 7 y luego EXE)
  - Observamos que nos aparece el resultado final del evento, siendo 0,4627

$$P_{(x=0)} = \binom{5}{0}(1/7)^0\,(6/7)^5 = 0{,}4627 = 46{,}27\%$$

- Casio fx - 350 TL y Casio fx - 82 TL, con ellas se procede asi:
  5 nCr 0 x (1 ÷ 7 1) x^y 0 x (6 ÷ 7) x^y 5 =
- Casio -350 M5 opera igual a los dos anteriores modelos, con la única diferencia, que no aparece la tecla x^y y debe ser reemplazada por la tecla Λ que indica lo mismo.
- Casio 350 HB requiere emplear la tecla SHIFT y luego x^y los demás pasos o tecleado es igual.
- Casio 991 MS tiene dos situaciones diferentes, así:
  5 SHIFT nCr 0 x (1 ÷ 7 Λ) Λ 0 x (6 ÷ 7) Λ 5 =
- En muchas calculadoras que no tienen la tecla nCr se debe operar paso a paso ((5 ! ÷ (0 ! x 5 !)) x (1/7 x^y 0) x (6/7 x^y 5) =

Ejemplo 5. Al inspeccionar 2.340 soldaduras producidas por cierto tipo de máquina, se encontraron 448 uniones defectuosas. Al efectuar 5 soldaduras ¿cuál es la probabilidad de: a) obtener 3 o más defectuosas; b) Como máximo dos buenas?

o :

a) $n=5 \qquad p=\dfrac{448}{2.340}=0{,}1 \qquad q=0{,}81 \qquad x=3,4,5$

$$P_{(x\geq 3)} = \binom{5}{3}(0{,}1)^3\,(0{,}81)^2 + \binom{5}{4}(0{,}1)^4\,(0{,}81)^1 + \binom{5}{5}(0{,}1)^5\,(0{,}81)^0$$

$$P_{(x\geq 3)} = 0{,}0450 + 0{,}0053 + 0{,}0002 = 0{,}0505 = 5{,}05$$

b) $n=5 \qquad p=0,\ 1 \qquad q=0{,}1 \qquad x=0,1,2$

$$P_{x\leq 2} = \binom{5}{0}(0,\ 1)^0\,(0{,}1)^5 + \binom{5}{1}(0,\ 1)^1\,(0{,}1)^4 + \binom{5}{2}(0,\ 1)^2\,(0{,}1)^3$$

$$P_{x\leq 2} = 0{,}0002 + 0{,}0053 + 0{,}0450 = 0{,}0505 = 5{,}05$$

Nota: Puede observarse en el ejercicio que obtener 3 o más defectuosos, es lo mismo que obtener 2 o menos buenos.

REPRESENTACIÓN GRÁFICA

Una *distribución binomial* se representa mediante el *diagrama de probabilidades*, el proceso es exactamente igual al *diagrama de frecuencias*.

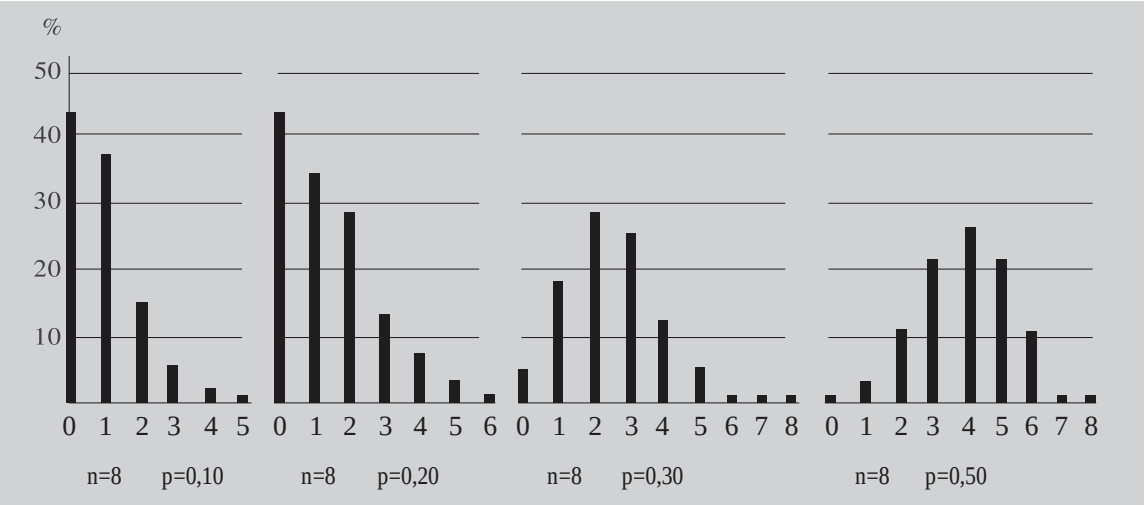
Las figuras 6.2 y 6.3 nos muestran varias distribuciones probabilísticas binomiales, además, permiten demostrar cómo se vuelven mas simétricas,tal como 10 podemos ver en los dos casos siguientes.

a) Cuando es constante y se hace cada vez más grande, hasta llegar a 0,50 = 50%. Supongamos que = 8 y que toma valores de 0,10 0,20 0,30 y 0,50

Para encontrar una forma más rápida, sin necesidad de realizar cálculos engorrosos al aplicar la fórmula de la binomial, podemos hacer uso de la tabla de distribución binomial, que se presenta en la tabla III cuando $n = 8$, siendo p: 0,05; 0,10; 0,15; 0,20; 0,25; 0,30; 0,35; 0,40; 0,45 y 0,50

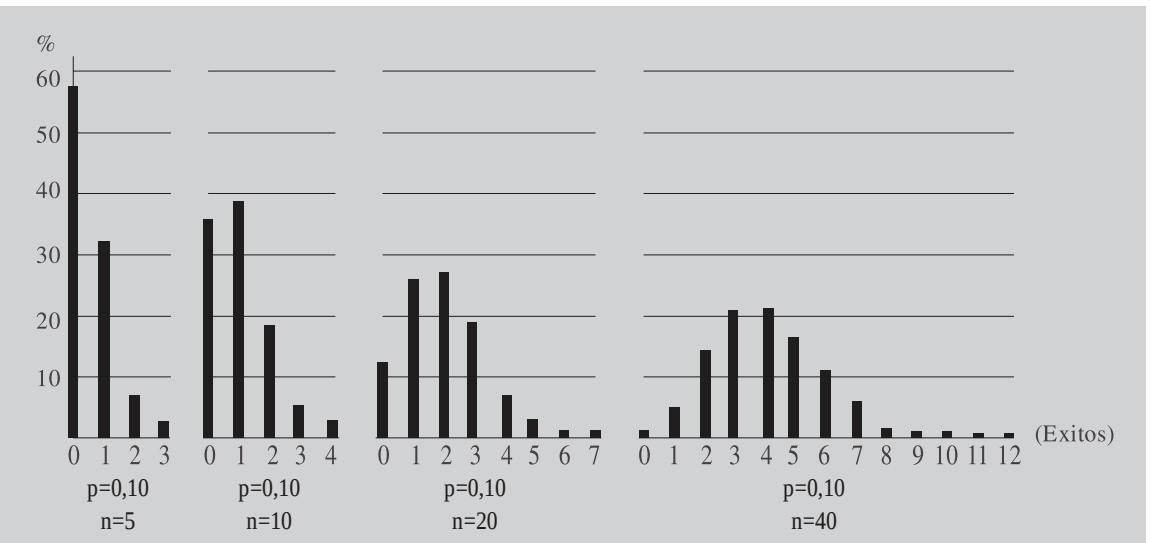
$$P_{(x)} = \binom{n}{x} p^x q^{n-x}$$

Figura 6.2 Diagrama de frecuencias



b) Cuando p es constante y n se hace cada vez más grande. Supongamos que $p = 0,10$ y n toma valores de 5, 10, 20 y 40.

Figura 6.3 Diagrama de frecuencias



EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

1. Al lanzar cuatro monedas, ¿cuál es la probabilidad de obtener exactamente dos caras?

2. Si en el ejercicio anterior, sólo queremos determinar la probabilidad de que aparezcan exactamente 3 caras.

3. En el caso de un dado, se quiere determinar la probabilidad de obtener exactamente 2 cincos en 4 lanzamientos.

4. En una facultad, la probabilidad de que un alumno apruebe el semestre es del 80%. Si consideramos 8 alumnos, ¿Cuál es la probabilidad de que:
   a) dos ganen. b) dos pierdan. c) por lo menos dos pierdan.
   d) como máximo 6 ganen. e) seis pierdan el semestre?

5. Se lanza 6 veces una moneda. Encontrar la probabilidad de obtener:
   a) exactamente 4 caras. b) máximo 4 caras.

6. Se lanzan 7 dados. Si el éxito consiste en sacar un 5 ó 6, encontrar la probabilidad de obtener:
   a) exactamente 4 éxitos. b) máximo 4 éxitos.

7. Se sabe que en la manufactura de cierto artículo, uno de cada 10 resulta defectuoso. ¿Cuál es la probabilidad de que una muestra aleatoria de 4 artículos contenga:
   a) ninguno defectuoso. b) exactamente uno defectuoso.
   c) exactamente dos defectuosos. d) no más de dos defectuosos?

8. Si un jugador que al batear tiene un promedio de 0,40, llega a batear 5 veces en un juego, ¿Cuál es la probabilidad de que obtenga:
   a) exactamente dos golpes. b) menos de dos golpes?

9. Al lanzar 8 monedas, ¿cuál es la probabilidad de obtener menos de 6 caras?

10. Según los registros universitarios, fracasa el 5% de los alumnos de cierto curso. ¿Cuál es la probabilidad que de 6 estudiantes seleccionados al azar, que hayan seguido dicho curso, menos de 3 hayan fracasado?

11. Los registros hospitalarios indican que el 10% de los casos de cierta enfermedad resultan fatales. Si hay 5 pacientes que sufren de la enfermedad, encontrar la probabilidad de que:
    a) todos sanen. b) por lo menos tres mueran. c) exactamente 3 mueran.

12. Si el 20% de los cerrojos producidos por una máquina son defectuosos, determinar la probabilidad, que de 4 cerrojos elegidos al azar,
    a) uno. b) cero. c) no más de dos cerrojos sean defectuosos.

13. La probabilidad de que se gradúe un estudiante que ingresa a una universidad es de 0,4. Calcular la probabilidad entre 5 estudiantes que ingresan.
    a) ninguno se gradúe. b) se gradúe uno. c) se gradúe al menos uno.

14. Un dado se lanza cinco veces. Calcular la probabilidad de que el 3 aparezca:
a) una vez. b) dos veces. c) tres veces. d) cuatro veces. e) ninguna vez.

15. El 10% de los artículos producidos mediante cierto proceso son defectuosos (no aceptables). Si se toma, al azar, una muestra de cuatro artículos, ¿cuál es la probabilidad de que contengan:
a) ninguno defectuoso. b) al menos uno defectuoso. c) menos de dos defectuosos?

16. En una fábrica el 20% de los artículos que produce cierta máquina resultan defectuosos. Si 10 artículos son elegidos al azar, de todos los producidos en el día por dicha máquina, calcular la probabilidad de que haya:
a) exactamente dos defectuosos. b) 3 ó más defectuosos.
c) más de cinco defectuosos. d) ninguno defectuoso.

17. Suponga que la mitad de los habitantes de cierto pueblo ven regularmente televisión. De 100 investigadores, cada uno encuesta a 10 personas. ¿Cuántos se espera que mencionen 3 o menos que sean televidentes regulares?

18. En promedio cierto estudiante puede resolver la mitad de los problemas que se le presentan; para aprobar es necesario solucionar 7 de 10 problemas de un examen. ¿Cuál es la probabilidad de que el estudiante apruebe el examen?

19. En promedio el 10% de las varillas de madera usadas en cierto producto se encuentran demasiado nudosas para ser usadas. ¿Cuál es la probabilidad de que en un paquete de 15 varillas:
a) exactamente 5 estén demasiado nudosas. b) por lo menos 10 estén demasiado nudosas.
c) más de 4 estén demasiado nudosas?

20. Por regla, el 25% de ciertos productos manufacturados por un cierto torno, son defectuosos. ¿Cuál es la probabilidad de que en 20 de estos productos haya:
a) exactamente 15 defectuosos. b) menos de 5 defectuosos. c) por lo menos 8 defectuosos?

21. De un total de 2.000 familias con 4 hijos cada una, ¿en cuántas de ellas cabe esperar que haya:
a) al menos 1 niño. b) 2 niños. c) ningún niño?

22. Siendo $n = 15$ $p = 0{,}05$, hallar la probabilidad de 2 o menos acontecimientos favorables.

23. Si el 40% de los votantes de una gran ciudad favorecen al candidato A, ¿cuál es la probabilidad para que, en una muestra al azar de 20 votantes, la mayoría favorezca al candidato A?

24. Supongamos que Jorge dispara su revólver 18 veces, ¿cuál es la probabilidad de que logre exactamente 8 centros, sabiendo que la probabilidad de que logre centro es de 0,20?

25. Diez monedas son extraídas del bolsillo de una persona y puestas sobre una mesa. ¿Cuál es la probabilidad de que el número de caras esté entre 5 y 7 inclusive?

26. ¿Cuál es la probabilidad de contestar correctamente, por lo menos tres de las cinco preguntas de un test de falso y verdadero?

27. Si se sabe que nueve de cada diez personas tienen caries, al tomar al azar un grupo de cinco personas, ¿cuál es la probabilidad de que:

a) cuatro tengan caries. b) por lo menos dos tengan caries.
c) por lo menos dos no tengan caries. d) por lo menos una tenga caries?

28. Si el 20% de los estudiantes de una universidad pierden el primer año y se toma al azar un grupo de seis estudiantes, ¿cuál es la probabilidad de que:
a) máximo dos aprueben. b) todos aprueben. c) ninguno apruebe?

29. De los 6.000 estudiantes matriculados en la universidad, se sabe que 4.800 se trasladan al claustro utilizando el transporte urbano (servicio público). Si se selecciona una muestra de ocho estudiantes, ¿cuál es la probabilidad de que:
a) no más de dos utilicen dicho servicio. b) por lo menos tres no lo utilicen.
c) exactamente dos no lo utilicen. d) exactamente dos lo utilicen?

30. Se sabe que el 60% de los alumnos de una universidad asisten a clases el día viernes; en una encuesta a ocho alumnos de la universidad, ¿cuál es la probabilidad de que:
a) por lo menos 7 asistan a clase el día viernes.
b) por lo menos dos no asistan a clases el día viernes?

31. Se sabe que en una universidad de 2.000 estudiantes, ochocientos usan gafas. Si se realiza una encuesta a cinco estudiantes, ¿cuál es la probabilidad de que:
a) por lo menos dos usen gafas. b) por lo menos dos no usen gafas.
c) de 2.000 estudiantes, ¿cuántos no usan gafas?

32. Si un tercio de los estudiantes de un curso de contabilidad son repitentes, calcule la probabilidad de que en una muestra al azar de cuatro estudiantes:
a) no más de dos sean repitentes b) al menos uno no sea repitente.

33. ¿Cuál es la probabilidad de que se produzcan diez o más acontecimientos desfavorables en 16 ensayos de una experiencia binomial, si la probabilidad de acontecimiento favorable en cada ensayo es de 0,4? (Utilice las tablas para el cálculo).

34. Una compañía de seguros considera que alrededor del 25% de los carros de servicio público se accidentan cada año. ¿Cuál es la probabilidad de que por lo menos tres de una muestra de siete vehículos afiliados, hayan tenido accidentes en el año?

35. De la producción de envases metálicos en una fábrica, se sabe que el 3% son defectuosos. ¿Cuál es la probabilidad de que en una muestra de siete envases:
a) por lo menos tres sean buenos. b) por lo menos tres sean defectuosos?

36. Si la probabilidad de que un niño enferme de sarampión es de 1%, ¿cuál es la posibilidad de que en una familia con cinco hijos, resulten:
a) dos enfermos. b) por lo menos uno enfermo. c) por lo menos dos no se enfermen?

37. Cierta enfermedad tiene un 20% de mortalidad. Si existen cinco pacientes con la enfermedad, ¿cuál será la probabilidad de que:
a) ninguno sobreviva. b) todos sobrevivan.
c) al menos uno sobreviva. d) al menos uno no sobreviva?

38. En una ciudad se publican 25 revistas de las cuales cinco son científicas. Si se eligen cuatro al azar, ¿cuál será la probabilidad de que:
    a) por lo menos una sea científica. b) por lo menos dos no sean científicas.
    c) una sea científica?

39. El 30% de las familias de un barrio de Quito, son consideradas posibles clientes para comprar cierto producto. Se toma una muestra de ocho familias. ¿Cuál es la probabilidad, en la muestra, de que:
    a) tres o menos sean clientes. b) tres o más no sean clientes?

40. Los dos quintos de los votantes inscritos en ciertos barrios de Bogotá apoyan al grupo A. ¿Cuál es la probabilidad de que en una muestra de siete votantes de ese puesto de votación:
    a) exactamente cinco apoyen a A. b) por lo menos dos apoyen a A.
    c) por lo menos dos no apoyen a A?

41. Un examen de tipo *verdadero-falso,* contiene 16 preguntas; suponiendo que un alumno conteste cada pregunta al azar, ¿cuál es la probabilidad de que:
    a. A lo más dos preguntas sean contestadas correctamente.
    b. Por lo menos dos preguntas sean contestadas correctamente.
    c. Por lo menos dos no sean contestadas correctamente?
    Nota: *se puede utilizar la tabla*

42. En la producción de un determinado artículo encontramos que por cada veinte que se producen, tres de ellos resultan defectuosos. Si se toma una muestra de ocho artículos, ¿cuál es la probabilidad de que:
    a) por lo menos dos sean defectuosos.
    b) por lo menos dos no sean defectuosos.
    c) En una producción de 2.000 artículos, ¿en cuántos de ellos esperamos que dos sean defectuosos?

43. Se sabe que el 70% de los miembros de la universidad son fumadores; en una muestra aleatoria de 18 fumadores, ¿cuál es la probabilidad de que haya exactamente 16 fumadores? (utilizar la tabla).

44. Una firma de pedidos por correo, envía una circular y espera una tasa de respuesta del 20%; suponga que diez circulares se envían como prueba de mercado en una nueva área geográfica. Si se aplica la tasa de respuestas del 20% en la nueva región, determine la probabilidad de que:
    a) la mayoría responda b) menos del 30% de la gente no responda c) nadie responda.

45. El 17 % de los artículos producidos mediante cierto proceso, son defectuosos. Se toma al azar una muestra de diez artículos, ¿cuál es la probabilidad de que:
    a) ninguno sea defectuoso. b) por lo menos dos no sean defectuosos.
    c) como máximo dos sean defectuosos?

46. Existe un 80% de probabilidad de que un tipo determinado de componentes se comporte adecuadamente bajo las condiciones de alta temperatura. Si el dispositivo en cuestión tiene cuatro de tales componentes, determine la probabilidad en cada uno de los siguientes eventos, por medio de la fórmula para probabilidades binomiales:
    a) todos los componentes se comportan adecuadamente y por lo tanto el dispositivo es operacional.

b) el dispositivo no es operacional porque falla uno de los cuatro componentes.
c) el dispositivo no es operacional porque falla uno o más de los componentes.

47. Suponga que un conductor de automóvil que maneja con exceso de velocidad, puede ser detectado por un sistema de radar. Se dice que de cada diez con exceso de velocidad, seis son detectados. Un automovilista va con exceso de velocidad, en viaje entre Bogotá y Tunja. Durante el trayecto hay ocho estaciones de vigilancia por radar.
a) ¿Qué probabilidad hay de que este automovilista, por lo menos cinco veces sea detectado con exceso de velocidad?
b) ¿Por lo menos dos no sean detectados?

48. Se sabe que la probabilidad de que una familia tome jugo de naranja en el desayuno es del 9,9%. Se toma una muestra de cinco familias y se quiere determinar la probabilidad de que:
a) por lo menos dos familias tomen jugo de naranja al desayuno.
b) como máximo tres familias no tomen jugo de naranja al desayuno.

49. Un vendedor supone que cada entrevista le produce una venta con probabilidad de ¼. En un determinado día el vendedor entrevista cinco posibles clientes. ¿Cuál es la probabilidad de que por lo menos tres compren?

50. Supongamos que el 16% de los estudiantes que se presentan a un examen de estadística, con menos de tres ausencias, en el semestre, lo pierden. Si se seleccionan 10 estudiantes con menos de tres ausencias, ¿cuál es la probabilidad que no lo pierda un número de alumnos no mayor a cinco ni menor a tres?

51. Una secretaria que debe llegar a su trabajo todas las mañanas a las 8:00 se tarda como mínimo 15 minutos, el 20% de las veces. El jefe de la oficina, que no llega sino a las 9:00, la mayoría de las veces, llama ocasionalmente entre las 8:00 y las 8:15 para enterarse de las novedades.
a. ¿Cuál es la probabilidad, en dos mañanas de las cinco de la semana, que el jefe llame y la secretaria no esté en la oficina?
b. ¿Que por lo menos dos mañanas de las cinco, la secretaria esté?

52. En una población de trabajadores del Estado, que gozan pensión de jubilación, el 36% tiene más de 65 años. ¿Cuál es la probabilidad, en una muestra de 18 pensionados, de que 15 o más tengan más de 65 años?

53. La oficina de personal en una fábrica, indica que el 30% de los empleados de la línea de montajes se retiran durante los primeros tres años de haber sido contratados. Se acaban de contratar 12 empleados nuevos. ¿Cuál es la probabilidad de que:
a) por lo menos nueve sigan trabajando después del tercer año.
b) como mínimo tres se retiren antes del tercer año?

54. Si se sabe que uno de cada diez libros de texto representan un éxito financiero y una casa editora ha decidido publicar seis textos nuevos, ¿cuál es la probabilidad de que por lo menos dos sean un éxito financiero?

55. En una región se tienen 25 colegios de enseñanza secundaria, de los cuales cinco son técnicos. Si se eligen al azar cuatro colegios, ¿cuál es la probabilidad de que:
a) por lo menos uno sea técnico. b) por lo menos dos no sean técnicos?

DISTRIBUCIÓN PROBABILÍSTICA DE POISSON

En una *distribución binomial* cuando es grande, por lo general mayor de cincuenta, y , la probabilidad de éxito de un suceso, se acerca a cero, mientras que la probabilidad de fracaso se aproxima a 1, de tal manera que el producto de , simbolizado por lambda λ, sea menor o igual a 5, debe utilizarse la *distribución de Poisson.* Algunos autores, consideran no sólo el hecho de que sea muy pequeña, sino también cuando es tan grande que se aproxima a 1, y $\lambda \leq 5$, en ambos casos, se puede aplicar esta distribución. La tabla que se adjunta al final dá la opción donde $\lambda \leq 10$, como algunos lo aceptan, contrario a un valor menor o igual a 5.

Su fórmula es: $$P_{(x)} = \frac{\lambda^{x} e^{-\lambda}}{X!}$$

Siendo
- $e = 2{,}71828$ (*base de los logarítmos* **neperianos**)
- $\lambda = np$
- $X =$ *número de casos favorables*
- $P_{(x)} =$ *Probabilidad que se va a calcular para un valor dado de* X

Generalmente se dice, que la ***distribución de Poisson*** tiene su mayor aplicación, cuando en el experimento que se realiza ocurren sucesos llamados raros, los cuales se identifican con una probabilidad de éxito sumamente pequeña (p) y el número de observaciones (n) grande; pero la verdad es que esta distribución se aplica a una variedad de situaciones diferentes, como las ocurrencias respecto a un campo continuo, como área o tiempo. Algunos de estos eventos aleatorios ocurren en forma independiente a una velocidad dentro de un campo o intervalo, generalmente de espacio (área y tiempo).

Son ejemplos para aplicación de la ***distribución de Poisson***: el número de personas que llegan a un almacén, banco o aeropuerto en un tiempo determinado; el número de llamadas telefónicas por minuto; el número de defectos en piezas similares en el material, ya sea por centímetro cuadrado o centímetro lineal; número de bacterias en un cultivo; insectos por kilómetro cuadrado; el número de fallas de una máquina durante una hora o un día; el número de accidentes por día; el número de reclamaciones o solicitudes a una compañía de seguros en un determinado período, etc. Como se puede observar, se trata de hallar la probabilidad de ocurrencia de cualquier número de éxitos (X) por unidad de medición (minuto, hora, día, centímetro, metro, etc.). Algunos de los ejercicios que se darán para su solución, solo tendrán el valor del parámetro lambda (λ) o sea el ***promedio o razón*** de ocurrencia del evento aleatorio por unidad de tiempo o espacio y el número de éxitos que se solicitan, por lo tanto no habrá necesidad de conocer los valores de *p* y de *n*.

Ejemplo 1. Si el 1% de las bombillas fabricadas por una compañía son defectuosas, hallar la probabilidad de que, en una muestra de 100 bombillas, 3 sean defectuosas.

o :

$\lambda = 100(0{,}01) = 1$ $X = 3$

$$P_{x=3} = \frac{1^3 e^{-1}}{3} = \frac{1(0{,}367\quad)}{3\ 2\ 1} = 0{,}06131 = 6{,}13$$

Ejemplo 2. Si la probabilidad de que una persona adquiera la enfermedad como consecuencia de una vacuna contra la misma, es 0,0002, ¿cuál es la probabilidad de que la adquieran exactamente 5 personas en una población de 10.000 vacunados?

o **:**

$\lambda = 10.000(0{,}0002) = 2 \qquad X = 5 \qquad 5\,! = 5.\,4.3.\,2.1 = 120$

$$P_{(x=5)} = \frac{2^5 e^{-2}}{5} = \frac{32\ (0{,}13534)}{120} = 0{,}0360$$

$P_{(x=5)} = 3{,}61$

Ejemplo 3. Se estima que una de cada 10.000 personas es alérgica a cierta sustancia utilizada en la fabricación de tintes para el cabello. ¿Cuál es la probabilidad de que en 20.000 usuarias de tintes, más de cinco sufran reacciones alérgicas, debido a su uso?

o **:**

$$p = \frac{1}{10\,000} = 0{,}0001 \qquad \lambda = np = 20\,000\ 0{,}0001 = 2$$

$$P_{x\geq 6} = 1 - \left[\frac{2^0 e^{-2}}{0} + \frac{2^1 e^{-2}}{1} + \frac{2^2 e^{-2}}{2} + \frac{2^3 e^{-2}}{3} + \frac{2^4 e^{-2}}{4} + \frac{2^5 e^{-2}}{5}\right]$$

$$P_{x\geq 6} = 1 - \left[\frac{1\ 0{,}13534}{1} + \frac{2\ 0{,}13534}{1} + \frac{4\ 0{,}13534}{2} + \frac{0{,}13534}{6} + \frac{16\ 0{,}13534}{24} + \frac{32\ 0{,}13534}{120}\right]$$

$$P_{x\geq 6} = 1 - 0,\quad 344 = 0{,}01656 = 1{,}66$$

El valor de e^{-x} se calcula generalmente mediante el uso de la calculadora, cuyo procedimiento varía de acuerdo a su marca y modelo:

USO DE CALCULADORA

- Casio fx - 5.000F (y en la mayoría de modelos) se procede así:
  SHIFT e^x (-) 3 EXE 0,135335283
  También mediante el uso de tablas (ver al final del libro tabla VII)

Ejemplo 4. El número de Clientes que llegan a una corporación de ahorro y vivienda los días sábados es, en promedio, 40 por hora ¿Cuál es la probabilidad de que lleguen por lo menos dos clientes (a) en un período de 14 minutos; (b) en 5 minutos?

o ***:*** utilizaremos la tabla para calcular la probabilidad en cada suceso.

a) $\lambda = 40 \qquad \lambda = \frac{40}{60} \times 14 = \ ,33 \qquad ,3$

$$P_{x\geq 2} = 1 - \left[P_{x=0} + P_{x=1}\right]$$

$$P_{x\geq 2} = 1 - \left[\frac{,3^0 e^{-\ ,3}}{0} + \frac{,3^1 e^{-\ ,3}}{1}\right] = 1 - \left[0{,}0001 + 0{,}000\ \right]$$

$$P_{x\geq 2} = 1 - 0{,}000\ = 0,\quad 1 = \ ,\ 1$$

Tabla 6.6 Distribución de Poisson

x	λ									
	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10.0
0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0000
1	0,0010	0,0009	0,0009	0,0008	0,0007	0,0007	0,0006	0,0005	0,0005	0,0005
2	0,0046	0,0043	0,0040	0,0037	0,0033	0,0031	0,0029	0,0027	0,0025	0,0023
3	0,0140	0,0131	0,0123	0,0115	0,0100	0,0100	0,0093	0,0087	0,0081	0,0076
4	0,0319	0,0302	0,0285	0,0269	0,0254	0,0240	0,0226	0,0213	0,0201	0,0189

b) $\lambda = 40$ $\lambda = \frac{40}{60} \times 5 = 3{,}33$ 3,3

$$P_{(x\geq 2)} = 1 - \left[\frac{3{,}3^0 e^{-3{,}3}}{0} + \frac{3{,}3^1 e^{-3{,}3}}{1}\right] = 1 - [0{,}036 + 0{,}1217]$$

$$P_{(x\geq 2)} = 1 - 0{,}1586 = 0{,}8414$$

$$P_{(x\geq 2)} = 84{,}14$$

Tabla 6.6A Distribución de Poisson

Esta tabla sólo sirve cuando se tiene un solo decimal; como se observa, y la empleamos cuando λ = 3,3 y no para λ = 3,33

x	λ					
	3.1	3.2	3.3	3.4	3.5	3.6
0	0,0450	0,0408	0,0369	0,0334	0,0302	0,0273
1	0,1397	0,1304	0,1217	0,1135	0,1057	0,0984
2	0,2165	0,2087	0,2008	0,1929	0,1850	0,1771
3	0,2237	0,2226	0,2209	0,2186	0,2158	0,2125
4	0,1734	0,1781	0,1923	0,1858	0,1888	0,1912

Nota: Si se hubiese trabajado correctamente con λ = 9,33 y 3,33, los resultados respectivos serían:

a) $P_{x\geq 2} = 1 - \left[\frac{,33^0 e^{-,33}}{0} + \frac{,33^1 e^{-,33}}{1}\right] = 1 - [0{,}0000 + 0{,}000\ 3]$

$P_{x\geq 2} = 1 - 0{,}000 = 0,\quad 1 = \ ,\ 1$

b) $P_{x\geq 2} = 1 - \left[\frac{3{,}33^0 e^{-3{,}33}}{0} + \frac{3{,}33^1 e^{-3{,}33}}{1}\right] = 1 - [0{,}035 + 0{,}11\ 2]$

$P_{x\geq 2} = 1 - 0{,}1550 = 0,\ 450 = \ 4{,}50$

Ejemplo 5. Supongamos que λ = 2 y deseamos hallar la $P_{(x)}$ cuando x = 0, 1, 2, 3 y 4

$$P_{(x\leq 4)} = P_{(x=0)} + P_{(x=1)} + P_{(x=2)} + P_{(x=3)} + P_{(x=4)}$$

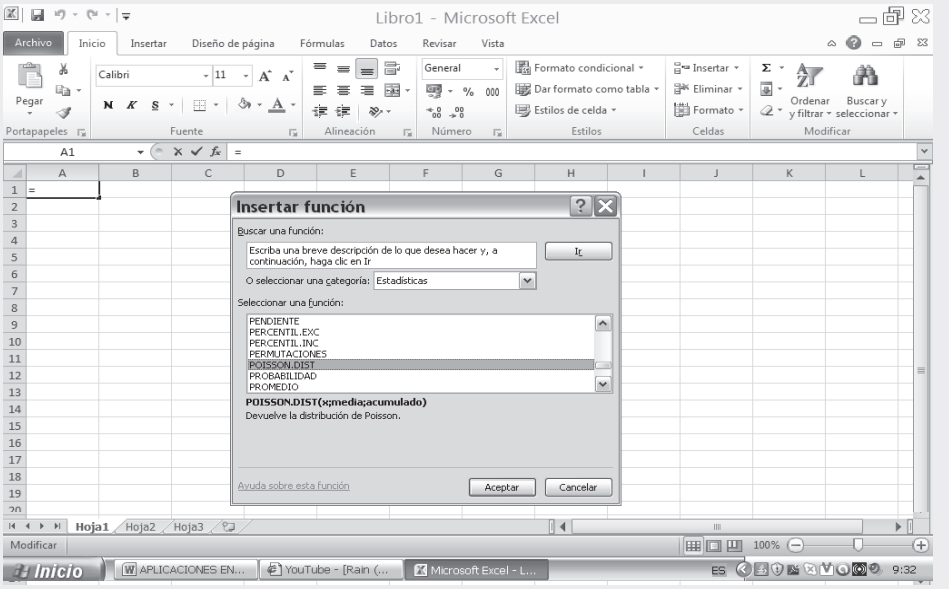
APLICACIÓN DEL EXCEL

Complemento virtual SIL

Las operaciones se harán más rápido y más confiable su resultado utilizando el programa Excel. Veamos en nuestro ejercicio cómo se opera:

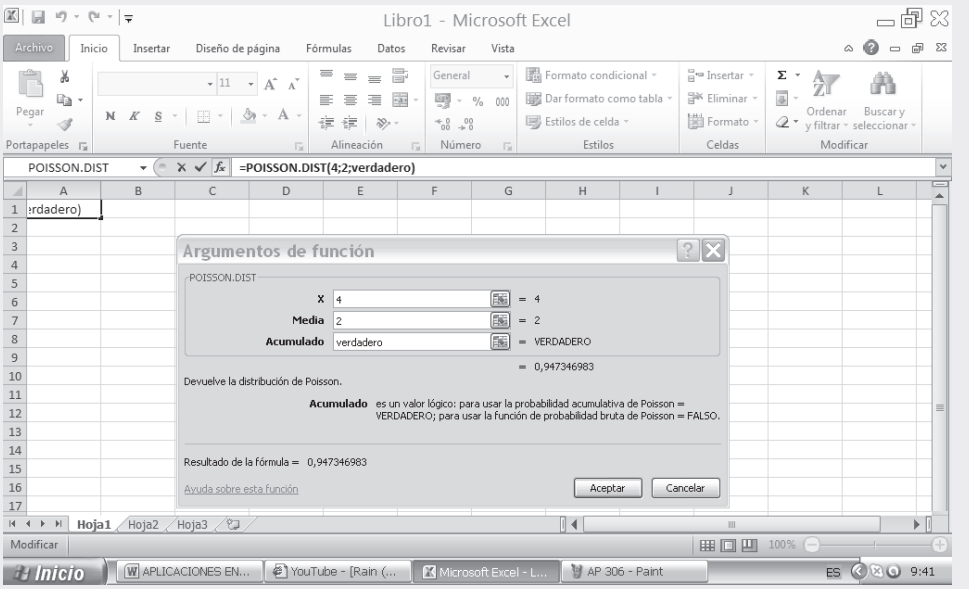
- |Nuevamente nos ubicamos en la función *fx*, haciendo CLIC en ella, tal como hicimos en la distribución binomial, sólo que ahora hacemos CLIC en POISSON, tal como se presenta en la aplicación siguiente:

Figura No. 4. Insertar Función.



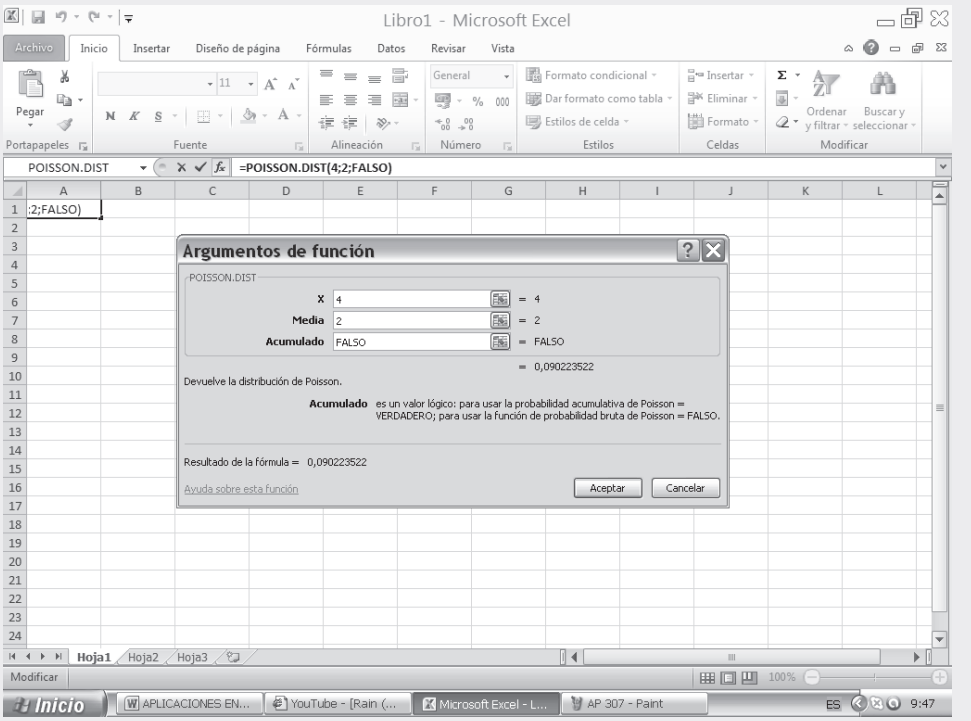
- Al presionar ACEPTAR aparece un cuadro de diálogo para lo cual en la casilla X digitamos el número correspondiente de éxitos, en este caso 4.
- En la siguiente casilla o renglón aparece la palabra MEDIA, digitamos el valor correspondiente a lambda λ siendo igual a 2.
- Nos queda por último el renglón ACUMULADO que corresponde a un valor lógico. Si digitamos en él, la palabra VERDADERO, nos estaremos refiriendo a la probabilidad de obtener como máximo, en este caso cuatro (4), como se presenta a continuación:

Figura No. 5. Argumentos de la Función. Parámetro Verdadero



- Si por el contrario, digitamos la palabra FALSO, se obtiene la probabilidad de que ocurra que X sea igual al parámetro establecido, así, en nuestro caso X es igual a 4.

Figura No. 6. Argumentos de la Función. Parámetro Falso



$$P_{(x)} = 1 - \left[\frac{\lambda^{x} e^{-\lambda}}{x!} \right] \qquad P_{(x\leq 4)} = 0{,}9473 \text{ (PALABRA VERDADERO)} \qquad P_{(x=4)} = 0{,}0902 \text{ (PALABRA FALSO)}$$

Ejemplo 6. En un proceso de manufactura textil, se tiene que el promedio de defectos es de 6 por cada 30 metros de tela. ¿Cuál es la probabilidad de obtener por lo menos 2 fallas o defectos, (a) en una pieza de 30 metros?; (b) en una pieza de 10 metros?

a) $\lambda = 6$ por cada 30 metros de tela; x = 2, 3, 4...n

$$P_{x\geq 2} = 1 - \left[\quad_{x=0} + \quad_{x=1} \right]$$

$$P_{(x\geq 2)} = 1 - \left[\frac{6^0 e^{-6}}{0} + \frac{6^1 e^{-6}}{1} \right] = 1 - [0{,}0025 + 0{,}014 \ ]$$

$$P_{(x\geq 2)} = 1 - 0{,}0174 = 0, \quad 26 = \quad ,26$$

b) $\lambda = 2$ por cada 10 metros de tela; x = 2, 3, 4...n

$$P_{(x\geq 2)} = 1 - \left[P_{(x=0)} + P_{(x=1)} \right]$$

$$P_{(x\geq 2)} = 1 - \left[\frac{2^0 e^{-2}}{0} + \frac{2^1 e^{-2}}{1} \right] = 1 - [0{,}1 \ 5 \ + 0{,}2 \ 0$$

$$P_{(x\geq 2)} = 1 - 0, \ 0 \ 0 = 0{,}5 \quad 0 = 5 \ , \ 0$$

USO DE CALCULADORA

El *uso de la calculadora* nos ayuda a agilizar los *cálculos*, veamos como se procede con alguna de ellas, por ejemplo al calcular $_{x=0} = \frac{6^0 e^{-6}}{0}$, es decir, cuando X = 0 y $\lambda = 6$

- Casio fx - 5000F

- ❑ Se teclea 104 y luego FMLA
- ❑ Nos pide el valor de μ? que equivale a λ, procedemos a digitar el valor de 6 y luego EXE
- ❑ Aparece X? en nuestro caso 0 o digitamos y luego EXE y nos da el resultado de 0,0024787 aproximadamente 0,0025.

■ Casio fx -350TL y Casio fx -82TL, se procede para el cálculo así:
(6 x^y 0 x SHIFT e^x (–) x^y 6) ÷ 0 SHIFT X! =

■ Casio 991 MS y 350 MS opera al igual que la calculadora anterior sólo que no aparece la tecla x^y sino Λ

■ Casio 350 HB también se modifica el teclado x^y ahora por SHIFT x^y

■ Casio -360 PV opera un poco diferente a las anteriores, así:
(6 x^y 0 x (6 +/- SHIFT e^x) ÷ 0 SHIFT ! =

b) λ = 6 por cada 30 metros; $\lambda = \frac{6}{30} \times 10 = 2$ 10

$$P_{(x\geq 2)} = 1 - \left[\frac{2^0 e^{-2}}{0} + \frac{2^1 e^{-2}}{1}\right] = 1 - [0{,}1353 + 0{,}2707]$$

$$P_{(x\geq 2)} = 1 - 0{,}4060 = 0{,}5 \quad 40 = 5 \quad ,40$$

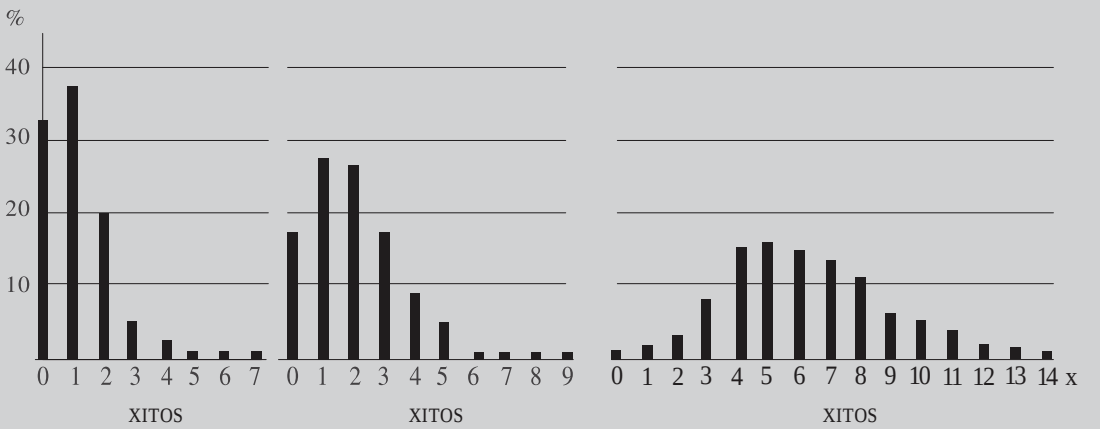
REPRESENTACIÓN GRÁFICA

Tal como se hizo en la distribución de *probabilidades binomial*, se puede realizar la representación gráfica en una distribución *probabilística de Poisson* mediante un diagrama de probabilidades parecido al diagrama de frecuencias.

Utilizando la Tabla VII que aparece al final del libro o la calculadora elaboraremos las gráficas para:

$$\lambda = 1{,}1 \quad \lambda = 2{,}0 \quad \text{y} \quad \lambda = 6$$

Figura 6.4 Diagrama de frecuencias



Obsérvese que a medida que aumenta el valor de lambda (λ, la distribución se vuelve casi simétrica. Además, el valor de n no tiene un límite específico.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

56. Un 10% de los utensilios producidos en un cierto proceso de fabricación resulta ser defectuoso. Hallar la probabilidad que de una muestra de diez utensilios, seleccionados al azar, sean exactamente dos los defectuosos, mediante:
a) distribución binomial. b) distribución de Poisson.

57. Si la probabilidad de que un individuo sufra una reacción por una inyección de un determinado suero es 0,001, determinar la probabilidad que de un total de 2.000 individuos:
a) exactamente 3 tengan reacción. b) más de 2 individuos tengan reacción.

58. Si el 3% de las bombillas fabricadas por una compañía son defectuosas, hallar la probabilidad de que en una muestra de 100 bombillas:
a) 0 b) 1 c) 2 d) 3 e) 4 f) 5 sean defectuosos.

59. El número de ahogados en un accidente, por año, en un país X es de tres por cada 100.000 habitantes. Hallar la probabilidad de que en una ciudad cuya población es de 200.000 haya:
a) 0 b) 2 c) 6 d) 8 e) entre 4 y 8 f) menos de 3 ahogados por año.

60. Suponga que una empresa aérea, en promedio, uno de cada 150 vuelos se retrasan más de una hora; si se programan 1.500 vuelos en un mes:
a) ¿Cuál es la probabilidad de que exactamente 3 vuelos se retrasen una hora?
b) ¿Cuál es la probabilidad, que más de 3 vuelos se retarden más de una hora?
c) ¿Cuál es la probabilidad de que menos de 5 vuelos se retrasen más de una hora?

61. Si la probabilidad de que una persona adquiera la enfermedad como consecuencia de una vacuna contra la misma, es 0,0001 ¿Cuál es la probabilidad de que la adquieran exactamente 5 personas en una población de 10.000 vacunados?

62. Si la probabilidad de que un individuo sufra una reacción por una inyección de un determinado suero es 0,001, determinar la probabilidad que de un total de 2.000 individuos:
a) exactamente 3 tengan reacción. b) más de dos individuos tengan reacción.

63. Una distribución de Poisson es dada: $P = \dfrac{(0{,}72)^x e^{-0{,}72}}{x\,!}$ Hallar la probabilidad cuando:
a) x = 0 b) x = 1 c) x = 2 d) x = 3

64. Entre las 2 y las 4 de la tarde, el promedio de llamadas telefónicas que recibe un conmutador de una empresa, por minuto, es 2,5. Hallar la probabilidad de que en un determinado minuto haya:
a) 0 b) 1 c) 2 d) 3 llamadas.

65. Supongamos que de cada 5.000 vehículos, dos tienen problemas con las llantas en una autopista. Si 1.000 vehículos transitan por la autopista durante cierto día, ¿cuál es la probabilidad de que por lo menos dos vehículos tengan problemas con las llantas?

66. Supongamos que, en promedio, una casa, de cada 2.000, en cierta zona de Buenos Aires, se incendia durante el año. Si hay 6.000 casas en dicha zona, ¿cuál es la probabilidad de que:

a) más de 3 casas se incendien durante el año.
b) exactamente dos se incendien durante el año?

67. El promedio de atracos en cierta ciudad es de dos por día. Utilizando la distribución de Poisson, determinar la probabilidad de que en un día dado, haya:
a) no más de tres atracos. b) a lo más dos atracos.

68. Una compañía de seguros considera que solamente alrededor del 0,01% de la población le ocurre cierto tipo de accidentes cada año. La empresa tiene 10.000 asegurados contra este tipo de accidentes, ¿Cuál es la probabilidad de que máximo tres de ellos sufran accidente?

69. Se toma una muestra de 1.500 artículos de un lote de producción que arroja el 24% de defectuosos, ¿Cuál es la probabilidad de obtener:
a) dos o menos artículos defectuosos. b) más de dos defectuosos?

70. Las estadísticas sobre la aplicación de normas de seguridad en una fábrica indican que, en promedio, se presentan 10 accidentes cada semestre. Utilice la distribución de Poisson para determinar la probabilidad de que no haya más de dos accidentes de trabajo en un trimestre.

71. El número de demandas presentadas a una compañía de seguros, en promedio (np) es de tres por día. ¿Cuál es la probabilidad de que en un día cualquiera:
a) no se presente ninguna demanda. b) por lo menos se presentan dos demandas?

72. La probabilidad de que un cajero se equivoque en el pago de un cheque es de 0,0005. ¿Cuál es la probabilidad de que en 800 cheques pagados por dicho cajero:
a) por lo menos se equivoque en el pago de tres cheques.
b) máximo se equivoque en dos cheques?

73. La tasa de mortalidad de cierta enfermedad es de tres por mil. ¿Cuál es la probabilidad de que en un grupo de 500 personas,
a) más de dos mueran. b) como máximo dos mueran?

74. En promedio doce personas por hora consultan a un especialista en decoración en un almacén de telas. ¿Cuál es la probabilidad de que durante un período de diez minutos,
a) por lo menos dos se acerquen al especialista.
b) no más de dos se acerquen al especialista?

75. Se estima que una de cada 10.000 personas es alérgica a cierta sustancia utilizada en la fabricación de tintes para el cabello, ¿Cuál es la probabilidad de que 30.000 usuarias de tintes,
a) por lo menos una sufra reacciones alérgicas. b) más de una sufra reacciones alérgicas?

76. El conmutador de una clínica recibe un promedio de 20 llamadas cada 2 minutos, ¿Cuál es la probabilidad de que lleguen,
a) exactamente 4 llamadas en un período de 30 segundos.
b) como máximo dos llamadas en un período de 15 segundos?

77. Los clientes llegan a una exhibición a razón de 6,8 clientes/hora. Calcule la probabilidad de que:
a) en la primera media hora, por lo menos llegue un cliente.

b) en el primer cuarto de hora, no llegue ningún cliente.
c) en cualquier hora dada, llegue más de uno

78. Los clientes de una cafetería llegan a razón de nueve personas, en un período de 30 minutos. Calcule la probabilidad,
a) de que en la primera media hora por lo menos lleguen 4 personas.
b) de que en los 10 primeros minutos no llegue ningún cliente.

79. El cierre de bancos por problemas financieros ha ocurrido a razón de 5,7 clausuras por año,
a) encuentre la probabilidad de que ningún banco sea cerrado durante un período de cuatro meses,
b) por lo menos un banco sea cerrado durante el semestre.

DISTRIBUCIÓN IPERGEOM TRICA

Al igual que la distribución *Binomial* y de *Poisson*, la *Hipergeométrica* corresponde a variables aleatorias discretas.

La distribución *Hipergeométrica* está asociada, generalmente, con un proceso de muestreo sin reposición en una población finita.

Las características o condiciones que debe reunir una distribución *Hipergeométrica*, son:

a) La información de la muestra se toma sin reposición de una población finita.
b) La probabilidad de éxito no es constante, cambia para cada observación.
c) El resultado de una prueba es dependiente de la prueba anterior, siempre se verá afectado por el resultado de observaciones previas.
d) El tamaño de la muestra (n) debe ser superior en un 5% con respecto al tamaño poblacional N.
e) Se relaciona con situaciones que tengan que ver con dos o más resultados
f) La distribución es adecuada, cuando el tamaño de la población es pequeña.
Esta última condición limita su aplicación.

La fórmula que se debe emplear para la distribución *Hipergeométrica* es:

$$P_x = \frac{\binom{A}{}\binom{N-A}{n-}}{\binom{N}{n}}$$

Siendo:
N = Tamaño de la población
A = Número de éxitos en la población
n = Tamaño de la muestra
x = Número de éxitos en la muestra

Es necesario tener en cuenta que *X* no puede exceder a A ni a n. Recordemos que:

$$\binom{A}{} = \frac{A}{(A- \quad)} \qquad \binom{N}{n} = \frac{N}{(N-n)\ n}$$

Ejemplo 1. En la producción de cierto artículo, se sabe que por cada 50 producidos, en 43 su terminado es excelente. Si se toma una muestra de 12 artículos, ¿cuál es la probabilidad: (a) de que exactamente dos no sean clasificados como excelentes; b) por lo menos dos no sean clasificados como excelentes; c) ¿diez sean clasificados como excelentes?

o :

a) $P_{x=2} = \dfrac{\binom{7}{2}\binom{50-7}{12-2}}{\binom{50}{12}} = \dfrac{\binom{7}{2}\binom{43}{10}}{\binom{50}{12}} = 0{,}3317 = 33{,}17$

$N = 50$
$A = 7$
$n = 12$
$= 2$

b) $P_{x\geq 2} = 1 - \left[P_{(0)} + P_{(1)}\right]$

$$P_{x\geq 2} = 1 - \left[\frac{\binom{7}{0}\binom{50-7}{12-0}}{\binom{50}{12}} + \frac{\binom{7}{1}\binom{50-7}{12-1}}{\binom{50}{12}}\right] = 1 - \left[0{,}1263 + 0{,}3317\right]$$

$$P_{x\geq 2} = 1 - 0{,}45\ \ 0 = 0{,}5420 = 54{,}20$$

$N = 50$
$A = 7$
$n = 12$
≥ 2

c) $P_{(x=10)} = \dfrac{\binom{43}{10}\binom{50-43}{12-10}}{\binom{50}{12}} = \dfrac{\binom{43}{10}\binom{7}{2}}{\binom{50}{12}} = 0{,}3317 = 37{,}17$

$N = 50$
$A = 43$
$n = 12$
$= 10$

Ejemplo 2. Un colegio tiene a su disposición para el transporte de sus estudiantes 10 buses. Por información llegada a las directivas del plantel, se sabe que 4 no se encuentran en óptimas condiciones. Si se selecciona una muestra de 5 buses,

a) ¿Cuál es la probabilidad de que dos de ellos no se encuentren en óptimas condiciones?

b) ¿Qué dos de ellos se encuentren en perfectas condiciones?

o :

a) $P_{x=2} = \dfrac{\binom{4}{2}\binom{10-4}{5-2}}{\binom{10}{5}} = \dfrac{\binom{4}{2}\binom{6}{3}}{\binom{10}{5}} = 0{,}4762 = 47{,}62$

$N = 10$
$A = 4$
$n = 5$
$= 2$

b) $P_{x=2} = \dfrac{\binom{6}{2}\binom{10-6}{5-2}}{\binom{10}{5}} = \dfrac{\binom{6}{2}\binom{4}{3}}{\binom{10}{5}} = 0{,}23\ \ 1 = 23,\ 1$

$N = 10$
$A = 6$
$n = 5$
$= 2$

En los dos ejemplos anteriores, se pueden agilizar los cálculos mediante el uso de la tabla respectiva o la calculadora. Veamos su aplicación en el ejemplo 2, utilizando la tabla 6.7

Tabla 6.7 Distribución hipergeométrica

N		A (ó k)	x	P(x)
10	1	1	0	0,900000
10	1	1	1	0.100000
10	2	1	0	0.800000
10	2	1	1	0.200000
10	2	2	0	0.622222
10	2	2	1	0.355556
10	2	2	2	0.022222
10	3	1	0	0.700000
10	3	1	1	0.300000
10	3	2	0	0.466667

N	n	A (o k)	x	P(x)
10	5	3	0	0,083333
10	5	3	1	0.416667
10	5	3	2	0.416667
10	5	3	3	0.083333
10	5	4	0	0.023810
10	5	4	1	0.238095
10	5	4	2	0.476190
10	5	4	3	0.238095
10	5	4	4	0.023810
10	5	5	0	0.003968

USO DE CALCULADORA

En este capítulo se dan algunas indicaciones sobre el uso de aquellas calculadoras que más se utilizan en clase.

- Casio fx - 5000f
  - Primero se teclea el número 106 FMLA con lo cual aparece la fórmula para calcular la probabilidad.
  - En pantalla aparece un interrogante K? que equivale a A por lo tanto tecleamos (de acuerdo al último ejemplo y operación), 6 y EXE
  - Aparece X?, así que tecleamos 2 EXE y observamos la pantalla
  - Nos pide el valor de N? en este caso tecleamos 10 EXE
  - Ahora, pide n? siendo 5 y luego EXE ¿ aparece en pantalla el mismo resultado esperado del ejercicio 0,238095, ver ejemplo 2b.

  De acuerdo a las *marcas y modelos* de las calculadoras que fueron utilizadas, tanto en la *Binomial* como en *Poisson*, en todas ellas se procede así:

  (6 x^y nCr 2 X 4 nC_r 3) ÷ 10 nCr 5 =

  En algunas de de las calculadoras, antes de teclear el símbolo de las combinaciones se deberá proceder así: SHIFT nCr

- Cuando las calculadoras no tienen la tecla nCr se deberá proceder a teclear paso a paso, así:
  ((6 ! ÷ (4 ! x 2 !) x (4 ! ÷ (3 ! x 1 !)) ÷ (10 ! ÷ (5 ! x 5 ! x)) =
  Es necesario tener en cuenta con algunas calculadoras, que el uso de factorial requiere de dos teclas SHIFT ! ; También en otras, el factorial aparece en la tecla así X! ; equivalente a lo mismo.

Ejemplo 3. Aplicar la distribución hipergeométrica:

a) Dada la siguiente información:

N = 42 n = 12 A = 36 x = 8 buenos

b) N = 42 n = 12 A = 42 - 36 = 6 malos X = 12 - 8= 4 malos

o :

a) $$P_{(x=8)} = \frac{\binom{36}{8}\binom{42-36}{12-8}}{\binom{42}{12}} = \frac{\binom{36}{8}\binom{6}{4}}{\binom{42}{12}} = 0{,}0410 = 4{,}10$$

La probabilidad de obtener exactamente 8 buenos, es del 4,10%

b) $$P_{(x=4)} = \frac{\binom{6}{4}\binom{42-6}{12-4}}{\binom{42}{12}} = \frac{\binom{6}{4}\binom{36}{}}{\binom{42}{12}} = 0{,}0410 = 4{,}10$$

La probabilidad de obtener exactamente 4 malos, es del 4,10%

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

80. Un profesor dispone en su archivo de 15 preguntas sobre un tema específico de la materia; seis de ellas son de teoría. Si desea preparar un cuestionario de 5 preguntas:
    a) ¿Cuál es la probabilidad de que 2 de las preguntas sean de teoría?
    b) ¿Qué dos no sean de teoría?

81. Un Instituto Técnico recibe un pedido de 12 computadores que requiere para dotar la división de sistemas. Por información de otro Instituto que realizó ese mismo pedido, se supo que 4 aparatos habían sufrido desperfectos en el transporte, situación frecuente con la firma contratada para distribuir la mercancía. El director del instituto decidió examinar 3 de los computadores recibidos. ¿Cuál es la probabilidad de que ninguno de los tres resulte averiado al ser transportado?

82. Un nuevo sabor de helado fue probado por un grupo de 15 alumnos de un colegio en el que se vende, 10 encontraron agradable el nuevo sabor. Si quisiéramos repetir la prueba, se haría seleccionando 5 alumnos de los 15, ¿Cuál es la probabilidad de que, a) a dos les guste el nuevo sabor; b) a dos no les guste el nuevo sabor?

83. Un jefe de almacén sabe que 6 de las 25 bicicletas que tiene para la venta presentan fallas en los frenos y necesitan ajuste. Si el vendedor, que no tenía conocimiento de lo anterior, vendió en el día, 4 bicicletas, ¿Cuál es la probabilidad de que vendiera dos que requerían ser ajustadas?

84. Una Facultad de Economía cuenta con una nómina de 40 profesores, de los cuales 35 son economistas. Si seleccionan al azar, cinco profesores para constituir un comité, con el fin de revisar el perfil que debe tener el egresado, ¿cuál es la probabilidad de que por lo menos uno sea economista?

VARIABLE ALEATORIA CONTINUA

DISTRIBUCIÓN NORMAL

Corresponde a una Distribución de *variable aleatoria continua*, que se extiende sobre un campo de variabilidad infinito y está dada por la función:

$$Y = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{Z^2}{2\sigma^2}}$$

Siendo:

N = Número de datos
σ = Desviación estándar de la distribución binomial = $\sqrt{npq}$
e = Base de los logaritmos naturales = 2,71828
π = 3,1415926535…..
μ = Media de la distribución binomial = np

Se le suele denominar: Gaussiana, Laplaciana, Distribución de Laplace-Gauss o de Gauss-Laplace o bien la segunda ley de Laplace. Aparentemente fue descubierta por De Moivre (1756) como forma límite de la *Distribución Binomial*.

Todo ejercicio planteado como *Binomial*, en especial cuando (*n*) el número de experimentos realizados es grande, se debe resolver mediante la *distribución Normal*. Este procedimiento permite facilitar y agilizar las operaciones, cuyo resultado no será exacto (método binomial) sino un valor bastante aproximado, razón por la cual algunos lo identifican como *método aproximado*.

Tal como se hizo en la *distribución binomial*, el experimento de lanzar 12 monedas, nos permite elaborar una tabla de frecuencias teóricas o probabilisticas, con lo cual se obtendrán las probabilidades de aparición de 0 caras, 1 cara, 2 caras.....12 caras, con los resultados se podrá calcular la *media*, la *varianza* y la *desviación típica*, al igual como lo hacíamos en la *estadística descriptiva*, a fin de compararlo con los resultados obtenidos a través de la distribución binomial, donde:

$$\mu = np \qquad \sigma^2 = npq \qquad \sigma = \sqrt{npq}$$

Los resultados del experimento son:

Tabla 6.8 Distribución frecuencias

VARIABLE CONTINUA $X'_{i-1} - X'_i$	VARIABLE DISCRETA NÚMERO DE CARAS X_i	FRECUENCIA ESPERADA O TEÓRICA n_i	PROBABILIDAD P(x) (h_i)	$X_i n_i$	$X_i^2 n_i$
−0.5 — 0,5	0	1	1/4.096 = 0,000244	0	0
0,5 — 1,5	1	12	12/4.096 = 0,002930	12	12
1,5 — 2,5	2	66	66/4.096 = 0.016113	132	264

Continúa...

Continuación

2,5 — 3,5	3	220	220/4.096 = 0.053711	660	1.980
3,5 — 4,5	4	495	495/4.096 = 0,120850	1.980	7.920
4,5 — 5,5	5	792	792/4.096 = 0,193359	3.960	19.800
5,5 — 6,5	6	924	924/4.096 = 0,225586	5.544	33.264
6,5 — 7,5	7	792	792/4.096 = 0,193359	5.544	38.808
7,5 — 8,5	8	495	495/4.096 = 0,120850	3.960	31.680
8,5 — 9,5	9	220	220/4.096 = 0,053711	1.980	17.820
9,5 — 10,5	10	66	66/4.096 = 0,016113	660	6.600
10,5 — 11,5	11	12	12/4.096 = 0,002930	132	1.452
11,5 — 12,5	12	1	1/4.096 = 0,000244	12	144
	Σ	4.096	1,000 = 1,000000	24.576	159.744

$$\overline{X} = \frac{\sum X_i n_i}{N} \qquad \overline{X} = \frac{24.576}{4.096} = 6 \qquad \boxed{\overline{X} = \frac{\sum X_i f_i}{N}}$$

$$S^2 = \frac{\sum X_i^2 n_i}{N} - \overline{X}^2 \qquad S^2 = \frac{15.744}{4.096} - 36 = 3 - 36 = 3 \qquad \boxed{\sigma^2 = \frac{\sum X_i^2 f_i}{N} - \overline{X}^2}$$

> Teorema. Si p es la probabilidad de éxito de un suceso en un solo experimento y q es la probabilidad de fracaso, entonces la distribución binomial que da las frecuencias esperadas para 0, 1,....n éxitos en n experimentos, tiene como media y desviación típica σ

Se tiene: $= np$ y $\sigma = \sqrt{npq}$, fórmulas que aplicaremos al ejercicio anterior, en la siguiente forma:

$= np$ $\qquad = 12\ (1/2) = 6$

$\sigma^2 = npq$ $\qquad \sigma^2 = 12\ (1/2)\ (1/2) = 3$

$\sigma = \sqrt{npq}$ $\qquad \sigma = \sqrt{12\ (1/2)\ (1/2)} = \sqrt{3} = 1{,}732$

Representación gráfica

Con los datos de la Tabla 6.8 correspondiente a las frecuencias teóricas, elaboramos un *histograma de frecuencias*. Teniendo en cuenta que el número de aparición de caras corresponde a una *variable discreta* y al no haber establecido intervalos, se deberá colocar cada valor de la variable en el centro del rectángulo.

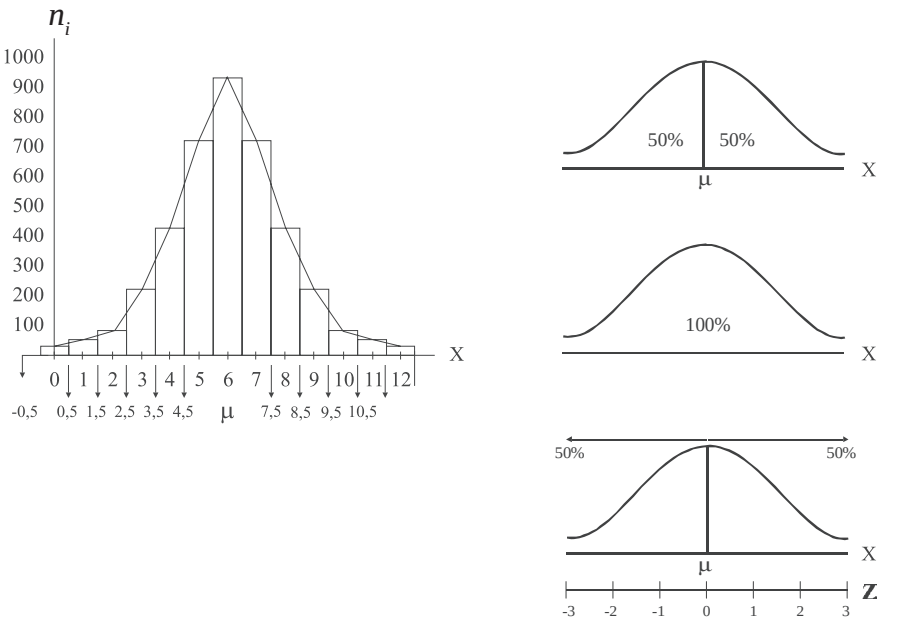
Uniendo los puntos medios en la parte superior de cada rectángulo, obtenemos una curva, conocida como ya se dijo con los nombres de *curva normal, curva de Gauss, curva de error o curva de probabilidad*.

El histograma binomial es *simétrico*, cuando $p = q = 1/2$. Tal como se presenta en el lanzamiento de una moneda; pero en el lanzamiento de un dado, por lo general, es diferente: $p \neq q$.

Se debe recordar que la variable en una distribución *binomial es discreta* y se debe transformar en una *variable continua*, cuando se quiere utilizar la *distribución normal*; dado que la amplitud del intervalo siempre va a ser uno ($c = 1$), se divide entre dos ($c/2$), siendo igual a 0,5. Ahora, al restarlo establecemos el límite inferior y al sumarlo se obtiene el límite superior de cada intervalo, así elaboramos la primera columna que aparece en la Tabla 6.8.

Con la *Variable continua,* es decir, la variable corresponde a intervalos, con ellos elaboramos el histograma y el polígono de frecuencias. Es necesario observar algunas condiciones que debe reunir la *curva normal* o *probabilística*.

Figura 6.5



Condiciones que debe reunir la curva normal o probabilística

- La curva es simétrica.
- El área bajo la curva es igual al 100%.
- La curva no toca el eje horizontal (X) ya que es asintótica, se prolonga indefinidamente.
- La media μ se localiza en el centro de la curva, es decir, cada parte es igual al 50%
- X toma valores de menor a mayor, es decir, de izquierda a derecha.
- Al estandarizar, convertir los valores de X en valores de Z, ésta tendrá una media $\mu_z = 0$ y $\sigma_z = 1$ y Z tomará valores desde -3 hasta +3 que cubre un área en el 99,7%, casi igual al 100%.
- La variante estadística $z = \frac{X - \mu}{\sigma}$ es una medida de las desviaciones estándar o de las llamadas unidades estandarizadas, conocida como desviación normal.
- Al usar la tabla 6.9 correspondiente a Áreas bajo la curva normal, ésta tomará a valores del centro a la derecha y del centro hacia la izquierda, dado que la tabla sólo presenta la mitad de la curva, así que el área de $Z = 0$ hasta $Z = 2$, será la misma de $Z = 0$ hasta $Z = -2$. Hay tablas donde el valor del área se presenta acumulada.

Ejemplo 1. En el lanzamiento de 12 monedas ¿Cuál es la probabilidad de obtener: a) exactamente 4 caras; b) por lo menos 2 caras; c) como máximo 9 caras?

o :

a) $p = 0,5 \quad q = 0,5 \quad n = 12$

Binomial: $P_{(x=4)} = \binom{12}{4} (0,5)^4 (0,5)^8 = 0,1208 = 12,08\%$ (Resultado exacto)

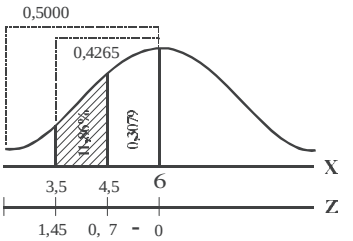
Normal: $= 12\,(0,5) = 6 \qquad \sigma = \sqrt{12\,(0,5)(0,5)} = 1,73$

Con la media y desviación típica en una distribución binomial, X es la variable discreta, ($X = 4$) se transforma en continua, sumando y restando 0,5 al valor de 4, siendo $P_{(3,5 \le x \le 4,5)}$ (no hay problema si se omite el igual).

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{4,5 - 6}{1,73} = -0,\ 7 \rightarrow A\ 0,307$$

$$Z = \frac{3,5 - 6}{1,73} = -1,45 \rightarrow A\ 0,4265 \quad \textit{(ver tabla)}$$



En este caso restamos:

$A\ (0,4265) - A\ (0,3079) = 0,1186$

$P_{(3,5 \le x \le 4,5)} = 11,86\ \%$ (resultado aproximado)

(*Ver tabla 6.1 Áreas bajo la curva normal*)

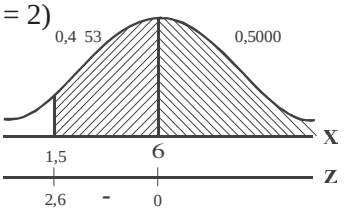
b) Por lo menos dos caras $X = 2, 3, 4, 5, 6.....12$

$p_{(x \ge 2)} = 1 - [p_{(x\,=0)} + p_{(x=1)}]$ (binomial)

$$P_{(x \ge 2)} = 1 - \left[\binom{12}{0} (0,5)^0\ (0,5)^{12} + \binom{12}{1} (0,5)^1\ (0,5)^{11} \right] = 99,69\% \text{ (resultado exacto)}$$

Normal: $= 6$ y $\sigma = 1,73 \qquad P_{(x>1,5)}$ (restamos 0,5 a $X = 2$)

$$Z = \frac{1,5 - 6}{1,73} = -2,60 \Rightarrow A(0,4\ 53)\ (\textit{ver tabla})$$



En este caso sumamos las dos áreas:

A(0,4953) + A(0,5000) = 0,9953 = 99,53%

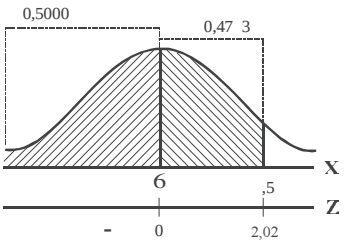
$P_{(x>1,5)} = 99,53\%$ (resultado aproximado)

c) Como máximo 9 caras $X = 0, 1, 2, 3, ...9$

$P_{(x \le 9)} = 1 - [P_{(10)} + P_{(11)} + P_{(12)}] = 98,08\%$ (Binomial)

Normal: $= 6 \qquad \sigma = 1,73$

$$Z = \frac{,5 - 6}{1,73} = 2,02 \Rightarrow A\ 0,47\ 3 \quad \textit{ver tabla} \qquad P_{(x \le 9,5)} = 97,83\%$$



Se suman las dos áreas: 0,5000 + 0,4783 = 0,9783 = 97,83%

Ahora procedemos a desarrollar un ejercicio, cuando el valor de μ y σ ya están calculados, en este caso x queda exactamente igual como lo dá el problema (como variable continua), en este caso se omite: *restar y/o sumar* 0,5

Para el manejo y cálculo de áreas, se le recomienda al lector, obligatoriamente hacer los ejercicios numerados desde el 83 hasta el 92 y así podrá saber cuando se suma y/o se resta.

Ejemplo 2. Un profesor manifiesta que el promedio que los estudiantes obtienen en su asignatura es de 3,9, con una desviación típica 0,35. ¿Cuál es la probabilidad que uno de sus alumnos obtenga:

a) una calificación superior a 4,4.

b) inferior a 3,2.

c) que gane la asignatura (mayor o igual a 3)?

o :

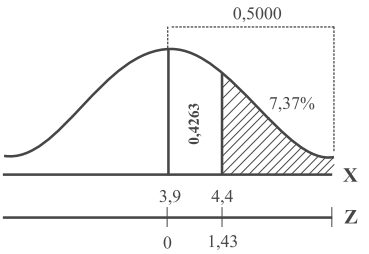
a) $\mu = 3, \qquad \sigma = 0,35 \qquad P_{(x>4,4)}$

$$Z = \frac{X - \mu}{\sigma}$$

$$Z = \frac{4,4 - 3,}{0,35} = 1,43 \Rightarrow A(0,4263)$$

$$0,5000 - 0,4263 = 0,0737 = 7,37$$

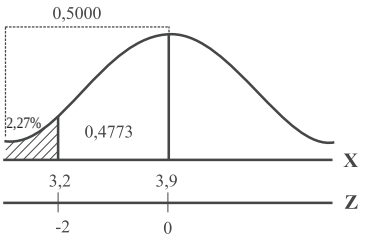
$$P_{(x \geq 4,4)} = 7,37$$



b) $P_{(x<3,2)} \qquad Z = \frac{3,2 - 3,}{0,35} = -2 \Rightarrow A(0,4773)$

$$0,5000 - 0,4773 = 0,0227 = 2,27$$

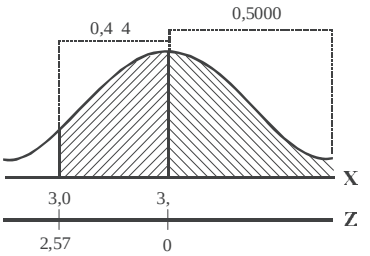
$$P_{(x \leq 3,2)} = 2,27$$



c) $P_{(x \geq 3)} \quad Z = \frac{3,0 - 3,}{0,35} = -2,57 \Rightarrow A(0,4\ \ 4\ \ )$

$$0,5000 + 0,4\ \ 4\ \ = 0, \quad 4\ \ = \quad ,4$$

$$P_{(x \geq 3)} = \quad ,4$$



Cada cantidad de la tabla corresponderá a la proporción bajo la curva que se encuentre entre Z = 0 y un valor positivo de Z. Las áreas para valores negativos de Z se obtienen por simetría.

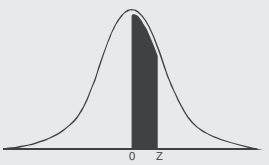
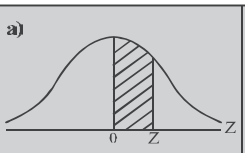
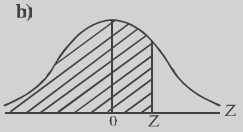
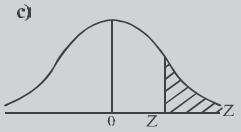


Tabla 6.9 Áreas de una distribución normal ordinaria

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,0
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,01	0,023	0,027	0,031	0,035
0,1	0,03 8	0,0438	0,0478	0,0517	0,0557	0,05 6	0,0636	0,0675	0,0714	0,0754
0,2	0,07 3	0,0832	0,0871	0,0 10	0,0 48	0,0 87	0,1026	0,1064	0,1103	0,1141
0,3	0,117	0,1217	0,1255	0,12 3	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,15 1	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,187
0,5	0,1 15	0,1 50	0,1 85	0,201	0,2054	0,2088	0,2123	0,2157	0,21 0	0,2224
0,6	0,2258	0,22 1	0,2324	0,2357	0,238	0,2422	0,2454	0,2486	0,2518	0,254
0,7	0,2580	0,2612	0,2642	0,2673	0,2704	0,2734	0,2764	0,27 4	0,2823	0,2852
0,8	0,2881	0,2 10	0,2 3	0,2 67	0,2 6	0,3023	0,3051	0,307	0,3106	0,3133
0,	0,315	0,3186	0,3212	0,3238	0,3264	0,328	0,3315	0,3340	0,3365	0,338
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,35	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,372	0,374	0,3770	0,37 0	0,3810	0,3830
1,2	0,384	0,386	0,3888	0,3 07	0,3 25	0,3 44	0,3 62	0,3 80	0,3 7	0,4015
1,3	0,4032	0,404	0,4066	0,4082	0,40	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,41 2	0,4207	0,4222	0,4236	0,4251	0,4265	0,427	0,42 2	0,4306	0,431
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,43 4	0,4406	0,4418	0,4430	0,4441
1,6	0,4452	0,4463	0,4474	0,4485	0,44 5	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,45 1	0,45	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,464	0,4656	0,4664	0,4671	0,4678	0,4686	0,46 3	0,4700	0,4706
1,	0,4713	0,471	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4762	0,4767
2,0	0,4773	0,4778	0,4783	0,4788	0,47 3	0,47 8	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4865	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,48 0
2,3	0,48 3	0,48 6	0,48 8	0,4 01	0,4 04	0,4 06	0,4 0	0,4 11	0,4 13	0,4 16
2,4	0,4 18	0,4 20	0,4 22	0,4 25	0,4 27	0,4 2	0,4 31	0,4 32	0,4 34	0,4 36
2,5	0,4 38	0,4 40	0,4 41	0,4 43	0,4 45	0,4 46	0,4 48	0,4 4	0,4 51	0,4 52
2,6	0,4 53	0,4 55	0,4 56	0,4 57	0,4 5	0,4 60	0,4 61	0,4 62	0,4 63	0,4 64
2,7	0,4 65	0,4 66	0,4 67	0,4 68	0,4 6	0,4 70	0,4 71	0,4 72	0,4 73	0,4 74
2,8	0,4 74	0,4 75	0,4 76	0,4 77	0,4 77	0,4 78	0,4 7	0,4 80	0,4 80	0,4 81
2,	0,4 81	0,4 82	0,4 83	0,4 83	0,4 84	0,4 84	0,4 85	0,4 85	0,4 86	0,4 86
3,0	0,4 87	0,4 87	0,4 87	0,4 88	0,4 88	0,4 8	0,4 8	0,4 8	0,4 0	0,4 0

Tabla 6.9 Áreas bajo la curva normal

	a)	b)	c)
			
0,00	0,0000	0,5000	0,5000
0,50	0,1 15	0,6 15	0,30 5
1,00	0,3413	0, 413	0,15 7
1, 6	0,4750	0, 750	0,0250
2,00	0,4772	0, 772	0,022
2,5	0,4 51	0, 51	0,004
3,00	0,4 7	0, 7	0,0000

En el primer caso: (a) el área comienza a crecer, desde el centro (donde Z=0) hacia la derecha (también hacia la izquierda por ser simétrica). En el segundo (b) crece de izquierda a derecha y hasta Z=0 el área corresponderá a la mitad de la curva siendo igual al 50%. Luego se prolonga hasta Z=3 donde el valor total del área será de 0,9987. En (c) el área correspondiente a la parte sombreada, su valor disminuye, como consecuencia de la diferencia entre el valor de la mitad de la curva (50%) y el área de columna, para obtener el valor de la zona sombreada.

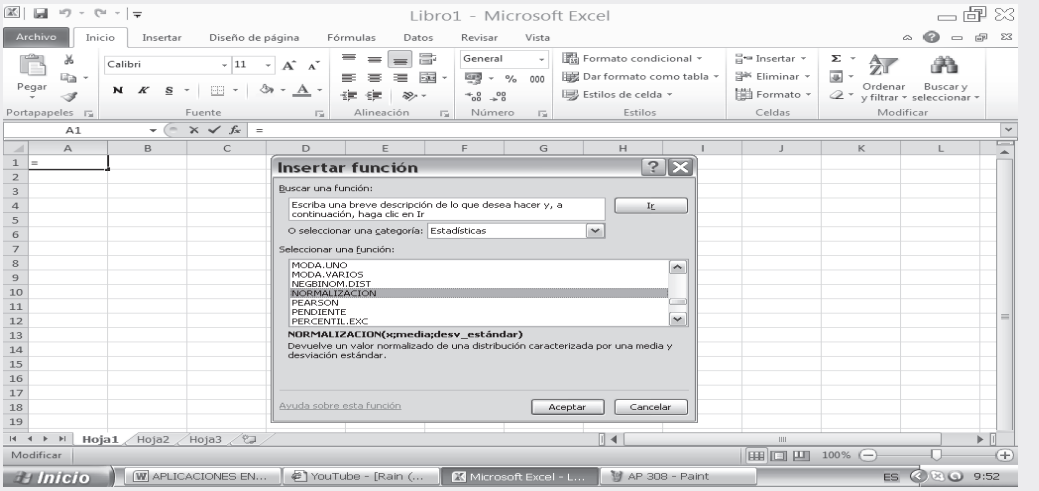
APLICACIÓN DEL EXCEL

Ejemplo 3.

A. Cuando solamente se requiera calcular el valor correspondiente a Z, se procederá de la siguiente forma:

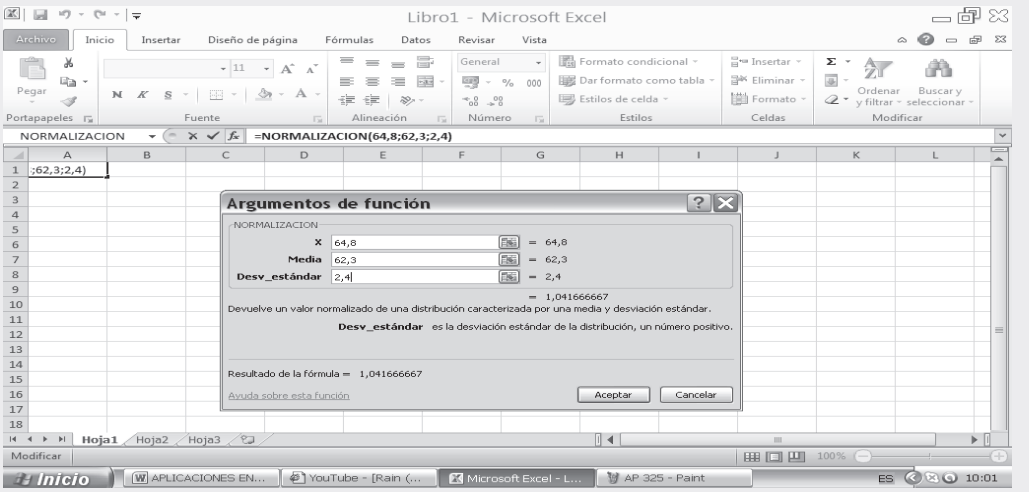
- Nos ubicamos en la función *fx*, haciendo CLIC en ella y seleccionamos NORMALIZACIÓN tal como se presenta en la siguiente figura:

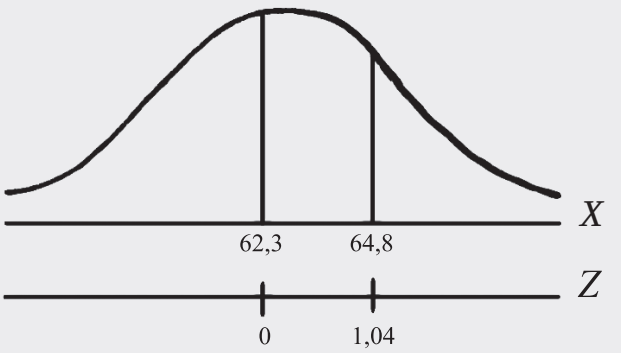
Figura No.7. Insertar Función



- Supongamos que se tiene la información: X = 64,8 = 62,3 σ = 2,4
- Digitamos el valor correspondiente X siendo en nuestro caso igual a 64.8. De la misma forma, en la siguiente casilla identificada como MEDIA, digitamos el valor correspondiente a 62.3 y finalmente en la casilla DESV_ ESTANDAR digitamos el valor de 2,4
- Al presionar ACEPTAR, observamos que el resultado de Z corresponde a 1.0416666667 con todos los decimales, este resultado sería exactamente igual al resultado que obtendríamos al utilizar las tablas que aparecen en el libro para el cálculo de Z

Figura No. 8. Argumentos de la Función





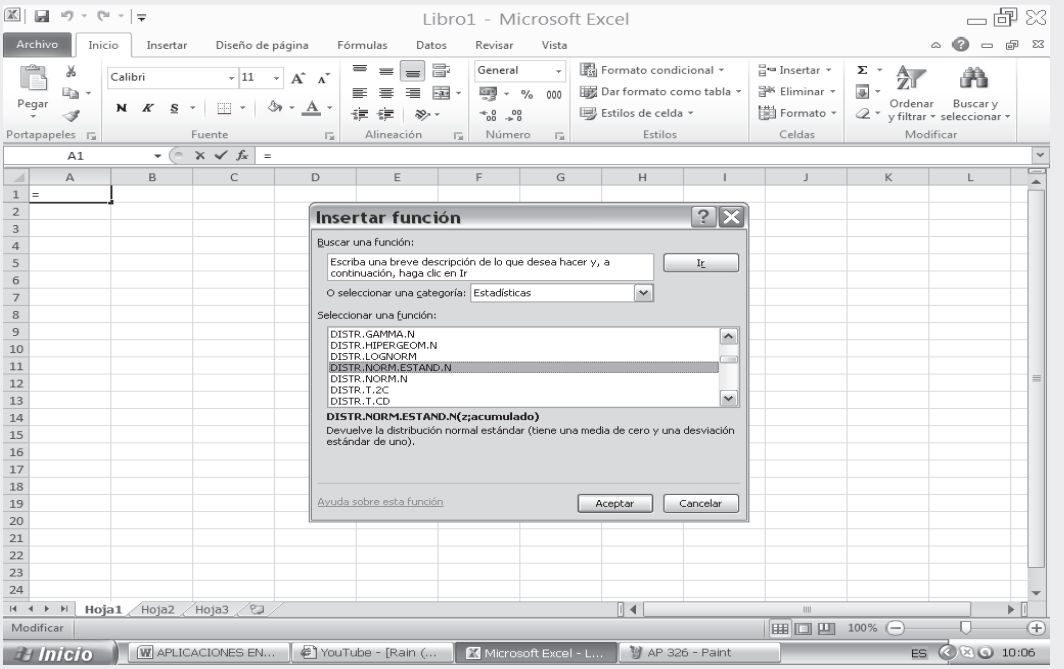
$$Z = \frac{X - \mu}{\sigma}$$

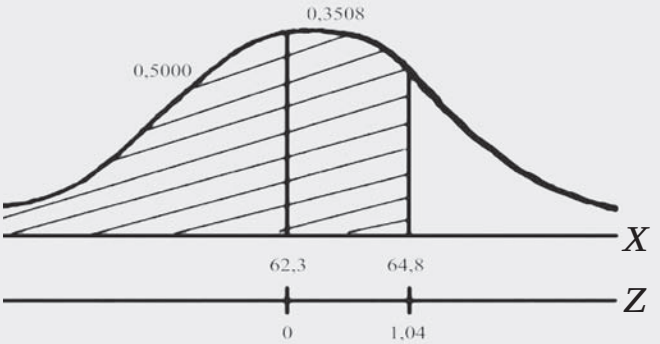
$$Z = \frac{64{,}8 - 62{,}3}{2{,}4} = 1{,}04$$

B. Si por el contrario nuestro objetivo es calcular el ÁREA BAJO LA CURVA NORMAL, utilizamos la función correspondiente a DIST.NORM.ESTAND.

P(X<64,8) = P(Z< 1,04) = 0,8508 = 85,08%

Figura No.9. Insertar Función

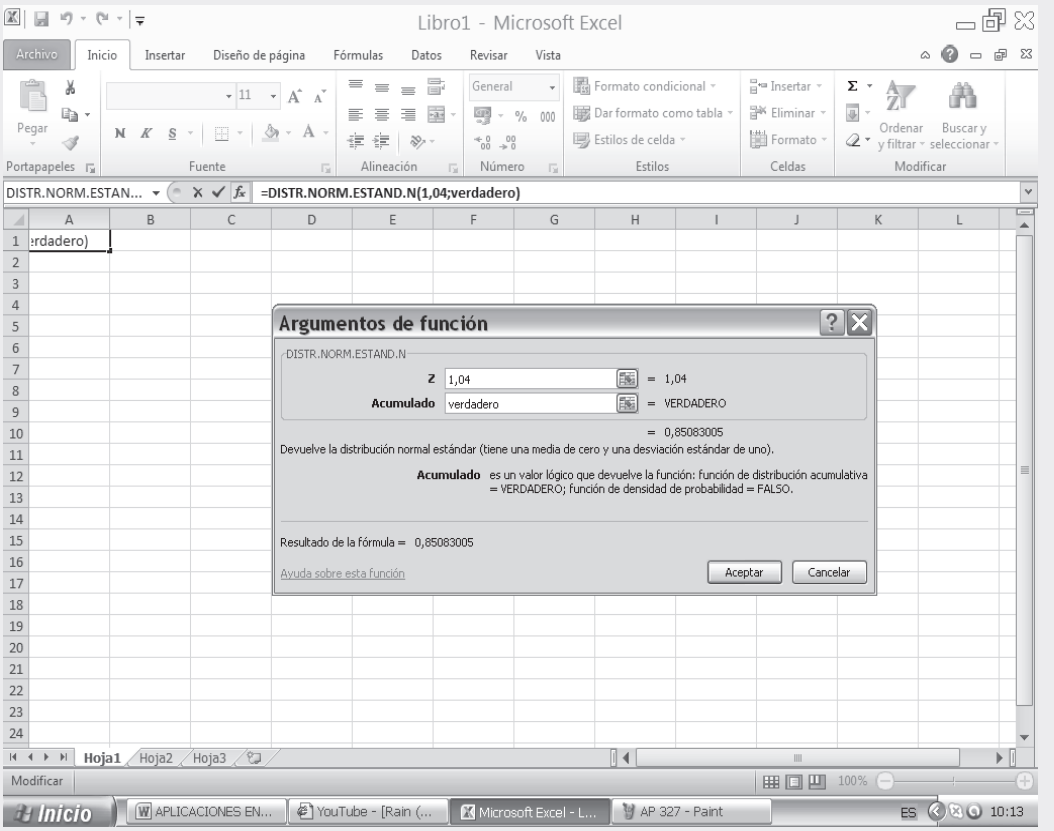




Observemos el cuadro siguiente donde aparece la casilla correspondiente al valor que tomará Z y un valor lógico que para nuestro caso lo consideraremos VERDADERO.

- Al digitar el valor de 1.04 aparece como resultado de esta operación el valor de 0.85083005, correspondiente al acumulado, es decir P(Z<1.04).

Figura No.10. Argumentos de la Función



C. La forma de DIST.NORM, nos permite agilizar las operaciones realizadas cuando utilizamos el procedimiento B, para hallar el área bajo la curva normal, así:

- Nos ubicamos en *fx*, y hacemos CLIC en DIST.NORM.
- En la figura correspondiente a DIST.NORM digitamos los siguientes datos: La X correspondiente al valor de la variable será de 64.8; la MEDIA en nuestro ejercicio es de ($\mu = 62,3$); y la DESV_ESTANDAR ($\sigma = 2,4$).
- Una vez digitados los anteriores datos, se tienen dos (2) alternativas: VERDADERO correspondiente al valor lógico, nos dará como resultado el valor de 0.851216854 (85.12%) correspondiente a la función distributiva acumulativa $P_{(X<64,8)} = 85,12\%$. Si por el contrario, utilizamos la palabra FALSO nos dará como resultado 0,09662266 (9,66%), correspondiente a la función de probabilidad bruta.
- Los dos casos anteriores pueden ser observados en los 2 cuadros que aparecen a continuación:

■ En la SEGUNDA PARTE DEL SIL que acompa a al libro, encontrará en conjunto toda la aplicación del EXCEL a la ESTADÍSTICA, que igual al capítulo 13, ofrece una BASE DE DATOS para trabajar.

Figura No. 11. Argumentos de la Función. Parámetro Verdadero

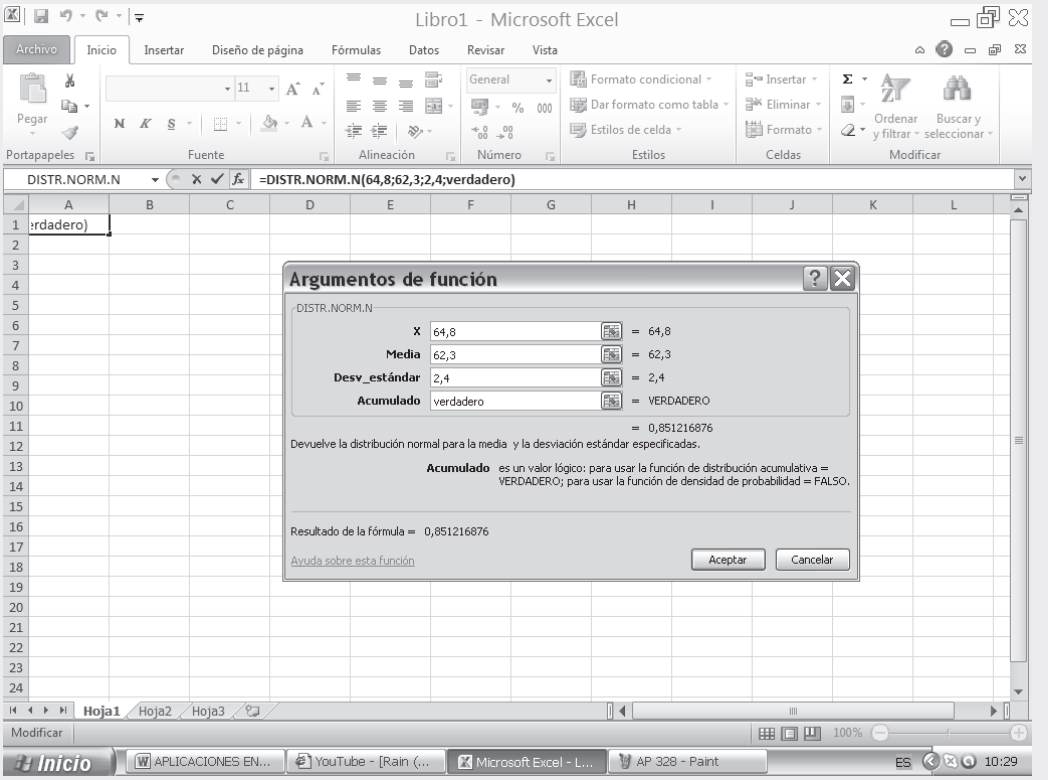
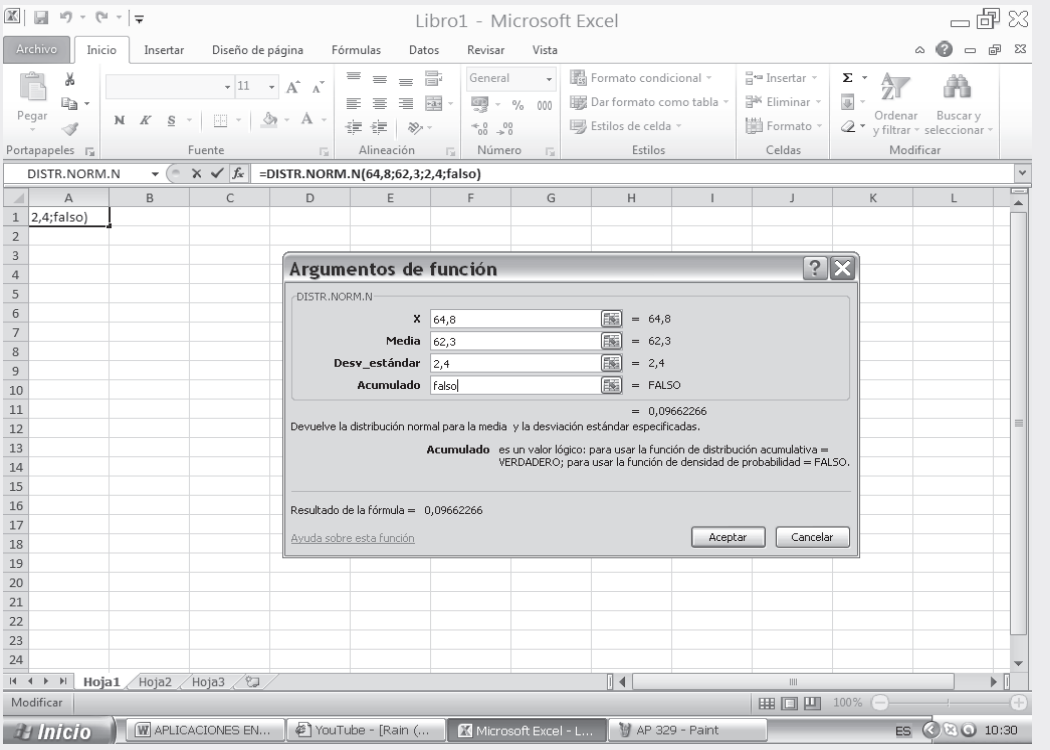


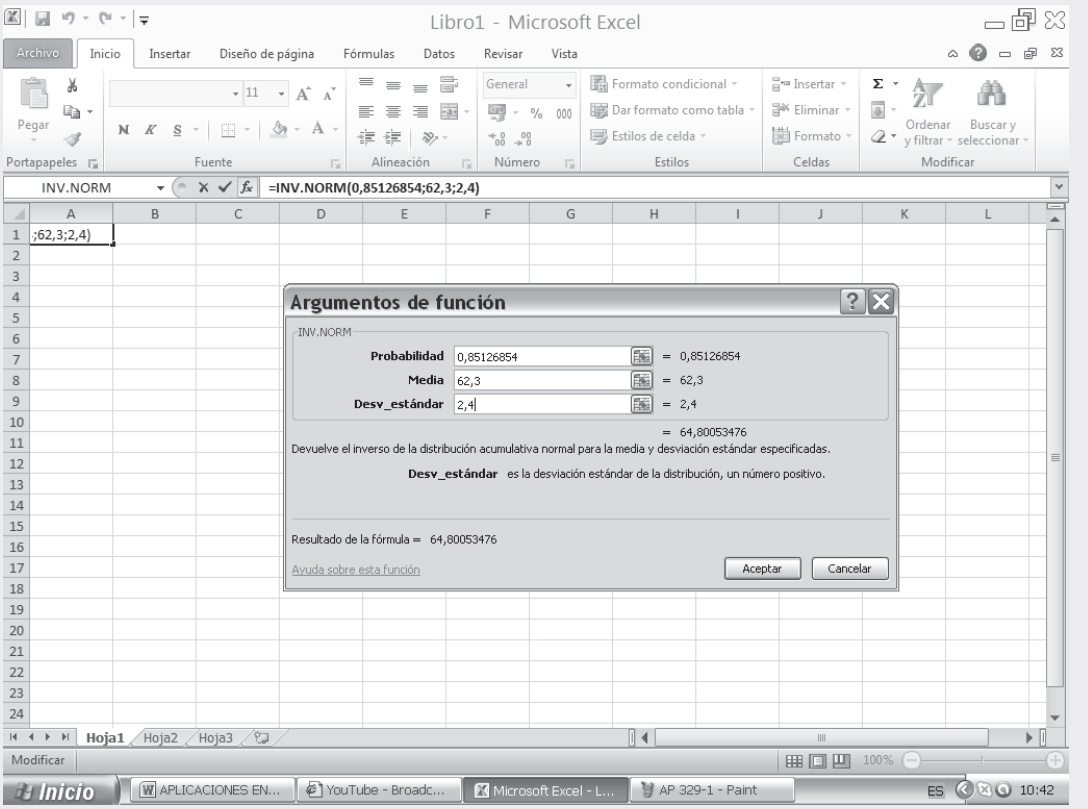
Figura No.12. Argumentos de la Función. Parámetro Falso (pantallazo adicional)



D. Otro procedimiento que podemos realizar, conociendo la probabilidad correspondiente al área bajo la curva normal, además de la MEDIA y DESVIACIÓN TÍPICA, corresponde INV. NORM Veámosla:

- ❑ Hacemos CLIC en función *fx* seleccionando INV. NORM.
- ❑ En el renglón PROBABILIDAD digitamos el resultado obtenido con la explicación del procedimiento C, correspondiente al valor de 0.85126854.
- ❑ En el segundo renglón digitamos el valor de la MEDIA, siendo igual a 62.3.
- ❑ Finalmente en el tercer renglón digitamos el valor correspondiente a la DESV_ ESTANDAR de 2.4.

Figura No.13. Argumentos de la Función



Nota: Para mayor comprensión, ver ejemplo 3. Con aplicaciones de Excel completo en el archivo Capítulo 6 ejemplo 3 en el SIL en internet.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

85. Hallar el área bajo curva normal:
a) Z = – 1,20 y Z = 2,40 b) Z = 1,23 y Z = 1,87; c) Z = -2,35 y Z = – 0,50

86. Determinar el área bajo la curva normal
a) a la izquierda de Z = –1,78 b) a la izquierda de Z = 0,56
c) a la derecha de Z = –1,45 d) correspondiente a Z < 2,16
e) correspondiente a – 0,80 < Z < 1,53

87. Encontrar el valor de Z:
a) el área a la derecha de Z es igual a 0,2266 b) el área a la izquierda de Z es 0,0314
c) el área entre – 0,23 y Z es igual a 0,5722 d) el área entre 1,15 y Z es 0,0730

88. Hallar Z si el área bajo la curva normal:
a) entre 0 y Z es 0,4515 b) a la derecha de Z es 0,3121 c) a la derecha de Z es 0,8023
d) a la izquierda de Z es 0,4562 e) entre – Z y Z es 0,7436

89. Hallar el área bajo la curva normal:
a) a la derecha de $Z = 2,68$ b) a la izquierda de $Z = 1,73$ c) a la derecha de $Z = – 0,66$
d) a la izquierda de $Z = –1,88$ e) entre $Z = 1,25$ y $Z = 1,67$ f) entre $Z = 0,90$ y $Z = – 1,85$
g) entre $Z = –1,45$ y $Z = 1,45$ h) entre $Z = 0,90$ y $Z = 1,58$

90. Si la media y la desviación estándar de una distribución normal: = 35,7 y σ = 2,8 cmts. Obtener los valores de Z (redondear en dos decimales las respuestas):
a) $X = 38,7$ cmts. b) $X = 31,5$ cmts. c) $X = 53,9$ cmts. d) $X = 29,6$ cmts.

91. Una variable aleatoria Z tiene distribución normal reducida (media 0 y varianza 1). Determinar las probabilidades utilizando la tabla de áreas bajo la curva:
a) $P_{(z < 0)}$ b) $P_{(1< z < 3)}$ c) $P_{(z > 3}$ d) $P_{(z = -1)}$ e)$P_{(2< z < -2)}$

92. Si X se encuentra distribuida normalmente con media 10 y desviación estándar 2, emplear la tabla de áreas para calcular la probabilidad de:
a) $X < 12$ b) $X > 11$ c) $X > 9$ d) $X > 9,5$ e) $9< X < 12$

93. Suponiendo que las estaturas (X) de varones de un colegio se encuentran distribuidas normalmente con media igual a 169 cm. y desviación estándar igual a 3 cm. (Emplear la tabla de áreas bajo la curva para calcular la probabilidad).
a) ¿Cuál es la probabilidad de que un estudiante tenga una estatura inferior a 165 cm?
b) ¿Qué porcentaje de alumnos tendrá una estatura entre 1,65 y 1,70?

94. En una distribución binomial de frecuencias, donde $p = ¼$, encuentre la probabilidad de obtener 25 o más éxitos en 80 experimentos.

95. Un fabricante de bombillas eléctricas ha encontrado que, en promedio, un 2% son defectuosas. ¿Cuál es la probabilidad que en 1.000 bombillas seleccionadas al azar, se encuentren 15 o más defectuosas?

96. Use la curva normal para encontrar la probabilidad de obtener exactamente 16 veces el seis en 96 lanzamientos de un dado; compare el resultado con el valor 0,110 obtenido con la distribución binomial.

97. Dada una curva normal con $\mu = 25,3$ y $\sigma = 8,1$. Hallar el área bajo la curva normal entre 20,6 y 29,1

98. Dada la curva normal con $\mu = 17,2$ y $\sigma = 3,5$. Hallar: a) el área bajo la curva normal a la derecha de 20, b) el área a la izquierda de 19,4 y c) el área entre 9,3 y 11, 7

99. El peso medio de las frutas de un gran cargamento es de 15,00 onzas, con una desviación estándar de 1,62 onzas; si sus pesos están distribuidos normalmente, ¿qué porcentaje de frutas tendrá un peso entre 15,00 y 18,00 onzas?

100. Si la vida media de cierta marca de baterías es de 30 meses, con una desviación estándar de 6 meses, ¿qué porcentaje de estas baterías puede esperarse que tengan una duración de 24 a 36 meses? Se supone que su duración sigue una distribución normal.

101. Se sabe que la duración media de los tubos de los receptores de televisión es de 3,0 años, con una desviación estándar de 1,5 años. Los tubos que duran menos de un año se reemplazan sin costo. Por cada 100 receptores vendidos (un tubo por receptor), ¿cuántos tubos deberán reemplazarse gratis?

102. En cierto negocio de construcción el salario medio mensual es de $686.000 y la desviación estándar de $4.500. Si se supone que los salarios tienen una distribución normal, ¿qué porcentaje de obreros perciben salarios entre $680.000 y $685.000.

103. Si una distribución normal de variable continua tiene $\mu = 21,2$ y $\sigma = 3,1$, encuentre la probabilidad de que una variable, seleccionada al azar, sea mayor de 30 o menor de 15.

104. La lluvia estacional media en cierto pueblo es de 18,75 pulgadas, con una desviación estándar de 6,50 pulgadas. Se supone que la lluvia estacional tiene una distribución normal. ¿En cuántos años, de un período de 50, se podrá esperar una lluvia de 15,00 a 25,00 pulgadas?

105. Dos estudiantes fueron informados de que habían recibido referencias tipificadas de 0,8 y – 0,4, respectivamente, en un examen de inglés. Si sus puntuaciones fueron 88 y 64, respectivamente, hallar la media y desviación típica de las puntuaciones del examen.

106. La media del peso de 500 estudiantes en un cierto colegio es de 151 libras y la desviación típica de 15 libras. Suponiendo que los pesos se distribuyen normalmente, hallar el número de estudiantes que pesan: a) entre 120 y 155 libras; b) más de 185 libras.

107. Las puntuaciones de un ejercicio de biología 0, 1, 2, ... dependiendo del número de respuestas correctas a 10 preguntas formúladas. La puntuación media fue de 6,7 y la desviación típica de 1,2. Suponiendo que las puntuaciones se distribuyen normalmente, determinar: a) el porcentaje de estudiantes que obtuvo 6 puntos; b) la puntuación máxima del 10% más bajo de la clase; c) la puntuación mínima del 10% superior de la clase.

108. La puntuación media en un examen final fue de 72 y la desviación típica de 9. El 10% de los mejores alumnos recibió la calificación A. ¿Cuál es la puntuación mínima que un estudiante debió tener para recibir una A?

109. Si las estaturas de 10.000 alumnos universitarios tienen una distribución normal, con media de 169 centímetros y desviación estándar de 2,5 centímetros.
a) ¿Cuántos alumnos tendrán por lo menos 172 centímetros?
b) ¿Cuál es el intervalo que incluye al 75% central de alumnos?

110. Tres estudiantes presentan varios exámenes. A, obtiene un puntaje de 72; B, de 85 y C, de 17. Todos los estudiantes que presentaron el examen A, obtuvieron un promedio de 85, los que presentaron B, de 90 y los de C, promediaron 25. Las respectivas desviaciones estándar fueron 7,3 y 7. Disponga a los estudiantes en orden de capacidad, juzgada por estos resultados.

111. Al calibrar ciruelas cocotas, cuyos pesos están distribuidos normalmente, un 20% es pequeño; 55% mediano; 15% grande y 10% extra grande. Si el peso medio de las ciruelas es de 4,83 onzas, con una desviación estándar de 1,20 onzas. ¿Cuáles son los límites superior e inferior del peso de las ciruelas medianas?

112. Un conjunto de 10.000 observaciones tiene una distribución normal con media de 450. Si 1.700 de ellas están comprendidas entre 450 y 460, ¿cuál es la desviación estándar?

113. En una distribución normal que tiene una desviación estándar de 2,00, la probabilidad de que el valor de una variable, elegida al azar, sea mayor de 28, es 0,03.
   a) Calcule la media de la distribución.
   b) Obtenga el valor de la variable que supera el 95% de los valores.

114. En un examen la nota media fue de 70,0 y la desviación estándar 10,0. El profesor da a todos los estudiantes con notas de 61,0 a 79,0 la calificación C. Hubo 24 alumnos con C. Si se supone que las calificaciones siguen una distribución normal, ¿cuántos alumnos se examinaron?

115. En una distribución normal con media 15,00 y desviación estándar 3,50, se sabe que 647 observaciones son mayores a 16,25. ¿Cuál es el número total de observaciones?

116. Un profesor califica un 10% de los exámenes con A; 20% con B; 40% con C; 20% con D y 10% con E. Se obtiene 68,0 como promedio en un examen. Si el límite entre C y B es de 78,0 en ese examen y si las calificaciones están normalmente distribuidas, ¿cuál es la desviación estándar del curso?

117. La estatura de un número de estudiantes está distribuida normalmente con una media de 168,50 centímetros. Si un 12% mide por lo menos 171,20 centímetros. ¿Cuál es la desviación de esta distribución?

118. En una distribución normal con media de 100 y desviación estándar de 53, existen 135 observaciones mayores de 200. ¿Cuántas observaciones hay entre 150 y 200?

119. En una distribución normal, con media 72,0 y desviación estándar 12,0, existen 220 observaciones entre 42,0 y 90,0. ¿Cuántas observaciones comprende toda la distribución?

120. En una distribución normal con media de 120 y desviación estándar de 30,0 existen 300 observaciones entre 130 y 150. ¿Cuántas observaciones existen entre 130 y 145?

121. En un examen, 13% del curso obtuvo una calificación de A; 20% de B; 48% de C; 10% de D, y 9% de E. La calificación C incluye desde 55 a 79 puntos. En el supuesto de una distribución normal, ¿cuál es la media y la desviación estándar de las calificaciones?

122. Si la distribución de los períodos de duración de cajas telefónicas metálicas es tal que el 9,51% excede los 15 años y el 62,55% los 9 años, ¿cuál es la media y la desviación estándar, si se admite que la distribución es normal?

123. En promedio, el 10% de las varillas de madera usadas en cierto producto son demasiado nudosas para usarlas. ¿Cuál es la probabilidad de que en un paquete de 15 varillas: a) exactamente 5 sean demasiado nudosas; b) por lo menos 10 estén demasiado nudosas; c) no más de 4 estén demasiado nudosas?

124. Como regla, 25% de los productos manufacturados por un cierto torno, son defectuosos, ¿cuál es la probabilidad de que en 20 de estos productos haya: a) exactamente 15 defectuosos; b) menos de 6 defectuosos; c) por lo menos 8 defectuosos?

125. En promedio, el 20% de las varillas de madera usadas en cierto producto son demasiado nudosas. ¿Cuál es la probabilidad de que en un paquete de 200 varillas, exactamente 5 estén demasiado nudosas?

126. Como regla el 15% de ciertos productos manufacturados por un torno, son defectuosos. ¿Cuál es la probabilidad de que en 1.000 de estos productos haya 100 defectuosos?

127. El 10% de ciertas unidades compradas por un almacén son inadecuadas para la venta. ¿Cuál es la probabilidad de que:
a) 42 o menos resulten inadecuadas en un lote de 500;
b) menos de 18 sean inadecuadas en un lote de 300?

128. Al inspeccionar 2.330 soldaduras producidas por cierto tipo de máquina se encontraron 448 uniones defectuosas. Al revisar 5 soldaduras, ¿cuál es la probabilidad de encontrar 3 ó más defectuosas?

129. La probabilidad de que un alumno que entra al primer semestre se gradúe, es de 0,4. Encuentre la probabilidad de que de 5 amigos que entran al primer semestre:
a) solamente uno se gradúe; b) ninguno se gradúe.

130. Si en general el 25% de los pacientes afectados por una enfermedad mueren de ella, ¿Cuál es la probabilidad de que, en un grupo de 50, mueran 6?

131. Una caja contiene 1.280 bombillas. El 15% de ellas son defectuosas; ¿cuál es la probabilidad de que una muestra al azar de 120 bombillas contenga menos de 5 defectuosas?

132. La tasa de mortalidad para cierta enfermedad es 18 por 1.000. ¿Cuál es la probabilidad de que 5 mueran de esta enfermedad, en un grupo de 500?

133. Si la distribución de X es normal, con media $\mu = 1$ y $\sigma = 0,4$; hallar: a) $P_{(x>0)}$ b) $P_{(0,2 < x < 1,8)}$

134. Pruebas realizadas en bombillas eléctricas de cierta marca, indican que el período de duración se distribuye normalmente con media igual a 1.860 horas y desviación estándar igual a 68 horas. Estimar el porcentaje de bombillas con una duración de:
a) más de 2.000 horas; b) menos de 1.750 horas.

135. El tiempo empleado para ir de un hotel al aeropuerto por la ruta A se distribuye normalmente, con media igual a 27 minutos y desviación típica igual a 5 minutos; por la ruta B la distribución es normal con media igual a 30 y desviación típica igual a 2. ¿Qué ruta conviene utilizar si se dispone de:
a) 30 minutos; b) 34 minutos?

136. Se sabe que la duración de las conversaciones telefónicas en una oficina tiene una distribución normal, cuya media es 4 y desviación estándar de 2 minutos, ¿Cuál es la probabilidad de que una conversación dure más de cinco minutos?

137. Los ingresos en cierto sector económico se distribuyen normalmente. La clasificación de los grupos económicos de mayor a menor ingreso son los siguientes:

GRUPO	% DE INDIVIDUOS EN EL GRUPO
A	10
B	19
C	40
D	20
E	11

Si el ingreso del grupo C está comprendido entre $ 750.000 y $ 780.000 mensuales, calcule la media y la desviación estándar de los ingresos mensuales del sector.

138. Encontrar la probabilidad de que las 5 primeras personas que se encuentren cierto día, por lo menos 3 hayan nacido en domingo:
a) mediante el método exacto (binomial).
b) mediante el método aproximado (normal).
c) mediante la distribución de Poisson.

139. ¿Cuál es la probabilidad de que 36 personas de una población de 360, de las cuales 252 son fumadoras, sean todas fumadoras?

140. En una distribución normal que tiene una desviación estándar de 2, la probabilidad de que el valor de una variable al azar sea mayor de 30 es de 0,05. a) Calcule la media de la distribución;
b) Obtenga el valor de la variable que es superado por el 95% de los valores.

141. Un industrial sabe que, en promedio, un 8% de sus productos es defectuoso. ¿Cuál es la probabilidad de que un lote de 100 piezas tenga por lo menos 15 defectuosas?

142. Un conjunto de 10.000 observaciones tiene una distribución normal con media 450. Si 1.900 de ellas están comprendidas entre 440 y 460, a) ¿cuál es la desviación estándar?; b) ¿cuántas observaciones hay entre 460 y 470?

143. Un par de dados perfectos se lanzan 50 veces. ¿Cuál es la probabilidad de que la suma de sus caras dé un valor de 6, de 7 o de 8, aparezcan 25 veces o más?

144. Una fábrica de aluminio produce, entre otras cosas, cierto tipo de canal de una aleación de aluminio. Por experiencia se sabe que la rapidez, medida en psi efectivos, está normalmente distribuido, con media 2.425 psi y desviación estándar de 115 psi. ¿Cuál es la probabilidad de que la canal seleccionada tenga un valor superior a 2.500 y menos de 2.250 psi?

145. Al hacer girar (impulsar) la flecha de una ruleta, su punta indicadora tiene la misma probabilidad de detenerse en cada uno de los cuatro cuadrantes. Si la flecha se hace girar 18 veces, ¿cuál es la probabilidad de que su punta se detenga menos de 3 veces en el primer cuadrante?

146. En una distribución con media 15 y desviación estándar 3,5, se sabe que 647 observaciones son mayores que 16,30. ¿Cuál es el número total de observaciones?

147. El gerente de crédito de un almacén de artículos electrodomésticos estima las pérdidas por malos clientes durante el año, en la siguiente forma: la pérdida tiene distribución normal con media de $50.000; además, la probabilidad de que sea mayor de $60.000 y menor de $40.000 es de 0,40, ¿cuál es la desviación estándar?

148. La vida útil de las pilas de cierta marca están distribuidas normalmente. Si el 7,68% de las pilas duran más de 54 horas y 39,80% duran menos de 50 horas, ¿cuál es la media y la desviación estándar?

149. La taberna Sancho ha instalado una máquina automática para la venta de sifón. La máquina puede regularse de modo que la cantidad media de sifón sea la que se desee; sin embargo, en cualquier caso esta cantidad tendrá una distribución normal, con una desviación estándar de 5,9 cm^3.
   a) Si el nivel se ajusta 304,6 cm^3, ¿qué porcentaje de los vasos contendrán menos de 295,7 cm^3?
   b) ¿A qué nivel medio debe ajustarse la máquina para que sólo el 2,28% de los vasos contengan menos de 295 cm^3?
   c) ¿A qué nivel medio debe ajustarse la máquina para que el 84,13% de los vasos contengan menos de 313,46 cm^3?

150. La duración de las pilas Gato están distribuidas normalmente con una media de 80 horas y una varianza de 100 horas. El fabricante garantiza que reemplazará cualquier pila que falle antes de cumplirse la garantía.
   a) ¿Cuánto tiempo debe dar de garantía de modo que no más del 5% de las pilas fallen antes de ese tiempo?
   b) Si en una semana vende 500 pilas, ¿cuántas reemplazará?

151. En una fábrica, el tiempo para producir un artículo está distribuido normalmente, con un promedio de 50 minutos y una varianza de 25 minutos. Se debe fabricar una partida de 80.000 artículos.
   a) ¿Cuántos artículos requerirán de un tiempo de fabricación mayor de 53 minutos?
   b) ¿Cuántos artículos requerirán de un tiempo de fabricación no inferior a los 48 minutos, ni superior a los 53 minutos?
   c) El 50% de los artículos requerirán de un tiempo de fabricación entre X_1 y X_2 minutos. Determine los valores de X_1 y X_2, si ellos son simétricos con respecto al tiempo medio.

152. Un fabricante de transformadores de corriente, asegura que los aparatos que vende tienen una vida útil media de 80.000 horas y una desviación estándar de 8.000. Suponiendo que esta vida útil está distribuida normalmente.
   a) ¿Cuál es la probabilidad de que un transformador dure más de 96.000 horas?
   b) El 50% de los transformadores duran entre X y X horas. Encuentre los valores de X_1 y X_2, si ellos son simétricos respecto a la media.
   c) El fabricante garantiza que reemplazará gratis cualquier transformador cuya duración sea inferior a X. Determine el valor de X de modo que tenga que reemplazar sólo el 1% de los transformadores.

153. Una compañía de seguros considera que más o menos al 0,05 de la población le ocurre cierto tipo de accidentes cada año. La empresa tiene 1.000 asegurados contra este tipo de accidentes. ¿Cuál es la probabilidad de que, como máximo, 35 de ellos sufran este accidente?

154. Si en la producción diaria de envases metálicos de una fábrica se sabe que el 1% son defectuosos, ¿Cuál es la probabilidad que de 1.000 envases producidos en un día:
a) por lo menos tres no sean buenos.
b) cómo máximo 980 buenos?

155. Se sabe que el 25% de los estudiantes de una universidad usa anteojos. ¿Cuál es la probabilidad de que en un curso de 40 alumnos, a) por lo menos cinco de ellos usen gafas. b) más de 33 no las usen?

156. En examen practicado a 1.500 estudiantes, la calificación promedio fue de 3,6 y la desviación típica de 0,6. Si las calificaciones se distribuyen normalmente,
¿Cuántos estudiantes obtuvieron:
a) calificaciones entre 3,5 y 4,0. b) calificaciones entre 2,0 y 3,0. c) calificaciones de 4,0 y más?

157. Si la estatura promedio de un grupo de 1.000 personas fueron de 160 centímetros y la varianza de 100; además se sabe que se distribuyeron normalmente, ¿Cuántas personas miden:
a) entre 140 y 165 centímetros. b) entre 170 y 180 centímetros.
c) 185 y más. d) 130 y menos?

158. Un taller de reparación de televisores a color, gasta en promedio 45 minutos en el arreglo de un aparato, con una desviación típica de ocho minutos. Si la población se distribuye normalmente. ¿cuál es la probabilidad de que en el arreglo de un televisor se gasten más de 50 minutos?

159. La vida útil de las pilas de cierta marca están distribuidas normalmente; si el 6,88% de las pilas duran más de 56 horas y el 30,85% duran menos de 52 horas,
a) ¿Cuál es la media y la desviación típica?
b) Si se tienen 400 pilas, ¿cuántas de ellas esperamos que tengan una duración superior a las 56 horas?

160. Una cuarta parte de los documentos archivados diariamente por un empleado de un departamento de ventas se hace equivocadamente. Si en un día se archivan 100 documentos, ¿cuál es la probabilidad de que:
a) por lo menos 18 documentos sean mal archivados.
b) exactamente 16 documentos sean mal archivados.
c) exactamente 86 documentos sean correctamente archivados?

161. Si la vida media de una batería de 12 voltios, es de 30 meses, con una desviación típica de 6 meses:
a) determine qué porcentaje de baterías dura menos de 18 meses;
b) si la garantía establece el cambio de la batería, si su duración es menor de 18 meses, ¿cuántas baterías tendrá que cambiar un almacén, de 360 vendidas en el trimestre?

162. La duración de ciertas pilas para radio transistor, están distribuidas normalmente; si el 2,28% duran más de 4,26 meses y 5,36% duran menos de 1,25 meses, determine la duración media y la desviación estándar.

163. El peso medio de una fruta es de 4 libras. El 15% de esas frutas pesan menos de 3 libras. Suponiendo que los pesos están distribuidos normalmente, ¿cuál es la desviación típica?

164. Un fabricante de juguetes considera que el lanzamiento de un nuevo juguete para navidad producirá una venta promedio de 80.000 unidades, si además piensa que las ventas están distribuidas normalmente y que existe una probabilidad del 25 % de vender más de 100.000 unidades, ¿Cuál es la desviación estándar?

165. Una asignatura se califica de 0 a 100 y al final del curso se observa que las calificaciones están distribuidas normalmente con una media de 72 y una desviación estándar de 10.
   a) Si el porcentaje mínimo para aprobar es 60, ¿cuál es la probabilidad de fracasos?
   b) Si el 80% de los estudiantes de esa asignatura deben ser aprobados, ¿Cuál debe ser la calificación mínima aprobatoria?

166. Se ha comprobado que el tiempo necesario para atender a un contribuyente en una oficina de recaudación de impuestos nacionales está normalmente distribuida, con media de 15 minutos y desviación típica de 5 minutos. ¿Cuál es la probabilidad de que un contribuyente seleccionado aleatoriamente:
   a) requiera menos de 10 minutos para terminar su diligencia.
   b) dure más de 20 minutos en la ventanilla?

167. Una fábrica de cemento empaca su producto en bolsas cuyo peso está distribuido normalmente, con una media de 50 kilos y una varianza de 4 kilos. Encuentre la probabilidad de sacar una bolsa que contenga:
   a) por lo menos 51 kilos. b) como máximo 51 kilos. c) de 49 a 52 kilos.

168. La inspección de bolsas de 20 kilos de arroz conducen al resultado de que el 60% pesan entre 19,20 y 20,80 libras; calcule la varianza, suponiendo que la distribución es normal.

169. La fábrica de refresco Tan envasa su producto en frascos cuyo peso neto tiene distribución normal, con desviación típica de 6,3 gramos. Si el 8% de los frascos tienen un peso mayor de 142 gramos. ¿Cuál es el peso medio de ellos?

170. El gerente de producción de una fábrica de bombillas, estima que la vida útil del producto está distribuida normalmente, con una media de 5.000 horas. Si, además, el gerente considera que hay una probabilidad del 60% de que la bombilla dure más de 5.568 y menos de 4.432 horas,
   a) ¿Cuál es la desviación típica?
   b) Si en un día se producen 20.000 unidades, ¿cuántas esperamos que tengan una duración entre 4.500 y 5.500 horas?

171. En una clínica de los Seguros Sociales, se establece que el período de hospitalización está distribuido normalmente, con una media de 7,6 días y una desviación típica de 2,2 días. ¿Cuál es la probabilidad de que un individuo que sea internado permanezca,
   a) por lo menos cuatro días. b) no más de nueve días?

172. Dado un cuestionario de cierto o falso de 100 preguntas, determinar la probabilidad de que una persona acierte,
   a) por lo menos 60 preguntas.
   b) no menos de cuarenta ni más de 60.
   c) si se presentan 1.000 estudiantes al examen, cuántos esperamos que presenten el resultado del punto b?

173. Los ingresos de un grupo económico se distribuyen normalmente. La clasificación de los grupos económicos, de mayor a menor ingreso, es la siguiente:

GRUPO:	A	B	C	D	E
PORCENTAJE:	8	16	42	20	14

El grupo C está comprendido entre $736.500 y $760.000 quincenales.
a) Calcular la media y desviación estándar.
b) Si en ese sector se ocupan 25.000 personas, ¿cuántas se esperan que tengan un ingreso superior a $770.000 quincenales?

174. Los ingresos en cierto sector económico se distribuyen normalmente; la clasificación de los grupos económicos de mayor a menor ingreso, son los siguientes:

GRUPO:	A	B	C	D	E
PORCENTAJE:	8	16	38	24	14

Si el ingreso del grupo C está comprendido entre $590.000 y $860.000 semanales,
a) calcular la media aritmética y la desviación típica.
b) si en dicho sector se ocupa un total de 23.000 personas ¿cuántas esperamos que tengan un ingreso inferior a los $500.000?

175. La duración de ciertas pilas de radio transistor, están distribuidas normalmente. Si el 2,30% duran menos de 3,8 meses y 54,8% más de 4,36 meses, a) ¿Cuál es la media y la desviación estándar? b) Si se tienen 2.000 pilas, ¿cuántas esperamos que tengan una duración superior a los 5 meses?

176. Se sabe que el 85% de los estudiantes de un curso aprueban el semestre. ¿Cuál es la probabilidad de que 9 o más de un total de 50, no aprueben?

177. Un conjunto de 20.000 observaciones tiene una distribución normal con media de 480. Si 4.000 están comprendidas entre 470 y 490,
a) ¿cuál es la desviación estándar?
b) ¿cuántas observaciones hay entre 460 y 470?

178. Si las calificaciones obtenidas en un examen, con notas entre 0 y 100, tienen una distribución normal, con media $\mu = 72$ y desviación estándar igual 4, a) ¿Cuál es la calificación mínima que debe recibir un alumno, para estar dentro del 10% de los alumnos que obtuvieron las notas más altas en el examen? b) ¿Si entre las calificaciones 60 y 80, hay 20 alumnos, cuántos tiene el curso? c) ¿Cuál es la máxima nota que debe recibir un estudiante, para encontrarse entre el 72% de los alumnos, con las notas más bajas del examen?

179. El propietario de un restaurante ha determinado que la demanda diaria de carne molida en su negocio tiene una distribución normal, con una media de 240 kg y una varianza de 529 kg^2. ¿Qué cantidad de carne molida debe estar disponible diariamente, para que la probabilidad de que se agote la dotación no sea mayor del 2%?

180. El control de inventarios en las librerías universitarias es un problema bastante complicado. Si se piden pocos ejemplares de un determinado libro, los costos aumentan, por lo que es necesario hacer pedidos grandes. Si se piden demasiados ejemplares, se corre el riesgo de que el libro no se requiera como

texto en el futuro, lo que dificultará su venta. Suponga que el promedio de alumnos matriculados en estadística, en la universidad durante un semestre, tiene una distribución normal, con media de 150 alumnos por semestre y una desviación típica de 20 estudiantes. ¿Cuántos libros de texto se deben pedir, para que la probabilidad de que se terminen las existencias no sea mayor de 0,10?

181. La duración de un determinado tipo de lavadora automática, tiene una distribución aproximadamente normal, con una media de 3,1 años y una desviación típica de 14,4 meses. Si la lavadora está garantizada por un año,
a) ¿qué proporción del total de unidades vendidas tendrán que ser reemplazadas?
b) Si se vendieron 500, ¿cuántas de ellas serán reemplazadas?

182. En un examen la nota media fue de 3,5 y la desviación estándar de 0,5. La calificación C corresponde a notas entre 3,05 y 3,95. Hubo 24 alumnos con nota C.
a) ¿Cuántos alumnos se examinaron?
b) ¿Cuántos perdieron la materia, si para ganar se requiere obtener un 3?

183. En una distribución normal, que tiene desviación típica 10, la probabilidad de que el valor de la variable sea mayor de 65 es 0,19.
a) Calcular la media de la distribución
b) Obtener el valor de la variable que supere al 75% de los valores.

184. El gerente de producción de una fábrica de bombillas, estima que la vida útil del producto está distribuida normalmente con una media de 5.000 horas. Se estima que hay una probabilidad del 60% de que la bombilla dure menos de 5.568 horas.
a) ¿Cuál es la desviación típica?
b) Si en un día se producen 20.000 unidades. ¿Cuántas esperamos que duren más de 4.000 horas?

185. En una distribución normal que tiene una desviación típica de 10, la probabilidad de que el valor de una variable sea mayor de 65 es de 0,15,
a) Calcular la media de la distribución.
b) Obtener el valor de la variable que es superado por el 75% de los valores.

186. Las cajas de cartón que contienen un determinado producto pesan en promedio 300 kg cada una y tienen una varianza de 2.500 kg ¿Cuál es la probabilidad de que el peso de una caja sea superior a 328 kg?

EJERCICIOS MISCELÁNEOS

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

187. Una máquina vendedora de café puede regularse de modo que, en promedio, sirva μ onzas de café por taza. Si las onzas servidas por taza tiene una distribución normal, con desviación estándar 0,25 onzas por taza, encuentre el valor de μ que permitirá que una taza de 7 onzas se derrame el 0,5% de las veces.

188. Si los ingresos de los miembros de una comunidad se distribuyen normalmente con una media de $454.000 y una desviación estándar de $50.000, a) ¿Cuál es el ingreso mínimo que deberá tener

un miembro de esta comunidad para pertenecer al 10% superior? b) ¿Cuál es el máximo ingreso correspondiente al 50% central?

189. El gerente de producción de una fábrica de bombillas estima que la vida útil del producto está distribuida normalmente con una media de 5.000 horas. Si, además el gerente estima que hay una probabilidad del 60% de que la bombilla dure más de 5.568 horas y menos de 4.432,
a) ¿Cuál es la desviación estándar?
b) Si en un día se producen 20.000 unidades, ¿cuántas de ellas esperamos que tengan una vida inferior a las 4.500 horas?

190. La vida útil de cierto producto está distribuida normalmente. Si el 17,68% de ellas duran más de 540 horas y el 69,83% duran menos de las 486 horas,
a) ¿Cuál es la media y desviación estándar?
b) Si son reemplazadas aquellas unidades con una vida inferior de 450 horas, ¿cuántas se reemplazarán si se han vendido 20.000?

191. Explique brevemente qué se entiende por:
a) Distribución probabilística. b) Variable aleatoria discreta. c) Variable aleatoria continua.

192. Establezca los requisitos o condiciones requeridas en la:
a) Distribución Binomial. b) Distribución de Poisson. c) Distribución Hipergeométrica.

193. Explique brevemente en qué forma sirven de modelos las distribuciones probabilísticas.

194. Se quiere promocionar un producto especialmente para colocar en la lonchera de los niños que inician sus estudios. Se probó el producto en un grupo de 16 niños; 10 afirmaron que les agradaba y los restantes indicaron lo contrario. Suponga que se seleccionaron 4 personas en dos oportunidades, la primera con reposición y en la segunda sin reposición ¿cuál es la probabilidad:
a) de que, máximo, a dos les agrade. b) por lo menos a tres no les agrade?

195. En un intento por burlar la vigilancia en la aduana de un aeropuerto, un viajero colocó 6 tabletas de narcótico en un frasco que contiene 9 pastillas de vitamina de apariencia semejante. Si el agente antinarcóticos selecciona al azar tres tabletas, ¿cuál es la probabilidad de no arrestar al viajero por posesión ilegal de narcóticos?

196. El número de radios que en promedio vende una compañía por día es de 1,5. Calcular la probabilidad de que la compañía venda por lo menos cuatro radios durante un período,
a) de dos días. b) de tres días. c) de cuatro días.

197. En una reciente encuesta gubernamental se encontró que el 80% de las familias que viven en una comunidad sub-urbana, y cuyos ingresos son superiores a los $860.000, poseen dos hijos. Suponiendo que el estudio esté en lo cierto, si se selecciona una muestra de 10 familias de esta categoría, obtenga la probabilidad de que:
a) por lo menos el 20% de los que integran dicha muestra tengan dos hijos.
b) como máximo, dos tengan un número de hijos diferentes a dos.

198. En una planta industrial, los lotes grandes de artículos recibidos se inspeccionan para detectar los defectuosos por medio de un esquema de muestreo. Se examinan 10 artículos y el lote será

rechazado si se encuentran dos o más artículos defectuosos. Si un lote contiene exactamente 5% de defectuosos.
a) ¿Cuál es la probabilidad de que el lote sea aceptado?
b) ¿Cuál es la probabilidad de que el lote sea rechazado?

199. Supongamos que por cierta intersección de tránsito pasan, en promedio 30 autos por hora, ¿Cuál es la probabilidad de que:
a) en un instante de dos minutos no pase ningún auto por la intersección?
b) por lo menos dos autos transiten en un intervalo de seis minutos?

200. De un inventario de 48 automóviles que se embarcan a distribuidores locales, 12 tienen instalados radios defectuosos, ¿Cuál es la probabilidad de que cierto distribuidor reciba 8 automóviles,
a) todos con radios buenos.
b) por lo menos uno con radio defectuoso.
c) cuántos automóviles con radios defectuosos se esperan recibir?

201. Por informes suministrados por las compañías de seguros, se sabe que en la ciudad, de cada 50 conductores de vehículo particular, 40 tienen seguro de responsabilidad civil. Si en un fin de semana 8 autos de esta ciudad se ven involucrados en accidentes de tránsito, ¿cuál es la probabilidad de que:
a) más de dos tengan seguro.
b) por lo menos seis tengan seguro.
c) cómo máximo dos no tengan seguro?

202. El promedio mensual de robos en grandes almacenes de la ciudad, a pesar de la severa vigilancia es de 3,2, ¿Cuál es la probabilidad de que:
a) no haya robos en una quincena.
b) más de dos robos en un período de dos meses.
c) cómo máximo dos robos en un mes?

203. El 0,5% de las cartas que se envían por correo no llevan los timbres postales correctos. En 4.000,
a) ¿Cuántas cartas con timbre incorrectos esperaría encontrar?
b) ¿Cuál es la probabilidad de encontrar más de dos cartas con timbres incorrectos?

204. El 12% de las personas que reservan cupos en vuelos de avionetas, a menudo no llegan a tiempo para abordarlas. La avioneta tiene capacidad para ocho pasajeros. (1) Obtener la probabilidad de que las 8 personas aborden la avioneta. (2) Si hicieron reservaciones cinco personas, encuentre la probabilidad de que:
a) se quede una persona. b) no se quede ninguna.

205. Como resultado de la apertura económica, el 20% de las ventas de automóviles nuevos en Colombia son importados. Suponga que se seleccionan al azar 6 personas que han comprado automóvil nuevo durante la semana pasada. ¿Cuál es la probabilidad de que:
a) más de 5 personas hayan comprado automóvil importado.
b) por lo menos 2 personas hayan comprado vehículo de fabricación nacional?

206. Una de cada 100 lámparas incandescentes fabricadas por una compañía se funde antes del final de su período de una semana, si se dejan encendidas todo el tiempo. Se instala una lámpara en cada uno de los cincuenta pisos de un edificio.

¿Cuál es la probabilidad aproximada de que:
a) se funda una lámpara al final de la semana.
b) más de tres lámparas se fundan al final de la semana.
c) menos de 48 lámparas no se fundan al final de la semana?

207. El tiempo necesario para atender un automóvil en una estación de gasolina, tiene una distribución normal, con media de 4,5 minutos y desviación estándar de 72 segundos.
a) ¿Cuál es la probabilidad de que un automóvil seleccionado en forma aleatoria requiera, cuando máximo, 3,5 minutos para el servicio?
b) ¿Cuál debe ser el tiempo necesario para el servicio, si el 8% de todos los automóviles requieren una cantidad mayor de tiempo?

208. La puntuación media en una prueba de ingreso a la universidad es de 500 y la desviación estándar de 75. Las puntuaciones se distribuyen en forma normal.
a) ¿Indique la puntuación por debajo de la cuál quedó el 15% de los estudiantes?
b) ¿Indique la puntuación por encima de la cual obtuvo calificación el 32% de los estudiantes?
c) La universidad establece que los aspirantes con un puntaje entre 350 y 420 se pondrán en reserva o espera, si hay 3.000 aspirantes, ¿cuántos quedarán en la clasificación de "espera"?

209. Un almacén de electrodomésticos acaba de recibir una remesa de 20 aparatos de T.V. Poco después de recibirlos, el fabricante llamó para informar que por descuido se habían enviado 6 aparatos defectuosos. Se decidió probar dos de los aparatos.
a) ¿Cuál es la probabilidad de que ninguno de los dos esté defectuoso?
b) ¿Qué los dos estén buenos?

210. En una Universidad el Departamento de economía y matemáticas está conformado por 40 profesores, 12 de los cuales son de planta. El decano selecciona al azar 5 profesores del departamento para revisar el plan de estudios ¿Cuál es la probabilidad de que:
a) todos sean de planta.
b) todos no sean de planta.
c) por lo menos dos sean de planta?

211. Un colegio de la ciudad, cuenta con 18 buses para recoger y regresar los alumnos a su casa. Supongamos que 6 de ellos tienen problemas mecánicos. Se seleccionan 6 vehículos al azar para probarlos. ¿Cuál es la probabilidad de que:
a) ninguno tenga fallas mecánicas?
b) dos de ellos tengan fallas mecánicas?
c) por lo menos dos no tengan fallas?

212. El grupo de secretarias de una empresa está conformada por 30 personas, 12 de las cuales llevan trabajando más de cinco años. Si la dirección requiere para un trabajo especial cinco secretarias y son elegidas al azar, ¿cuál es la probabilidad de que:
a) ninguna de las secretarias tendrán más de 5 años de experiencia.
b) dos de las secretarias tendrán más de 5 años de experiencia.
c) por lo menos una no tendrá más de 5 años de experiencia?

213. Un determinado producto industrial se embarca en lotes de 30 unidades, con el propósito de minimizar el número de unidades defectuosas enviados a los diferentes distribuidores, se elaboró un

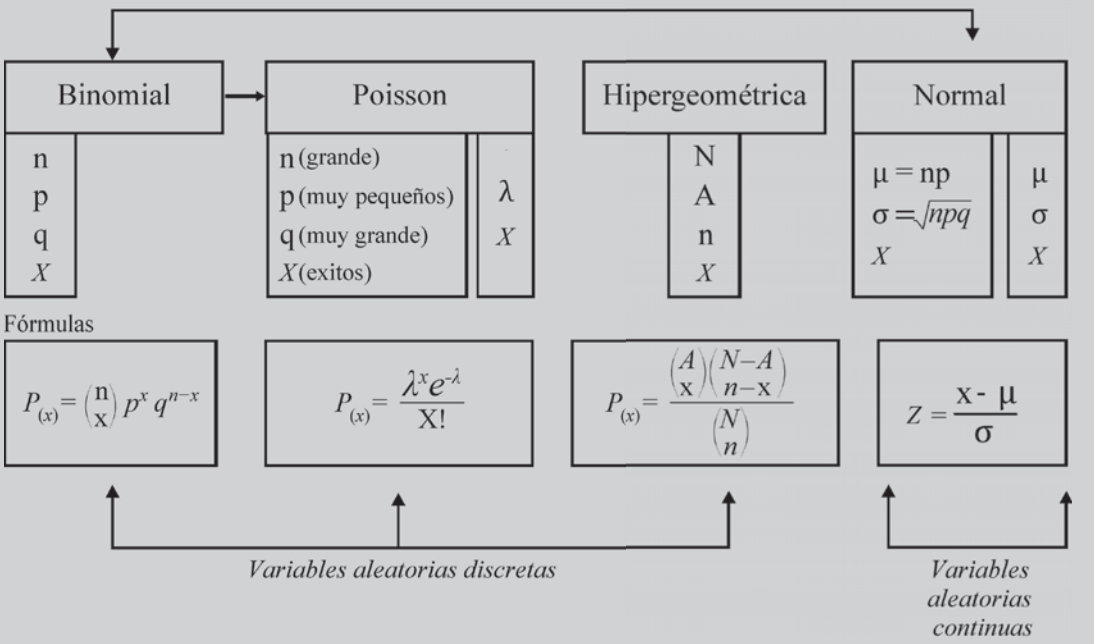
programa de inspección que consiste en tomar una muestra de 8 unidades de cada lote y rechazarlo si se observa más de cierto número de unidades defectuosas:

a) Si la decisión es de rechazar el lote si se observa más de un artículo defectuoso en la muestra (a sabiendas de que el fabricante estima que cada lote contiene 4 defectuosos) ¿Cuál es la probabilidad de que sea aceptado?
b) Si el fabricante estima que cada lote tiene 17 artículos buenos, ¿cuál es la probabilidad de aceptar el lote con menos de dos defectuosos, en una muestra de ocho artículos?

Síntesis de capítulo

1 Información que debe tenerse en cuenta en la solución de ejercicios sobre distribución probabilística, ya sea discreta o continua.

DISTRIBUCIONES DE PROBABILIDAD



2. La distribución Binomial, Poisson y la Normal se aplican en aquellos experimentos que sólo tienen dos resultados, éxito o fracaso.

3. La distribución normal presenta las siguientes características:
   a) Simétrica.
   b) Es asintótica.
   c) El área bajo la curva, es aproximadamente del 100%.
   d) La media, se localiza en el centro de la campana.
   e) Se puede aplicar en los ejercicios de distribución Binomial, siendo su resultado un valor aproximado.

4. La distribución *Hipergeométrica* se aplica al muestrear sin reposición. Por tanto la probabilidad de éxito no es constante. Su cálculo es tedioso cuando n es grande. En muchos casos sus resultados son casi iguales al de la *Binomial*, por lo tanto se procede a resolverlos por este último método. Las pruebas son dependientes.

5. La distribución de *Poisson* se aplica en ejercicios de *Binomial*, sólo cuando n es muy grande y la probabilidad de éxito es muy pequeña.

6. En la distribución *Binomial*, la probabilidad de éxito debe ser constante para cada ensayo sucesivo.

7. La *Hipergeométrica* se aplica cuando pueden ocurrir dos o más resultados.

8. Cada distribución *normal* queda totalmente especificada por su media y desviación estándar; por lo tanto se tendrá una distribución normal, diferente para cada media y su desviación estándar.

7

CAPÍTULO

Distribuciones muestrales

MUESTREO ALEATORIO

Aprender sin reflexionar es
malgastar la energía

o o

CONTENIDO

- Algunos conceptos generales y específicos sobre el muestreo.
- Estimación puntual y de intervalos. Propiedades de un buen estimador.
- Teorema sobre el límite central.
- Distribuciones muestrales: Medias, Proporciones, Diferencias entre dos medias y dos proporciones.
- Tamaño óptimo en muestras aleatorias simples.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el Sistema de Información en Línea SIL.

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- ❑ Manejar, comprender y utilizar términos, métodos y conceptos de muestreo.
- ❑ Entender, calcular y aplicar las distribuciones muestrales en casos prácticos.
- ❑ Explicar y utilizar el teorema del Límite central en el desarrollo de esta unidad.
- ❑ Comprender y explicar las condiciones requeridas por un buen estimador.
- ❑ Calcular el tamaño óptimo para una muestra.

ALGUNOS CONCEPTOS SOBRE MUESTREO

Hoy, la *estadística* está considerada como la teoría de la información, no sólo como función descriptiva, sino con el objeto básico de hacer estimaciones acerca de los valores estadísticos de la población o en la comprobación de hipótesis de aquellas características que han sido investigadas.

Se recomienda volver a leer el Capítulo 1, lo mismo que el Capítulo 13, donde se podrán ampliar estos conceptos de Muestreo, importantes para el desarrollo no sólo de este sino de los siguientes.

De los anteriores contenidos se observa que la estadística cubre dos aspectos de gran importancia: la *estadística descriptiva* a través de la *recolección, clasificación, presentación*, ya sea en forma de cuadros o gráficas, la *aplicación de medidas* como promedios, desviaciones, etc., y la *interpretación* y *análisis* de datos a fin de obtener conclusiones. Se realiza un *proceso deductivo* de lo general a lo particular.

El segundo aspecto es la *inferencia estadística* o *método inductivo*, el cual, mediante investigaciones por muestreo, logra obtener resultados considerados como estimadores de los valores estadísticos, correspondientes a las características de las unidades que conforman la población, correspondiendo al método inductivo, que denominaremos inferencia.

Se podría afirmar que la tarea más importante de la *estadística* es la realización de *inferencias* acerca de una población objetivo, con base en los resultados obtenidos a través de una muestra.

Hacemos una pequeña síntesis de los que consideramos que el alumno debe recordar y afianzar.

Población o universo se puede definir como un conjunto de elementos o de unidades en una población.

El *elemento* o *unidad* puede ser una persona, familia, empresa, zona, animal u objeto.

Del *elemento* se estudian sus *características*. Éstas se clasifican en: *cualitativas* o *atributos*, expresadas por palabras y se cuantifican mediante el conteo o recuento; las *cuantitativas o variables* expresadas en forma numérica, que pueden ser medidas (*variable continua*) o contadas (*variable discreta*).

De acuerdo con lo anterior, la *población* puede definirse como un conjunto de medidas, o el recuento de todas las unidades que presentan una característica común. Se podría definir como un conjunto de mediciones, finito o infinito, real o conceptual.

Marco muestral. Es un listado, actualizado y revisado, de todos los elementos que constituyen la población que va a ser objeto de investigación. También puede ser un mapa o croquis con las unidades de selección plenamente identificadas.

Encuesta preliminar, piloto o pretest. Antes de iniciar la investigación, se recomienda realizar una pequeña encuesta preliminar con el fin de probar el cuestionario, conocer mejor la población, entrenar al entrevistador, determinar el tiempo que requiere la entrevista y en especial tener un mayor conocimiento acerca de algunos parámetros.

La población se clasifica en: *finita o infinita*

Cuando se investigan las características de todas las unidades que constituyen la población o universo, nos referimos a una *investigación total, exhaustiva* o *censo*.

> Factores tales como: *costo, tiempo, recursos humanos, poblaciones muy grandes o infinitas, destrucción de la unidad sometida a control, características con gran homogeneidad,* impiden la realización del censo. Se sustituye, entonces, por una investigación *parcial o muestra*.

El objetivo principal de muestreo es considerar el mayor número de unidades con el menor costo posible.

La *muestra*, para que sea representativa de la población, requiere que todas las unidades de la población tengan la misma probabilidad de ser seleccionadas, es decir, debe ser *aleatoria, al azar o probabilística.*

El *muestreo aleatorio,* realizado bajo ciertas condiciones y sometido a cumplir ciertos requisitos, se constituye en un procedimiento práctico, económico y rápido para generalizar conclusiones obtenidas a través de una muestra, aplicables a toda la población de la que forma parte, dentro de ciertos límites de confiabilidad, establecidos de antemano.

Dentro del *Muestreo aleatorio,* se tienen los siguientes métodos:

- *Muestreo aleatorio simple* o muestreo aleatorio irrestricto, en el cual se da igual oportunidad de selección a cada elemento o unidad dentro de la población.
- *Muestreo aleatorio estratificado* (Asignación igual, proporcional y óptimo), garantiza la representatividad, reduciendo el error de la muestra al formar grupos o subpoblaciónes más o menos homogéneas, en cuanto a su composición interna y heterogénea cuando se comparan entre sí.
- *Muestreo por conglomerados*, *por áreas o geográfico*. Cuando la unidad básica de muestreo se encuentra en la población en grupos o conglomerados y la selección de la unidad permite la observación del total de elementos de cada conglomerado elegido. Cada conglomerado tiene las mismas características de la población; puede hacerse un segundo muestreo dentro del conglomerado seleccionado, denominándose de doble etapa o bietápico.

  Generalmente es muy aplicado cuando no se dispone de un marco de referencia completo. El área total se divide en pequeñas áreas que son muestreadas. Cada área seleccionada podrá ser subdividida y enumerada para una nueva selección, si es necesario, y así sucesivamente dando origen al muestreo por etapas o polietápicos.
- *Muestreo por fases*. En ocasiones, es conveniente y económico recoger cierta información de la totalidad de elementos de una muestra, la cual se extrae de la población en tal forma que sea lo suficientemente grande. Además, la necesidad de información más detallada obliga a una nueva muestra proveniente de la anterior, ocasionando un muestreo de dos fases o bifásico. Puede considerarse, también, de varias fases o polifásico.
- ❑ *Muestreo sistemático*. La selección de las unidades se hace a intervalos regulares, en un orden sistemático.

Las condiciones del muestreo aleatorio implican consideraciones importantes: (a) se debe seguir un *diseño* estadístico específico (muestreo aleatorio, simple, estratificado, etc.); el mejor es aquél que proporciona la precisión necesaria, en términos de un límite, en cuanto al *error de estimación* a un menor costo; (b) *la selección de los elementos* al azar, para luego recolectar la información por cualquiera de los métodos: entrevista, observación directa, correo, teléfono, etc.; (c) *el error muestral*, es decir, la diferencia que debe haber entre el resultado realizado mediante la muestra y el obtenido posiblemente mediante la investigación total o censo.

El error de estimación, es la diferencia que puede haber entre la estimación puntual y el parámetro. Cuando la estimación no representa bien al parámetro, a pesar de estar perfectamente diseñada, nos referiremos a *errores muestrales*; los errores no muestrales son ocasionados por el mal diseño del formulario, errores cometidos en el proceso de recolección, procesamiento y análisis de los datos.

Parámetro (poblacional), son las medidas descriptivas numéricas aplicadas a las características en las unidades de la población. También se les denomina como valores estadísticos de la población.

E*stimador puntual,* son las medidas descriptivas numéricas aplicadas a las características en las unidades de la muestra. Se podrá decir que el estimador es una norma o método para estimar una constante perteneciente a una población. La estimación hace referencia a los valores numéricos de los parámetros poblacionales desconocidos, a los cuales se llega mediante una muestra.

El *estimador por intervalos,* es una regla que nos indica cómo calcular dos puntos o valores a través de una muestra. La estimación por intervalos es la estimación del parámetro mediante la especificación de un intervalo de valores, determinado por un límite inferior y otro superior (límites de confianza) dentro del cual estará comprendido el valor verdadero o parámetro poblacional.

Se dice que un buen estimador debe ser:

- *Insesgado* es decir que no tenga sesgo, error o bias, cuando el valor del estimado es igual al del parámetro. En caso contrario la estimación será sesgada.
- *Consistente* es aquel estimador que, al aumentar el tamaño de la muestra, converge en probabilidad al parámetro que estima.
- *Eficiente* es el estimador que tiene la menor varianza entre todos los estimadores posibles.
- *Suficiente* cuando incluye toda la información que la muestra puede proporcionar acerca del parámetro.

El *intervalo de confianza*, corresponde a un intervalo de valores, dentro de los cuales se espera que esté el parámetro con cierto grado de confianza o con riesgo de error conocido; para ello es necesario determinar primero la estimación puntual.

La probabilidad de que un *intervalo de confianza* contenga el parámetro que se estima, se denomina *coeficiente de confianza.*

Se han visto una de serie de términos y definiciones con respecto al ***muestreo aleatorio.*** También existe el *muestreo no aleatorio, circunstancial* o *errático,* método cuyos resultados o estimaciones no son de ninguna manera confiables, dado que la selección de las unidades que conforman la muestra, se realiza en forma caprichosa o por conveniencia, primando en muchos casos el juicio personal del investigador.

Dentro del muestreo no aleatorio existen algunos métodos tales como:

- *Muestreo a juicio, intencional* u *opinático,* donde los elementos seleccionan a juicio o en opinión del investigador; se podría decir que prima la intención de que estas unidades sean incluidas dentro de la muestra.
- *Muestreo por conveniencia,* donde se eligen los elementos más al alcance del investigador.
- *Muestreo voluntario,* donde el informante, voluntariamente, suministra información sin ser previamente seleccionado.
- *Muestreo por cuotas,* es un número de entrevistas, encuestas, condiciones o cuotas que se le fijan al encuestador para que a su vez seleccione los elementos en la forma que considere oportuna.

Encuestas descriptivas y analíticas. Algunos autores clasifican las encuestas, en términos generales como: encuestas descriptivas y encuestas analíticas. En las primeras, el interés se especifica en la obtención de alguna información correspondiente a una población. En las otras, la finalidad es analizar ciertas hipótesis o supuestos acerca de la población, que el investigador se fijó de antemano. Hay encuestas que sirven para ambos propósitos.

DISTRIBUCIONES MUESTRALES

Corresponden a una distribución de todas las muestras que pueden ser escogidas conforme a un esquema de muestreo especificado, que implique selección a azar y, a una función de un número fijo de variables aleatorias independientes.

De una población a estudiar, se selecciona una sola muestra de todas las muestras posibles de igual tamaño, con el fin de obtener conclusiones sobre la población, no sobre la muestra.

La selección de las unidades que van a conformar la muestra debe hacerse al azar, mediante un generador de números aleatorios, usando cualquier método: sorteo, tablas de números aleatorios, sistemático, calculadora o EXCEL.

La aplicación de EXCEL en la selección de elementos, se dan tres (3) opciones o métodos que usted puede consultar en el capítulo 13.

Se presentan a continuación cuatro tipos de distribución:

DISTRIBUCIÓN DE MEDIAS MUESTRALES $\overline{x}$

Simbología que será utilizada en este capítulo:

MEDIDAS	POBLACIÓN	MUESTRA
Media aritmética	μ	$\overline{x}$
Varianza	σ^2	s^2
Desviación típica	σ	s
Tamaño	N	n

$\mu_{\overline{x}}$ = Media de todas las medias muestrales

$\sigma_{\overline{x}}$ = Desviación típica de todas las medias muestrales.

M = Número de muestras posibles.

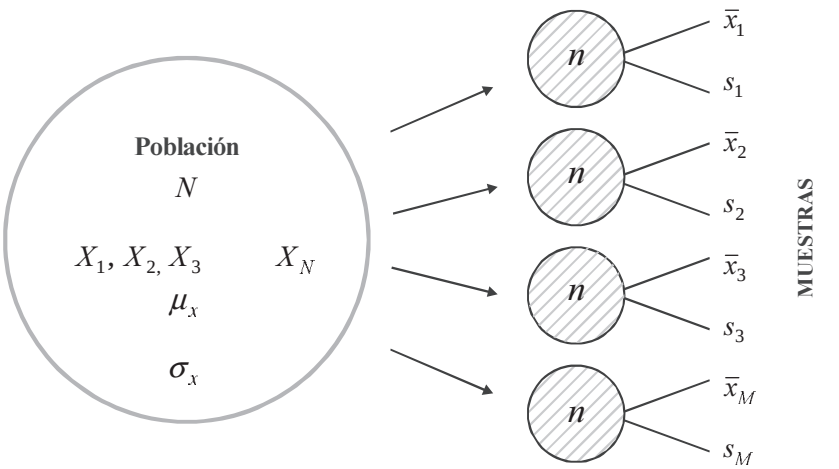
$$M = \binom{N}{n} = \frac{N}{(N-n)\ n}$$ Cuando la selección se hace sin reposición.

$M = N^n$ Cuando se hace la selección con reposición.

En el capítulo 13 se da orientación mucho más amplia sobre el uso de algunos de los métodos de selección de elementos o unidades, más utilizados.

Si consideramos una población de *N* elementos, con media μ y desviación típica σ, si se obtienen M número de muestras posibles, de tamaño n, simbolizamos a cada media muestral,

por: $\overline{x}_1$ $\overline{x}_2$ $\overline{x}_3$ $\overline{x}_M$ y cada desviación típica muestral, por: s_1 s_2 s_3 ..... s_M



Teorema. *Dada una población, si extraemos todas las muestras posibles de un mismo tamaño, entonces la media de la distribución de todas las medias muestrales posibles, será igual a la media poblacional.*

Simbolizaremos la *media de todas las medias muestrales* por $\mu_{\bar{x}}$, la cual será igual a la media poblacional

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{M} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 \quad + \bar{x}_M}{M} = \mu$$

$\mu = \mu_{\bar{x}}$ = Media de la distribución de muestreo.

La media de todas las medias muestrales debe ser exactamente igual a la media poblacional (μ), debido a que la distribución de muestreo resulta de todas las muestras posibles que se pueden extraer de una población; por tal razón incluye todos sus elementos. De ahí, que calcularla es un método indirecto para obtener el valor del parámetro μ.

La varianza de todas las medias muestrales se simboliza por $\sigma^2_{\bar{x}}$ y el error estándar de la media, será igual $\sigma_{\bar{x}}$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum \bar{x}_i - \mu^{\,2}}{M}} = \sqrt{\frac{(\bar{x}_1 - \mu)^2 + (\bar{x}_2 - \mu)^2 \quad + (\bar{x}_M - \mu)^2}{M}} = \frac{\sigma}{\sqrt{n}}$$

Siendo $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ (para muestras grandes o sea n > 30 el cual se denomina: *error estándar de la media*)

$\sigma_{\bar{x}}$ = Desviación estándar de la distribución de muestreo de medias, que depende de la dispersión de la población y del tamaño de la muestra.

Si el tamaño muestral (n) no es una fracción pequeña del tamaño poblacional, lo más adecuado es considerar el *error estándar de la media*, como $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

La expresión $\frac{N-n}{N-1}$ se conoce como *factor de corrección* para población finita.

Ejemplo 1. Supongamos una población de 5 elementos (N=5) y los valores que toma la variable, dependiendo de la unidad de medida utilizada, ya sean kilómetros, metros, valores, etc., son:

N = 5 ⇒ siendo: $X_1 = 7$ $X_2 = 3$ $X_3 = 5$ $X_4 = 8$ $X_5 = 2$

Con los anteriores valores se puede calcular la media, la varianza y desviación típica poblacional.

$$\mu = \frac{\sum X_i}{M} = \frac{25}{5} = 5 \qquad \sigma^2 = \frac{\sum X_i^2 - N\bar{X}^2}{N} = \frac{151 - 5(25)}{5} = 5{,}2 \qquad \sigma = \sqrt{5{,}2} = 2{,}2$$

Ahora determinaremos el número de muestras posibles (M) de esta población, si el tamaño de la muestra que fijamos arbitrariamente es 2 y si la selección se hace sin repetición o reposición. En el primer caso, se tendrá que:

$$M = C_n^N = \frac{N}{N-n \quad n} \qquad M = C_2^5 = \frac{5}{2 \quad 3} = 10$$

El número de las combinaciones que se pueden obtener considerando los 5 elementos, son:

X_1X_2 X_1X_4 X_2X_3 X_2X_5 X_3X_5
X_1X_3 X_1X_5 X_2X_4 X_3X_4 X_4X_5

Calculemos la *media aritmética* para cada una de las posibles muestras.

$$\bar{x}_1 = \frac{x_1 + x_2}{n} = \frac{7+3}{2} = 5 \qquad \bar{x}_6 = \frac{x_2 + x_4}{n} = \frac{3+}{2} = 5{,}5$$

$$\bar{x}_2 = \frac{x_1 + x_3}{n} = \frac{7+5}{2} = 6 \qquad \bar{x}_7 = \frac{x_2 + x_5}{n} = \frac{3+2}{2} = 2{,}5$$

$$\bar{x}_3 = \frac{x_1 + x_4}{n} = \frac{7+}{2} = 7{,}5 \qquad \bar{x} = \frac{x_3 + x_4}{n} = \frac{5+}{2} = 6{,}5$$

$$\bar{x}_4 = \frac{x_1 + x_5}{n} = \frac{7+2}{2} = 4{,}5 \qquad \bar{x} = \frac{x_3 + x_5}{n} = \frac{5+2}{2} = 3{,}5$$

$$\bar{x}_5 = \frac{x_2 + x_3}{n} = \frac{3+5}{2} = 4 \qquad \bar{x}_{10} = \frac{x_4 + x_5}{n} = \frac{+2}{2} = 5{,}0$$

La *media de todas las medias muestrales* será igual a:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{M} = \frac{5+6+....+5}{10} = \frac{25}{5} = 5$$

Se podrá observar que se cumple el teorema, que la *media* de todas las *medias muestrales* es igual a la *media poblacional*. $\mu_{\bar{x}} = \mu = 5$

En cuanto a la desviación típica de todas las *medias* muestrales, encontramos que:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu)^2}{M}} = \sqrt{\frac{(5-5)^2 + (6-5)^2 + \quad + (5-5)^2}{10}} = \sqrt{1,\ 5} = 1{,}3$$

Lo anterior debería ser igual a $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ denominado como *error estándar* de la media, el cual se cumple tan sólo cuando el tamaño de la muestra (n) es grande. Algunos autores consideran que n es grande, cuando es mayor a 30 ($n > 30$) y es pequeña, cuando n es menor o igual a 30 ($n \leq 30$). Es ésta la razón por la cual, al calcular el error estándar de la media, utilizando la desviación típica poblacional (σ), el resultado es diferente a 1,39. Veámoslo:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2{,}2}{\sqrt{2}} = \frac{2{,}2}{1{,}414} = 1{,}61 \neq 1{,}3$$

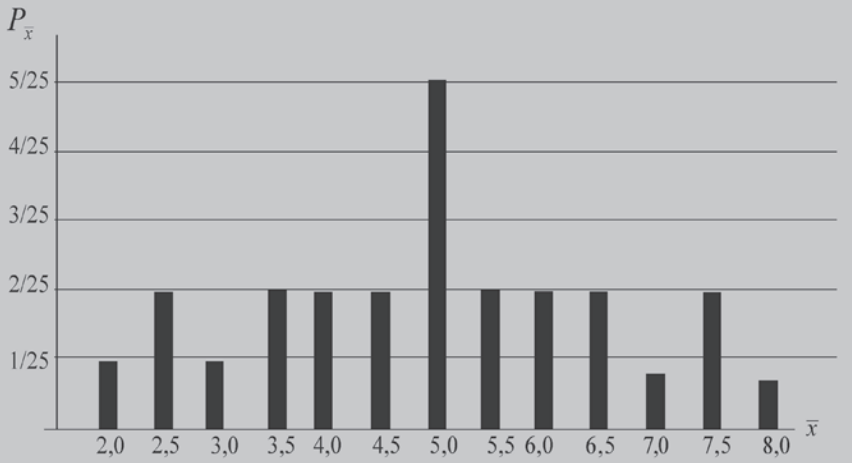
En el caso de hacerse la *selección con reposición*, se tendrá que $M = N^n = 5^2 = 25$ y las combinaciones posibles son:

X_1X_2	X_2X_1	X_2X_3	X_3X_2	X_3X_4	X_4X_3	X_4X_5
X_1X_3	X_3X_1	X_2X_4	X_4X_2	X_3X_5	X_5X_3	X_5X_4
X_1X_4	X_4X_1	X_2X_5	X_5X_2	X_3X_3	X_5X_5	
X_1X_5	X_5X_1	X_1X_1	X_2X_2	X_4X_4		

$M = 25$
$\mu = 5$
$\sigma_{\bar{x}} = 1{,}3$
$n = 2$

Los resultados serán los mismos, es decir, $\mu = 5$ y $\sigma_{\bar{x}} = 1{,}39$. Es de anotar que, cuando se hace la selección sin reposición, la población se vuelve infinita y que en cada extracción o selección, el elemento o unidad puede ser nuevamente seleccionado y así sucesivamente.

Figura 7.1



La representación gráfica de la distribución muestral de las 25 medias muestrales, tomando como ejemplo las $P_{\bar{x}}$ obtenidas, cuando la selección la hacemos con reposición

$$P_{(\bar{x}=5)} = \frac{5}{25} \qquad P_{(\bar{x}=4)} = \frac{2}{25} \qquad P_{(\bar{x}=7{,}0)} = \frac{1}{25}$$

$$P_{(\bar{x}=6)} = \frac{2}{25} \qquad P_{(\bar{x}=5{,}5)} = \frac{2}{25} \qquad P_{(\bar{x}=3)} = \frac{1}{25}$$

$$P_{(\bar{x}=7{,}5)} = \frac{2}{25} \qquad P_{(\bar{x}=2{,}5)} = \frac{2}{25} \qquad P_{(\bar{x}=8)} = \frac{1}{25}$$

$$P_{(\bar{x}=4{,}5)} = \frac{2}{25} \qquad P_{(\bar{x}=6{,}5)} = \frac{2}{25} \qquad P_{(\bar{x}=2)} = \frac{1}{25}$$

$$P_{(\bar{x}=3{,}5)} = \frac{2}{25}$$

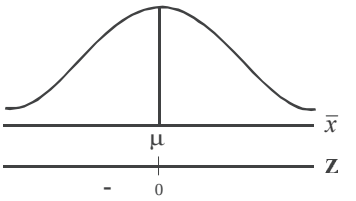
Teoría del límite central. Se cumple, cuando independientemente de la población origen, la distribución de las medias aleatorias se aproximan a una distribución normal a medida que el tamaño de la muestra crece. Se podrá decir también, que si las muestras provienen de una población que no es normal, es de importancia tener en cuenta el tamaño de la muestra, si el tamaño muestral es pequeño, la distribución obtenida con sus medias muestrales tendrán un comportamiento similar al de la población de donde se extrajeron. Por el contrario, si el tamaño muestral es grande, el comportamiento de estas medias muestrales será igual al de una distribución normal, independientemente de la población de donde fueron extraídas.

En su forma más simple, el teorema indica que si n variables aleatorias independientes tienen varianzas finitas, su suma, cuando se le expresa en medida estándar, tienden a estar normalmente distribuidas cuando n tiende a infinito. Se debe observar que ninguna de las varianzas sea mayor comparada con el total.

De acuerdo con el teorema anterior, la variante estadística *Z* para distribuciones de medias muestrales será:

$$Z=\frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x}-\mu}{\sigma\ \sqrt{n}}$$

Por la cual consideramos que se aproxima a un distribución normal.



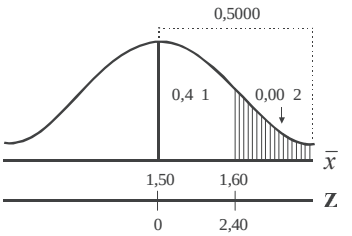
Ejemplo 1. La altura media de 400 alumnos de un plantel de secundaria es de 1,50 mts, y su desviación típica es de 0,25 mts. Determinar la probabilidad de que en una muestra de 36 alumnos, la media sea superior a 1,60 mts.

Solución: $P_{(\bar{x}>1,60)} =$

$$Z=\frac{1,60-1,50}{\frac{0,25}{\sqrt{36}}}=\frac{0,10}{\frac{0,25}{6}}=\frac{0,60}{0,25}=2,40$$

$Z=2,40 \rightarrow A\ 0,4\ 1$

$P=0,5000-0,4\ 1\ =0,00\ 2=0,\ 2$



$P_{(\bar{x}>1,60)}=0,\ 2$

Ejemplo 2. Se tiene para la venta un lote de 1.000 pollos, con un peso promedio de 3,50 kg y una desviación estándar de 0,18 kg, ¿Cuál es la probabilidad de que en una muestra aleatoria, 100 pollos de esta población, pesen entre 3,53 y 3,56 kg?

Solución: $\mu=3,5 \qquad \sigma=0,1 \qquad n=100 \qquad P_{(3,53<\bar{x}<3,56)} =$

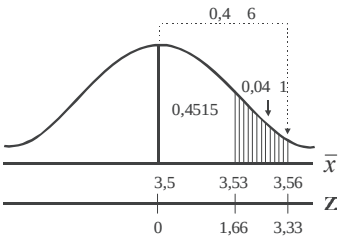
$$Z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{3,56-3,5}{\frac{0,1}{\sqrt{100}}}=\frac{0,06\ 10}{0,1}=3,33$$

$Z=3,33 \rightarrow A\ 0,4\ \ 6$

$$Z=\frac{3,53-3,5}{\frac{0,1}{\sqrt{100}}}=\frac{0,03\ 10}{0,1}=1,66$$

$Z=1,66 \rightarrow A\ 0,4515$

$P=0,4\ \ 6-0,4515=0,04\ 1$



$P_{3,53<\bar{x}<3,56}=4,\ 1$

Ejemplo 3. Un fabricante de cierto champú para el cabello, distribuye el tamaño *profesional* de su producto en 100 salones de belleza de Caracas. Se ha determinado que el consumo promedio de su producto es de 2.800 cojines mensuales, con desviación estándar de 280 cojines. Si se toma una muestra probabilística de 36 salones, ¿cuál es la probabilidad de que el consumo promedio en un mes sea inferior a 2.700?

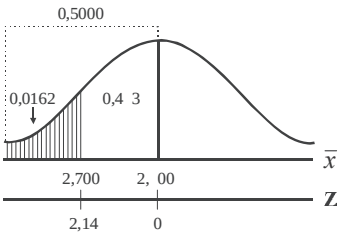
Solución: $\mu=2\ \ 00 \qquad \sigma=2\ \ 0 \qquad n=36 \qquad P_{(\bar{x}<2\,700)} =$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2\,700 - 2\ \ 00}{\frac{2\ 0}{36}}$$

$$Z = \frac{-100\ \ 6}{2\ \ 0} = -2{,}14$$

$$Z = -2{,}14 \rightarrow A\ \ 0{,}4\ \ 3$$

$$P = 0{,}5000 - 0{,}4\ \ 3\ \ = 0{,}0162$$



$$P_{\bar{x}<2\,700} = 1{,}62$$

En aquellos casos de *poblaciones finitas*, es decir, cuando se da información sobre el tamaño poblacional, y cuando el tamaño de la muestra es mayor del 5% de la población, se puede aplicar el factor de corrección, representado de diferentes maneras; cualquiera de las siguientes formas podrán ser aplicadas:

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{N-n}{N}} = \sqrt{\frac{N}{N} - \frac{n}{N}} = \sqrt{1 - \frac{n}{N}} = \sqrt{1-f}$$

Factor de corrección para poblaciones finitas.

La *f* se conoce como fracción de muestreo. $f = \frac{n}{N}$

Por lo tanto la desviación estándar de la muestra se convierte en $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

En distribuciones de medias muestrales, la estandarización de Z, incluyendo el factor de corrección para poblaciones finitas será:

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N-1}}}$$

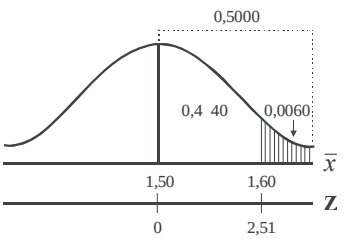
Ejemplo 4. Si en el ejemplo (1) se considera que dicho plantel de secundaria tiene un total de N = 400 alumnos, ¿Cuál es la probabilidad, en una muestra de 36 alumnos, de que la media sea superior a 1,60m.?

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)\sqrt{\frac{N-n}{N-1}}}$$

$$Z = \frac{1{,}60 - 1{,}50}{\left(\frac{0{,}25}{\sqrt{36}}\right)\sqrt{\frac{400-36}{400-1}}} = 2{,}51$$

$$Z = 2{,}51 \rightarrow A\ \ 0{,}4\ \ 40$$

$$P = 0{,}5000 - 0{,}4\ \ 40 = 0{,}006$$



$$P_{\bar{x}>1{,}60} = 0{,}6$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

1. En una población normal, con media 72,1 y desviación estándar 3,1, encuentre la probabilidad de que en una muestra de 90 observaciones, la media sea menor que 71,7.

2. En un banco de ahorros, la cuenta media es de $659.320, con una desviación de $18.000. ¿Cuál es la probabilidad de que un grupo de 400 cuentas, elegidas al azar, tenga un depósito medio de $660.000 o más?

3. En cierta región los salarios diarios de los mineros del carbón están distribuidos normalmente con una media de $864.500 y una desviación estándar de $15.000. ¿Cuál es la probabilidad de que una muestra representativa de 25 mineros, tenga un promedio diario inferior a $857.500?

4. Las estaturas de cierto grupo de adultos tienen una media de 167,42 y una desviación estándar de 2,58 centímetros. Si las estaturas están normalmente distribuidas y se eligen aleatoriamente 25 personas del grupo, ¿cuál es la probabilidad de que su media sea de 168,00 centímetros o más?

5. Si se extrae una muestra aleatoria de 36 elementos de una población, ¿cuántos elementos debe contener otra muestra de la misma población, para que el error estándar de la media de la segunda muestra, sea 2/3 del error estándar de la media de la primera muestra?

6. Si se extraen dos muestras aleatorias de una misma población, y si el error estándar de la media de una de ellas, es k veces el error estándar de la media de otra, ¿cuál es la relación entre los tamaños de ambas muestras?

7. Los salarios diarios en cierta industria están distribuidos normalmente con un media de $23.000. Si el 9% de las medias de los salarios diarios en muestras de 25 obreros, es inferior a $22.500, ¿cuál es la desviación estándar de los salarios diarios en esta industria?

8. Supongamos que se tienen, en una urna, 500 fichas numeradas 1, 2, 3,.... 499, 500. Después de mezclarlas completamente, se sacan16 fichas aleatoriamente. ¿Cuál es la probabilidad de que la suma sea mayor de 3.000?

9. Si los pesos individuales de las personas que viajan en avión se distribuyen normalmente con media de 68 kilos y desviación típica de 3,5 kilos, ¿cuál es la probabilidad de que un Boeing 707 con 81 pasajeros pesen más de 5.700 kilos?

10. Las estaturas de los estudiantes de una universidad, se distribuyen normalmente con media de 170 centímetros y desviación típica de 18 centímetros. Si se toma una muestra de 81 estudiantes, ¿cuál es la probabilidad de que tengan una estatura superior a 175 centímetros?

11. Quinientos cojinetes de bolas tienen un peso medio de 5,02 onzas y una desviación de 0,30 onzas. Hallar la probabilidad de que una muestra al azar de 100 cojinetes, elegidos entre este grupo, tengan un peso de más de 5,10 onzas.

12. Una siderúrgica está produciendo actualmente cables para suspensión de puentes. La característica más importante de este producto es su resistencia, el peso que puede soportar antes de que se

reviente. Por experiencias pasadas se sabe que el promedio de la resistencia es de 6 toneladas con desviación típica de ¾ de tonelada. Para efectos de control, se selecciona una muestra de 9 cables y se adopta la siguiente regla de decisión:

Si la resistencia promedio está por encima de 6,5 toneladas o por debajo de 5,5, se suspende el proceso. Si está entre 5,5 y 6,5 se deja tal y como está.

a ¿Cuál es la probabilidad de detener el proceso, si la media de la producción es aún de 6 toneladas?
b) ¿Cuál es la probabilidad de detener el proceso, si la media de la producción no es 6 toneladas sino 6,18 toneladas?
c) ¿Cuál es la probabilidad de continuar el proceso, si el promedio es en realidad de 6,4 toneladas?
d) ¿Si es de 5,8 toneladas?

13. Suponga que una máquina produce tornillos cuyos diámetros se distribuyen normalmente con media μ = ½ pulgada y una desviación típica σ = 0,01 pulgadas.
¿Cuál es la probabilidad de que el diámetro medio esté comprendido entre 0,49 y 0,51, para una muestra de 4 tornillos?

14. Una compañía productora de maíz híbrido afirma que sus productos darán, por término medio, 120 bultos por hectárea. Veinticinco hectáreas producen, en promedio, 115 bultos. Si se supone que la desviación típica σ, es de 10 bultos por hectárea, ¿cuál es la probabilidad de obtener una media muestral de 115 o menos?

15. En una distribución normal se seleccionan todas las posibles muestras de tamaño 10; si el 2% de estas muestras tienen medias que difieren de la media poblacional en más de 4 en valor absoluto, encontrar la desviación estándar de la población.

16. Ciertos tubos fabricados por una compañía tienen una duración media de 900 horas y una desviación típica de 70 horas. Hallar la probabilidad, al seleccionar al azar 36 tubos, que tengan una duración media entre 870 y 925 horas.

17. Se sabe que en cierta gran ciudad, los clientes de los restaurantes gastan en promedio $32.900 en comida, con una desviación estándar de $1.500. Si se pide a cada uno de 50 restaurantes que seleccionen al azar las cuentas de 100 personas y que informen sobre el consumo medio de esas 100 personas, ¿de cuántos restaurantes debe esperarse que informen sobre cuentas promedio, superiores $33.259,30?

18. Los pesos de los paquetes recibidos en una bodega tienen una media de 580 libras y una desviación típica de 80 libras, ¿cuál es la probabilidad de que el peso de 49 paquetes recibidos al azar y cargados en un montacargas, supere su capacidad de 30.000 libras?

19. En una universidad el promedio de calificación, en exámenes de admisión, ha sido de 3,5 con una desviación típica de 1. ¿Cuál es la probabilidad, si el examen lo presentan 36 estudiantes, de que obtengan un promedio mayor de 3,7?

20. El valor promedio de los pedidos que hacen los detallistas de una ciudad a cierto mayorista es de $25.900 diarios, con desviación estándar de $1.800. Si elegimos al azar una muestra de 200 pedidos, ¿cuál es la probabilidad de que la media de valor del los pedidos sea superior a $26.100?

21. Una firma de ingenieros ha estimado que el peso promedio de los adultos que vivirán en un edificio de apartamentos es de 68 kilos, con desviación estándar de 15 kg. De acuerdo con la anterior estimación, instalan en el edificio un ascensor para 36 personas con capacidad de 2.700 kilos; si la estimación es correcta, ¿cuál es la probabilidad de que un cupo completo exceda la capacidad del ascensor?

22. Un distribuidor mayorista recibe mensualmente 70.000 pilas de 1,5 voltios. Para decidir si acepta o rechaza las pilas, utiliza la siguiente regla de decisión: mide la vida útil de 36 pilas. Si la media de la muestra es 60 o más horas, se acepta la totalidad; en caso contrario, se rechaza. ¿Cuál es la probabilidad de:
    a) aceptar una remesa que tiene una vida útil de 59 horas y una desviación estándar de 3 horas?
    b) rechazar un cargamento que tiene una vida útil de 60,5 horas y una desviación estándar de 3 horas?

23. El gerente de una cooperativa de ahorro y vivienda, estima que el promedio de ahorro por cliente en un mes es de $520.000 con una desviación típica de $222.960. ¿Cuál es la probabilidad de que al examinar 36 cuentas, el promedio de ahorro se mayor de $630.000?

24. Un profesor sabe por experiencia que el examen final realizado a sus alumnos, proporciona una calificación promedio de 68 y una varianza de 441. El curso actual lo conforman 36 alumnos. ¿Cuál es la probabilidad de que el rendimiento medio sea menor de 60?

25. El gerente de una empresa asegura que su programa de entrenamiento en ventas permite aumentarlas más del 10% sobre $400.000 pesos diarios actuales, con una desviación estándar de $78.608. Si selecciona un grupo de 25 vendedores, ¿cuál es la probabilidad de que el gerente tenga la razón?

26. Un almacén ofrece cargos de empacadores de mercancía con un promedio de 58 cajas y desviación estándar de 16. En cada caja se pueden empacar 5 docenas de camisas, colocándole a cada una su marca distintiva. Si se toma una muestra aleatoria de 16 aspirantes a los cargos y se les somete a un día de prueba, ¿cuál es la probabilidad de que empaquen, en promedio, entre 50 y 70 cajas en el día?

27. A vendedores de libros, una distribuidora les asegura que además de su sueldo básico de $660.000 mensuales, obtendrán un promedio mensual de $240.000 por comisión. Si la distribuidora emplea a 25 vendedores, ¿cuál es la probabilidad de que el ingreso promedio por comisión sea inferior a $237.000? Se admite que la comisión por venta de libros tiene una desviación típica de $8.200.

28. En un supermercado se establece que los paquetes de café de libra, tienen un promedio 1,03 libras, con una desviación típica de 0,05 libras. ¿Cuál es la probabilidad de seleccionar 28 paquetes (marcados de libra) con un peso promedio superior a 1,02 libras?

29. Un auditor toma una muestra de tamaño 49, de una población de 800 cuentas por cobrar. La desviación estándar de la población es de $93.800 y la media es $226.000. ¿Cuál es la probabilidad de que la media de la muestra, sea menor igual a $206.000?

30. Un analista financiero selecciona una muestra aleatoria del 8% de 500 cuentas. La media y la desviación típica, de los saldos encontrados en las 500 cuentas son: $417.500 y $17.000 respectivamente. ¿Cuál es la probabilidad de que con dicha muestra se obtenga una media superior a $420.000?

31. El jefe de un departamento de ventas sabe que en el almacén principal, el promedio de compra por cliente es de $112.000 con una desviación estándar de $5.500. Si se toma una muestra de tamaño 36, ¿cuál es la probabilidad:

a) que la media de la muestra sea superior a $113.500.
b) que sea superior a $113.200 e inferior a $111.500?

32. Una fábrica produce correas cuyo diámetro promedio es de 16 mm y desviación típica de 3,5 mm. Si se selecciona una muestra de 36 correas, ¿cuál es la probabilidad de que la media muestral sea menor o igual a 15,3 mm?

33. Se sabe por experiencia que el rendimiento promedio por hectárea, en un cultivo, es de 70 bultos y la desviación típica de 20 bultos. Si se selecciona una muestra de 36 hectáreas, ¿cuál es la probabilidad de que el rendimiento medio sea superior a 75 bultos?

34. Las cajas de cartón que contienen un determinado artículo producido por un establecimiento industrial tienen un peso medio de 300 kg , y una varianza d 2.500 kg^2. ¿Cuál es la probabilidad de que 25 paquetes, tomados al azar y cargados en un camión de distribución del producto, excedan la capacidad del camión, de 8.200 kg?

35. Si la desviación típica de las estaturas de niños de kinder es de 5 centímetros, ¿cuál es la probabilidad de que la estatura promedio de una muestra al azar de 100 niños, difieran en más de un centímetro?

36. Si la desviación típica de las calificaciones de estadística en el quinto semestre es de 8 décimas, ¿cuál es la probabilidad, en una muestra de 20 estudiantes, de que la diferencia entre la media muestral y la población sea mayor en 4 décimas?

37. Un fabricante de bombillas dice que su producto tiene una media de duración de 700 horas, con varianza de 14.400 horas. El dueño de un taller compró 144 bombillas de esta marca con la idea de comprar más, si la duración media de la muestra le resultara superior a 680 horas; ¿qué probabilidad hay de que el dueño del taller no vuelva a hacer compras de esa marca?

38. Un distribuidor de botas especiales para el trabajo, garantiza al gerente de una empresa que el promedio de duración de las botas es de 8,10 meses, con una desviación típica de dos meses y cinco días. Si se decide la empresa a comprar 36 pares de botas, ¿cuál es la probabilidad, de que en promedio la duración sea inferior a los 7 meses y quince días?

o o a a t o t a t o a a

DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN p

En el análisis de una característica cualitativa o atributo, se emplea la proporción de éxitos y no el número de éxitos como en la distribución binomial.

Anteriormente se definió la proporción de éxitos como: $$p = \frac{\text{Número de casos favorables o éxitos}}{\text{Total de casos posibles}}$$

Ahora en vez de expresar la variable en términos de éxitos (X) nos referiremos, al número de elementos con el atributo en la muestra (a) y lo dividimos por el tamaño de la muestra (n).

$$p = \frac{\sum a_i}{n} = _____$$

Simbología

$A=\sum A_i$ Total de elementos que presentan la característica investigada en la población $A=\sum A_i=NP$

$\mu_{\bar{p}} = P = \overline{P}$ $P = \frac{A}{N} = \frac{\sum A_i}{N}$ Proporción de elementos que presenta la característica investigada en la población.

$Q = \frac{N-A}{N} = 1-P$ Proporción de elementos que no presenta la característica estudiada $P+Q=1$

σ_p^2 = Varianza de la proporción en la población $\sigma_p^2 = PQ$

$\sigma_{\bar{p}} =$ Desviación estándar $\sigma_p = \sqrt{PQ}$

$\sigma_{\bar{p}} = \frac{\sigma_{\bar{p}}}{\sqrt{n}} = \sqrt{\frac{PQ}{n}}$ Error estándar de la proporción

Demostración $\sigma_p^2 = PQ$

Variable: supongamos que se tiene una población de 500 personas para analizar su peso en kilogramos

$\mu_x = \frac{\sum X_i}{N} = \frac{70+50+4\ +60}{500}$; por lo tanto $\sum X_i = 70+50+48+60+.....$

$\sigma_x^2 = \frac{\sum X_i^2 - N\mu^2}{N}$ y se tendrá que: $\sum X_i^2 = 70^2 + 50^2 + 48^2 + 60^2 + ...$

Ya conocemos cómo se procede en el cálculo de la μ y de σ_x^2 en la variable X; ahora, observemos cómo se debe calcular en el atributo.

Atributo: Consideremos que se desea investigar la proporción de mujeres en 500 personas de una población; para ello se cuenta el número de mujeres observadas.

$\sum A_i = 1+1+1+1+1 \quad = 200$ equivale a $\rightarrow \sum X_i$

$P = \mu_{\bar{p}} = \frac{\sum A_i}{N} = \frac{200}{500} = 0{,}40$ (*Media proporcional*) equivale a μ_x

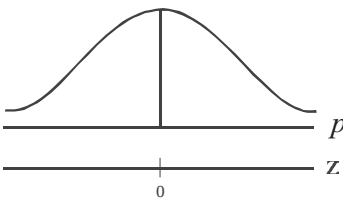
Ahora si se requiere $\sum A_i^2 = 1^2+1^2+1^2+1^2 = \quad = \sum A_i = NP_i$ reemplazando se tiene que :

$\sigma_{\bar{p}}^2 = \frac{\sum A_i^2 - NP^2}{N} = \frac{NP - NP^2}{N} = \frac{NP\ 1-P}{N} = P\ 1-P\ = PQ$

$\sigma_p^2 = PQ$ lo cual queda demostrado. $\sigma_p = \sqrt{PQ}$

Variante estadística. En muchos casos, podemos utilizar la distribución normal para evaluar la distribución muestral de proporciones, siendo:

$$Z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{p - \mu_p}{\sigma_{\bar{p}}}$$



Vale la pena observar la simbología que se utiliza para la muestra

$p = \frac{a}{n} = \frac{\sum a_i}{n}$	= Proporción muestral $\Leftrightarrow \overline{p}$
$P = \frac{A}{N}$	= Proporción poblacional $\Leftrightarrow \mu_{\overline{p}} = \overline{P}$
$\sigma_p^2 = PQ$	= Varianza de una proporción en la población
$\sigma_p = \sqrt{PQ}$	= Desviación proporcional en la población
$s_p^2 = pq$	= Varianza de una proporción muestral
$s_p = \sqrt{pq}$	= Desviación típica proporcional en la muestra
$\sigma_{\overline{p}} = \sqrt{\frac{PQ}{n}}$	= Error estándar en una proporción

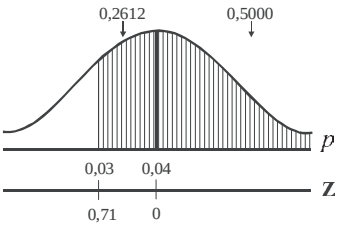
Ejemplo 1. Se tiene que el 4% de las piezas producidas por cierta máquina es defectuosa, ¿Cuál es la probabilidad de que en un grupo de 200 piezas, el 3% más sea defectuosa?

$$\sigma_{\overline{p}} = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{(0{,}04)(0,\ 6)}{200}} = 0{,}014$$

a) Se desea determinar la $P_{(p \geq 0{,}03)} =$

$$Z = \frac{p - \mu_p}{\sqrt{\frac{PQ}{n}}} = \frac{0{,}03 - 0{,}04}{\sqrt{\frac{(0{,}04)(0,\ 6)}{200}}} = -0{,}71$$

$$Z = -0{,}71 \rightarrow A(0{,}2612)$$
$$P = 0{,}2612 + 0{,}5000 = 0{,}7612$$



$P_{(p \geq 0{,}03)} = 76{,}12$

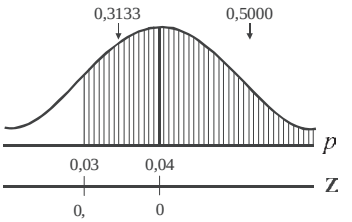
b) Con corrección. Si se quiere obtener una buena aproximación a la distribución normal, debe hacerse la corrección en la variable discreta, siendo igual $\frac{1}{2n}$. Si se va obtener una área hacia la derecha, se restará este factor de corrección; en el caso de que sea a la izquierda, se sumará ese factor al valor de *p*.

$$\frac{1}{2n} = \frac{1}{2\ 200} = \frac{1}{400} = 0{,}0025$$

$$Z = \frac{\left(p - \frac{1}{2n}\right) - \mu_p}{\sigma_{\overline{p}}}$$

$$Z = \frac{0{,}03 - 0{,}0025\ - 0{,}04}{0{,}014} = -0,$$

$$Z = -0, \quad \rightarrow A(0{,}3133)$$
$$P = 0{,}3133 + 0{,}5000 = 0,\ 133$$



$P_{(p \geq 0{,}03)} =$ 1,33

c) Resolveremos el anterior problema mediante la distribución normal

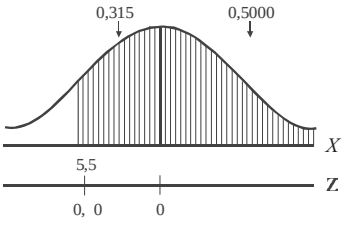
$\mu = np = 200\ 0{,}04\ =$

$\sigma = \sqrt{npq} = \sqrt{200\ 0{,}04\ \ 0,\ 6} = 2{,}77$

$P_{(x \geq 6)} =$ *siendo* $P_{x>5{,}5} =$

$Z = \dfrac{5{,}5 -}{2{,}77} = -0,\ 0 \rightarrow A\ 0{,}315$

$P = 0{,}315\ + 0{,}5000 = 0,\ 15$



$P_{x>5{,}5}\ =\ 1{,}5$

Resultado bastante aproximado al obtenido en el punto (b)

d) Podríamos plantear y aún resolver el mismo problema aplicando la distribución binomial.

$n = 200 \quad p = 0{,}04 \quad q = 0,\ 6 \quad P_{(x \geq 6)} =$

$P = 1 - \left[C_0^{200}\ 0{,}04 \quad 0,\ 6\ ^{200} \quad + C_0^{200}\ 0{,}04\ ^{5}\ 0,\ 6\ ^{1\ 5} \right]$

Ejemplo 2. Se desea estudiar una muestra de 49 personas para saber la proporción de las mayores de 40 años; sabiendo que la proporción en la población es 0,4. ¿Cuál es la probabilidad de que la proporción en la muestra sea menor de 0,5?

o **:**

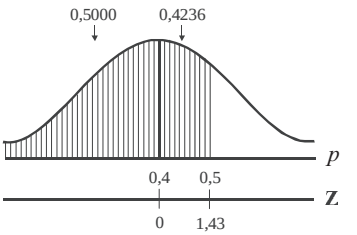
$n = 4 \qquad P = 0{,}4 \qquad P_{(\bar{p}<0{,}5)} =$

$Z = \dfrac{\bar{p} - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0{,}5 - 0{,}4}{\sqrt{\dfrac{(0{,}4)(0{,}6)}{4}}} =$

$Z = \dfrac{0{,}10}{\sqrt{\dfrac{0{,}24}{4}}} = \dfrac{0{,}10}{\sqrt{0{,}004}} = \dfrac{0{,}10}{0{,}07} = 1{,}43$

$Z = 1{,}43 \rightarrow A\ 0{,}4236$

$P = 0{,}5000 + 0{,}4236 = 0,\ 236$



$P_{p<0{,}5}\ =\ 2{,}36$

Ejemplo 3. Cuarenta y seis por ciento de los sindicatos del país están en contra de comerciar con la China Continental; ¿cuál es la probabilidad de que una encuesta a 100 sindicatos muestre que más del 52% tenga la misma posición?

o **:**

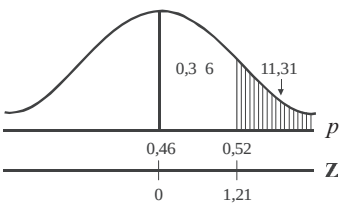
$P = 0{,}46 \qquad p = 0{,}52 \qquad n = 100 \qquad P_{p>0{,}52}\ =$

$Z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} = \dfrac{0{,}52 - 0{,}46}{\sqrt{\dfrac{(0{,}46)(0{,}54)}{100}}}$

$Z = \dfrac{0{,}06}{\sqrt{\dfrac{0{,}24\ \ 4}{100}}} = 1{,}21$

$Z = 1{,}21 \rightarrow A\ 0{,}3\ \ 6$

$P = 0{,}5000 - 0{,}3\ \ 6\ \ = 0{,}1131$



$P_{p>0{,}52}\ = 11{,}31$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

39. Se ha determinado que el 65% de los estudiantes universitarios de Medellín prefieren los cuadernos marca Profesional. ¿Cuál es la probabilidad de que en un muestra de 100 universitarios de dicha ciudad, encontremos que:
    a) como máximo el 68% sean usuarios de ese tipo de cuaderno.
    b) exactamente 66% sean usuarios (utilizar medio punto de porcentaje para los límites)?

40. Un fabricante de desodorantes recibe cada semana lotes de 10.000 válvulas para los tarros rociadores. Para aceptar o rechazar dichos lotes, selecciona al azar 400 válvulas de cada lote; si el 2% o más resultan defectuosos, se rechaza el lote. En caso contrario se acepta el lote. ¿Cuál es la probabilidad de rechazar un lote que contenga el 1% de válvulas defectuosas?

41. Se ha encontrado que el 4% de las piezas producidas por cierta máquina son defectuosas. ¿Cuál es la probabilidad, al seleccionar 400 piezas, que el 5% o más sean defectuosas?

42. Para elegir presidente de un sindicato, un candidato obtuvo el 46% de los votos. Determinar la probabilidad de que entre 200, elegidos al azar, de un total de 1.000 afiliados, se obtenga la mayoría de votos para dicho candidato.

43. En cierta facultad de una universidad de Caracas 1/6 de los alumnos son mujeres. Si se extrae una muestra aleatoria de 200 estudiantes de la facultad, ¿cuál es la probabilidad de que el 20% o más sean mujeres?

44. Hallar la probabilidad de que en 200 lanzamientos de una moneda, el número de caras esté comprendido entre el 40% y 60%.

    (En el SIL se ha dejado planteado como distribución binomial y se ha resuelto mediante la distribución normal y de proporciones).

45. Se sabe que el 25% de los estudiantes de un colegio usan anteojos. ¿Cuál es la probabilidad de que 8 o menos usen anteojos en una muestra de 36 estudiantes?

46. Un nuevo tratamiento con rayo láser asegura su eficacia en el 90% de los casos. Si se selecciona una muestra de 40 enfermos, ¿qué probabilidad hay de que se presente una diferencia mayor del 8% en cuanto a su eficacia?

47. Según datos anteriores, se sabe que la efectividad de una vacuna es del 90%. ¿Cuál es la probabilidad de que al vacunar a 64 personas la proporción sea mayor del 95%?

48. Se ha demostrado, por reclamos que han hecho, que el 20% de las encomiendas llegan averiadas al utilizar una compañía de transporte intermunicipal. ¿Cuál es la probabilidad, al enviar 100 encomiendas, de que la proporción sea menor de 25%?

49. Por datos obtenidos con anterioridad, se sabe que el 70% de las familias que tienen teléfono no se encuentran en las horas de la tarde del día domingo. Se toma una muestra aleatoria de 36 familias del directorio telefónico y se les llama. ¿Cuál es la probabilidad de que el 50% o más estén ausentes?

50. Se sabe que el 7% de los niños que nacen en cierta región, mueren antes de alcanzar un mes de vida. Si durante un cierto período de tiempo nacieron 40 niños, ¿cuál es la probabilidad de que 6 o más de ellos mueran antes de alcanzar el primer mes de vida?

51. Según registros del Departamento de Circulación y Tránsito, el 25% de los heridos en accidentes de tránsito quedan con alguna incapacidad de por vida. En un mes cualquiera se registran 150 personas que sufrieron lesiones. ¿Cuál es la probabilidad de que 42 o más víctimas queden con alguna incapacidad?

52. 1/3 de los alumnos matriculados en las facultades de publicidad son hombres. Se extrae una muestra aleatoria de 150 alumnos matriculados en dichas facultades, ¿cuál es la probabilidad de que 40 o menos sean de sexo masculino?

53. Se toma una muestra aleatoria de 200 unidades producidas en una hora por un máquina. Se sabe que el 10% de las unidades producidas son defectuosas. ¿Cuál es la probabilidad de que 16 o menos resulten defectuosas?

54. Se sabe que el 70% de los empleados públicos usan corbata; ¿cuál es la probabilidad de que, al seleccionar 64 empleados, menos del 56% no usen la corbata?

55. Si se realiza una investigación entre 36 personas, sobre la preferencia en el uso de desodorante en barra y en atomizador, ¿cuál es la probabilidad de que el 82% más de las personas entrevistadas prefieran el desodorante en barra? Se sabe por experiencia que dicha proporción es del 74% y 26%, respectivamente.

56. En cierto proceso de producción se utiliza la siguiente regla de decisión: se elige una muestra al azar de 36 piezas; si el porcentaje de piezas defectuosas de la muestra excede de p, se detiene el proceso para localizar las fallas. Si se sabe que el proceso ocasiona un 10% de piezas defectuosas, en promedio, determine el valor de p, para que exista un 22,5% de probabilidad de detener el proceso, cuando la proporción de piezas defectuosas exceda de p.

57. Se sabe que el 65% de los estudiantes de primaria, en una concentración escolar, son nacidos en esa misma ciudad. ¿Cuál es la probabilidad de que en una muestra de 100 estudiantes de esa concentración, cuando menos el 68 % sean oriundos de esa ciudad?

58. Un fabricante de insecticidas domésticos recibe cada semana lotes de 10.000 válvulas para los tarros rociadores. Para aceptar o rechazar dichos lotes, ha establecido el siguiente procedimiento de muestreo: selecciona al azar 400 válvulas de cada lote; si el 20% o más resultan defectuosas, rechaza el lote. En caso contrario lo acepta. ¿Cuál es la probabilidad de rechazar un lote que contenga el 15% de válvulas defectuosas?

59. Si se toma una muestra aleatoria de 80 artículos producidos por una máquina, sabiendo que el 15% resulta defectuoso, ¿cuál es la probabilidad de que el 20% o más de los artículos observados en la muestra sean defectuosos?

60. En una elección departamental o provincial, el 55% de los electores están a favor del candidato del partido A, ¿cuál es la probabilidad de que, en una muestra de 100 electores, el resultado no muestre una mayoría a favor del candidato?

61. Se sabe que el 40% de las familias, en uno de los barrios al norte de la ciudad, tiene aparatos receptores de televisión a color. Si se toma una muestra aleatoria de 50 familias, ¿cuál es la probabilidad de que haya 25% o más familias con dicha clase de receptores?

62. Se sabe que el 70% de la población económica activa del país tiene ingreso mensual menor de $1.900.000. Si se toma una muestra de 1.000 personas de tal población, encontrar la probabilidad de que entre 680 y 730 tengan ingresos menores de $1.900.000.

63. El gerente de un supermercado considera que de un total de 50 clientes que realizan compras a mediodía, 7 incluyen leche en su compra; efectúa una muestra de 100 clientes. ¿Cuál es la probabilidad de que menos de 12 clientes compre leche a mediodía?

64. En cierto proceso de producción se utiliza el siguiente sistema de control de calidad: se elige una muestra de 36 unidades; si el porcentaje de unidades de la muestra no excede el valor de *p*, se continúa el proceso. Si se sabe que el proceso ocasiona un 10% de unidades defectuosas, en promedio, determine el valor de *p*, para que exista un 6 2% de probabilidad de continuar el proceso, cuando la proporción de piezas defectuosas es inferior a *p*.

65. El jefe de bodega de una cadena de almacenes, recibe semanalmente 15.000 unidades de un determinado artículo, que debe ser examinado para su aceptación. El tiempo disponible para esta revisión es pequeño dado el volumen de artículos, por lo cual se consideró necesario la selección al azar de 300 artículos, con la recomendación de que si 15 o más de ellos no están en buen estado, se devuelve la mercancía. ¿Cuál es la probabilidad de devolver las 15.000 unidades, si sabemos que el 3% de los artículos se consideran en mal estado?

66. Se toma una muestra aleatoria de 200 unidades producidas en una hora por un máquina. Se sabe que el 10 % de las unidades producidas son defectuosas. ¿Cuál es la probabilidad de que en la muestra tomada, 16 o más artículos, resulten defectuosos?

67. Un laboratorio lanza una nueva droga al mercado. Se asegura que en el 80% de los casos la droga es eficaz. Si en una clínica se seleccionan 49 pacientes que padecen de la misma enfermedad, ¿qué probabilidad hay de que se presente una diferencia superior al 10%, a lo asegurado por el laboratorio?

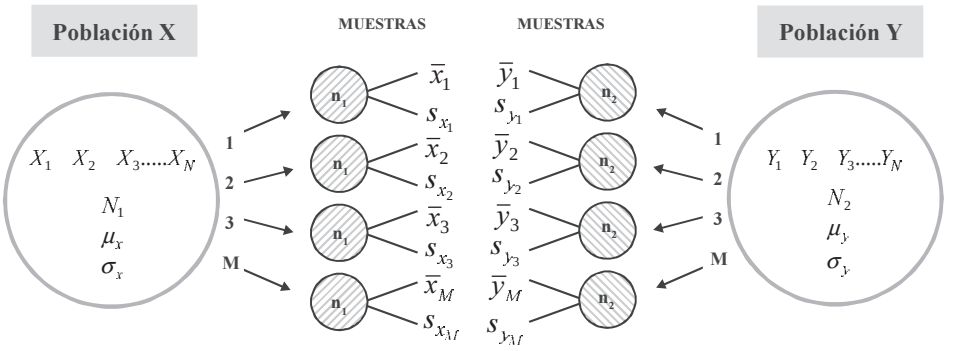
o o a a t o t a t o a a

DISTRIBUCIÓN DE DIFERENCIAS ENTRE DOS MEDIAS MUESTRALES $\bar{x} - \bar{y}$

Se tiene dos poblaciones normales e independientes, identificadas la primer por X y la segunda por Y, de tamaños N_1 y N_2, cuyas medias se simbolizan por μ_x y μ_y y sus desviaciones típicas por σ_x y σ_y. Se obtiene un número (M) de pares de muestras posibles. Las medias muestrales de la primera población se identifican por:

$\bar{x}_1 \quad \bar{x}_2 \quad \bar{x}_3 \ldots\ldots \bar{x}_M$ y las muestras de la segunda variable por $\bar{y}_1 \quad \bar{y}_2 \quad \bar{y}_3 \quad \bar{y}_M$

Las desviaciones típicas muestrales respectivas serán: $s_{x_1} \quad s_{x_2} \quad s_{x_3} \quad s_{x_M}$ y $s_{y_1} \quad s_{y_2} \quad s_{y_3} \quad s_{y_M}$,



Ahora, si consideramos las diferencias para cada par, la media aritmética de dichas diferencias se simbolizará por $\mu_{\bar{x}-\bar{y}}$ donde:

$$\mu_{\bar{x}-\bar{y}} = \frac{\sum(\bar{x}_i - \bar{y}_i)}{M} = \frac{(\bar{x}_1 - \bar{y}_1) + (\bar{x}_2 - \bar{y}_2) + \quad + (\bar{x}_M - \bar{y}_M)}{M}$$

$$\mu_{\bar{x}-\bar{y}} = \frac{\sum \bar{x}_i}{M} - \frac{\sum \bar{y}_i}{M}$$

Se puede demostrar que la media de las diferencias de todos los pares de medias muestrales posibles, es igual a la diferencia entre las medias poblacionales.

$$\mu_{\bar{x}-\bar{y}} = \mu_{\bar{x}} - \mu_{\bar{y}} \qquad \mu_{\bar{x}-\bar{y}} = \mu_x - \mu_y$$

La desviación típica de las diferencias entre los pares de medias muestrales se simboliza por: $\sigma_{\bar{x}-\bar{y}}$ y se calcula mediante la fórmula:

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sum\left[(\bar{x}_i - \bar{y}_i) - \ \mu_x - \mu_y\ \right]^2}{M}}$$

La desviación típica de las diferencias entre los pares de medias muestrales, denominada también como error estándar de las diferencias entre las medias muestrales, es igual a:

siendo:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n_1} \qquad \sigma_{\bar{y}}^2 = \frac{\sigma_y^2}{n_2}$$

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2} \qquad \sigma_{\bar{x}-\bar{y}} = \sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}$$

Suponiendo que la distribución de diferencias entre las medias muestrales tenga un comportamiento similar a la distribución normal, la variante estadística estará dada por la fórmula:

$$Z = \frac{(\bar{x} - \bar{y}) - \mu_{\bar{x}-\bar{y}}}{\sigma_{\bar{x}-\bar{y}}} \qquad Z = \frac{(\bar{x}_1 - \bar{x}_2) - \ \mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

También se pueden presentar por una de las siguientes formas:

$$Z=\frac{(\bar{x}-\bar{y})-\ \mu_x-\mu_y}{\sqrt{\sigma_{\bar{x}}^2+\sigma_{\bar{y}}^2}} \qquad Z=\frac{(\bar{x}-\bar{y})-\ \mu_x-\mu_y}{\sqrt{\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2}}}$$

Utilizamos la distribución normal como una aproximación, para resolver problemas de distribuciones de diferencias entre dos medias muestrales.

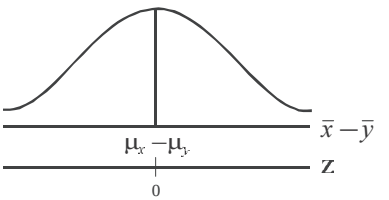
μ_x	μ_y
σ_x	σ_y
n_1	n_2
$\bar{x}$	$\bar{y}$

$$Z=\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2}}}$$

μ_1	μ_2
σ_1	σ_2
n_1	n_2
$\bar{x}_1$	$\bar{x}_2$

Se puede aplicar esta distribución cuando no se conocen las varianzas poblacionales las cuales pueden ser sustituidas por varianzas muestrales siempre que sean mayores de 30. Hay autores que consideran su utilización si $n_1 + n_2 > 30$. Siendo su fórmula:

$$Z=\frac{(\bar{x}-\bar{y})-\ \mu_x-\mu_y}{\sqrt{\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}}}$$



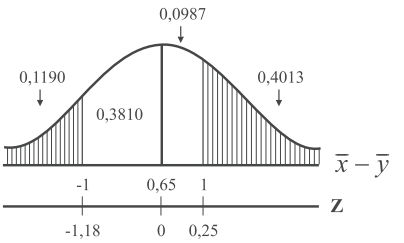
Ejemplo 1. Se tienen dos poblaciones normales e independientes, donde la media de la segunda población es 0,65 menor que la de la primera; si se seleccionan muestras de tamaño 100 y 120 y si las respectivas desviaciones típicas poblacionales son 12 y 8, se pide determinar la probabilidad de que, en un par de muestras, la diferencia entre ambas medias muestrales sea superior a 1 en valor absoluto.

o **:** $\mu_x-\mu_y=0{,}65 \quad n_1=100 \quad n_2=120 \quad \sigma_x=12 \quad \sigma_y=8$

$\mu_1-\mu_2=0{,}65 \quad n_1=100 \quad n_2=120 \quad \sigma_1=12 \quad \sigma_2=8$

$P_{(\bar{x}-\bar{y}>|1|)}=?$

$$\sigma_{\bar{x}-\bar{y}}=\sqrt{\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2}}=\sqrt{\frac{144}{100}+\frac{64}{120}}=1{,}40$$



$$Z=\frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sigma_{\bar{x}-\bar{y}}}$$

$$Z=\frac{1-0{,}65}{1{,}40}=0{,}25 \qquad Z=\frac{-1-0{,}65}{1{,}40}=-1{,}18$$

$Z=0{,}25 \rightarrow A(0{,}0987)$ $\qquad$ $Z=-1{,}18 \rightarrow A(0{,}3810)$

$P=1-[0{,}0987+0{,}3810]$

$P_{(\bar{x}-\bar{y}>|1|)}=1-(0{,}4797)=0{,}5203=52{,}03\%$

$P_{(\bar{x}-\bar{y}\geq|1|)}=52{,}03\%$

Ejemplo 2. Se obtiene una muestra aleatoria de 100 elementos de una población normal, que tiene media 50 y desviación estándar 8. Luego se selecciona otra muestra aleatoria de 400 elementos de una población normal, que tiene media 40 desviación estándar 12. Encontrar la probabilidad de que:

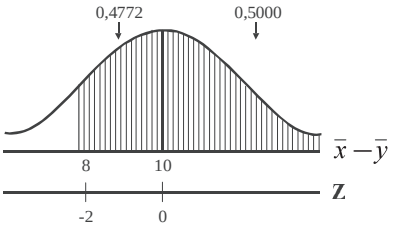
a) la media de la primera muestra exceda a la de la segunda en 8 o más.
b) ambas medias difieran, en valor absoluto, en 12 o más.

o **:** $n_1 = 100 \quad n_2 = 400 \quad \mu_x = 50 \quad \mu_y = 40 \quad \sigma_x = 8 \quad \sigma_y = 12$

$n_1 = 100 \quad n_2 = 400 \quad \mu_1 = 50 \quad \mu_2 = 40 \quad \sigma_1 = 8 \quad \sigma_2 = 12$

$P_{(\bar{x}-\bar{y}\geq 8)} = ?$

$$Z = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{\sigma_X^2}{n_1}+\frac{\sigma_Y^2}{n_2}}} = \frac{8-(50-40)}{\sqrt{\frac{64}{100}+\frac{144}{400}}} = -2$$



$Z = -2 \rightarrow A(0{,}4772)\ ó\ A(0{,}4773)$

$P_{(\bar{x}-\bar{y}\geq 8)} = 0{,}5000 + 0{,}4772$

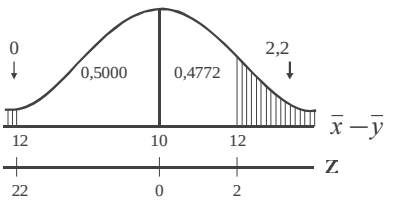
$P_{(\bar{x}-\bar{y}\geq 8)} = 0{,}9772 = 97{,}72\%$

b) $P_{\bar{x}-\bar{y}\geq|12|}$

$$Z = \frac{12-10}{1} = 2$$

$Z = 2 \rightarrow A\ 0{,}4772$

$P_{(\bar{x}-\bar{y}\geq|12|)} = 1 - 0{,}4772 + 0{,}5000 = 0{,}0228$



$$Z = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

$\sigma_{\bar{x}-\bar{y}} = 1$

$$Z = \frac{-12-10}{1} = -22$$

$Z = -22 \rightarrow A\ 0{,}5000$

$P_{(\bar{x}-\bar{y}\geq -12)} = 0{,}5000 - 0{,}5000 = 0$

$P_{(\bar{x}-\bar{y}\geq|12|)} = 0{,}022 \quad 0 = 0{,}022 \quad = 2{,}2$

$P_{(\bar{x}-\bar{y}\geq|12|)} = 2{,}2$

Ejemplo 3. En un restaurante, el consumo medio por desayuno es de $4.980, con una desviación estándar de $950. En un segundo restaurante las correspondientes cifras son $4.238 y $820. Si se eligen al azar 80 boletas de pago del primer restaurante y una muestra aleatoria de 60 del segundo, ¿cuál es la probabilidad de que la diferencia entre los consumos medios de ambas muestras sea mayor que $1.000 en valor absoluto?

$\mu_x = 4 \quad 0 \quad \mu_y = 4\,23 \quad \sigma_x = \ 50 \quad \sigma_y = \ 20 \quad n_1 = \ 0 \quad n_2 = 60$

$\mu_1 = 4 \quad 0 \quad \mu_2 = 4\,23 \quad \sigma_1 = \ 50 \quad \sigma_2 = \ 20 \quad n_1 = \ 0 \quad n_2 = 60$

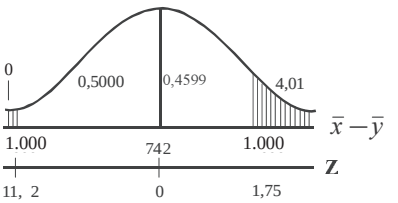
Solución: $P_{(\bar{x}-\bar{y}\geq|1\,000|)} =$ $\mu_x - \mu_y = 4 \quad 0 - 4\,23 \;= 742$

$$Z = \frac{1\,000 - 742}{\sqrt{\frac{02\,500}{0} + \frac{627\,400}{60}}}$$

$$Z = \frac{25}{\sqrt{21\,737,\ 2}} = 1{,}75$$

$$Z = \frac{-1\,000 - 742}{\sqrt{11\,2\ \ 1{,}25 + 10\,546{,}67}}$$

$$Z = \frac{-1\,742}{\sqrt{21\,737,\ 2}} = -11,\ 2$$



$P_{(\bar{x}-\bar{y}\geq|1\,000|)} = 4{,}01$

$Z = 1{,}75 \rightarrow A\ 0{,}45 \qquad Z = -11,\ 2 \rightarrow A\ 0{,}5000$

$\text{A} = 0{,}5000 + 0{,}45 \;= 0,\ 5$

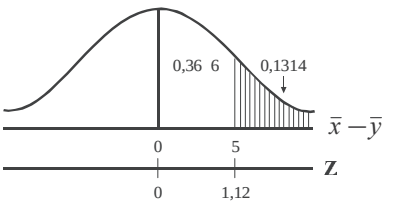
$P = 1 - 0,\ 5 \;= 0{,}0401 = 4{,}01$

Ejemplo 4. Dos marcas, A y B de tabletas antiácidas efervescentes registran el mismo promedio de disolución en agua, con desviación estándar de 12 segundos para la marca A y 24 segundos para B. Suponiendo que el tiempo de disolución esté normalmente distribuido, ¿cuál es la probabilidad de que, con una muestra de 36 tabletas de cada marca, las tabletas B registren un promedio de tiempo de disolución, cuando menos 5 segundos más rápido que A?

o **:** $\mu_x = \mu_y \quad \sigma_x = 12 \quad \sigma_y = 24 \quad n_1 = 36 \quad n_2 = 36 \quad P_{(\bar{x}-\bar{y}>5)} =$

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}}$$

$$Z = \frac{5 - 0}{\sqrt{\frac{144}{36} + \frac{576}{36}}} = \frac{5}{\sqrt{4 + 16}} = 1{,}12$$



$P_{(\bar{x}-\bar{y}>5)} = 0{,}1314 = 13{,}14$

$Z = 1{,}12 \rightarrow A\ 0{,}36\ 6$

$P = 0{,}5000 - 0{,}36\ 6 = 0{,}1314$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

68. De cada una de dos poblaciones normales e independientes con iguales medias desviaciones estándar de 6,40 y 7,20, se extraen muestras de 64 elementos. Encontrar la probabilidad de que la diferencia entre las medias de las muestras exceda de 0,60 en valor absoluto.

69. El rendimiento medio de los autos de la marca A es de 20 kilómetros por galón de gasolina, con una desviación estándar de 6 k.p.g. Las cifras comparables para los autos B son 25 y 5,5 k.p.g. Se

supone que el rendimiento de cada una de ambas marcas está normalmente distribuido. ¿Cuál es la probabilidad de que en un concurso, el rendimiento medio para 10 autos de la marca A sea mayor que el de 9 autos de la marca B?

70. En promedio, los estudiantes de la Universidad A se levantan 50 minutos después de la salida del sol, con una desviación estándar de 15 minutos. Los estudiantes de la Universidad B se levantan 60 minutos después de la salida del sol, con una desviación estándar de 18 minutos. Un grupo de 25 estudiantes de la Universidad A realiza un viaje junto con 20 alumnos de la B. Encontrar la probabilidad de que la hora media de levantada del grupo de la Universidad B sea más temprana que la del grupo de la Universidad A.

71. A y B fabrican dos tipos de cables, con una resistencia media a la rotura de 4.000 y 4.300 libras, con desviaciones típicas de 980 y 850 libras, respectivamente. Si se prueban 70 cables de A y 40 cables de B, ¿cuál es la probabilidad de que la media de resistencia a la rotura de A sea, al menos, 300 libras más que B?

72. El sueldo mensual de los licenciados en educación en Bogotá es de $920.000 como promedio, con una desviación estándar de $31.500. En la misma ciudad, el salario mensual de los economistas en la administración pública es de $925.000, con una desviación estándar de $52.500. Se toma una muestra aleatoria de 100 de cada población. ¿Cuál es la probabilidad de que la diferencia en las asignaciones B sea superior a las de A en $12.510?

73. Dos fábricas A y B, productoras de bombillas afirman que el promedio de duración de ellas es de 1.500 y 1.450 horas, respectivamente, con desviaciones típicas de 90 y 100 horas. Si se compran 100 bombillas de cada fábrica, ¿cuál es le probabilidad de que la diferencia en duración entre las dos marcas sea mayor de 40 horas?

74. Dos marcas de bombillas de alumbrado público, A y B tienen una duración promedio de 1.400 y 1.200 horas, respectivamente, y sus varianzas de 40.000 y 10.000 horas. Se extrae una muestra aleatoria de 125 por cada marca. Determine la probabilidad de que:
    a) La marca A tenga una vida media de por lo menos 160 horas más que B.
    b) La marca A tenga una vida media de por lo menos 250 horas más que B.

75. El tiempo promedio requerido para ejecutar un trabajo de ensamblaje es de 2 horas, con una desviación típica de 40 minutos y, el tiempo requerido para ejecutar otro trabajo o etapa en el ensamblaje es de una hora con cuarenta minutos, con una desviación típica de 32 minutos. Suponiendo que se distribuyen normalmente, ¿qué porcentaje de veces será mayor el promedio del primer trabajo con relación al segundo, si se toman muestras de tamaño 28 y 30 respectivamente?

76. Dada la siguiente información acerca de las vidas útiles de dos marcas de pilas de 9 voltios Marca A (horas): media = 51; Desviación típica = 8; Marca B (horas): media = 50; desviación típica = 6. Si se selecciona una muestra aleatoria de 100 pilas de cada marca:
    a) ¿cuál es la probabilidad de que la vida útil media de la marca A sea superior a de la marca B en 0,6 horas o más.
    b) Sea inferior a la de la marca B en 0,6 horas o más.

77. Por experiencia se sabe que el tiempo de trabajo promedio en un artículo con los tornos existentes es de 38,6 minutos (desviación típica de 13,8). El tiempo de trabajo promedio con los nuevos tornos

es de 35,5 minutos, (desviación típica 14,1). Si se toman dos muestras de 18 tornos cada una. ¿Cuál es la probabilidad, al producir 900 artículos, que el promedio de diferencia de B con respecto a A, sea superior en dos minutos?

78. El tiempo promedio requerido para ejecutar un determinado trabajo es de 2 horas, mientras que para otro trabajo, otro grupo de trabajadores, gasta una hora cuarenta y cinco minutos. Se sabe que la desviación estándar para cada uno de estos trabajos es de 30 y 20 minutos respectivamente. ¿Cuál es la probabilidad, si se toman dos muestras de 30 trabajadores cada una, de que el promedio requerido para ejecutar el segundo trabajo sea mayor que el primero?

79. El promedio de los obreros, en una fábrica textil, es de 34 años con desviación estándar de 6 años, además, el promedio de edad de las obreras es de 30 años con desviación estándar de 4 años. ¿Cuál es la probabilidad, al tomar muestras de 20 obreros de cada sexo, de que el promedio de edad de las mujeres resulte mayor que el de los hombres?

80. Cierto artefacto eléctrico tiene una vida promedio de 2.600 horas, con desviación típica de 200 horas; mientras que el mismo artefacto de fabricación extranjera tiene un promedio de duración de 2.400 horas, con desviación estándar de 180 horas. Si se toman dos muestras de 125 y 100 artefactos, respectivamente, ¿cuál es la probabilidad de que la diferencia registrada entre las dos muestras sea superior en 150 horas?

81. Se sabe que los pesos de ciertas bolsitas aromáticas tienen una media de 25 gramos, con desviación típica de 5 gramos. Si se toman dos muestras de 100 bolsitas cada una, ¿cuál es la posibilidad de que los resultados de estas muestras difieran en más de 2 gramos?

o o a a t o t a t o a a

DISTRIBUCIÓN DE DIFERENCIAS ENTRE DOS PROPORCIONES MUESTRALES $P_1 - P_2$

En el caso de dos poblaciones independientes, de tamaño N_1 y N_2, distribuidas binomialmente, con parámetros, medias proporcionales P_1 y P_2 (también se le pueden representar las medias por μ_{P1} y μ_{P2} y desviaciones proporcionales σ_{P1} y σ_{P2} siendo: $\sigma_{p_1} = \sqrt{P_1 Q_1}$ y $\sigma_{p_2} = \sqrt{P_2 Q_2}$. El error estándar de las diferencias entre las dos medias proporcionales estará dada por σ_{P1-P2} siendo:

$$\sigma_{p_1 - p_2} = \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}$$ Cuando se trabaja con parámetros o valores poblacionales

Las *proporciones* P_1 y P_2 son poblacionales, y se simbolizan con letras mayúsculas. Luego cuando son muestrales, se utilizan letras minúsculas p_1 y p_2.

Cuando n_1 y n_2 corresponden a muestras grandes, es decir, ambas superiores a 30, se tendrá, que el error estándar de las diferencias entre dos proporciones es:

$$s_{p_1 - p_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

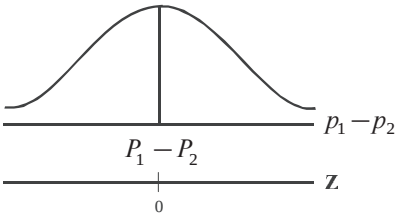
La *media de las diferencias* entre dos medias proporcionales, se simboliza; indistintamente por:

$$\mu_{P_1} - \mu_{P_2} = \mu_{P_1} - \mu_{P_2} = P_1 - P_2$$

La variante estadística Z, estará dada en la misma forma que fue presentada para diferencias entre dos medias muestrales:

$$Z = \frac{(p_1 - p_2) - (\mu_{p_1} - \mu_{p_2})}{\sqrt{\frac{P_1\ Q_1}{n_1} + \frac{P_2\ Q_2}{n_2}}}$$

En este Capítulo, en todos los ejercicios de proporciones se trabajará con P_1 y P_2 (mayúsculas), en vez de p_1 y p_2 aunque las muestras sean pequeñas, menores de 30, cuando se calcula el error estandar proporcional.



Ejemplo 1. Consideremos dos máquinas que producen un determinado artículo; la primera produce por término medio un 14% de artículos defectuosos, en tanto que otra, produce el 20% de artículos defectuosos; si se obtienen muestras de 200 unidades en la primera y 100 unidades en la segunda, ¿cuál es la probabilidad de que difiera A de B en 8% o más?

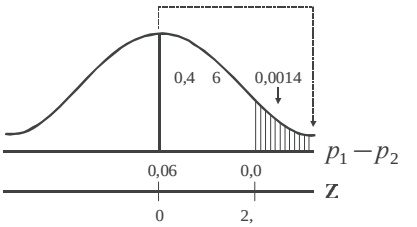
o : $P_{P_1 - P_2 \geq 0,0}$ = $n_1 = 200$ $n_2 = 100$ $P_1 = 14$ $P_2 = 20$

$$Z = \frac{(p_1 - p_2) - (\mu_{p_1} - \mu_{p_2})}{\sqrt{\frac{P_1\ Q_1}{n_1} + \frac{P_2\ Q_2}{n_2}}}$$

$$Z = \frac{0,0\ \ -\ -0,06}{\sqrt{\frac{0,14\ 0,\ 6}{200} + \frac{0,2\ 0,}{100}}}$$

$$Z = \frac{0,14}{0,047} = 2,$$

$$Z = 2, \quad \rightarrow A\ 0,4\quad 6$$



$$P_{P_1 - P_2 \geq 0,0} = 0,5000 - 0,4\quad 6$$

$$P_{P_1 - P_2 \geq 0,0} = 0,0014 = 0,14$$

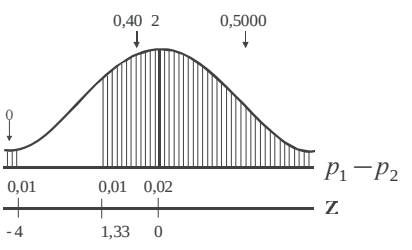
$$P_{P_1 - P_2 \geq 0,0} = 0,14$$

Ejemplo 2. Dos fábricas A y B, producen artículos similares. La producción de A contiene 7% de defectuosos, y la de B contiene, 5%. Si se extrae una muestra aleatoria de 2.000 de cada una de las producciones de las fábricas, ¿cuál es la probabilidad de que las dos muestras revelen una diferencia en el número de los defectuosos del 1% o más?

o : $P_A = 0,07$ $P_B = 0,05$ $n_1 = 2\ 000$ $n_2 = 2\ 000$ $P_{(P_1 - P_2 \geq |0,01|)}$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1\ Q_1}{n_1} + \frac{P_2\ Q_2}{n_2}}}$$

$$Z = \frac{0,01 - 0,02}{\sqrt{\frac{0,07\ \ 0,\ 3}{2\ 000} + \frac{0,05\ \ 0,\ 5}{2\ 000}}}$$



$$Z = \frac{-0,01}{0,0075} = -1,33 \qquad P = 0,5000 + 0,40\ \ 2 = 0,\ \ 0\ \ 2$$

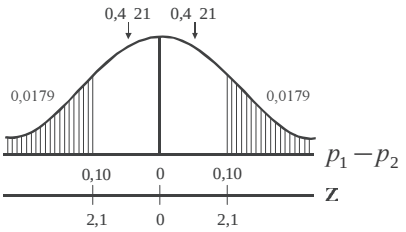
$$Z = \frac{-0,01 - 0,02}{0,0075} = -4 \qquad P_{(P_1 - P_2 \geq |0,01|)} = \ \ 0,\ \ 2$$

$$Z = -1,33 \rightarrow A(0,40\ \ 2)$$
$$Z = -4\ \ \rightarrow A(0,5000)$$

Ejemplo 3. Se sabe que cierta marca de crema para las manos satisface e 65% del mercado. ¿Cuál es la probabilidad de que dos muestras aleatorias de 200 usuarios cada una, muestre una diferencia mayor del 10% en las proporciones de uso de la crema?

o **:** $P_1 - P_2 = 0 \qquad P_{(P_1 - P_2 > |0,10|)} = \qquad P_1 = 0,65 \qquad P_2 = 0,65$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\dfrac{P_1\ Q_1}{n_1} + \dfrac{P_2\ Q_2}{n_2}}}$$



$$Z = \frac{0,10 - 0}{\sqrt{\dfrac{0,65\ 0,35}{200} + \dfrac{0,65\ 0,35}{200}}} = 2,1$$

$$A = 0,4\ \ 21 + 0,4\ \ 21 = 0,\ \ 642$$

$$P_{(P_1 - P_2 > |0,10|)} = 1,0000 - 0,\ \ 642 = 0,035$$

$$Z = \frac{-0,10 - 0}{0,0477} = \frac{-0,10}{0,0477} = -2,1$$

$$P_{(P_1 - P_2 > |0,10|)} = 3,5$$

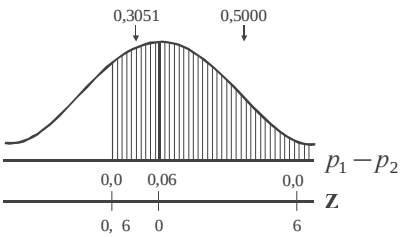
$$Z = 2,1 \rightarrow A\ \ 0,4\ \ 21$$
$$Z = -2,1 \rightarrow A(0,4\ \ 21)$$

Ejemplo 4. Suponga que una máquina A produce, por término medio, un 12% de piezas defectuosas, en tanto que la máquina B, el 18% de piezas con defectos. Se desea hallar la probabilidad de que el promedio de defectuosas de una muestra de 600 unidades, tomada de la máquina A, no difiera en más de un 8 % de otra muestra de 400 unidades, de la máquina B.

o **:** $P_1 = 0,12 \qquad P_2 = 0,18 \qquad n_1 = 600 \qquad n_2 = 400 \qquad P_{(P_1 - P_2 < |0,08|)} =$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\dfrac{P_1\ Q_1}{n_1} + \dfrac{P_2\ Q_2}{n_2}}}$$

$$Z = \frac{0,0\ \ -\ -0,06}{\sqrt{\dfrac{0,12\ \ 0,}{600} + \dfrac{0,1\ \ \ 0,\ \ 2}{400}}}$$



$$Z = \frac{0{,}14}{0{,}0233} = 6$$

$$Z = \frac{-0{,}0 \;\; - \;\; -0{,}06}{\sqrt{\frac{0{,}12 \;\; 0{,}}{600} + \frac{0{,}1 \;\; 0{,} \;\; 2}{400}}}$$

$$Z = \frac{-0{,}02}{0{,}0233} = -0{,}\;\; 6$$

$$Z = 6 \rightarrow A \;\; 0{,}5000$$

$$Z = -0{,}\;\; 6 \rightarrow A(0{,}3051)$$

$$P_{(P_1 - P_2 < |0{,}0\;\;|)} = 0{,}5000 + 0{,}3051 = 0{,}\;\; 051$$

$$P_{(P_1 - P_2 < |0{,}0\;\;|)} = \;\; 0{,}51$$

Ejemplo 5. Ciertas encuestas a televidentes, revelan que el 25% de los hombres y 33% de las mujeres de clase media, ven la telenovela de las 11 y 1/2 de la mañana. ¿Cuál es la probabilidad de que en dos muestras aleatorias de 150 hombres y 100 mujeres respectivamente, pertenecientes a dicho estrato social, se encuentre que la proporción de hombres que han visto el programa sea igual o mayor que la proporción de mujeres?

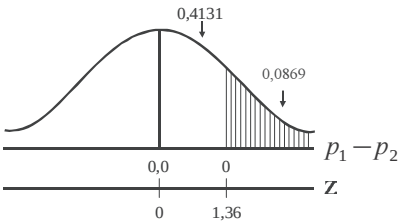
o **:** $P_1 = 0{,}25 \quad P_2 = 0{,}33 \quad n_1 = 150 \quad n_2 = 100 \quad P_{(P_1 - P_2 > 0)} =$

$$Z = \frac{0 - \;\; 0{,}25 - 0{,}33}{\sqrt{\frac{0{,}25 \;\; 0{,}75}{150} + \frac{0{,}33 \;\; 0{,}67}{100}}}$$

$$Z = \frac{0{,}0}{0{,}05 \;\; 3} = 1{,}36$$

$$Z = 1{,}36 \rightarrow A \;\; 0{,}4131$$

$$P_{(P_1 - P_2 > 0)} = 0{,}5000 - 0{,}4131 = 0{,}0 \;\; 6$$



$$P_{(P_1 - P_2 > 0)} = 8{,}6$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

82. Ciertas encuestas realizadas en una ciudad de la costa, revelan que el 25% de los hombres y el 33% de las mujeres escuchan cierto programa radial. ¿Cuál es la probabilidad de que en dos muestras de 150 hombres y 100 mujeres respectivamente, domiciliados en dicha ciudad, se encuentre que la proporción de mujeres que escuchan el programa sea menor o igual a la proporción de hombres?

83. En dos fábricas A y B, que producen camisas para hombre, se sabe que el 17% y 15% de la producción es defectuosa. Si se extrae una muestra de 200 camisas de cada lote producido en las fábricas, ¿Cuál es la probabilidad de que las dos muestras revelen una diferencia superior al 3%?

84. Se sabe que cierta marca de dentífrico satisface el 65% del mercado. ¿Cuál es la probabilidad de que dos muestras aleatorias de 200 usuarios cada una, revele una diferencia mayor del 10% en las proporciones del uso del dentífrico?

85. Ciertas encuestas sobre un programa de televisión revelan que el 28% de los hombres y el 38% de las mujeres de clase media, ven dicho programa. ¿Cuál es la probabilidad de que en dos muestras aleatorias de 150 hombres y 100 mujeres respectivamente, pertenecientes a dicho estrato, se encuentre que la proporción de hombres que han visto el programa sea igual o mayor que la proporción de mujeres?

86. Se sabe que cierto producto satisface el 72% del mercado. Seleccionamos dos muestras (independientes) de la misma población, de tamaño 150 cada una. ¿Hallar la probabilidad de que revelen una diferencia:
a) mayor del 6%.
b) en la segunda muestra la diferencia sea superior en un 5%?

87. El 12% de la producción de una máquina es defectuosa, mientras que en otra similar es del 15%. Si se extraen dos muestras de tamaños 80 y 100 respectivamente, ¿Cuál es la probabilidad (en cuanto al porcentaje de defectuosos):
a) que las dos muestras revelen una diferencia superior al 3%.
b) que el porcentaje en la muestra A, sea superior a la de B?

88. Es sabido que los porcentajes de familias con ingreso superior a $570.000 en las ciudades A y B es de 25% y 20%, respectivamente. Se seleccionan al azar dos muestras de 100 familias en cada una de las dos ciudades y se comparan las proporciones muestrales de las familias con ingreso superior a $570.000. ¿Cuál es la probabilidad de que la proporción muestral correspondiente a la ciudad B sea:
a) mayor que la de A en 3% o más. b) menor que la de A en 3% o más?

89. En una agencia de empleo se sabe que por cada 100 personas que lo solicitan, 50, además de ser bachilleres, tienen alguna experiencia sobre el trabajo a desarrollar. Se extraen 2 muestras de la misma población, en forma independiente, de tamaño 36 cada una. ¿Cuál es la probabilidad de que las dos muestras difieran en 8 o más personas que tengan alguna experiencia sobre el trabajo?

90. En dos fábricas A y B, que producen determinados artículos para hombre; se sabe que el 8 y 12% de la producción es defectuosa. Si se extrae una muestra de 400 artículos de cada lote producido, en cada fábrica, cuál es la probabilidad de que las dos muestras revelen una diferencia inferior al 3%?

o o a a t o t a t o a a

TAMAÑO DE LA MUESTRA

Hasta ahora, se ha venido considerando el *tamaño* (n) *de la muestra* conocido, pero para determinarlo, es necesario identificar los siguientes componentes o elementos técnicos.

La varianza (σ_x^2). Corresponde al grado de variabilidad que presentan las unidades de la población. Mientras más grande sea σ^2 mayor será el tamaño de la muestra. El valor de σ_x^2 supuestamente conocido, de lo contrario se debe estimar a través de una investigación preliminar. En el caso de $\sigma_p^2 = PQ$, sucede algo similar, pero se tiene la costumbre de tomar P = 0,50 con lo cual se obtiene el máximo valor posible de n.

Nivel de confianza. Tiene relación directa con el tamaño de la muestra, por lo tanto se dirá que a mayor nivel de confianza más grande debe ser el tamaño de la muestra. Los valores de Z se obtienen mediante el uso de tablas tal como se ha venido haciendo. El nivel de significación es fijado por el investigador, de acuerdo con su experiencia.

Precisión de la estimación. Corresponde al margen de error que el investigador fija de acuerdo con el conocimiento que tenga acerca del parámetro que piensa estimar. Se le conoce como error de muestreo (E), siendo:

$$E = Z\frac{\sigma}{\sqrt{n}} \qquad E = Z\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Recursos humanos, financieros y tiempo, no entran dentro de la determinación técnica del tamaño de la muestra, pero es de suma importancia al establecer el tamaño de las investigaciones.

Cálculo de **en poblaciones infinitas**

La fórmula para calcular el tamaño óptimo en el muestreo aleatorio simple, en una población infinita se obtiene así:

$$Z = \frac{\bar{x}-\mu}{\sigma\ \sqrt{n}} \rightarrow \quad E = \bar{x}-\mu \qquad Error = \frac{Z\sigma}{\sqrt{n}} \rightarrow \quad de\ donde\ \ \sqrt{n} = \frac{Z\sigma}{E}$$

Siendo:

(En la variable)

$$n = \left(\frac{Z\sigma}{E}\right)^2 = \frac{Z^2\sigma^2}{E^2}$$

(En la proporción)

$$n = \frac{Z^2PQ}{E^2}$$

Ejemplo 1. El mantenimiento de cuentas puede resultar demasiado costoso, si el promedio de compra por cuenta, baja de cierto nivel. El gerente de un gran almacén por departamentos desea estimar el promedio de lo comprado mensualmente por los clientes que usan la cuenta de crédito, con un error de $2.500, y una probabilidad aproximada de 0,95. ¿Cuántas cuentas deberá seleccionar, si sabe que la desviación estándar es de $30.000, la cual fue obtenida de los balances mensuales de las cuentas de crédito?

o ***:***

$$n = \left(\frac{Z\sigma}{E}\right)^2 = \frac{1,\ 6^2\ \ 30\ 000\ ^2}{2\ 500^2} = 553{,}1\ \ = 554 \qquad n = 554\ cuentas$$

Nota: siempre el resultado obtenido se redondea al n mero superior, por peque a que sea la fracción.

Ejemplo 2. Un auditor desea tener un nivel de confianza del 95%, para que la verdadera proporción de error no exceda del 2%. Si la población es muy grande, ¿qué tamaño tendrá la muestra que va a tomarse, si el auditor estima que la proporción de error es del 5%?

$$n = \frac{Z^2PQ}{E^2} = \frac{1,\ 6^2\ \ 0{,}05\ \ 0,\ 5}{0{,}02^2} = 456{,}1\ \ = 457\ \text{cuentas} \qquad n = 457\ cuentas$$

Ejemplo 3. De una remesa, de la cual se tomó una muestra de 200 artículos, se encontró que 20 de ellos eran defectuosos. Con una confianza del 95%, calcular el error de la muestra.

o :

$$n = \frac{Z^2 PQ}{E^2} \qquad E^2 = \frac{Z^2 PQ}{n} \qquad E = Z\sqrt{\frac{PQ}{n}} = 1,\ 6\sqrt{\frac{0,10\ \ 0,\ \ 0}{200}} = 0,04157$$

$$p = \frac{20}{200} = 0,10 \Rightarrow P$$

$$E = 4,16$$

Ejemplo 4. Una firma constructora desea estimar la resistencia promedio de las barras de acero utilizadas en la construcción de edificios de apartamentos. ¿Qué tamaño de muestra se requiere para garantizar que habrá un riesgo de sólo 0,001 de sobrepasar un error de 5 kg. o más en la estimación? La desviación estándar de la resistencia de este tipo de barras se estima en 50 libras.

o :

$$P = 1 - 0,001 = 0,$$

$$P = \frac{0,}{2} = 0,4 \quad 5 \ \} \rightarrow Z = 3,27$$

$$n = \frac{3,27^2 (2.500)}{10^2} = 267,32 = 268$$

$$E = 5\,Kg = 10\ libras$$

$$\sigma = s = 50\ libras$$

$$\sigma^2 = s^2 = 2.500\ libras^2$$

$$n = 268\ barras$$

Ejemplo 5. En un día, la producción de tarjetas perforadas por una persona se encuentra en una gaveta. Se desea estimar el porcentaje de tarjetas que tiene al menos un error, mediante una muestra aleatoria simple. ¿Qué tamaño de muestra es necesario si se piensa que el porcentaje está entre el 4 y 8%? Se acepta un error estándar del 3% y confianza del 95%.

o :

$$n = \frac{Z^2 PQ}{E^2} = \frac{1,\ 6^2\ \ 0,0\ \ \ 0,\ 2}{0,03^2} = 314,15 \cong 315$$

$$n = 315\ tarjetas$$

En este caso se toma el valor de P más cercano a 0,50. También se puede tomar el promedio, es decir $P = \frac{4\ \ +}{2} = 6$ por lo tanto el tamaño n será:

$$n = \frac{1,\ 6^2\ \ 0,06\ \ 0,\ 4}{0,03^2} = 240,74 \cong 241\ tarjetas$$

$$n = 241\ tarjetas$$

Ejemplo 6. ¿Qué tamaño de muestra es necesario, si se considera una confianza del 90% para la proporción de la población, y el error es del 8%?

o :

$$n = \frac{1,64^2\ \ 0,50\ \ \ 0,50}{0,0^{\ 2}} = 105,06 \cong 106$$

$$n = 106$$

⇒ Observe que no se conoce el valor de *P*; en estos casos, se acostumbra tomar el valor de 0,50 con lo cual se obtiene el máximo valor de *n*.

Ejemplo 7. ¿Qué tamaño de muestra parece razonable, para un nivel de confianza del 99% con un error de 2 en cualquier sentido (± 2), si la desviación estándar es 12?

o :

$$n = \frac{2{,}57^2 \; 12^{\;2}}{2^2} = 237{,}7 \;\cong 23$$

$$n = 23$$

Tamaño óptimo en poblaciones finitas

La fórmula para el tamaño óptimo en el muestreo aleatorio simple, cuando la población es finita, se obtiene:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}} \Rightarrow E = \frac{Z\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N}} \Rightarrow E^2 = \left(\frac{Z^2\sigma^2}{n}\right)\left(\frac{N-n}{N}\right)$$

Siendo $n = \dfrac{Z^2 N\sigma^2}{NE^2 + Z^2\sigma^2}$ también: $n = \dfrac{\sigma^2}{\left(\frac{E}{Z}\right)^2 + \frac{\sigma^2}{N}}$

La fórmula más utilizada $n = \dfrac{n_0}{1 + \frac{n_0}{N}}$ donde $\rightarrow$ $n_0 = \dfrac{Z^2\sigma^2}{E^2} = \left(\dfrac{Z\sigma}{E}\right)^2$

Las anteriores fórmulas que se aplican en variables, también con una pequeña modificación pueden ser utilizadas en atributos, es decir, cuando se calculan proporciones. Las fórmulas son:

$$n = \frac{Z^2 N\,PQ}{(N-1)E^2 + Z^2\,PQ} \qquad \text{ó} \qquad n = \frac{PQ}{\left(\frac{E}{Z}\right)^2 + \frac{PQ}{N}}$$

$$\text{ó} \qquad n = \frac{n_0}{1 + \frac{n_0}{N}} \qquad \text{siendo} \rightarrow \qquad n_0 = \frac{Z^2\,PQ}{E^2}$$

$\Rightarrow n_0$ se considera como primera aproximación, dado que algunos investigadores utilizan su valor, por considerarlo que económicamente, al igual que el tiempo y los recursos humanos disponibles lo permiten.

Ejemplo 8. Si el ejemplo (1) en poblaciones infinitas dijera que el total de cuentas de crédito es de 4.000, ¿cuál sería el tamaño de la muestra?

o :

$$n = \frac{Z^2\,N\sigma^2}{NE^2 + Z^2\sigma^2} = \frac{1,\;6^2\;\;4\,000\;\;\;30\,000^{\;2}}{4\,000\;\;1\,500^{\;2} + 1,\;6^2\;\;30\,000^{\;2}} = 1\,110{,}16 = 1\,111\;\textit{cuentas}$$

también: $n = \dfrac{\sigma^2}{\left(\frac{E}{Z}\right)^2 + \frac{\sigma^2}{N}} = \dfrac{30\,000^2}{\left(\frac{1\,500}{1\;\;6}\right)^2 + \frac{30\,000^2}{4\,000}} = 1\,111\;\textit{cuentas}$

$$n = 1\,111\;\textit{cuentas}$$

Ejemplo . Un investigador está interesado en estimar la ganancia en peso total, en 0 a 4 semanas de 1.000 pollitos alimentados con una nueva ración. Obviamente, pesar cada ave sería tedioso y llevaría demasiado tiempo. Por lo tanto, se debe determinar el número de pollitos a seleccionar en una muestra, para estimar el total con un límite para el error de estimación igual a 1.000 gramos. Muchos estudios similares sobre nutrición de pollitos se han llevado a cabo en el pasado. Usando datos de estos estudios, el investigador encontró que la varianza es, aproximadamente, de 36 gramos. Determine el tamaño de muestra requerido, con un confianza del 95%.

o **:** $N = 1\,000$ $P = 5$ $Z = 1,\ 6$ $s^2 = \sigma^2$

$E = 1.000$ *(error total)* $E = \dfrac{1.000}{1.000} = 1$ *(error promedio)*

$$n = \frac{N Z^2 S^2}{(N-1)E^2 + Z^2 S^2} = \frac{1.000(1,\ 6)^2(36)}{(1)^2 + 1,\ 6^2(36)} = 121,60 \cong 122 \ \textit{pollitos}$$

$n = 122$ *pollitos*

Ejemplo 10. Se desea realizar una investigación sobre el número de unidades que se encuentran en mal estado en 4.000 cajas y la proporción de cajas que contienen unidades en mal estado. Se realiza una encuesta preliminar de 80 cajas con el siguiente resultado, presentado en una tabla de frecuencias:

No. DE UNIDADES DEFECTUOSAS:	0	1	2	3	4	5	10	12
No. DE CAJAS EXAMINADAS:	37	16	8	8	4	2	2	3

Determinar el tamaño de muestras que cumpla con las dos condiciones. Para ello el investigador debe establecer un error del 6% para el promedio, del 12 % para la proporción y una confianza del 95%, para ambos casos.

o **:**

a) $Z = 1,\ 6$ $N = 4\,000$ $\bar{x} = \dfrac{\sum x_i n_i}{m} = \dfrac{13}{0} = 1,725 \cong 1,73$

$$s^2 = \frac{\sum x_i^2 n_i - n\bar{x}^2}{n} = \frac{66 - 0\ 1,725^{\ 2}}{0} = 7,\ 5 \qquad E = 0,06\ \ 1,73 = 0,10$$

$$n = \frac{1,\ 6^2\ \ 4\,000\ \ \ 7,\ 5}{4\,000\ \ 0,10^{\ 2} + 1,\ 6^2\ \ 7,\ 5} = 1\,720 \ \textit{cajas}$$

$n = 1\,720$ *cajas*

b) $p = \dfrac{0 - 37}{0} = \dfrac{43}{0} = 0,5375 = 0,54$

$$n = \frac{4\,000\ \ 1,\ 6^{\ 2}\ \ 0,54\ \ \ 0,46}{4\,000\ \ 0,12^{\ 2} + 1,\ 6^2\ \ 0,54\ \ \ 0,46} = 66 \ \textit{cajas}$$

$n = 66$ *cajas*

El tamaño óptimo es de 1.720 cajas, ya que se toma el mayor valor de n calculado..

Ejemplo 11. El Departamento de Tránsito y Transporte requiere estimar la proporción de conductores con experiencia de un año o menos, que puedan clasificarse como conductores descuidados. ¿De qué tamaño deberá ser la muestra a fin de que los resultados estén dentro de un 2%, con una confianza del 95%? Se espera observar que aproximadamente ¼ del total de conductores sean descuidados.

$$n = \frac{Z^2 PQ}{E^2} \quad n = \frac{1,\ 6^2\ 0,25\ \ 0,75}{0,02^2} = 1\ \ 00,75 \cong 1\ \ 01$$ Conductores con experiencia de un año o menos $n = 1.801$ *conductores*

Ejemplo 12. Si en el ejercicio anterior se dijera que el número de conductores a investigar, es de 10.000, ¿cuál será el tamaño de la muestra? (como ejercicio de aplicación, utilizamos las diferentes fórmulas vistas).

$$n = \frac{PQ}{\left(\frac{E}{Z}\right)^2 + \frac{PQ}{N}} \quad n = \frac{0,25\ \ 0,75}{\left(\frac{0,02}{1\ ,6}\right)^2 + \frac{0,25\ \ 0,75}{10\ 000}} = 1\ 526$$ Conductores con experiencia de un año o menos

$n = 1\ 526$ *conductores*

$$n = \frac{Z^2\ N\ PQ}{NE^2 + Z^2 PQ} \quad n = \frac{1,\ 6^2\ 10\ 000\ \ 0,25\ \ 0,75}{10\ 000\ \ 0,02^{\ 2} + 1,\ 6^2\ \ 0,25\ \ 0,75} = 1\ 526\ \ conductores$$

$$n_0 = \frac{Z^2 PQ}{E^2} \quad n = \frac{1,\ 6^2\ 0,25\ \ 0,75}{0,02^2} = 1\ \ 00,75 \qquad n = \frac{1\ \ 00,75}{1 + \frac{1\ \ 00,75}{10\ 000}} = 1\ 526\ \ conductores$$

Ejemplo 13. Un investigador asegura que el promedio salarial de los obreros, en cierto sector industrial es de $822.000 y sus edades oscilan entre 17 y 38 años; además, sus gastos en alimentación deben encontrarse entre el 40% y 60% de su salario. Se desea estimar el salario promedio, (suponiendo una desviación típica d $33.600); al mismo tiempo el porcentaje de sus gastos en alimentación. Se considera un error del 1,5% para el promedio y del 8% para la proporción; además la confianza será del 95% y un total de 4.300 obreros; ¿cuál será el tamaño óptimo para las dos características?

a) $n_0 = \frac{Z^2 S^2}{E^2} \quad n = \frac{1,\ 6^2\ 33\,600^{\ 2}}{12\,330^2} = 2\ ,53 \qquad n = \frac{2\ ,53}{1 + \frac{2\ ,53}{4\ 300}} = 2$ *obreros*

$E = \quad \bar{x} = 0,015\ \ 22\ 000 = 12\ 330$

$n = 2$ *obreros*

Tama o con corrección. Si al calcular el tamaño de la muestra (n) o la primera aproximación (n_o), se utiliza la desviación típica o la varianza proveniente de un encuesta preliminar o piloto (n_p), de tamaño menor de 30, algunos prefieren aplicar un factor de corrección a fin de que n o n_o, sean un poco mayor, dado que la desviación típica estimada (s) va a ser menor que la poblacional (σ). De acuerdo a lo anterior se tendrá que:

$$n = n_0 = \left[\frac{z^2 s^2}{E^2}\right]\left[1 + \frac{2}{n_p}\right] \qquad n = \frac{n_0}{1 + \frac{n_0}{N}}$$

n_p = tamaño de la muestra piloto

Nota: Esta fórmula fue aplicada en algunos *ejercicios para resolver,* problemas que se presentan desarrollados en el SIL.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

En el capítulo 13 se explica nuevamente el tema, con más ejercicios de aplicación, por la importancia que tiene para el investigador

91. De una población, N = 10.000 personas nos proponemos obtener una muestra, para estimar el ingreso promedio por persona. Se quiere que la estimación muestra, no se aparte en más de $5.000 del promedio verdadero y que esto se cumpla en 95 de cada 100 casos. La desviación típica es de $30.000. ¿Cuál será el tamaño óptimo?

92. Supongamos que en un área dada, la proporción de explotaciones agropecuarias que poseen energía es de 0,36. ¿Cuál será el error de muestreo de la estimación, utilizando una muestra al azar de 300 explotaciones, con una confianza del 95% un total de 8.000 explotaciones?

93. ¿Qué tamaño deberá tener una muestra para estimar dentro del 3%, la proporción de mujeres casadas que van periódicamente a consulta ginecológica, en una población de 5.000 mujeres y una seguridad del 95%?

94. Se desea estimar el costo promedio de matrículas de los estudiantes universitarios de la ciudad. Por estudios anteriores y a precios actuales se sabe que la desviación típica es de $18.000. a) Calcular el tamaño muestral fijando para ello un error de $\pm$3.000 y una confianza del 99%; b) Si se considera que la población estudiantil que se desea investigar es de 12.000, ¿cuál sería el valor de n?; c) Calcular el valor de n si se desea estimar el valor total de la matrícula cancelada por los 12.000 estudiantes.

95. La gerencia de una empresa manufacturera desea hacer una investigación entre sus trabajadores, con el fin de establecer si a través de cursos de entrenamiento y programas de mejoramiento de las condiciones del trabajo, tanto en la empresa como en su vida familiar, se logra elevar el rendimiento del personal. Con la ayuda de un experto en estudios de tiempos y movimientos, además de una trabajadora social, se realiza una encuesta preliminar en 70 de los 3.600 trabajadores. Algunos resultados de la encuesta fueron:
    a) el tiempo promedio necesario para realizar una operación fue de 40 minutos, con una varianza de 2,4 horas.
    b) 44 de los trabajadores son casados o de unión libre.
    c) El total de gastos mensuales en recreación de los hijos fue de $59.000, con una desviación típica del promedio igual a $325,oo.

    Nota: Las 3 características anteriores se consideran importantes en la determinación del tamaño de la muestra. Se ha fijado un coeficiente de confiabilidad de 95%, un error del 5% para el promedio y 10% para la proporción. ¿Qué tamaño de n recomendaría usted?

96. En un barrio residencial se espera que el 60% de las familias tengan vehículo propio. Se desea hacer una investigación para estimar la proporción de familias propietarias de vehículo, con un intervalo de confianza cuya amplitud no sea mayor de 0,03 y un coeficiente de confianza del 95,5%
    a) Determinar el tamaño de la muestra.
    b) ¿Qué sucedería si P = 0,50?; ¿si es igual a 0,90?
    c) En el caso del aparte a), si se conoce el número de familias en el barrio (N = 10.000), ¿cuál sería el tamaño de la muestra?

97. Un estimativo de la proporción de artículos alterados de un inventario en depósito, bajo condiciones desfavorables, es obtenido con un error máximo de 0,03 y un coeficiente de confianza del 97,5%. El total consta de 20.000 artículos y se estima por anticipado que la proporción de artículos no alterados es del 85%. ¿Cuál debe ser el tamaño de la muestra para asegurar un estimativo dentro de la precisión deseada?

98. Una oficina de investigaciones sobre salud considera que el 20% de las personas adultas de una región, padecen cierta enfermedad parasitaria. ¿Cuántas personas tendrán que seleccionarse en la muestra al azar, para que el error del estimado de la proporción sea del 7% y tenga una confianza del 99%?

99. Interesa estimar el número promedio de accidentes de tránsito en una ciudad. Durante un año (365 días) se determina una desviación típica de 12 accidentes diarios. ¿Cuántos días (tamaño de la muestra) se requieren observar para no errar en más de dos accidentes, con un 90% de confianza?

100. Se selecciona una muestra aleatoria simple de familias de clase media baja en un barrio de la ciudad, con el fin de estimar el ingreso promedio mensual. El error debe estar en el rango de $5.000.oo, con un riesgo de 0,045. ¿De qué tamaño debe ser seleccionada la muestra, si la desviación normal ha sido calculada en $28.000.oo?

101. Entre los estudiantes de cierta universidad privada, se desea tomar una muestra aleatoria para estimar la proporción de alumnos que utilizan la biblioteca. El error debe conservarse en un 4%, con un riesgo del 0,045. ¿Cuál es el tamaño de la muestra, si la universidad tiene 3.200 alumnos matriculados?

102. Determine el tamaño máximo de una muestra para estimar una proporción, con una confianza del 99%, sin que el error en la estimación exceda del 2%, para una población de 10.000.

103. Un investigador a cuyo cargo está un departamento de educación física, desea hacer una estimación del consumo de oxígeno (en litros por minuto) de los estudiantes normales entre 17 y 21 años de edad, después de un tiempo especial de ejercicios. El investigador desea que su estimación se encuentre por lo menos a 0,10 litros de la media verdadera, con una confianza del 95%. Estudios realizados indican que la varianza del consumo de oxígeno, para grupos similares, es de 0,90. ¿Qué tamaño debe tener la muestra que necesita este investigador?

104. Un investigador de un Instituto descentralizado que tiene 1.500 empleados, desea hacer una estimación del tiempo promedio que gastan los funcionarios, entre el instituto y la casa. El investigador desea un intervalo de confianza del 99% y una estimación comprendida entre 10 minutos y la media verdadera. Un pequeña muestra preliminar dio una varianza de tres horas y 15 minutos. ¿Qué tamaño debe tener la muestra que necesita el investigador?

105. Se desea hacer una investigación sobre el ingreso familiar promedio semanal de los 12.500 hogares en una ciudad intermedia. Por investigaciones anteriores, se considera que la desviación típica de los ingresos es de $30.000. ¿Qué tamaño debe tener la muestra, si se desea hacer una estimación de la media que se encuentra a $3.000.oo de la media verdadera, con un nivel de confianza de 95,5%?

106. Una trabajadora social que presta sus servicios en una Junta de Acción Comunal, desea hacer un estudio para determinar las actitudes de la comunidad frente a los programas que desea emprender. Se debe calcular el tamaño de la muestra, para hacer una estimación de la proporción de ciudadanos

que están de acuerdo con los programas. Investigaciones realizadas anteriormente en zonas similares, demuestran que el 72% de las personas entrevistadas contestaron afirmativamente. Además, se desea que el valor de la estimación esté a 0,12 del valor verdadero, con un 95% de confianza.

107. Un veterinario quiere hacer una estimación, en una población de ganado vacuno, sobre la proporción de reses infestadas por un parásito intestinal. a) ¿Qué tamaño de muestra se debe tomar, si se quiere que su estimación esté 0,05 de la proporción real, con 95% de confianza? No se tiene conocimiento de P, ni se puede obtener una encuesta preliminar; b) Hallar el tamaño de la muestra, suponiendo que un estudio anterior encontró que el 28% del ganado estaba contaminado; c) Si el veterinario quiere que su estimación esté a 0,02 de la proporción real y, la población es de 2.000 animales, ¿cuál es el valor de n? (Considere P = 0,5 y P = 0,28).

108. Un especialista en publicidad desea calcular el tamaño de la muestra de hogares en un barrio de la ciudad, para determinar en qué proporción, por lo menos, uno de sus miembros ve el programa musical. Se desea que la estimación esté a 0,04 de la proporción verdadera, con un 90% de confianza. En una encuesta preliminar a 30 hogares, el 30% de los entrevistados indicó que alguien veía regularmente dicho programa.

109. Una universidad desea ofrecer una nueva carrera profesional; para ello debe calcular la proporción de alumnos de último año de secundaria que piensa estudiar dicha carrera. ¿Qué tamaño debe tener la muestra si su estimación debe estar a 0,03 del valor verdadero, con 95% de confianza? a) el año anterior el 20 % de los alumnos encuestados se inclinaba por una carrera similar; b) El número de alumnos que cursan el último semestre en la ciudad donde se realiza la investigación, es de 6.000.

110. Un analista de departamento quiere estimar el número promedio de las horas de entretenimiento anual para los supervisores de una división de la compañía, con un factor de error de $\pm$ 3 horas y un 95% de confiabilidad. Toma información de otras divisiones para calcular la desviación típica de horas de capacitación anual en s = 20 horas. ¿Cuál es el tamaño mínimo requerido, si la compañía tiene 200 supervisores?

111. Se requiere determinar el tamaño de muestra para el estudio del gasto semestral en productos de tocador. Para estimar la desviación típica de la población, el investigador realiza una encuesta preliminar en 80 familias de un barrio de la ciudad, habiendo encontrado una desviación típica de $13.000.oo. a) ¿Cuál es el valor de n, si se desea que el valor estándar de la media inferida de la población no exceda de $2.400 a un nivel de confianza del 95%. b) ¿Cuál es el tamaño de la muestra, si se sabe que en el barrio investigado, viven 1.200 familias?

112. Un contador desea hacer un estudio sobre los profesores universitarios en la ciudad de México; para saber la cantidad de dinero por mes que cada profesor dedica a la alimentación de la familia. Realiza un inventario del número de profesores vinculados a las diferentes universidades y obtiene un listado de 2.000. El contador dice que el promedio de gastos semanales en alimentación que a él interesa, debe encontrarse en alrededor de $30.000.oo, ya que la mayoría son casados, entre 30 y 50 años de edad y el nivel de sueldos es aceptable. Se tiene un desviación estándar de $2.980, error del 3% y una confianza del 99%. ¿Cuál debe ser el tamaño de la muestra?

113. Se selecciona una muestra aleatoria de familias para estimar el ingreso promedio diario. El error debe estar en el rango de $22.000, con un riesgo de 0,045. ¿De qué tamaño debe ser seleccionada la muestra? La desviación normal ha sido calculada en $88.000.

114. En una región, se realizó una encuesta preliminar para determinar el promedio de empleo en explotaciones agrícolas; para ello se tomó una muestra simple al azar de 40 explotaciones, con los siguientes resultados:

2	2	3	2	3	5	2	4	3	4
2	4	1	6	6	7	2	5	4	2
4	2	2	2	5	2	2	2	2	6
2	3	5	4	2	4	2	6	5	3

Determinar el tamaño de la muestra, si se tiene un total de 4.000 explotaciones agrícolas, una confianza del 95% y un error del 5%.

115. Una encuesta preliminar realizada entre 30 estudiantes de primaria, correspondiente a cinco concentraciones escolares localizadas al sur de una ciudad, arrojó los siguientes resultados:

No. DE PIEZAS CON CARIES:	0	1	2	3	4	5	6	7
No. DE ESTUDIANTES:	10	9	5	2	2	0	1	1

El objetivo es poder realizar una investigación parcial (muestreo), para calcular estimativos del promedio de caries por alumno y la proporción de alumnos con caries. ¿Cuál deberá ser el tamaño óptimo, si el número de alumnos matriculados en estas cinco concentraciones es de 7.800? Se desea un error del 5% y un confianza del 95,5%.

116. Se realiza una encuesta preliminar a 17 familias en cuanto a tamaño, ingreso diario, costo de alimentos diarios y si tiene suscripción a algún periódico de la ciudad. Los resultados fueron:

No. FAMILIAS	NÚMERO DE PERSONAS POR FAMILIAS	INGRESO DIARIO (miles)	COSTOS DE ALIMENTOS DIARIO (miles)	SUSCRIPCIÓN PERIÓDICO
1	2	26.2	8.2	Si
2	3	18.4	9.6	No
3	3	19.3	10.1	No
4	5	14.5	7.6	No
5	4	26.2	12.2	Si
6	7	30.1	11.3	Si
7	2	24.2	10.4	No
8	4	26.2	12.3	No
9	2	22.1	10.2	Si
10	5	20.6	10.2	No
11	3	24.2	11.3	No
12	6	20.3	8.4	No
13	4	18.6	7.6	Si
14	4	15.2	7.8	No
15	2	17.9	7.9	No
16	5	19.4	8.3	No
17	3	28.2	11.4	Si

Se quiere determinar el tamaño muestral para una población de 1.200 familias, en cuanto a:
a) estimación del promedio de personas por familia.
b) proporción de familias con suscripción P = 95% E = 5%.

117. Suponga que una compañía desea estimar la proporción de cuentas que incluyen gastos por trabajo y monto total del año anterior. Suponga que se ha fijado una confianza del 95% y un error del 5%. ¿Qué tamaño de muestra se debe seleccionar, si una encuesta preliminar de 30 cuentas dio como resultado 12 tarjetas que incluyen gastos por trabajo y el total fue de $5.400.000, con un desviación típica de $20.000?

118. Se tomó una encuesta preliminar de 100 colegios de secundaria, de un total de 4.680, dando los siguientes resultados:

TIPO DE COLEGIO	CANTIDAD DE COLEGIOS	NÚMERO DE ESTUDIANTES		NÚMERO DE PROFESORES		$\sum x_i y_i$
		$\sum x_i$	$\sum x_i^2$	$\sum y_i$	$\sum y_i^2$	
PÚBLICOS	54	31.281	29.881.219	2.024	111.090	1.729.349
PRIVADOS	46	13.707	6.366.785	1.075	33.119	431.041

Se quiere determinar el tamaño de la muestra de los colegios (públicos y privados) con una confianza del 95% y un error del 8%, para estimar:
a) El promedio de alumnos por colegio
b) La proporción de colegios privados

119. Una cooperativa de ahorro y préstamo desea determinar la cantidad promedio que tienen los clientes en sus cuentas. La desviación estándar de todas las cuentas de ahorro es estimada por el gerente en $44.000.
   a) ¿Qué tamaño de muestra se requiere para afirmar, con una confianza de 95%, que el error en la estimación no excede de $20.000?
   b) ¿Qué tamaño de muestra se requiere para afirmar, con una confianza de 95%, que el error en la estimación no excede de $30.000?

120. Determinar el tamaño máximo de una muestra de proporciones siendo la población de 30.000 amas de casa, con un error estándar de ±2 % y un nivel de confianza del 99%.

121. Interesa estimar el número de accidentes de tránsito en una ciudad cualquiera. Durante un año (365 días) se determina una desviación típica de 12 accidentes diarios. ¿Cuántos días (tamaño de muestra) debemos observar para no errar, en más de 3 accidentes, con un 90% de confianza?

122. Se quiere determinar el tamaño de la muestra para un estudio sobre el gasto mensual en productos de limpieza para el hogar. Para estimar la desviación típica de la población, el investigador utilizó una muestra preliminar de 80 familias, habiendo encontrado que la desviación estándar era de $20.500.
   a) ¿Cuál es el valor de n si se desea que el error estándar de la media inferida de la población no exceda de ± $2.400 con un nivel de confianza del 95%?
   b) ¿Cuál es el tamaño de muestra en el punto anterior, si se sabe que N = 2.000?

123. Se selecciona una muestra aleatoria simple de familia de clase media en un barrio de la ciudad, con el fin de estimar el ingreso promedio semanal. El error debe estar en el rango de $20.000, con un riesgo de 0,045. ¿De qué tamaño debe ser seleccionada la muestra, si la desviación normal ha sido calculada en $80.000.

124. Entre los estudiantes de cierta universidad privada, se desea una muestra aleatoria para estimar la proporción de lectores de obras literarias. El error debe conservarse en un 5%, con un riesgo de 0,045. ¿Cuál es el tamaño de muestra requerida?

125. Una fábrica de gaseosas tiene un total de 628 camiones repartidores; la varianza de las ventas diarias de todos los camiones es de 25.000 (miles de $). Se va a seleccionar una muestra aleatoria de camiones para realizar con base en ella estimación de la media poblacional de las ventas. ¿Cuál será el tamaño de la muestra, para que 0,95 sea la probabilidad (confianza) de que la diferencia entre la media muestral y la población no exceda de 80 (miles $)?

126. En una zona donde hay 4.000 viviendas, el porcentaje de propietarios va a ser estimado con una muestra no mayor del 2% y el porcentaje de propietarios de autos no mayor al 1%. Con una confianza del 90%.
Se piensa que el verdadero porcentaje de propietarios de vivienda puede estar entre el 45% y el 65%, y el porcentaje de propietarios de 2 autos entre el 5 y el 10%, ¿qué tan grande debe ser una muestra para satisfacer las dos finalidades?

127. Una empresa productora de bombillas de 75W, quiere estimar su vida media de tal manera que el error estándar de la media inferida no sea mayor a las 25 horas. Se sabe que la vida media tiene una desviación típica de 100 horas y se desea una confianza del 95% para la estimación. ¿Cuán grande debe ser la muestra? ¿Cuál sería la respuesta si la población fuera de 1.000 bombillas?

128. Una biblioteca pública quiere calcular el porcentaje de libros de que dispone con fechas de publicación de 2005 o anteriores. ¿De qué tamaño debe tomar la muestra aleatoria para que se tenga un 90% de seguridad de quedar dentro del 5% de la proporción real de la muestra? Se sabe que la biblioteca tiene 5.000 títulos.

129. El administrador de un restaurante quiere determinar el tiempo promedio que los clientes tardan en tomar sus alimentos con el fin de efectuar una ampliación de sus instalaciones. Considera que la desviación estándar razonable es de 10 minutos basados en una muestra anterior hecha a 30 clientes. En esa encuesta se encontró que el 75% de los que frecuentan el establecimiento son hombres. Se quiere determinar el tamaño de la muestra, con un intervalo de confianza de 95%, para tener la seguridad de que el promedio quede dentro del 6%. Es de anotar, el dueño afirma, que el promedio de permanencia de sus clientes es aproximadamente de una hora.

130. ¿Qué tamaño de muestra se requiere para una población de 5.000 unidades, si se desea un error del 8% y un intervalo de confianza del 95%?

131. A partir de una muestra de 200 observaciones se encontró que, en una remesa, había 20 acumuladores defectuosos. Utilizando un nivel de confianza del 99%, calcule el error de muestreo.

132. La Secretaría de Tránsito y Transporte, requiere estimar la proporción de conductores con experiencia de un año o menos que pueden clasificarse como *conductores descuidados*. ¿De qué tamaño debería ser la muestra a fin de que los resultados estén dentro de un 2%, con una confianza del 95%? Se espera aproximadamente observar que ¼ del total de conductores sean descuidados.

133. El mantenimiento de cuentas de crédito puede resultar demasiado costoso si el promedio de compra por cuenta, baja de cierto nivel. El gerente de un almacén desea estimar el promedio de cantidad

comprada por sus clientes mensualmente, que usan cuenta de crédito, con un error no más de $12.500, con una probabilidad aproximada de 0,95. ¿Cuántas cuentas deberán ser seleccionadas de un total de 7.000, si se sabe que la desviación estándar de los balances mensuales de las cuentas de créditos es de $95.000?

134. La producción, en un día, de tarjetas perforadas es de 3.000. Se quiere estimar el porcentaje de tarjetas que tienen, al menos, dos errores mediante una muestra aleatoria simple: a) ¿Qué tamaño de muestra es necesario si se acepta un error del 3%?; b) Si alguien que conoce el trabajo, considera que el porcentaje podría encontrarse entre el 72 y 80% , ¿cuál sería el tamaño? Deberá ser utilizada una confianza del 95%.

135. Al examinar el directorio telefónico de una localidad que contiene 50 mil suscriptores, se desea estimar el porcentaje de ellos cuyo apellido empieza con la letra V. Desgraciadamente por error de elaboración del directorio, no se hizo teniendo en cuenta el orden alfabético sino localización o dirección. Se pide encontrar el tamaño de la muestra, si el porcentaje es considerado en un 10%, con un error del 0,5% y una confianza del 95%.

136. Un fabricante de automóviles, produce una gran cantidad de vehículos antes de recibir pedidos. Le agradaría poder calcular la proporción de automóviles que debe pintar de verde pampa en su producción por adelantado. Este fabricante ofrece una opción entre 7 colores. a) ¿Qué tamaño de muestra necesitará? Si desea hacer una estimación dentro de un margen del 1% y un nivel de confianza del 98% para una producción de 5.000 vehículos; b) Defínase la población de interés para el fabricante.

137. Decir si son ciertas o falsas las siguientes aseveraciones:
    a) Se dice que en todo proceso de selección debe haber sustitución de las unidades que no se han podido entrevistar.
    b) El parámetro describe una determinada característica de las unidades que conforman la población.
    c) El muestreo sistemático consiste en seleccionar uno a uno los elementos de la muestra en un orden indeterminado.
    d) A medida que aumenta el error, manteniendo constante la varianza, se aumenta el tamaño de la muestra.
    e) Una muestra aleatoria es aquélla en la cual ciertas unidades tienen mayor posibilidad que otras de ser seleccionadas.

138. La Asociación de egresados de la universidad está trabajando para formular una propuesta relativa a la consecución de una sede en la universidad. Se tienen 5.600 egresados registrados, de los cuales se toma una muestra preliminar del 1%, con la cual se obtiene que 28 de ellos piensan que la idea deberá ser llevada a cabo. ¿Cuál deberá ser el tamaño de la muestra a fin de que la asociación pueda estimar la proporción de egresados que están a favor de la propuesta, dentro de un 3%, con una confianza del 95%?

139. Un analista del Departamento de personal, quiere estimar el promedio de horas de entrenamiento para los supervisores de la compañía y el porcentaje de supervisores que están de acuerdo con la efectividad de esos entrenamientos. ¿Cuál es el tamaño mínimo sugerido, si la compañía tiene 500 supervisores? La confianza será del 95%, los factores de error son ± 3horas y varianza de 400 horas^2 de entrenamiento, se sabe que 6 de cada 10 supervisores se mostraron partidarios, siendo el error del 5%.

140. Se tiene la siguiente población de 15 valores:

x_i:	6	2	8	7	2	4	10	9	10	65	4	3	8	6

Información adicional $\sum x_i = 90$ $\sum x_i^2 = 640$

Determine el tamaño de la muestra, si se desea un error de $\pm$ 2 y una confianza del 90%.

141. El gerente de una tienda de departamentos de tamaño medio, ha observado que el total de cuentas pendientes por tarjetas de crédito de la propia firma ha aumentado en forma alarmante. Para formular una política de descuento por pago oportuno, se desea estimar la cantidad total en pesos ($) de cuentas pendientes que tengan por lo menos 90 días de vencidas. ¿Cuál debe ser el tamaño de la muestra si se fija el error tolerable de $10.000, con probabilidad de 0,90? Supóngase que hay 1.500 cuentas distintas con 90 o más días de vencidas y se piensa que la desviación típica de esas cuentas vencidas es inferior a $20.000.

142. Un estimativo del total de artículos alterados de un inventario de depósito de cierto almacén, bajo condiciones desfavorables es obtenido dentro de un error de más o menos $\pm$ 0,03, con un 97,5% de nivel de confianza. El inventario total se debe hacer sobre un total de 20.000 artículos. Para ello se realizó una encuesta preliminar de 100 artículos de los cuales 85 no están alterados. ¿Cuál será el valor de n?

143. Un publicista desea calcular el tamaño de la muestra de hogares, que va a tomar en cierta región para determinar en qué proporción de hogares, por lo menos uno de sus miembros ve un programa musical de televisión. En esa región hay un total de 700 hogares y se desea que la estimación esté a 0,05 de la proporción verdadera, con una confianza del 90%. Anteriormente se realizó una encuesta en 40 hogares, en la cual 14 de los entrevistados contestaron que alguien en su casa veía regularmente dicho programa. ¿Cuál debe ser el tamaño de la muestra?

144. Una entidad social dedicada a la protección del niño realizó una muestra al azar de 225 mujeres, de una población de 6.000, con el fin de determinar sus actitudes ante ciertos problemas sociales. Una de las preguntas que tiene relación directa con el objetivo de la encuesta decía: ¿Cree usted que las madres de los niños con una edad inferior a los 7 años deben trabajar fuera del hogar? 90 contestaron negativamente. Si se desea que esa muestra tenga una confianza del 95%, ¿cuál deberá ser el error en la muestra?

145. El administrador de una taberna desea estimar la cantidad promedio gastada por su clientela los fines de semana, con un error menor de $12.000 y con una probabilidad del 95%. ¿Cuántos clientes que cancelan la cuenta deberán ser entrevistados, si se sabe que el número de clientes es de 400? Suponga que es administrador sabe que esos gastos varían entre $80.000 y $150.000.

146. Una firma constructora de apartamentos, desea estimar el promedio de resistencia de las varillas de acero utilizadas en la construcción. ¿Qué tamaño se requiere para garantizar un riesgo de sólo 0,05 de sobrepasar un error de 5 kg o más en la estimación? La desviación típica de la resistencia de este tipo de varillas se estima en 25 kg. y se tiene en el momento de realizar el estudio un total de 5.000 varillas.

147. Un fabricante de automóviles desea estimar el kilometraje promedio por galón de gasolina que los clientes obtendrán con un nuevo modelo de carburador. ¿Cuántos viajes de muestra debe efectuar,

a fin de que su estimación tenga una precisión de 0,5 kilómetros por galón, con una confianza del 95%, suponiendo que la desviación típica es de 1,9 kilómetros por galón?

148. Suponga que una compañía desea estimar la proporción de cuentas que incluye gastos específicos por trabajo y el valor promedio por cuenta. Además se ha fijado una confianza de 95% y un error del 6%. ¿Qué tamaño de muestra se debe fijar, si una encuesta preliminar de 30 cuentas dio como resultado 12 tarjetas que incluyen gastos específicos, por un valor de $5.400.000 y una desviación estándar de $20.000.

149. En una ciudad hay 360 fábricas de helados y se requiere realizar una muestra para estimar la proporción de consumo de acuerdo a una nueva forma de presentación. ¿Cuál debe ser el tamaño de la muestra si se desea una confianza del 95,5% y un error del 9%.

150. Un investigador asegura que el salario promedio de los obreros, en cierto sector industrial, es de $750.000 y sus edades oscilan entre los 18 y los 40 años, además sus gastos en alimentación deben encontrarse entre los 40 y 60%. Se desea estimar el salario promedio, (suponiendo para ello una desviación típica de $40.000) y el porcentaje de los gastos en alimentos. Considere un error del 2% para el promedio y un 8% para la proporción, además una confianza del 95% un total de 10.000 obreros, para calcular el tamaño de la muestra.

151. Se estima que el sesgo en la proporción de ejecutivos subalternos que renuncian en compañías manufactureras grandes, después de tres años es del 3%. Se utiliza el grado de confianza de 0,95. Un estudio realizado hace varios años reveló que el porcentaje de ejecutivos subalternos que renuncian después de tres años fue del 21%.
    a) Para actualizar este estudio, ¿cuántos ejecutivos subalternos deberían estudiarse de los archivos?
    b) ¿Cuántos funcionarios, de un total de 520 deberían considerarse, si no se contara con una estimación previa?

152. Se estimará el número medio de días de viaje al año de los vendedores foráneos empleados por una empresa. Se utiliza el grado de confianza del 90%. La media de un estudio piloto realizado a 25 vendedores fue de 5,2 meses y la proporción de viajeros con más de 100 días fue del 62%. La empresa emplea un total de 620 vendedores que recorren todo el país. Dicha encuesta permitió conocer o estimar la desviación estándar de 14 días. ¿Cuántos vendedores foráneos deberán considerarse si se establece un error del 1,5% para el promedio de viajes y del 12% para la proporción con más de 100 días de viaje?

153. Un investigador está interesado en estimar la ganancia total en peso, en un período de 0 a 4 semanas de 5.000 pollitos alimentados con una nueva ración. Obviamente pesar cada vez sería tedioso y llevaría demasiado tiempo. Por lo tanto, se debe determinar el número de pollitos a seleccionar en una muestra para estimar el total con un límite para el error de estimación igual a 5.000 gramos. Muchos estudios similares sobre nutrición de pollitos se han llevado a cabo en el pasado. Usando datos de estos estudios, el investigador encontró que la varianza es aproximadamente de 36 gramos. Determine el tamaño de la muestra requerido.

154. Se desea realizar una investigación sobre el número de unidades en mal estado, en 4.000 cajas que se encuentran en bodega, además, la proporción de cajas con unidades en mal estado. Se realiza una encuesta preliminar de 80 cajas con el siguiente resultado, que se presenta en una tabla de frecuencias.

No. DE UNIDADES EN MAL ESTADO:	0	1	2	4	5	8	10	12
No. DE CAJAS EXAMINADAS:	37	16	6	8	4	2	2	3

Determinar el tamaño óptimo para una muestra que cumpla con las dos condiciones anteriores, Para ello el investigador establece un error del 6% para el promedio de unidades en mal estado por caja y del 12% para la proporción de cajas con unidades en mal estado. La confianza será del 95%.

155. Los gerentes de una cadena de almacenes, han decidido actuar rápidamente y en forma drástica, para reducir el número (frecuencias) y el valor de las pérdidas debidas a hurtos. Se han propuesto varias alternativas algunas de ellas son costosas. Al investigarse cada alternativa se hace patente que la firma requiere de una estimación con la mayor precisión posible, del porcentaje de compradores que cometen hurto y el valor que ello representa. Un examen en fuentes secundarias, revela estimaciones de compradores que han hurtado un artículo en la semana pasada que van desde 0,01% al 10%, por un valor promedio de $23.000. Mediante una muestra aleatoria simple de compradores, serán seguidos durante toda su visita al almacén, mediante la utilización de detectives especialmente adiestrados. Si la gerencia quiere estar «prácticamente segura» que la estimación no difiera de la cifra real del 0,5% en ambos casos, ¿cuál ha de ser el tamaño de la muestra, si para el valor de las pérdidas, piensa que la desviación estándar es de $1.500 y el número de clientes que entran durante la toma de la muestra es de 5.000 personas? La confianza en ambos casos es de 95%.

156. A partir de una muestra de 200 observaciones se encontró que, en una remesa habrá 20 acumuladores defectuosos. Con una confianza del 95%, calcular el error de la muestra.

157. Un auditor desea tener un nivel de confianza del 95%, de tal manera que la verdadera proporción de errores no exceda del 6%. Si la población es de 2.000 cuentas, ¿qué tamaño tendrá la muestra que va a tomar, si el auditor estima que la proporción de errores es del 4%?

158. Explique brevemente la diferencia que hay entre:
a) Estimador y parámetro.
b) Población y muestra.
c) Estadística descriptiva e inferencia.
d) Muestreo aleatorio. y no aleatoria .

159. Conteste brevemente:
a) ¿Cuál es la esencia del teorema del límite central y cuál es su importancia?
b) ¿Qué condiciones debe reunir un buen estimador?
c) ¿Qué significa distribución de muestreo?

EJERCICIOS MISCELÁNEOS

160. Una muestra aleatoria de 150 apartamentos de un área metropolitana, produce una renta mensual promedio de $775.000. Se calculó una renta promedio de $780.000, con base en una muestra aleatoria independiente para 120 apartamentos de otra área metropolitana. Supongamos que no hay diferencia entre las dos áreas respecto a las rentas mensuales promedio de los apartamentos. ¿Cuál es la probabilidad de observar una diferencia entre las medias muestrales, más grande de la que se acaba de presentar, si se supone que la desviación poblacional es de $20.000 para ambas áreas?

161. Suponga que la media de la población de las utilidades después del impuesto, de todos los pequeños fabricantes, durante un cierto período, es en realidad de 4,8 millones de pesos, con desviación estándar de 1,5 millones de pesos. ¿Cuál es la probabilidad de que la utilidad media después de pagar impuestos, calculada mediante una muestra aleatoria de 100 pequeños fabricantes, exceda en $300.000?

162. Un procesador de alimentos envasa café en frascos de 400 gramos. Para controlar el proceso, se utiliza la siguiente regla de decisión. Se seleccionan 64 frascos cada hora; si su peso medio es superior a un valor crítico, se detiene el proceso y se reajusta; en caso contrario, se continúa la operación sin detener el proceso. Determinar el valor crítico de modo que haya una probabilidad de 0,05 de determinar el proceso cuando está envasado a un promedio de 407,5 gramos, con una desviación estándar de 2,5 gramos.

163. Una oficina gubernamental, toma una muestra aleatoria de 400 trabajadores de una fábrica para obtener el número de ellos a favor de la sindicalización. Calcule la probabilidad de obtener una proporción muestral que difiera en más del 3% de la realidad, si la proporción real de trabajadores a favor de la sindicalización se le considera de 20%.

164. Se desea estimar el gasto mensual en energía eléctrica por familia de clase media de una ciudad, que cuenta con 10.000 familias clasificadas en ese grupo de ingresos. Con base en estudios realizados en otras ciudades con estratificación y comportamiento de consumo similar, se supone que la desviación estándar es de $3.000. Se quiere estimar con una aproximación de ±$380 del promedio real, con 95% de confianza. ¿Qué tamaño de muestra se requiere para dicha investigación?

165. La asociación de egresados de la universidad, está trabajando para formular una propuesta relativa a la consecución de una sede en el área de la universidad. Se tiene un total de 3.600 egresados registrados, del cual se toma una muestra preliminar del 1%, con lo que se obtiene que 28 de ellos piensan que la idea deberá ser llevada a cabo. ¿Cuál deberá ser el tamaño de la muestra a fin de que la asociación pueda estimar la proporción de egresados que están a favor de la propuesta dentro de un 3% con una confianza del 95%?

166. Explique brevemente:
a) El objetivo del muestreo
b) “Si se utilizan métodos de muestreo no probabilístico, los resultados, pueden estar sesgados”
c) Qué significa «error de muestreo»
d) ¿En qué condiciones es más apropiado el muestreo no probabilístico que el probabilístico?

167. Un gerente de una compañía, realiza dos muestras de tamaño de 120 empleados, cada una, con el fin de determinar el porcentaje de ausentismo. En la primera fábrica, durante la semana de observación no asistieron 12 personas, mientras que en la segunda fábrica no fueron 16 trabajadores. ¿Cuál es la probabilidad de que la ausencia observada sea mayor, si por experiencia se sabe que es el 3%?

168. Una máquina vendedora de gaseosas, se encuentra regulada de modo que la cantidad servida tenga una distribución normal, con media de 10 onzas y desviación típica de 0,82 onzas. Si se toma una muestra de 25 vasos, ¿qué valor excedería el 90% de la media de la muestra?

169. La duración del viaje que realiza una empresa de transporte intermunicipal, están distribuida normalmente con media de 8 horas y desviación estándar de 2 horas. Se seleccionan al azar 20

de los viajes realizados durante el mes: a) ¿Cuál es la desviación estándar de la media? b) ¿Qué proporción de los viajes estaría entre 7 y 8 ½ horas? c) ¿Cuál es la probabilidad de que la duración del viaje sea superior a las 9 horas?

170. Se sabe que el 10% de los estudiantes de una facultad no presenta exámenes en la fecha en que se programan. Si se selecciona a 30 alumnos de la facultad, ¿cuál es la probabilidad de que 8 o más de ellos hayan dejado de presentar exámenes durante el semestre?

171. El 12% de la producción de una máquina es defectuosa, mientras que en otra similar es del 15%. Si se extraen dos muestras de tamaño 70 y 90 respectivamente ¿cuál es la probabilidad de que:
a) las dos muestras difieran en más del 3%.
b) el porcentaje en la muestra A sea superior al de la muestra B?

Síntesis de capítulo

Sobre dos aspectos se fundamenta la teoría de las distribuciones muestrales:

a) *La ley de los grandes números*. Si el tamaño de la muestra es lo suficientemente grande, la probabilidad de que la media muestral difiera de la media poblacional en más de una diferencia positiva prescripta en forma arbitraria se acerca a cero (0).

b) *Teorema central del límite.* La distribución de las medias muestrales al azar, se aproxima a una distribución normal a medida que aumenta el tamaño de la muestra.

El objeto de la *Inferencia estadística* es la de recoger información mediante la realización de una muestra, con el fin de utilizarla en el cálculo de estimaciones y así obtener conclusiones respecto a los valores estadísticos de la población.

Distribución de muestreo es aquélla que puede obtenerse como resultado de un número infinito de muestras aleatorias independientes, cada una de tamaño n, provenientes de la misma población.

Las muestras se clasifican en *aleatorias y no aleatorias*. En la primera, todos los elementos tienen la misma posibilidad de ser seleccionados. En las no aleatorias, las unidades se seleccionan en forma caprichosa, generalmente por conveniencia.

El *parámetro* es el resultado que se obtiene al aplicar una medida a las características de las unidades de la población; por lo tanto será una cantidad constante. Cuando la media describe la característica de los elementos en una muestra, se denomina estimador puntual, estadígrafo o valor estadístico de la muestra.

Las cuatro (4) propiedades de un buen estimador son: *insesgado, consistente, suficiente y eficiente.*

El *error estándar* de la media, es la diferencia que puede haber entre el estimado y el parámetro.

$$\sigma_{\bar{x}} = \sigma / \sqrt{n} \qquad \sigma_{\bar{p}} = \sqrt{\frac{PQ}{n}}$$

Las distribuciones muestrales que se estudiaron en el capítulo son:
a) Distribuciones de medias muestrales.
b) Distribuciones de medias proporcionales.
c) Distribuciones de diferencias entre dos medias muestrales.
d) Distribuciones de diferencias entre dos medias proporcionales.

El *tamaño de la muestra* depende del comportamiento de sus cuatro (4) componentes:

a) El grado de variabilidad de los datos.

b) El nivel de confianza.

c) El margen de error establecido.

d) El tamaño poblacional.

CAPÍTULO

8

Pruebas de ipótesis y límites de confianza

uien no duda no puede conocer la verdad

o aa a aa o

CONTENIDO

- Conceptos generales, usos y procedimientos de aplicación.
- Pruebas de hipótesis con aplicaciones en distribuciones de: Medias, Proporciones.
- Diferencias entre dos medias y entre dos Proporciones muestrales.
- Teoría de las muestras pequeñas. Distribución "t" de Student.
- Límites de confianza.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el Sistema de Información en Línea SIL.

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Entender la utilidad que éstas tienen en muchos casos de la vida real.
- Distinguir, describir y aplicar estas pruebas de hipótesis.
- Realizar pruebas unilaterales y bilaterales cuando las condiciones lo exijan.
- Entender, aplicar y explicar Límites de confianza.
- Distinguir los casos en que se pueda cometer errores de Tipo I y II.
- Aplicar estas pruebas en muestras grandes y pequeñas.

CONCEPTOS GENERALES

Es evidente que las distribuciones muestrales, vistas en el capítulo anterior, basadas en la teoría de la distribución normal, desarrollan un papel de gran importancia en la *inferencia estadística*.

La *inferencia estadística* comprende dos partes principales, a saber: la estimación de parámetros y la *prueba o docimasia de hipótesis*. En este capítulo serán estudiadas, con el fin de desarrollar procedimientos y observar su aplicación a problemas corrientes de la vida diaria.

La *inferencia estadística* está basada en el supuesto de tomar muchas muestras, todas con igual probabilidad de ser seleccionadas y a través de una de ellas sabremos algo acerca de la población, mediante el cálculo de estimadores, que nos permitan hacer aseveraciones, incorrectas algunas veces, estableciéndose la probabilidad de error.

Este método se basa en la aplicación de *técnicas de muestreo*, para lo cual se requiere de un buen diseño, además de la aplicación de métodos aleatorios de selección, cuando las probabilidades son iguales para cada elemento de una población. En algunos casos no requieren ser iguales, siempre que se conozcan y sean diferentes a cero.

En este capítulo, buscamos establecer si el valor del estimador obtenido en la muestra es representativo del parámetro o valor estadístico poblacional de la cual se extrajo; además, la identificación de la distribución en la que debe ser utilizada.

PRUEBA DE IPÓTESIS

Las *pruebas de hipótesis*, denominadas también *pruebas de significación,* tienen como objeto principal evaluar suposiciones o afirmaciones acerca de los valores estadísticos de la población, denominados *parámetros*.

La palabra *docimar* significa *probar*. Cuando se hace indispensable tomar una decisión sobre la validez de la representación en una población, con base en los resultados obtenidos a través de una muestra, se dice que se toman decisiones estadísticas. Para tomar una decisión, es necesario ante todo, plantear posibilidades acerca de la característica o características a estudiar en una población determinada. La suposición puede ser cierta o falsa. Estas suposiciones se llaman *hipótesis estadísticas*.

Hipótesis estadística: son supuestos acerca de un parámetro o de algún valor estadístico de una población. Con esta definición, encontramos que no todas las hipótesis son *hipótesis estadísticas*. Se debe tomar con referencia a un parámetro, ya sea una media aritmética, una proporción (porcentaje), varianza, etc., para que sea hipótesis estadística.

Una hipótesis estadística, también puede considerarse como la afirmación acerca de una característica ideal de una población sobre la cual hay inseguridad en el momento de formularla y que, a la vez, es expresada de tal forma que puede ser rechazada.

Ejemplo 1. Para entender en qué consiste una prueba o dócima estadística, supongamos que se efectúan 100 lanzamientos de una moneda y que se obtienen 60 caras (40 sellos). Vamos a probar la *legitimidad de la moneda* (que no esté cargada) tomando en cuenta que al lanzar 100 monedas, lo lógico sería que cayeran 50 caras y 50 sellos. Sin embargo, al realizar el experimento, encontramos que en vez de obtener las 50 caras, se presentaron 60; esta pequeña diferencia puede llevarnos a pensar, que la probabilidad de presentación de cara es mayor que la de sello. Dicho en otras palabras, que la moneda está cargada.

Determinemos en primer lugar la probabilidad de que se obtengan 60 caras o más,

$\mu = np$ $\qquad$ $\sigma = \sqrt{npq}$

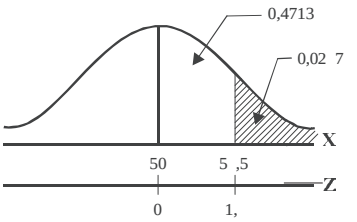
$$\mu = \left(\frac{1}{2}\right)100 = 50 \qquad \sigma = \sqrt{100\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{25} = 5$$

Al determinar la probabilidad de que en el lanzamiento de 100 monedas obtengamos 60 caras o más, utilizamos la Distribución normal, como aproximación a la Binomial.

$$Z = \frac{X - \mu}{\sigma} = \frac{5\ ,5 - 50}{5} = 1,$$

El área es igual a 0,4713

0,5000 – 0,4713 = 0,0287



$$P_{(x \geq 60)} = 2,87\%$$

Se observa que la probabilidad de que se obtengan 60 caras o más, al lanzar 100 monedas, es de 2,87%, lo que se considera como una probabilidad muy pequeña.

El resultado anterior de haber obtenido un número mayor o menor de caras, nos lleva a pensar sobre lo sucedido y para ello tenemos dos explicaciones:

- La moneda o monedas utilizadas en este experimento es o son fabricadas perfectamente, por lo tanto el resultado obtenido es correcto, pero ha sucedido algo raro, pues esperábamos igual número de caras y sellos.
- La moneda tiene fallas de fabricación, por lo que se explica que una de las caras puede aparecer con mayor frecuencia. Estamos diciendo que la moneda no está equilibrada, es decir, está cargada en uno de los lados.

Con las anteriores explicaciones estamos formulando dos clases de hipótesis.

La hipótesis puede ser formulada con el fin de rechazarla de acuerdo con el análisis estadístico. Esta clase de hipótesis se denomina *hipótesis nula* y se representa por o. Se tiene también la *hipótesis alternativa* por $_a$. En nuestro ejemplo inicial del lanzamiento de 100 monedas, la hipótesis nula corresponde a la legitimidad de la moneda, $_0$: $\mu = 50$ obtención de 50 caras, ya que $p = 1/2$ y $\mu = np = 100\ (1/2) = 50$, la hipótesis alternativa será $_a$: $\mu \neq 50$ (que el número de caras sea mayor o menor de 50), es decir, la moneda no es legítima o está cargada.

H_0: Es considerada como la *hipótesis nula*, ya que hace referencia al valor del parámetro que se quiere probar como verdadero.

H_1: Corresponde a la *hipótesis alternativa* o falsa, estableciendo que el parámetro puede ser mayor, menor o igual, de acuerdo con la propuesta hecha en la hipótesis nula.

Tipo de error

En la decisión anterior de aceptar o rechazar la moneda (o monedas) de acuerdo con el resultado obtenido, puede conducir a error. Se consideran dos tipos de error:

- Error tipo II — *Aceptar* la hipótesis nula (H*a*) cuando se ha debido rechazar. En el ejercicio que estamos desarrollando, sería: "Aceptar la moneda como correcta, cuando en verdad no lo es".
- Error tipo I — *Rechazar* la hipótesis nula (H*a*) cuando se ha debido aceptar. "Rechazar la moneda como incorrecta, cuando en verdad está equilibrada".

Existen por lo tanto, dos posibles decisiones: *aceptar* o *rechazar* la hipótesis que a la vez puede ser *cierta* o *falsa*. Veámoslo en la tabla No. 8.1

Tabla 8.1 Tipo de error

Decisiones		Verdadera	Falsa
	Aceptar		
	Rechazar		

Los dos tipos de error son inherentes al proceso de la prueba de significación y la probabilidad de cometer error será igual al nivel de significación.

Decisiones en cuanto a los tipos de error

- Si se acepta una hipótesis verdadera, la decisión es correcta.
- Si se acepta una hipótesis falsa, cometemos el error de tipo II.
- Si rechazamos una hipótesis verdadera, cometeremos error de tipo I.
- Si rechazamos una hipótesis falsa, la decisión es correcta.

Ejemplo 2. Lo más indicado y utilizado para comprender mejor lo anterior, es el siguiente ejemplo: supongamos que se detiene a una persona por robo y se le envía al juez quien podrá declararlo inocente o culpable. Al juez se le presentan los pro y los contra y con base en toda la información, decide dejarlo libre o condenarlo. El juez, no sabrá si hubo error en su decisión, sólo lo podrá saber la persona que ha sido juzgada.

Tabla 2.1

		Persona juzgada	
Decisiones	Del juez	Inocente	Culpable
	Libre		
	Condenado		

Si lo deja libre y el acusado es culpable se estará cometiendo un error (tipo II), si por el contrario,

lo condena siendo inocente, también se estará cometiendo error (tipo I); hasta ahí va nuestra decisión: *aceptar* o *rechazar* pero no sabemos exactamente el error cometido, sólo sabemos que se puede cometer. Un comentario adicional: para la sociedad ¿Cuál de las decisiones es la mejor? ¿Qué opina usted?

Ejemplo 3. Supongamos con base en una muestra aleatoria, que deseamos probar la hipótesis de que el precio medio de un apartamento con tres dormitorios y dos baños, en un barrio de la ciudad es de $180.000.000. Explique ¿en qué condiciones cometeríamos un error y de qué tipo?

o :

- Error de tipo II, si aceptamos que el precio del apartamento descrito es de $180.000.000 de pesos, cuando no lo es.
- Error tipo I, si rechazamos el precio de $180.000.000, cuando es cierto.

El que un error de *tipo I* o *tipo II* se cometa, depende de cómo sea formulada la hipótesis. Veámoslo mediante el siguiente ejemplo.

Ejemplo 4. Supongamos que el decano de una universidad desea contratar los servicios del profesor Pedro Pérez, para ello es sometido a una entrevista, bajo las siguientes hipótesis:

a) El profesor Pérez es competente para desarrollar la labor.

b) El profesor Pérez no es competente para esa labor ¿En qué condiciones el decano cometerá errores de tipo I y de tipo II, en cada caso?

o :

a) Error tipo I: decidir que el señor Pérez no es competente si realmente lo es.

Error tipo II: decidir que el señor Pérez es competente si realmente no lo es.

b) Error tipo I: decidir que el señor Pérez es competente cuando realmente no lo es.

Error tipo II: decidir que el señor Pérez no es competente cuando realmente lo es.

Nota: Se aconseja realizar, las veces que sean necesarias, los ejercicios que se dan al principio como ejemplos y los primeros ejercicios dados para resolver, cuya solución está en el SIL.

ipótesis nula y alternativa

Se ha dicho que una *hipótesis estadística* es un supuesto, concerniente a los parámetros o a la forma de la distribución de probabilidad, correspondiente a una o más poblaciones dadas. En otras palabras, se resume diciendo que corresponde a un enunciado acerca de un *valor estadístico* (parámetro) *poblacional*.

La hipótesis se debe formular en forma correcta o lógica y debe ser enunciada antes de obtener los datos muestrales. Son ejemplos de hipótesis estadística:

- Que el promedio de calificación que tendrán los alumnos en mi curso de estadística, sea superior a 4,0.
- El 90% de mis estudiantes aprobarán el curso.
- 5% de las unidades producidas por una máquina, serán defectuosas.

La *hipótesis nula* es aquélla por medio de la cual se hace una afirmación sobre un parámetro, que se va a constatar con el resultado muestral. Cuando el fabricante dice que su producto tiene una duración de 5.000 horas, se le considera como hipótesis nula, pues es lo que se quiere probar.

La *hipótesis alternativa*, es toda aquella hipótesis que difiere de la hipótesis nula, es decir, ofrece una alternativa, afirmando que la hipótesis nula es falsa. Para el ejemplo, se podrá decir que la hipótesis alternativa podrá ser:

- El fabricante ha *exagerado* la duración de su producto *(prueba unilateral* a la *izquierda*).
- El producto tiene una duración *superior* al señalado por el fabricante (prueba unilateral a la *derecha*). La duración del producto no es la señalada por el fabricante (prueba bilateral).

Prueba unilateral y bilateral

La *prueba de hipótesis unilateral* es aquélla en la cual la zona de rechazo o zona crítica está completamente comprendida en uno de los extremos de la distribución. La prueba es *unilateral a la derecha* (de la curva); cuando la hipótesis alternativa de lo que se quiere probar, hace mención por ejemplo a que los salarios que paga una empresa son *mayores*; que la calidad de un producto es *superior*; que el rendimiento académico es *mejor*, etc. Si por el contrario, la hipótesis alternativa se refiere a que los salarios son *inferiores*, que el producto es de *menor* calidad, que el rendimiento académico es *bajo* o que el productor *exagera* la duración del producto, etc., corresponderá a una *prueba unilateral a la izquierda*.

En el caso de que la prueba comprenda áreas o zonas de rechazo en ambos extremos de la distribución, se dice que la *prueba es bilateral* o sea que la hipótesis alternativa *es diferente*; por lo tanto se omiten los términos: *superior*, *mayor*, *mejor*, *inferior*, *bajo*, *menor*, etc.

Nivel de significación y puntos críticos

Se entiende por *nivel de significación*, la máxima probabilidad de que se especifique con el fin de hacer mínimo el primer tipo de error. Generalmente, esta probabilidad se fija antes de escoger la muestra.

El nivel de significación se simboliza por alfa ($\propto$) siendo generalmente del 1%, 5% o 10%, pero se puede usar cualquier nivel, dependiendo del tipo de investigación que se adelante. Existe la costumbre de trabajar con el nivel de 0,05 o sea del 5%, especialmente cuando el enunciado del problema no lo da.

Cuando se trabaja con un nivel del 5%, el resultado es *significativo*; si se emplea el 1 %, el resultado es *altamente significativo*, y si es del 10%, se considera *poco significativo.*

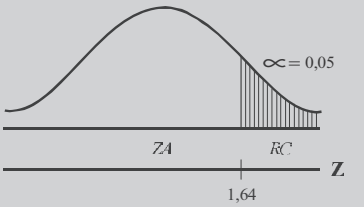
El valor del *nivel de significación* corresponde a un área bajo la curva de probabilidad o normal, denominada *región crítica* o *zona de rechazo*. Se tendrán casos en que la región crítica esté situada únicamente la derecha de la curva, se dirá que se trata de una dócima unilateral a la derecha. Si se sitúa a la izquierda, será una dócima unilateral hacia la izquierda. En caso de considerar dos regiones críticas, se hablará de una dócima bilateral.

En las dócimas unilaterales se tomará el valor total de alfa; para las dócimas bilaterales alfa se dividirá por dos. La región no sombreada o no cubierta por el nivel de significación se denominará *zona de aceptación* o de *no rechazo*.

Figura 8.1 Regiones críticas

a) Dócima *unilateral* hacia la *derecha,* con $\propto = 0{,}05$

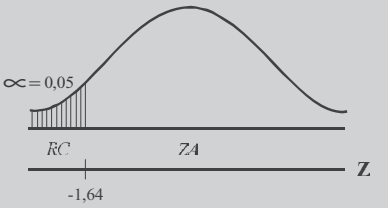
$A(0{,}4500) \rightarrow Z = 1{,}64$ ó 1,65



b) Dócima *unilateral* hacia la *izquierda,* con $\propto = 0{,}05$

$0{,}5000 - 0{,}0500 = 0{,}4500$

$A(0{,}4500) \rightarrow Z = -1{,}64$ ó $-1{,}65$



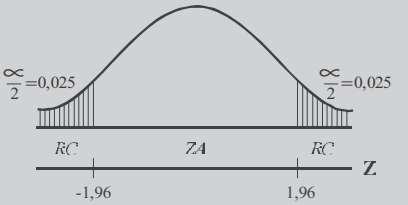
c) Dócima *Bilateral,* con $\propto = 0{,}05$

$0{,}5000 - 0{,}0250 = 0{,}4750$

$A(0{,}4750) \rightarrow Z = 1{,}96$

$Z_s = 1{,}96$

$Z_i = -1{,}96$



ZA = Zona de aceptación
RC = Región crítica
$\propto$ = Nivel de significación
Z = Punto crítico

Procedimiento a seguir en las pruebas de ipótesis

Se presenta en primer lugar una síntesis de los pasos a seguir, luego se amplía cada uno de ellos para que se tenga una mejor visión del procedimiento y finalmente, se dan ejemplos para ayudar a resolver las inquietudes que se hayan creado.

:

1. Formular la hipótesis nula y la alternativa.
2. Seleccionar el nivel de significación.
3. Conocer o estimar la varianza.
4. Determinar la técnica y la prueba estadística.
5. Determinar los valores críticos y sus regiones de rechazo.
6. Calcular los datos muestrales, utilizando las fórmulas correspondientes.
7. Tomar la decisión estadística, de aceptar o rechazar.

Ampliemos el resumen anterior; para ello se recomienda tener presente, además de los siguientes pasos, las alternativas que se dan para cada situación especial.

1. *Establecer las hipótesis*:

H_0 : Hipótesis nula

H_a : Hipótesis alternativa

$\begin{bmatrix} Ho \\ Hl \end{bmatrix}$

a) En el caso de la moneda (normal) y en distribuciones de medias muestrales se podrían presentar las hipótesis de las siguientes formas:

$H_0 \ \mu = 50$ $H_a \ \mu \neq 50$ *(dócima bilateral)*	$H_0 \ \mu = 50$ $H_a \ \mu < 50$ *(dócima unilateral a la izquierda)*	$H_0 \ \mu = 50$ $H_a \ \mu > 50$ *(dócima unilateral a la derecha)*

b) En el caso de una distribución de diferencias entre *dos medias muestrales*, puede plantearse así:

$H_0 \ \mu_x = \mu_y$ $H_a \ \mu_x \neq \mu_y$ *(dócima bilateral)*	$H_0 \ \mu_x = \mu_y$ $H_a \ \mu_x > \mu_y$ *(dócima unilateral a la derecha)*	$H_0 \ \mu_x = \mu_y$ $H_a \ \mu_x < \mu_y$ *(dócima unilateral a la izquierda)*

c) En las *proporciones* se escribirá para cada caso, así:

$H_0 \ \mu_p = 0{,}50$ $H_a \ \mu_p \neq 0{,}50$	$H_0 \ \mu_p = 0{,}50$ $H_a \ \mu_p > 0{,}50$	$H_0 \ \mu_p = 0{,}50$ $H_a \ \mu_p < 0{,}50$

d) Diferencias entre *dos proporciones*:

$H_0 \ \mu_{p_1} = \mu_{P_2}$ $H_a \ \mu_{p_1} \neq \mu_{P_2}$	$H_0 \ \mu_{p_1} = \mu_{P_2}$ $H_a \ \mu_{p_1} > \mu_{P_2}$	$H_0 \ \mu_{p_1} = \mu_{P_2}$ $H_a \ \mu_{p_1} < \mu_{P_2}$

2. *Elegir el riesgo*: $\propto = \%$

Los niveles de significación más utilizados son:

a) $\propto = 0{,}05$ o 5%
b) $\propto = 0{,}01$ o 1%
c) $\propto = 0{,}10$ o 10%

3. *Se establecen ciertos supuestos:*

a) La muestra es aleatoria.
b) La población es normal.
c) La varianza poblacional es conocida (en la mayoría de los casos no se conoce, por lo tanto debe ser estimada).

4. Se formula la respectiva variante estadística:

a) *Distribución normal*: $Z = \dfrac{X - \mu}{\sigma}$

b) Distribución de *medias muestrales*:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \text{o} \quad Z = \frac{\bar{x} - \mu}{S / \sqrt{n}} \quad Cuando \quad n > 30$$

c) Distribución de *proporciones muestrales*:

$$Z = \frac{p - P}{\sqrt{\frac{p\,q}{n}}} \quad \text{Siendo} \quad n > 30$$

d) Distribución de diferencias entre *dos medias muestrales*:

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}} \quad \text{o} \quad Z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}}$$

(Siendo n_1 y n_2 mayores que 30)

e) Distribuciones de diferencias entre *dos proporciones muestrales*:

$$Z = \frac{(p_1 - p_2) - (\mu_{p_1} - \mu_{p_2})}{\sqrt{\frac{p_1\,q_1}{n_1} + \frac{p_2\,q_2}{n_2}}} \quad \text{Cuando } n_1 \text{ y } n_2 > 30$$

5. F*ormular los puntos críticos*: Z_i y Z_s

a) Al trabajar con un nivel de significación del 5% de dócima *bilateral*, se tendrá:

$$Z_s = 1{,}96 \quad \text{y} \quad Z_i = -1{,}96$$

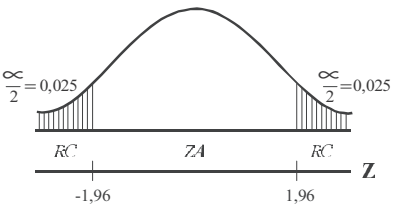
b) Con el mismo nivel de significación y una dócima *unilateral*: $Z_s = 1{,}64$ ó $1{,}65$; si la prueba es hacia la derecha; $Z_i = -1{,}64$ ó $-1{,}65$, si la prueba se hace a la izquierda.

6. *Descripción de la región crítica:*

(Bilateral con un $\propto = 0{,}05$)

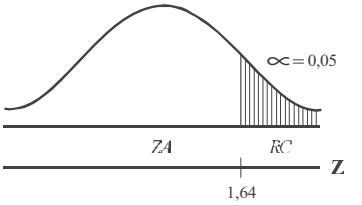
$Z_s \geq 1{,}96$

$Z_i \leq -1{,}96$



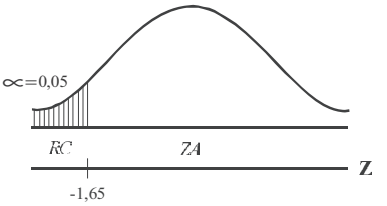
(Dócima unilateral a la derecha)

$Z_s < 1{,}64$ o $1{,}65$



(Dócima unilateral a la izquierda)

$Z_i < -1,64$ ó $-1,65$



7. Se le dan valores a la variante estadística con el fin de obtener el valor de Z.

   En el ejemplo de la moneda, se tiene que: $Z = \frac{5\ ,5-50}{5} = 1,$

8. *Adoptar una decisión*, se acepta o se rechaza la hipótesis nula, de acuerdo al nivel de significación dado.

Siguiendo el ejemplo 1, donde $Z = 1,9$, trabajando con una dócima bilateral, además, con un nivel de significación del 5%, encontramos que el 1,9 se sitúa en la zona de aceptación, por lo tanto aceptamos la hipótesis nula ($H_0 : \mu = 50$), es decir, la moneda es legítima o correcta (no está cargada). En otras palabras la diferencia que hay entre el valor esperado (50 caras) y el obtenido en el experimento (60 caras) no es significativa.

DISTRIBUCIÓN DE MEDIAS MUESTRALES ( )

Veamos algunos ejemplos, cuando se conoce la *varianza poblacional* y cuando *se desconoce*. Como orientación en este último caso, por lo general, después de señalar el tamaño de la muestra, y su media, vendrá la identificación de la desviación típica, evitando de esta manera que se confunda la desviación o la varianza muestral con la poblacional.

Prueba cuando se conoce la varianza poblacional

En los pasos descritos anteriormente, específicamente en el número 3, se establecerá que la varianza poblacional (σ) supuestamente se conoce. Cuando no es conocida, deberá ser sustituida por la varianza muestral (s^2), siempre y cuando la muestra sea grande, es decir, mayor a 30. En los dos casos el proceso de cálculo es igual, sólo se sustituye a σ por s.

En un ejercicio, cuando tengamos duda sobre la varianza, si es poblacional o muestral, debemos observar que después de *mencionar el tamaño de la muestra*, lo que sigue son resultados muestrales, tal como lo presentamos a continuación.

a) $\mu = 360$	$\sigma = 25$	$n = 25$	$\rightarrow$	$\bar{x} = 372$
b) $\mu = 360$	$n = 36$	$\rightarrow$ $\bar{x} = 372$		$s = 25$

Ejemplo 1. Un inspector de calidad investiga las acusaciones contra una embotelladora por su deficiente llenado que debe ser, en promedio, de 32,5 onzas. Para ello toma una muestra de 60 botellas, encontrando que el contenido medio es de 31,9 onzas de líquido. Se sabe que la máquina embotelladora debe producir un llenado con una desviación típica de 3,6 onzas. ¿Puede el inspector llegar a la conclusión, a un nivel de significación del 5%, que se están llenando las botellas por *debajo* de su especificación de contenido?

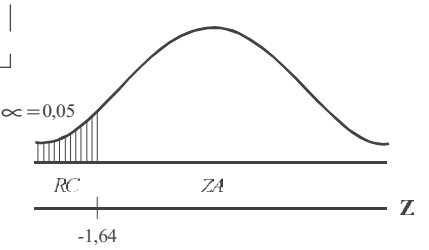
o **:**

$\mu = 32{,}5 \qquad \sigma = 3{,}6 \qquad n = 60 \qquad \bar{x} = 31{,}9$ $\qquad Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

a) $H_0 : \mu = 32{,}5$ $\quad H_a = \mu = 32{,}5$
$H_a : \mu < 32{,}5$ $\quad H_1 = \mu < 32{,}5$

b) $\propto = 0{,}05$

c) $\sigma = 3{,}6$



$$Z = \frac{31{,}9 - 32{,}5}{3{,}6/\sqrt{60}} = -1{,}29$$

Como -1,29 se sitúa en la zona de aceptación, es válida la hipótesis nula, lo cual significa que el inspector no debe llegar a la conclusión, que se está llenando y vendiendo un producto por debajo de su especificación, al nivel del 5%.

Ejemplo 2. Un proceso está programado para empacar la cantidad media, de una libra (16 onzas) de café. Se toma una muestra aleatoria de 36 paquetes; resulta una media de 14,2 onzas y desviación típica de 5,3 onzas. Al nivel del 5%, ¿se podrá afirmar que *no se está cumpliendo* con lo indicado en el empaque?

o **:**

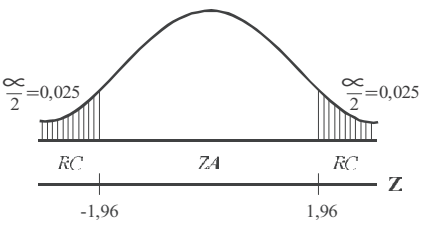
Observe que no se está diciendo que el peso está por debajo de lo establecido, ya que la media podría ser cualquier valor. En este caso la prueba es *bilateral*.

$\mu = 16 \qquad n = 36 \qquad \bar{x} = 14{,}2 \qquad s = 5{,}3$

1) $H_0 : \mu = 16$
$H_a : \mu \neq 16$

2) $\propto = 0{,}05$

3) $s = 5{,}3$



$$Z = \frac{14{,}2 - 16}{5{,}3/\sqrt{36}} = -2{,}04$$

Al nivel del 5%, se podrá afirmar que no se está cumpliendo con lo indicado por la fábrica. Se puede ver que -2,04 se ubica en la región crítica, por lo tanto se estará rechazando la hipótesis nula y aceptando la hipótesis alternativa.

Ejemplo 3. En una oficina de remesas se tienen 500 paquetes con diferentes pesos en kilogramos para ser remitidas a diferentes partes del país. Para conocer el peso promedio de los paquetes se selecciónó una muestra de 16 de ellas, con los siguientes resultados:

62	43	60	49	72	56	45	46	37	56	41	43	36	45	56	49

Se sabe que la desviación estándar poblacional es de 10, pero es desconocida la media poblacional (μ = 50 verdadera). Cometiendo un riesgo alfa (nivel de significación) del 5%, dócimar ó probar la hipótesis de que la media poblacional sea igual:

a) 40 b) 49 c) 50 d) 51 y e) 60.

Observación: se trata de un ejercicio relativo a dócimas o pruebas bilaterales.

Como se conoce la media poblacional verdadera, se puede comprobar si la decisión adoptada es correcta o no. En efecto, para cada uno de los 5 casos anteriores establecer si la decisión es correcta y en caso contrario indicar el tipo de error cometido.

La intención de este ejercicio es mostrar que a medida que la media poblacional hipotética se aproxima a la media poblacional verdadera se comete error tipo II.

o :

a) 1) $H_0 : \mu = 40$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$ $\bar{x} = \frac{796}{16} = 49{,}75$

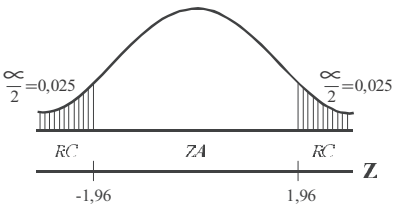
$H_a : \mu \neq 40$

4) $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{49{,}75 - 40}{10/\sqrt{16}} = 3{,}9$

$A(0{,}4750) \rightarrow Z = 1{,}96$

$Z_s \geq 1{,}96$

$Z_i \leq -1{,}96$



Z = 3,9 cae en la región critica (Z > 1,96) por tal motivo se rechaza la hipótesis de que μ = 40, sabiendo que la media verdadera μ = 50, estamos rechazando algo falso, por lo tanto no hay error, al nivel del 5%.

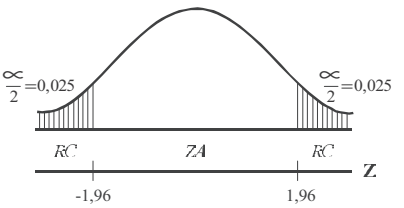
b) 1) $H_0 : \mu = 49$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$

$H_a : \mu \neq 49$

4) $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{49{,}75 - 49}{10/\sqrt{16}} = 0{,}30$

$0{,}5000 - 0{,}0250 = 0{,}4750$

$A(0{,}4750) \rightarrow Z = 1{,}96$



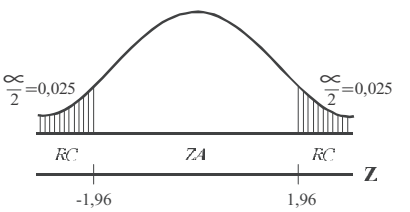
Siendo Z = 0,30 cae en la zona de aceptación, es decir, que aceptamos que μ = 49, como sabemos que la media poblacional verdadera es 50, estamos aceptando algo falso (error de tipo II), al nivel del 5%.

c) 1) $H_0 : \mu = 50$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$

$H_a : \mu \neq 50$

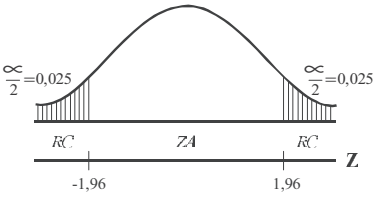
4) $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{49{,}75 - 50}{10/\sqrt{16}} = -0{,}10$

Z = –0,10 se ubica en la zona de aceptación. No hay error, ya que aceptamos algo verdadero (μ = 50)



d) 1) $H_0 : \mu = 51$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$
$H_a : \mu \neq 51$

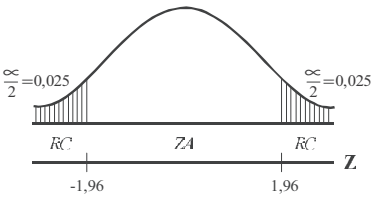
4) $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{49{,}75 - 51}{10 / \sqrt{16}} = -0{,}50$



Z = -0,50 se ubica en la región de aceptación. El error es de tipo II, ya que aceptamos algo falso, al nivel del 5%.

e) 1) $H_0 : \mu = 60$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$
$H_a : \mu \neq 60$

4) $Z = \dfrac{49{,}75 - 60}{10 / \sqrt{16}} = -4{,}10$



No hay error, se rechaza algo falso, al nivel del 5%.

Ejemplo 4. De la "población normal" mencionada en el ejercicio anterior, se extrajo una muestra aleatoria de 25 observaciones cuya media muestral es de 54. Empleando los niveles de significación del 1% y del 5%, docimar ó probar la hipótesis donde la media poblacional sea: a) 50 b) 49.

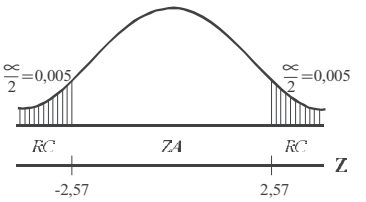
Observación: se trata de cuatro dócimas. En cada una de ellas calcular el valor de Z y adoptar la decisión.

o :

❑ Cuando ∝ = 0,01

a) 1) $H_0 : \mu = 50$ 2) $\propto = 0{,}01$ 3) $\sigma = 10$
$H_a : \mu \neq 50$

4) $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{54 - 50}{10 / \sqrt{25}} = 2$

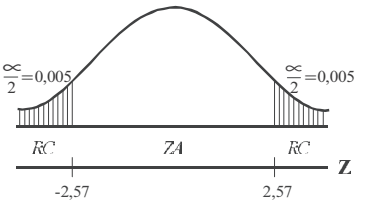


$0{,}5000 - 0{,}005 = 0{,}4950$
$A\ (0{,}4950 \rightarrow Z = 2{,}57)$

$Z = 2$ El 2 se ubica en la zona de aceptación. No hay error, al nivel del 1%.

b) 1) $H_0 : \mu = 49$ 2) $\propto = 0{,}01$ 3) $\sigma = 10$
$H_a : \mu \neq 49$

4) $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \dfrac{54 - 49}{10 / \sqrt{25}} = 2{,}5$

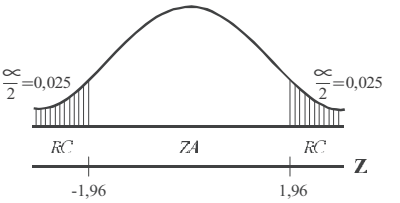


Hay error de tipo II, se acepta algo falso, al nivel del 1%.

❑ Ahora, cuando $\propto = 0{,}05$

a) 1) $H_0 : \mu = 50$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$
$H_a : \mu \neq 50$

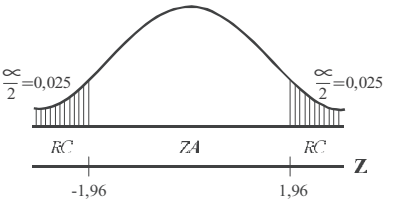
4) $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{54 - 50}{10/\sqrt{25}} = 2$



$Z = 2$ Se ubica en la zona de rechazo. Se comete un error de tipo I por rechazar algo verdadero, al nivel del 5%.

b) 1) $H_0 : \mu = 49$ 2) $\propto = 0{,}05$ 3) $\sigma = 10$
$H_a : \mu \neq 49$

4) $Z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{54 - 49}{10/\sqrt{25}} = 2{,}5$



$Z = 2{,}5$ Se ubica en la zona rechazo. No hay error, al nivel del 5%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

1. Dado $\bar{x} = 82$ $\sigma = 15$ $n = 25$ Docimar la hipótesis: $\mu = 86$

2. Dado $\bar{x} = 82$ $S = 15$ $n = 100$ Docimar la hipótesis: $\mu = 86$

3. Muchos años de experiencia en un examen de ingreso a la universidad en inglés, arroja una calificación promedio de 64, con una desviación estándar de 8. Todos los estudiantes de cierta ciudad, en la cual existen 64, han obtenido una calificación promedio de 68. ¿Puede tenerse la certeza de que los estudiantes de esta ciudad son superiores en inglés?

4. Docimar la hipótesis, de que la distancia media requerida para poder detener un automóvil que va a 20 km/h, es de 25 metros. Con base en una muestra de 100 conductores se obtiene que la distancia media es $\bar{x} = 27{,}3$ metros, con una desviación estándar de $s = 6{,}1$ metros. Utilizar un nivel de significación del 5%.

5. Dado $\bar{x} = 86$ $s = 16$ y $n = 100$, docimar la hipótesis $\mu = 80$ frente a la alternativa bilateral $\mu \neq 80$, al nivel de significación del 5%.

6. Cuatrocientos estudiantes, elegidos aleatoriamente, se someten a un "test" de rendimiento, obteniéndose los siguientes resultados: $\bar{x} = 76$ y $s = 16$, con base en esta información docimar la hipótesis $\mu = 74$ frente a la alternativa $\mu \neq 74$, al nivel de significación del 1%.

Nota: cuando se conoce σ no debe preocuparse si es peque a

7. Dado $\bar{x} = 23{,}5$; $\sigma = 3{,}2$ y $n = 25$, docimar la hipótesis $\mu = 22$ frente a la alternativa $\mu \neq 22$, al nivel de significación del 5%.

8. Una encuesta revela que los 100 autos particulares, que constituyen una muestra aleatoria, se condujeron a un promedio de 12.500 km durante un año, con una desviación estándar de 2.400 km. Con base en esta información, docimar la hipótesis donde, en promedio, los autos particulares se condujeron a 12.000 km. durante un año, frente a la alternativa de que el promedio sea superior. Utilizar el nivel de significación del 5%.

9. Una muestra aleatoria de zapatos ($n = 40$) usados por los soldados en campaña en un desierto, revela una vida media de 1,08 años, con una desviación estándar de 0,5 años. Se sabe que en condiciones normales dichos zapatos tienen una vida media de 1,28 años. Al nivel de significación del 5%, ¿hay razón para sostener que la disminución de la vida media de los zapatos se debe a su uso en el desierto?

10. Un fabricante de cuerdas, ha establecido con base en una experiencia de muchos años, que las cuerdas tienen una fuerza de ruptura de 15,9 libras, con una desviación estándar de 2,3 libras. Se efectúa un cambio en el proceso de fabricación, y se obtiene una muestra de 64 artículos cuya fuerza media de ruptura es de 15,0 libras, con una desviación estándar de 2,2 libras. ¿Debe considerarse que el nuevo proceso tiene un efecto significativamente negativo respecto a la resistencia de las cuerdas?

11. Un fabricante sostiene que sus autos consumen en promedio 5,50 galones cada 100 kilómetros. Un vendedor de la compañía comprueba el consumo de gasolina de 35 autos y encuentra que el consumo medio de ese grupo es de 5,65 galones cada 100 kilómetros con una desviación estándar de 0,35 galones. Con estos datos, ¿puede dudarse de lo sustentado por la compañía (Nivel del 1%)?

12. Los salarios diarios en un sector de la industria están distribuidos normalmente con una media de $23.200 y una desviación estándar de $2.500. Si una empresa del sector, que cuenta con 40 obreros, paga, en promedio, $22.200, ¿puede decirse que esta compañía paga salarios inferiores, al nivel de la significación del 1%?

13. Un investigador asegura que el gasto promedio semanal de las familias de una determinada región en alimentación es de $81.000. Con base en los siguientes resultados para la variable X (gastos semanales). $\bar{x} = 80.600$ $s = 1.100$ $n = 100$

    ¿Con el nivel de significación del 5%, se puede aceptar lo que dice el investigador?

14. Un distribuidor de botas especiales de trabajo, le garantiza al jefe de una empresa que el promedio de duración de las botas es de 8 meses, con una desviación típica de mes y medio. El cliente de la empresa decide comprar 36 pares de botas, que en promedio duran 8 meses 10 días. ¿Tiene razón el fabricante, con un nivel de significación del 5%?

15. Una compañía que vende camarón congelado, imprime sobre el envase: “contiene 14 onzas”. Una muestra de 25 envases da una media de 13,83 onzas. De las experiencias anteriores se deduce que la población del peso de los envases tiene una desviación típica de 0,5 onzas. Utilizando un nivel de significación del 5%, ¿se puede aceptar lo ofrecido por la empresa?

16. En una ciudad se usan miles de bombillas cada año. La marca que se ha usado en el pasado tiene duración media de 1.000 horas, con una desviación típica de 100 horas. Se ofrece una marca a un

precio mucho más bajo del que están pagando por la antigua. Deciden que se cambiarán a la nueva marca, a menos que se demuestre que ésta no tiene una duración media inferior a la de la bombilla anterior; el nivel de significación es del 5%. Como consecuencia, se prueba una muestra de 100 bombillas de la marca, obteniendo una duración media de 985 horas. Suponiendo que la desviación típica de la nueva marca es la misma que la antigua, ¿qué conclusión debería obtenerse?

17. Un fabricante de bombillas de destello para fotografía, asegura que la duración media de su producto pasa de las 40 horas. Una compañía desea comprar un lote muy grande de dicho artículo, si la aseveración es cierta. Se prueba una muestra aleatoria de 36 bombillas y se halla que la media muestral es de 46 horas. Si la población de bombillas tiene una desviación típica de 9 horas, ¿es posible que se compren las lámparas?

18. Una máquina produce en promedio 12 libras de nitrato por minuto. Al adicionar un compuesto químico a 60 máquinas, éstas produjeron un promedio de 15 libras de nitrato por minuto y una $s = 5$ ¿Se puede concluir que la solución aumenta la productividad? ($\propto = 1\%$)

19. Un cierto tipo de fusible está diseñado para fundirse cuando la corriente llega a 20 amperios. Se toma una muestra de 36 fusibles de un lote de 500 y se encuentra que el punto promedio de fusión es de 20,8 amperios, con desviación estándar de 1,5 amperios. Utilizando una dócima bilateral, ¿a qué conclusiones se puede llegar respecto a las especificaciones del lote, a un nivel de significación del 1%?

20. Un fabricante de aparatos de calefacción, tiene un proveedor que le suministra ciertos componentes que tienen como especificaciones una resistencia de 400°C al calor. El fabricante realiza una selección al azar de 64 de dichos componentes y comprueba que su resistencia media al calor es de 395°C, con desviación estándar de 20°C. ¿Puede llegar a la conclusión de que su proveedor no sostiene las especificaciones acordadas, a un nivel de significación de 5%?

21. Un test de psicología tenía una puntuación media de 78 puntos y una desviación típica de 6. Se seleccionó al azar un grupo de 16 estudiantes siendo su calificación de 74 puntos ¿Puede afirmarse al nivel de significación del 1% que el resultado de este grupo fue inferior?

22. Una muestra de 200 artículos producidos por una máquina, que debe tener como especificación un diámetro de 3,6 cmts, revela un diámetro promedio de 3,62 cmts, con desviación estándar de 0,21 cm. ¿Podría afirmarse que el anterior resultado se ajusta a las especificaciones de producción?

23. Un proceso está programado para empaquetar la cantidad media de una libra de azúcar. Se toma una muestra aleatoria de 36 paquetes y se encuentra una media de 13 onzas y desviación típica de 8 onzas. Al nivel del 5% se podrá afirmar:
    a) ¿que se está vendiendo un producto por debajo de su peso?
    b) ¿si el promedio verdadero es de una libra, en la prueba anterior, se cometió un error de tipo II?

24. Una compañía afirma que el tiempo necesario para fabricar un artículo es de 53 minutos, con una varianza de 1,35 horas. Se toma una muestra de 128 artículos, con un tiempo promedio de 56 minutos.
    a) Al nivel del 5% ¿se podrá afirmar que el producto requiere un tiempo mayor de fabricación que lo afirmado por la compañía?
    b) ¿Si el tiempo real de fabricación es de 50 minutos, en la prueba de significación se cometió un error de tipo I?

25. Un gimnasio recién inaugurado en la capital, invita a su afiliación argumentando una reducción de peso, al menos de 4,6 kilos. Una muestra aleatoria de 34 personas revela que el promedio de reducción de peso es de 4,1 kilos, con desviación típica de 1,8 kilos. A un nivel del 1%, ¿se puede creer lo tan anunciado por el gimnasio?

26. Un concesionario de automóviles promociona un nuevo modelo de vehículo, ofreciendo economía en el combustible; asegura un promedio de 50 kilómetros por galón de gasolina. Se tomó una muestra de 35 vehículos de los 500 que se han vendido y se encontró que el promedio es de 43,8 kilómetros por galón, con desviación típica de 15 kilómetros por galón. Al nivel del 2%, ¿se podrá decir que el concesionario ha exagerado?

27. Una muestra aleatoria de 60 obreros de un sector de la industria de la construcción, encontró que el promedio de edad es de 24 años. Sin embargo un alto ejecutivo asegura que el promedio de edad es superior a los 22 años, con una desviación estándar de 8 años, ¿Se puede aceptar la afirmación del ejecutivo?

28. Una compañía promociona un nuevo ambientador asegurando que mantiene el ambiente agradable por lo menos durante 8 horas. Un análisis realizado en 20 ambientes, seleccionados aleatoriamente, indicó un promedio de 8 horas y media. ¿Se puede aceptar la anterior aseveración? Se sabe por experiencia que la desviación típica es de una hora y cuarenta y cinco minutos.

29. Una máquina produce en promedio 650 libras de nitrato por hora. Al adicionar un compuesto químico a 40 máquinas, éstas produjeron un promedio de 700 libras de nitrato por hora y una varianza de 12.960 libras ¿Se puede concluir que la solución aumenta la producción de nitrato? (nivel de 1%).

30. Una muestra aleatoria de 100 profesores universitarios de la capital, dio los siguientes resultados:

$$\sum x_i = 4.000 \qquad \sum (x_i - \bar{x})^2 = 9.900$$

    a) La autoridad educativa afirma que el promedio de edad de un profesor es de 43 años, promedio que consideramos diferente. Pruebe dicha hipótesis al nivel del 5%.
    b) Si en verdad el promedio en la población es de 39 años, ¿qué tipo de error cometeríamos?

31. Un profesor examina a su curso, sabe por experiencia que proporciona un rendimiento con calificación de 78. Su curso actual es de 35 alumnos, tomada como muestra; obtiene una media de 82 y desviación típica de 21. ¿Acierta al suponer que es un curso superior? (Nivel del 1%).

Nota: los problemas que aparecen **al final de este capítulo, permiten** repasar lo aprendido en las diferentes unidades del capítulo 8.

DISTRIBUCIÓN DE PROPORCIONES MUESTRALES

Los procedimientos de decisión, aplicados a las *proporciones* son similares a los ya indicados para las medias muestrales; salvo que por lo general, la desviación típica y por ende el error estándar de la proporción, se calcula con datos obtenidos en la muestra, con más de 30 elementos, ya que estamos trabajando con muestras grandes.

Veamos su aplicación con algunos ejemplos, luego se tratará de resolver otros ejercicios, con el deseo de cubrir algunas situaciones particulares que se presentan en esta distribución.

Nota: La **varianza** de la proporción se calculará con valores muestrales

Ejemplo 1. Por estadísticas que se tienen, se ha podido establecer que por lo menos el 40% de los jóvenes toman regularmente Coca-Cola, cuando tienen sed. Una muestra aleatoria de 450 jóvenes reveló que 200 de ellos solían tomar dicha bebida, cuando tenían sed. ¿Cuál podría ser su conclusión al nivel del 1%, acerca de lo que muestran las estadísticas?

o **:**

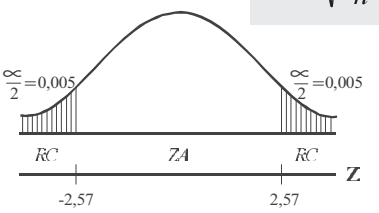
$p = 200/450 = 0{,}44 = 44\%$ $\quad n = 450$ $\quad q = \frac{250}{450} = 0{,}56$

$$Z = \frac{p - P}{\sqrt{\frac{pq}{n}}}$$

1) $H_0 : P = 0{,}40$ $\quad \boxed{H_0 : P = 0{,}40}$
$H_a : P \neq 0{,}40$ $\quad \boxed{H_1 : P \neq 0{,}40}$

2) $\propto = 0{,}01$ $\quad$ 3) $s_p = \sqrt{pq}$

4) $Z = \dfrac{0{,}44 - 0{,}40}{\sqrt{\dfrac{0{,}44(0{,}56)}{450}}} = 1{,}71$



Z = 1,71 se ubica en la zona de aceptación, luego al nivel del 5% la conclusión a que se llega, es la de aceptar el 40% que arrojan las estadísticas; por lo tanto, no hay razón para rechazarlas.

Nota: En todos estos ejercicios siempre se tomará la proporción en la **muestra,** para el cálculo de la v**arianza**

Ejemplo 2. Un gerente de una compañía afirma, que el porcentaje de atrasos en las horas de llegada al trabajo cobija al 25% de sus empleados. Solicita al jefe de personal la revisión de 40 tarjetas marcadas, con las horas de llegada, en la quincena y encuentra que 8 han llegado tarde. Al nivel del 5%, ¿hay razón para concluir que el gerente de la compañía está exagerando?

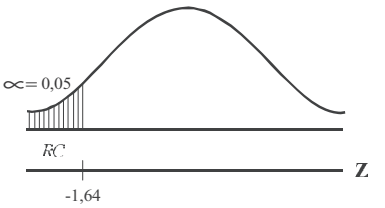
o **:**

$p = \frac{8}{40} = 0{,}20$ $\quad n = 40$

1) $H_0 : P = 0{,}25$ $\quad$ 2) $\propto = 0{,}05$
$H_a : P < 0{,}25$

3) $s_p = \sqrt{0{,}2(0{,}8)} = 0{,}40$

4) $Z = \dfrac{0{,}20 - 0{,}25}{\sqrt{\dfrac{0{,}2(0{,}8)}{40}}} = -0{,}79$



Se ubica Z = -0,79 en la zona de aceptación; por lo tanto podemos concluir, que al nivel del 5% el gerente no está exagerando.

Ejemplo 3. Una empresa, al seleccionar su personal lo somete a un curso de entrenamiento. Por experiencia, el 76% de los aspirantes aprueban el curso. Se efectúan ciertos cambios en el programa para el cual se inscriben 40 y 24 lo aprueban. ¿Podría afirmarse que los cambios introducidos reducen la selección? ($\propto = 1\%$).

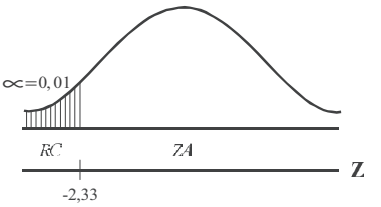
o :

$$p = \frac{24}{40} = 0{,}60$$

1) $H_0 : \mu_{\bar{p}} = 0{,}76$ 2) $\propto = 0{,}01$
$H_a : \mu_{\bar{p}} < 0{,}76$

3) $s_{\bar{p}} = \sqrt{\frac{pq}{n}}$

4) $Z = \frac{p - \mu_{\bar{p}}}{\sqrt{\frac{pq}{n}}} = \frac{0{,}60 - 0{,}76}{\sqrt{\frac{(0{,}6)(0{,}4)}{40}}} = -2{,}07$



Como − 2,07 cae en la región de aceptación, la selección no se reduce con los cambios introducidos, al nivel del 1%.

Ejemplo 4. El ministro de educación de un país, asegura que el 80% de los estudiantes universitarios tienen un ingreso mensual para su sostenimiento, superior a $570.000; usted quiere refutar al ministro con un nivel de confianza del 99% y para hacerlo toma una muestra de 300 estudiantes, encontrando 231 con ingresos mayores a $570.000. ¿Tiene razón el señor ministro?

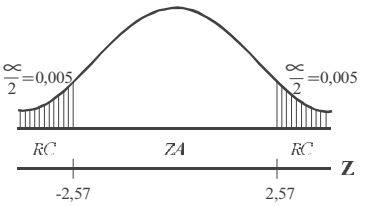
o :

$$p = \frac{231}{300} = 0{,}77 \qquad q = 0{,}23$$

1) $H_0 : P = 0{,}8$ 2) $\propto = 0{,}01$
$H_a : P \neq 0{,}8$

3) $s_{\bar{p}} = \sqrt{\frac{(0{,}77)(0{,}23)}{300}} = \sqrt{\frac{0{,}177}{300}} = 0{,}024$

4) $Z = \frac{0{,}77 - 0{,}80}{0{,}024} = \frac{-0{,}03}{0{,}024} = -1{,}25$



Se acepta la hipótesis de que el 80% de estudiantes tienen ingreso mensual superior a $570.00, al nivel del 1%.

Nota: la palabra superior (en este caso) no implica prueba unilateral derecha, ya que forma parte de la caracterís tica y no de aqu llo que se quiere probar.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

32. La fracción de artículos defectuosos de cierto proceso supervisado es 0,14. Un proveedor de materia prima ofrece un nuevo producto, asegurando que reduce la fracción de defectuosos. Con las muestras que el proveedor suministra, se hace un ensayo en la producción con el resultado de 48 defectuosos de un total de 360. Contrastar si el proveedor tiene o no razón en la calidad de la nueva materia prima, con un 5% de significación.

33. El fabricante de cierto producto estima tener el 50% del mercado de la categoría de dicho producto. Al realizar un sondeo en una muestra probabílistica de 400 consumidores de la categoría del producto, 180 de ellos indicaron ser consumidores. ¿Es correcta la estimación hecha por el fabricante? (5%).

34. Por evidencia experimental, se sabe que las pruebas de ingreso a la universidad son eficaces en un 80% de los casos, cuando está correctamente elaborada. Se aplica dicha prueba a 400 estudiantes y se obtiene únicamente 300 buenos resultados ¿Puede considerarse este resultado, como evidencia de que la prueba no estuvo bien elaborada? ($\propto = 1\%$).

35. Suponga que se está estudiando la compra de una nueva máquina para fabricar tornillos. Se comprará la máquina si la proporción de tornillos que necesiten rehacerse es igual o menor del 10%. Se examina una muestra de 40 tornillos fabricados por la máquina y necesitan rehacerse 3. Con $\propto = 0{,}05$, ¿se puede concluir que no se satisface la condición exigida?

36. Un cirujano desarrolló una técnica quirúrgica nueva para una enfermedad, en la cual la mortalidad postoperatoria usual es del 20%. Aplica la técnica en 50 pacientes de los cuales mueren 9. ¿Cree usted que la nueva técnica disminuye la mortalidad post-operatoria? ($\propto = 5\%$).

37. Por evidencia experimental, se sabe que cierto tratamiento pediátrico es eficaz en un 80% de los casos, cuando está correctamente suministrado. Se aplica el tratamiento a 400 niños y se obtienen únicamente 300 resultados positivos. ¿Puede considerarse dicho resultado como evidencia de que el tratamiento estuvo bien administrado? (Nivel del 1%).

38. En una muestra de 50 neumáticos, fabricados por cierta compañía, cinco de ellos no cumplieron las normas de calidad. El gerente sostiene que el 12% de sus neumáticos no satisfacen dicha norma. Al nivel del 5%. ¿Se podrá decir que el gerente exagera el porcentaje?

39. El gerente de un supermercado considera que de 50 clientes que realizaron compras a mediodía, siete incluyeron leche en su compra. Realiza una muestra aleatoria a 100 clientes y revela que tan sólo 10 compraron leche. ¿Al nivel 5% se podrá deducir que el número de compradores de leche, al mediodía, es inferior al anotado por el gerente?

40. Una muestra aleatoria de 225 habitantes de apartamentos reveló que 25 poseían perros. ¿Proporcionan estos datos evidencia suficiente como para concluir, que menos del 15% de los habitantes de apartamentos poseen perros? (Nivel del 5%).

41. Una muestra de ventas con descuento, está considerando la compra de una gran partida de discos de un proveedor que afirma que, en promedio, sólo el 2% de los discos tienen fallas. Al examinar 400 de ellos, la firma encuentra 15 imperfectos. ¿Rechazará la empresa esta afirmación del proveedor? (Nivel 5%).

42. Se sabe que el 25% de los alumnos de una facultad, habilitan asignaturas para la aprobación del curso. Si se toma una muestra de 36 estudiantes, de los cuales 8 habilitan, ¿se podrá afirmar que el porcentaje es inferior?

43. En una conferencia de prensa, una alta autoridad anuncia que el 90% de los habitantes adultos del país están en favor de cierto proyecto económico del gobierno. Una muestra aleatoria de 650 adultos indica que 570 están a favor del proyecto. Si usted desea rechazar una hipótesis verdadera no más de una vez en cien, ¿concluiría que la popularidad del proyecto ha sido exagerada por la autoridad?

44. Una compañía desea saber si es válida la información, de que el 52% o más de sus clientes aún poseen los vehículos que compraron hace 8 años. Una muestra aleatoria de 100 compradores seleccionados al azar mostró que 48 de ellos aún los poseen. ¿Es válida esta afirmación, al nivel del 10%?

45. El jefe de admisiones de una universidad, afirmó en una reunión con las directivas que el 15% de los estudiantes que ingresan, se retiran antes de haber completado cuatro semestres académicos. En una revisión a los registros de los últimos años, mediante una muestra aleatoria de 300 alumnos, se encontró que 54 se retiraron antes de haber completado los 4 semestres académicos. ¿Al nivel del 1% es válida la afirmación?

DISTRIBUCIÓN DE DIFERENCIAS ENTRE DOS MEDIAS MUESTRALES y x

Esta prueba está indicada en aquellos casos cuando se quiere establecer si la diferencia entre dos medias muestrales, extraidas de dos *poblaciones independientes*, son significativas o si una media es mayor o menor que la otra.

Son ejemplos, cuando se quiere probar ¿si la accidentalidad vehicular es mayor en la población femenina que en la masculina? ¿si hay alguna diferencia en los hábitos de fumar de los hombres y mujeres? ¿si la duración de un producto de la marca A es diferente al de la marca B? Como se observa, estos interrogantes encierran la comparación entre dos poblaciones.

Mediante varios ejemplos, indicaremos el proceso a seguir en la realización de pruebas de diferencias cuando las ***dos muestras son grandes*** (mayores de 30) o cuando se dan las *desviaciones típicas poblacionales*, en este último caso no nos interesa los tamaños muestrales, pues los procesos son iguales, sólo que las varianzas poblacionales son sustituidas por las varianzas muestrales.

$$Z=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2}{n_1}+\frac{\sigma_y^2}{n_2}}} \quad \text{ó} \quad Z=\frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}}} \qquad Z=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

Veamos estos ejemplos que nos ayudaran a explicar el proceso a seguir e interpretar los resultados obtenidos.

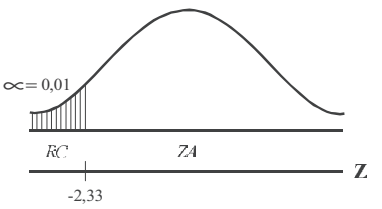
Ejemplo 1. Supongamos que la empresa desarrolló un curso de entrenamiento para sus técnicos, formando dos grupos y aplicando métodos distintos de entrenamiento. Los dos grupos se consideran homogéneos en capacidad. El primer grupo lo componen 36 técnicos que obtuvieron un puntaje de 6 (en una escala de 0 a 10 puntos) y una desviación típica de 4 puntos y el segundo grupo de 40 técnicos cuyo promedio fue 8,2 y desviación típica de 4,3 puntos. ¿Se puede concluir que el método aplicado al segundo grupo fue superior al primero? Nivel 1%

o **:**

1) $H_0:\mu_x=\mu_y$
$H_a:\mu_x<\mu_y$

2) $\propto=0{,}01$

3) $s_{\bar{x}-\bar{y}}=\sqrt{\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}}$



4) $Z = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{s_x^2}{n_1} + \dfrac{s_y^2}{n_2}}}$ $\quad Z = \dfrac{6 - 8{,}2}{\sqrt{\dfrac{16}{36} + \dfrac{18{,}49}{40}}} = -2{,}31$ $\quad Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

El resultado obtenido nos indica que no existe una diferencia significativa que permita concluir que el método aplicado en B fue superior al de A al nivel de 1%; observemos que -2,31 se ubica en la zona de aceptación.

Ejemplo 2. Una prueba de resistencia al esfuerzo de dos tipos diferentes de cables, que presentan desviaciones típicas de 35 y 45 respectivamente, se llevó a cabo, seleccionando dos muestras de tamaño 32 y 40, con medias de 905 y 925. ¿Proporcionan estos resultados, al nivel del 10%, suficiente evidencia de que la resistencia de B es superior a la de A?

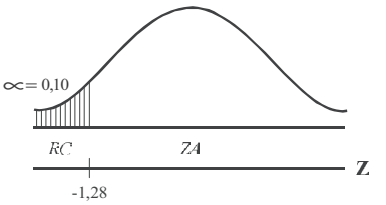
o **:**

$\sigma_x = 35$ $\quad \sigma_y = 45$ $\quad \bar{x} = 905$ $\quad n_1 = 32$ $\quad n_2 = 40$ $\quad \bar{y} = 925$

1) $H_0 : \mu_x = \mu_y$ $\quad$ 2) $\propto = 0{,}10$
$H_a : \mu_x < \mu_y$

3) $\sigma_{\bar{x}-\bar{y}} = \sqrt{\dfrac{\sigma_x^2}{n_1} + \dfrac{\sigma_y^2}{n_2}}$

4) $Z = \dfrac{(905 - 925) - 0}{\sqrt{\dfrac{35^2}{32} + \dfrac{45^2}{40}}} = -2{,}12$



Al nivel del 10%, si nos permite llegar a la conclusión de que la resistencia al esfuerzo del cable B es superior a la del cable A.

Ejemplo 3. Una firma que tiene dos fábricas ubicadas en dos regiones del país, desea establecer el promedio de antigüedad que tienen sus trabajadores, a fin de instaurar un programa para sus pensionados. Se toma de la primera fábrica una muestra de 60 obreros, la cual reflejó un promedio de trabajo de 16,4 años, con desviación estándar de 5 años, mientras que en la segunda fábrica una muestra de 40, fue de 15,8 años, con desviación estándar de 4,2 años ¿Al nivel del 5% se podrá afirmar que hay una diferencia significativa en cuanto a la antigüedad en la empresa?

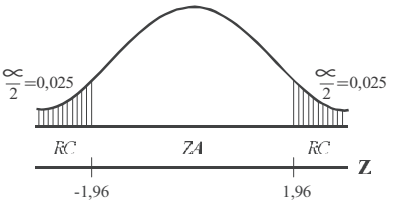
o **:**

$n_1 = 60$ $\quad n_2 = 40$ $\quad \bar{x} = 16{,}4$ $\quad \bar{y} = 15{,}8$ $\quad s_x = 5$ $\quad s_y = 4{,}2$

1) $H_0 : \mu_x = \mu_y$ $\quad$ 2) $\propto = 0{,}05$
$H_a : \mu_x \neq \mu_y$

3) $S_{\bar{x}-\bar{y}} = \sqrt{\dfrac{S_x^2}{n_1} + \dfrac{S_y^2}{n_2}}$ $\quad Z = \dfrac{\bar{x} - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

4) $Z = \dfrac{(16{,}4 - 15{,}8) - 0}{\sqrt{\dfrac{5^2}{60} + \dfrac{4{,}2^2}{40}}} = 0{,}65$



No hay diferencia significativa, en cuanto a la antiguedad en la empresa, al nivel del 5%

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

46. Se comparan dos procesos de fabricación. Una muestra de 100 artículos del primer proceso tiene una media de 107 y una desviación estándar de 17. En el segundo proceso, una muestra de tamaño 90 tiene una media de 103 y desviación estándar de 16. ¿Existe alguna diferencia significativa entre las medias de ambos procesos?

47.

	TIPO I	TIPO II
Bombillas en las muestras:	46	64
Media de la muestra:	1.070 horas	1.041 horas
Suma de los cuadros de las desviaciones respecto a la media:	$\sum (x_i - \bar{x})^2 = 21.000$	$\sum (y_i - \bar{y})^2 = 23.200$

¿Son significativamente diferentes, las medias de ambos tipos, con base en el nivel de significación del 1%?

48. Se requiere comparar el nivel salarial de los empleados de dos empresas. La primera reporta que en una muestra aleatoria de 46 empleados, su salario promedio fue de $818.000, con una desviación estándar de $32.000. Se elige una muestra aleatoria de 60 empleados de la segunda empresa obteniéndose un salario promedio de $842.000 y desviación estándar de $41.000 ¿Con los anteriores resultados podemos concluir que los salarios en la primera empresa son inferiores? (Nivel del 1%).

49. Los siguientes datos pertenecen al desarrollo del trigo bajo los tratamientos A y B:

	TRATAMIENTO A	TRATAMIENTO B
Número de plantas	44	36
Altura media	15,6	14,1
Suma de los cuadros de las desviaciones respecto a la media:	$\sum (x_i - \bar{x})^2 = 167,52$	$\sum (y_i - \bar{y})^2 = 159,89$

¿Existe alguna diferencia significativa entre ambas medias?

50. X_1 = ingresos mensuales (en cientos de $) de los empleados de una Cía. A y X_2 = ingresos mensuales (en cientos de $) de los empleados de una Cía. B. Con respecto a las variables se conocen los siguientes datos:

$\sum_{i=1}^{100} x_{1i}^2 = 2.500.200.000$	$\sum_{i=1}^{100} x_{2i} = 490.000$	$n_1 = 100$	$n_2 = 100$
$\bar{x}_1 = 5.000$	$M_{[x_2^2]} = 24.011.000$		

Se sabe además, que con la Cía. A el ahorro (y_1) en función del ingreso está dado por $y_1 = 20 + 0,2\,x_1$ y en la Cía. B el ahorro (y_2) en función del ingreso está dado por $y_2 = 520 + 0,1x_2$. Se puede aceptar, con un nivel de significación del 5%, que el ahorro promedio de la Cia. A es mayor que la Cia. B?

51. En una universidad se dice que el rendimiento de la Facultad A es inferior al de la Facultad B. Se le ha preguntado al jefe de notas sobre los respectivos promedios y desviaciones típicas y tan sólo recuerda estos últimos, 0,7 y 0,86 respectivamente. Frente a lo anterior se plantea la realización de

dos muestras de tamaño 20 y 28 en cada facultad. Obteniendo promedios de 3,32 y 3,50. ¿Al nivel del 5% se podrá aceptar lo que generalmente se dice?

52. Una cadena de almacenes ofrece dos planes de cuentas de cargo, disponibles para sus clientes con cuenta corriente de crédito. El departamento de crédito desea información acerca del comportamiento de cada plan, a fin de estudiar sus diferencias. Se consideró que una de las variables a estudiar debía ser el saldo mensual promedio. Se tomaron dos muestras de tamaño de 36 y 40 cuentas, de cada plan, con los siguientes resultados: promedios de $95 y $110 mil pesos respectivamente, siendo sus desviaciones estándar de $15 y $18 mil pesos. ¿Al nivel del 5%, se podrá afirmar que existe diferencia en el comportamiento de estos planes?

53. Un departamento de producción desea determinar si hay alguna diferencia en el rendimiento entre el turno diurno y el nocturno. Una muestra de 80 obreros del turno diurno revela una producción promedio de 94,3 partes por hora, con una desviación estándar de 14 partes por hora, mientras que otra muestra de 60 obreros de la jornada nocturna alcanza un promedio de 89,7 partes por hora, con una desviación estándar de 17 partes por hora. ¿Se podrá afirmar, al nivel del 5%, un mayor rendimiento en el turno diurno?

54. Una fábrica de salsa de tomate desea saber, si una nueva presentación de su producto aumentaría sus ventas. Una muestra de 40 tiendas reveló un promedio mensual de ventas de 310 botellas, con la nueva presentación y una desviación estándar de 20 botellas. Otra muestra realizada en 34 tiendas donde se expende el producto con el envase tradicional, revela unas ventas promedio de 292 botellas con desviación estándar de 26 botellas. ¿Al nivel del 10%, se puede decir que aumentaron las ventas?

55. Un departamento de investigación desea saber, si un producto con marca conocida se vende a un precio superior al de una marca desconocida. Se procedió a seleccionar 36 almacenes que distribuyen el producto cuya marca es ampliamente conocida y 32 almacenes que venden el mismo producto, de una marca que apenas entra al mercado. Los resultados fueron en promedio de $86.000 y de $80.000, con desviación estándar de $6.200 y $4.800 respectivamente ¿Al nivel del 1%, se puede aceptar tal información?

56. Se lanzan al mercado dos productos de diferentes marcas, cada uno como el "mejor" para bajar rápidamente de peso, sin causar daños colaterales. Se desarrollan dos encuestas de tamaño de 46 personas que usan el producto A y 35 del producto B, las cuales arrojan los siguientes datos: con el producto A se comprueba una reducción promedio de 10 libras en un mes, con una desviación típica de 2,4 libras, mientras que con el producto B, la reducción promedio es de 12 libras, con desviación estándar de 3 libras. ¿Al nivel del 5% se puede decir que hay una diferencia significativa en su efectividad?

57. Supongamos que dos muestras tomadas al azar, dan los siguientes resultados:

$n_1 = 82$	$n_2 = 41$	$\sum x_i = 4.100$
$\sum y_i = 2.225$	$\sum x_i^2 = 282.210$	$\sum y_i^2 = 213.284$

Probar la hipótesis de que el promedio de la segunda variable es superior a la primera (Nivel del 5%).

DISTRIBUCIÓN DE DIFERENCIAS ENTRE DOS PROPORCIONES MUESTRALES $_1 - _2$

Se ha dicho en repetidas ocasiones, que las *proporciones* son utilizadas como medidas aplicadas a *características cualitativas* (atributos). La prueba de hipótesis, como en los casos anteriores, implica el uso de la distribución normal, que nos permite establecer si hay o no alguna diferencia entre dos proporciones, obtenidas en dos poblaciones independientes, o si un grupo tuvo una proporción mayor que el otro.

Como en el caso de la distribución de diferencias entre dos medias, consideramos como hipótesis nula (Ho) que las dos proporciones de la población son iguales. Veamos su aplicación con varios ejemplos.

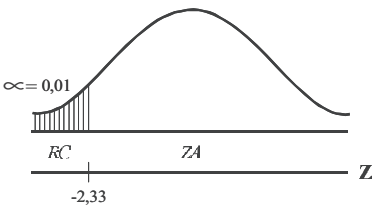
Ejemplo 1. En una encuesta se preguntó sobre los hábitos de lectura, utilizando una muestra aleatoria de 350 señoras que trabajan y otra muestra independiente de 325 que no lo hacen. En el primer caso, 105 manifestaron que estaban suscritas a cierto tipo de revista. En el segundo, la respuesta fue de 130 que no estaban suscritas ni mostraban interés por ninguna revista, argumentando la falta de tiempo. ¿Al nivel del 1%, se podrá afirmar que las señoras que trabajan leen menos que las señoras que no trabajan?

o **:**

1) $H_0 : P_1 = P_2$ $\quad p_1 = \frac{105}{350} = 0{,}30 \quad p_2 = \frac{130}{325} = 0{,}40 \quad 2) \propto = 0{,}01$
$H_a : P_1 < P_2$

3) $s_{\bar{P}_1 - \bar{P}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

4) $Z = \frac{0{,}30 - 0{,}40}{\sqrt{\frac{0{,}3(0{,}7)}{350} + \frac{0{,}4(0{,}6)}{325}}} = -2{,}73$



Al nivel del 1%, si se puede afirmar que las señoras que trabajan leen menos, que las que no trabajan.

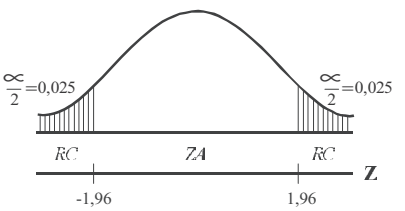
Ejemplo 2. Un gerente de una compañía, realiza dos muestras de tamaño de 120 empleados, una en cada fábrica, con el fin de determinar el porcentaje de accidentes de trabajo en el trimestre. En la primera fábrica durante el trimestre de observación se presentaron 12 casos, mientras que en la segunda, 16. ¿Al nivel del 5%, se podrá afirmar que los accidentes de trabajo son iguales en las dos fábricas?

o **:**

1) $H_0 : P_1 = P_2$ $\quad p_1 = \frac{12}{120} = 0{,}10 \quad p_2 = \frac{16}{120} = 0{,}13$
$H_a : P_1 \neq P_2$

2) $\propto = 0{,}05$

3) $Z = \frac{0{,}10 - 0{,}13}{\sqrt{\frac{0{,}1(0{,}9)}{120} + \frac{0{,}13(0{,}87)}{120}}} = -0{,}73$



Podemos concluir que la accidentalidad en el trabajo es igual en las dos fábricas, al nivel del 5%.

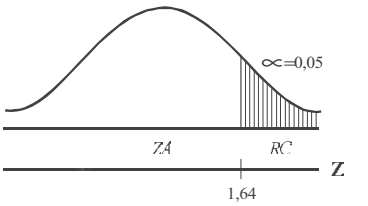
Ejemplo 3. Dos grupos A y B de 100 personas cada uno tienen determinada enfermedad. Un suero es dado al grupo A, pero no al B. Por otra parte, los grupos son tratados idénticamente. Si encontramos que en el grupo A, 75 personas se recobran de la enfermedad y en el B, 65, pruebe la hipótesis de que el suero cura la enfermedad.

o :

$$n_1 = 100 \qquad n_2 = 100 \qquad p_1 = \frac{75}{100} = 0{,}75 \qquad p_2 = \frac{65}{100} = 0{,}65 \qquad \propto = 0{,}05$$

1) $H_0 : P_1 = P_2$
$H_a : P_1 > P_2$

2) $\propto = 0{,}05$



3) $s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

4) $Z = \dfrac{p_1 - p_2}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}$ $\qquad Z = \dfrac{0{,}75 - 0{,}65}{\sqrt{\dfrac{(0{,}75)(0{,}25)}{100} + \dfrac{(0{,}65)(0{,}35)}{100}}} = \dfrac{0{,}10}{0{,}064} = 1{,}56$

Al nivel de investigación del 5%, no podemos aceptar que el suero cure la enfermedad.

Ejemplo 4. Al investigar la "imagen" de calidad de cierta marca de reloj de pulso, se seleccionó una muestra de 120 profesionales y 80 talleres de reparación y se obtuvo la siguiente información. Con los resultados que se dan a continuación, ¿se podría interpretar, al nivel del 5% que existe una diferencia significativa, en cuanto a opinión sobre la marca del reloj?

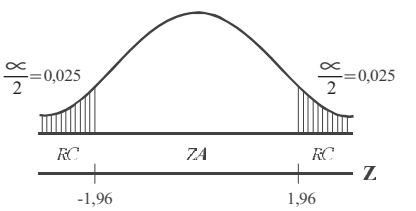
IMAGEN DE CALIDAD	PROFESIONALES	TALLERES
La marca es buena. Rara vez sufre descompostura en los primeros años de uso.	60	48
La marca no es buena. Se descompone frecuentemente, desde el primer año de uso.	60	32
Total	120	80

o :

$$n_1 = 120 \qquad n_2 = 80 \qquad p_1 = \frac{60}{120} = 0{,}50 \qquad p_2 = \frac{48}{80} = 0{,}60 \qquad \propto = 0{,}05$$

1) $H_0 : \mu_{p_1} = \mu_{p_2}$
$H_a : \mu_{p_1} \neq \mu_{p_2}$

2) $\propto = 0{,}05$



3) $s_{\bar{p}_1 - \bar{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

4) $$Z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \qquad Z = \frac{0{,}5 - 0{,}6}{\sqrt{\frac{0{,}5(0{,}5)}{120} + \frac{(0{,}6)(0{,}4)}{80}}} = \frac{-0{,}1}{0{,}071} = -1{,}41$$

Al nivel del 5%, se puede concluir que no hay una diferencia significativa en esos resultados.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

58. Generalmente se escucha que el equipo de la ciudad debe ganar más partidos, cuando juega de local que de visitante. Se toman al azar 40 partidos de local y 40 jugados por fuera durante las últimas cinco temporadas, con los resultados favorables de 30 y 22 partidos, respectivamente. Al nivel del 5%, ¿con los resultados obtenidos se puede aceptar tal afirmación?

59. Un estudio del consumo de café en el trabajo, por sexo, mostró en una muestra aleatoria de 200 mujeres que 128 lo toman durante su trabajo, mientras que una muestra de 150 hombres reveló que 106 lo toman. ¿Hay alguna diferencia entre la proporción de los dos grupos, en cuanto al hábito de tomar café en el trabajo?

60. El departamento de matemáticas desea determinar si la proporción de alumnos que pierden la materia se debe a su estado civil. Se tomó una muestra de 60 alumnos casados de los cuales 12 la habían perdido, mientras que en otra muestra de 60 alumnos solteros, tan solo 10 la perdieron. ¿Al nivel del 5%, el estado civil influye en el rendimiento?

61. Un jefe de personal quiere comprobar que la ausencia al trabajo es mayor en el turno nocturno que en el diurno. Selecciona en cada jornada 40 y 50 trabajadores y encuentra que en el mes, para el turno diurno 7 trabajadores se excusaron de ir a trabajar, mientras que en la noche faltaron 12. ¿Los anteriores resultados dan la razón al jefe de personal? (Nivel del 10%).

62. Un distribuidor de dos marcas de baterías para automóvil, asegura que la marca A es mejor que la B. Una muestra aleatoria de 50 baterías de la marca A, muestra que 38 de ellas duraron más de tres años, mientras que de 70 de la marca B, sólo 50 duraron más de tres años. ¿Estos resultados confirman la afirmación del distribuidor?

63. Un fabricante de helados quiere saber si su producto tiene igual aceptación entre niños y niñas de secundaria. Se tomaron dos muestras de 40 alumnos cada una, encontrando que 26 niños y 30 niñas prefieren el producto. ¿Se podrá decir que la proporción de aceptación por sexo es igual?

64. Dos grupos A y B de 500 personas, cada una, presenta la misma sintomatología. Al grupo A se le da una droga que contiene un nuevo compuesto y la droga del grupo B carece de él. Por otra parte, los grupos son tratados idénticamente. Si encontramos que en el grupo A, 375 personas se recobran de la enfermedad y en el B, 325, pruebe la hipótesis de que la droga aplicada al grupo A es mejor que la del B, (Nivel del 5%).

65. Una empresa fabricante de equipos mecánicos patrocina un curso de adiestramiento para su personal de ventas y servicio. Los anales del curso mostraron en una encuesta anterior realizada a 100

estudiantes que el 64 % lo aprobaron. Se efectúan ciertos cambios en el programa de adiestramiento con el objeto de mejorarlo y al año siguiente, de 100 alumnos del curso, 70 aprobaron. Al nivel del 1% ¿Podría considerarse este hecho, como una demostración de efectividad de las reformas introducidas al programa?

66. Una fábrica tiene dos plantas dedicadas al proceso de llenado; cuando no cumplen con las especificaciones deben ser reprocesadas. Cada tres horas se aplica el muestreo y se seleccionan 100 cajas de cada planta. Encontramos que en la primera planta 8 necesitan reprocesamiento, mientras que en la segunda 6 lo requieren. ¿Hay alguna diferencia entre A y B en la proporción de cajas que requieren ser reprocesadas?

67. Un sociólogo cree que la proporción de hombres que pertenecen a un grupo socio-económico determinado (grupo A) y que ve regularmente deportes en televisión, es superado por un segundo grupo de hombres (grupo B) que también ve deportes. Muestras aleatorias simples de los dos grupos arrojan los siguientes resultados:

GRUPO	TAMAÑO	No. HOMBRES
A	$n_1 = 130$	80
B	$n_2 = 100$	96

¿Proporcionan estos datos, evidencia suficiente como para apoyar la tesis del sociólogo (Nivel del 5%)?

68. Un líder sindical, considera que la proporción de ejecutivos a favor de una política de control de precios, es menor que la proporción de obreros a favor del control. Se toman dos muestras aleatorias independientes de igual tamaño, 100 personas en cada una y se encuentran 42 ejecutivos y 61 obreros a favor. ¿Al nivel del 1% se podrá aceptar la consideración hecha por el líder sindical?

69. En un estudio, para averiguar los efectos de usar modelos femeninos en la publicidad para automóviles, a un grupo de 50 hombres, (grupo A), se le mostró la fotografía de un automóvil con una modelo femenina, y la de otro automóvil del mismo precio pero sin modelo. A otro (grupo B), de 50 hombres, se le mostraron ambos automóviles sin modelo femenino. En el grupo A el automóvil que aparecía con la modelo fue considerado más lujoso por 37 de los entrevistados; en el grupo B el mismo automóvil fue juzgado como el más lujoso por 23 de los entrevistados. ¿Se considera que estos datos indican, que usar una modelo influye en el lujo aparente de un automóvil? (Nivel del 5%).

70. Un gerente de una compañía realiza dos muestras de tamaño 120 (una en cada fábrica) con el fin de determinar el porcentaje de ausentismo al trabajo. En la primera fábrica, durante la semana de observación no asistieron 12, mientras que en la segunda faltaron 16 trabajadores. ¿Al nivel del 5%, se podrá afirmar que la no asistencia en la segunda fábrica es superior a la primera?

71. Una programadora de TV. afirma que el hábito a la programación deportiva es indiferente al sexo. Realiza dos muestras a 120 mujeres y a 150 hombres, encontrando que el 25% de las mujeres no ven ningún programa deportivo, mientras que el 15% de los hombres, le encuentran alguna satisfacción. ¿Al nivel del 5% se debe aceptar dicha afirmación?

TEORÍA DE LAS MUESTRAS PE UEÑAS

DISTRIBUCIÓN t DE STUDENT

En problemas referentes a la prueba de hipótesis, *cuando se conoce la desviación típica poblacional* no importa que el tamaño de la muestra sea grande o pequeña. Se dice que una muestra es grande, si el número de unidades es mayor a treinta y es pequeña si es menor o igual a treinta.

En el caso de que se *desconozca* la desviación típica poblacional, se le podrá reemplazar por la *desviación típica muestral,* siempre que la muestra *sea grande*.

Cuando n >30 (muestra grande), la desviación típica se simboliza por s y se considera como un buen estimador de la desviación típica poblacional, debido a que existe una mayor probabilidad de que los valores extremos que toma la variable, queden incluidos en el cálculo de la varianza en la muestra, tal como ocurre al obtener la varianza poblacional.

Siendo: $s = \sigma$ la fórmula para calcular s será: $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

Si $n \leq 30$ (muestra pequeña), la desviación típica se simboliza por s cuando no se ha efectuado ninguna corrección. Se considera que $\hat{s}$, por lo general, es menor que σ, debido a la poca posibilidad de que se incluyan valores extremos de la variable poblacional en la muestra. Por lo tanto se hace necesario efectuar algunas correcciones en su cálculo, con el fin de convertirla en un buen estimador de la desviación típica poblacional.

También se puede expresar como $s = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}}$. En las calculadoras se puede observar: $_x\sigma_n$ y $_y\sigma_n$ para muestras grandes (n > 30) o sin corregir cuando la muestra es pequeña. Las teclas $_x\sigma_{n-1}$ y $_y\sigma_{n-1}$ indican desviaciones típicas corregidas, es decir, dividida por n –1 y no por n.

A pesar de las correcciones que se le puedan hacer a las desviaciones típicas, no es efectiva en todas las muestras; por tal razón la distribución de todas las medias muestrales, no tiene un comportamiento similar a la distribución normal, a pesar de ser una distribución continua. A esta distribución se le conoce como distribución ”t” de Student, en honor al estadístico irlandés W. S. Gosset, quien escribía bajo el seudónimo de Student. Fue enunciada en 1908 y ligeramente modificada por R. A. Fisher (1925).

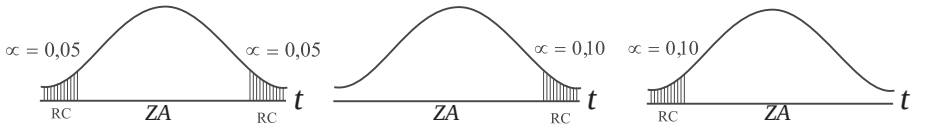
Esta distribución se expresa en forma de campana y simétrica, pero más achatada y con más área en los extremos, es decir, las áreas que corresponden a las regiones críticas o de rechazo. Se puede considerar que no hay una distribución “t”, sino más bien una familia de distribuciones “t”, dado que las desviaciones estándar se modifican a medida que se va aumentando el tamaño de la muestra, acercándose a la normal.

La función de la distribución “t” es $Y = C\left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$

Grados de libertad. Esta expresión fue introducida por Fisher; dice que los *grados de libertad* de un conjunto de observaciones, están dados por el número de valores que pueden ser asignados arbitrariamente, antes de que el resto de las variables queden completamente determinadas. Ejemplo: Si tengo cuatro valores: 8 + 10 + 6 + 20 = 44, se observa que puedo asignar tres (3) valores arbitrariamente: 16 + 4 + 10 + 14 = 44, pero el cuarto debe ser 14 y no lo puedo asignar libremente pues la suma debe ser 44. Por lo tanto $v = n$ - 1 = 4 - 1 = 3. “υ ” se lee “nu” o “niu” y se emplea para simbolizar los *grados de libertad.*

Veamos ahora qué tratamiento se le debe dar a la varianza o a la desviación estándar muestral, en especial, cuando se van a realizar pruebas de hipótesis, trabajando con muestras pequeñas en las diferentes distribuciones.

En la determinación de los puntos críticos, t_i (inferior) y t_s (superior), se utiliza la tabla, "t" de Student. En primer lugar se fija el nivel de significación, por ejemplo $\propto = 0{,}05$; luego se calculan los grados de libertad, siendo en distribuciones muestrales: $\nu = n - 1$ y en diferencias de medias muestrales: $\nu = n_1 + n_2 - 2$. Si la dócima es *bilateral* se tomará el total, suponiendo que sea del 5%, del nivel de significación, para cada una de las regiones críticas $\propto = 0{,}05$ y si es *unilateral*, se tomará el doble del nivel de significación asignado, en este caso, será $\propto = 0{,}10$.



bilateral *unilateral*

Es necesario observar que algunos textos consideran que las muestras son pequeñas si n_1 y n_2. son menores o iguales a 100, pero aquí vamos a considerar que son pequeñas si n_1 y n_2. son menores o iguales a 30.

DISTRIBUCIÓN DE MEDIAS MUESTRALES (Muestras peque as)

Cuando el problema nos da la desviación típica muestral y a la vez el tamaño de la muestra es menor o igual a 30 ($n < 30$), se considera que la desviación está sin corregir, procediendo a su corrección para ser aplicada en la variante estadística "t". Simbolicemos a: $\hat{S}$ (desviación típica sin corregir) y (desviación típica corregida).

1) $s = \hat{s}\sqrt{\dfrac{n}{n-1}}$ 2) $s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\dfrac{\sum x_i^2 - n\bar{x}^2}{n-1}}$ 3) $t = \dfrac{\bar{x} - \mu}{\hat{s}/\sqrt{n-1}}$

En la (3) se puede corregir directamente, al calcular la variante estadística:

Ejemplo 1. Una muestra de 25 observaciones tiene una media de 42,0 y una desviación de 8. Trabajando con un nivel de significación del 1%. ¿Existe razón para rechazar la hipótesis de que la media de la población es 46,0?

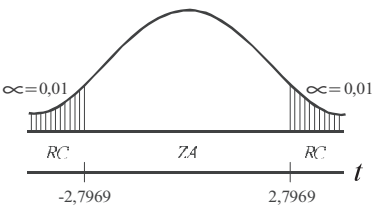
o :

$n = 25 \qquad \bar{x} = 42 \qquad \mu = 46 \qquad \hat{s} = 8$

1) $H_0 : \mu_x = 46$ 2) $\propto = 0{,}01$

$H_a : \mu_x \neq 46$

3) $\hat{s} = 8 \qquad s = 8\sqrt{\dfrac{25}{24}} = 8(1{,}021) = 8{,}17$



4) $t = \dfrac{42-46}{8{,}17/\sqrt{25}} = \dfrac{-4(5)}{8{,}17} = \dfrac{-20}{8{,}17} = -2{,}45$ ó $t = \dfrac{42-46}{8/\sqrt{25-1}} = -2{,}45$

$\upsilon = 25-1 = 24$ y $\propto = 0{,}01$ corresponde una $t = 2{,}7969$

Se acepta la hipótesis de que μ = 46, es decir, no existe razón para rechazar, que la media de la población es 46, al nivel del 1 %.

Es necesario observar que si el nivel de significación, como en este ejemplo, es de $\propto = 0{,}01$ se tomará el mismo nivel para cada una de las zonas críticas (zona de rechazo) tal como aparece en la tabla donde dice "nivel de significación para pruebas de dos colas". (ver tabla 8.3 ; página 358).

En caso de que sea de una cola, suponiendo que $\propto = 0{,}05$ debe observarse en la tabla, donde dice "Nivel de significación para pruebas de una cola", es decir, debe corresponder al doble para pruebas de dos colas, así que a pesar de ser $\propto = 0{,}05$ se está trabajando con $\propto = 0{,}10$.

RESUMEN

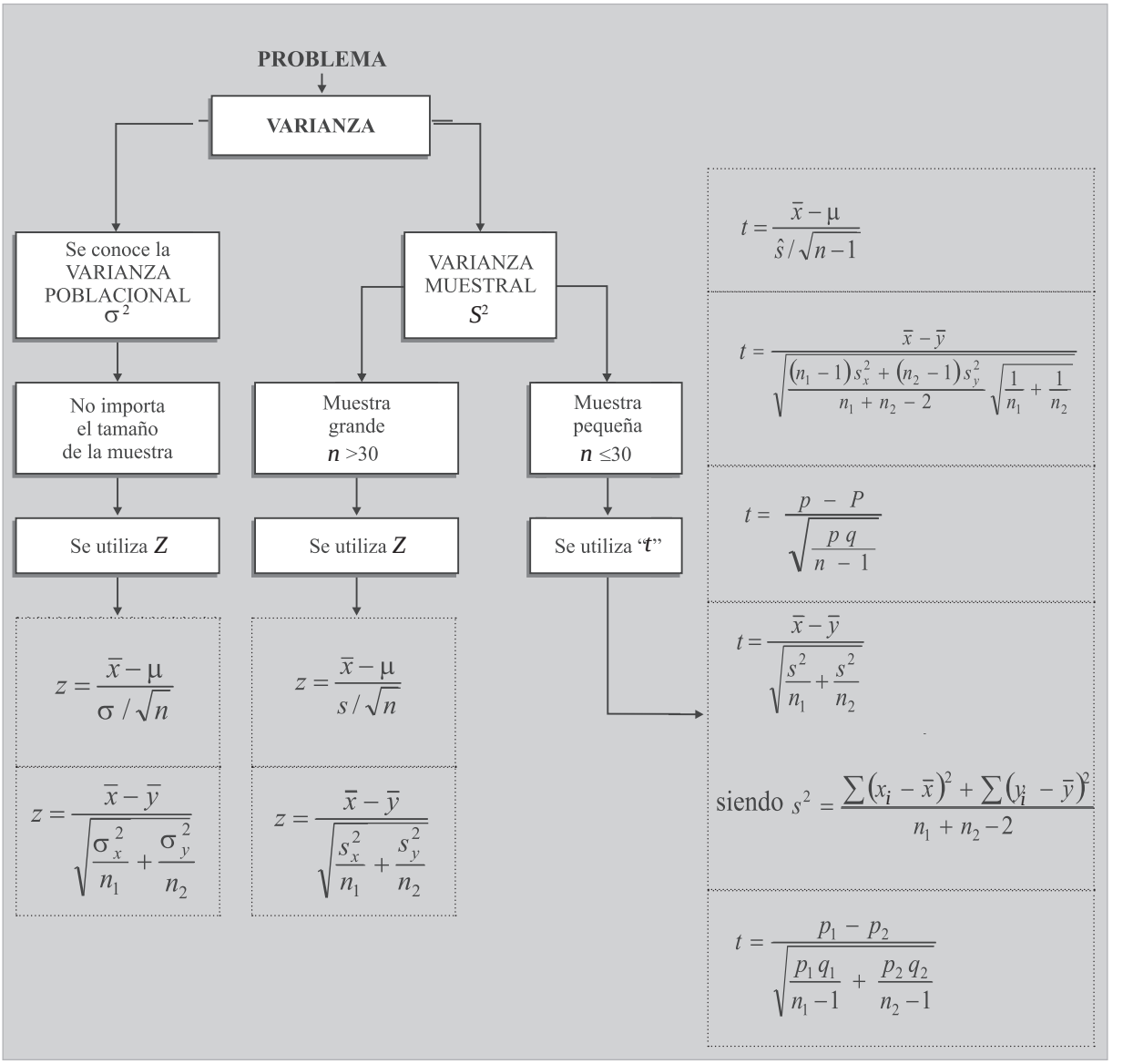


Tabla 8.3 Tabla de distribución "t" de Student

NIVEL DE SIGNIFICACIÓN PARA PRUEBAS DE UNA COLA						
Grados de libertad υ	**0,10**	**0,05**	**0,025**	**0,01**	**0,005**	**0,0005**
	Nivel de significación para pruebas de dos colas					
	0,20	**0,10**	**0,05**	**0,02**	**0,01**	**0,001**
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,681	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646
40	1,303	1,684	2,021	2,423	2,704	3,551
60	1,296	1,671	2,000	2,390	2,660	3,460
120	1,289	1,658	1,980	2,358	2,617	3,373
∞	1,282	1,645	1,960	2,326	2,576	3,291

Fuente: Esta tabla es un resumen de la Tabla III de Fisher y Yates: *Statistical Tables for Biological, Agricultural, and Medical Research*, obra publicada por Oliver & Boyd Ltd., Edimburgo.

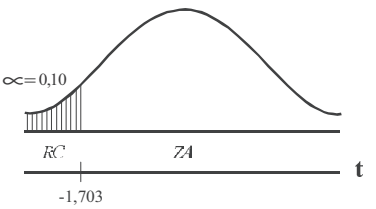
Ejemplo 2. Una inspectora de calidad, investiga las acusaciones contra una fábrica de cerveza, porque no

llena bien sus envases, afirmando que contienen 35 onzas de líquido. Se muestrearon 28 latas de cerveza, encontrando un contenido medio de 33,2 onzas, con una desviación estándar de 2,2 onzas ¿Debe la inspectora llegar a la conclusión, al nivel del 5%, que se está exagerando su contenido?

o :

$n = 28 \qquad \bar{x} = 33{,}2 \qquad \hat{s} = 2{,}2 \qquad \mu = 35 \qquad \upsilon = 28 - 1 = 27$

1) $H_0 : \mu = 35$ 2) $\propto = 0{,}05$
$H_a : \mu < 35$



3) $s = \hat{s}\sqrt{\dfrac{n}{n-1}} = 2{,}2\sqrt{\dfrac{28}{27}} = 2{,}24$

4) $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{33{,}2 - 35}{2{,}24/\sqrt{28}} = -4{,}25$

Tambien: $t = \dfrac{\bar{x} - \mu}{\hat{s}/\sqrt{n-1}} = \dfrac{33{,}2 - 35}{2{,}2/\sqrt{28-1}} = -4{,}25$

La respuesta es: al nivel del 5%, el inspector si puede llegar a la conclusión de que se está exagerando el llenado.

> Obsérvese en la tabla 8.3 cómo se obtiene el valor del punto crítico, cuando ∝= 0,05 (nivel de significación); localizamos el 0,05 donde dice "prueba de una sola cola" y encontramos que equivale al 10% en las "pruebas de dos colas", es decir el doble; por lo tanto, siendo el grado de libertad de 27 (υ = 28 - 1 = 27) y ∝= 0,10 el valor de "t" (punto crítico) será de 1,703 y al ser localizada a la izquierda de la curva se le da signo negativo.

Ejemplo 3. Un fabricante desea hacer público, a fin de aumentar sus ventas, que el contenido de nicotina de sus cigarrillos tiene un promedio inferior a los 22 mg. Una oficina gubernamental de salud y del medio ambiente, realiza un análisis de 10 cigarrillos y obtiene los contenidos de nicotina para cada uno, siendo de:

21	24	18	16	22	23	20	20	24	16

a) ¿Tiene justificación lo aseverado por el fabricante?; b) ¿Cuándo se cometerá un error de Tipo II?

o :

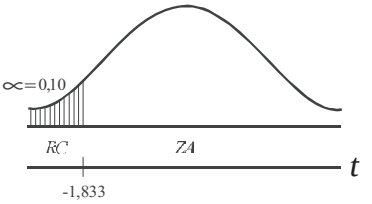
$n = 10 \qquad \bar{x} = \dfrac{\sum x_i}{n} = \dfrac{204}{10} = 20{,}4$

$s = \sqrt{\dfrac{\sum x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\dfrac{4.242 - 10(20{,}4)^2}{10-1}} = 2{,}99$

1) $H_0 : \mu = 22$ 2) $\propto = 0{,}05$ 3) $s = 2{,}99$
$H_a : \mu < 22$

4) $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ $\quad t = \dfrac{20{,}4 - 22}{2{,}99/\sqrt{10}} = -1{,}69$

$$\left.\begin{array}{l} \upsilon = n - 1 = 10 - 1 = 9 \\ \propto = 0{,}10 \end{array}\right\} t = 1{,}833$$



(a) No hay justificación, aceptar lo aseverado, al nivel del 5%.
(b) Aceptar que μ = 22 mg, cuando el contenido de nicotina es en realidad diferente.

Ejemplo 4. Una muestra aleatoria de 10 vigas de acero, tiene una resistencia media a la compresión de 57.498 libras por pulgada cuadrada (l.p.c.), con una desviación típica de 539 l.p.c. Docimar la hipótesis, de que la verdadera resistencia media a la compresión de las vigas de acero de las que se extrajo la muestra es μ = 57.000. Utilizar la alternativa bilateral μ ≠ 57.000 y un nivel de significación del 1%.

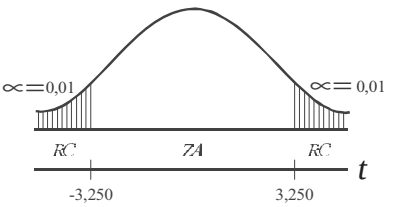
o :

$n = 10$ $\quad \bar{x} = 57.498$ $\quad \hat{s} = 539$ $\quad \mu = 57.000$ $\quad \upsilon = 10 - 1 = 9$

1) $H_0 : \mu = 57.000$ $\quad$ 2) $\propto = 1\%$
$H_a : \mu \neq 57.000$

3) $\hat{s} = 539$

4) $t = \dfrac{57.498 - 57.000}{539/\sqrt{10-1}} = \dfrac{498(3{,}00)}{539} = 2{,}77$



Aceptamos que μ = 57.000, es decir, que es la verdadera resistencia media a la compresión de las vigas de acero, con un nivel de significación del 1%.

Nota: Se trabajó con "t" debido a que n es pequeña y se da (s) la desviación típica de la muestra, sin corregir.

Ejemplo 5. Un fabricante de ciertas piezas de proyectiles sostiene que, en condiciones normales de operación, tienen una duración media μ = 320 horas. Docimar esta afirmación frente a la alternativa μ ≠ 320, si 16 piezas duran un promedio de 308 horas con una desviación estándar de 29 horas. Utilizar un nivel de significación del 5%.

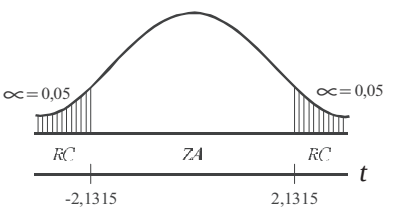
o :

$n = 16$ $\quad \bar{x} = 308$ $\quad \hat{s} = 29$ $\quad \mu = 320$ $\quad \upsilon = 15$ $\quad \propto = 5\%$

1) $H_0 : \mu = 320$
$H_a : \mu \neq 320$

2) $\propto = 0{,}05$ $\quad$ 3) $\hat{s} = 29$

4) $t = \dfrac{308 - 320}{29/\sqrt{15}} = \dfrac{-12(3{,}87)}{29} = -1{,}60$



Se ubica en la zona de aceptación, o sea que la duración promedio es de 320, a un nivel de significación del 5%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

72. Un estudio a 29 familias de una zona residencial de una ciudad de Colombia, revela que el ingreso medio por familia durante 2011 fue de $808.000, con una desviación estándar de $16.000. Docimar la hipótesis, de que el verdadero ingreso medio por familia en Colombia durante 2011 fue de $800.000, frente a la alternativa de que no fue de $800.000. Utilizar un nivel de significación del 5%.

73. Un pescador decide que necesita un sedal que resista más de 10 libras si ha de capturar el tamaño de pescado que desea. Prueba 16 piezas de sedal de la marca G y halla una media muestral de 10,4 libras. Si en la muestra se obtiene que la desviación típica es de 0,5 libras, ¿qué conclusión se puede sacar del sedal de la marca G.? (Nivel de significación del 5%).

74. Un jefe de personal, está dispuesto a contratar una secretaria para ocupar un puesto a menos que ella cometa más de una equivocación por página mecanografiada. Se elige una muestra aleatoria de cinco páginas de las escritas por la aspirante. Las equivocaciones por página son: 3 3 4 0 1. Utilizando $\propto = 0,05$ ¿qué decisión se debe tomar?

75. Durante un período de 12 meses, el número de nacimientos de mellizos por mes registrados en un hospital son: 2 0 1 1 0 0 3 2 1 1 0 1. ¿Contradicen estos resultados la hipótesis de que el promedio de nacimiento de mellizos es de 0,5 por mes? (Nivel del 5%).

76. Supongamos que una persona quiere tener desconectado su teléfono, si el promedio de llamadas que hace el día es menor de 2. Elige aleatoriamente 5 días y anota el número de llamadas, así: 0, 2, 1, 1, 2. Utilizando $\propto = 0,05$, ¿debería retirar el teléfono?

77. Un dispensador de gaseosas está diseñado para descargar 7 onzas. Si se selecciona una muestra de 16 vasos para medir su llenado, observando que el promedio es de 5,8 onzas con una desviación de 1,6 onzas, ¿se puede concluir que la máquina no funciona correctamente?

78. Un test de psicología tenía un puntaje de 82 puntos; en un grupo de 20 estudiantes el puntaje fue de 75, con una desviación de 9 puntos. Al nivel del 5%, ¿se puede afirmar que la diferencia es significativa?

79. Un profesor realiza un examen a su curso y sabe, por experiencia que proporciona un rendimiento de 78 puntos. Si su curso actual lo conforman 25 alumnos y obtienen una nota promedio de 82 y desviación típica de 21, ¿acierta al suponer que es un curso superior? ($\propto$ =1%).

80. El departamento de admisiones de una universidad, sostiene que el promedio de calificación de los alumnos admitidos es de 380 puntos, de un total de 450 que exige. Si se toma una muestra de 25 alumnos cuyo promedio es de 360, con una desviación típica de 40 puntos. Al nivel del 1%, ¿se puede afirmar que el puntaje de los admitidos es inferior al señalado?

81. El sindicato de empleados de establecimientos comerciales detallistas, sostiene que el salario medio por día de sus asociados es de $18.250. La asociación de dueños de establecimientos comerciales

detallistas piensan, que el sindicato está subestimando el salario medio por día. En una muestra aleatoria de 30 empleados, la Asociación encuentra que el salario medio por día es de $19.500, con una varianza de $9.000.000. Si la asociación está dispuesta a rechazar una afirmación verdadera no más de 5 en 100 (95%), ¿rechazará la afirmación del sindicato?

82. Un distribuidor de botas especiales para trabajo, garantiza al gerente de una empresa, que el promedio de duración de las botas es de 8,10 meses. La empresa decide comprar 25 pares de botas, que en promedio duran 7 meses y 15 días, con una varianza de 12,5 meses. ¿Al nivel del 5% se podrá decir que la duración de las botas es inferior al señalado por el distribuidor?

83. Anteriormente, un determinado grupo de máquinas había tenido una duración media entre reparaciones de 200 horas de operación. El personal de este departamento recientemente terminó un programa de entrenamiento en el que se destacaba el mantenimiento preventivo del equipo. Las siguientes 15 reparaciones tuvieron una media de 210 horas entre cada una, con una desviación estándar de 11 horas. ¿Proporciona esto, una prueba suficiente para concluir que el tiempo medio entre descompostura aumenta? (Nivel del 1%).

84. Una muestra aleatoria simple de tamaño 12 de una población, presentó los siguientes resultados:

128	158	174	201	193	160	172	120	125	136	184	192

¿Proporcionan estos datos, evidencia suficiente para poder concluir que la media poblacional es superior a 160? (Nivel del 1%).

85. El fabricante de un cierto modelo de automóvil afirma, que el kilometraje medio de este modelo es de 42 km por galón de gasolina corriente. Un organismo de defensa del consumidor piensa que ese kilometraje promedio ha sido exagerado por el fabricante. Nueve automóviles de este modelo son conducidos del mismo modo con un galón de gasolina corriente. Los kilómetros recorridos por los diversos automóviles son:

42	41	40	40	30	41	35	43	42,5

Si el organismo desea rechazar una afirmación verdadera no más de una vez en 100, ¿rechazará la afirmación del fabricante?

86. Una distribuidora de gas ofrece a sus clientes el servicio en un máximo de espera de 48 horas. Se toma una muestra de seis hogares que hicieron pedidos y se encontró lo siguiente: 24 20 60 72 40 30 ¿se puede creer lo ofrecido por la distribuidora?

DISTRIBUCIÓN DE UNA PROPORCIÓN MUESTRAL

En *muestras pequeñas*, correspondiente a distribuciones *proporcionales*, se aplica un procedimiento casi igual que en las muestras grandes, con la diferencia de ser una distribución "t" de Student, que a la vez implica para *varios autores* corregir el error estándar de la proporción. La fórmula a utilizar es:

$$t = \frac{p - P}{\sqrt{\frac{pq}{n-1}}}$$

Veamos algunos ejercicios con aplicación a este tipo de distribución.

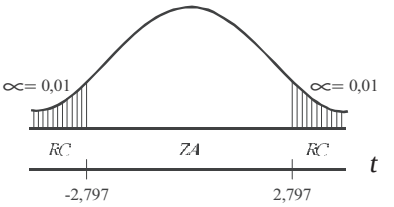
Ejemplo 1. Se dice con frecuencia, que la proporción de funcionarios públicos que tienen el hábito de fumar en horas de trabajo, es del 42%. La oficina gubernamental de salud desea realizar una campaña a fin de disminuir este porcentaje, para ello debe comprobarlo; así que decide realizar una investigación de muestreo a 25 funcionarios, encontrando que 13 de ellos fuman. ¿A nivel del 1% la oficina puede aceptar el porcentaje del 42% como indicador?

Solución:

$$\mu_p = P = 42\% \qquad p = \frac{13}{25} = 0{,}52 \qquad n = 25 \qquad \upsilon = n - 1 = 24$$

1) $H_0 : \mu_p = 0{,}42$ $\quad H_a : \mu_p \neq 0{,}42$ 2) $\propto = 0{,}01$ 3) $s_p = \sqrt{pq}$

4) $t = \dfrac{0{,}52 - 0{,}42}{\sqrt{\dfrac{0{,}52(0{,}48)}{25-1}}} = 0{,}98$



Sí hay razón para aceptar el 42%, como indicador de fumadores en horas de trabajo, al nivel del 1%.

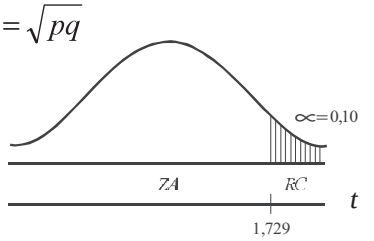
Ejemplo 2. El distribuidor de una máquina afirma, que el máximo de elementos defectuosos por hora que presenta su funcionamiento es del 3%. En una determinada hora, se toma como muestra 20 artículos producidos, los que a su vez son sometidos a control, encontrando un artículo defectuoso. ¿Al nivel del 5%, se podrá decir que el porcentaje de defectuosos es superior al señalado por el distribuidor?

o **:**

$$P = 3\% \qquad p = \frac{1}{20} = 0{,}05 = 5\% \qquad n = 20 \qquad \propto = 0{,}05$$

1) $H_0 : P = 0{,}03$ $\quad H_a : P > 0{,}03$ 2) $\propto = 0{,}05$ 3) $s_p = \sqrt{pq}$

4) $t = \dfrac{0{,}05 - 0{,}03}{\sqrt{\dfrac{0{,}05(0{,}95)}{20-1}}} = 0{,}4$



Con este resultado no se puede concluir, que el porcentaje de defectuosos sea superior al señalado por el distribuidor, al nivel del 5%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

87. El gerente de una Corporación de ahorros, argumenta que el 30% de los clientes posee saldo aproximado a los $150.000. Se desea confirmar tal operación, mediante una muestra aleatoria a 20 clientes elegidos al azar, de los cuales 8 tienen saldo superior a los $150.000. Al nivel del 5% ¿se podrá decir que el porcentaje de ahorradores tienen saldos superiores a la afirmación hecha por el gerente?

88. Una papelería con un alto volumen de ventas de elementos para estudiantes universitarios, recibe de un proveedor cierto artículo, con una garantía del 3% de defectuosos. El dueño de la papelería desea saber, si en los pedidos que hace se cumple la garantía, para ello toma una muestra de 28 artículos encontrando 2 con algún defecto. ¿Se puede decir que el fabricante no cumple con lo prometido?

89. Al frente de la universidad se instala una venta de comida rápida. El encargado del negocio encuentra que el 70% de los estudiantes prefieren el pastel de pollo, en vez de carne de res. Selecciona una muestra de 30 estudiantes para estudiar su preferencia, encontrando que 19 de ellos prefieren el pastel de pollo. ¿Se podrá afirmar que el encargado del negocio exagera el porcentaje?

90. Un fabricante de automóviles afirma que sus autos de tipo familiar, en el 86% de los casos pueden resistir un choque de frente a una velocidad inferior a los 70 k/h, si utilizan cierto equipo. Se toma una muestra de 18 vehículos que tienen este equipo; se encuentran que 16 autos resisten un choque de frente. ¿Se puede decir, al nivel del 1% que el equipo es mucho más efectivo que la afirmación del fabricante?

91. Una agencia de empleos, critica el hecho de que el 30% de las personas que son colocadas no, pasan la prueba de trabajo en los tres meses. Se quiere comprobar esta crítica y del archivo de colocación de empleados selecciona una muestra de 25 y se encuentra que 7 no pasaron la prueba. ¿Se puede justificar esta crítica?

92. Una dieta especial asegura su efectividad en un 70% de los casos. Se selecciona una muestra de 30 personas que la han seguido, encontrando que en 18 de ellos ha sido efectiva. ¿Se puede asegurar que el porcentaje de efectividad es inferior? (Nivel del 5%).

93. La oficina de control de tránsito sostiene que el 30% de conductores de vehículos de servicio particular tienen el pase de conducción vencido. Se lleva a cabo una muestra de 20 conductores, encontrando que 9 de ellos tienen el pase vencido. ¿Al nivel del 10%, se puede afirmar que el porcentaje es mayor que el señalado por la oficina?

94. Un cirujano desarrolla una técnica quirúrgica nueva para una enfermedad, en la cual la mortalidad postoperatoria usual es del 25%. Aplica la técnica en 25 pacientes de los cuales mueren 8 ¿Cree usted que el cirujano considera que su técnica constituye un progreso con respecto al método de tratamiento anterior? ($\propto = 1\%$).

DISTRIBUCIÓN DE DIFERENCIAS ENTRE DOS MEDIAS MUESTRALES

De acuerdo con el *teorema del Límite Central*, cuando ambas variables presentan tamaños muestrales (n_1 *y* n_2) superiores a 30, tendrán un comportamiento similar a la *distribución normal*, por lo tanto aplicamos Z. En el caso, que ambos tamaños muestrales sean menores o iguales a 30, se aplicará la distribución "t" de Student, utilizando diferentes fórmulas, de acuerdo a la forma en que se de la información. Veámoslos, bajo el supuesto de que las dos poblaciones de donde se extraen las dos muestras, tienen varianzas poblacionales iguales ($\sigma_x^2 = \sigma_y^2$), supuesto que se cumple en la mayoría de los casos, lo cual permite calcular el error estándar de las diferencias entre las dos medias muestrales así:

A) Cuando el ejercicio suministra información para cada una de las observaciones muestrales, es preferible calcular una varianza común para las dos muestras, siendo:

$$s^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2} \quad \text{ó} \quad s^2 = \frac{\left[\sum x_i^2 - n_1 \bar{x}^2\right] + \left[\sum y_i^2 - n_2 \bar{y}^2\right]}{n_1 + n_2 - 2}$$

Con el resultado obtenido al aplicar la fórmula anterior, nos permite calcular el error estándar para la diferencia entre las medias muestrales.

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$$ Y la variante estadística será: $$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

Algunos ejemplos correspondientes a este primer procedimiento, se darán a continuación.

Ejemplo 1. Un fabricante de cigarrillos analiza el tabaco de dos marcas diferentes, para determinar el contenido en nicotina y obtiene los resultados siguientes (en miligramos):

MARCA	A:	24	26	25	22	23
MARCA	B:	27	28	25	29	26

¿Los resultados anteriores, señalan que existe una diferencia en el contenido medio de nicotina en ambas marcas?

o :

2

Siendo $s^2 = \frac{\sum (x_i - \bar{x}) + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2}$ $s_{\bar{x}-\bar{y}} = \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}$

$$\sum (x_i - \bar{x})^2 = 10 \qquad \sum (y_i - \bar{y})^2 = 10$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{120}{5} = 24 \qquad \bar{y} = \frac{\sum y_i}{n} = \frac{135}{5} = 27 \qquad s^2 = \frac{10+10}{5+5-2} = \frac{20}{8} = 2{,}5$$

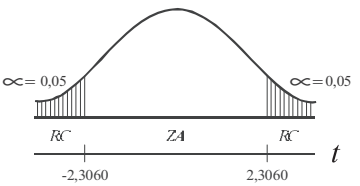
$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{2{,}5}{5} + \frac{2{,}5}{5}} = \sqrt{1} = 1$$

1) $H_0 : \mu_x = \mu_y$ $\quad$ $H_0 : \mu_1 = \mu_2$
$H_a : \mu_x \neq \mu_y$ $\quad$ $H_1 : \mu_1 \neq \mu_2$

2) $\propto = 0{,}05$ $\qquad$ 3) $s_{\bar{x}-\bar{y}} = 1$

4) $t = \frac{(24-27)}{1} = \frac{-3}{1} = -3$

$\upsilon = n_1 + n_2 - 2 \quad \Rightarrow \quad \upsilon = 5 + 5 - 2 = 8$



Los resultados anteriores, señalan que existe una diferencia significativa en el contenido medio de nicotina en ambas marcas, al nivel de 5%.

Ejemplo 2. Las calificaciones obtenidas a través de dos muestras a estudiantes diurnos y nocturnos, respectivamente fueron:

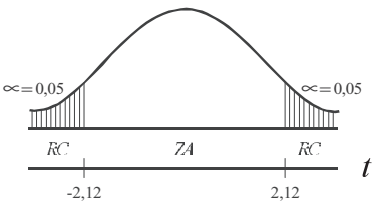
DIURNO:	3,5	4,2	3,8	3,6	4,0	3,6	4,2	4,8	4,0	3,6
NOCTURNO:	3,2	3,6	3,4	3,8	4,0	3,2	3,6	3,4		

¿Los anteriores resultados, permiten concluir, que hay una diferencia en el rendimiento de los estudiantes diurno y nocturno?

o :

3) $$s^2 = \frac{\left[155{,}89 - 10(3{,}93)^2\right] + \left[99{,}96 - 8(3{,}53)^2\right]}{10 + 8 - 2} = 0{,}11$$

4) $$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{3{,}93 - 3{,}53}{\sqrt{\frac{0{,}11}{10} + \frac{0{,}11}{8}}} = 2{,}54$$



Al nivel del 5%, se puede concluir que hay diferencias en el rendimiento entre los estudiantes según la jornada de estudio.

B) Ahora, cuando las ***o a a a t a a a a a*** éstas supuestamente están corregidas, es decir, fueron obtenidas aplicando las siguientes fórmulas:

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \frac{\sum x_i^2 - n_1 \bar{x}^2}{n_1 - 1} \quad \text{y} \quad s_y^2 = \frac{\sum (y_i - \bar{y})^2}{n_2 - 1} = \frac{\sum y_i^2 - n_2 \bar{y}^2}{n_2 - 1}$$

Estos dos resultados se pueden utilizar, para calcular el error estándar de la diferencia entre las dos medias muestrales, así:

$$s_{\bar{x}-\bar{y}} = \sqrt{\frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Por lo tanto la variante estadística en este caso, será:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Observemos que este procedimiento empleado, debe arrojar un resultado exactamente igual a aquél en donde se dan los valores para las variables muestrales, dado que las fórmulas son las mismas, pero presentadas en forma diferente. Veámoslo:

$$s^2 = \frac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2} = \frac{\cancel{(n_1 - 1)}\left[\frac{\sum (x_i - \bar{x})^2}{\cancel{n_1 - 1}}\right] + \cancel{(n_2 - 1)}\left[\frac{\sum (y_i - \bar{y})^2}{\cancel{n_2 - 1}}\right]}{n_1 + n_2 - 2}$$

Ahora, si anulamos a $(n_1 - 1)$ y $(n_2 - 1)$, nos queda que:

$s^2 = \dfrac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2}$, con lo cual se demuestra que los dos procedimientos de cálculo son iguales.

Ahora, consideremos la información del ejemplo 1. (pág., 361)

$\bar{x} = 24$ $\qquad$ $\bar{y} = 27$

$s_x^2 = \dfrac{\sum (x_i - \bar{x})^2}{n_1 - 1} = \dfrac{10}{4} = 2{,}50$ $\qquad$ $s_y^2 = \dfrac{\sum (y_i - \bar{y})^2}{n_2 - 1} = \dfrac{10}{4} = 2{,}5$

Nota: Hubo en este ejercicio coincidencia en los resultados de las varianzas, ya que en ambas variables resultaron iguales a 2,5.

Procedemos al cálculo del error de las diferencias entre las medias muestrales, siendo:

$s_{\bar{x}-\bar{y}} = \sqrt{\dfrac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}} \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ $\quad S_{\bar{x}-\bar{y}} = \sqrt{\dfrac{(5-1)2{,}5 + (5-1)2{,}5}{5+5-2}} \sqrt{\dfrac{1}{5} + \dfrac{1}{5}} = 1$

Como $t = \dfrac{\bar{x} - \bar{y}}{s_{\bar{x}-\bar{y}}}$ $t = \dfrac{24 - 27}{1} = \dfrac{-3}{1} = -3$, llegamos a los mismos resultados y conclusiones anteriores.

Ejemplo 2. Se toma como muestra 6 mujeres y 10 hombres fumadores. Se quiere saber si el número de cigarrillos que consumen los hombres diariamente es superior al de las mujeres, los datos fueron en promedio 8 cigarrillos en el grupo de mujeres y 11 en los hombres; las desviaciones típicas son 2,1 y 1,8 respectivamente. Al nivel del 5% ¿se puede llegar a la conclusión de que los hombres fuman más que las mujeres?

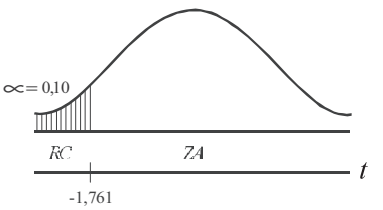
o :

$\bar{x} = 8$ $\qquad$ $\bar{y} = 11$ $\qquad$ $S_x = 2{,}1$ $\qquad$ $S_y = 1{,}8$

1) $H_0 : \mu_x = \mu_y$
$H_a : \mu_x < \mu_y$

2) $\propto = 0{,}05$ $\qquad$ $\upsilon = 6 + 10 - 2 = 14$

3) $t = \dfrac{\bar{x} - \bar{y}}{\sqrt{\dfrac{(n_1 - 1)s_x^2 + (n_2 - 1)s_y^2}{n_1 + n_2 - 2}} \quad \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$



4) $t = \dfrac{8 - 11}{\sqrt{\dfrac{(6-1)2{,}1^2 + (10-1)1{,}8^2}{6+10-2}} \sqrt{\dfrac{1}{6} + \dfrac{1}{10}}} = -3{,}04$

El resultado de -3,04 se ubica en la región crítica. Al nivel del 5%, se puede aceptar la conclusión de que los hombres fuman más que las mujeres.

Sin embargo, hay muchos casos en que las *varianzas poblacionales son desiguales* o no se puede comprobar su igualdad. En estos casos no se deben aplicar los dos procedimientos anteriores, pues no se ha encontrado una distribución precisa que la describa, tan sólo se han elaborado procedimientos aproximados, entre estos la "t" de Student con ciertos grados de libertad, mediante la aplicación de la siguiente fórmula:

$$\upsilon = \frac{\left(s_x^2 / n_1\right) + \left(s_y^2 / n_2\right)}{\frac{\left(s_x^2 / n_1\right)}{n_1 - 1} + \frac{\left(s_y^2 / n_2\right)}{n_2 - 1}}$$

Ejemplo 3. Consideremos los datos sobre dos tipos de amortiguadores para vehículos MAZDA, las pruebas de duración resultaron en dos muestras aleatorias, de tamaños 16 y 12 respectivamente, los primeros con una duración de 22,6 meses, y desviación típica de 7 meses, mientras que en el segundo es de 18,2 meses y desviación típica de 5,2 meses. No hay ninguna prueba de que las varianzas respecto a la duración sean iguales. Probar que la duración no presenta diferencias significativas, al nivel del 5%.

o :

Calculamos en primer lugar los grados de libertad.

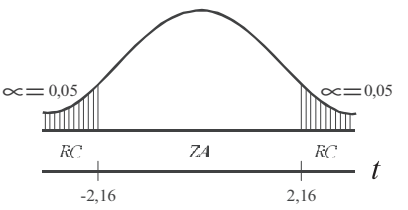
$$\upsilon = \frac{\left(\frac{7^2}{16} + \frac{5{,}2^2}{12}\right)}{\left[\left(\frac{7^2}{16}\right) \Big/ (16-1)\right] + \left[\left(\frac{5{,}2^2}{12}\right) \Big/ (12-1)\right]}$$

$$\upsilon = \frac{(49/16) + (27{,}04/12)}{\frac{(49/16)}{16-1} + \frac{(27{,}04/12)}{12-1}} = \frac{5{,}32}{\frac{3{,}06}{15} + \frac{2{,}25}{11}} = 13{,}02$$

Con un nivel de significación del 5%, el valor de "*t*" obtenido en la Tabla es de 2,16 para υ = 13 (grados de libertad).

1) $H_0 : \mu_x = \mu_y$ 2) $\propto = 0{,}05$
$H_a : \mu_x \neq \mu_y$

3) $t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} = \frac{22{,}6 - 18{,}2}{\sqrt{\frac{49}{16} + \frac{27{,}04}{12}}} = 1{,}91$



El resultado de 1,91, nos permite concluir que las diferencias sobre la duración de estos amortiguadores, no son significativas, al nivel del 5%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

95. El coeficiente de digestión, al utilizar maíz ensilado como alimento para ovinos, arrojó los siguientes resultados: $\bar{x}_1 = 58\%$ y $s_1^2 = 5\%$. El mismo resultado en bovinos arrojó: $\bar{x}_2 = 63\%$ y $s_2^2 = 8\%$. El número de animales, por muestra, fue de 15. Se pide determinar si hay diferencias entre los coeficientes de digestión de los ovinos y bovinos al nivel del 5%.

96. Los resultados del control de la calidad en dos procesos manufactureros son los siguientes:

TIPO I:	1,5	2,5	3,5	2,5		
TIPO II:	2,5	3,0	3,0	4,0	3,5	2,0

¿Permiten estos resultados concluir que los precios den como resultado calidades diferentes? (Nivel de 10%)

97. Se propone un nuevo método para fabricar concreto. Para docimar (someter a prueba) si el nuevo método ha aumentado la resistencia a la compresión, se elaboran cinco bloques con cada método. La resistencia a la compresión, en libras por pulgada cuadrada, resultan ser las siguientes:

NUEVO M TODO:	154	143	132	147	139
M TODO USUAL:	144	131	155	126	134

¿Diría usted que el nuevo método ha aumentado la resistencia a la compresión?

98. La tabla siguiente presenta la ganancia en pesos en dos lotes de ratas:

DIETA	RICA EN PROTEINA	POBRE EN PROTEINAS
Número de ratas	12	7
Ganancia media (gramos)	120	101
Suma de cuadros de las desviaciones respecto a la media	5.032	2.552

¿Indican estos datos una diferencia significativa entre las ganancias medias en peso correspondiente a ambos grupos?

99. Se requiere comparar el nivel salarial de los empleados de dos empresas. Se solicita dicha información a la primera de ellas y reporta que en una muestra aleatoria de 17 empleados, su ingreso promedio fue de $428.000, con una desviación estándar de $80.000. Se elige otra muestra aleatoria de 12 empleados de la segunda empresa y sus salarios son $338; $393; $416; $363; $375; $420; $447; $412; $510; $436; $476 y $380 (miles). ¿Presentan los datos anteriores evidencia suficiente que indique una diferencia en el nivel salarial de los empleados de las dos empresas?

100. A continuación se presentan las cifras correspondientes al análisis de proteínas realizadas sobre una misma variedad de trigo cosechada en dos distritos. Si éstas son las únicas cifras disponibles, docime (someta a prueba) si existe o no una diferencia significativa entre la proteína promedio de los dos distritos:

DISTRITO 1:	12,6	13,4	11,9	12,8	13,0		
DISTRITO 2:	13,1	13,4	12,8	13,5	13,3	12,7	12,4

101. Dos muestras de 15 especímenes cada una, de dos tipos de género de lana fueron sometidos a una prueba de resistencia. La media de la primera muestra, fue de 131 libras por pulgada cuadrada y de 136 la de la segunda. Las respectivas desviaciones estándar fueron 6,25 y 4,65 libras por pulgada cuadrada. ¿Indica lo anterior que la segunda calidad de género es superior?

102. Una compañía desea comprar una máquina que fabrica neumáticos para llantas de bicicleta. Ha recibido ofertas de dos productores y quiere determinar si le es indiferente comprar la maquinaria A o la B. Basándose en la producción de las máquinas de cada uno de ellos. Durante 10 días, elegidos al azar, observa la producción de la máquina ofrecida por A y obtiene que:

$\sum x_i = 250 \quad \sum x_i^2 = 6.500$. Donde X es el número de neumáticos producidos por A. Para la máquina que ofrece B se observa su producción durante 15 días escogidos al azar y se obtiene que:

$s_y^2 = 35 \quad \frac{\sum y_i^2}{n_2^2} = 80$ Siendo (y) el número de neumáticos producidos por esa máquina. Usted es el gerente de la compañía, tome una decisión con un nivel de significación del 5%.

103. Una compañía está tratando de decidir cuál de dos tipos de neumáticos va a comprar para sus camiones. Le gustaría adoptar la marca B a menos que haya alguna evidencia de que la marca A sea mejor. Se hace un experimento en el que se utilizan 16 neumáticos de cada marca. Los neumáticos se prueban bajo condiciones iguales hasta que se desgastan; los resultados son:

MARCA A:	$\bar{x}$ = 26.000 km	s_x = 4.200 km
MARCA B:	$\bar{y}$ = 25.000 km	s_y = 2.800 km

A un nivel del 5% existe alguna evidencia de que la marca A sea superior a B?

104. Dos tipos de soluciones químicas, A y B, fueron ensayadas para ver su PH (grado de acidez de la solución). Análisis de 20 muestras de A dieron un PH de 7,5 con una s = 0,24. Análisis de 12 muestras de B dieron un PH de 7,4 con una s = 0,30. Docimar si existe una diferencia significativa. ($\propto$ = 5%)

105. De un estudio efectuado en 22 animales, a once de ellos se les dio proteína de maní crudo y a los otros once proteínas de maní tostado. A un nivel del 5%, pruebe la hipótesis de que la proteína de maní tostado ha tenido menor efecto sobre el valor proteínico circulante.

TIPO DE MANÍ	PROTEÍNAS CIRCULANTES (A B)										
a) Crudo	44	63	63	59	61	61	60	56	56	59	56
b) Tostado	61	59	63	55	58	54	47	51	57	54	54

106. El cuociente de inteligencia de 16 estudiantes de secundaria, jornada nocturna, dio un promedio de 112 con una varianza de 64, mientras que para 14 estudiantes del mismo colegio, de jornada diurna, dio un promedio de 107 con una desviación estándar de 10. ¿Existe una diferencia entre el I.Q. de los dos grupos?

107. Con base en los registros de personal, dos grandes compañías toman muestras relativas a la edad que se pensiona su personal. En ambas compañías se consideran muestras aleatorias de 25 ex-empleados. Las compañías encuentran que:

$$\sum x_i^2 = 44.500 \qquad \bar{x} = 40 \qquad \sum y_i^2 = 78.000 \qquad \bar{y} = 50 \quad (\text{Nivel del } 5\%)$$

¿Cree usted que los empleados de la compañía B se pensionan a una edad mayor que los de A?

108. Un inversionista desea decidir entre dos tipos de valores. Para su decisión cuenta con la siguiente información acerca del rendimiento expresado como porcentaje del precio:

VALOR A:	7,9	8,3	6,3	8,2	7,3	7,6	10,3	8,9
VALOR B:	9,6	9,1	5,8	10,3	9,4	8,2		

¿Sugieren estos datos que el valor B produce mayor utilidad que el valor A?

109. Una prueba de lectura fue rendida por los alumnos del primer curso de una escuela elemental. El resultado consistió en que de 10 alumnos que aprendieron a leer con el método A, su promedio fue de 3,5 y desviación típica de 0,5; con 12 alumnos por el método B, su promedio de calificación fue de 3,7 y desviación estándar de 0,4. ¿Al nivel del 5%, se podría afirmar que el método B es más efectivo que el A?

110. El gerente de una empresa, asegura que su programa de entrenamiento en ventas le permite un promedio diario en venta superior a $110.000 por vendedor; selecciona una muestra de 10 agentes que han recibido el programa de entrenamiento y obtiene los siguientes resultados (en miles de $):

100	90	120	130	130	110	125	122	115	135

a) Si usted fuera el superior de estos agentes, ¿adoptaría el programa de entrenamiento propuesto por el gerente? ($\propto = 5\%$).
b) Si el verdadero valor promedio diario de ventas de la empresa es de $112.000, ¿qué tipo de error se comete, si se docima al 1% que el promedio es diferente?
c) Si usted selecciona 14 vendedores de la empresa sin someterse a dicho programa de entrenamiento y compara dicha muestra con la de los diez (10) que recibieron entrenamiento, ¿puede afirmarse, a un nivel de significación del 5%, que dicho programa no produce los efectos que sostiene el gerente?

80	110	130	115	120	110	100	70
130	75	140	85	120	95	(miles de $)	

111. Un fabricante registró el número de artículos producidos diariamente, durante 10 días, para un grupo de 15 obreros que operaban con base en un plan de salarios fijos. El fabricante introdujo un plan de incentivos salariales para otros 15 obreros y registró la producción durante 10 días. El número de artículos producidos diariamente fueron:

PRIMER GRUPO:	75	76	74	80	72	78	76	73	72	75
SEGUNDO GRUPO:	86	78	86	84	81	79	78	84	88	80

Suponiendo que los salarios pagados a cada grupo fueron iguales, ¿puede el fabricante concluir que el plan de incentivos fue efectivo?

112. Una prueba de resistencia al esfuerzo de dos tipos diferentes de cable, se llevó a cabo, seleccionando dos muestras de tamaños 15 y 20, respectivamente; las medias fueron de 905 y 925, con desviaciones típicas de 25 y 35. ¿Proporcionan estos datos, al nivel del 10%, suficiente evidencia para indicar que la resistencia al esfuerzo de los cables de B es superior a la de A?

113. En un laboratorio de psicología hicieron llegar, por diferentes conductos, una sustancia tóxica hasta el sistema nervioso central en varios animales. La variable de interés fue el tiempo, en horas, entre la administración de la toxina y la iniciación de los síntomas. Se obtuvieron los siguientes datos:

CONDUCTOS	n	$\bar{x}$	VARIANZA
A	11	40	10,0
B	7	31	20,9

¿Proporcionarán estos datos evidencia como para concluir, en promedio, que la iniciación de los síntomas operó pronto cuando la toxina se administró por el conducto B? (Nivel del 5%).

DISTRIBUCIÓN DE DIFERENCIAS ENTRE DOS PROPORCIONES MUESTRALES $_{1 \quad 2}$

En la aplicación de estas distribuciones en *muestras pequeñas,* se sigue el mismo procedimiento utilizado para muestras grandes, con la diferencia en la presentación de la fórmula, en la cual se trabaja con $n_1 - 1$ y $n_2 - 1$

$$t = \frac{p_1 - p_2}{\sqrt{\frac{p_1 q_1}{n_1 - 1} + \frac{p_2 q_2}{n_2 - 1}}}$$

Ejemplo 1. En una muestra probabilística de l2 amas de casa, el 20% indicó preferencia por la marca A de margarina. Con posterioridad a una campaña intensiva de radio y televisión, se seleccionó una nueva muestra entre amas de casa del mismo tamaño y clase social. En esta muestra el 22% indicó preferencia por la marca A. De acuerdo con estos resultados y a un nivel de 5%, ¿podría rechazarse la hipótesis de que la campaña de publicidad no fue efectiva?

o :

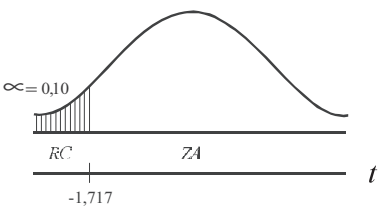
$n_1 = 12 \qquad n_2 = 12 \qquad p_1 = 0{,}20 \qquad p_2 = 0{,}22 \qquad \propto = 0{,}05$

1) $H_0 : \mu_{\bar{p}_1} = \mu_{\bar{p}_2}$ 2) $\propto = 0{,}05$

$Ha : \mu_{\bar{p}_1} < \mu_{\bar{p}_2}$

3) $S_{\bar{p}1-\bar{p}2} = \sqrt{\frac{p_1 q_1}{n_1 - 1} + \frac{p_2 q_2}{n_2 - 1}}$

4) $t = \frac{0{,}2 - 0{,}22}{\sqrt{\frac{(0{,}2)(0{,}8)}{12} + \frac{(0{,}22)(0{,}78)}{12}}} = -0{,}12$



Se ubica − 0,12 en la zona de aceptación. Al nivel del 5%, se puede rechazar la hipótesis que la campaña publicitaria no fue efectiva.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

114. El Jefe de cartera de una compañía asegura, que utilizando la llamada telefónica logra acelerar más que una carta de cobro, el pago de cuentas morosas. Dos grupos de clientes morosos fueron contactados por cada método, fijando un plazo no mayor de 10 días para recibir el pago. En el primer grupo conformado por 20 personas, se obtuvieron respuestas de 14: en el segundo grupo, se tomaron 16 clientes morosos y sólo respondieron 10. ¿A un nivel de significancia del 0,01, debe concluir la compañía que el jefe de cartera tiene razón?

115. En una empresa se considera que la proporción de permisos para ausentarse de la oficina en horas de trabajo, es mayor en los hombres que en las mujeres. Para comprobar la anterior afirmación se escogió una muestra de 20 hombres y 24 mujeres, observando que el número de permisos solicitados y concedidos fue de 14 y 10 respectivamente. La investigación se realizó sobre el total de permisos en un mes. ¿Al nivel del 5%, es válida la afirmación?

116. Un distribuidor de acuerdo a su experiencia, dice que existe igual preferencia por la marca de computador que ofrece, dependiendo de la Universidad en la que estudia el consumidor. Con base en una muestra aleatoria en 25 alumnos de la primera Universidad, se encuentra que 20 prefieren dicha marca, mientras que en una muestra de 16 alumnos en la segunda Universidad, sólo 10 la prefieren. ¿Es aceptable la afirmación de igual preferencia por la marca del computador?

117. Una embutidora de carne ofrece para el consumo, dos clases de productos diferentes por el contenido de grasa. Se tomaron dos muestras teniendo en cuenta la clase del producto, para preguntar si la aceptación dependía del contenido de grasa. En el primero, en una muestra de 16 consumidores, 6 compraron poco contenido; en la muestra de 10, sólo 3 no tuvieron en cuenta el contenido de grasa. ¿Es válida la afirmación de que la preferencia depende del nivel de grasa?

APLICACIÓN CON EXCEL

PRUEBA DE IPOTESIS

Consideremos en primer lugar la aplicación de Excel en las pruebas de hipótesis, en una distribución de MEDIA MUESTRALES, cuando es bilateral, con un nivel de significancia del 5% (α=0.05), para la característica que hemos venido utilizando, es decir, CALIFICACIONES obtenidas por los alumnos en las pruebas realizadas por el ICFES. La hipótesis nula, es decir, la media poblacional supuesta es de 310 (μ=310), frente a la alternativa de que sea diferente.

Muestra grande: **> 30 → ( )**

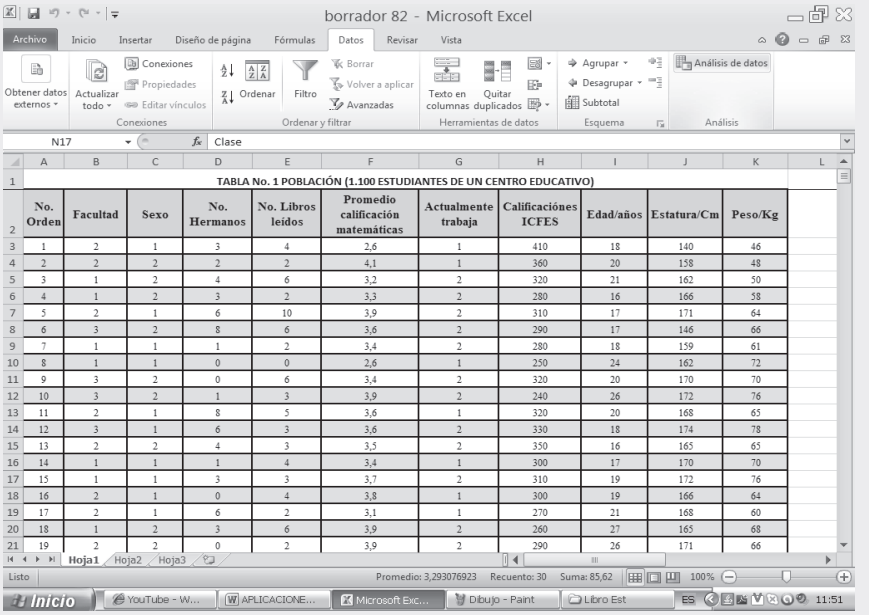
El tamaño de la muestra será de N=80 ubicada en la COLUMNA H, donde la identificación de la variable aparece en la CELDA H2, los datos van desde H3 hasta H82 (Ver Tabla No. 1 en el SIL).

1) $H_o : \mu = 310$ 2) $\alpha = 0,05$
$H_a : \mu \neq 310$

Nota: Ver información y todo el proceso en el SIL, segunda parte.

- En los procedimientos anteriores, se hará CLIC en la ficha DATOS-ANÁLISIS DE DATOS

Figura No. 1. Análisis de Datos



- Seleccionamos ESTADÍSTICA DESCRIPTIVA y hacemos CLIC en ACEPTAR.
- En el cuadro desplegable ESTADÍSTICA DESCRIPTIVA, teclearemos en la primera casilla RANGO DE ENTRADA H2:H82.
- Seleccionamos RÓTULOS EN PRIMERA ENTRADA, lo mismo en lo que respecta a RESUMEN DE ESTADÍSTICA, haciendo CLICK en ambos casos.
- Establecemos el RANGO DE SALIDA, tecleando por ejemplo N3 y luego haremos CLIC en ACEPTAR, tal como se puede observar a continuación:

Figura No. 2. Estadística Descriptiva.

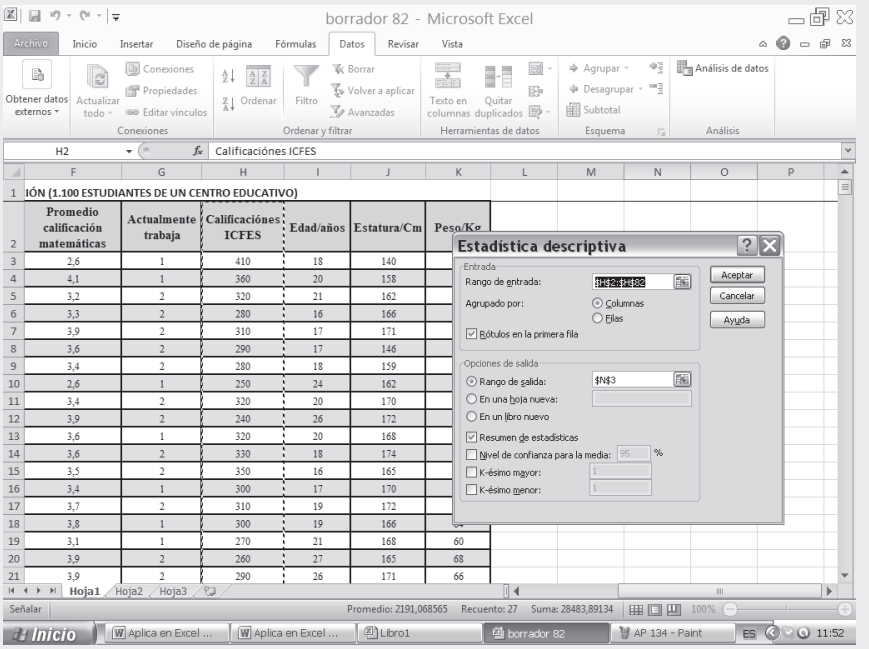
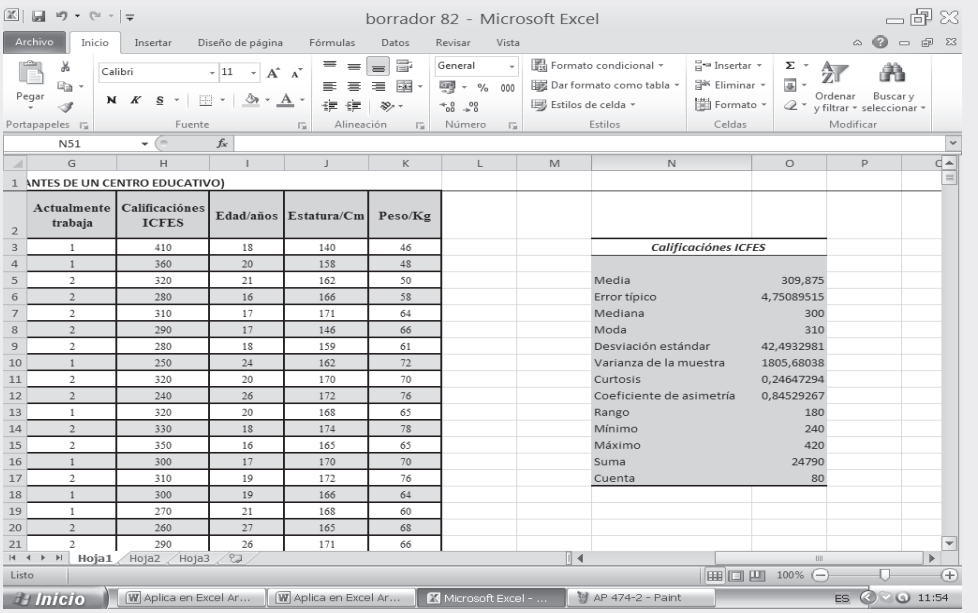
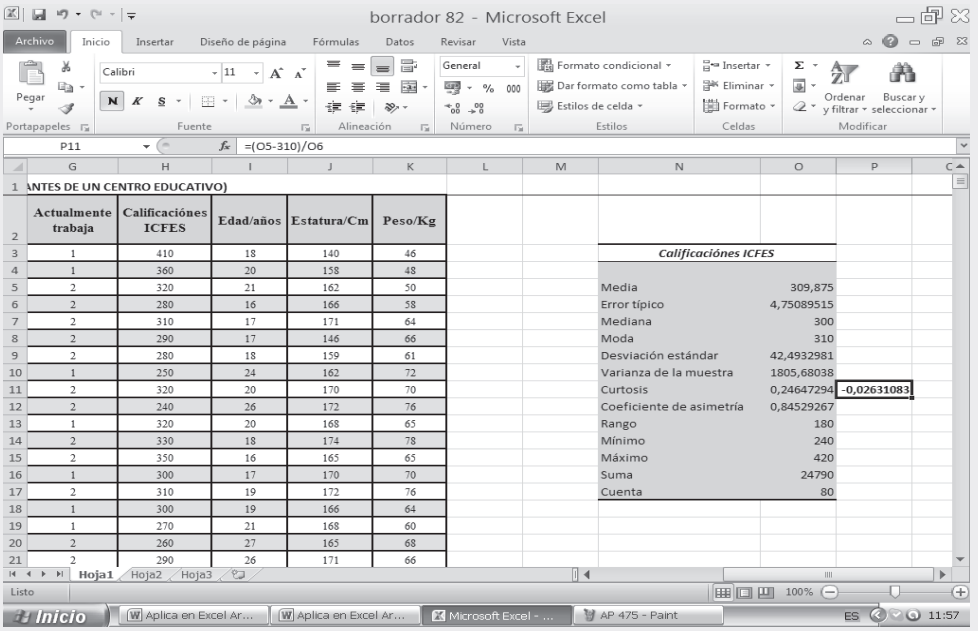


Figura No. 3. Cuadro de Resultados.



- El valor de la MEDIA ARITMÉTICA (x) aparecerá en la CELDA N5 acompañada de resultados, correspondientes a otras medidas.
- Ahora procedemos a seleccionar la CELDA O11 en la cual tecleamos la fórmula =N5-310/N6.

Figura No. 4. Cálculo del Valor Z



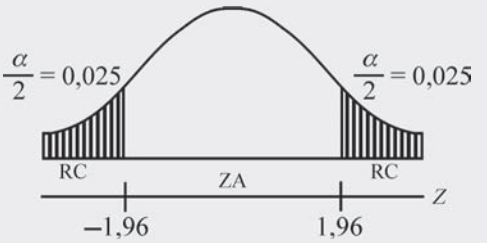
La fórmula estadística que se utilizó para el cálculo es: $$Z = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Los resultados obtenidos, serán utilizados para ser reemplazados en la fórmula

$$\text{Siendo} \quad Z = \frac{(309{,}9 - 310)}{\frac{42{,}49}{\sqrt{80}}} = \frac{-0{,}1}{4{,}75} = -0{,}022$$

$$\bar{x} = 309{,}9 \qquad \frac{S}{\sqrt{n}} = 4{,}75 \qquad \mu = 310$$

El resultado de Z será aproximadamente igual a – 0,022



Observamos que – 0,022 cae en la región aceptable o zona de aceptación, lo que indica que estaremos aceptando la hipótesis nula (H_0).

Muestras peque as ≤ 30 (t)

Corresponde a muestras de $n \leq 30$ y la prueba con aplicación de EXCEL, es casi igual al procedimiento indicado para muestras grandes ($n > 30$), tal como se verá a continuación

$$H_o: \mu = 310$$
$$H_a: \mu \neq 310$$

$$\alpha = 0{,}05 \qquad t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

- Hacemos CLIC en ANÁLISIS DE DATOS de la ficha DATOS-ESTADÍSTICA DESCRIPTIVA.
- En el RANGO DE ENTRADA solamente tomaremos los primeros 20 datos de la variable CALIFICACIONES ICFES, siendo H2 el título de la columna y de H3:H22 el rango de datos a seleccionar.
- Seleccionamos RÓTULO PRIMERA FILA y RESUMEN DE ESTADÍSTICA.
- Establecemos el RANGO DE SALIDA, en nuestro caso será M3.
- Finalmente hacemos CLIC en ACEPTAR.

Nota: En la segunda parte del SIL se encuentra la Tabla No.1 con la información completa incluida la columna de CALIFICACIONES ICFES.

Figura No.5. Selección de Análisis de Datos.

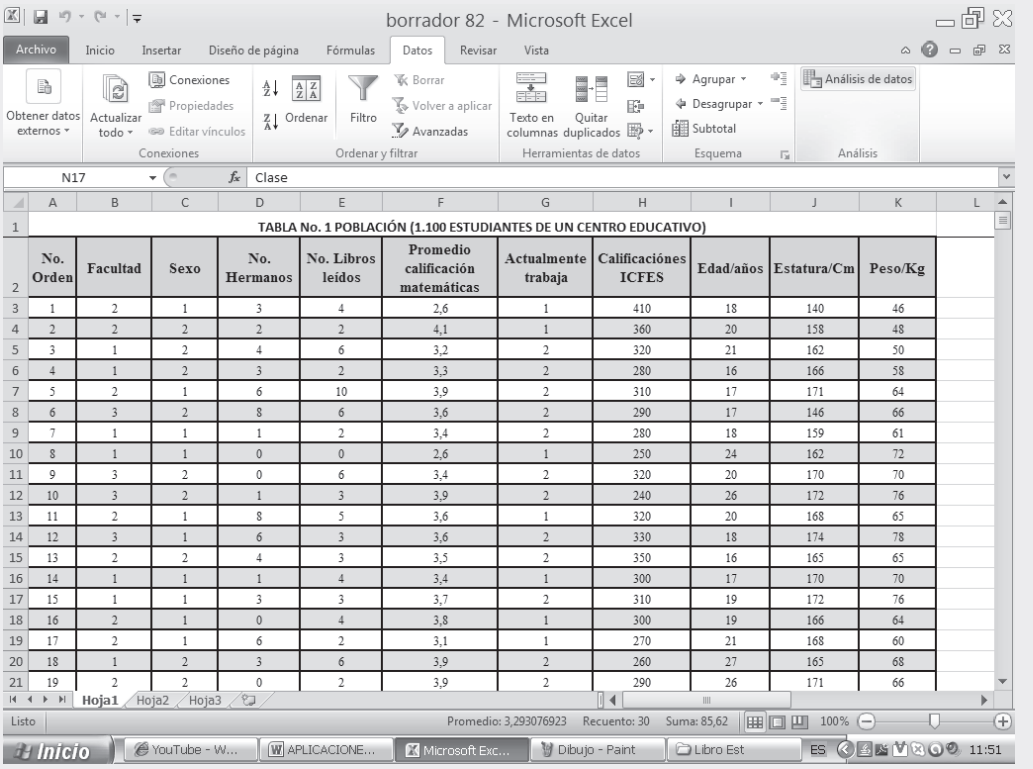


Figura No.6. Estadística Descriptiva – Datos de la Muestra.

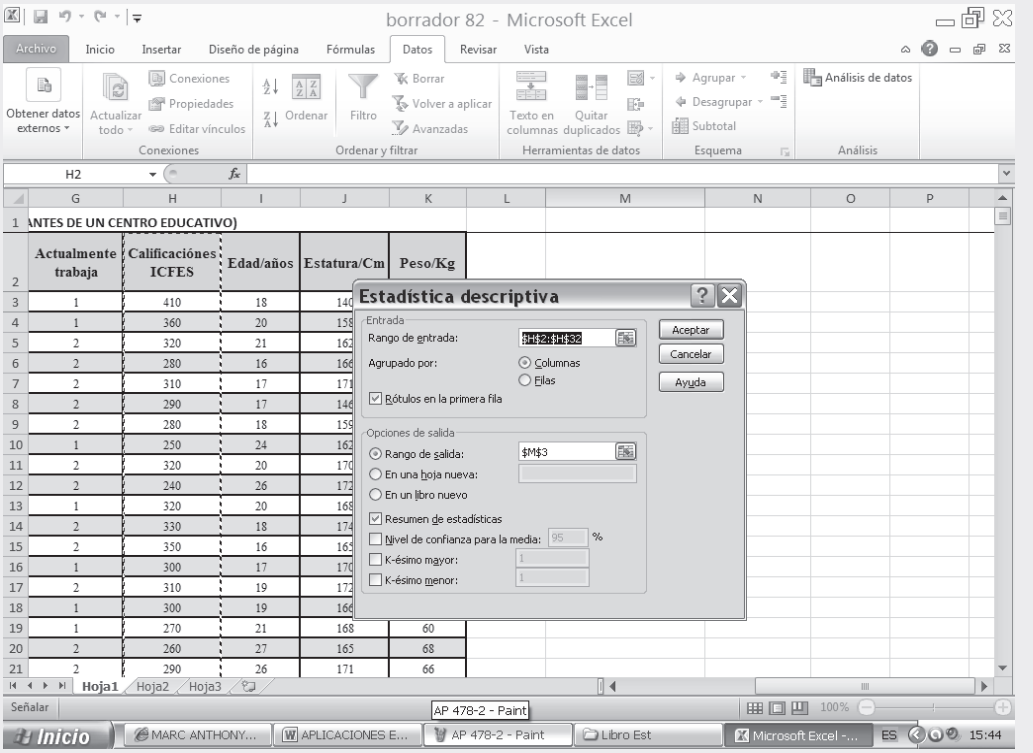
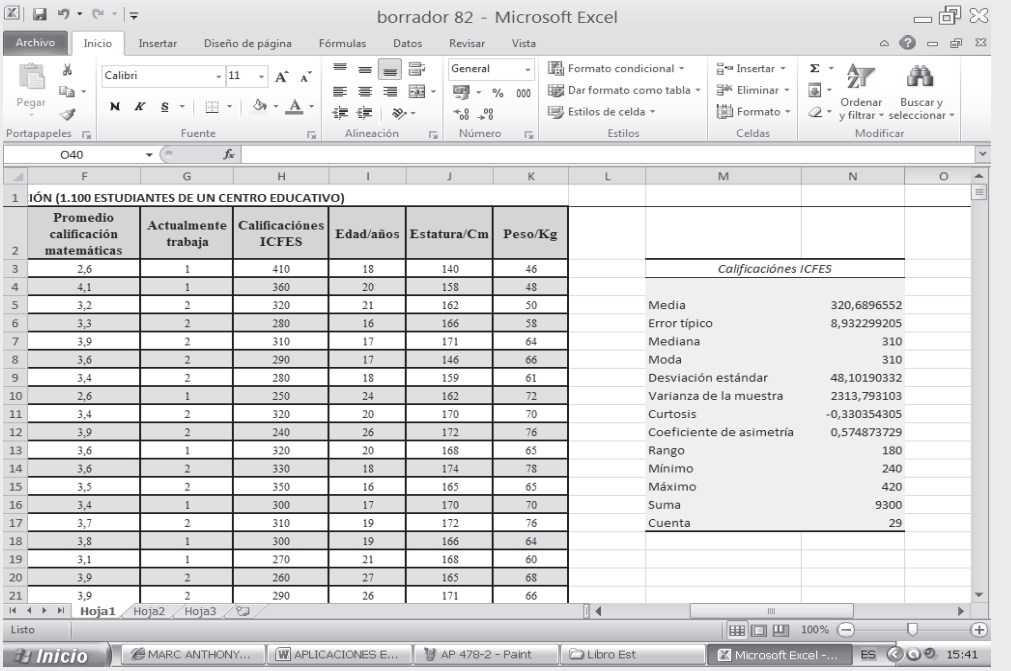


Figura No. 7. Cuadro de Resultados.



borrador 82 - Microsoft Excel

IÓN (1.100 ESTUDIANTES DE UN CENTRO EDUCATIVO)

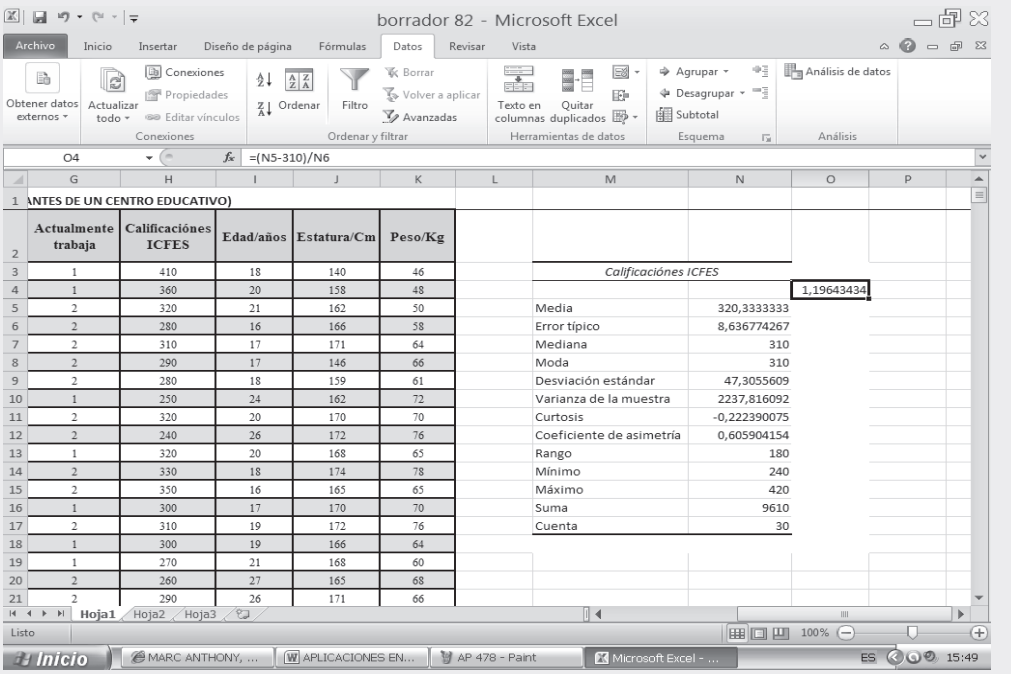
Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
2,6	1	410	18	140	46
4,1	1	360	20	158	48
3,2	2	320	21	162	50
3,3	2	280	16	166	58
3,9	2	310	17	171	64
3,6	2	290	17	146	66
3,4	2	280	18	159	61
2,6	1	250	24	162	72
3,4	2	320	20	170	70
3,9	2	240	26	172	76
3,6	1	320	20	168	65
3,6	2	330	18	174	78
3,5	2	350	16	165	65
3,4	1	300	17	170	70
3,7	2	310	19	172	76
3,8	1	300	19	166	64
3,1	1	270	21	168	60
3,9	2	260	27	165	68
3,9	2	290	26	171	66

Calificaciónes ICFES

Media	320,6896552
Error típico	8,932299205
Mediana	310
Moda	310
Desviación estándar	48,10190332
Varianza de la muestra	2313,793103
Curtosis	-0,330354305
Coeficiente de asimetría	0,574873729
Rango	180
Mínimo	240
Máximo	420
Suma	9300
Cuenta	29

- Como se puede observar, el cuadro de resultados aparece a partir de las celdas M5 y N5. Por ejemplo, para el caso de la MEDIA MUESTRAL, éste aparece en la Celda M5 y su valor corresponde a 320.69 ubicado en la celda N5.
- Finalmente, seleccionamos la CELDA P5 y en ella tecleamos la fórmula t=(N5-310)/N6, en la CELDA P5 debe aparecer el valor estadístico de "t".

Figura No. 8. Cálculo del Valor del Estadístico t.



borrador 82 - Microsoft Excel

=(N5-310)/N6

ANTES DE UN CENTRO EDUCATIVO)

Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	410	18	140	46
1	360	20	158	48
2	320	21	162	50
2	280	16	166	58
2	310	17	171	64
2	290	17	146	66
2	280	18	159	61
1	250	24	162	72
2	320	20	170	70
2	240	26	172	76
1	320	20	168	65
2	330	18	174	78
2	350	16	165	65
1	300	17	170	70
2	310	19	172	76
1	300	19	166	64
1	270	21	168	60
2	260	27	165	68
2	290	26	171	66

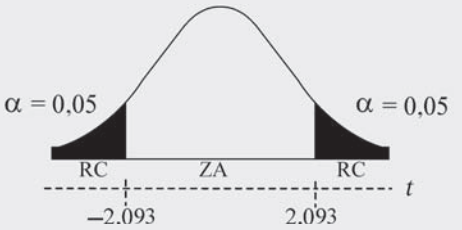
Calificaciónes ICFES

1,19643434

Media	320,3333333
Error típico	8,636774267
Mediana	310
Moda	310
Desviación estándar	47,3055609
Varianza de la muestra	2237,816092
Curtosis	-0,222390075
Coeficiente de asimetría	0,605904154
Rango	180
Mínimo	240
Máximo	420
Suma	9610
Cuenta	30

Siendo: $t = \dfrac{(\bar{x}-\mu)}{\dfrac{S}{\sqrt{n}}}$ $\quad t = \dfrac{(320.33-310)}{\dfrac{47.30}{\sqrt{20}}} = \dfrac{10.33}{8,63} = \cong 1,19$

Se concluye aceptando la hipótesis nula H_0, ya que 1,19 cae en la zona de aceptación



PRUEBA IPÓTESIS ENTRE DOS MEDIAS MUESTRALES

Ho: μ=310 $\quad \alpha = 0,05$ $\quad Z = \dfrac{(\bar{x}-\bar{y})}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}$

Ha: μ≠ 310

Muestras **grandes (z)**

Digitamos la siguiente información en una hoja de cálculo, correspondiente a dos (2) muestras de tamaños $n_1 = 32$ y $n_2 = 40$ acerca de la duración en miles de kilómetros de dos (2) marcas de llantas para automóviles.

Cuadro No. 1. Muestra Sobre la Duración de las Llantas en Kilómetros.

Marca A	43	38	36	42	39	56	85	25	45	85	96	75	36	15	78	87	98	96	96	95	35	48	56	35	45	57	48	25	36	45	25	36			
Marca B	42	45	65	35	48	68	45	15	45	35	32	43	47	48	65	35	15	25	35	42	58	65	26	35	34	39	34	44	48	49	47	49	52	35	24	28	45	59	45	48

Figura No.9. Información Digitada en Hoja de Cálculo.

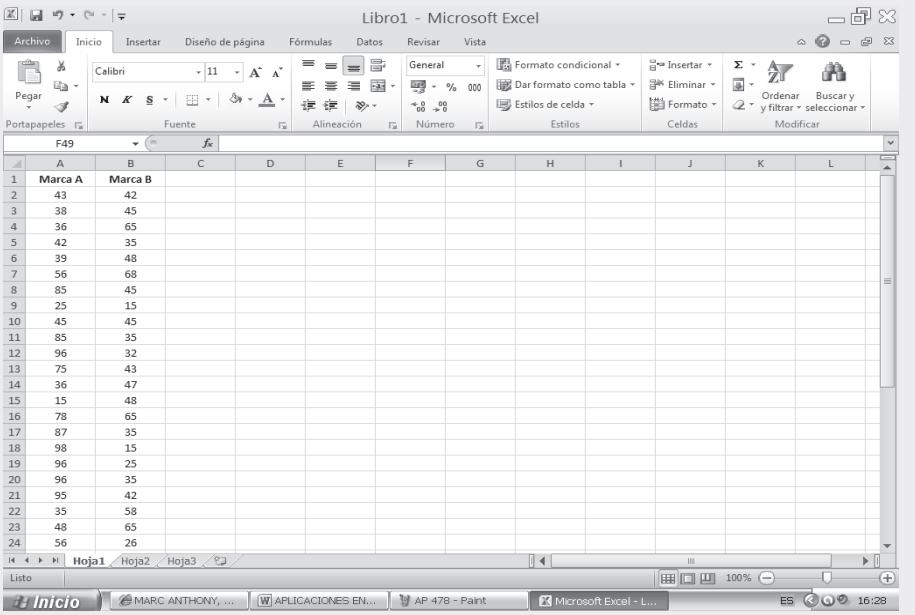
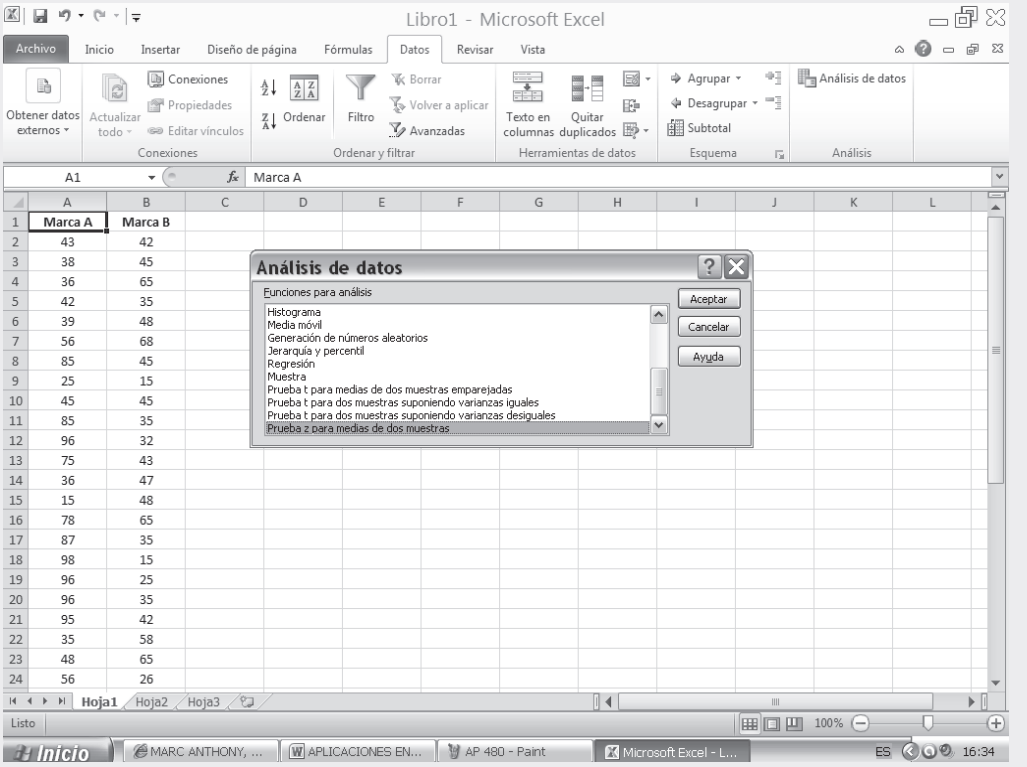


Figura No. 10. Análisis de Datos.



- Seleccionamos la ficha INSERTAR-ANÁLISIS DE DATOS.
- En ANÁLISIS DE DATOS elegimos PRUEBA Z PARA MEDIAS DE DOS MUESTRAS y hacemos CLICK en aceptar.
- En los campos correspondientes a RANGO PARA LA VARIABLE, elegimos los datos de las celdas Marca A y Marca B según sea el caso.
- Dado que la hipótesis nula planteada corresponde a Ho: μ_x μ_y , es decir, que $\mu_x - \mu_y = 0$ por lo tanto en el espacio dado para DIFERENCIA HIPOTÉTICA ENTRE LAS MEDIAS, tecleamos 0.
- En este procedimiento se considera que las varianzas poblacionales estimadas que sirven a los campos VARIANZA PARA LA VARIABLE 1 y 2 son: $\sigma_x^2 = 65{,}36$ y $\sigma_y^2 = 51{,}82$.
- Procedemos a seleccionar RÓTULOS.
- En cuanto al nivel de significancia casi siempre se trabaja con un nivel del 5% que corresponde al valor que vamos a digitar en el campo sugerido para ALFA.
- Finalmente seleccionamos el RANGO DE SALIDA, en nuestro caso corresponde a la CELDA C3.

Figura No. 11. Prueba Z para Medias de Dos Muestras

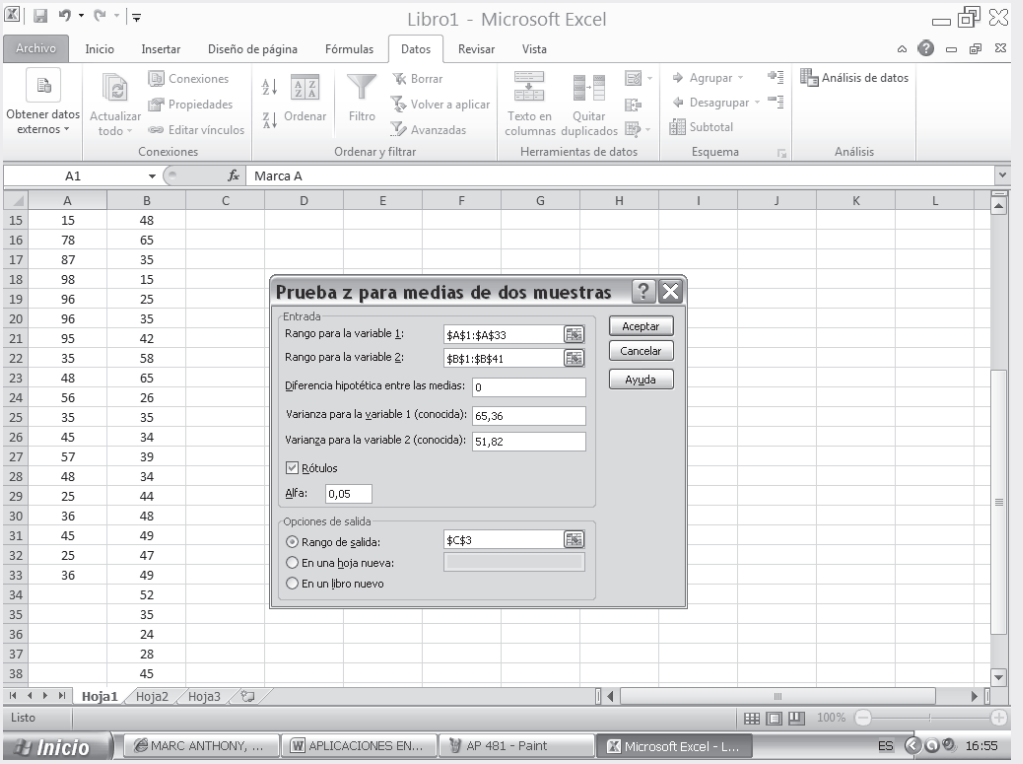
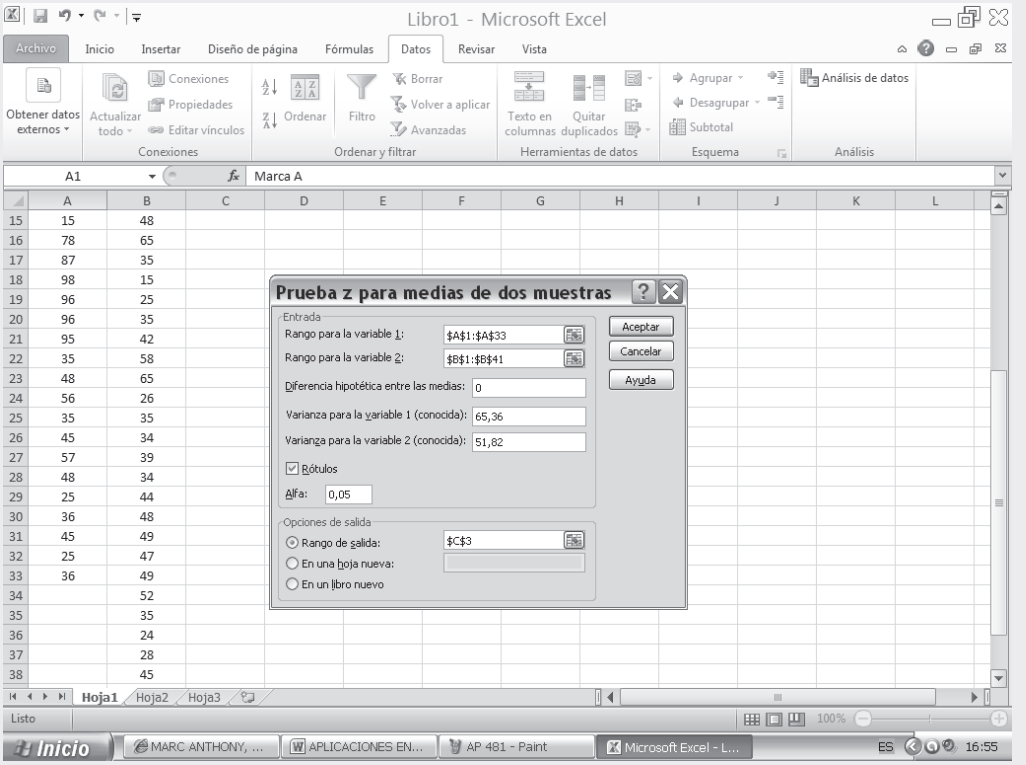
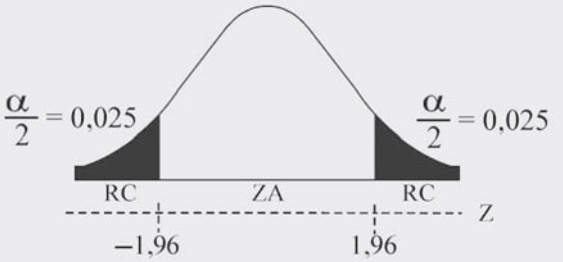


Figura No. 12. Cuadro de Resultados.



Como Z es igual a 6.87, cae en la zona de rechazo, por lo tanto al nivel del 5% rechazamos H_o. Es decir, que las llantas para automóviles presentan una duración diferente en kilómetros según la marca.



Muestras peque as (t)

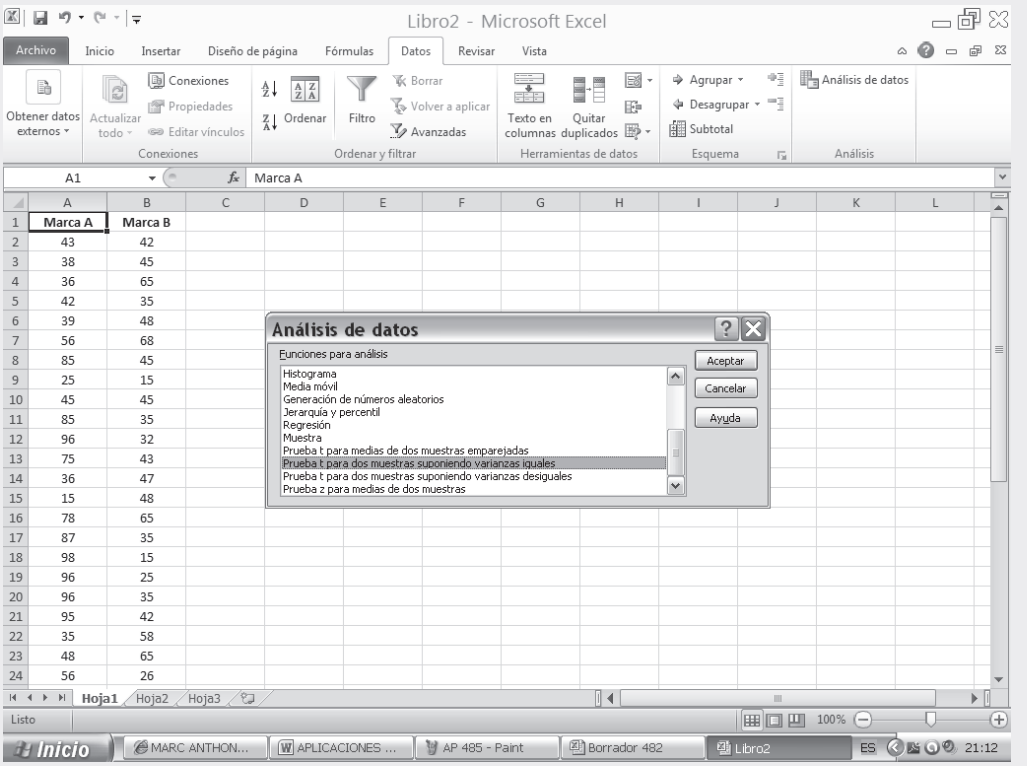
Tomando como base la tabla propuesta del ejemplo anterior y seleccionando sólo n_1=12 y n_2=16, de los datos allí incluidos, desarrollaremos ahora la prueba bilateral para pruebas pequeñas.

$Ho: \mu_x = \mu_y$ $\quad$ $\alpha = 0{,}05$ $\quad$ $t = \dfrac{(\bar{x} - \bar{y})}{\sqrt{\dfrac{s^2}{n_1} + \dfrac{s^2}{n_1}}}$

$Ha: \mu_x \neq \mu_y$

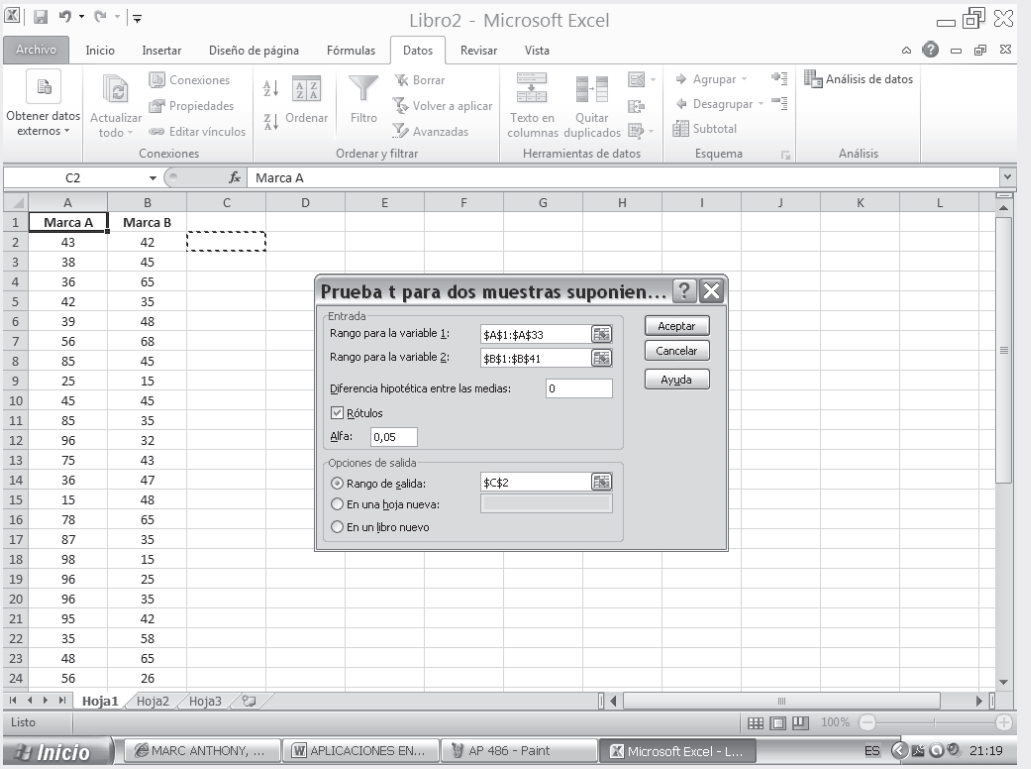
- Se calcula una VARIANZA MUESTRAL para ambas muestras tal como se presenta en el proceso siguiente.
- En la Ficha DATOS-ANÁLISIS DE DATOS seleccionamos la función PRUEBA "t" para media de dos (2) muestras suponiendo varianzas iguales.

Figura No. 13. Análisis de Datos



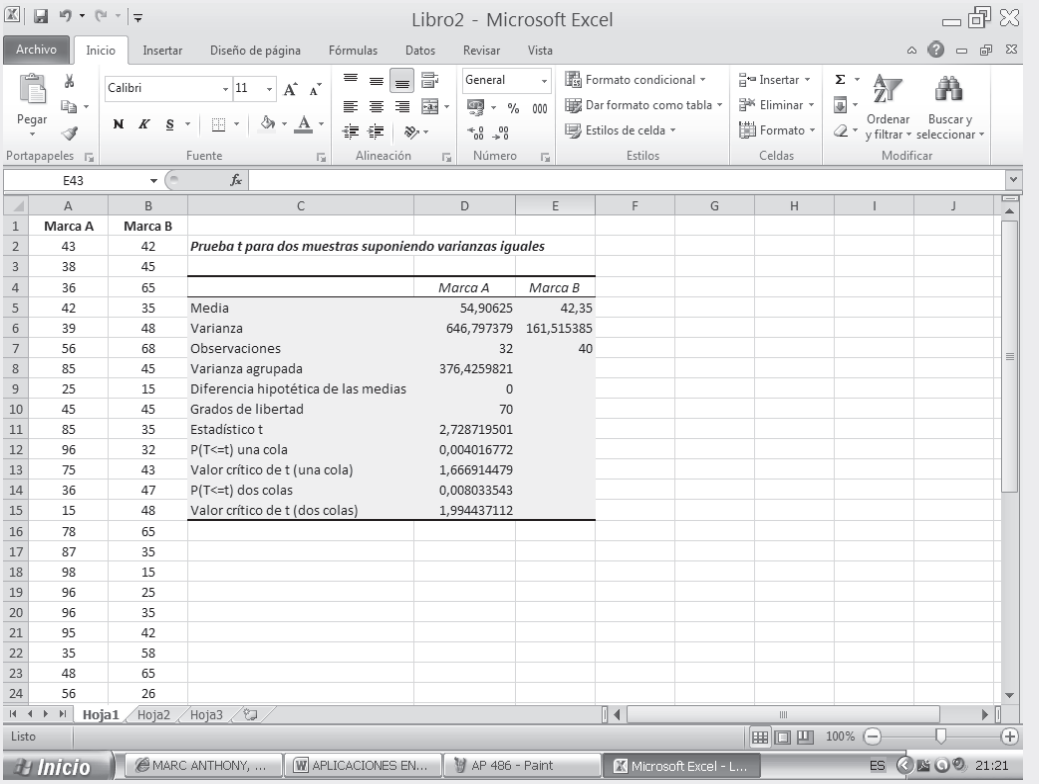
- En el cuadro de diálogo que aparece, se establece en los dos primeros campos la selección de cada una de las columnas utilizadas para este ejemplo, en este caso, primero seleccionamos el rango de datos de la columna A denominada Marca A y luego el rango correspondiente a la columna B que denominamos como Marca B.
- La hipótesis nula (H_0) corresponde a $\mu_x - \mu_y = 0$, por lo tanto tecleamos cero (0) en el campo DIFERENCIA HIPOTÉTICA ENTRE LAS MEDIAS.
- Se preselecciona RÓTULOS.
- En la casilla ALFA digitamos 0.05, dado que el NIVEL DE SIGNIFICACIÓN será $\propto = 0{,}05$.
- Finalmente procedemos a seleccionar el RANGO DE SALIDA, en nuestro caso tecleamos la celda D3 en la casilla correspondiente y hacemos CLIC en ACEPTAR, obteniéndose el resultado para t.

Figura No. 14. Prueba t para Dos Muestras



■ **En la SEGUNDA PARTE del SIL que acompaña al libro, encontrará en conjunto toda la aplicación del EXCEL a la Estadística, que igual a1 capítulo 13, ofrece una BASE DE DATOS para trabajar.**

Figura No. 15. Cuadro de Resultados.



	A	B	C	D	E
1	Marca A	Marca B			
2	43	42	*Prueba t para dos muestras suponiendo varianzas iguales*		
3	38	45			
4	36	65		*Marca A*	*Marca B*
5	42	35	Media	54,90625	42,35
6	39	48	Varianza	646,797379	161,515385
7	56	68	Observaciones	32	40
8	85	45	Varianza agrupada	376,4259821	
9	25	15	Diferencia hipotética de las medias	0	
10	45	45	Grados de libertad	70	
11	85	35	Estadístico t	2,728719501	
12	96	32	P(T<=t) una cola	0,004016772	
13	75	43	Valor crítico de t (una cola)	1,666914479	
14	36	47	P(T<=t) dos colas	0,008033543	
15	15	48	Valor crítico de t (dos colas)	1,994437112	
16	78	65			
17	87	35			
18	98	15			
19	96	25			
20	96	35			
21	95	42			
22	35	58			
23	48	65			
24	56	26			

LÍMITES DE CONFIANZA

Se habla con frecuencia de *límites de probabilidad,* siendo dos valores (superior e inferior) asignados a un valor estimado, con el objeto de establecer el intervalo dentro del cual se supone está comprendido el valor verdadero (parámetro) conforme a algún valor de carácter probabilístico, generalmente denominado *nivel de confianza*.

Los *límites de confianza* se tendrá los valores Z_s y Z_i cuando la muestra es grande, (también t_s y t_i cuando la muestra es pequeña) que forman los extremos superior e inferior respectivamente de los intervalos de confianza, con respecto a un valor puntual.

$$L_s = Z_s = \text{Límite superior} \qquad L_i = Z_i = \text{Límite inferior del intervalo}$$

Recordemos que en alguna oportunidad se habló de dos tipos de estimación: *puntual* y por *intervalos*. El primero, corresponde a *un solo valor* o *punto* que resulta de una investigación por muestreo, como por ejemplo, la media aritmética, una proporción, una razón, etc. Al realizar una muestra, se obtienen conclusiones sobre la población, mediante la utilización de alguna medida de posición o de dispersión, pero este valor puede ser igual o diferente al del parámetro. Por lo tanto se debe establecer dos puntos, dentro de los cuales estará el parámetro con cierto grado de confiabilidad.

Ejemplo 1. El propietario de una pequeña fábrica de artesanías, toma una muestra aleatoria de la producción semanal de 6 de los 35 empleados, obteniendo un promedio de 22,5 figuritas, con una desviación estándar de 3,1.

o :

Encontramos que el estimador es de 22,5, el cual puede ser igual o diferente del promedio (parámetro), considerando un total de 35 empleados. Para muchos, es más fácil y más representativo fijar los límites de confianza para ese estimativo, con la probabilidad que se considere conveniente, supongamos del 95%. Por lo tanto se tendrán, partiendo de la variante estadística, los límites de confianza para ese estimativo, con la probabilidad más indicada, generalmente del 95%. Su fórmula se obtiene, partiendo de la variante estadística, en este caso "*t*":

$$t = \frac{\bar{x} - \mu}{\hat{s}/\sqrt{n-1}} \qquad \text{despejando:} \qquad \mu = \bar{x} \pm t\frac{\hat{s}}{\sqrt{n-1}}$$

$$\text{reemplazando: } \mu = 22{,}5 \pm 2{,}571\frac{3{,}1}{\sqrt{6}} = 22{,}5 \pm 3{,}25 \quad = \begin{cases} 25{,}75 = \mu_{\text{superior}} \\ 19{,}25 = \mu_{\text{inferior}} \end{cases}$$

Se tendrán dos límites $\mu_{\text{superior}} = 25{,}75$ y $\mu_{\text{inferior}} = 19{,}25$. Con estos resultados se dirá que la producción real semanal por empleado debe estar entre 19 y 26 figuras, con una seguridad o confianza del 95%. Observamos que nos queda un 5% considerado como error, de tal manera que el promedio de figuras por empleado puede ser superior a 26 o inferior a 19. La media poblacional deberá estar incluida (en algún punto) dentro del intervalo.

Podría preguntarse ¿por qué se toma 90%, 99% y no el 100% de confianza? Se puede responder al interrogante con el anterior ejercicio, observemos que a medida que se aumenta el nivel de confianza, más amplio va a ser el intervalo, con lo cual se obtendría una información menos precisa y menos útil.

> Con los intervalos de confianza pueden realizarse pruebas de hipótesis, teniendo en cuenta que la *prueba* debe ser siempre *bilateral.* Si la hipótesis nula (Ho) se ubica dentro del intervalo, se acepta; en caso contrario, se aceptará la alternativa (Ha).

DISTRIBUCIÓN DE MEDIAS MUESTRALES

Con las siguientes fórmulas se pueden determinar los límites de confianza para cada caso, dependiendo de la desviación típica y del tamaño de la muestra, son:

$\mu_{\substack{s \\ i}} = \bar{x} \pm Z\frac{\sigma}{\sqrt{n}}$	$\mu_{\substack{s \\ i}} = \bar{x} \pm Z\frac{s}{\sqrt{n}}$	$\mu_{\substack{s \\ i}} = \bar{x} \pm t\frac{s}{\sqrt{n}}$	$\mu_{\substack{s \\ i}} = \bar{x} \pm t\frac{\hat{s}}{\sqrt{n-1}}$
Cuando se da σ	se tiene s y $n > 30$	se tiene S (corregida) y $n \leq 30$	se tiene $\hat{s}$ (sin corregir) y $n \leq 30$

Ejemplo 2. Un muestra de 50 observaciones tiene una media de 65 y una desviación estándar de 4,2. Se piden límites de confianza del 95%.

$$\mu_{\substack{s\\i}} = 65 \pm 1{,}96\,\frac{4{,}2}{\sqrt{50}}; \qquad \frac{0{,}9500}{2} = 0{,}4750 \qquad \mu_{\substack{s\\i}} = \bar{x} \pm Z\frac{\sigma}{\sqrt{n}}$$

$$A\,(0{,}4750) \rightarrow 1{,}96 = Z$$

$$Z = 1{,}96$$

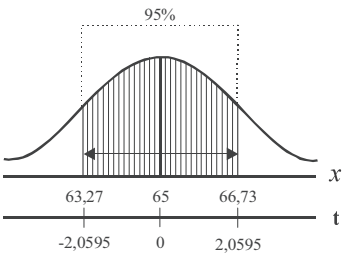
$$\mu_{\substack{s\\i}} = 65 \pm 1{,}96\,\frac{4{,}2}{7{,}07} = 65 \pm 1{,}96\,(0{,}59) = 65 \pm 1{,}16 \;\begin{matrix}\nearrow\ 66{,}16\\ \searrow\ 63{,}84\end{matrix}$$

Ejemplo 3. Una muestra de 26 observaciones tiene una media de 65 y una desviación de 4,2. Se piden límites de confianza del 95%.

o **:** $\bar{x} = 65 \qquad \hat{s} = 4{,}2$

$$\mu_{\substack{s\\i}} = \bar{x} \pm t\,\frac{\hat{s}}{\sqrt{n-1}}$$

$$\mu_{\substack{s\\i}} = 65 \pm 2{,}0595\,\frac{4{,}2}{\sqrt{25}}$$



$$\mu_{\substack{s\\i}} = 65 \pm 2{,}0595(0{,}84) = 65 \pm 1{,}73 \;=\; \begin{cases} 66{,}73 \\ 63{,}27 \end{cases}$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

118. Si se obtienen todas las posibles muestras de tamaño 25 en una distribución normal con media 20 y desviación estándar 4, ¿dentro de qué límite se encuentra el 90% central de las medias muestrales?

119. Suponga que la estatura media de los hombres tiene una desviación estándar de 2,48 centímetros. Se miden 100 estudiantes, hombres, elegidos aleatoriamente y se obtiene una estatura media de 168,52 centímetros. Determine los límites de confianza del 99 % para la estatura media de los hombres de esta universidad.

120. En una muestra de 60 observaciones, la media es 35 y la desviación estándar, 4,2. Obtenga los límites entre los cuales debe encontrarse la media poblacional, con una probabilidad del 95%.

121. Una muestra de 36 observaciones tiene una media de 40 y una desviación estándar de 2,1. Encuentre los límites de confianza del 95% para la media de la población.

122. Una muestra de 80 láminas de hierro galvanizado, con la cual determinamos su peso, con una media de 4,82 onzas y una desviación típica de 0,1 onzas. Calcular los límites de confianza, al nivel del 90%.

123. Un fabricante de metros metálicos, con el fin de controlar la exactitud de los mismos, tomó una muestra de 10 de ellos y los midió con toda precisión. Las medidas obtenidas fueron: 0,99 1,04 0,98 0,97 1,02 1,01 0,99 0,95 1,03 y 1,02 metros. Estime con un 90% de confianza los límites para la media poblacional.

124. Quince trozos de alambre de cobre tomados de 15 rodillos de alambre tienen las siguientes medidas de resistencia a la rotura (medida en libras): 1.672 2.854 1.680 1.570 2.020 1.568 872 980 1.080 870 896 812 930 850 864. Calcular los límites de confianza con un 95%, para la media de la población.

125. Las determinaciones químicas del porcentaje de hierro de 15 minerales de un depósito elegido al azar, dieron un promedio de 38,2% y una desviación típica del 5,2%. Establecer los límites de confianza con un 99%, para el porcentaje medio de hierro en el depósito.

126. Una muestra de 14 observaciones tiene una media de 34,86 y una desviación estándar de 4,23. Encuentre los límites que en el 95% de los casos permiten acertar, al afirmar que la media poblacional queda incluida entre ellos.

127. En una muestra de 23 observaciones, la media es 26,3 y la desviación estándar 1,9. Obtenga los límites de confianza del 99% para la media poblacional.

128. Encuentre los límites de confianza del 90% para la media de una población normalmente distribuida, con desviación estándar 3,83, si una muestra obtenida en dicha población presenta los siguientes valores: 18,5 20,6 12,9 14,6 19,8 15,0.

129. En una muestra de 5 artículos se ha obtenido, al nivel del 95%, que los límites de confianza para peso de todos los artículos embarcados, son de 12,39 y 17,11 onzas. Obtenga la media y la desviación estándar de dicha muestra.

130. Los pesos netos, en onzas, de una muestra aleatoria de 8 tarros de cerveza, son los siguientes:

12,1	11,9	12,4	12,3	11,9	12,1	12,4	12,1

a) Encuentre los límites de confianza del 99% para el peso medio por tarro de cerveza correspondiente a la población de la cual se obtuvo la muestra.
b) Empleando el resultado obtenido, señale en esta muestra (al nivel del 1%) que se mantiene la producción corriente de 12,35 onzas de promedio por tarro. ¿Explique su respuesta?

131. Un sondeo efectuado a 400 familias de clase media reveló un gasto trimestral promedio de $374.000 en productos de tocador, con desviación de $80.000.
a) Determine un intervalo de confianza del 95%.
b) ¿Cuál es el máximo error, cuando se afirma que dicha media es de $374.000, con una confianza del 99%?

132. En una muestra de tamaño l0, tomada de una población normal, los límites de confianza del 90% para la media de la población son $\bar{x} \pm 5{,}2$ se pide determinar la desviación estándar de la población.

133. Supongamos que un grupo de médicos ha establecido, que si los cigarrillos contienen por término medio, 30 miligramos o más de nicotina, es seguro que se produce cáncer de pulmón en el fumador. Un ensayo sobre 100 cigarrillos de la marca A, muestra que $\bar{x} = 26$ miligramos de nicotina. Si se sabe que $\sigma = 8$ miligramos, a) calcular el intervalo de confianza de 99% para el consumo medio de nicotina de la marca A de cigarrillos. b) El fumador está dispuesto a correr el riesgo si la media es diferente de 30, ¿Qué decisión tomaría al nivel del 1%?

134. El salario medio mensual para una muestra de 30 operarios textiles es de $612.000, con una varianza de $935.000. Fije los límites de confianza del 95% para el salario medio mensual de los operarios en fábricas de textiles.

135. Se prueba una muestra aleatoria de 8 fusibles de una cierta marca, para determinar el punto medio de ruptura. Los puntos de ruptura medidos en amperios fueron:

24	18	22	14	26	20	22	19

a) Fije los límites de confianza del 99%
b) Al nivel del 1%, podría afirmarse que estos resultados son diferentes a los señalados por la empresa, de 22 amperios de punto medio de ruptura?

136. Una muestra aleatoria de 18 operadoras de máquinas perforadoras en bancos comerciales de una ciudad, da la siguiente información:
X = Salario por semana de una operadora;

$$\sum x_i = \$2.700.000 \qquad \sum (x_i - \bar{x})^2 = 1.351.000$$

a) Obténgase un intervalo de confianza del 95% para estimar el salario medio por semana de los operadores de máquinas perforadoras empleadas por los bancos comerciales de dicha ciudad.
b) A un nivel del 5%, ¿podría afirmarse que el promedio de salarios por semana es de $152.000.
c) Si se sabe que el verdadero promedio de salario semanal es de $160.000, ¿qué tipo de error se estaría cometiendo?

137. Una muestra de 9 frascos de café instantáneo, seleccionados de un proceso de producción, dio los siguientes valores para el contenido, medido en gramos:

285	291	265	270	279	288	290	279	284

a) Obténgase un intervalo de confianza del 95%, para estimar el peso neto de los frascos de café instantáneo.
b) Por larga experiencia, se sabe que el contenido promedio es de 280 gramos. Pruebe, al nivel de 5%, si es cierto.

138. El departamento de tránsito y transporte llevó a cabo un estudio sobre velocidad a la que conducen los automovilistas en una autopista durante las primeras horas de la mañana (2 am - 6 am). Durante el tiempo de estudio se detectó que 100 automóviles tuvieron un promedio de velocidad de 120 km/h con una desviación estándar de 25 km/h.
a) Estime la media verdadera.
b) Establezca un intervalo de confianza del 98% para la media de la población. c) Pruebe al nivel del 2% que el promedio de velocidad es diferente al de 100 km/h fijado para esa vía, durante ese período.

DISTRIBUCIÓN DE MEDIAS PROPORCIONALES MUESTRALES

De la fórmula $Z = \dfrac{p - P}{\sqrt{\dfrac{pq}{n}}}$ se despeja P ó μ_p para obtener $\mu_p = p \pm Z\sqrt{\dfrac{pq}{n}}$

Y el procedimiento de cálculo es casi similar al utilizado en *las Medias muestrales*. Veamos algunos ejemplos que ayudarán a establecer los límites de confianza, tanto para z cuando la muestra es grande ($n > 30$) y "t cuando es pequeña ($n \leq 30$) :

Ejemplo 1. Una muestra de 100 votantes elegidos al azar entre los habitantes de un barrio, indicaba que el 45% de ellos estaban a favor de un candidato. Hallar los límites de confianza del 95%.

o :

$$\mu_p = 0{,}45 \pm 1{,}96\sqrt{\frac{0{,}45(0{,}55)}{100}} = 0{,}45 \pm 0{,}10 \begin{matrix} \nearrow 0{,}55 \\ \searrow 0{,}35 \end{matrix}$$

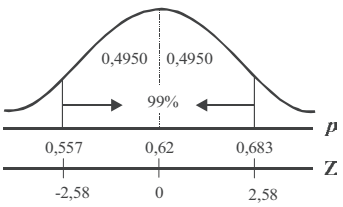
$$\mu_p = p \pm Z\sqrt{\frac{pq}{n}}$$

Ejemplo 2. Una investigación efectuada a 400 familias de clase media reveló que en la realización de fiestas familiares, un 62% prefería el aguardiente a cualquier otra clase de licor. Determine los límites de confianza del 99%.

o :

$$\mu_p = 0{,}62 \pm 2{,}58\sqrt{\frac{(0{,}62)(0{,}38)}{400}} =$$

$$\mu_p = 0{,}62 \pm 0{,}063 \begin{matrix} \nearrow 0{,}683 \\ \searrow 0{,}557 \end{matrix}$$



Ejemplo 3. Una compañía asegura que el 80% de sus semillas de zanahoria germinan. Se plantan 50 semillas de las cuales 8 no germinan.

a) Hállese un intervalo de confianza del 90%, para la proporción de semillas que germinaron en la muestra
b) Con base en lo anterior docime, al nivel del 10%, lo que afirma la compañía.

o :

a) $p = 1 - q \qquad p = 1 - \frac{8}{50} = 1 - 0{,}16 = 0{,}84$

$$\mu_{p_s} = 0{,}\ 4 \pm 1{,}65\sqrt{\frac{0{,}\ 4\ \ 0{,}16}{50}} = 0{,}\ 4 \pm 0{,}0\ \ 6 \begin{matrix} \nearrow 0{,}\ 26 \\ \searrow 0{,}754 \end{matrix}$$

b) 1) $H_0 : \mu_p = 0{,}80$ 2) $\propto = 0{,}10$ 3) $s_p = 0{,}052$

$H_a : \mu_p \neq 0{,}80$

Se puede aceptar lo ofrecido por la empresa, ya que 0,80 se encuentra dentro de los límites de confianza, por tal motivo se acepta la hipótesis nula, al nivel del 10%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

139. Una muestra al azar de 5.000 casas en una ciudad, indica que 188 estaban desocupadas. Estime la proporción de casas desocupadas en la ciudad, con un nivel de confianza del 90%, del 95 y del 99%.

140. En una muestra al azar de 826 teléfonos de residencias del directorio de Medellín, 95 no respondieron a la llamada entre 7 y 8 de la noche, el día que se realizó la muestra. Determinar los límites de confianza del porcentaje de suscriptores en cuyas residencias hubo alguien entre 7 y 8 de la noche. (Se admite que no se contestó porque no había nadie en casa). El nivel de confianza adoptado es del 90%.

141. Un fabricante de máquinas de afeitar desechables, solicita que se haga una investigación de mercados, tomando una muestra de 100 personas empleadas del sector público, obteniéndose que el 60% prefieren esa marca y tipo de máquina. Con la anterior información se pide.
   a) Fijar los límites de confianza del 99%, de las personas que prefieren esta máquina.
   b) El fabricante estima que el 62% de los empleados públicos prefieren dicha máquina; se pide, al nivel del 1%, probar la hipótesis de que tiene razón.
   c) Si el porcentaje verdadero de preferencia es del 65%, determinar qué tipo de error se está cometiendo.

142. En una encuesta de opinión pública, 640 personas de un total de 800, se pronunciaron en favor de cierto producto.
   a) Establecer un intervalo de confianza del 99%, para estimar la proporción de personas que están a favor de este producto.
   b) Si el productor afirma que el 85% de la opinión es favorable, ¿se puede aceptar dicha aseveración al nivel del 1%?
   c) ¿Con qué grado de confianza, puede afirmarse que la proporción de personas que están a favor del producto se sitúa entre 77% y 83%?

143. Una muestra aleatoria de 100 clientes que realizan sus compras entre las 12 a.m y 2 p.m, revela que todos, excepto 15, incluyen la leche en sus compras.
   a) Establezca un intervalo de confianza del 90%.
   b) A nivel del 10% y de acuerdo con los resultados del punto (a) ¿se podrá afirmar, que el porcentaje de personas que incluyen leche en sus compras es diferente al afirmado por el gerente, quien asegura que es del 80%?

144. En un barrio de la ciudad se efectúa un muestreo para determinar la proporción de familias que poseen un televisor a color. La muestra de tamaño 80 indica que 52 tienen televisión a color.
   a) Establecer un intervalo de confianza del 95%, para estimar la proporción de familias que poseen televisor a color.
   b) Si una información dada por la Cámara de Comercio afirma que de 1.000 familias, 560 poseen televisor a color, pruebe esta afirmación al nivel del 5%.

DISTRIBUCIÓN DE DIFERENCIA ENTRE DOS MEDIAS MUESTRALES ( )

Se presentan en resumen, fórmulas que deben ser aplicadas de acuerdo al enunciado del problema, dado que cada uno de ellos tiene información diferente.

$$\mu_{\bar{x}-\bar{y}} = \mu_x - \mu_y = (\bar{x} - \bar{y}) \pm z\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}$$

Cuando se conoce σ_x y σ_y

$$\mu_{\bar{x}-\bar{y}} = (\bar{x}-\bar{y}) \pm z\sqrt{\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}}$$

Si se tiene s_x y s_y; además n_1 y n_2 son mayores de 30

$$\mu_{\bar{x}-\bar{y}} = (\bar{x}-\bar{y}) \pm t\sqrt{\frac{(n_1-1)s_x^2+(n_2-1)s_y^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Se tiene s_x y s_y (ambas sin corregir) además n_1 y n_2 son menores o iguales a 30

$$\mu_{\bar{x}-\bar{y}} = (\bar{x}-\bar{y}) \pm t\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}}$$

Cuando n_1 y $n_2 \leq 30$

$$s^2 = \frac{\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2}{n_1+n_2-2}$$

Ejemplo 1. En un experimento llevado a cabo para medir la efectividad de dos abonos en dos parcelas, de 10 hectáreas cada una cultivadas con maíz, se obtuvieron las siguientes medias, para la primera parcela abonada con el producto A, 81,7 cargas por hectárea; en la segunda parcela, 88,3 cargas por hectárea con el B y las desviaciones típicas de 0,7 y 0,8 respectivamente. ¿Cuáles son los límites de confianza para la diferencia de los promedios con una probabilidad del 95%?

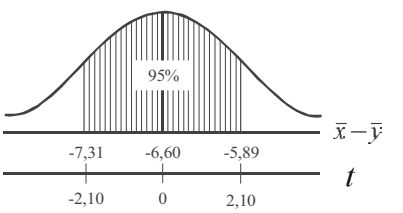
o :

$n_1 = 10 \quad n_2 = 10 \quad \bar{x} = 81{,}7 \quad \bar{y} = 88{,}3 \quad s_x = 0{,}7 \quad s_y = 0{,}8$

$$\mu_{\bar{x}-\bar{y}_{s}^{i}} = (81{,}7-88{,}3) \pm 2{,}10\sqrt{\frac{9(0{,}49)+9(0{,}64)}{10+10-2}}\sqrt{\frac{1}{10}+\frac{1}{10}}$$

$$\mu_{\bar{x}-\bar{y}_{s}^{i}} = -6{,}6 \pm 2{,}10\sqrt{\frac{4{,}41+5{,}76}{18}}\sqrt{\frac{2}{10}}$$

$$\mu_{\bar{x}-\bar{y}_{s}^{i}} = -6{,}6 \pm 2{,}10\sqrt{(0{,}565)\ (0{,}20)}$$



$$\mu_{\bar{x}-\bar{y}_{s}^{i}} = -6{,}6 \pm 2{,}10\ (0{,}34) = -6{,}6 \pm 0{,}71 \begin{matrix} \nearrow -5{,}89 \\ \searrow -7{,}31 \end{matrix}$$

Ejemplo 2. Pruebas de resistencia en dos tipos de géneros de lana dieron los siguientes resultados (en libras por centímetro cuadrado).

TIPO I:	138	127	134	125		
TIPO II:	134	137	135	140	130	134

Determinar los límites de confianza del 95%

o :

x_i	y_i	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
138	134	7	49	-1	1
127	137	-4	16	2	4
134	135	3	9	0	0
125	140	-6	36	5	25
	130			-5	25
	134			-1	1
524	810	0	110	0	56

$$\bar{x}=\frac{524}{4}=131 \qquad \bar{y}=\frac{810}{6}=135 \qquad s^2=\frac{110+56}{8}=20{,}75$$

$$s_{\bar{x}-\bar{y}}=\sqrt{\frac{20{,}75}{4}+\frac{20{,}75}{6}}=\sqrt{5{,}19+3{,}46}=\sqrt{8{,}65}=2{,}94$$

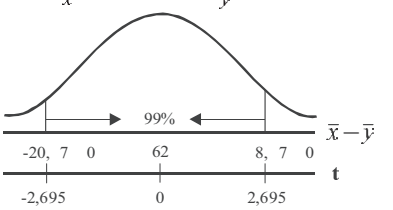
$$\mu_{\bar{x}-\bar{y}}=(131-135)\pm 2{,}306\ (2{,}94)=-4\pm 6{,}78 \begin{matrix}\nearrow 2{,}78 \\ \searrow -10{,}78\end{matrix}$$

Ejemplo 3. Se someten a una prueba 45 alumnos de un curso de estadística, para determinar las diferencias de rendimiento entre hombres y mujeres. Las 20 mujeres tienen un puntaje medio de 60 con una desviación estándar de 19 y los 25 hombres, una media de 66 con una desviación estándar de 16. ¿Cuáles son los límites de confianza del 99%, para la diferencia entre los puntajes medios de los hombres y las mujeres?

o :

$$n_1=20 \qquad n_2=25 \qquad \bar{x}=60 \qquad \bar{y}=66 \qquad s_x=19 \qquad s_y=16$$

$$s_{\bar{x}-\bar{y}}=\sqrt{\frac{(n_1-1)s_x^2+(n_2-1)s_y^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$



$$s_{\bar{x}-\bar{y}}=\sqrt{\frac{(20-1)361+(25-1)256}{20+25-2}}\sqrt{\frac{1}{20}+\frac{1}{25}}=5{,}22$$

$$\mu_{\bar{x}-\bar{y}_{s_i}}=(\bar{x}-\bar{y})\pm t\,\hat{s}_{\bar{x}-\bar{y}} \qquad \mu_{\bar{x}-\bar{y}_{s_i}}=-6\pm 2{,}695\ (5{,}22)=-6\pm 14{,}07 \begin{matrix}\nearrow 8{,}07 \\ \searrow -20{,}07\end{matrix}$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

145. A partir de dos muestras obtenidas de dos poblaciones normalmente distribuidas se obtuvieron los siguientes datos:

MUESTRA A:	$n_1 = 10$	$\sum x_i = 21$	$\sum x_i^2 = 92$
MUESTRA B:	$n_2 = 16$	$\sum y_i = 32$	$\sum y_i^2 = 136$

Encuentre el intervalo de confianza del 95%, para la diferencia entre las medias poblacionales.

146. En dos muestras se obtienen los siguientes datos:

$\bar{x} = 75$ $\bar{y} = 60$ $n_1 = 36$ $n_2 = 64$

$\sum (x_i - \bar{x})^2 = 1.482$ $\sum (y_i - \bar{y})^2 = 1.830$

Encontrar los límites de confianza del 98%, para la diferencia entre las medias de las poblaciones de donde se seleccionaron las muestras.

147. Se utiliza una escala especial para medir las actitudes de un grupo de hombres y mujeres, con respecto al aroma de cierta marca de desodorante personal. Mientras más altos son los puntos, más favorable es la actitud hacia el aroma. Una muestra de 9 hombres, arroja una media de 5 puntos y una desviación típica de 2,83 y otra muestra del mismo tamaño, aplicado en mujeres, presentó un puntaje medio de 7,5 y una desviación típica de 2,27.

a) Fije los límites de confianza del 95%

b) Con los datos anteriores, docime si la actitud de las mujeres hacia el aroma es diferente a la de los hombres, (5%).

148. a) Construya un intervalo de confianza del 99%, para estimar la diferencia entre las medias problacionales, basado en dos muestras cuyos resultados fueron:

$n_1 = 9$ $\sum x_i = 360$ $s_x = 6$ $n_2 = 20$ $\sum y_i = 960$ $s_y^2 = 100$

b) Docime si la diferencia es significativa utilizando un nivel de significación del 1%.

149. Se hizo una entrevista a siete (7) subdirectores y a cinco (5) analistas de mercado de una gran empresa. Se les preguntó a cada uno sobre el porcentaje óptimo de cobertura de mercado para su compañía. Se obtuvieron las siguientes respuestas:

SUBDIRECTORES:	25,0	31,3	27,8	20,3	24,2	22,5	26,3
ANALISTAS DE MERCADO:	23,1	28,6	24,2	23,5	26,3		

a) Establezca límites de confianza del 95%

b) Al nivel del 5%, ¿sugieren estos datos que los subdirectores y los analistas están en desacuerdo cuando estiman la cobertura óptima de mercado para la empresa?

150. El advenimiento de materiales sintéticos tales como nylon, poliester y latex, y su introducción al mercado ha suscitado debates acerca de la calidad y resistencia de estas fibras comparadas con las fibras naturales; se seleccionaron 10 fibras sintéticas y 10 naturales, siendo:

$\bar{x} = 272\ g$ $s_y^2 = 1.636\ g^2$ $\bar{y} = 335$ kg $s_y^2 = 1.892\ g^2$

a) Encuentre un intervalo de confianza del 95% para la diferencia en resistencia media entre la fibra sintética y la natural.

b) Al nivel del 5%, ¿se podrá aceptar una diferencia entre la resistencia media de estas fibras?

151. El auditor de un banco toma dos muestras, una de 41 cuentas corrientes y otra de 35 cuentas de ahorro, para estudiar el saldo al finalizar un año, encontrando lo siguiente:

MEDIAS	CUENTAS CORRIENTES	CUENTAS DE AHORRO
Saldo promedio (en miles)	800	1.200
Desviación típica	150	200
No. de personas con cuentas con más de dos (2) años	41	35

a) Hallar intervalo de confianza del 98% para el saldo promedio de cuentas corrientes.
b) ¿Se puede aceptar al nivel de 1%, que el saldo promedio de las cuentas de ahorro es mayor a $1.000.000?
c) ¿Se puede aceptar que la proporción de los que tienen cuenta corriente hace más de dos años es mayor del 50% al nivel del 5%?
d) ¿Se puede aceptar, al nivel del 5%, que la variabilidad en los saldos de cuentas corrientes es diferente a la de las cuentas de ahorro?
e) ¿Se puede aceptar al nivel del 5%, que el saldo promedio de las cuentas corrientes es menor que el de las cuentas de ahorro?

152. Los estudiantes que se matricularon en un curso de investigación educativa fueron distribuidos al azar en dos grupos. El grupo A conformado por 10 alumnos, utilizó numerosas técnicas y actividades para enriquecer el curso, obtuvo como rendimiento una calificación promedio de 4 y una desviación típica de 0,4; mientras el grupo B, estudió mediante el método tradicional de apuntes y conferencias, conformado por 12 alumnos tuvo un rendimiento con promedio de 3,6 y desviación típica de 0,03. Construir el intervalo de confianza del 95%.

DISTRIBUCIÓN DE DIFERENCIA ENTRE DOS PROPORCIONES MUESTRALES $_1 - _2$

Generalmente se aplican dos fórmulas, dependiendo del tamaño de las muestras, éstas son:

$$\mu_{p_1-p_2} = (p_1 - p_2) \pm t\sqrt{\frac{p_1 q_1}{n_1 - 1} + \frac{p_2 q_2}{n_2 - 1}}$$

Cuando n_1 y $n_2 \leq 30$

$$\mu_{p_1-p_2} = (p_1 - p_2) \pm Z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Cuando n_1 y $n_2 > 30$

Ejemplo 1. En una muestra al azar de 200 adultos y 300 adolescentes que veían cierto programa de televisión, 50 adultos y 150 adolescentes dijeron que les gustaba. Hallar los límites de confianza del 95%, para la diferencia de proporción de adultos y adolescentes que ven televisión.

o :

$$\mu_{p_1-p_2} = (p_1 - p_2) \pm Z\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$p_1 = \frac{50}{200} = 0{,}25$$

$$p_2 = \frac{150}{300} = 0{,}50$$

$$\mu_{p_1-p_2} = (0{,}25-0{,}50) \pm 1{,}95\sqrt{\frac{0{,}25(0{,}75)}{200}+\frac{0{,}5(0{,}5)}{300}} = -0{,}25 \pm 0{,}08 \begin{matrix} \nearrow -0{,}17 \\ \searrow -0{,}33 \end{matrix}$$

Ejemplo 2. Se desea determinar si hay alguna diferencia significativa, según el sexo, en la preferencia por margarina o mantequilla. Se realizó una muestra aleatoria entre 26 hombres y 18 mujeres, indicando que 16 y 10 respectivamente, preferían la margarina a la mantequilla.

a) Establezca límites de confianza del 95% para estas diferencias

b) Pruebe la hipótesis de que la diferencia es significativa, al nivel del 5%.

o :

$$p_1 = \frac{16}{26} = 0{,}62 \qquad p_2 = \frac{10}{18} = 0{,}55 \qquad v = 26+18-2 = 42 \qquad \propto = 0{,}05$$

$$\mu_{p_1-p_2} = (0{,}62-0{,}55) \pm 2{,}018\sqrt{\frac{0{,}62(0{,}38)}{25}+\frac{0{,}55(0{,}45)}{17}} = 0{,}07 \pm 0{,}31 \begin{matrix} \nearrow 0{,}38 \\ \searrow -0{,}24 \end{matrix}$$

$H_0 : P_1 = P_2 \qquad \propto = 0{,}05$
$H_a : P_1 \neq P_2$

$P_1 - P_2 = 0$; el cero está dentro de los límites de confianza, por lo tanto aceptamos H_0, es decir, no hay diferencia en la preferencia de acuerdo al sexo, al nivel del 5%. En este caso hemos utilizado "t" para muestras pequeñas.

Ejemplo 3. En una universidad de Bogotá se toma una muestra aleatoria de 600 mujeres, 300 de las cuales indican que están a favor de la supresión de los exámenes preparatorios y la sustentación de tesis en público. En una muestra de 400 hombres, 240 indican que están a favor de lo mismo.

a) Construya los límites de confianza del 95%, para la diferencia en las proporciones de todas las alumnas y alumnos, que están a favor de la supresión de los preparatorios y sustentación.

b) Docime al nivel del 5% si la diferencia entre las proporciones es significativa, para ello utilice el punto (a).

o :

$$p_1 = \frac{300}{600} = 0{,}50 \qquad p_2 = \frac{240}{400} = 0{,}60$$

a) $\mu_{P_1-P_2} = (0{,}50-0{,}60) \pm 1{,}95\sqrt{\frac{(0{,}5)(0{,}5)}{600}+\frac{(0{,}6)(0{,}4)}{400}} = -0{,}10 \pm 0{,}0622 \begin{matrix} \nearrow -0{,}0378 \\ \searrow -0{,}1622 \end{matrix}$

b) $H_0 : \mu_{p_1} = \mu_{p_2} \qquad \propto = 0{,}05 \qquad s_{P_1-P_2} = 0{,}032$
$H_a : \mu_{p_1} \neq \mu_{p_2}$

$\mu p_1 - \mu p_2 = 0$ no se ubica dentro de los límites de confianza y se rechaza la hipótesis nula, es decir, que la diferencia entre las proporciones es significativa, al nivel del 5%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

153. Un candidato a la alcaldía de la capital, realiza dos encuestas de tamaño 120, a personas nacidas en la ciudad y 150 que viven permanentemente pero nacieron en otras ciudades, para determinar la proporción de personas que estarían dispuestas a votar por él. Los resultados a favor fueron de 40 y 55 respectivamente.

    a) Fijar límites de confianza del 99% para la diferencia entre estas proporciones.

    b) ¿Se puede afirmar, al nivel del 1%, que hay una deficiencia significativa entre los porcentajes?

154. Se quiere poner en marcha un plan de incentivos al personal de una empresa. Por lo tanto, se realizan dos muestras aleatorias, una de tamaño 16 del personal de oficina y 22 del personal de operarios (planta). Recogida la información se encontró que el número de personas a favor de la propuesta de la directiva fue de 10 y 12 respectivamente. a) Fije límites de confianza del 95% para la diferencia del porcentaje de estas opiniones; b) ¿cree usted que los anteriores resultados permiten concluir que la aceptación de la nueva política de incentivos es igual para empleados y operarios? (Nivel del 5%).

155. El 20 de agosto, una encuesta a 200 electores, en un barrio de la ciudad, indicó que 120 estaban a favor del candidato X. Un mes después, una encuesta a 300 electores indicó sólo 150 a favor del mismo candidato.

    a) Fije los límites de confianza del 95%

    b) Utilizando un nivel de significación del 5%, ¿indica la anterior información que la popularidad del candidato disminuyó entre las dos muestras?

156. El director de una editorial de textos para secundaria, debe decidir sobre la publicación de un texto, siempre que las preferencias entre los colegios privados y públicos sean iguales. Selecciona dos muestras de tamaño 10 y 18 respectivamente, encontrando aceptación de 6 y 10 colegios. De acuerdo a esos resultados fije límites de confianza del 90% para la diferencia.

157. Una empresa comercializadora que ofrece dos sistemas de pago a sus clientes, desea saber la proporción de cuentas por cobrar, con más de 60 días de vencidas. En el primer sistema la proporción es del 18%, mientras que en el segundo, del 10%. Fije límites de confianza del 95% para la diferencia entre estas dos proporciones, cuyos tamaños muestrales fueron 16 y 20 clientes respectivamente.

158. En una ciudad hay dos ligas de baloncesto, con rivalidad en cuanto a la calidad de sus jugadores y de encestadores. Un deportista propone seleccionar al azar 20 jugadores de cada liga, de tal manera que cada jugador lance una sola vez desde un determinado punto. Realizados los lanzamientos se encontró que 14 de la liga A y 17 de la liga B encestaron. Fije límites de confianza del 95% para la diferencia entre las proporciones.

Nota: De aqui en adelante se presentan ejercicios misceláneos lo cual permite repasar lo **aprendido, pues no se especifica a que distribución corresponde.**

159. Un proceso está programado para empacar la cantidad media de una libra de café. Se toma una muestra aleatoria de 20 paquetes, resultando una media de 15,1 onzas y desviación estándar de 6,3 onzas. Al nivel del 5% se podría afirmar que: a) ¿se está vendiendo un producto por debajo de su peso? b) ¿si el valor verdadero es de 15,5 onzas, en la prueba anterior se cometió error?; c) ¿de qué tipo fue el error?

160. Un inversionista desea decidir entre dos tipos de valores. Para ello cuenta con la siguiente información acerca del rendimiento expresado como porcentaje del precio.

VALOR A:	8,1	8,5	6,6	8,6	7,8	7,6	10,2	9,2
VALOR B:	9,8	9,3	6,1	10,7	9,9	8,2		

Al nivel del 1%, ¿sugieren estos datos que el valor B produce una utilidad mayor que el valor A?

161. El gerente de una compañía realiza dos muestras de tamaño 120 empleados, cada una, con el fin de determinar el porcentaje de ausentismo. En la primera fábrica durante la semana de observación no asistieron 12 personas, mientras que en la segunda fábrica no asistieron 16 empleados. ¿Al nivel del 2%, se podrá afirmar que la asistencia en la primera fábrica es superior a la segunda?

162. Se requiere comparar el nivel salarial de los empleados de dos empresas. Se solicita información a la primera de ella y reporta que en una muestra aleatoria de 17 empleados, el ingreso promedio mensual es de $828 (miles de $), con una desviación estándar de $63 (miles de $).

Se elige en la segunda empresa una muestra aleatoria de 8 empleados con sueldos, en miles de pesos, así: $738 $8l6 $763 $775 $820 $812 $910 $836.

Fijar límites de confianza del 99% y probar con dicho resultado, si existe una diferencia salarial en las poblaciones de empleados de las dos empresas, al nivel del 1%.

163. Un distribuidor de leche afirma que el contenido medio de grasa es de 3,07 mg./ cc. Se toma una muestra en 10 recipientes y se encuentran los siguientes resultados: 3,5 2,8 3,5 4,0 3,2 2,5 2,8 3,0 2,5 y 2,6. ¿Al nivel del 1% se puede aceptar la afirmación?

164. Se hizo una entrevista a 7 subdirectores y a cinco analistas de mercado de una gran empresa. Se les preguntó, a través de estas dos muestras aleatorias, cuál consideraban que debería ser el porcentaje óptimo de cobertura de mercado para su compañía. Los resultados fueron:

SUBDIRECTORES:	35,0	31,3	37,8	30,3	34,2	32,5	36,3
ANALISTAS DE MERCADO:	33,1	28,6	34,2	33,5	36,3		

a. Establezca límites de confianza del 95%, para la diferencia entre los promedios de porcentajes.

b. Al nivel del 5% ¿sugieren estos resultados que los subdirectores y analistas están en desacuerdo cuando estiman la cobertura de mercado para la empresa?

165. Una empresa de manufacturas, considera que controla el 10% del mercado de cigarrillos. Para comprobar lo anterior se toma una muestra de 2.500 fumadores y se les pregunta sobre la marca de cigarrillo que prefieren, con un resultado de 218 que prefieren esa marca. Al nivel de 5%, ¿se podrá afirmar que ese porcentaje es correcto?

166. Un fabricante garantiza que la resistencia promedio a la tensión de los tornillos A, excede a los de B. Se toman dos muestras de 50 cada una, con los siguientes resultados: tornillo A con resistencia promedio a la tensión de 95,7 kg. y desviación estándar de 8,20 kg., mientras que los tornillos de la marca B, mostraron 87,8 kg. y 6,64 kg., respectivamente. Al nivel del 1% comprobar la afirmación del fabricante.

167. Dos plantas utilizadas por una fábrica se cree que presentan diferentes porcentajes de contaminación, como consecuencia de ser alimentadas con combustibles diferentes. Durante un determinado período se toman 100 muestras de aire de la planta A con niveles del 32% de las veces, mientras que otra muestra de 80 refleja niveles del 24% de las veces. ¿Se puede considerar que la segunda planta presenta menores niveles de contaminación? Nivel de significancia del 2%.

168. La secretaria de salud quiere saber si está en lo cierto, cuando afirma que la proporción de fumadores en su localidad, para personas mayores de 16 años, es del 40%. La secretaria lleva a cabo una muestra de 600 personas mayores de 16 años, de las cuales 210 son fumadoras. Pruebe la hipótesis de que la proporción de fumadores mayores de 16 años es diferente del 40%, para α= 0,05.

169. El gerente de producción, basado en su experiencia ha determinado que una operación específica requiere de un tiempo promedio de 12 minutos. El gerente sospecha que su personal requiere de un tiempo promedio diferente a los 12 minutos y por tal razón realiza una muestra en 8 de sus trabajadores con los siguientes resultados: 12,3 12,2 11,9 11,7 12,7 11,9 11,8 y 12,6. ¿Es válida la sospecha que tiene el gerente?

170. Un investigador de mercados utiliza dos medios para impulsar las ventas de su producto y desea saber si un medio es superior al otro en el aumento de las ventas. Discuta de manera completa cómo debe llevar el experimento, estableciendo las suposiciones necesarias que le permitan obtener conclusiones.

171. Un economista que trabaja en el departamento de recursos humanos, desea determinar si la tasa de desempleo en dos grandes ciudades del país, son diferentes. Con base en muestras aleatorias de 250 personas en cada ciudad, encuentra 18 empleados en una y 13 en la otra. Con un nivel 5%, ¿existe razón para afirmar que la tasa de desempleo en la segunda ciudad es más baja que en la primera?

172. Explique cada uno de los siguientes puntos:

a) ¿Cuál es la finalidad de una prueba de hipótesis?.

b) ¿Cuándo se debe utilizar la distribución "t" de Student?

c) ¿Cuándo se comete un error de tipo I?

d) ¿Como se define la inferencia estadística?

173. El tiempo promedio para que una persona se intoxique, estando expuesto a una sustancia química, es de 24 minutos. En una muestra de 10 personas, se encuentra que el tiempo promedio es de 26 minutos, con una desviación estándar de 3 minutos. Pruebe la afirmación contra la alternativa de que el tiempo medio es superior a los 24 minutos. ($\propto = 0{,}01$).

174. Discuta los siguientes puntos:

a. ¿Qué se debe entender por estimación puntual?

b. ¿En qué difiere la estimación puntual de la estimación por intervalo?

c. ¿Qué se entiende por grados de libertad?

d. ¿Qué es un intervalo de confianza?

e. ¿Qué significan los términos hipótesis nula e hipótesis alternativa?

f. ¿Por qué es posible que la hipótesis nula sea rechazada cuando de hecho es cierta?

g. ¿Cuándo se presenta una prueba unilateral a la derecha?

175. Las normas de una compañía sobre el servicio a los clientes, especifican que no más del 5% de los pedidos directos por correo no llegan al almacén, para ser cumplidos dentro de las 48 horas de su recibo. Se tomó una muestra de 191 pedidos, de los cuales 14 quedaban fuera del límite de las 48 horas. ¿Se justifica suponer que el requisito del 5% no se estaba cumpliendo?

APLICACIÓN CON EXCEL

Complemento virtual SIL

LÍMITES O INTERVALOS DE CONFIANZA PARA LA MEDIA ARITM TICA MEDIANTE EXCEL

Muestra grande n 30

Fórmula: $\hat{\mu} = \overline{x} \pm z \frac{s}{\sqrt{n}}$

Para el cálculo de estos límites, tomaremos como ejemplo la columna CALIFICACIONES obtenidas por los alumnos en las pruebas del ICFES que aparecen en la Tabla No. 1. De esta selección hemos tomado para el desarrollo de la aplicación en Excel una muestra de 50 alumnos de una población total de 1100 alumnos. (Ver información en el SIL segunda parte).

En la Tabla No. 1 la información se ubica en la columna H, para la selección se tiene en cuenta un rango entre H2 hasta H52, siendo H2, la celda correspondiente al nombre de la columna, que para este caso, corresponde a CALIFICACIONES; y el resto, es decir, el rango contemplado entre H3 hasta H52 son los datos de la muestra.

- Seleccionamos en la ficha DATOS el ícono ANÁLISIS DE DATOS.
- Aparece un cuadro de diálogo, correspondiente a ANÁLISIS DE DATOS, seleccionamos ESTADÍSTICA DESCRIPTIVA y hacemos CLIC en ACEPTAR.

Figura No. 16. Análisis de Datos

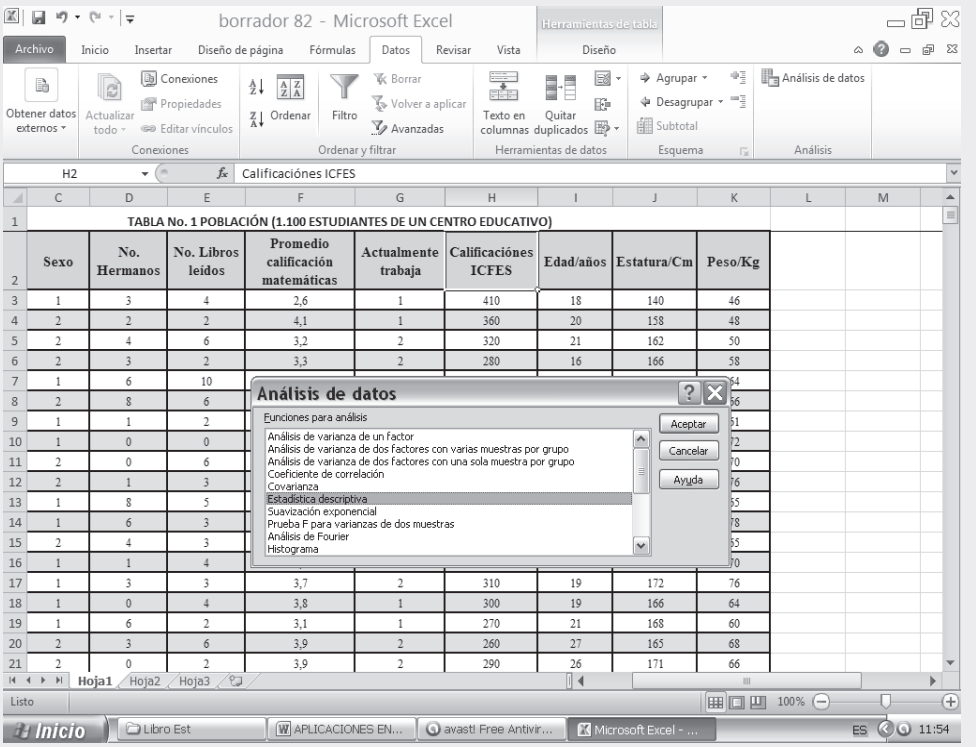


TABLA No. 1 POBLACIÓN (1.100 ESTUDIANTES DE UN CENTRO EDUCATIVO)

Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	3	4	2,6	1	410	18	140	46
2	2	2	4,1	1	360	20	158	48
2	4	6	3,2	2	320	21	162	50
2	3	2	3,3	2	280	16	166	58
1	6	10						64
2	8	6						56
1	1	2						51
1	0	0						72
2	0	6						70
2	1	3						76
1	8	5						55
1	6	3						78
2	4	3						55
1	1	4						70
1	3	3	3,7	2	310	19	172	76
1	0	4	3,8	1	300	19	166	64
1	6	2	3,1	1	270	21	168	60
2	3	6	3,9	2	260	27	165	68
2	0	2	3,9	2	290	26	171	66

❑ En el cuadro de diálogo de la figura que sigue, digitamos los datos que nos solicitan:

Figura No. 17. Estadística Descriptiva.

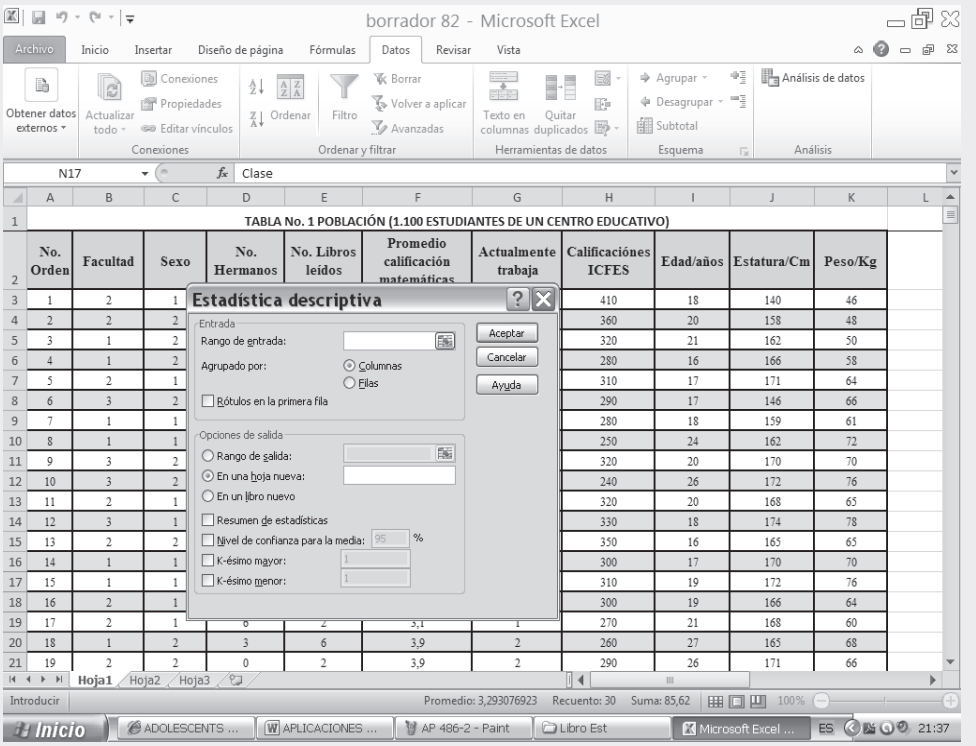
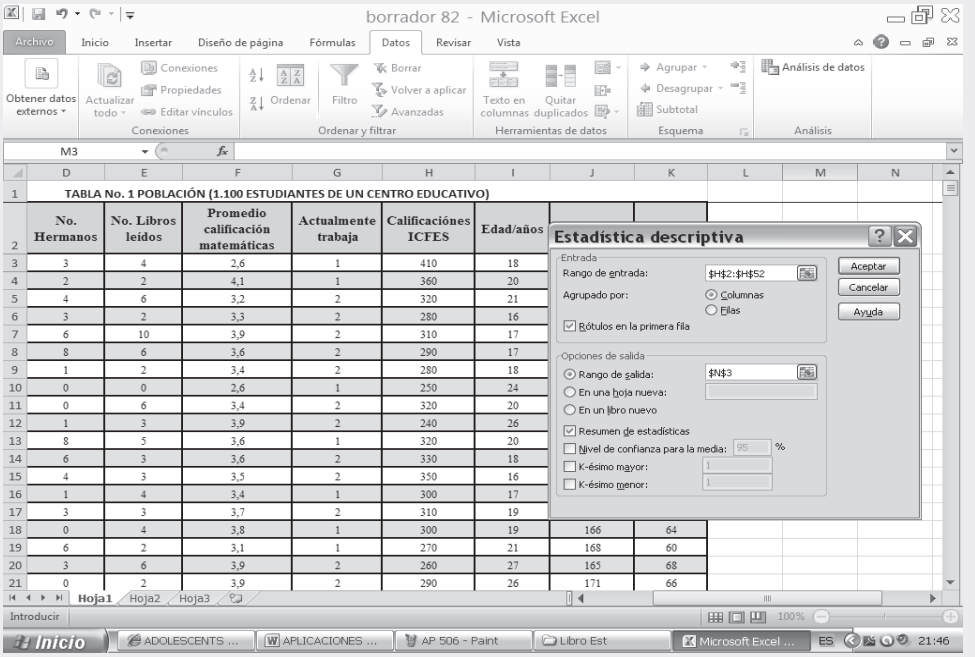


TABLA No. 1 POBLACIÓN (1.100 ESTUDIANTES DE UN CENTRO EDUCATIVO)

No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1					410	18	140	46
2	2	2					360	20	158	48
3	1	2					320	21	162	50
4	1	2					280	16	166	58
5	2	1					310	17	171	64
6	3	2					290	17	146	66
7	1	1					280	18	159	61
8	1	1					250	24	162	72
9	3	2					320	20	170	70
10	3	2					240	26	172	76
11	2	1					320	20	168	65
12	3	1					330	18	174	78
13	2	2					350	16	165	65
14	1	1					300	17	170	70
15	1	1					310	19	172	76
16	2	1					300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
18	1	2	3	6	3,9	2	260	27	165	68
19	2	2	0	2	3,9	2	290	26	171	66

- RANGO DE ENTRADA en esta casilla tecleamos H1:H52 que corresponde al rango de la muestra seleccionada.
- Activamos haciendo CLIC en RÓTULOS PRIMERA FILA y RESUMEN DE ESTADÍSTICAS.
- Finalmente establecemos el RANGO DE SALIDA en nuestro caso, ubicamos la celda en N3 y hacemos CLIC en ACEPTAR.

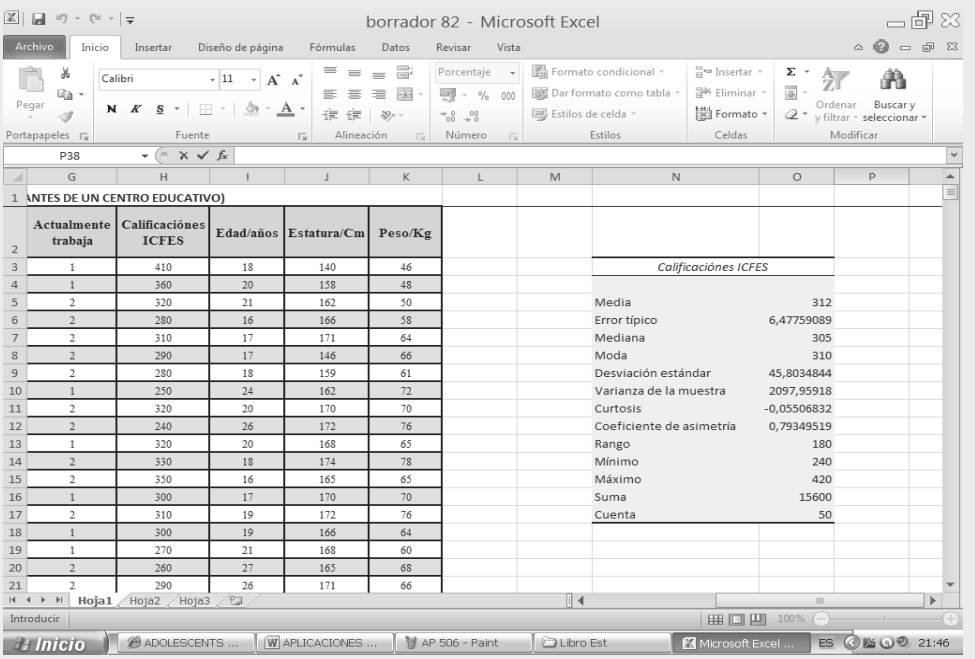
Figura No. 18. Cuadro de Información Estadística Descriptiva.



No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años
3	4	2,6	1	410	18
2	2	4,1	1	360	20
4	6	3,2	2	320	21
3	2	3,3	2	280	16
6	10	3,9	2	310	17
8	6	3,6	2	290	17
1	2	3,4	2	280	18
0	0	2,6	1	250	24
0	6	3,4	2	320	20
1	3	3,9	2	240	26
8	5	3,6	1	320	20
6	3	3,6	2	330	18
4	3	3,5	2	350	16
1	4	3,4	1	300	17
3	3	3,7	2	310	19
0	4	3,8	1	300	19
6	2	3,1	1	270	21
3	6	3,9	2	260	27
0	2	3,9	2	290	26

166	64
168	60
165	68
171	66

Figura No. 19. Cuadro de Resultados



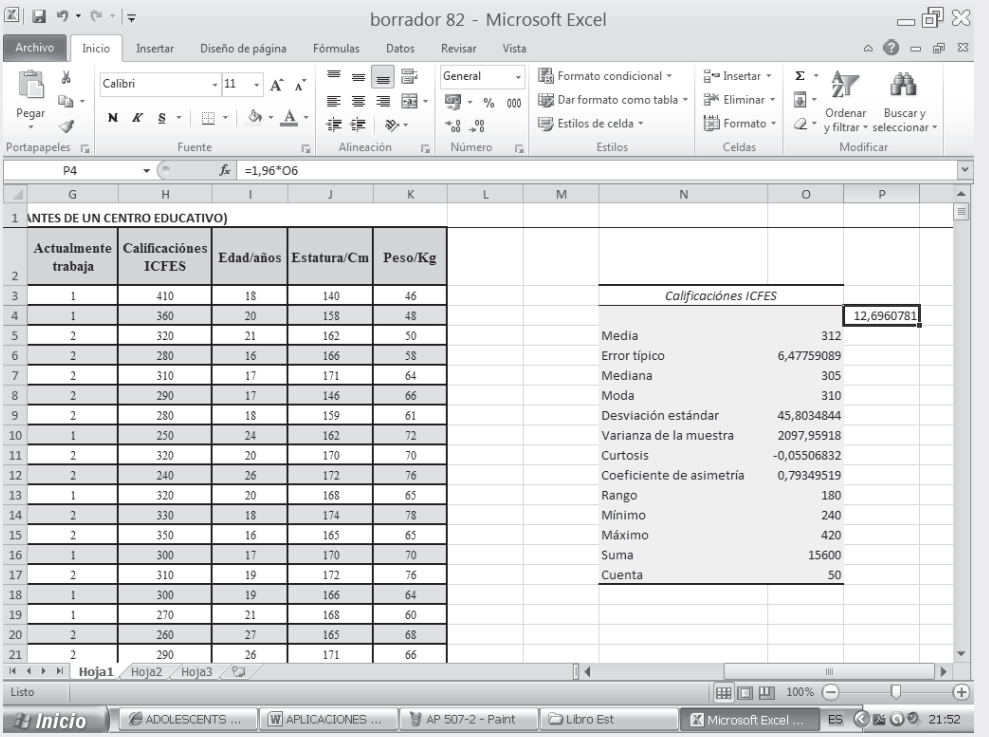
Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	410	18	140	46
1	360	20	158	48
2	320	21	162	50
2	280	16	166	58
2	310	17	171	64
2	290	17	146	66
2	280	18	159	61
1	250	24	162	72
2	320	20	170	70
2	240	26	172	76
1	320	20	168	65
2	330	18	174	78
2	350	16	165	65
1	300	17	170	70
2	310	19	172	76
1	300	19	166	64
1	270	21	168	60
2	260	27	165	68
2	290	26	171	66

Calificaciónes ICFES

Media	312
Error típico	6,47759089
Mediana	305
Moda	310
Desviación estándar	45,8034844
Varianza de la muestra	2097,95918
Curtosis	-0,05506832
Coeficiente de asimetría	0,79349519
Rango	180
Mínimo	240
Máximo	420
Suma	15600
Cuenta	50

- Si reemplazamos los resultados en la siguiente expresión $z\ \frac{s}{\sqrt{n}}$ tenemos: Con un nivel de confianza del 95%, z toma un valor de 1.96; y $\frac{s}{\sqrt{n}}$ corresponde al error típico, y toma un valor de 6,48. Por tanto, el producto de 1.96 * 6.48, nos da como resultado, 12,70, tal como aparece en la siguiente figura, en la celda P3.
- Lo anterior implica que podemos cambiar el valor de Z, dependiendo de la confianza establecida. Algunos de los resultados que aparecen en el cuadro, serán utilizados para determinar los límites.

Figura No.20. Cálculo Límites de Confianza.



$$\hat{\mu} = 312 \pm 12{,}70 \begin{cases} \hat{\mu}_s = 324{,}7 \\ \hat{\mu}_i = 299{,}3 \end{cases}$$

Lo anterior son los límites de confianza del 95% para $\hat{\mu}$

Para muestras peque as ≤ 30

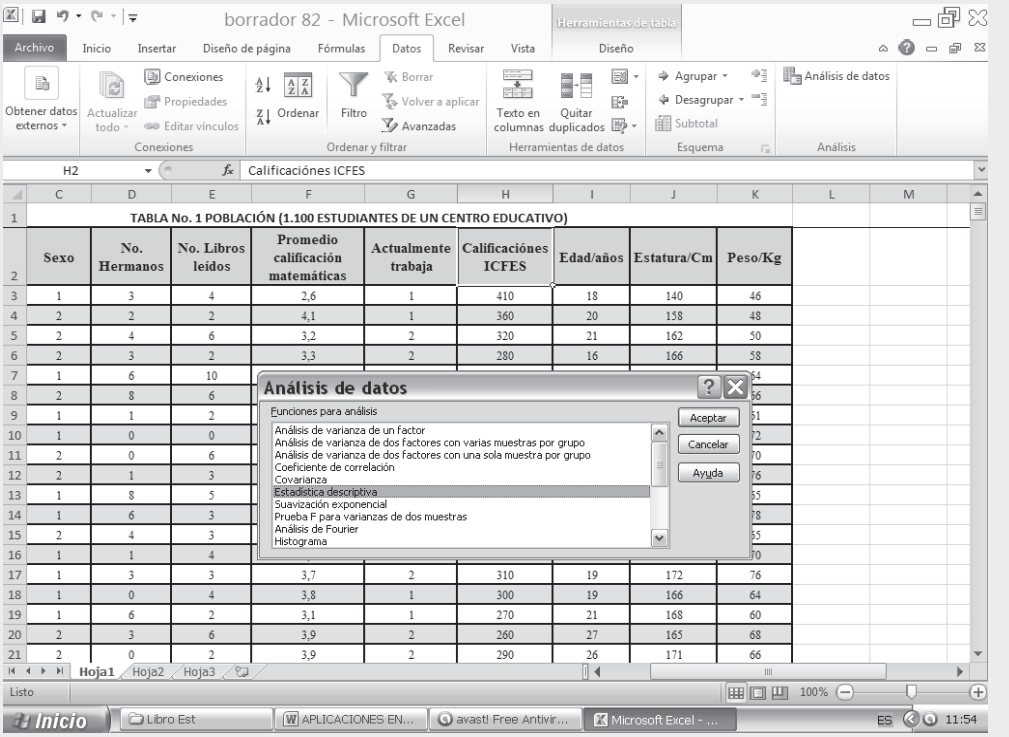
Fórmula: $\hat{\mu} = \bar{} \pm \sqrt{}$

Dado que una muestra pequeña corresponde a datos menores o iguales a 30, sólo tomaremos los primeros veinte (20) datos de la Tabla No. 1 de la columna H, esto es, una muestra de tan solo veinte estudiantes. La característica seguirá siendo CALIFICACIONES y el rango que vamos a seleccionar corresponde a

H3:H22. Además la confianza y la hipótesis son las mismas que utilizamos para el ejemplo de muestras grandes.

- Hacemos CLIC en la ficha DATOS y seleccionamos la opción ANÁLISIS DE DATOS.
- Hacemos CLIC en la opción ESTADÍSTICA DESCRIPTIVA en el cuadro de diálogo que nos aparece.

Figura No. 21. Análisis de Datos.



	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
3	1	3	4	2,6	1	410	18	140	46
4	2	2	2	4,1	1	360	20	158	48
5	2	4	6	3,2	2	320	21	162	50
6	2	3	2	3,3	2	280	16	166	58
7	1	6	10						54
8	2	8	6						56
9	1	1	2						61
10	1	0	0						72
11	2	0	6						70
12	2	1	3						76
13	1	8	5						65
14	1	6	3						78
15	2	4	3						65
16	1	1	4						70
17	1	3	3	3,7	2	310	19	172	76
18	1	0	4	3,8	1	300	19	166	64
19	1	6	2	3,1	1	270	21	168	60
20	2	3	6	3,9	2	260	27	165	68
21	2	0	2	3,9	2	290	26	171	66

En el cuadro de diálogo que resulta del proceso anterior, tecleamos y seleccionamos las siguientes opciones:

- RANGO DE ENTRADA, en nuestro caso el rango deberá ser seleccionado tecleando H2:H22.
- RÓTULOS EN LA PRIMERA FILA, este ícono deberá ser seleccionado haciendo CLIC en el cuadro que le antecede.
- Lo mismo deberá hacerse con el ítem RESUMEN DE ESTADÍSTICAS.
- De otra parte se procede a determinar el NIVEL DE CONFIANZA que en este caso como en los anteriores ejemplos es del 95%, por lo tanto, debemos teclear 95 en esta casilla.
- Nos falta establecer nuestro RANGO DE SALIDA, para ello seleccionamos la celda donde lo ubicaremos, en nuestro caso, señalamos la celda N3 y haciendo CLIC en ACEPTAR es posible observar los siguientes resultados:

Figura No. 22. Estadística Descriptiva.

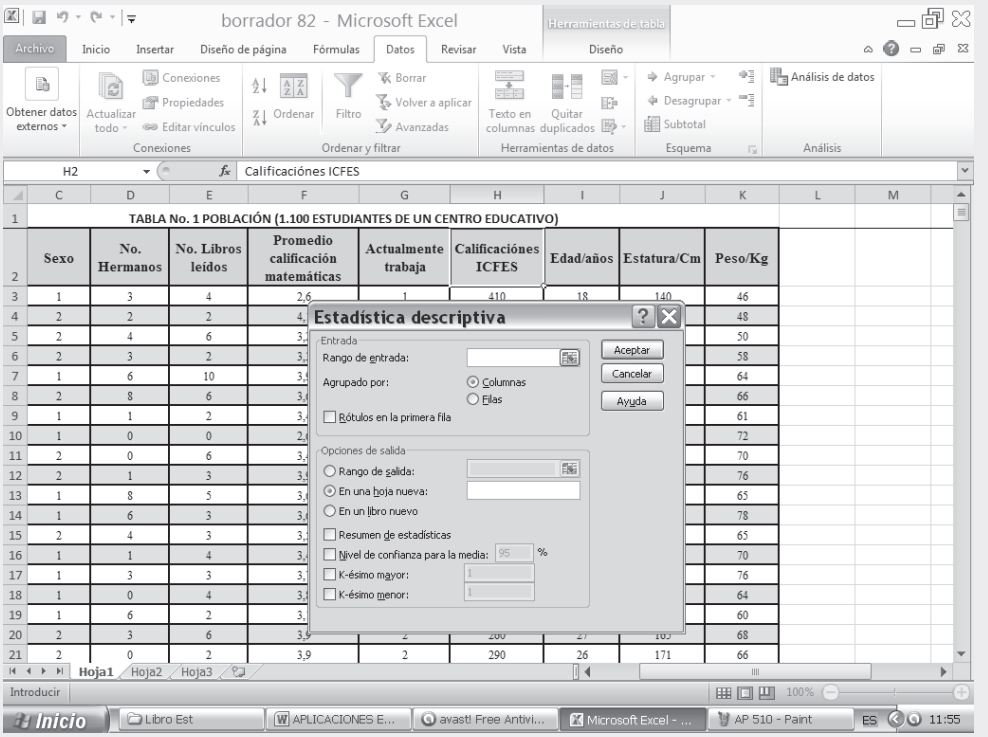
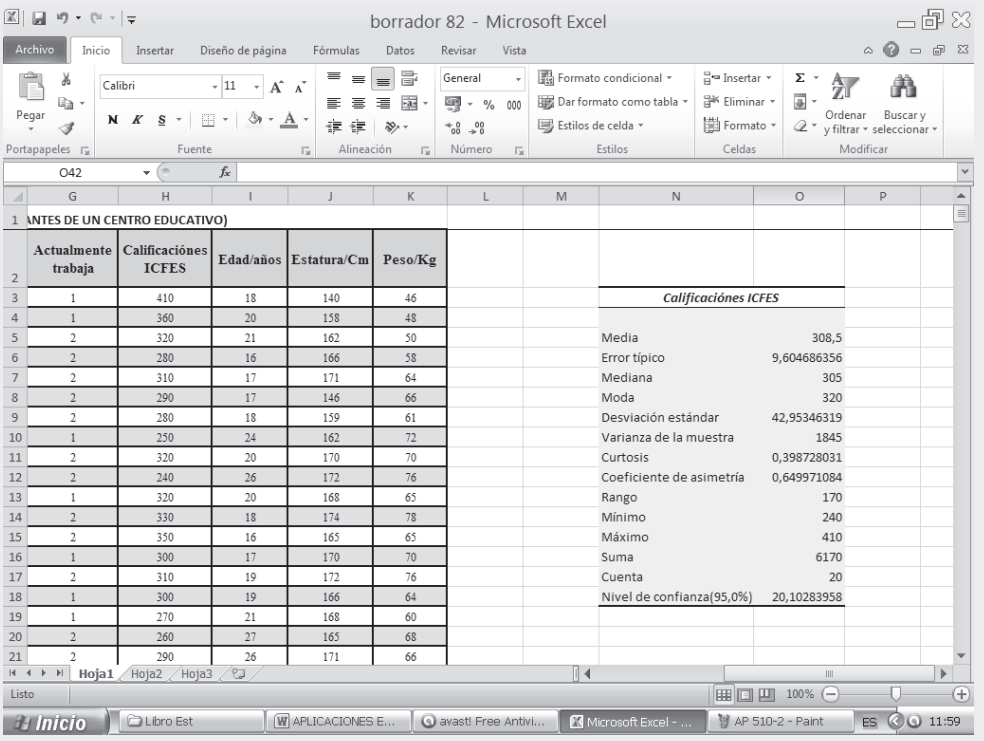


TABLA No. 1 POBLACIÓN (1.100 ESTUDIANTES DE UN CENTRO EDUCATIVO)

	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
3	1	3	4	2,6	1	410	18	140	46
4	2	2	2	4,					48
5	2	4	6	3,					50
6	2	3	2	3,					58
7	1	6	10	3,					64
8	2	8	6	3,					66
9	1	1	2	3,					61
10	1	0	0	2,					72
11	2	0	6	3,					70
12	2	1	3	3,					76
13	1	8	5	3,					65
14	1	6	3	3,					78
15	2	4	3	3,					65
16	1	1	4	3,					70
17	1	3	3	3,					76
18	1	0	4	3,					64
19	1	6	2	3,					60
20	2	3	6	3,9	2	200	27	165	68
21	2	0	2	3,9	2	290	26	171	66

Observemos que el cuadro de resultados aparece en la columna N y O a partir de la 3 hasta la línea 18.

Figura No. 23. Cuadro de Resultados.



	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
3	1	410	18	140	46
4	1	360	20	158	48
5	2	320	21	162	50
6	2	280	16	166	58
7	2	310	17	171	64
8	2	290	17	146	66
9	2	280	18	159	61
10	1	250	24	162	72
11	2	320	20	170	70
12	2	240	26	172	76
13	1	320	20	168	65
14	2	330	18	174	78
15	2	350	16	165	65
16	1	300	17	170	70
17	2	310	19	172	76
18	1	300	19	166	64
19	1	270	21	168	60
20	2	260	27	165	68
21	2	290	26	171	66

Calificaciónes ICFES	
Media	308,5
Error típico	9,604686356
Mediana	305
Moda	320
Desviación estándar	42,95346319
Varianza de la muestra	1845
Curtosis	0,398728031
Coeficiente de asimetría	0,649971084
Rango	170
Mínimo	240
Máximo	410
Suma	6170
Cuenta	20
Nivel de confianza(95,0%)	20,10283958

Reemplazando los resultados obtenidos en la fórmula siguiente podemos observar que:

$$\hat{\mu} = \bar{x} \pm t \frac{s}{\sqrt{n}}$$

$$\hat{\mu} = 308{,}5 \pm 2{,}093 \frac{9.60}{\sqrt{20}} = \qquad \hat{\mu} = 308{,}5 \pm 4{,}48 \begin{cases} 312.48 \\ 303{,}52 \end{cases}$$

Lo anterior nos indica que con un nivel de confianza del 95% los límites de confianza van desde 303.52 hasta 312,48.

Síntesis de capítulo

Hoy en día, se cuenta con cientos de procedimientos de *pruebas estadísticas* que pueden ser aplicadas a diferentes situaciones. Las más importantes se describen ampliamente en los diferentes textos de estadística, así, quien esté interesado no tendrá dificultad de encontrarlos junto con la guía de cómo hacerla. En esta unidad sólo se tomaron en cuenta cuatro tipos de pruebas, a saber: de *una media aritmética;* de *una proporción*; de la *diferencia entre dos medias* y la *diferencia entre dos proporciones*.

Además del desarrollo de estas pruebas, aplicadas en *muestras grandes* ($n > 30$) y *muestras pequeñas* ($n \leq 30$) que comprenden la distribución "t" de Student, se explicaron los procedimientos en la estimación puntual y de intervalos.

Se habla de *estimación puntual,* para designar el simple número que uno da a su mejor estimación correspondiente a una cantidad desconocida. *Estimación de intervalo* son dos números entre los cuales se pretende que ha de estar la cantidad desconocida, con cierto grado de confianza o de seguridad.

Tales intervalos suelen llamarse *límites de confianza* y fueron introducidos en EEUU por primera vez por Jerzy Newman en 1937.

La noción de *prueba estadística*, casi siempre, es nueva para el estudiante que se inicia en este campo del saber; generalmente, sin conocimiento, en muchas oportunidades las aplica. En general, se puede decir que el objeto de una prueba es proporcionar respuestas a diversos cuestionamientos, a fin de obtener conclusiones. En este capítulo, se emplea la palabra *prueba* para designar el procedimiento que se utiliza para establecer situaciones de *cierto o falso*, es decir, siempre se van a tener dos posibilidades: que sea diferente (*mayor o menor*) o igual.

Las *hipótesis*, para que sean de carácter estadístico, requieren hacer referencia a un valor estadístico, luego no todas las hipótesis son de carácter estadístico. Para ello es necesario la experimentación en el sentido de que la conclusión con respecto a la validez de tal hipótesis, se obtiene reuniendo datos a través de muestras aleatorias, para luego ser examinadas, calculando el estimado correspondiente.

Se puede decir que la prueba estadística sólo es, por tanto, el procedimiento experimental que se sigue cuando se quiere saber si es probable que una hipótesis estadística sea cierta.

Dentro de la realización de estas pruebas, es necesario identificar y saber definir términos como: *hipótesis nula*; *hipótesis alternativa*; *pruebas unilaterales* y *bilaterales*; *nivel de significancia*; *regiones criticas* (de rechazo) y *regiones de aceptación.*

9 CAPÍTULO

Otras pruebas de hipótesis

La inteligencia consiste no sólo en el conocimiento sino también en la destreza de aplicar los Conocimientos en la práctica.

Aristóteles

CONTENIDO

Pruebas de hipótesis aplicadas a:

- Una varianza y Límites de confianza. Varianzas de dos poblaciones.
- Coeficientes de correlación. Observaciones apareadas. Chi cuadrado.
- Prueba del signo. Correlación de Spearman. Prueba de T de Wilcoxon.
- Prueba de U de Mann-Whitney y de Kruskal- Wallis.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- ❑ Distinguir en qué casos en particular se debe aplicar cada una de las pruebas.
- ❑ Comprender la razón de su aplicación y hacer correctamente su interpretación.
- ❑ Entender con claridad los conceptos de poblaciones independientes y dependientes.
- ❑ Identificar las distribuciones paramétricas y no paramétricas.
- ❑ Manejar correctamente las tablas que se deban utilizar en cada prueba.

ALGUNOS ASPECTOS GENERALES

En los capítulos que anteceden, los temas tratados correspondían a *distribuciones muestrales*, todas ellas seleccionadas al azar, provenientes de *distribuciones paramétricas* y *poblaciones independientes.*

En este capítulo, nos dedicaremos en gran parte a las *Distribuciones no paramétricas* y algunas provenientes de poblaciones independientes de gran aplicación en las ciencias sociales.

PRUEBA DE HIPÓTESIS DE UNA VARIANZA

En gran parte, por no decir que en su totalidad, las unidades dedicadas a la *Inferencia estadística* van encaminadas hacia el promedio; sin embargo, las inferencias realizadas con la varianza, pueden constituirse en medidas más importantes que el promedio, pues este último, vale la pena recordar, es un punto de estimación, siendo aquel valor que consideramos típico, pero que no siempre lo va a ser, dado que depende del grado de variabilidad para ser representativo de un conjunto de observaciones.

Primero, se hará referencia a la prueba de una *sola varianza* obtenida mediante una muestra aleatoria, luego se realizarán con *dos varianzas*, supuestamente iguales en la población, cuando corresponde a dos muestras aleatorias independientes provenientes de distribuciones normales.

Partimos suponiendo que se tiene una *población normal* con media μ y varianza σ, generalmente desconocida. La simbología que se utilizará en el proceso será:

Hipótesis nula: $H_0 : \sigma^2 = \sigma_0^2$ equivale a → $H_0 : \dfrac{\sigma^2}{\sigma_0^2} = 1$

La alternativa puede ser: $H_a : \sigma^2 \neq \sigma^2$ $H_a : \sigma^2 > \sigma_0^2$ $H_a : \sigma^2 < \sigma_0^2$ que pueden ser presentadas también como:

$H_a : \dfrac{\sigma^2}{\sigma_0^2} \neq 1$	$H_a : \dfrac{\sigma^2}{\sigma_0^2} > 1$	$H_a : \dfrac{\sigma^2}{\sigma_0^2} < 1$
Diferente	*Mayor*	*Menor*

Para la realización de esta prueba se emplea la variante estadística *chi-cuadrado* con n –1 grados de libertad y está dada de la siguiente manera:

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma_0^2}$$

Recordemos que n –1 lo simbolizamos por la letra υ (nu) que significa grados de libertad.

Veamos el proceso que se sigue en esta prueba de hipótesis y el uso de la tabla para calcular los valores críticos.

▸ Establecer la hipótesis: $H_0 : \dfrac{\sigma^2}{\sigma_0^2} = 1$ $H_a : \dfrac{\sigma^2}{\sigma_0^2} \neq 1$

▸ Nivel de significación: $\propto$

- Establecer los supuestos: $\begin{cases} a.\ muestra\ \ aleatoria \\ b.\ población\ \ normal \end{cases}$
- Formular la variante estadística: $\dfrac{\chi^2}{\upsilon} = \dfrac{s^2}{\sigma_0^2}$
- Determinar los valores críticos $\left(\dfrac{\chi^2}{\upsilon}\right)_{\frac{\propto}{2};\upsilon}$ $\left(\dfrac{\chi^2}{\upsilon}\right)_{1-\frac{\propto}{2};\upsilon}$
- Calcular el valor estadístico, analizar y adoptar la decisión.

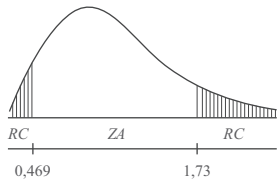
Ejemplo 1. Docimar la hipótesis de que $\sigma = 8$, dada $\hat{S} = 10$ para una muestra de tamaño 20.

Solución:

1) $H_0 : \dfrac{\sigma^2}{\sigma_0^2} = \dfrac{\sigma^2}{64} = 1$ $\qquad H_a : \dfrac{\sigma^2}{64} \neq 1$

2) $\propto = 0{,}05$

3) a) La muestra es aleatoria
   b) La población es normal



4) $\dfrac{\chi^2}{\upsilon} = \dfrac{s^2}{64}$ $\qquad$ e) $\upsilon = n - 1 = 20 - 1 = 19$

(Ver Tabla 9.1 página 409 de Percentiles de la distribución $\dfrac{\chi^2}{\upsilon}$)

$$\left(\frac{\chi^2}{\upsilon}\right)_{0{,}025\ ;\ 19} = 0{,}469 \qquad \left(\frac{\chi^2}{\upsilon}\right)_{0{,}975\ ;\ 19} = 1{,}73$$

5) $\left(\dfrac{\chi^2}{\upsilon}\right) = \dfrac{s^2}{64} = \dfrac{100}{64} = 1{,}56$

0, 469 < 1,56 < 1,73, se sitúa en la zona de aceptación; se puede afirmar al nivel del 5% que la varianza de la muestra, puede corresponder a una población cuya varianza es 64.

Límites de confianza

Ejemplo 2. Para obtener los límites para S con una confianza del 95%, requiere del siguiente proceso: siendo: $S = 10$ $S^2 = 100$ $\upsilon = 19$ $n = 20$

Solución:

$$\left(\frac{\chi^2}{\upsilon}\right)_{0{,}025\ ;\ 19} = 0{,}469 \qquad \left(\frac{\chi^2}{\upsilon}\right)_{0{,}975\ ;\ 19} = 1{,}73$$ (Ver tabla 9.1 de percentiles)

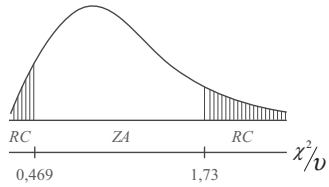
$$0{,}469 < \frac{\chi^2}{\upsilon} < 1{,}73 \ \Rightarrow\ 0{,}469 < \frac{S^2}{\sigma^2} < 1{,}73 \ \Rightarrow\ \frac{1}{0{,}469} > \frac{\sigma^2}{S^2} > \frac{1}{1{,}73} \ \Rightarrow\ \frac{1}{0{,}469} > \frac{\sigma^2}{100} > \frac{1}{1{,}73}$$

$$\frac{100}{0,469} > \sigma^2 > \frac{100}{1,73} \Rightarrow \frac{10}{\sqrt{0,469}} > \sigma > \frac{10}{\sqrt{1,73}} \Rightarrow \frac{10}{1,315} < \sigma < \frac{10}{0,684} \Rightarrow 7,60 < \sigma < 14,62$$

Ejemplo 3. En una ciudad se utilizan bombillas de la marca A durante años, pero no se contempla cambio a la nueva marca B debido a su mejor precio. Los vendedores de la marca B, afirman que su producto es tan bueno como el de la marca A. A fin de docimar la afirmación de los vendedores de la marca B, se probaron 81 de tales bombillas, que fueron adquiridas en los establecimientos del ramo, obteniéndose una desviación estándar de 80 horas en el tiempo de encendido, mientras que la larga experiencia con ampolletas de la marca A, ha mostrado una desviación estándar de 90 horas. Si se admite que no existe experiencia entre las dos marcas, se trata de docimar la hipótesis de que la desviación estándar de las bombillas marca B es igual al de la marca A.

Solución:

a) $\begin{cases} H_0 : \dfrac{\sigma^2}{\sigma_0^2} = \dfrac{\sigma^2}{90^2} = 1 \\ H_a : \dfrac{\sigma^2}{90^2} \neq 1 \end{cases}$



b) $\propto = 0,05$ c) $\dfrac{\chi^2}{\upsilon} = \dfrac{\hat{s}^2}{90^2}$ d) $\left(\dfrac{\chi^2}{\upsilon}\right)_{0,025\ ;\ 80} = 0,714$; $\left(\dfrac{\chi^2}{\upsilon}\right)_{0,025\ ;\ 80} = 1,33$

e) $\left(\dfrac{\chi^2}{\upsilon}\right) \leq 0,714$; $\left(\dfrac{\chi^2}{\upsilon}\right) \geq 1,33$ f) $\left(\dfrac{\chi^2}{\upsilon}\right) = \dfrac{80^2}{90^2} = 0,79$

Aceptamos $H_0 : \dfrac{\sigma^2}{90^2} = 1$, por lo tanto las desviaciones estándar, pueden considerarse iguales.

Tabla 9.1 De percentiles de la distribución χ^2

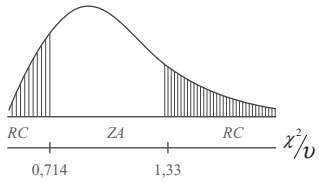
υ	.05	.1	.5	1.0	2.5	5.0	1.0	....	90	95	97.5	99	99.5	99.9	99.95	υ
1	$.0^639$	$.0^5157$	$.0^439$	$.0^316$	$.0^398$	$.0^239$	.016	....	2.71	3.84	5.02	6.64	7.88	10.83	12.12	1
2	.001	.001	.005	.010	.025	.052	.106	....	2.30	3.00	3.69	4.61	5.30	6.91	7.60	2
3	.005	.008	.024	.038	.072	.117	.195	....	2.08	2.60	3.12	3.78	4.28	5.42	5.91	3
4	.016	.023	.052	.074	.121	.178	.266	....	1.94	2.37	2.79	3.32	3.72	4.62	5.00	4
5	.032	.042	.082	.111	.166	.229	.322	....	1.85	2.21	2.57	3.02	3.35	4.10	4.42	5
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
16	.221	.246	.321	.363	.432	.498	.582	....	1.47	1.64	1.80	2.00	2.14	2.45	2.58	16
17	.234	.260	.335	.337	.445	.510	.593	....	1.46	1.62	1.78	1.97	2.10	2.40	2.52	17
18	.247	.272	.348	.390	.457	.522	.604	....	1.44	1.60	1.75	1.93	2.06	2.35	2.47	18
19	.258	.285	.360	.402	.469	.532	.613	....	1.43	1.59	1.73	1.90	2.03	2.31	2.42	19
20	.270	.296	.372	.413	.480	.543	.622	....	1.42	1.57	1.71	1.88	2.00	2.27	2.37	20
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
60	.506	.529	.592	.625	.675	.720	.774	....	1.24	1.32	1.39	1.47	1.53	1.66	1.71	60
70	.535	.558	.618	.649	.697	.739	.790	....	2.22	1.29	1.36	1.43	1.49	1.60	1.65	70
80	.560	.582	.640	.669	.714	.755	.803	....	1.21	1.27	1.33	1.40	1.45	1.56	1.60	80
90	.581	.602	.658	.686	.729	.768	.814	....	1.20	1.26	1.31	1.38	.143	.152	1.56	90
100	.599	.619	.673	.701	.742	.779	.824	....	1.18	1.24	1.30	1.36	1.40	1.49	1.53	100

Nota: las cifras en cuadros blancos corresponden a los dos ejercicios anteriores.

Ejemplo 4. Docimar la hipótesis $\sigma = 8$, dada $\hat{s} = 10$ para una muestra de tamaño $n = 20$

Solución:

$n = 20 \qquad \sigma = 8 \qquad s = 10$



a) $H_0 : \dfrac{\sigma^2}{\sigma_0^2} = \dfrac{\sigma^2}{64} = 1\,; \qquad H_\alpha : \dfrac{\sigma^2}{\sigma_0^2} \neq 1$

b) $\propto = 0{,}05$ c) $\dfrac{\chi^2}{\upsilon} = \dfrac{\hat{s}^2}{64^2}$

d) $\left(\dfrac{\chi^2}{\upsilon}\right)_{0{,}025\ ;\ 19} = 0{,}469 \qquad \left(\dfrac{\chi^2}{\upsilon}\right)_{0{,}975\ ;\ 19} = 1{,}73$

e) $\left(\dfrac{\chi^2}{\upsilon}\right) \leq 0{,}469 \qquad \left(\dfrac{\chi^2}{\upsilon}\right) \geq 1{,}73$

f) $\left(\dfrac{\chi^2}{\upsilon}\right) = \dfrac{100}{64} = 1{,}5625$ g) Se acepta la hipótesis $H_0 : \dfrac{\sigma^2}{\sigma_0^2} = 1$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. Determinar los límites de confianza del 95% para $\sigma = 8$, y $\hat{s} = 10$ para una muestra de tamaño: (a) 20 (b) 51.

2. Efectuar el ejercicio anterior, considerando muestras de tamaño: (a) 70 (b) 120.

3. Dada X normalmente distribuida y los valores de la muestra $n = 15$ y $\hat{s} = 7$, docimar la hipótesis de $\sigma = 5$.

4. Con los datos del ejercicio anterior, encontrar los límites de confianza de s. Utilizar:

   (a) el 95% (b) el 99%.

5. Docimar la hipótesis $\sigma = 20$, dada $\hat{s} = 10$ para una muestra de tamaño 25.

6. Los siguientes datos representan las cantidades de un compuesto químico determinado, obtenidos en análisis diarios.
   La experiencia de mucho tiempo, da el valor 0,5 para la desviación estándar.
   (a) Docimar la hipótesis de que $\sigma = 0{,}5$ para estos análisis.
   (b) Hallar los límites de confianza del 95%, para σ de estos análisis.

12,7	12,3	13,2	12,8	13,6	13,1	12,6	12,4	14,1	13,3	13,4
13,1	12,6	12,9	13,0	12,4	14,6	13,8	13,4	12,7	13,5	12,5

7. Para 100 empleados de una Cía. se observa el salario mínimo mensual expresado en miles de pesos

(X), con el fin de determinar si la varianza anterior de todos los salarios fue de $\sigma^2 = 8$; se obtuvieron los siguientes resultados de esa observación:

$$\Sigma x_i^2 = 2.000 \qquad \frac{(\Sigma x_i)^2}{n} = 1.000$$

Suponiendo que se corre un riesgo de equivocarse del 5% ¿qué conclusión obtiene la Cía.?

PROCEDIMIENTO MÁS UTILIZADO EN LA PRUEBA CON UNA VARIANZA

Anteriormente se realizaron pruebas de hipótesis sobre *la varianza de una población*, usando para ello la tabla de Percentiles χ^2/v. Sin embargo, este procedimiento a pesar de lo fácil en su aplicación, es costumbre generalizada utilizar en el proceso la tabla de *Chi – cuadrado*.

Si consideramos, como ejemplo, la fabricación de algún instrumento de medición, que requiera alta precisión, el técnico dedicado al control de calidad, además de interesarle la obtención de un promedio o una proporción, le concede mayor importancia a calcular el *grado de variabilidad*, con el fin de mantener o reducir las diferencias que se puedan presentar.

En la contrastación de hipótesis, el procedimiento a seguir, es el mismo que se ha venido utilizando en éste y en el anterior capítulo. Veamos como se procede con la varianza:

Se formula la hipótesis:	1) $H_0 : \sigma^2 = \sigma_0^2$ (hipótesis nula) $H_a : \sigma^2 \neq \sigma_0^2$ (hipótesis alternativa)
Grado de significación:	2) $\propto = 0{,}05$
Variante estadística:	3) $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$

Así como se realiza una prueba bilateral, también pueden efectuarse pruebas unilaterales (a la derecha o izquierda), teniendo en cuenta las necesidades o el enunciado del ejercicio.

Unilateral a la derecha	*Unilateral a la izquierda*
$H_0 : \sigma^2 = \sigma_0^2$	$H_0 : \sigma^2 = \sigma_0^2$
$H_a : \sigma^2 > \sigma_0^2$	$H_a : \sigma^2 < \sigma_0^2$

Ahora procedemos a su aplicación, mediante ejercicios prácticos.

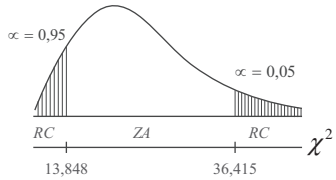
Ejemplo 1. Un buen profesor, confecciona un tema de examen que permita determinar si se logró un buen aprendizaje, a través de un nuevo método de enseñanza establecido gracias a su experiencia. Considera que los resultados deben presentar muy poca variabilidad, establece una $\sigma = 0,65$. Realiza una prueba en 25 de sus estudiantes, con los siguientes resultados: $\bar{x} = 3,6$ y $S = 0,80$. Determinar, al nivel del 5%, que la dispersión obtenida en la muestra, es igual a la establecida.

Solución:

1) $H_0 : \sigma^2 = 0,4225$
$H_a : \sigma^2 \neq 0,4225$

2) $\propto = 0{,}05$

3) $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} = \dfrac{(25-1)\ 0{,}8^2}{0{,}65^2} = 36{,}36$



En la tabla de observamos que para $v = n - 1 = 25 - 1 = 24$, el valor de $\chi^2_{0,95}$ =13,848 y $\chi^2_{0,05}$ = 36,415 (Ver tabla IV - Chi-cuadrado). Como x^2 (calculado) es de 36,36, cae en la zona de aceptación, es decir, entre los límites 13, 848 y 36, 415, permitiendo concluir que el examen cumple con la condiciones fijadas por el profesor al nivel del 5%.

Ejemplo 2. Supongamos que se aplica la prueba, con la expectativa de que el resultado de la varianza obtenida en la muestra de 25 estudiantes, sea superior al esperado por el profesor al nivel del 5%.

Solución:

$\sigma = 0{,}65 \qquad \sigma^2 = 0{,}4225 \qquad s = 0{,}8 \qquad s^2 = 0{,}64 \qquad \propto = 0{,}05$

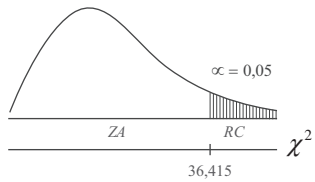
1) $H_0 = \sigma^2 = 0{,}4225$

$H_a : \sigma^2 > 0{,}4225$

2) $\propto = 0{,}05$

3) $\chi^2 = \dfrac{(n-1)\ s^2}{\sigma^2} = \dfrac{(25-1)\ 0{,}64}{0{,}4225} = 36{,}36$

$\left.\begin{matrix} v = n-1 = 24 \\ \alpha = 0{,}05 \end{matrix}\right\} \chi^2_{0,05} = 36{,}42$



Se concluye, que el grado de variabilidad obtenida en la muestra, al nivel del 5%, no es superior al señalado por el profesor. Sin embargo, algunos autores cuando realizan una prueba unilateral, toman el doble para la región crítica, tal como lo hacíamos con la "t" *de Student*, de acuerdo a este procedimiento, el *Chi* de tabla sería $\chi^2_{0,10}$ = 33,20 , por lo tanto concluiríamos, que los resultados obtenidos son superiores al esperado por el profesor, al nivel del 5%.

Límites de confianza

Recordemos que el fijar límites, siempre debe considerarse como bilateral, además, se trabaja con la varianza muestral. Es recomendable, antes de aceptar o rechazar la hipótesis nula, fijar límites de confianza que sirvan para establecer dentro de qué valores debe estar la varianza poblacional, con cierto grado de confiabilidad. La fórmula empleada es:

$$\frac{(n-1)\ s^2}{\chi^2_s} < \sigma^2 < \frac{(n-1)\ s^2}{\chi^2_i}$$

χ^2_i = Valor inferior

χ^2_s = Valor superior

Ejemplo 3. Sean los valores:

$s = 0,8$ $s^2 = 0,64$ $n = 25$ $P = 95\%$

Solución:

$$\frac{(n-1)\ s^2}{\chi_s^2} < \sigma^2 < \frac{(n-1)\ s^2}{\chi_i^2}$$

$$\frac{(25-1)\ 0,64}{36,415} < \sigma^2 < \frac{(25-1)\ 0,64}{13,848} \qquad \frac{15,36}{36,415} < \sigma^2 < \frac{15,36}{13,848}$$

Por lo tanto los límites para la varianza poblacional son: $0,42 < \sigma^2 < 1,11$. La varianza poblacional, estará entre 0,42 y 1,11, con una confianza del 95%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

8. Todos los ejercicios dados para resolver en el tema anterior sobre varianza, aplicando la tabla de percentiles de χ^2/υ, pueden ser utilizados para trabajar con la tabla de *Chi- cuadrado* (χ^2). Se le deja al lector su aplicación.

9. Un fabricante de un determinado producto, establece una duración promedio de trabajo continuo de 5.000 horas y que $\sigma \leq 400$ horas. Se examinó un muestra de 40 de este producto y se encontró que su media es de 5.500 y varianza de 172.000 horas^2. Al nivel del 1% ¿estos datos proporcionan una evidencia que σ sea mayor a 400?

10. Un vendedor experto que trabaja en un almacén, asegura que el tiempo dedicado a la atención de un cliente en promedio es de 25 minutos, con una desviación típica (σ) de 8 minutos. Se pregunta, si el grado de variación en la atención al cliente es diferente. Una muestra a 20 clientes da una desviación típica de 10 minutos. Pruebe la hipótesis al nivel del 10%.

11. En una muestra de $n = 25$, se obtiene una $s^2 = 67,8$. Al nivel del 5%
    a) ¿Indica este resultado que s^2 es superior a 65?
    b) Fije los límites de confianza del 95% para la varianza poblacional.

12. Una muestra de $n = 28$, produce una desviación típica de 168,5. Al nivel del 10%
    a) ¿Se puede deducir que $\sigma^2 = 171$?
    b) Fije los límites de confianza del 90% para la varianza poblacional.

COMPARACIÓN ENTRE VARIANZAS DE DOS POBLACIONES

DISTRIBUCIÓN F

Cuando se realizó la aplicación de la "t" de Student, en (muestras pequeñas), pruebas de hipótesis para diferencias entre medias muestrales, las establecíamos con base en dos supuestos:

a) El primero, el más utilizado, que las muestras provienen de dos poblaciones con *varianzas idénticas*, aceptando a ciegas que fuera cierto.

b) El otro, bajo el supuesto, que las varianzas poblacionales fueran diferentes. Ahora, se trata de probar mediante la aplicación de la *Distribución F*, la validez de que las varianzas poblacionales, sean idénticas o diferentes.

Es de anotar que la *Distribución* F, desarrollada por Sir Ronald Fisher, es una medida muy diferente a las otras hasta aquí desarrolladas, *no es simétrica* y su forma depende del número de grados de libertad asociadas con s_1^2 y s_2^2.

Siguiendo procedimientos conocidos, primero se plantean las hipótesis H_o y H_a; luego el nivel de significación; se determina el valor F en la variante estadística y finalmente se toma la decisión de aceptar o rechazar que las varianzas poblacionales sean idénticas.

$$1)\ \begin{matrix} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_a : \sigma_1^2 \neq \sigma_2^2 \end{matrix} \qquad 2)\ \propto = 0{,}05 \qquad 3)\ F = \frac{s_1^2}{s_2^2} = \frac{\text{Varianza mayor}}{\text{Varianza menor}}$$

Consulta de la tabla F

Al final del libro aparece la Tabla XVI cuyo proceso de utilización es el siguiente: en la parte superior (nominador) correspondiente a la fórmula de F, aparece la varianza de mayor valor y en su parte inferior (denominador) la de menor valor. Establecido el orden de las varianzas, determinamos los respectivos grados de libertad ($\upsilon_1 = n_1 - 1$ y $\upsilon_2 = n_2 - 1$); el primero, se leerá en la primera línea o renglón de la tabla y el segundo υ_2 en la primera columna. Una vez localizados los dos valores, buscamos su punto de intersección siendo éste el valor correspondiente a F.

Ejemplo 1. Localizar los valores de F, utilizando la tabla respectiva, dependiendo del nivel de significación.

a) $\upsilon_1 = 12$, $\upsilon_2 = 16$, $\propto = 0{,}05$

b) $\upsilon_1 = 15$, $\upsilon_2 = 20$, $\propto = 0{,}01$

c) $\upsilon_1 = 40$, $\upsilon_2 = 80$, $\propto = 0{,}10$

d) $\upsilon_1 = 24$, $\upsilon_2 = 17$, $\propto = 0{,}025$

Solución:

a) $F = 2{,}42$ b) $F = 3{,}09$ c) $F = 1{,}40$ d) $F = 2{,}56$

Para los *límites de confianza*, tendrá que ser bilateral, por lo tanto el proceso para determinar los puntos críticos, será:

a) $\upsilon_1 = 15$, $\upsilon_2 = 20$, $\propto = 0{,}05$, $F_{\upsilon_1,\upsilon_2} = 2{,}20$, $F_{\upsilon_2,\upsilon_1} = 2{,}33$, $F_s = 2{,}20$, $F_t = \frac{1}{2{,}33} = 0{,}429$

b) $\upsilon_1 = 30$, $\upsilon_2 = 40$, $\propto = 0{,}01$, $F_{\upsilon_1,\upsilon_2} = 2{,}20$, $F_{\upsilon_2,\upsilon_1} = 2{,}30$, $F_s = 2{,}20$, $F_t = \frac{1}{2{,}30} = 0{,}435$

c) $\upsilon_1 = 8$, $\upsilon_2 = 15$, $\propto = 0{,}025$, $F_{\upsilon_1,\upsilon_2} = 3{,}20$, $F_{\upsilon_2,\upsilon_1} = 4{,}10$, $F_s = 3{,}20$, $F_t = \frac{1}{4{,}10} = 0{,}24$

Los anteriores cálculos se hicieron con fines de explicación, dado que los valores críticos a la izquierda de *F*, no aparecen en las Tablas (F_s = punto crítico superior; F_i = punto crítico inferior); por lo tanto el límite inferior se obtiene calculando el recíproco, al mismo tiempo que se intercambian los grados de libertad, υ_1 como υ_2 y υ_2 como si fuese υ_1.

Conociendo el manejo de la Tabla F, veamos cómo se procede en la realización de las pruebas de hipótesis, con el fin de establecer si *hay o no igual variabilidad* en las dos poblaciones, de donde se extrajeron las muestras.

Ejemplo 2. Supongamos la selección de dos muestras, de tamaños $n_1 = 21$ y $n_2 = 16$, cuyas varianzas muestrales son S_1^2=57,12 y S_2^2=25,68 respectivamente. ¿Dan estos resultados suficiente evidencia, que nos haga pensar que las varianzas poblacionales sean distintas? Nivel del 5%.

Solución:

$\upsilon_1 = 21-1 = 20$ $\quad S_1^2 = 57{,}12$ $\quad F_{\upsilon_1,\upsilon_2} = 2{,}57$

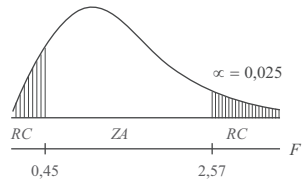
$\upsilon_2 = 16-1 = 15$ $\quad S_2^2 = 25{,}68$ $\quad F_{\upsilon_2,\upsilon_1} = 2{,}20$

1) $H_0 : \sigma_1^2 = \sigma_2^2$

$H_\alpha : \sigma_1^2 \neq \sigma_2^2$

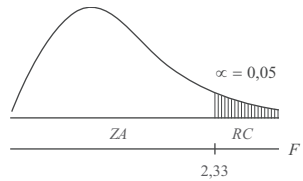
2) $\propto = 0{,}05$

3) $F = \dfrac{S_1^2}{S_2^2} = \dfrac{57{,}12}{25{,}68} = 2{,}22$



Estos resultados no muestran evidencia, al nivel del 5%, que las varianzas poblacionales sean distintas.

Nota: No hay necesidad de localizar el punto crítico inferior: (1/2,20 = 0,45), dado que *F* calculado siempre va a ser positivo, ($F \geq 1$), pues en el numerador se coloca el mayor valor de S^2.



Sin embargo, numerosos autores, cuando la *prueba es bilateral*, toman el total del nivel de significación para cada zona crítica, dándole el mismo tratamiento que se dio en la distribución "t" de Student, por lo tanto el valor para el punto crítico será $F_{\upsilon_2,\upsilon_2} = 2{,}33$. De acuerdo con este procedimiento *F* = 2,22, cae en la zona de aceptación, es decir, aceptamos H_0, luego podemos concluir que las varianzas poblacionales son idénticas, al nivel del 5%.

Ejemplo 3. Se realizan dos procesos diferentes en la fabricación de un determinado producto. Durante un período de observación de 31 días, se determinó el número de unidades producidas con sus respectivas varianzas, estos resultados en su orden fueron de: 290 y 296 unidades, con varianzas de 1.102,24 y 528,70 respectivamente ¿Presentan estos resultados suficiente evidencia para concluir, al nivel del 5%, que la variabilidad en la producción diaria, es superior en uno de los procesos?

Solución:

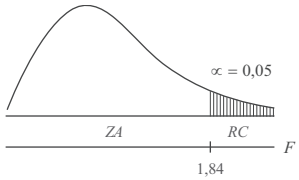
$n_1 = 31 \qquad \upsilon_1 = 30 \qquad F_{\upsilon_1,\upsilon_2} = 1{,}84 \qquad s_1^2 = 1.102{,}24$

$n_2 = 31 \qquad \upsilon_2 = 30 \qquad\qquad\qquad\quad s_2^2 = \ 528{,}70$

1) $H_0 = \sigma_1^2 = \sigma_2^2$ 2) $\propto = 0{,}05$

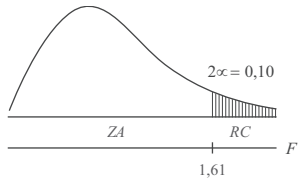
$H_a = \sigma_1^2 > \sigma_2^2$

3) $F = \dfrac{1.102{,}24}{528{,}70} = 2{,}08$



Al nivel del 5%, hay base suficiente para concluir que una de las varianzas es significativamente superior a la otra.

Como se vió en el caso bilateral, también hay divergencias entre los autores en la determinación del valor correspondiente a la región crítica. En este ejercicio se siguió el procedimiento aplicado en la normal (trabajando en *z*), sin embargo otros toman el doble del nivel de significación, es decir, en este caso sería del 10%, tal como se trabajó con la "t" de Student, así, que el valor del punto crítico sería $F_{\upsilon_1,\upsilon_2} = 1{,}61$ en vez de 1, 84.



Límites de confianza

También con la *distribución F*, se pueden establecer límites de confianza para el cuociente de las varianzas, aplicando la siguiente fórmula:

$$\frac{s_1^2}{s_2^2}\ \frac{1}{F_{\upsilon_1,\upsilon_2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}\ F_{\upsilon_2,\upsilon_1}$$

Ejemplo 4. Durante una jornada en la bolsa, se tomaron dos grupos de industrias, con 30 y 41 empresas cada una, registrándose los precios de cierre en sus acciones, siendo sus medias de $403,3 y $425,4 y varianzas de 15,4 y 29,6 respectivamente. Encuentre un intervalo de confianza del 95% para el cuociente de las varianzas.

$s_1^2 = 29{,}6 \qquad \upsilon_1 = 30 \qquad F_{\upsilon_1,\upsilon_2} = 1{,}94 \qquad \dfrac{\propto}{2} = \dfrac{0{,}05}{2} = 0{,}025$

$s_2^2 = 15{,}4 \qquad \upsilon_2 = 40 \qquad F_{\upsilon_2,\upsilon_1} = 2{,}01$

$$\frac{29{,}6}{15{,}4}\left[\frac{1}{1{,}94}\right] < \frac{\sigma_1^2}{\sigma_2^2} < \frac{29{,}6}{15{,}4}\ (2{,}01)$$

$$\frac{1{,}92}{1{,}94} < \frac{\sigma_1^2}{\sigma_2^2} < 1{,}92\ (2{,}01) \quad \rightarrow \quad 0{,}99 < \frac{\sigma_1^2}{\sigma_2^2} < 3{,}86$$

El intervalo para el cuociente poblacional, estará entre 0,99 y 3,86 con el 95% de confianza. Aquí también hacemos la misma observación establecida para la prueba bilateral y la unilateral, sobre la forma como se determina el valor de la región crítica. Observemos que α= 0,05 se distribuye la mitad para cada

lado, por lo tanto utilizamos la tabla de F para determinar el valor correspondiente a $\alpha = 0{,}0250$, de ahí en adelante el proceso es igual al que se hizo para la normal (muestras grandes), sin embargo algunos proceden, al igual que en el caso de las muestras pequeñas, así, cuando $\alpha = 0{,}05$:

$$\frac{29{,}6}{15{,}4}\left[\frac{1}{1{,}74}\right] < \frac{\sigma_1^2}{\sigma_2^2} < \frac{29{,}6}{15{,}4}\ (1{,}79) \qquad F_{\upsilon_1,\upsilon_2} = 1{,}74 \qquad F_{\upsilon_2,\upsilon_1} = 1{,}79$$

$$\frac{1{,}92}{1{,}74} < \frac{\sigma_1^2}{\sigma_2^2} < 1{,}92\ (1{,}79) \qquad \frac{1{,}92}{1{,}74} < \frac{\sigma_1^2}{\sigma_2^2} < 1{,}92\ (1{,}79)$$

Nota: Se exponen estos procedimientos que pueden confundir (en algunos casos), para que el lector esté prevenido y conozca, que existen criterios diferentes en su aplicación. De acuerdo con la orientación dada por el profesor y a su criterio, podrá identificarse con alguno de ellos.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

13. Una serviteca desea corresponder a sus clientes por su confianza, en especial con los neumáticos para camiones que más acreditan el negocio. Ofrece dos marcas que tienen la misma duración, pero no está muy seguro de su variabilidad. Selecciona muestras de tamaño 16 y 21 neumáticos de cada marca, con varianzas de 36.000 y 42.000 km^2.
    a) ¿Hay evidencia de una diferencia entre las varianzas poblacionales?
    b) Fije los límites de confianza del 95%.

14. Se comparan dos métodos para realizar cierta operación. Supongamos que los resultados obtenidos en las dos muestras fueron: $\bar{x} = 725$; $s_x^2 = 61$ y $\bar{y} = 661$; $s_y^2 = 86$; y los tamaños muestrales son: 7 y 13 respectivamente. Docime o pruebe la hipótesis de que la segunda muestra (y), presenta mayor variabilidad que x.

15. Dos muestras sobre rendimiento académico, una aplicada a los estudiantes del diurno, con notas de 3,8 4,1 3,7 4,3 3,2 3,8 4,3 y otra al nocturno con calificaciones de: 3,3 3,8 3,4 4,9 4,8 3,1 4,7 4,9. Pruebe la hipótesis al 2%, que no hay diferencia en cuanto al grado de variabilidad.

PRUEBA DEL COEFICIENTE DE CORRELACIÓN DE PEARSON r = R

Cuando en un problema de *análisis de regresión*, como se verá en los capítulos 11 y 12, se calcula el *Coeficiente de correlación*, es importante saber si ese resultado pudo haberse dado al azar, con una buena probabilidad en una muestra de tamaño considerado.

Para efectuar la prueba de hipótesis en la correlación, cuando la muestra es pequeña, se emplea la distribución "t", y "z" para n >30 obteniéndose mediante la aplicación de la siguiente fórmula:

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

Siendo r el coeficiente de correlación y n el número de pares observados.

También se puede escribir la fórmula de "t" y la del coeficiente de correlación de la siguiente manera:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \qquad r = \frac{n\Sigma x_i\, y_i - (\Sigma x_i)\,(\Sigma y_i)}{\sqrt{\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right]\left[n\Sigma y_i^2 - (\Sigma y_i)^2\right]}}$$

El valor de r se obtiene generalmente mediante la aplicación de esta fórmula y otras más, con las cuales se llega al mismo resultado, como ya se dijo, podrá observarse en los capítulos 11 y 12 en los temas *"Regresión y Correlación" y "Series Cronológicas"*.

La anterior fórmula se aplica, cuando la *Hipótesis nula* (H_0) se formula en el sentido de que no existe correlación entre las dos variables, es decir, que la correlación es igual a 0, frente a la hipótesis alternativa de que sea diferente (*dócima bilateral*); que sea mayor de 0 (*dócima unilateral a la derecha*) o menor de cero (*dócima unilateral a la izquierda*).

Ejemplo 1. En una muestra de 28 pares de observaciones de x y y, se obtuvo un Coeficiente de correlación de 0,42, ¿indica esto a un nivel de significación que existe una correlación significativa?

Solución:

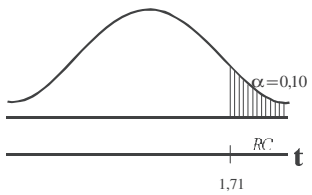
1) $H_0 : R = 0$ (No hay correlación)
$H_a : R > 0$ (Si hay correlación)

También: $H_0 : \rho = 0$, $H_1 : \rho > 0$

2) $\propto = 0{,}05$ 3) $t = r\sqrt{\frac{n-2}{1-r^2}}$

$$t = 0{,}42\sqrt{\frac{28-2}{1-0{,}42^2}} = 2{,}36$$

4) $\upsilon = n - 2 = 26$ $t = 1{,}706$



Rechazamos la hipótesis nula H_0 y aceptamos la hipótesis alternativa $H_a : R > 0$, es decir, que al nivel de 5% se puede aceptar una correlación mayor de 0, que puede ser significativa.

Cuando en la hipótesis nula se plantea que $R \neq 0$, sólo puede utilizarse para decidir si en una determinada muestra el coeficiente de correlación es significativamente diferente a 0 (*dócima bilateral*), aunque se puede efectuar la prueba unilateral de que sea mayor de 0, cuando R tiene signo positivo o menor de 0 cuando el signo es negativo. No hay procedimiento alguno para establecer límites de confianza de R o de ρ(*rho*).

Para muestras grandes (n =31 o más) se utiliza Z, siendo la fórmula más utilizada $Z = \frac{r}{\frac{1}{\sqrt{n-1}}}$; también se aplica otra fórmula diferente a la anterior y se trabaja con Z, de la siguiente manera: $Z = \frac{Z - \mu_z}{\sigma_z}$

Siendo: $\mu_z = \frac{1}{2}\ \log_n\left(\frac{1+r}{1-r}\right) = \frac{1}{2}\ (2{,}3026)\ \log\left(\frac{1+r}{1-r}\right)$ que también se puede presentar así:

$\mu_z = 1{,}1513 \log\left(\frac{1+r}{1-r}\right)$ para hallar el valor de $\boldsymbol{Z}$, se puede aplicar la fórmula anterior. Ambos valores se pueden obtener directamente sin necesidad de cálculo, consultando la tabla 9.2 de transformación de r a Z, o la tabla V al final del libro.

La desviación estándar se simbolizará por σ_z y se calcula así: $\sigma_z = \frac{1}{\sqrt{n-3}}$

Ejemplo 2. En una muestra de 40 pares de observaciones, el Coeficiente de correlación obtenido es 0,86. Use el nivel de significación del 5% para probar la hipótesis de que el coeficiente de correlación poblacional es 0,90.

Solución:

$$\mu_z = 1{,}1513 \ \log\left(\frac{1+0{,}90}{1-0{,}90}\right) = 1{,}472 \qquad Z = 1{,}1513 \ \log\left(\frac{1+0{,}86}{1-0{,}86}\right) = 1{,}293$$

$$\sigma_z = \frac{1}{\sqrt{37}} = \frac{1}{6{,}08} = 0{,}1644$$

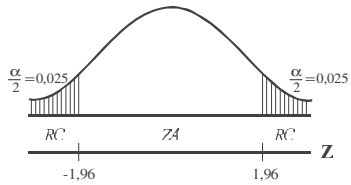
1) $H_0: R = 0{,}90$ $\quad \overline{H_0 : \rho = 0{,}90}$
$H_a: R \neq 0{,}90$ $\quad \overline{H_1 : \rho \neq 0{,}90}$

2) $\propto = 0{,}05$

3) $\sigma_z = 0{,}1644$

4) $Z = \frac{Z - \mu_z}{\sigma_z}$

$$Z = \frac{1{,}293 - 1{,}472}{0{,}1644} = -1{,}09$$



Z = –1,09 cae en la zona de aceptación: se acepta la hipótesis nula, es decir que el coeficiente de correlación 0,86 en una muestra, puede pertenecer a una población cuyo coeficiente de correlación es de 0,90, con un nivel del 5%.

Agreguemos que el *Coeficiente de correlación* es una medida establecida por Karl Pearson, que proporciona la fuerza de asociación entre una variable independiente y otra dependiente.

Transformación de *r* a *z* $\left(es\ decir,\ z = 0{,}5 \quad Log_e \frac{1+r}{1-r}\right)$

Tabla 9.2 Transformaciones de r a z

r	z	r	z	r	z
.00	.000				
.01	.010	.36	.377	.71	.887
.02	.020	.37	.388	.72	.908
.03	.030	.38	.400	.73	.929
.04	.040	.39	.412	.74	.950
.05	.050	.40	.424	.75	.973
...	...	...	...	...	...
.16	.161	.51	.563	.86	1.293
.17	.172	.52	.576	.87	1.333
.18	.182	.53	.590	.88	1.376
.19	.192	.54	.604	.89	1.422
.20	.203	.55	.618	.90	1.472

Ejemplo 3. En una muestra de 42 pares de valores de x y y se obtuvo un coeficiente de correlación de 0,22, ¿Indica esto una correlación significativa con base en un nivel de significación del 5%?

Solución:

1) $H_0 : R = 0$ *(No existe correlación)*
   $H_a : R \neq 0$ *(El coeficiente de correlación es significativo)*

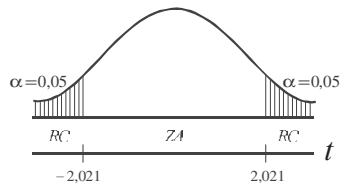
$$\boxed{\begin{array}{l} H_0 : \rho = 0 \\ H_a : \rho \neq 0 \end{array}}$$

2) $\propto = 0{,}05$

3) $\upsilon = n - 2 = 40$; $t_{0{,}05} = 2{,}021$

4) $t = r\sqrt{\dfrac{n-2}{1-r^2}}$

$$t = 0{,}22\sqrt{\frac{40}{1-(0{,}22)^2}} = 1{,}43$$



1,43 < 2,021 por lo tanto aceptamos la hipótesis nula, es decir, que el Coeficiente de correlación 0,22 no es significativo, al nivel del 5%.

Ejemplo 4. Al efectuar un estudio sobre la marca de cierto producto, se encontró que 50 personas habían usado y cambiado anteriormente dicha marca.

La relación entre el tiempo en que habían usado la marca, antes de sustituirla por otra y el número de ex-usuarios en cada caso, fue la siguiente:

AÑOS DE USO:	(*x*) =	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
No. EX-USUARIOS:	(*y*) =	8	6	9	4	6	5	3	2	4	3

Sabiendo que el coeficiente de correlación de este ejercicio es –0,8 (r = –0,8), docime al nivel del 5% que el coeficiente es significativo.

Solución:

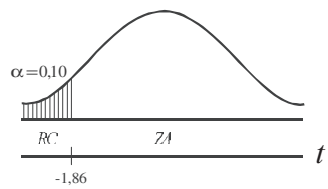
1) $H_0 : R = 0$ *(No existe correlación entre el tiempo de uso y número de usuarios)*
   $H_a : R < 0$ *(Existe correlación negativa entre las variables y es significativa)*

2) $\propto = 0{,}05$ $\quad t_{0{,}10} = -1{,}86$ $\quad \boxed{R = \rho}$

3) $\upsilon = n - 2 = 10 - 2 = 8$

4) $t = r\sqrt{\dfrac{n-2}{1-r^2}}$

$$t = -0{,}8\sqrt{\frac{8}{1-(-0{,}8)^2}} = -3{,}77$$



-3,77 < -1,86. Se rechaza la hipótesis nula. Aceptamos la hipótesis alternativa de que existe correlación negativa entre las variables.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

16. Un coeficiente basado en una nuestra de tamaño 20 resultó ser 0,37. ¿Se puede deducir al nivel de significación del 5%, que el Coeficiente de correlación de la población correspondiente difiere de 0?

17. En un estudio de correlación entre el número de accidentes infantiles en relación con las zonas verdes, se obtuvo un Coeficiente de correlación de −0,92 (r = −0,92), sobre la base de 18 barrios de Cali. Mediante una dócima bilateral y un nivel de significación del 5%, ¿qué se puede concluir respecto al Coeficiente de correlación?

18. Se eligen aleatoriamente once estudiantes entre los que han terminado el curso de matemáticas II y estadística II. Sus calificaciones finales (de 0 a 100) son:

MATEMÁTICAS II:	(x)	67	81	65	42	53	40	71	64	60	68	49
ESTADÍSTICA II :	(y)	92	65	81	75	85	78	77	79	81	82	85

Suponiendo que los pares de calificaciones se extraen aleatoriamente de una población normal bidimensional, docime la hipótesis de que la calificación de matemática II de un estudiante, es independiente de su calificación en estadística II ($\propto = 0{,}05$).

19. Un fabricante de productos alimenticios hace una prueba previa con cierto tipo de salsa picante envasada, que puede preparar en una forma más espesa (X) o en otra forma menos espesa (Y). Para medir la preferencia por uno y otro tipo de salsa, utiliza una muestra de 10 amas de casa, quienes manifiestan sus preferencias por dichos tipos de salsa, con los siguientes resultados:

AMA DE CASA No. :	1	2	3	4	5	6	7	8	9	10
SALSA A :	3	1	5	2	0	4	3	3	2	5
SALSA B :	2	4	4	7	3	4	6	5	5	8

Aplique el test de correlación al nivel del 5%

20. En una muestra de 35 pares, el Coeficiente de correlación obtenido es 0,80. Use el nivel de significación del 5%, para docimar la hipótesis de que $r = 0{,}90$.

21. Un Coeficiente de correlación en una nuestra de tamaño 24, resultó ser $r = 0\,,75$. ¿Puede rechazarse la hipótesis de que el coeficiente de correlación poblacional sea igual a 0,60?

22. En una muestra de 52 pares, el Coeficiente de correlación obtenido es 0,53. Se piden los límites de confianza del 95% para el Coeficiente de correlación.

23. Docimar al nivel del 5%, la hipótesis de que $R = 0{,}60$ ($\rho = 0{,}60$). Si en una muestra de 103 pares se ha obtenido un Coeficiente de correlación igual a 0,90.

24. Se toma una muestra de tamaño 52 en una población cuyo Coeficiente de correlación es 0,59. Encontrar la probabilidad de que el coeficiente de correlación de la muestra sea 0,40 o menos.

25. En una población con Coeficiente de correlación 0,65, se toman muchas muestras de igual tamaño. Si el 15% de estas muestras tienen un coeficiente de correlación de 0,75 o más, ¿cuál es el tamaño de la muestra?

26. Dos Coeficientes de correlación obtenidos de tamaño $n_1 = 28$; $n_2 = 35$ resultaron ser: $r_1 = 0,5$ y $r_2 = 0,30$ respectivamente. ¿Hay una diferencia significativa entre los dos coeficientes al nivel de 0,05?

27. El servicio de control de calidad somete dos muestras de mercancía a una prueba, para determinar si existe una relación entre la resistencia y la duración de utilización del producto. La primera muestra está formada de 15 unidades y la segunda de 27. Se encuentra un coeficiente de correlación de $r_1 = 0,80$ y $r_2 = 0,60$. ¿Podemos afirmar al nivel de significación del 5% que si existe una diferencia significativa?

28. Supongamos dos variables *X* y *Y*. Probar la hipótesis al nivel del 5% de que el Coeficiente de regresión (Coeficiente angular) de la ecuación de regresión sea tan bajo como 0,70.

x_1:	5	8	10	12	20
y_1:	3	2	5	6	14

Nota: este ejercicio es de Prueba de Regresión y no de Correlación (capítulo 12)

PRUEBAS CON OBSERVACIONES APAREADAS

Anteriormente, se realizaron pruebas de diferencias entre dos medias muestrales, es decir, se trabajó con muestras provenientes de dos *poblaciones independientes*, como por ejemplo, cuando se desea probar la efectividad de un fertilizante; el procedimiento que se siguió consistió en tomar una población de lotes o de hectáreas, a las cuales se les ha aplicado un fertilizante, se extraía de ella una muestra de lotes a fin de obtener el promedio de rendimiento por lote o por hectárea. Este resultado es comparado con otro obtenido de una muestra de lotes, provenientes de una población, que no han sido fertilizados. De esta manera se puede probar si con el uso del fertilizante se logra obtener un mayor rendimiento por lote.

El anterior procedimiento conlleva, en algunos casos, a cometer errores en el análisis de los resultados, pues en el ejemplo señalado, puede ser que las variaciones en el rendimiento no sean consecuencias del uso del fertilizante, sino de otros factores ajenos a él, como: condiciones del clima, calidad del terreno, etc. Esto nos obliga a realizar un procedimiento diferente que elimine estos factores ajenos, utilizando para ello las diferencias entre los pares de observaciones, provenientes de poblaciones relacionadas, denominados también *poblaciones dependientes*. Veamos el proceso a seguir en este último caso: se toma una muestra de lotes, y en cada uno de ellos se procede a fertilizar la mitad, comparando su rendimiento con la otra mitad que quedó sin fertilizar; de esta manera eliminamos factores que puedan afectar los resultados, ya que las condiciones serán iguales, tanto para el área fertilizada como para la no fertilizada.

Podríamos dar otro ejemplo, para entender un poco más el nuevo procedimiento que se puede utilizar, dependiendo de la población que se examina. Supongamos que un empresario desea aumentar el rendimiento de su personal, para lo cual diseña cursos de capacitación para mejorar las condiciones del trabajador. La prueba consiste en determinar si esos cursos son efectivos, para ello selecciona una

muestra de trabajadores que no han asistido al curso, luego determina el promedio de rendimiento para el grupo o muestra, ya sea respecto al tiempo de duración en su ejecución o en cuanto a la cantidad de unidades o tareas realizadas en un determinado período. Este resultado, se compara con otro obtenido en una muestra proveniente de una población de trabajadores que han asistido a entrenamiento o capacitación. El anterior procedimiento puede distorsionar los resultados, pues dejamos aparte aspectos o condiciones que presenta cada persona, por ejemplo sentirse bien en el trabajo, tener habilidad innata para desarrollarlo, responsabilidad, condiciones emocionales, etc. Para ello se puede preferir otro procedimiento, que consiste en medir el rendimiento, en cada uno de los trabajadores que no han sido sometidos al curso y que forman parte de la muestra; luego a cada trabajador seleccionado se le compara su rendimiento *antes y después* del entrenamiento, lo cual nos permite eliminar factores que pueden conducir a resultados erróneos.

Anteriormente se comparaban dos muestras elegidas aleatoriamente, es decir, las dos poblaciones de donde se extraían las muestras se consideraban ***independientes***. En cambio en las *observaciones apareadas*, se debe tomar una muestra aleatoria de pares, de manera que cada observación esté asociada con alguna observación en particular. La muestra de pares de observaciones da base para considerar la existencia de dos muestras ***dependientes***, donde las observaciones de un par estarán relacionadas entre sí.

La mayoría de autores consideran de gran importancia su aplicación, tanto en problemas que comprenden datos *antes y después,* como en aquellos casos donde se desea comprobar si se producen cambios de actitud.

Consideremos en primer lugar la simbología que se utilizará: d_i = diferencia para cada par de observaciones, $(d_i = x_i - y_i)$.

$\bar{d}$ = Media aritmética de las diferencias $\bar{d} = \frac{\Sigma d_i}{n}$

n = Tamaño de la muestra. Será igual al número total de pares que conforman la muestra.

N = Número de pares de observación en la población.

a_d ó $a_{\bar{d}}$ = Media de las diferencias en la población. $a_{\bar{d}} = \frac{\Sigma d_i}{N}$ Se considerará que la media de las diferencias en la población es igual a 0. $a_{\bar{d}} = 0$

s_d = Desviación típica en observaciones apareadas. $s_d = \sqrt{\frac{\Sigma(d_i - \bar{d})^2}{n-1}}$

La *variante estadística* estará dada en la siguiente forma: $t = \frac{\bar{d} - a_{\bar{d}}}{\frac{s_d}{\sqrt{n}}} = \frac{\bar{d} - 0}{s_{\bar{d}}}$

$s_{\bar{d}}$ = Error estándar en observaciones apareadas. $s_{\bar{d}} = \frac{s_d}{\sqrt{n}}$

Ejemplo 1. En una escuela pública se seleccionaron 10 pares de niños de primer año para comparar similitud de inteligencia y preparación. Un niño de cada par fue enseñado a leer con un método y el otro niño con otro método. Después del período de aprendizaje, los niños fueron sometidos a una prueba de lectura con los siguientes resultados, (el puntaje utilizado fue de 0 a 100):

NIÑO Nº :	1	2	3	4	5	6	7	8	9	10
MÉTODO I : (x)	65	68	70	63	64	62	74	72	70	66
MÉTODO II : (y)	63	68	68	60	68	66	70	78	70	70

¿A un nivel de significación del 5%, existe alguna diferencia significativa en la mayor efectividad de alguno de los métodos aplicados?

Solución:

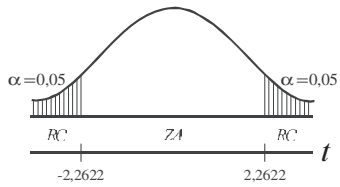
PAREJA	x_i	y_i	$d_i = x_i - y_i$	$d_i - \bar{d}_i$	$(d_i - \bar{d}_i)^2$
1	65	63	2	2,7	7,29
2	68	68	0	0,7	0,49
3	70	68	2	2,7	7,29
4	63	60	3	3,7	13,69
5	64	68	-4	-3,3	10,89
6	62	66	-4	-3,3	10,89
7	74	70	4	4,7	22,09
8	72	78	-6	-5,3	28,09
9	70	70	0	0,7	0,49
10	66	70	-4	-3,3	10,89
Σ	674	681	–7	0	112,10

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{-7}{10} = -0,7$$

$$s_d = \sqrt{\frac{\Sigma(d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{112,1}{9}} = 3,53$$

1) $H_0 : a_d = 0$ 2) $\propto = 0,05$
$H_a : a_d \neq 0$

3) $s_d = 3,53$



4) $t = \dfrac{\bar{d}}{s_{\bar{d}}} = \dfrac{-0,7}{\dfrac{3,53}{\sqrt{10}}} = -0,63$ $\upsilon = n - 1 = 9$ $\propto = 0,05$ $t = 2,2622$

t = – 0,63 se ubica en la zona de aceptación, por lo tanto aceptamos la hipótesis nula ($a_d = 0$) Se puede concluir que ninguno de los métodos es superior al otro, al nivel del 5%.

Ejemplo 2. Supongamos que se quiere estudiar la efectividad de una dieta y se nos proporciona la siguiente información referente a los pesos, antes y después en una muestra al azar de 8 mujeres adultas con edades de 35 a 40 años o más (datos en libras):

MUJERES Nº :	1	2	3	4	5	6	7	8
ANTES : (x)	137	130	124	138	149	140	168	152
DESPUÉS : (y)	132	121	126	130	147	141	159	147

Docimar al nivel del 5% que la dieta fue efectiva.

Solución:

PAREJA	x_i	y_i	$d_i = x_i - y_i$	$d_i - \bar{d}_i$	$(d_i - \bar{d}_i)^2$
1	137	132	–5	– 0,625	0,39
2	130	121	–9	– 4,625	21,39
3	124	126	2	6,375	40,64
4	138	130	–8	–3,625	13,14
5	149	147	–2	2,375	5,64
6	140	141	1	5,375	28,89
7	168	159	–9	– 4,625	21,39
8	152	147	–5	–0,625	0,39
Σ	–	–	–35	0	131,87

Nota: Se hubiera podido trabajar con diferencias positivas y la gráfica sería al contrario, lado derecho. El resultado es el mismo.

$$\bar{d} = \frac{\Sigma d_t}{n} = \frac{-35}{8} = -4{,}375 \qquad s_d = \sqrt{\frac{\Sigma (d_t - \bar{d})^2}{n-1}} = \sqrt{\frac{131{,}87}{7}} = 4{,}34$$

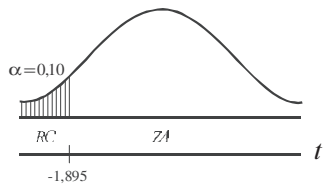
$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{4{,}34}{\sqrt{8}} = \frac{4{,}34}{2{,}83} = 1{,}53$$

1) $H_0 : a_d = 0$ (la dieta no es efectiva)
   $H_a : a_d < 0$ (la dieta es efectiva)

2) $\propto = 0{,}05$ 3) $s_{\bar{d}} = 1{,}53$

4) $t = \dfrac{\bar{d}}{s_{\bar{d}}} = \dfrac{-4{,}375}{1{,}53} = -2{,}86$

$\upsilon = 8 - 1 = 7 \rightarrow t = -1{,}895$



t = – 2,86 se ubica en la zona crítica y se rechaza la hipótesis nula, por lo tanto aceptamos que la dieta ha sido efectiva, al nivel del 5%.

Nota: Se había dicho que $d_i = x_i - y_i$ pero en este caso cuando la efectividad de una dieta corresponde a una disminución del peso, si se quiere se puede cambiar el signo para cada diferencia; el resultado debe ser el mismo, la prueba se hace al lado contrario.

En el cálculo de los *límites de confianza*, se aplicará la siguiente fórmula:

$$a_{\bar{d}_i} = \bar{d} \pm t \frac{s_d}{\sqrt{n}}$$

Ejemplo 3. En un experimento referente a variedades de maíz sobre 16 pares de lotes, la diferencia entre las medias de los rendimientos de las variedades fue de 3 arrobas por hectárea. El error estándar de esta diferencia fue de 1,1 arrobas por hectárea.

a) Determinar los límites de confianza del 95% para la diferencia media de rendimiento entre las dos variedades.

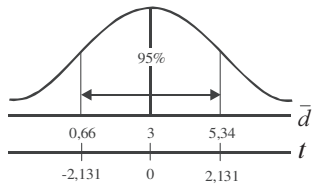
b) Use la respuesta del punto a) para decidir si existe una diferencia significativa entre los rendimientos de las dos variedades ($\propto = 0{,}05$).

Solución:

1) $\bar{d} = 3 \qquad s_{\bar{d}} = 1{,}1 \qquad n = 16 \qquad \upsilon = 15 \qquad t = 2{,}131$

$$a_{\bar{d}_i} = \bar{d} \pm t s_{\bar{d}}$$

$$a_{\bar{d}_i} = 3 \pm 2{,}131\ (1{,}1) \begin{matrix} \nearrow 5{,}34 \\ \searrow 0{,}66 \end{matrix}$$



2) $H_0 : a_d = 0$
$H_a : a_d \neq 0$

3) $\propto = 0{,}05$

Siendo que $a_{\bar{d}} = 0$, no se sitúa dentro de los límites de confianza (0,66 y 5,34), se rechaza la hipótesis nula y aceptamos la alternativa de que existe una diferencia significativa, al nivel del 5%.

Nota: Para aceptar la hipótesis nula se requiere que uno de los límites sea positivo y el otro negativo. En caso de tener los mismos signos, siempre estaremos rechazando la hipótesis nula.

Los anteriores procedimientos corresponden a pruebas de diferencias para pares de observaciones, en que cada elemento de estudio de la primera muestra se encuentra en la segunda, razón por la cual no son muestras independientes. En sí hay una sola muestra, no dos como aparentemente se puede pensar, por lo tanto se tendrá una muestra constituida por la diferencia entre dos observaciones para el mismo elemento. Generalmente es aplicado para aquellas observaciones tomadas *antes* y *después* de cada uno de los elementos en la muestra.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

29. En un experimento realizado en 10 lotes de terreno, la mitad de cada lote se sembró con una semilla resistente al moho y la otra mitad con semilla corriente. Los resultados siguientes (peso del grano en libras), ¿señalan que existe una diferencia significativa entre ambas semillas?

Continuación

LOTE	ESPECIAL	CORRIENTE
1	49	47
2	58	57
3	53	49
4	60	57
5	45	44

LOTE	ESPECIAL	CORRIENTE
6	49	44
7	66	67
8	55	52
9	44	42
10	52	53

30. En un experimento alimenticio apareado, las ganancias de peso (en libras) de cerdos alimentados con dos clases de dietas, son las siguientes:

PAREJA	DIETA A	DIETA B
1	25	19
2	30	32
3	28	21
4	34	34
5	23	19

¿Puede considerarse alguna de las dietas como superior?

31. En un estudio con mellizos idénticos, a uno de ellos se le da una droga y luego se le somete a una prueba de inteligencia bajo los efectos de la droga; el otro mellizo es sometido a la prueba sin darle droga. Si el experimento se efectúa con 6 pares de mellizos y se obtiene una diferencia media de 5 puntos con una desviación estándar de 6,48, construya los límites de confianza del 95%, para la diferencia media en puntajes.

32. Una manzana recién descubierta tiene un sabor delicioso. Se ha decidido docimar (someter a prueba) su rendimiento, plantando este tipo de manzanas junto a otras corrientes, en ocho huertos diseminados en una región apropiada para la producción de ambas variedades. Cuando los árboles empiezan a rendir frutos, se mide su producción en cajas. Los datos obtenidos son los siguientes:

HUERTO	NUEVA MANZANA	MANZANA CORRIENTE
1	13	12
2	14	16
3	19	17
4	10	9
5	15	16
6	14	12
7	12	10
8	11	8

¿Señalan estos resultados una mayor producción para la nueva manzana que para la corriente?

33. Una compañía desea estudiar el efecto que tiene la pausa para el café, sobre la productividad de sus obreros. Seleccionaron 6 obreros, miden su productividad en un día corriente y luego la productividad de los mismos 6 obreros en un día en que se concede la pausa para el café. Las cifras que miden la productividad son las que siguen:

TRABAJADOR	SIN PAUSA	CON PAUSA
1	23	28
2	35	38
3	29	29
4	33	37
5	43	42
6	32	30

¿Indican estos resultados que la pausa para el café aumenta la productividad?

34. Para analizar dos promisorios nuevos híbridos de maíz bajo condiciones normales de cultivo, una compañía productora de semillas seleccionó aleatoriamente 8 fincas en un departamento del país y

plantó ambos híbridos en lotes experimentales en cada finca. El rendimiento en cargas por hectárea, en los 8 terrenos, fue:

FINCA :	1	2	3	4	5	6	7	8
HÍBRIDO A :	86	87	56	93	84	93	73	79
HÍBRIDO B :	80	79	58	91	77	82	74	66

Construya los límites de confianza del 95%, para la diferencia entre el rendimiento medio de ambos híbridos.

35. A diez atletas se les sometió a un programa de entrenamiento físico intensivo por parte del preparador físico. Se anotaron sus pesos (en libras) antes y después de dicho entrenamiento con los resultados siguientes:

ATLETA :	1	2	3	4	5	6	7	8	9	10
PESO ANTES :	127	195	162	170	143	205	168	175	197	136
PESO DESPUÉS :	135	200	160	182	147	200	172	186	194	141

¿Afecta el programa el peso medio de los atletas?

36. Suponga que en un colegio oficial de Guayaquil, se están comparando dos métodos de enseñanza utilizando 50 estudiantes divididos en dos grupos con igual número de alumnos, agrupándolos por parejas basados en su coeficiente de inteligencia, de tal manera que cada estudiante del primer grupo se analiza con un estudiante del segundo grupo. Al finalizar el curso se realiza un examen y se calculan las diferencias entre las calificaciones apareadas. Los resultados son $\bar{d} = 5{,}6$ y $s_d = 9{,}6$ ¿Al nivel del 5%, puede demostrarse que el segundo método es inferior al primero?

37. Un fabricante de productos alimenticios, hace una prueba previa con cierto tipo de salsa picante que puede preparar en una forma más espesa (x) o en otra forma menos espesa (y). Para medir la preferencia por uno y otro tipo de salsa picante, utiliza una muestra de 10 amas de casa, quienes manifiestan sus preferencias por dichos tipos de salsa picante, con los siguientes resultados (se utilizó una escala de 1 a 10) en puntos.

AMA DE CASA No.:	1	2	3	4	5	6	7	8	9	10
SALSA (x) Puntos :	3	1	5	2	0	4	3	3	2	5
SALSA (y) Puntos :	2	4	4	7	3	4	6	5	5	8

Al nivel de significación del 5%, ¿se podría concluir que el tipo de salsa menos espesa (y) tiene mayores oportunidades de ser aceptada en el mercado, que el tipo más espeso (x)?

38. Una agencia de publicidad, desea probar dos formas diferentes de presentación de un anuncio para revistas de una loción para hombres. En el anuncio A el producto es enfocado desde el punto del *sex-appeal*. En el segundo anuncio es enfocado desde el punto de revista; después de un tiempo se seleccionó una muestra de 12 hombres, usuarios potenciales. El procedimiento seguido en el test consiste en dejar que los entrevistados hojeen libremente la revista y luego se sometan a un cuestionario diseñado para medir el grado de recuerdo de los anuncios en cuestión. Los resultados obtenidos en puntos (usando una escala de 1 a 10) son los siguientes:

ENTREVISTADO No. :	1	2	3	4	5	6	7	8	9	10	11	12
ANUNCIO A :	2	3	4	5	1	4	5	7	4	5	3	3
ANUNCIO B :	3	5	7	4	3	2	5	4	6	7	6	6

Al nivel del 5%, ¿podría concluirse que el anuncio B suscita más atención que el A?

39. Una empresa necesita un químico y los aspirantes son sometidos a diferentes pruebas psicotécnicas, de las cuales una es un test de análisis de un producto. Los dos mejores candidatos A y B, han obtenido los resultados siguientes en la prueba del análisis químico:

No. DE LA PRUEBA :	1	2	3	4	5	6	7	8	9	10
CANDIDATO A :	20	17	18	20	19	18	19	20	19	20
CANDIDATO B :	19	18	20	17	18	17	19	19	20	19

¿Puede considerarse que el químico A ha realizado mejores análisis que B o viceversa?

40. La dirección de una empresa de productos dietéticos para niños desea confrontar las propiedades nutritivas de una de sus especialidades con las de un producto de la competencia. Para estos efectos, encarga a un médico nutrir 6 niños con su producto (x) y después, alimentar a estos mismos niños con el producto de la competencia (y). Se ha decidido que el efecto de las propiedades nutritivas de los productos venga medido por el aumento de peso diario de los niños.

No. ORDEN NIÑOS :	1	2	3	4	5	6
PRODUCTO A :	21	20	20	22	16	21
PRODUCTO B :	17	18	18	16	14	13

Al nivel del 1%, ¿existe una diferencia entre el valor nutritivo de los productos?

41. Un ingeniero industrial, cree que ha perfeccionado un programa de entrenamiento que puede acortar el tiempo de ensamble, en forma considerable,para cierto mecanismo. Para verificar esta creencia, planea seleccionar diez obreros en forma aleatoria y realizar estudios de tiempo y movimiento con ellos antes y después de que hayan acabado su programa de entrenamiento (el tiempo se tomó en minutos).

No. DE OBREROS :	1	2	3	4	5	6	7	8	9	10
1o. ESTUDIO :	7	8	10	11	18	16	12	12	6	12
2o. ESTUDIO :	8	8	7	6	10	9	9	8	7	10

a) Obtenga los límites de confianza del 99%.
b) Docime al nivel del 1%, si el programa de entrenamiento resulta de utilidad para reducir el tiempo medio de ensamble.

42. Para comparar la eficiencia de máquinas de escribir corrientes y eléctricas, se eligen en forma aleatoria 8 mecanógrafas y se las instruye en forma completa para que operen ambos tipos de máquinas. Se les pide que escriban en cada tipo de máquina durante diez minutos y se observan sus velocidades, que se miden mediante el promedio de palabras escritas por minuto.

MECANÓGRAFA No. :	1	2	3	4	5	6	7	8
ELÉCTRICA :	75	89	79	85	102	115	97	6 9
CORRIENTE :	79	62	54	67	81	78	66	73

¿Se puede llegar a la conclusión de que los dos tipos de máquinas son diferentes en eficiencia? (Nivel del 5%).

PRUEBAS NO PARAMÉTRICAS

Una buena parte del contenido de este libro, ha sido dedicado al estudio de las *distribuciones paramétricas*, probando o docimando suposiciones acerca de la naturaleza de la población, a través de muestras que nos permitan obtener valores o estimativos, llámese *media*, *proporción*, *varianza*, *coeficiente de correlación*, etc. Dado que se trabaja con valores estadísticos de la población, conocidos como *parámetros*, las técnicas empleadas en estos casos, reciben el nombre de *paramétricas*.

Con las muestras se recoge información, algunas veces difícil de cuantificar, por lo tanto utilizamos como solución, escalas de medición de 0 a 5, de 0 a 10, etc. Este problema se presenta en casi todos los campos del saber; sin embargo con más frecuencia en aquellos estudios de comportamiento o de conducta, como sucede en las ciencias sociales. Observemos, como ejemplo, lo más común que encontramos en las universidades o institutos, la evaluación periódica del profesor; en la mayoría de los casos, su permanencia depende en gran parte de sus alumnos, donde algunas veces se ponen en tela de juicio los conocimientos y otros aspectos de carácter académico y personal, pero las opiniones son, de alguna forma, consecuencia de la subjetividad por parte de los estudiantes.

Los métodos *no paramétricos,* son de gran utilidad en éste y en muchos otros casos para el análisis de los datos, cuando las observaciones se pueden ordenar, ya que son imposibles de medir, también cuando se carece del conocimiento acerca del comportamiento de los parámetros de la población.

Se puede concluir que en todos aquellos casos, donde sea imposible establecer la forma de la *distribución poblacional* o cuando los datos están dados a *escala ordinal*, es decir, que pueden ordenarse por *rangos*, se hace indispensable la aplicación de métodos, denominados de diferentes maneras: *no paramétricos*, de *distribución libre, de rango o de orden*, técnicas que surgieron a mediados del siglo XIX.

Entre los numerosos métodos no paramétricos sobresalen:

- *Prueba del Chi* ó J_i *cuadrado.*
- *Prueba de signo*, donde sólo se toman los signos positivos y negativos de las diferencias en las observaciones apareadas.
- *Método de Spearman o Correlación por Rangos,* cuando los datos o información obtenida mediante muestras, se clasifican de acuerdo al rango obtenido.
- *Prueba T de Wilcoxon*, requiere que los datos estén dados a escala ordinal y que las dos muestras se relacionen por pares o parejas.
- *Prueba de U de Mann - Whitney*, usada para determinar si dos muestras han sido extraídas de una misma población, utilizando la suma de rangos.
- *Prueba H de Kruskal - Wallis*, utilizada para probar si tres o más muestras independientes han sido extraídas de poblaciones con la misma distribución.

Existen más pruebas que no se expondrán en este capítulo, sólo haremos referencia a las ya enunciadas y las desarrollaremos a continuación.

PRUEBA DE CHI-CUADRADO χ^2

Corresponde a las *distribuciones no paramétricas*, siendo una de las más utilizadas, establecida por primera vez por Helmert en 1875 y redescubierta en 1900 por Karl Pearson. En 1934, en el caso de tablas

de contingencia de "2 x 2", fue propuesta la aplicación de la *corrección por continuidad* de Yates, atendiendo la sugerencia hecha por Fisher, pero hoy en día, es considerada por algunos como innecesaria.

La *distribución normal*, se utiliza en todos aquellos casos en que el experimento ofrece dos resultados posibles; cuando se presentan más de dos resultados debe aplicarse la prueba de *Chi-cuadrado*. Un ejemplo típico de distribución normal, lo constituye el lanzamiento de una moneda con posibilidades de que aparezca cara o sello y la de *Chi-cuadrado* estará aplicada, en el lanzamiento de un dado con seis caras posibles, numeradas del 1 al 6.

Esta prueba se denomina J_i - *cuadrado*, derivada de la letra griega mayúscula J_i que se escribe χ y que se lee chi; el cuadrado se debe a que la suma de las diferencias entre los valores observados y esperados será igual a 0, por lo tanto se hace necesario elevarlos al cuadrado, para cuantificar la diferencia.

Por definición, J_i – *cuadrado*, es la suma de las fracciones que tienen por numerador el cuadrado de las diferencias entre las *frecuencias reales u observadas y las frecuencias esperadas o teóricas y por denominador la frecuencia esperada.*

$$\begin{bmatrix} F_o \end{bmatrix} = n_i = \text{Frecuencia observada o real}$$
$$\begin{bmatrix} F_e \end{bmatrix} = n_t^* = \text{Frecuencia teórica o esperada}$$

$$\chi^2 = \sum \frac{(n_t - n_t^*)^2}{n_t^*}$$

Se puede observar en la fórmula anterior, que mientras mayor sea la coincidencia entre las frecuencias observadas y las esperadas, menor será el valor de J_i - *cuadrado*. Si $x^2 = 0$, significa que hay una *completa concordancia* entre las *frecuencias observadas* y *las esperadas*.

La curva que constituye una excelente aproximación a la distribución J_i - *cuadrado*, está dada por la siguiente ecuación:

$$Y = C(\chi^2)^{\frac{\upsilon-2}{2}} e^{-\frac{\chi^2}{2}}$$

Donde v es el número de grados de libertad de la variable y C es una constante que depende de v elegida de tal forma, que el área bajo la curva sea igual a uno.

Esta distribución es semejante a la distribución "t", en cuanto al hecho de que para muestras menores de 30 unidades no es normal, pero a medida que se aumenta el tamaño de la muestra, va ajustándose a la normal, por lo cual, habrá también una distribución para cada tamaño de muestra.

Esta prueba es realizada, en todos aquellos casos en los cuales el análisis de los datos está basado en el conteo, es decir, provienen de una característica cualitativa (*atributo*), por no ser medidas tal como ocurre con la variable.

Para la aplicación de la prueba de *Chi-cuadrado*, se debe proceder así: primero: se formula la hipótesis, en seguida establecemos las diferencias entre las frecuencias observadas y las esperadas, se eleva cada diferencia al cuadrado y se divide a cada una de ellas por su frecuencia teórica o esperada; finalmente se suma y de esta manera nos resulta el valor de χ^2, denominada por muchos como J_i *calculada*: siendo:

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

$$\chi^2 = \sum \frac{(n_t - n_t^*)^2}{n_t^*}$$

Luego de obtenido el resultado de x^2 podemos especificar un criterio de decisión para juzgar si las frecuencias observadas, en promedio, difieren significativamente de las frecuencias esperadas. Al observar la fórmula anterior, como ya se dijo, que en aquellos casos en que las frecuencias observadas sean iguales a las frecuencias esperadas, la diferencia debe ser igual a cero, por tal razón la suma de estas diferencias será cero y el valor de *Chi-cuadrado* también debe ser cero, lo cual significa que hay una completa concordancia. De la misma manera, a medida que aumentan las diferencias, también se irá aumentando el valor de *Chi-cuadrado*, como indicador de que hay discrepancias entre lo observado y lo esperado.

Ejemplo 1. Supongamos que se lanza un dado 60 veces. Se sabe que las frecuencias teóricas, para este caso, son de 10 veces cada cara y las frecuencias reales son los resultados del lanzamiento. ¿Con base en un nivel de significación del 5%, permite suponer que el dado no es perfecto?

CARAS	n_t	n_t^*	$n_t - n_t^*$	$(n_t - n_t^*)^2$	$\frac{(n_t - n_t^*)^2}{n_t^*}$
1	7	10	– 3	9	0,9
2	14	10	4	16	1,6
3	8	10	– 2	4	0,4
4	5	10	5	25	2,5
5	16	10	6	36	3,6
6	10	10	0	0	0
Σ	60	60	0	-	9,0
—	F_o	F_e	$F_o - F_e$	$(F_o - F_e)^2$	$\frac{(F_o - F_e)^2}{F_e}$

Solución:

El problema nos da las frecuencias reales u observadas (n_i) para cada cara, tal como aparece en el cuadro anterior. Las frecuencias esperadas se obtienen multiplicando la probabilidad de presentación para cada suceso por n (número de lanzamientos).

$n_i^* = np = 60\left(\frac{1}{6}\right) = \frac{60}{6} = 10$ y así sucesivamente se obtienen las demás frecuencias teóricas o esperadas.

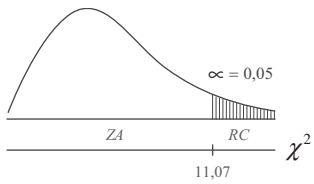
Se realiza la prueba teniendo en cuenta los siguientes pasos:

1) $H_0 : n_t = n_t^*$ $\quad$ $H_0 : F_o = F_e$

$H_a : n_t \neq n_t^*$ $\quad$ $H_1 : F_o \neq F_e$

2) $\propto = 0{,}05$

3) $\chi^2 = \sum \frac{(n_t - n_t^*)^2}{n_t^*} = 9{,}0$



4) $\upsilon = n - 1 \rightarrow \chi^2_{0,05} = 11{,}07$ (Ver tabla 9.3)

$\chi^2 < \chi^2_{0,05}$ es decir $9 < 11{,}07$

por lo tanto se acepta la hipótesis de que la diferencia no es significativa. En otras palabras, podemos afirmar al nivel del 5%, que las diferencias que presentan las frecuencias reales, con relación a las frecuencias teóricas no nos dan base para decir que el dado está cargado.

Al utilizar J_i-*cuadrado* para una dócima, el rechazo de la hipótesis nula o sea H_0, se presenta cuando el valor de J_i-*cuadrado* calculado mediante la realización del experimento, es igual o mayor que el valor J_i-*cuadrado* obtenido en la tabla, para ciertos grados de libertad y un nivel de significación fijado de antemano.

Tabla 9.3 Distribución de *Ji-cuadrado*

υ	A = 0,30	A = 0,20	A = 0,10	A = 0,05	A = 0,02	A = 0,01	A = 0,001
1	1,07	1,64	2,71	3,84	5,41	6,64	10,83
2	2,41	3,22	4,60	5,99	7,82	9,21	13,82
3	3,66	4,64	6,25	7,82	9,84	11,34	16,27
4	4,88	5,99	7,78	9,49	11,67	13,28	18,46
5	6,06	7,29	9,24	11,07	13,39	15,09	20,52

Ejemplo 2. De 64 crías en un cruzamiento entre conejos, 34 son rojos, 10 negros y 20 blancos. Conforme al modelo genético, estos números deberían estar en la relación 9:3: 4. ¿Al nivel del 5%, se podría afirmar que los datos son consistentes con el modelo?

Solución:

$9 + 3 + 4 = 16 \qquad p_1 = \frac{9}{16} = 0{,}56 \qquad p_2 = \frac{3}{16} = 0{,}19 \qquad p_3 = \frac{4}{16} = 0{,}25$

$n_t^* = np_1 = 64\,(0{,}56) = 36 \qquad n_2 = np_2 = 64\,(0{,}19) = 12 \qquad n_3 = np_3 = 64\,(0{,}25) = 16$

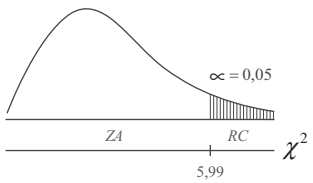
Siendo: $\upsilon = 3 - 1 = 2 \qquad \chi^2_{0{,}05} = 5{,}99$ *(J_i de tabla ó j_i teórico)*

n_t	n_t^*	$n_t - n_t^*$	$(n_t - n_t^*)^2$	$\frac{(n_t - n_t^*)^2}{n_t^*}$
34	36	–2	4	0,1111
10	12	–2	4	0,3333
20	16	4	16	1,0000
64	64	0	–	1,4444
F_o	F_e	$F_o - F_e$	$(F_o - F_e)^2$	$\frac{(F_o - F_e)^2}{F_e}$

1) $H_0 : n_i = n_i^* \qquad H_0 : F_o = F_e$

$H_a : n_i \neq n_i^* \qquad H_a : F_o \neq F_e$

2) $\propto = 0{,}05$



3) $\chi^2 = \sum \frac{(n_t - n_t^*)^2}{n_t^*} = 1{,}4444$

Como $\chi^2 = 1{,}4444$ se sitúa en la zona de aceptación, se puede afirmar al nivel del 5%, que los datos son consistentes con el modelo.

TABLAS DE CONTINGENCIA

Las dócimas o pruebas aplicadas a *cuadros de contingencia,* algunos las denominan también como *dócimas de independencia*. Sin embargo, permiten la realización de *pruebas de homogeneidad*. Un *cuadro de contingencia,* es un arreglo en el cual un conjunto de observaciones se dispone conforme a dos criterios de clasificación, uno de los cuales se expresa en columnas y el otro en renglones. Si cada uno de los criterios admite dos clasificaciones, se obtiene una *Tabla de contingencia de* "2 x 2". Si el primer criterio admite tres clasificaciones y el segundo criterio 4, se denominará como una Tabla de "3 x 4". Si designamos las columnas por K y los renglones por J, se tendrá una Tabla de "K x J".

Los grados de libertad serán iguales a $v = (K - 1)(J - 1)$, así que en una tabla de "2 x 2", el valor de v será igual a 1, o sea $v = (2 - 1)(2 - 1)$; en la tabla de 3x4 será: $v = (3 - 1)(4 - 1) = 6$.

Cuando el número de grados de libertad obtenidos es igual a 1 y la variable es discreta, tal como ocurre en la prueba de *Ji-cuadrado,* hay necesidad de introducir en los cálculos la llamada *Corrección de continuidad de Yates*, que consiste en restar el valor de 0,5 de valor absoluto de las diferencias entre las frecuencias observadas y las esperadas, antes de ser elevadas al cuadrado. La fórmula, aplicando la corrección de Continuidad de Yates será:

$$\chi^2 = \sum \frac{\left(\left|n_t - n_t^*\right| - 0{,}5\right)^2}{n_t^*} \qquad \chi^2 = \sum \frac{\left(\left|F_o - F_e\right| - 0{,}5\right)^2}{F_e}$$

Puede observarse que en muestras grandes, se obtienen prácticamente los mismos resultados que al aplicar la χ^2 sin corregir.

A pesar de que durante mucho tiempo se ha venido utilizando la corrección por continuidad de Yates, hoy su uso no es relevante pues se ha demostrado que, en la mayoría de los casos, la hipótesis nula (H_0) no se rechaza con la frecuencia requerida. De ahí que su aplicación sea limitada.

Se dice que la corrección debe hacerse, cuando los grados de libertad sean iguales a uno (1), es decir, se trata de una tabla de "2x2"; también se recomienda en aquellos casos en que una de las frecuencias observadas sea menor que 10.

Ejemplo 3. En un consultorio se trató a un grupo de personas que se quejaban de insomnio, dándole a unos pastillas para dormir y a otros pastillas de azúcar (que hacían creer que eran para dormir). Después de someterlos a observación, se obtuvo el siguiente resultado.

TRATAMIENTO	DURMIERON	NO DURMIERON	TOTAL
Pastillas para dormir	35	5	40
Pastillas de azúcar	45	15	60
Total	80	20	100

Pruebe al nivel del 5% que no existe diferencia

Solución:

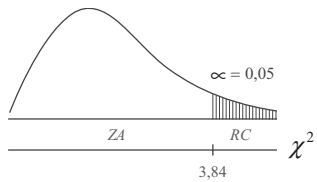
n_i	n_i^*	$n_i - n_i^*$	$\lvert n_i - n_i^* \rvert - 0{,}5$	$(\lvert n_i - n_i^* \rvert - 0{,}5)^2$	$\frac{(\lvert n_i - n_i^* \rvert - 0{,}5)^2}{n_i^*}$
35	32	3	2,5	6,25	0,1953
45	48	–3	2,5	6,25	0,1302
5	8	–3	2,5	6,25	0,7812
15	12	3	2,5	6,25	0,5208
100	100	0	–	–	1,6275
F_o	F_e	$F_o - F_e$	$\lvert F_o - F_e \rvert - 0{,}5$	$(\lvert F_o - F_e \rvert - 0{,}5)^2$	$\frac{(\lvert F_o - F_e \rvert - 0{,}5)^2}{F_i}$

$$n_1^* = \left(\frac{40}{100}\right)80 = 32 \quad n_2^* = \left(\frac{60}{100}\right)80 = 48 \quad n_3^* = \left(\frac{40}{100}\right)20 = 8 \quad n_4^* = \left(\frac{60}{100}\right)20 = 12$$

a) $H_0 : n_i = n_i^*$

$H_\alpha : n_i \neq n_i^*$

b) $\propto = 0{,}05$



c) $\chi^2 = \sum \frac{\left(\lvert n_i - n_i^* \rvert - 0{,}5\right)^2}{n_i^*} = 1{,}6275$

siendo $\upsilon = (2-1)\,(2-1) = 1$

$\chi^2_{0{,}05} = 3{,}84$

d) Se acepta la hipótesis nula $H_0 : n_i = n_i^*$, la diferencia no es significativa, al nivel del 5%.

Ejemplo 4. Dos grupos A y B cada uno de 100 individuos, padecen una enfermedad. Se administra un suero solo al grupo A; en todo lo demás, los dos grupos fueron tratados idénticamente. Se encuentra que en los grupos A y B, 75 y 65 individuos, respectivamente, se han recuperado de la enfermedad.

Probar la hipótesis de que el suero ayuda a curar la enfermedad. ($\propto = 0{,}05$).

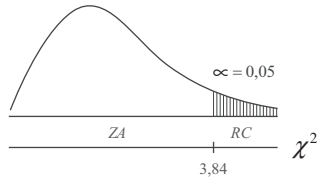
TRATAMIENTO	DURMIERON	NO DURMIERON	TOTAL
Grupo A (con suero)	75	25	100
Grupo B (sin suero)	65	35	100
Total	140	60	200

Solución:

n_i	n_i^*	$n_i - n_i^*$	$\lvert n_i - n_i^* \rvert - 0{,}5$	$(\lvert n_i - n_i^* \rvert - 0{,}5)^2$	$\frac{(\lvert n_i - n_i^* \rvert - 0{,}5)^2}{n_i^*}$
75	70	5	4,5	20,25	0,2893
65	70	–5	4,5	20,25	0,2893
25	30	–5	4,5	20,25	0,6750
35	30	5	4,5	20,25	0,6750
200	200	0	–	–	1,9286
F_o	F_e	$F_o - F_e$	$\lvert F_o - F_e \rvert - 0{,}5$	$(\lvert F_o - F_e \rvert - 0{,}5)^2$	$\frac{(\lvert F_o - F_e \rvert - 0{,}5)^2}{F_i}$

$$n_1 = \left(\frac{100}{200}\right)140 = 70 \quad n_2^* = \left(\frac{100}{200}\right)140 = 70 \quad n_3^* = \left(\frac{100}{200}\right)60 = 30 \quad n_4^* = \left(\frac{100}{200}\right)60 = 30$$

1) H_0: La recuperación es independiente del empleo del suero
   H_a: La recuperación y el empleo del suero están relacionados
2) $\propto = 0{,}05$
3) $\chi^2 = \sum \dfrac{\left(\left|n_t - n_t^*\right| - 0{,}5\right)^2}{n_t^*} = 1{,}93$



4) $\upsilon = (2-1)(2-1) = 1 \qquad \chi^2_{0{,}05} = 3{,}84$

Aceptamos la hipótesis nula H_0, y se deduce que la recuperación es independiente del empleo del suero, al nivel del 5%.

Nota: con la corrección de Yates siempre se disminuye el valor de χ^2. Observemos el resultado χ^2 cuando se hace sin corrección.

n_i	n_t^*	$n_t - n_t^*$	$(n_t - n_t^*)^2$	$\dfrac{(n_t - n_t^*)^2}{n_t^*}$
76	70	5	25	0,3571
65	70	–5	25	0,3571
25	30	–5	25	0,8333
35	30	5	25	0,8333
200	200	0	–	2,3808
F_o	F_e	$F_o - F_e$	$(F_o - F_e)^2$	$\dfrac{(F_o - F_e)^2}{F_e}$

χ^2 (*Corregida* = 1,93)

χ^2 (*Sin corrección* = 2,38)

Se podrá observar que $\chi^2 < \chi^2_{0{,}05}$, es decir, 2,38 < 3,84 de tal manera que aceptamos que la recuperación es independiente del empleo del suero.

Un procedimiento más rápido para realizar la prueba de hipótesis en la distribución de *Ji-cuadrado,* es como sigue:

$$\chi^2 = \frac{n(AD - BC)^2}{m_1\ m_2\ m_3\ m_4}$$

Cuando sea necesario aplicar la corrección de Yates, tal como lo habíamos recomendado, para casos en que $\upsilon = 1$ o cuando $n < 50$ y en aquellos casos en que alguna o todas las frecuencias sean menores que 5, se procederá así:

$$\chi^2 = \frac{n\,(\left|AD - BC\right| - 0{,}5n)^2}{m_1\ m_2\ m_3\ m_4}$$

Observemos las letras ó símbolos anteriores, ¿a qué valores corresponden en la Tabla de contingencia?

ATRIBUTOS	CLASIFICACIÓN		TOTALES
	1	2	
1	A	B	m_3
2	C	D	m_4
-	m_1	m_2	n

Ejemplo 5. Utilicemos las fórmulas del ejercicio anterior sobre la aplicación del suero y la recuperación.

A =75 B=25 C =65 D =35 m_1 =140 m_2 =60 m_3 =100 m_4 =100 n =200

Solución:

$AD = 75\ (35) = 2.625$

$BC = 25\ (65) = 1.625$

$m_1\ m_2\ m_3\ m_4 = 140\ (60)\ (100)\ (100) = 84.000.000$

$$\chi^2 = \frac{n(AD - BC)^2}{m_1\ m_2\ m_3\ m_4} \qquad \chi^2 = \frac{200\ (2.625 - 1.625)^2}{84.000.000} = 2{,}38 \quad \text{(sin corrección)}$$

Aplicamos ahora la fórmula para calcular J_i- *cuadrado* corregida:

$$\chi^2 = \frac{n(|AD - BC| - 0{,}5n)^2}{m_1\ m_2\ m_3\ m_4}\ ; \quad \chi^2 = \frac{200\ (|2.625 - 1.625| - 100)^2}{84.000.000} = 1{,}9286 \quad \text{(corregida)}$$

Notaremos que los resultados son exactamente iguales a los obtenidos al aplicar la fórmula general de J_i- *cuadrado*, con y sin corrección.

DÓCIMAS DE HOMOGENEIDAD Y DE INDEPENDENCIA

La prueba de ***homogeneidad*** se utiliza para determinar, si dos o más muestras aleatorias provienen de la misma población o de poblaciones diferentes. Cuando se dice que las clasificaciones son homogéneas, se quiere dar a entender que tienen algo en común. En cambio, la ***prueba de independencia*** es aplicada para establecer si hay alguna relación en cuanto a los criterios de clasificación de la información.

La diferencia existente entre la dócima de ***independencia*** y la de *homogeneidad,* se presenta cuando en la primera interesa saber si un atributo o clasificación es independiente del otro, y generalmente corresponde a resultados obtenidos a través de una muestra de una población; en la segunda prueba, interesa establecer si las muestras provienen de la misma población, por lo tanto estará conformada con dos o más muestras independientes, de cada una de las poblaciones sometidas a estudio.

Para mayor claridad, se podrá decir que en la *prueba de independencia,* se extrae una sola muestra de una población y sólo nos interesa probar si existe alguna relación entre los criterios de clasificación establecidos. Algunos, erróneamente, creen que se trata de establecer la relación o correlación que puede haber entre las variables; para este caso se aplican las ***técnicas de regresión*** y las *pruebas de correlación*.

En la ***prueba de homogeneidad***, al contrario de la de ***independencia***, se extraen dos o más muestras provenientes de dos o más poblaciones; algunas veces, de una muestra se pueden obtener dos o más categorías y el interés es el de probar si las poblaciones tienen cierta identidad con respecto a la característica analizada.

En resumen, el cálculo de *Ji-cuadrado* es igual para las dos pruebas y se diferencia en cuanto a la manera como se recolectan los datos y se interpretan los resultados.

Ejemplo 6. Una tienda vende lavadoras y aspiradoras eléctricas y tiene 3 vendedores. Las ventas realizadas en un mes por dichos vendedores, son las siguientes:

ATRIBUTOS	CLASIFICACIÓN			TOTALES
	1	2	3	
Lavadoras	20	8	15	43
Aspiradoras	17	16	5	38
Totales	**37**	**24**	**20**	**81**

¿Demuestra este resultado que la habilidad de cada vendedor depende del tipo de artículo que vende? Pruebe la hipótesis de homogeneidad al nivel del 5%.

Solución:

n_t	n_t^*	$n_t - n_t^*$	$(n_t - n_t^*)^2$	$\frac{(n_t - n_t^*)^2}{n_t^*}$
20	19,6	0,4	0,16	0,0082
17	17,4	– 0,4	0,16	0,0092
8	12,7	– 4,7	22,09	1,7393
16	11,3	4,7	22,09	1,9549
15	10,6	4,4	19,36	1,8264
5	9,4	– 4,4	19,36	2,0596
81	81,0	0	–	7,5976
F_o	F_e	$F_o - F_e$	$(F_o - F_e)^2$	$\frac{(F_o - F_e)^2}{F_e}$

$$n_1^* = \left(\frac{43}{81}\right)37 = 19{,}6 \qquad n_2^* = \left(\frac{38}{81}\right)37 = 17{,}4 \qquad n_3^* = \left(\frac{43}{81}\right)24 = 12{,}7$$

$$n_4^* = \left(\frac{38}{81}\right)24 = 11{,}3 \qquad n_5^* = \left(\frac{43}{81}\right)20 = 10{,}6 \qquad n_6^* = \left(\frac{38}{81}\right)20 = 9{,}4$$

1) H_0: Las muestras se extraen de la misma población.
H_a: Las muestras se extraen de poblaciones diferentes.

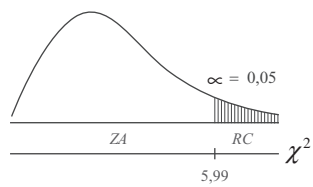
2) $\propto = 0{,}05$

$$\chi^2 = \sum \frac{(F_o - F_e)^2}{F_e}$$

3) $$\chi^2 = \sum \frac{(n_t - n_t^*)^2}{n_t^*} = 7{,}60$$

4) $\upsilon = (2-1)\,(3-1) = 2$

$\chi^2_{0{,}05} = 5{,}99$



$\chi^2_{0{,}05} = 5{,}99$, siendo 7,60 mayor a 5,99 se rechaza la hipótesis nula (H_0); las muestras se extraen de poblaciones diferentes. También podemos contestar que la habilidad de cada vendedor no depende del artículo, al nivel del 5%.

APLICACIONES DE EXCEL EN PRUEBAS DE CHI-CUADRADO

Consideremos otros dos (2) ejemplos, que serían resueltos mediante la aplicación de EXCEL.

n_i =F_o = Frecuencia observada o real
n_i^*=F_e = Frecuencia esperada o teórica

Fórmula: $\chi^2 = \sum \frac{(n_t - n_t^*)^2}{n_t^*} = 7{,}60$ $\quad \chi^2 = \frac{(F_o - F_e)^2}{F_e}$

Primero: Lanzamiento de un dado 360 veces, que equivale a lanzar 360 dados una sola vez, de esta manera se observarán los resultados obtenidos (n_i), los esperados (n_i^*), y además, el porcentaje o la proporción de presentación de cada cara (p).

CARAS	n_t	n_t^*	p
1	56	60	0,1666
2	58	60	0,1666
3	62	60	0,1666
4	64	60	0,1666
5	61	60	0,1666
6	59	60	0,1666
Σ	360	360	1,000

$n_t^* = np$
Siendo: $n = 360$ y p =1/6

n_t^* = 360(1/6) = 360/6 = 60

Se espera que cada cara se presente 60 veces en los 360 lanzamientos.

Segundo: En *Tablas de contingencia*, por ejemplo: una tabla de 3 x 3, con la siguiente información:

VARIABLE LÍNEA	VARIABLE COLUMNA			TOTAL
	I.	M	N	
A	400	200	100	700
B	300	350	150	800
C	300	450	250	1.000
TOTAL	1.000	1.000	500	2.500

p_1 = 700/2.500 = 0,28
p_2 = 700/2.500 = 0,32
p_3 = 700/2.500 = 0,40

Para la aplicación de Excel trabajaremos con la segunda tabla de 3x3:

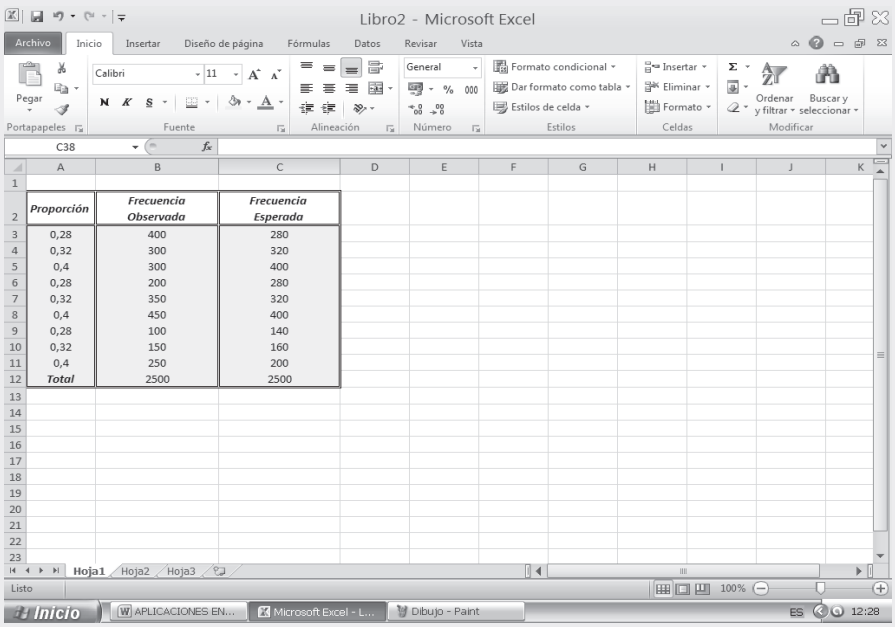
- Lo primero que hacemos es teclear en una hoja electrónica la información anterior:

Los pasos a seguir para el cálculo de CHI-CUADRADO (x^2) son:

- Seleccionamos la celda C14 en la hoja de cálculo que contiene los datos de la frecuencia esperada.
- Seleccionamos la opción función *fx,* con la cual aparece un cuadro de diálogo correspondiente a INSERTAR FUNCIÓN.

En la SEGUNDA PARTE DEL SIL que acompaña al libro, encontrará en conjunto toda la aplicación del EXCEL a la ESTADÍSTICA, que igual a1 capítulo 13, ofrece una BASE DE DATOS para trabajar.

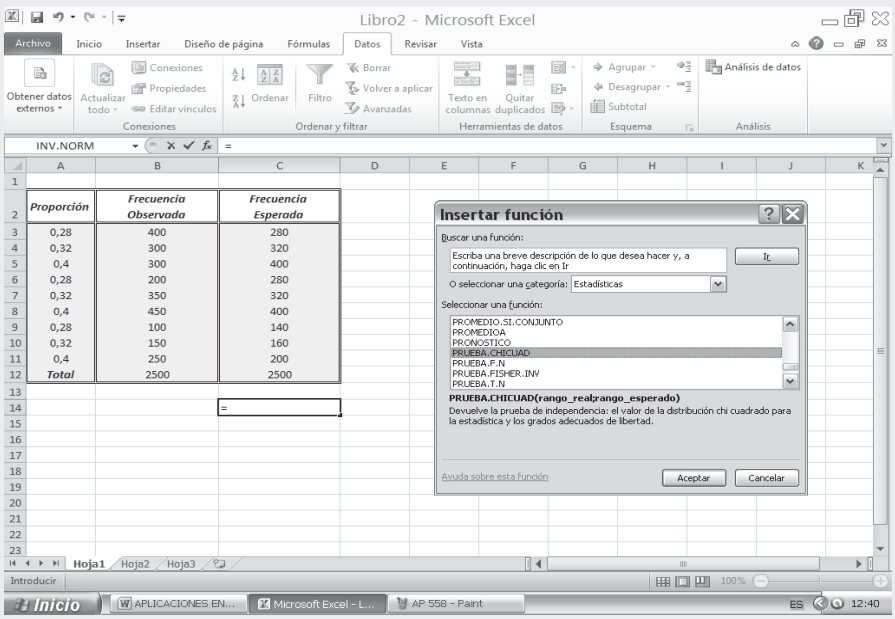
Figura No. 1.Información en Hoja de Cálculo.



Proporción	Frecuencia Observada	Frecuencia Esperada
0,28	400	280
0,32	300	320
0,4	300	400
0,28	200	280
0,32	350	320
0,4	450	400
0,28	100	140
0,32	150	160
0,4	250	200
Total	2500	2500

- Observemos dentro del cuadro de diálogo, el campo CATEGORÍA DE LA FUNCIÓN, en el cual seleccionamos la opción ESTADÍSTICAS y en el espacio correspondiente a SELECCIONAR FUNCIÓN, señalamos la función PRUEBA CHICUAD; finalmente hacemos CLIC en ACEPTAR.

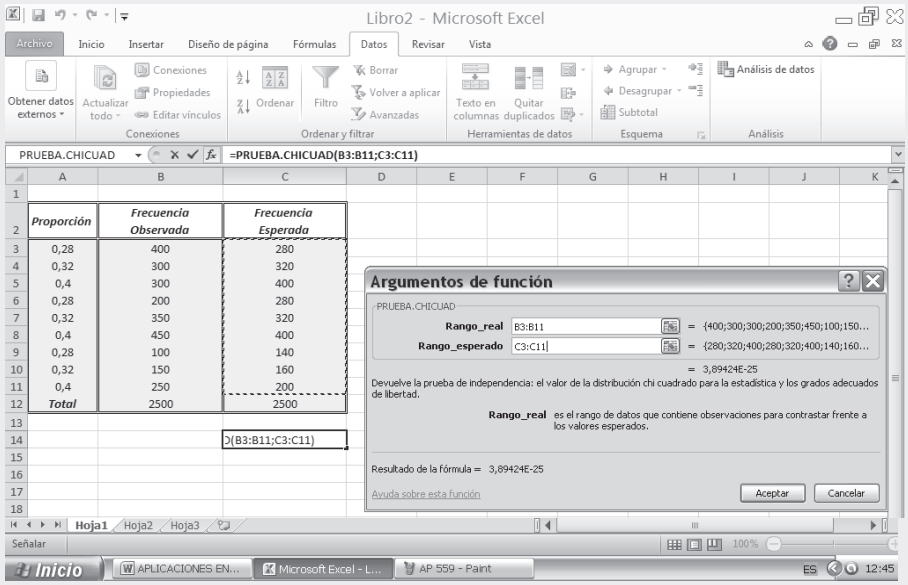
Figura No. 2. Insertar Función.



Proporción	Frecuencia Observada	Frecuencia Esperada
0,28	400	280
0,32	300	320
0,4	300	400
0,28	200	280
0,32	350	320
0,4	450	400
0,28	100	140
0,32	150	160
0,4	250	200
Total	2500	2500

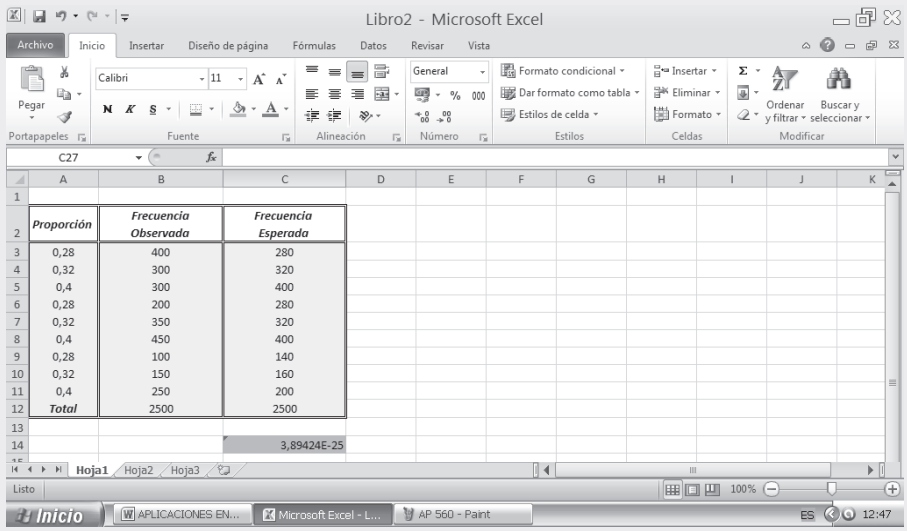
- Nos aparece en pantalla un CUADRO DE DIÁLOGO correspondiente a la función PRUEBA CHI CUADRADO, que nos solicita información sobre los RANGOS:
- RANGO ACTUAL, debemos teclear en la casilla respectiva, B3:B11, correspondiente a la FRECUENCIAS OBSERVADAS.
- RANGO ESPERADO, se teclea C3:C11, que constituye la FRECUENCIA ESPERADA.
- Y, finalmente hacemos CLIC en ACEPTAR. Este proceso se ve en las dos figuras que se presentan a continuación:

Figura No. 3. Argumentos de la Función.



Proporción	Frecuencia Observada	Frecuencia Esperada
0,28	400	280
0,32	300	320
0,4	300	400
0,28	200	280
0,32	350	320
0,4	450	400
0,28	100	140
0,32	150	160
0,4	250	200
Total	2500	2500

Figura No. 4. Resultados de la Prueba de Chi- Cuadrado.

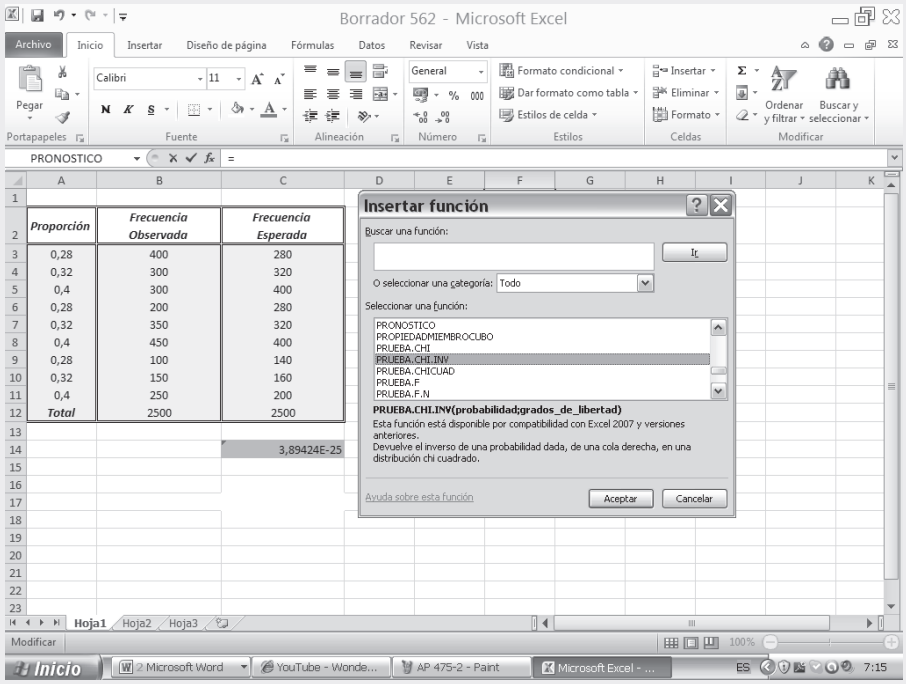


Proporción	Frecuencia Observada	Frecuencia Esperada
0,28	400	280
0,32	300	320
0,4	300	400
0,28	200	280
0,32	350	320
0,4	450	400
0,28	100	140
0,32	150	160
0,4	250	200
Total	2500	2500

Ahora vamos a determinar el inverso de una probabilidad dada. Esto es, PRUEBA.CHI.INV, para ello:

- Procedemos a seleccionar la CELDA C14 en la hoja de cálculo, cuya casilla presentará el valor de χ^2 calculado.
- Hacemos CLIC en la función *fx*. Aparece un cuadro de diálogo correspondiente a INSERTAR FUNCIÓN.
- En el cuadro de diálogo, CATEGORÍA DE LA FUNCIÓN, seleccionamos PRUEBA.CHI.INV y hacemos CLIC en ACEPTAR.

Figura No.5. Insertar Función.



- En la pantalla aparece un nuevo cuadro de diálogo correspondiente a PRUEBA.CHI.INV. En este cuadro tecleamos en el primer campo denominado PROBABILIDAD nuestro valor de CHI. CUAD del ejercicio anterior, es decir el valor de la celda C14.
- Finalmente establecemos los GRADOS DE LIBERTAD, υ = (J-1)(k-1) siendo:

  υ = (3-1) (3-1) =4 , razón por la cual tecleamos en la casilla GRADOS DE LIBERTAD, el número 4 y hacemos CLIC en ACEPTAR.

Figura No.6. Argumentos de la Función.

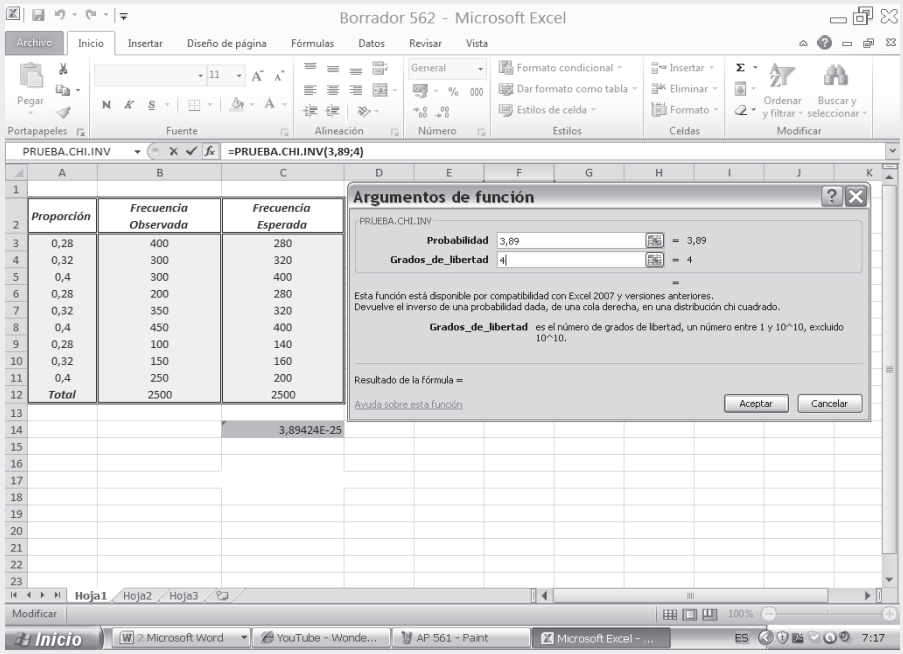
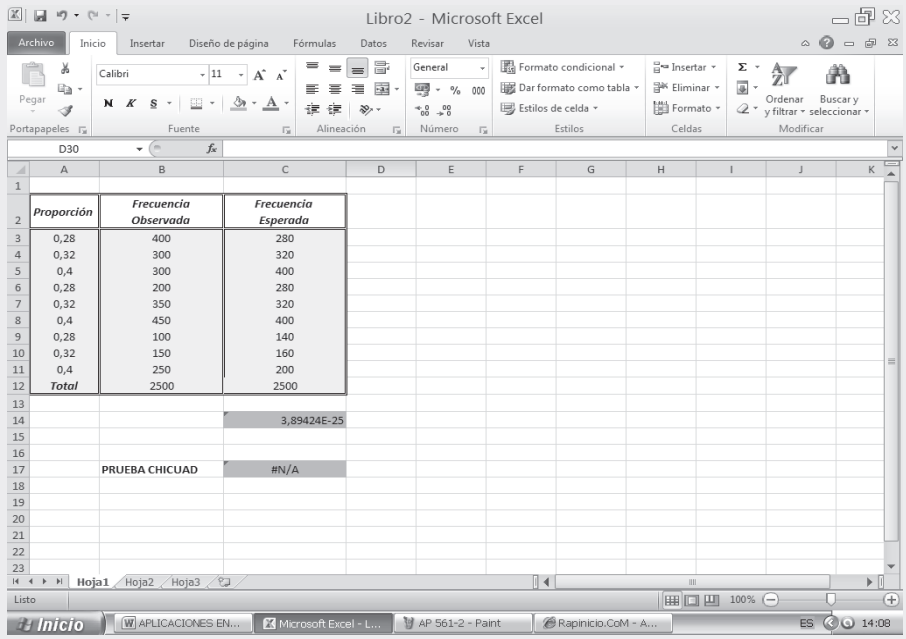


Figura No.6. Resultado de la Prueba de Chi-Inversa.



El resultado anterior, corresponde a una prueba exacta.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

43. Un dado se lanza 100 veces y se obtienen los siguientes resultados:

CARA:	1	2	3	4	5	6
n_i:	12	17	20	22	13	16

Determine si el dado puede considerarse perfecto, al nivel del 1% de significación.

44. En un cruzamiento entre flores roja, rosada y marfil, se obtuvieron los siguientes valores en la generación F_2.

FENOTIPO	NÚMERO DE PLANTAS
Roja	20
Rosada	55
Marfil	25

Basado en estos datos, ¿puede suponerse que la segregación se ha presentado en la relación Mendeliana simple de 1: 2: 1?

45. En un experimento relativo al cruce de dos híbridos de especies de flores, se obtuvo el resultado tal como se indica a continuación ¿Son estos resultados consistentes con la proporción esperada 9: 3: 3: 1?

FLOR MAGENTA ESTIGMA VERDE	FLOR MAGENTA ESTIGMA ROJO	FLOR ROJA ESTIGMA VERDE	FLOR ROJA ESTIGMA ROJO
120	49	36	12

46. En cierto cruzamiento de dos variedades de guisantes, la teoría genética lleva a esperar que la mitad de las semillas sean corrugadas y la mitad lisas. En una muestra de 800 semillas se encuentran 440 corrugadas.

a) Calcúlese el valor de χ^2

b) Si en una muestra de 400 semillas se cuentan 220 corrugadas, ¿será χ^2 la mitad del valor anterior?

c) Mediante un χ^2 no ajustado, calcule la probabilidad de obtener, en un total de 400 semillas, un número de semillas corrugadas que difieran en 20 o más del valor esperado conforme a la teoría (es decir, 200 ó más y 180 ó menos). Resuélvalo mediante la Distribución normal.

47. De 64 crías de un cruzamiento entre conejillos de Indias, 34 son rojos, 10 negros y 20 blancos. Conforme al modelo genético, estos números deberían estar en la relación: 9: 3: 4. ¿Los datos son consistentes con el modelo?

48. Durante una epidemia de fiebre amarilla se obtuvieron los siguientes datos, sobre la efectividad de una vacuna como medida preventiva.

TRATAMIENTO	ENFERMOS	NO ENFERMOS
Vacunados	192	4
No vacunados	113	34

Estos datos ¿indican la efectividad de la vacunación con base en el nivel de significación del 1%?

49. Se clasifican 50 individuos conforme al color de los ojos y del pelo. Con los datos mencionados ¿puede concluirse que en estos individuos existe una relación significativa entre ambas propiedades?

COLOR	PELO CLARO	PELO OSCURO
Ojos azules	23	7
Ojos castaños	4	16

50. Un grupo de 100 hombres y 100 mujeres son entrevistados respecto a su interés en cierto programa radial. 65 hombres no escuchan el programa y 35 sí, 80 mujeres no lo oyen y 20 son auditoras. ¿Existe una diferencia significativa entre los hábitos de este grupo de hombres y mujeres respecto al programa? ($\propto = 1\%$)

51. Los siguientes datos muestran e1 efecto de cierto tipo de fumigación sobre el deterioro de la fruta.

PROCESO	DETERIORADA	SANA	TOTAL
Sin fumigar	8	16	24
Fumigada	2	14	16
Total	**10**	**30**	**40**

¿Depende la cantidad de fruta deteriorada de su fumigación? ($\propto = 2\ \%$)

52. Un grupo de 30 hombres y 80 mujeres expresan su preferencia por el aroma de dos tabacos para pipa. Los resultados son los siguientes:

SEXO	MARCA A	MARCA B	TOTAL
Hombres	9	21	30
Mujeres	8	72	80
Total	**17**	**93**	**110**

¿Existe alguna diferencia significativa en cuanto a las preferencias por el aroma?
(Nivel de significación 5%).

53. En tres grupos de personas, elegidas en diferentes regiones geográficas, la distribución del color del pelo es:

GRUPOS	PELO ROJIZO	PELO CLARO	PELO OSCURO	TOTAL
Grupo A	2	9	9	20
Grupo B	3	6	21	30
Grupo C	15	15	20	50
Total	**20**	**30**	**50**	**100**

¿Indican estos datos que el color del pelo depende de la región geográfica? (Use un nivel de significación del 1%).

54. El gerente de ventas de cierta compañía está interesado en probar si los vendedores son igualmente ambiciosos, enérgicos, imaginativos, etc., realizan igual número de visitas durante el mismo tiempo.

El número de ventas que realiza un vendedor, por lo general depende del número de visitas. Se selecciona una muestra aleatoria de 6 registros de los vendedores en una semana dada y se halla que se realizó un número total de 210 visitas, distribuidas de la siguiente forma:

VENDEDORA	A	B	C	D	E	F	Total
No. DE VISITAS	35	20	47	32	51	25	210

55. Una estadística de accidentes leves, ocurridos en dos fábricas A y B muestran que de 102 accidentes, 59 han tenido lugar en la fábrica A y 43 en la fábrica B. Formulemos la hipótesis de que no existe relación entre el número de accidentes y el hecho de que ocurran en la fábrica A o en la fábrica B (hipótesis nula).

56. El servicio comercial de una compañía, ha determinado las cuotas para cada una de las cinco regiones donde posee sucursales. Este servicio puede suponer que el número de artículos de sonido digital vendidos por región, durante la semana en cuestión no está conforme con las cuotas de reparto? ($\propto$ =1%).

REGIÓN	**NÚMERO DE ARTÍCULOS VENDIDOS**	**CUOTAS A ALCANZAR**
Costa Atlántica	61	18%
Antioquia	83	30%
Santanderes (oriente)	54	21%
Centro del país	46	17%
Valle (Occidente y Sur)	56	14%
Total	**300**	**100%**

57. Dos procedimientos de fabricación, A y B, han sido ensayados con el fin de aumentar la duración de la conservación de productos enlatados. Los resultados obtenidos son los siguientes:

PROCEDIMIENTOS	**FRACASOS**	**ÉXITOS**	**TOTALES**
A	77	63	140
B	54	66	120
Totales	131	129	260

¿A qué conclusión podemos llegar? ($\propto$ = 0,05)

58. En una operación de control, se clasifican las piezas producidas por cuatro máquinas en tres grupos, según su calidad (buenas, regulares y malas). Al nivel del 2%, según los resultados obtenidos ¿puede admitirse que la distribución de las piezas buenas, regulares y malas es la misma, independientemente de la máquina que las fabrica?

MÁQUINA No.	**CLASIFICACIÓN**			**TOTALES**
	Buenas	**Regulares**	**Malas**	
1	229	216	105	505
2	186	119	63	368
3	110	92	49	251
4	150	127	67	344
Totales	675	554	284	1.513

59. Se eligen 1.100 personas al azar de una gran ciudad y se les interroga acerca de sus hábitos de fumar y beber. Los resultados de la encuesta fueron:

BEBER	FUMAR				TOTAL	%
	Nunca	Ocasional-mente	Moderada-mente	Excesiva-mente		
Nunca	85	23	56	36	200	18,18
Ocasionalmente	153	44	128	75	400	36,36
Moderadamente	128	26	101	45	300	27,28
Excesivamente	134	7	15	44	200	18,18
Total	**500**	**100**	**300**	**200**	**1.100**	**100,00**

¿Qué conclusión debe deducirse respecto de los hábitos de fumar y beber, entre los habitantes de la ciudad? ($\propto = 0{,}05$).

60. Se elige una muestra aleatoria de 200 alumnos, cada una de cuatro clases de estudiantes no graduados. Se pregunta a cada alumno su preferencia por el ajedrez, fútbol y tenis, los resultados son:

CURSO	PREFERENCIA			TOTALES
	Ajedrez	Fútbol	Tenis	
1º Año	51	110	39	200
2º Año	74	106	20	200
3º Año	25	124	51	200
4º Año	90	60	50	200
Totales	**240**	**400**	**160**	**800**

¿Tienen los cuatro cursos la misma preferencia? ($\propto = 0{,}05$).

61. Contestar si es cierto o falso, a los siguientes puntos:

- ❑ χ^2 puede tener valores negativos
- ❑ Si se tienen 4 filas y 3 columnas, los grados de libertad son 5.
- ❑ La suma de los valores esperados es igual a la suma de los valores observados.
- ❑ En el caso de independencia, los totales de las filas y de las columnas se obtienen al azar.
- ❑ En el caso de homogeneidad, los totales de las filas son números elegidos
- ❑ La corrección de Yates se hace solamente cuando el número de grados de libertad es igual a 1.

62. Aplicar el test de homogeneidad, al caso referente al recuerdo de un slogan publicitario en una muestra de 374 estudiantes: 276 hombres y 98 mujeres; de los cuales 209 varones y 65 mujeres recordaron el slogan. Determinar si esta proporción permite sostener la hipótesis, de que el recuerdo no es igual en hombres y mujeres, a un nivel de 10%.

63. Un analista de mercados, tiene el problema de si las amas de casa que no se encuentran en ella cuando llaman los sondeadores, difieren en sus opiniones acerca de un determinado producto. Para comprobar esta posibilidad, los sondeadores regresaron a las casas hasta que obtuvieron una entrevista. Los resultados fueron:

OPINIÓN DEL PRODUCTO	OPINIÓN		TOTAL	%
	Primera llamada	Llamadas posteriores		
Excelente	62	36	98	0,36
Satisfactorio	84	42	126	0,47
No satisfactorio	24	22	46	0.17
Totales	**170**	**100**	**270**	**1,00**

Probar si las amas de casa que no se encontraron, tienen la misma opinión que las que si se encontraron ($\propto = 0{,}05$).

64. A 400 personas de ambos sexos, se les dio a probar margarina y mantequilla y se les pidió que expresaran su preferencia. ¿Al nivel del 5%, se puede afirmar que se diferencian significativamente las preferencias de los hombres y de las mujeres respecto a dichos productos?

SEXO	MARGARINA	MANTEQUILLA
Hombres	86	74
Mujeres	144	96

65. Dos operadores de una máquina, A y B, produjeron cierto artículo con el consiguiente número de defectuosos.

RESULTADOS	A	B
Defectuosos	40	60
No defectuosos	300	500

¿Existe alguna diferencia entre la proporción de defectuosos para los dos operadores? (Nivel del 1%).

66. De acuerdo a experiencias pasadas, la distribución en porcentajes de la opinión sobre determinada fábrica ha sido la siguiente:

OPINIÓN	%
a) La mejor empresa	85
b) Buena pero podría ser mejor	11
c) Mediocre	3
d) Mala	1

Una muestra de 3.030 personas presentan opciones así:

OPINIÓN	PERSONAS
a	2.573
b	292
c	124
d	41
TOTAL	3.030

¿Cree usted, al nivel de 10%, que los porcentajes de opinión son aún los mismos?

67. Número de incendios en cierta ciudad, durante una semana:

LUNES	MARTES	MIÉRCOLES	JUEVES	VIERNES	SÁBADO	DOMINGO	TOTAL
8	7	9	6	13	12	10	65

Verificar que los incendios no están homogéneamente distribuidos en la semana. (Nivel del 5%).

68. Se toma una muestra de 220 estudiantes de la facultad de Mercadotecnia y se clasifican de acuerdo a las calificaciones y sexo, como sigue.

CALIFICACIÓN	HOMBRES	MUJERES
Aprueban	110	80
Reprueban	20	10

A un nivel de significación del 5%, ¿hay alguna relación entre sexo y la aprobación de curso?

69. Se toman muestras, aleatorias entre bebedores y no bebedores, clasificándolos de acuerdo con sus hábitos de fumar, como sigue:

HÁBITO DE FUMAR	BEBEDORES	ABSTEMIOS
Fumadores en exceso	40	20
Fumadores promedio	60	40
Poco fumadores	80	40
No fumadores	10	60
Total	190	160

¿Existe alguna diferencia en la distribución de los fumadores entre bebedores y abstemios? (Nivel del 5%)

70. En tres ciudades: Bogotá, Medellín y Cali, se seleccionaron muestras aleatorias de 240, 200 y 160 personas: respectivamente. Se les preguntó qué tipo de programa de televisión preferían: novelas, musicales, ciencia ficción o comedias. Las respuestas fueron:

TIPO DE PROGRAMA	BOGOTÁ	MEDELLÍN	CALI	TOTAL
Novelas	70	40	40	150
Musicales	100	80	60	240
Ciencia ficción	40	50	30	120
Comedias	30	30	30	90
Total	240	200	160	600

Al nivel del 5%, probar la hipótesis de que las preferencias por el tipo de programa son las mismas en las tres ciudades.

71. Se toma una muestra aleatoria de 230 pacientes que sufren cierta enfermedad y se clasifican de acuerdo al tratamiento por cierto método y su recuperación es:

PACIENTES	CURADOS	NO CURADOS
Tratados	140	30
No tratados	20	40

¿Existe alguna relación entre el tratamiento y la curación de los pacientes? (α = 1%).

72. Diga si son ciertos o falsos, los siguientes puntos:
    a) Si se tienen 4 filas y 3 columnas, lo grados de libertad son 5.
    b) χ^2 puede tener valores negativos.
    c) Si se tiene una tabla de "2x2", los grados de libertad son de 1 y se debe utilizar la corrección de Yates.
    d) La suma de los valores esperados puede ser diferente a la suma de los valores observados.

73. Se desea saber si existe alguna relación entre las diferentes clases de fumadores y la posición familiar.

CARACTERÍSTICAS	POSICIÓN FAMILIAR	
	Padres	Hijos
No fumadores	8	30
Ocasionales	10	24
Habituales	40	56
Exfumadores	20	12

(Nivel del 5%)

74. Se tomó un grupo de 160 animales para probar el efecto de una vacuna contra una enfermedad particular. De los 100 vacunados, 30 resultaron enfermos. De los no vacunados, 36 enfermaron. ¿Qué se puede opinar de la efectividad de la vacuna? (Nivel del 10%).

75. Se toma una muestra de 200 niños a fin de determinar si el resultado (positivo o negativo) guarda relación con la aplicación de una vacuna.

RESULTADOS	VÍA DE APLICACIÓN	
	* Intradérmica	**Escarificación
Positivo	30	20
Negativo	70	80

Aplique el nivel del 5%, para probar si existe alguna diferencia significativa, en los resultados obtenidos.
* Interior de la piel ** Incisiones superficiales en la piel

76. Suponga que se tienen los siguientes resultados de una muestra aleatoria de 1.500 electores en un barrio de la ciudad, clasificados por afiliación de partidos y preferencia sobre determinado programa.

RESULTADOS	PREFERENCIA	
	Pro	Contra
Gobiernistas	600	375
Opositores	225	300

Probar la hipótesis, al nivel del 5%, de que la afiliación al partido no tiene que ver con la preferencia de voto.

77. Un almacén desea determinar si el sexo de sus posibles clientes guarda relación con los artículos que vende.

ARTÍCULOS	SEXO	
	Masculino	Femenino
Neveras	380	400
Radios	260	300
Televisores	270	350

Para esa muestra de 1.960 clientes, al nivel del 5% probar la hipótesis de que el sexo no tiene relación con la preferencia.

78. Con el objeto de investigar la relación existente, entre los niveles de ingreso y el nivel de sintonía de los noticieros de televisión, se selecciona una muestra de 861 televidentes, estratificados por niveles de ingreso. Los resultados obtenidos son:

NIVEL DE INGRESO	TELEVIDENTES			Total
	Frecuente	Ocasional	Nunca	
Alto	220	70	30	320
Mediano	131	100	80	311
Bajo	30	80	120	230
Total	**381**	**250**	**230**	**861**

De acuerdo a lo anterior, pruebe que existe alguna relación entre los niveles de ingreso y la teleaudiencia en los noticieros, al nivel del 1%.

79. Por experiencia, en las ventas de televisores a color, se sabe que los porcentajes de preferencia según el tamaño de la pantalla es el siguiente:

14" el 18%	16" el 22%	19" el 40%	21" el 20%

Se realizó una muestra aleatoria en 200 compradores de este artículo y se obtuvo la siguiente información:

14" un total de 42 compradores.	16" un total de 40 compradores.
19" un total de 88 compradores.	21" un total de 30 compradores.

¿Cree usted, al nivel del 5%, que la preferencia por los tamaños de las pantallas es el mismo?

80. Se ha solicitado a 4.000 propietarios de vivienda, adquirida en los dos últimos años que señalen la razón principal por la cual la han comprado. Las dos principales razones son:

SEXO	RESPUESTA	
	Acabado	Precio
Masculino	950	550
Femenino	850	1.650

Pruébese con $\propto = 0{,}05$ el supuesto de que la respuesta es independiente del sexo del propietario entrevistado.

81. En una investigación de opinión pública, se hizo a 1.000 habitantes de una ciudad la siguiente pregunta: ¿cómo calificaría el desempeño del señor X, como alcalde de la ciudad: bueno, regular, malo? Las respuestas, clasificadas de acuerdo al nivel educativo de los encuestados, fueron:

DESEMPEÑO	NIVEL EDUCACIONAL			Total
	Elemental	Secundario	Universitario	
Bueno	82	427	191	700
Regular	10	110	60	180
Malo	8	63	49	120
Total	**100**	**600**	**300**	**1.000**

Utilizando un nivel del 5%, ¿podemos concluir que la calificación de su desempeño es independiente del nivel educacional de los encuestados?

82. El "temor a un examen" es una emoción que experimentan muchas personas, de todos los grupos de edad. Una encuesta a 276 individuos pertenecientes a cada uno de cinco grupos de edad, arrojó los siguientes resultados:

GRUPOS	PRIMARIA	SECUNDARIA	TECNOLÓGICA	UNIVERSITARIA	ESPECIALIZACIÓN
Número de temerosos	74	56	50	54	42

Pruebe, al nivel del 5%, la afirmación de que cada grupo de edad teme igualmente a los exámenes.

83. En un estudio sobre delincuencia juvenil, se clasificaron 100 jóvenes, en edad escolar, de acuerdo con el grado de delincuencia y el contacto con los padres durante los ratos libres. ¿Proporcionan estos datos evidencia suficiente, como para indicar que hay una relación en cuanto a los criterios de clasificación? Nivel del 5%.

TIEMPO LIBRE COMPARTIDO CON EL PADRE	DELINCUENTES	NO DELINCUENTES	TOTAL
Alto	10	29	39
Bajo	20	41	61
Total	**30**	**70**	**100**

84. Se ha tomado una muestra de cada una de dos máquinas que envuelven caramelos:

MÁQUINA	ROSA	LILA	AMARILLO	ANARANJADO	VERDE	BLANCO
A	11	77	6	5	3	3
B	6	7	10	7	6	6

Pruébese la independencia de la mezcla de colores con la máquina que los envuelve, con $\propto = 0{,}01$.

85. El número de reclamaciones recibidas en 5 establecimientos de una cadena de tiendas, por concepto de mercancía equivocada o defectuosa en un cierto mes del año, es la siguiente:

TIENDA No.	1	2	3	4	5
RECLAMACIONES	63	78	54	49	56

Sobre la base de estos resultados, ¿es sostenible la hipótesis de que el número de reclamaciones que recibe cada tienda, es aproximadamente el mismo? ($\propto = 0{,}05$)

86. Los siguientes datos muestran el efecto de cierto tipo de fumigación sobre el deterioro de la fruta. Siendo una tabla de "2x2", se obtuvo la siguiente información.

$m_1 = 10$	$m_2 = 30$	$m_3 = 24$	$m_4 = 16$	$n = 40$	$AD = 112$	$BC = 32$

Al nivel del 5%, ¿depende la cantidad de fruta deteriorada de su fumigación?

87. En un semestre, se ofrecieron siete secciones de un curso particular de matemáticas. Se programaron diversas horas, con distintos profesores. A continuación se presenta el número de alumnos que eligieron cada una las siete secciones:

SECCIÓN	1	2	3	4	5	6	7	Total
No. ALUMNOS	18	12	25	23	8	19	14	119

¿Al nivel del 5%, indican los datos que los alumnos muestran preferencia por algunas secciones?

DÓCIMA O PRUEBA DEL SIGNO

La prueba del signo es *el método no paramétrico* más sencillo, se emplea de preferencia, cuando se observan números pares en condiciones diferentes, utilizando únicamente el signo de las diferencias entre las observaciones apareadas.

> El análisis estadístico que no depende del conocimiento de la distribución, ni de los parámetros de la población, se denomina *método no paramétrico o de distribución libre*.

Esta prueba es aplicada, en aquellos experimentos cuyos resultados son observados *antes y después*; por lo tanto tal como se trabajó en *observaciones apareadas*, se establece la ***diferencia*** para cada par de observaciones, pero sólo tomando el signo, de ahí que se denomine *prueba del signo*, diseñada en especial para ***distribuciones no paramétricas***.

La *dócima del signo* es usada erróneamente en ***distribuciones paramétricas***, debido a la facilidad de cálculo. Sin embargo, se recomienda evitar su aplicación y sólo debe utilizarse en casos extremos, ya que esta prueba no es tan eficiente como las paramétricas, con frecuencia conlleva a aumentar considerablemente la posibilidad de cometer error de tipo II (aceptar algo falso). Además, para ser utilizada requiere que n (el número de pares), sea relativamente grande, 15 o más pares.

Al suponer que las dos poblaciones tienen medias idénticas, el número de signos positivos debe ser aproximadamente igual al número de signos negativos.

Se designa $p = 1/2$ (signos positivos), $q = 1/2$ (signos negativos) y D_i la diferencia entre x y y. Se cuenta el número de signos positivos y negativos; la suma corresponde a n, es decir, se excluyen las diferencias iguales a cero. Mediante la *Distribución binomial* se puede obtener la probabilidad (dado el supuesto de igualdad de presentación de los signos positivos y negativos), de que aparezcan más signos positivos, negativos o que la diferencia sea significativa. La simbología a utilizar en este método será:

$p = 0{,}5$	Probabilidad de presentación del signo positivo.
$q = 0{,}5$	Probabilidad de presentación del signo negativo.
n	Suma del número de positivos más los negativos, excluyendo los ceros.
x	El número de signos positivos (puede ser considerado el número de signos negativos).

La información anterior permite utilizar la *Distribución binomial*, pero es preferible trabajar con la *Distribución normal*, a fin de agilizar el procedimiento; para ello se debe recordar el proceso a seguir, para calcular la media y desviación estándar en una Distribución binomial.

μ = Media de la distribución binomial $\rightarrow \quad \mu = np$
σ = Desviación típica en la distribución binomial $\rightarrow \quad \sigma = \sqrt{n\ p\ q}$
X = Valor de la variable. (Número de positivos o negativos)

Se debe tener en cuenta que *X*, puede corresponder al número de signos positivos o negativos. Sin embargo, es preferible trabajar con el número de positivos, observando que si es mayor que μ se le debe restar 0,5; si por el contrario es menor que μ se le suma 0,5 al valor de X, uno de los componentes de la fórmula para la distribución normal.

La *variante estadística* estará dada: $Z = \dfrac{X - \mu}{\sigma}$

D_i= La diferencia entre x_i y y_i, se tomará únicamente el signo de cada par.

Recordemos que la prueba de significación depende de los signos y no de las magnitudes entre ciertas cantidades.

Ejemplo 1. Se detallan a continuación los rendimientos de dos tipos de maíz, germinados sin mezclas, obtenidos en diferentes experimentos. ¿Existe alguna diferencia significativa entre ambos tipos de maíz? (Nivel del 5%)

EXPERIMENTO	RENDIMIENTO A	RENDIMIENTO B	$D_i = x_i - y_i$
1	48,6	42,3	+
2	43,0	45,1	–
3	41,0	41,0	0
4	39,1	36,4	+
5	36,8	32,8	+
6	40,8	42,6	–
7	28,9	30,3	–
8	35,9	33,2	+
9	39,2	39,2	0
10	47,8	45,4	+
11	42,1	40,6	+
12	47,6	48,5	–
13	26,4	26,4	0
14	30,6	28,6	+
15	39,8	36,4	+
16	37,5	35,0	+
17	28,1	29,1	–
18	33,3	31,0	+
19	43,0	43,0	0
20	29,0	27,0	+

Solución:

$n = 16$ $\quad p = 1/2$ $\quad q = 1/2$ $\quad X = 11$ (Número de +)

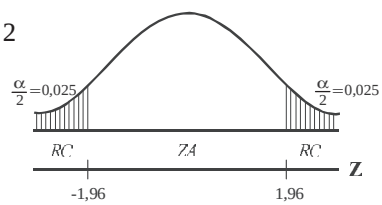
$\mu = np = 16\,(1/2) = 8$ $\quad \sigma = \sqrt{npq} = \sqrt{16\,(1/2)\,(1/2)} = 2$

Mediante la Distribución normal, como aproximación a la Binomial, tenemos que: $Z = \dfrac{X - \mu}{\sigma}$

1) $H_0 : P = 1/2$
$H_a : P \neq 1/2$

2) $\propto = 0{,}05$

3) $\sigma = 2$



4) $Z = \dfrac{10{,}5 - 8}{2} = 1{,}25$

Se acepta la hipótesis nula, ya que 1,25 se ubica en la zona de aceptación, se dice entonces, que la diferencia entre ambos tipos de maíz no es significativa, al nivel del 5%.

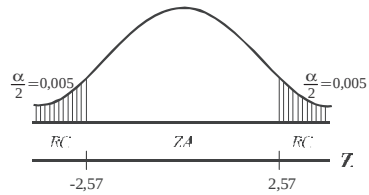
> *Nota*: se toma 10,5, correspondiente al límite inferior de $X = 11$; por ser 11 mayor que 8 (valor de μ). Recordemos que la distribución binomial requiere que la variable sea discreta y al utilizar la normal, se le debe convertir en continua.

El problema se hubiera podido resolver utilizando el número de signos negativos como *X*. Observemos el procedimiento a seguir:

1) $H_0 : P = 1/2$ 2) $\propto = 0{,}01$
$H_a : P \neq 1/2$

3) $\sigma = 2$

4) $Z = \dfrac{X - \mu}{\sigma} = \dfrac{5{,}5 - 8}{2} = -1{,}25$



Se acepta H_0, la diferencia no es significativa, al nivel del 5%.

Otro procedimiento de cálculo

Existe otro procedimiento mucho más fácil en la aplicación de la *dócima del signo* y en considerar, en primer lugar, el número de veces que ocurre el signo menos frecuente, simbolizado por *p*. El ejercicio en consideración presenta 5 signos negativos, por lo tanto $p = 5$.

El valor crítico para una dócima bilateral, con un nivel de significación del 5%, se puede obtener mediante la aplicación de la fórmula siguiente:

$k = \dfrac{(n-1)}{2} - (0{,}98)\sqrt{n+1}$ reemplazando se tiene:

$$k = \frac{16-1}{2} - 0{,}98\sqrt{16+1} = 7{,}5 - 0{,}98\,(\ ,12\ ) = \ ,\ 6$$

Es necesario aclarar, que esta fórmula sólo se aplica en *pruebas bilaterales*, con un nivel de significación del 5%.

> Si $S \leq k$ se rechaza la hipótesis nula, ahora si $S > k$ se acepta H_0. Teniendo en cuenta el ejercicio anterior donde $5 > 3{,}46$, es decir que $S > k$, se acepta la hipótesis nula, o sea que no existe una diferencia significativa.

Ejemplo 1. Se detallan a continuación los rendimientos de dos tipos de maíz, germinados sin mezclas, obtenidos en diferentes experimentos. ¿Existe alguna diferencia significativa entre ambos tipos de maíz, con base en el nivel de significación del 1%?

EXPERIMENTO	REND. A	REND. B	EXPERIMENTO	REND. A	REND. B
1	48,6	50,1	15	39,8	40,8
2	43,0	48,6	16	37,5	37,3
3	41,0	42,9	17	28,1	27,5
4	39,1	42,6	18	33,3	32,4
5	36,8	37,5	19	43,0	48,6
6	40,8	41,3	20	29,0	31,1

Continúa...

Continuación

7	28,9	38,6	21	27,4	28,0
8	35,9	37,3	22	28,0	28,7
9	39,2	42,6	23	28,3	28,8
10	47,8	46,1	24	26,8	26,8
11	42,1	43,4	25	42,2	42,2
12	47,6	48,2	26	38,9	39,1
13	26,4	26,3	27	39,0	39,4
14	30,6	31,7	28	33,6	34,0

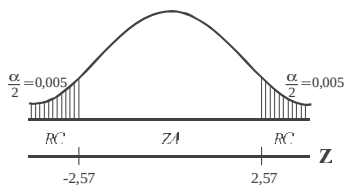
Solución:

EXPERIMENTO	x_i	y_i	$D_i = x_i - y_i$
1	48,6	50,1	-
2	43,0	48,6	-
3	41,0	42,9	-
4	39,1	42,6	-
5	36,8	37,5	-
6	40,8	41,3	-
7	28,9	38,6	-
8	35,9	37,3	-
9	39,2	42,6	-
10	47,8	46,1	+
11	42,1	43,4	-
12	47,6	48,2	-
13	26,4	26,3	+
14	30,6	31,7	-
15	39,8	40,8	-
16	37,5	37,3	+
17	28,1	27,5	+
18	33,3	32,4	+
19	43,0	48,6	-
20	29,0	31,1	-
21	27,4	28,0	-
22	28,0	28,7	-
23	28,3	28,8	-
24	26,8	26,8	0
25	42,2	42,2	0
26	38,9	39,1	-
27	39,0	39,4	-
28	33,6	34,0	-

Para obtener el valor de *n*, se toman únicamente los signos positivos y negativos.
Signos positivos = 5 Signos negativos = 21 $n = 26$

$$p = \frac{1}{2} \qquad q = \frac{1}{2} \qquad \mu = np = 26\left(\frac{1}{2}\right) = 1$$

$$\sigma = \sqrt{npq} = \sqrt{26\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 2{,}55$$



1) $H_0 : P = 0{,}5$ 2) $\propto = 0{,}01$ 3) $\sigma = 2{,}55$
$H_a : P \neq 0{,}5$

Continúa...

$$Z = \frac{X - \mu}{\sigma} = \frac{5{,}5 - 1}{2{,}55} = -2{,}9$$

Se rechaza la hipótesis nula a un nivel de significación del 1%. Se acepta que existe una diferencia significativa entre ambos tipos de maíz.

El ejercicio anterior se puede resolver aplicando la siguiente fórmula; en el caso de utilizar un nivel de significación del 5% y si la prueba es bilateral.

$$K = \frac{n-1}{} - \qquad \sqrt{n+1} \qquad K = \frac{(26-1)}{2} - 0{,}98\sqrt{26+1} = 12{,}5 - 5{,}09 = 7,\ 1$$

El número de signos positivos es el menos frecuente; en total 5 y lo simbolizamos por *p*.

Siendo $S = 5$, nos encontramos que $S \leq k$, por lo tanto rechazamos la hipótesis nula.

Ejemplo 2. Se realiza un experimento apareado de alimentación de cerdos, para determinar el valor relativo de calizas y harina de hueso, en el desarrollo óseo. En cada par se registra, en porcentaje, el contenido de ceniza en omoplatos, denominando con *X* los cerdos alimentados con calizas y con *y* aquéllos alimentados con harina de huesos. Si se realizan 35 experimentos, en 27 de los cuales *X* es mayor que *Y* y en 3 pares de observaciones, *X* es igual a *Y*. Señalar en el resultado que la caliza produce mayor contenido de ceniza que la harina de huesos. (Nivel del 5%)

Solución:

Signos positivos = 27 Signos negativos = 5 $n = 32$

$p = 0{,}5$ $q = 0{,}5$ $\mu = np = 2\left(\frac{1}{2}\right) = 16$

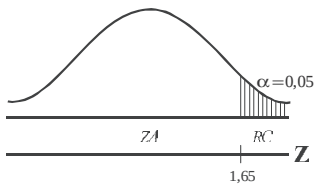
$$\sigma = \sqrt{n\,p\,q} = \sqrt{2\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 2{,}8$$

$H_0 : P = 0{,}5$ $2)\ \propto = 0{,}05$

$H_a : P > 0{,}5$

$\sigma = 2{,}8$

$$Z = \frac{X - \mu}{\sigma} = \frac{26{,}5 - 16}{2{,}8} = \ ,71$$



Aceptamos que la caliza produce mayor contenido de ceniza que la harina de huesos, al nivel del 5%.

La *prueba del signo* se debe aplicar en todos aquellos casos, en los cuales no nos interesa la *mensurabilidad* de los datos, sino su *comparabilidad*, cuando las observaciones están dadas en condiciones diferentes, por lo tanto se tienen distribuciones diferentes.

Ejemplo 3. Un experimento apareado realizado a 25 alumnos, sobre el rendimiento académico obtenido en una asignatura, aplicando dos sistemas de evaluación diferentes, arrojó los siguientes resultados:

Positivo = 13 pares Cero = 7 pares Negativo = 5 pares

¿Al nivel 5%, se puede afirmar que hay una diferencia significativa en la aplicación de los dos sistemas?

Solución:

$n = 1 + 5 = 18$ $\quad \mu = 18(1/2) = 9$ $\quad \sigma = \sqrt{n\,p\,q}$

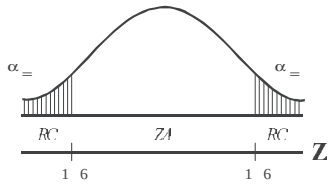
$n = 18$ $\quad \mu = np$ $\quad \sigma = \sqrt{18\ (1/2)\ (1/2)} = \sqrt{\ ,5} = 2{,}12$

Como X > μ, es decir, 13 > 9, se le resta 0,5 a X, siendo X = 12,5. Por lo tanto se tendrá que:

1) $H \quad P =$
$H_a \quad P \neq$

2) $\propto =$

3) $Z = \dfrac{X - \mu}{\sigma} = \dfrac{1 \quad -}{1} = 1\,6$



Como 1,65 cae en la zona de aceptación, se acepta H_0, por lo tanto la diferencia entre los dos sistemas no es significativa, al nivel del 5%.

En una hipótesis, correspondiente a una prueba bilateral, con un nivel de significación del 5%, se puede emplear la siguiente fórmula:

$$K = \frac{(n-1)}{2} - (0{,}98)\ \sqrt{n+1} \qquad K = \frac{(18-1)}{2} - (0{,}98)\ \sqrt{18+1} = \ ,7$$

Observando que el signo negativo es menos frecuente, se tendrá que *S* = 5. Ahora si *S* > k se acepta la hipótesis nula (5 > 4,73), aceptando que la diferencia entre los dos sistemas no es significativa.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

88. En un experimento apareado (*no paramétrico*) se dan las velocidades (*en número de elementos empaquetados por hora*) por dos sistemas distintos. Docime a un nivel del 5% que el primer sistema es superior.

PAR	SISTEMA I	SISTEMA II
1	10	6
2	11	7
3	9	12
4	8	9
5	15	13
6	12	7
7	11	13
8	13	9
9	15	8
10	10	10
11	13	11
12	12	9

Continúa...

Continuación

13	14	14
14	13	8
15	14	10
16	12	8
17	14	7
18	12	7
19	15	12
20	15	15

89. Los pesos, en libras, de 20 personas antes y después de haberse sometido a una dieta, cuya duración es de dos semanas, son:

PERSONAS	PESO (en libras)	
	Antes	Después
1	186	175
2	147	144
3	128	125
4	167	167
5	183	182
6	176	177
7	159	154
8	212	203
9	192	187
10	177	169
11	158	158
12	204	197
13	188	182
14	157	160
15	189	181
16	149	151
17	172	169
18	185	184
19	191	187
20	200	195

Utilizando la dócima del signo y un $\propto = 5\ \%$, docimar la hipótesis de que la dieta no es efectiva.

90. Las parejas de 15 matrimonios fueron sometidas a un "test" relativo al sentido del humor de una persona. Utilizar la dócima del signo, con una alternativa de que, en promedio, no hay diferencia en el sentido del humor de esposos y esposas, con base en los siguientes resultados obtenidos.

PAREJAS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Esposo:	56	90	38	47	85	49	55	58	68	74	83	87	60	31	89
Esposa:	49	88	51	50	83	41	52	69	83	89	77	62	65	44	92

91. El número de piezas defectuosas producidas por dos máquinas durante el mes de septiembre de 2011 fueron:

DÍAS:		1	2	3	4	5	6	7	8	9	10
Número defectuosos	A:	8	5	10	7	9	11	3	6	6	8
	B:	9	4	12	9	10	11	6	7	8	10

DÍAS:		11	12	13	14	15	16	17	18	19	20
Número defectuosos	A:	5	6	9	3	7	10	9	13	8	4
	B:	6	9	11	6	7	8	12	13	9	6

DÍAS:		21	22	23	24	25	26	27	28	29	30
Número defectuosos	A:	2	8	8	6	5	10	9	4	7	3
	B:	3	9	10	9	7	11	9	6	6	5

Atendiendo a que los métodos corrientes (paramétricos) no pueden utilizarse, debido a que en los días se emplearon materias primas diferentes, utilizar la prueba del signo, una alternativa bilateral y un nivel de significación del 5%, para docimar la hipótesis de que no hay diferencia significativa.

92. Clasificaciones de dos grupos de personas con cualidades de liderazgo (datos hipotéticos):

NÚMERO:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Experimental:	46	42	38	36	30	28	25	20	20	17	14	9	8	7	5	4
Control:	40	42	36	38	32	25	25	22	17	15	10	6	9	5	3	3

Aplique la prueba del signo y la normal (observaciones apareadas), para determinar si las diferencias son significativas. (Nivel del 5%).

93. El gerente de producción de una empresa, desea realizar un experimento para comparar dos métodos de ensamble de cierto mecanismo. Se seleccionan 18 pares de obreros, apareándose de acuerdo a la edad, se les asignan los métodos en forma aleatoria.

PARES:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Método I:	27	28	10	20	11	11	15	27	21	20	28	32	30	26	38	27	34	36
Método II:	20	25	10	21	11	13	18	20	16	20	23	29	26	28	29	23	26	33

Probar la hipótesis, al nivel del 5%, de que los datos anteriores indican una diferencia significativa.
a) Utilizar la prueba del signo.
b) Utilizar el método de las observaciones apareadas.

94. Se realiza un estudio de tiempos y movimientos, con el fin de probar la eficiencia de dos procedimientos al realizar ciertos trabajos en una entidad bancaria. Se aparean los empleados con base en los coeficientes de inteligencia y el número de años de estudio. Los miembros de cada par son asignados en grupos a uno de los dos procedimientos. El tiempo establecido en segundos para realizar el trabajo es el siguiente:

PAR:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
I:	440	400	400	310	300	410	260	260	320	380	360	370	290	290	286	310	252
II:	410	440	290	270	300	360	250	300	340	350	360	350	250	290	290	320	252

a) Aplique la prueba del signo, para determinar si la diferencia entre los procedimientos es significativa.
b) Aplique la prueba "*t*" para observaciones apareadas, a fin de determinar si la diferencia es significativa.
c) Compare los resultados anteriores obtenidos.
d) Fije los límites de confianza del 99% para los puntos (a) y (b).

95. En un estudio para determinar el efecto de una droga sobre la agresividad, se formaron dos grupos y se aparearon; al grupo A se le suministró la droga y al grupo B se le dio una droga que no producía ningún efecto. Inmediatamente después de administrar la droga, se realizó una prueba para comprobar la agresividad. Los puntajes obtenidos fueron: (*a mayor clasificación, mayor agresividad*).

PAREJAS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Grupo A:	10	8	12	16	5	9	7	11	8	16	8	5	8	6	12	14	8	4	9
Grupo B:	10	15	10	18	13	14	9	16	6	16	6	5	5	4	11	12	8	2	8

Utilizando la prueba del signo, al nivel del 5%, ¿se puede afirmar que la diferencia es significativa?

96. Una empresa realiza un estudio para determinar el número de unidades producidas por cada uno de los obreros, luego de ser sometidos a un curso de capacitación, a fin de aumentar el rendimiento por obrero.

PAR:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Antes:	45	44	29	27	30	36	35	30	34	40	43	29	32	38	34	28	32	36
Después:	45	48	36	27	28	32	31	42	36	42	44	29	30	42	40	28	30	40

Se pide:

a) Aplicar la prueba del signo, para determinar si en realidad se mejoró el rendimiento (Nivel 5%)
b) Aplicar la dócima de normalidad (observaciones apareadas) al Nivel del 5%, para determinar si el curso ayudó a aumentar el rendimiento; en el punto, (b) fijar límites de confianza del 95%.

97. Para evaluar la acción hipertensiva de una droga, se tomó la tensión arterial a un grupo de individuos, antes y después de su aplicación; los resultados fueron:

INDIVIDUOS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Antes:	65	65	76	84	68	70	66	82	64	65	88	82	70	70	75	80	92
Después:	72	70	78	84	74	76	72	79	65	70	88	80	76	76	75	86	80

Utilice el nivel de significación del 5%, considerando que la muestra proviene de poblaciones dependientes y luego trabájela como perteneciente a una distribución no paramétrica, que las diferencias no son significativas.

98. Una prueba de ansiedad, fue aplicada a un grupo de 18 estudiantes universitarios varones estando sobrios y se repitió dos horas después de que cada uno había ingerido 100 cc de aguardiente. Los resultados del índice de ansiedad fueron:

SUJETO :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Antes :	98	81	72	63	92	63	81	73	63	82	54	88	93	68	97	81	72	63
Después :	82	71	63	63	90	72	66	54	63	88	50	88	82	52	82	81	63	54

a) Al nivel del 5% mediante, la prueba del signo, ¿se puede concluir que el alcohol tiene efectos sobre la ansiedad?
b) Al nivel del 5%, mediante la prueba "*t*" (observaciones apareadas), ¿se podrá concluir que el alcohol no reduce la ansiedad?

99. Para probar los efectos de un curso de educación física sobre la condición corporal de los alumnos, se ha registrado el número de "sentadillas" que una persona puede realizar en un minuto. Diecisiete participantes escogidos al azar, muestran las siguientes puntuaciones:

PERSONAS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Antes :	29	22	25	29	26	24	21	20	46	38	28	30	29	25	21	22	23
Después :	30	26	25	35	33	36	32	20	54	58	43	30	30	20	20	20	23

Con las anteriores observaciones apareadas, provenientes de poblaciones dependientes:
a) Pruebe la hipótesis de que hubo mejoramiento, al nivel del 5%.
b) Utilizando la prueba del signo, pruebe la hipótesis de que hubo mejoramiento, al nivel del 5%.

100. Un químico asegura que ha descubierto un aditivo para la gasolina, que aumenta el rendimiento del combustible. Para probar el rendimiento se ponen a prueba 18 autos con y sin aditivo. Los resultados en kilómetros/ galón son:

AUTOMÓVIL:	1	2	3	4	5	6	7	8	9
Sin aditivo:	40,6	63,3	48,2	38,4	39,2	42,6	42,0	46,1	51,0
Con aditivo:	51,2	62,1	52,3	42,0	43,5	40,5	42,0	50,2	51,0

AUTOMÓVIL:	10	11	12	13	14	15	16	17	18
Sin aditivo:	39,1	38,6	42,0	38,6	38,4	37,3	38,6	39,3	37,6
Con aditivo:	42,6	40,4	42,4	37,5	39,6	40,2	41,2	38,4	40,2

a) Considere que las muestras provienen de poblaciones dependientes. (Observaciones apareadas) Se pide establecer los límites de confianza del 95% y probar si la diferencia es significativa (Nivel del 5%).
b) Considere que las muestras provienen de poblaciones no paramétricas (Aplicando la prueba del signo) y determine si la diferencia es significativa (Nivel del 5%).

101. En un esfuerzo por disminuir los costos de transporte, una empresa de taxis imparte clases de *manejo económico* a cada uno de sus 20 conductores. Ponga a prueba los resultados, utilizando un nivel de significación de 0,05 en observaciones apareadas y en la prueba de signos. (*Utilice los dos métodos de cálculo*).

CONDUCTOR:	1	2	3	4	5	6	7	8	9	10
Antes:	30,4	28,6	27,6	34,2	32,8	30,2	30,1	32,4	33,3	28,4
Después:	30,4	29,4	30,0	34,4	33,8	30,4	32,4	34,5	33,3	30,6

CONDUCTOR:	11	12	13	14	15	16	17	18	19	20
Antes:	33,6	36,4	36,6	35,4	30,6	37,6	37,6	38,6	34,5	33,8
Después:	31,4	34,2	36,6	34,5	33,5	36,6	39,4	38,6	37,3	31,4

Nota: El rendimiento se tomó en kilómetros / galón.

102. En un experimento apareado (no paramétrico) se dan las velocidades (número de elementos empaquetados por hora), aplicando dos sistemas diferentes con los siguientes resultados:

Positivo = 14 pares	Negativo = 3 pares	Cero = 3 pares

¿Al nivel de 5% la diferencia es significativa?

CORRELACIÓN POR RANGOS

Conocida también como *Método de Spearman*, algunos la denominan *Prueba del coeficiente de correlación de las categorías*, establecida y empleada por Pearson y Spearman desde 1936; esta medida nos determina la existencia del grado de asociación entre dos variables, utilizando para ello una escala ordinal, de tal manera que se puedan colocar en dos series ordenadas.

La fórmula a utilizar es: $r_s = 1 - \frac{6\Sigma d_i}{n - n}$

(En el capítulo 12 sobre "Regresión y Correlación" hay un complemento de este tema, que debe ser consultado).

Ejemplo 1. Supongamos que se tienen los siguientes 16 pares de observaciones, a 16 trabajadores que fueron evaluados con dos instrumentos.

No.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_i:	85	78	75	65	80	83	76	65	83	76	83	65	90	80	84	76
y_i:	50	38	30	25	38	42	35	25	45	35	42	25	55	40	48	35

Primero: Ordenamos los datos de menor a mayor, o de menor a mayor, de esta manera establecemos los rangos:

x_i :	65	65	65	75	76	76	76	78	80	80	83	83	83	84	85	90
Rango :	2	2	2	4	6	6	6	8	9,5	9,5	12	12	12	14	15	16
y_i :	25	25	25	30	35	35	35	38	38	40	42	42	45	48	50	55
Rango :	2	2	2	4	6	6	6	8,5	8,5	10	11,5	11,5	13	14	15	16

Segundo: Reemplazamos cada valor de *X* y *Y* en la tabla original (enunciado) por su respectivo rango, luego establecemos las diferencias entre cada rango y las elevamos al cuadrado. Finalmente sumamos la columna $d_i^2 = 357$

Rango x_i :	2	9,5	6	4	9,5	8	12	14	15	16	2	12	6	6	12	2
Rango y_i :	4	2	6	10	8,5	11,5	14	15	16	13	2	6	8,5	6	2	11,5
d_i :	-2	7,5	0	-6	1	-3,5	-2	-1	-1	3	0	6	-2,5	0	10	-9,5
d_i^2 :	4	56,25	0	36	1	12,25	4	1	1	9	0	36	6,25	0	100	90,25

$$r_s = 1 - \frac{6(57)}{16 - 16} = 1 - \frac{2,12}{,080} = 1 - 0,525 = 0,75$$

PRUEBA DE r_s

Cuando n es grande ($n \geq 25$) se considera que r_s tiene el comportamiento de una distribución aproximadamente Normal con Media 0 y Desviación típica $1/\sqrt{n-1}$, por tal razón calculamos *Z*, siendo:

$$Z = \frac{r_s - 0}{1/\sqrt{n-1}}$$

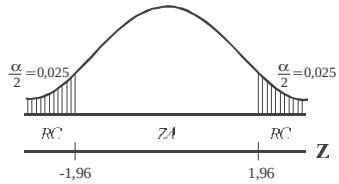
Ejemplo 2. Suponiendo que $n = 28$ $r_s = 0{,}48$, al nivel del 5% probar que el coeficiente de correlación es significativo.

Solución: $Z = \frac{r_s -}{1/\sqrt{n-1}} = r_s\sqrt{n-1} =$ $\sqrt{\quad -1} =$

1) H_0: $R_s = 0$ (*no es significativo, no hay correlación*)
   H_a: $R_s \# 0$ (*es significativo, hay correlación*)

2) $\propto = 0{,}05$

Al nivel del 5%, el coeficiente de correlación es significativo. Se puede concluir que estas dos variables están asociadas en la población de donde se seleccionó la muestra.



Cuando n es pequeña ($10 < n < 25$) o bien si el valor de r_s es pequeño, es aconsejable utilizar la distribución "t" de Student, donde:

$$t = \frac{r_s - 0}{\sqrt{(1 - r_s^2)/(n-2)}} = \frac{r_s\sqrt{n-2}}{\sqrt{1 - r_s^2}}$$

Si consideramos el ejercicio donde, $n = 16$ y $r_s = 0{,}475$, se observa que:

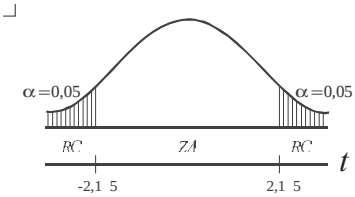
$$t = \frac{0,\ 75}{\sqrt{(1 - 0,\ 75^2)\ /\ (16-2)}} = 2{,}02 \quad \text{ó} \quad t = \frac{0,\ 75\sqrt{16-2}}{\sqrt{(1 - 0,\ 75^2)}} = 2{,}02$$

1) $H_0 : R_s = 0$ $\quad \upsilon = n - 2 = 1$
   $H_a : R_s \neq 0$ $\quad \propto = 0{,}05$

$$\begin{array}{|l} H_o : \rho_s = 0 \\ H_1 : \rho_s \neq 0 \end{array}$$

2) $\propto = 0{,}05$ $\quad t_{0{,}05} = 2{,}1\ \ 5$

Al nivel del 5%, se concluye que no hay correlación, es decir, que la correlación obtenida no es significativa. También se puede decir que estas dos variables no están asociadas en la población de donde se extrajo la muestra.



En *muestras pequeñas*, se puede utilizar la TABLA XVII, donde aparecen los valores teóricos para r_s, veamos su uso:

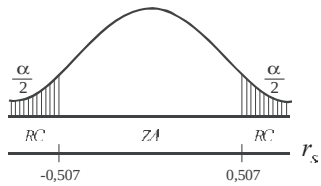
1) $H_0 : R_s = 0$
   $H_a : R_s \neq 0$

$$\begin{array}{|l} H_o : \rho_s = 0 \\ H_1 : \rho_s \neq 0 \end{array}$$

2) $\propto = 0{,}05$

   r_s = *calculado* $= 0,\ 75$

   $n = 16$



Observamos que $r_s = 0{,}475$ cae en la zona de aceptación, es decir, aceptamos H_0, donde $R_s = 0$, implica que no hay correlación. Se puede decir, que al nivel de 5%, estas dos variables no están asociadas, resultado igual al obtenido cuando aplicamos la TABLA II, correspondiente a la "t" de Student.

Para muestras pequeñas ($n \leq 30$) la distribución r_s no es normal; muchos autores utilizan la "t" de Student, tal como se hizo en el ejercicio anterior, sin embargo, no son pocos quienes sostienen que este proceso o método no es tan apropiado, por el contrario prefieren utilizar la Tabla XVII (al final del libro), correspondiente a los *Coeficientes de Spearman*, como el utilizado anteriormente para muestras pequeñas.

LA PRUEBA T DE WILCOXON

Denominada por algunos como *Prueba del signo de las categorías*, ya que no sólo establece las diferencias para cada par de observaciones, sino que las clasifica, aspecto éste que se tiene en cuenta en la *prueba del signo*, por tal razón se le considera de mayor eficacia entre las *pruebas no paramétricas*. Esta prueba fue establecida por Frank Wilcoxon en el año de 1945.

MUESTRAS PEQUEÑAS

Veamos primero el procedimiento que se sigue en muestras pequeñas para pares *igualados*.

- Se obtienen las diferencias para cada par de observaciones, $d_i = x_i - y_i$
- Luego establecemos categorías, ordenando sus diferencias de acuerdo con sus magnitudes en valor absoluto, es decir, no tomamos en consideración el signo.
- Procedemos a determinar el valor de T, para ello seleccionamos la suma más pequeña para uno de los signos.
- Determinamos el valor de n, con el fin de conocer el tamaño. Si $n \leq 25$ utilizamos la tabla XVIII (al final del libro), pero si $n > 25$, se puede admitir que la suma de los rangos tiene una *Distribución normal*, por lo tanto se debe utilizar la tabla *Z*.

Nota: Vale la pena anotar que algunos autores consideran una muestra pequeña cuando $n \leq 50$ y grande cuando $n > 50$.

Ejemplo 1. Para explicar el procedimiento e interpretación de los resultados obtenidos, partamos del supuesto de una muestra realizada a 12 parejas, para estimar su actitud frente al manejo de sus hijos en los problemas escolares, (un menor puntaje corresponde a una actitud menos favorable):

PAREJA:	1	2	3	4	5	6	7	8	9	10	11	12
Esposa:	35	46	29	32	40	43	36	30	56	35	48	29
Esposo:	29	32	36	35	46	45	35	18	41	37	40	25

¿Al nivel del 5% existe una misma actitud, en las parejas, frente a los problemas escolares?

Solución:

PAREJA	ESPOSA	ESPOSO	d_i	RANGO d_i	SIGNO MENOS FRECUENTE	No.	d_i ORDENADO	RANGO
1	35	29	6	6,5		1	1	1
2	46	32	14	11,0		2	2	2,5
3	29	36	–7	–8,0	–8	3	2	2,5
4	32	35	–3	–4,0	–4,0	4	3	4
5	40	46	–6	–6,5	–6,5	5	4	5
6	43	45	–2	–2,5	–2,5	6	6	6,5
7	36	35	1	1,0		7	6	6,5
8	30	18	12	10,0		8	7	8
9	56	41	15	12,0		9	8	9
10	35	37	–2	-2,5	–2,5	10	12	10
11	48	40	8	9		11	14	11
12	29	25	4	5		12	15	12
–	–	–	–	–	T = 23,5	–	–	–

1) H_0: la actitud de las parejas, no difiere
   H_a: la actitud de las parejas, difiere
2) $\propto = 0,05$ $T_{0,05} = 14$ (ver tabla al final del libro) Observamos que $T > T_{0,05}$ $(21 > 14)$

Con una $T_{0,05} = 14$ quiere decir que aceptamos a H_o, a1 nivel del 5%. Concluyendo que la actitud de los padres frente al problema escolar no difiere.

Si se hubiese presentado el caso en que $T \leq T_{0,05}$ estaríamos rechazando a H_0, al nivel del 5%, concluyendo que la actitud de los padres, frente al problema escolar de sus hijos difiere en forma significativa.

Ejemplo 2. Se diseña un estudio de tiempos y movimientos, con el fin docimar la eficiencia de dos procedimientos para montar cierto mecanismo. Se aparean los obreros con base en los coeficientes de inteligencia y el número de años de educación. Los miembros de cada par son asignados en grupos a uno de los dos procedimientos. El tiempo en segundos para terminar el montaje es como sigue:

PAR:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x:	44	40	40	31	44	41	26	26	32	38	37	29	20	40	31	44
y:	41	44	29	27	30	41	28	30	34	23	29	25	20	36	27	44

Solución:

EXPERIMENTO	PROCEDIMIENTOS		d_i	CATEGORÍA d_i		No.	ORDEN	RANGO
	I	II		+	–			
1	44	41	3	3		1	2	1,5
2	40	44	–4		–6,5	2	2	1,5
3	40	29	11	11		3	3	3
4	31	27	4	6,5		4	4	6,5
5	44	30	14	12		5	4	6,5
6	41	41	0	–	–	6	4	6,5

Continuación

7	26	28	–2		–1,5	7	4	6,5
8	26	30	–4		–6,5	8	4	6,5
9	32	34	–2		–1,5	9	4	6,5
10	38	23	15	13		10	8	10
11	37	29	8	10		11	11	11
12	29	25	4	6,5		12	14	12
13	20	20	0	–	–	13	15	13
14	40	36	4	6,5		Nota: Eliminamos las diferencias iguales a cero		
15	31	27	4	6,5				
16	44	44	0	75,0	16,0			

Observemos en las tablas anteriores la más pequeña, en la cual aparece d_i, ordenada de menor a mayor, sin tener en cuenta los signos. La columna RANGO se calculó teniendo en cuenta la columna No., así aparece dos veces el 2, por lo tanto el rango será: $1{,}5 = \frac{1+2}{2} = \frac{}{2}$; luego el 3 (una sola vez), su rango corresponde al número 3; el 4 aparece 6 veces, por lo tanto su rango es $\frac{+}{} = \frac{1}{} = 6{,}5$; el 10 será 10 y así sucesivamente.

Sabemos que la prueba es bilateral, que $n = 13$ $\alpha = 0{,}05$ y $T = 16$ por ser la suma de los valores negativos, siendo la más pequeña de las dos columnas.

En la Tabla No. XVIII encontramos que el *valor teórico* es de $T_{0,05} = 17$.

Si $T \leq T_{0,05}$ rechazamos => H_0: la diferencia no es significativa; no difiere
Si $T > T_{0,05}$ aceptamos => H_a: difiere; la diferencia es significativa

En el ejercicio que estamos desarrollando encontramos que $16 < 17$; es decir que $T \leq T_{0,05}$ por lo tanto aceptamos H_a: la *diferencia es significativa*, es decir que difiere la eficiencia en los dos procesos. Si nos devolvemos a observar la prueba del signo, el resultado es al contrario, es decir, que no existe una diferencia significativa (se acepta H_0) entre los dos procesos, al nivel del 5%.

MUESTRA GRANDE

En este caso, utilizamos el ejercicio No. 91 que aparece en la página 457, cuando se trabaja con muestras grandes ($n > 25$)

Ejemplo 3. Los datos que se dan a continuación corresponden al número de piezas defectuosas producidas por dos máquinas durante el mes de septiembre de 2011. Probar al nivel del 5%, que no hay una diferencia significativa entre los dos procesos.

NÚMEROS DEFECTUOSOS				RANGOS		ORDEN		
Días	A	B	$d_i = x_i - y_i$	+	–	No.	d_i	Rango
1	8	9	–1		–5,5	1	1	5,5
2	5	4	1	5,5		2	1	5,5
3	10	12	–2		–16,0	3	1	5,5
4	7	9	–2		–16,0	4	1	5,5
5	9	10	–1		5,5	5	1	5,5
6	11	11	0	–	–	6	1	5,5
7	3	6	–3		–24,0	7	1	5,5
8	6	7	–1		–5,5	8	1	5,5
9	6	8	–2		–16,0	9	1	5,5
10	8	10	–2		–16,0	10	1	5,5
11	5	6	–1		–5,5	11	2	16
12	6	9	–3		–24,0	12	2	16
13	9	11	–2			13	2	16
14	3	6	–3		–24,0	14	2	16
15	7	7	0	–	–	15	2	16
16	10	8	2	16,0		16	2	16
17	9	12	–3		–24,0	17	2	16
18	13	13	0	–	–	18	2	16
19	8	9	–1		–5,5	19	2	16
20	4	6	–2		–16,0	20	2	16
21	2	3	–1		–5,5	21	2	16
22	8	9	–1		–5,5	22	3	24
23	8	10	–2		–16,0	23	3	24
24	6	9	–3		–24,0	24	3	24
25	5	7	–2		–16,0	25	3	24
26	10	11	–1		–5,5	26	3	24
27	9	9	0	–	–			
28	4	6	–2		–16,0			
29	7	6	1	5,5				
30	3	5	–2		–16,0			
				T = 27				

Nota: Elminamos las diferencias iguales a cero, siendo $n = 26$

Observemos que el valor de $T = 27$. Además, como $n > 25$, la distribución muestral de T es aproximadamente normal (recordemos que algunos la consideran así cuando $n > 50$), procedemos a calcular la *media y la varianza* de T, cuando $n = 26$, de la siguiente manera:

$$M_{[T]} = \overline{T} = \frac{n\,(n+1)}{} = \frac{26(27)}{} = 175{,}5$$

$$V_{[T]} = \frac{(2n+1)\,M_{[T]}}{6} = \frac{n\,(n+1)(2n+1)}{2}$$ Reemplazando, se tendrá el mismo resultado

$$V_{[T]} = S^2_{[T]} = \frac{(52+1)\,(175{,}5)}{6} = \frac{26(27)(52+1)}{2} = 1\,550{,}25$$

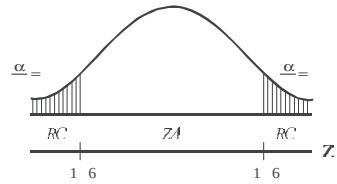
$$S_{[T]} = \sqrt{S^2_T} = \sqrt{1\,550{,}25} = \ 9{,}17$$

En consecuencia la variante estadística Z, está dada por la siguiente fórmula:

$$Z = \frac{T - M_{[T]}}{S_{[T]}} = \frac{27 - 175{,}5}{8{,}62} = - \ ,85$$

Es necesario observar que T = 27, la categorización la hacemos con valor absoluto, procediendo a elegir la suma más pequeña, sin tener en cuenta el signo, en este caso corresponde a la columna con signos positivos.

1) H_0: la diferencia no es significativa
   H_0: la diferencia es significativa
2) $\alpha = 0{,}05$
3) $Z = -3{,}85$ cae dentro de la región crítica, es decir, rechazamos H_0 y aceptamos H_a, la diferencia entre los dos procesos se considera significativa, al nivel del 5%, con un resultado igual al obtenido cuando aplicamos la Prueba del signo (ver página 451).



PRUEBA U DE MANN-WHITNEY

Es la prueba más aplicada, dentro de las distribuciones *no paramétricas*, debido a que toma la mayor información que poseen los datos, haciéndola mucho más eficiente que las otras pruebas, sobretodo cuando se seleccionan *dos conjuntos aleatorios e independientes*, en tal forma que se puedan ordenar por rangos. Tiene un proceso de cálculo muy parecido al utilizado en la *Prueba del signo*, siendo usada para probar si dos grupos muestrales proceden de una misma población.

El proceso de cálculo se hace engorroso, cuando los tamaños muestrales n_1 y n_2 son grandes, pues se trabaja con dos grupos independientes.

MUESTRAS GRANDES

El cálculo o desarrollo de esta prueba, requiere de los siguientes pasos:

- ❑ Se ordenan todos los datos, como si fuera una sola muestra, de menor a mayor.
- ❑ Se establece el rango, es decir, se categoriza.
- ❑ Se coloca el valor del rango al frente de cada observación, manteniendo la presentación original dada para las dos muestras.
- ❑ Se suma la columna del rango en forma independiente para cada muestra, obteniéndose R_1 y R_2 respectivamente.
- ❑ Se aplica la fórmula correspondiente a la prueba U, siendo:

$$(1)\ U = n_1\, n_2 + \frac{n_1\,(n_1+1)}{2} - R_1$$

$$(2)\ U = n_1\, n_2 + \frac{n_2\,(n_2+1)}{2} - R_2$$

La mayoría de los investigadores aplican la *Prueba* U, cuando n_1 o n_2 es mayor a 20, admitiendo de esta manera que la distribución tiene un comportamiento similar a la *normal*, siendo utilizada en estos casos la *variante estadística* Z, igual a:

$$Z = \frac{U - (n_1 - n_2/2)}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Ejemplo: 1. Supongamos dos muestras grandes e independientes (n_1 =23 y n_2 =12) correspondientes a dos pruebas cuyas puntuaciones fueron las siguientes:

PRIMERA PRUEBA x_i	SEGUNDA PRUEBA y_i	RANGO x_i	RANGO y_i
10	5	21,5	5,0
4	13	2,0	28,0
9	13	19,0	28,0
10	8	21,5	16,5
11	10	24,0	21,5
14	12	31,0	25,5
13	16	28,0	34,0
4	15	2,0	33,0
6	8	8,5	16,5
5	6	5,0	8,5
6	14	8,5	31,0
7	18	13,0	35,0
6	n_2=12	8,5	ΣR_2=282,5
7		13	
8		16,5	
14		31,0	
6		8,5	
4		2,0	
8		16,5	
6		8,5	
12		25,5	
10		21,5	
7		13,0	
n_1=23		ΣR_1 = 8	

CATEGORIZACIÓN					
No.	Orden	Rango	No.	Orden	Rango
1	4	2	19	9	19,0
2	4	2	20	10	21,5
3	4	2	21	10	21,5
4	5	4	22	10	21,5
5	5	5	23	10	21,5
6	6	8,5	24	11	24,0
7	6	8,5	25	12	25,5
8	6	8,5	26	12	25,5
9	6	8,5	27	13	28,0
10	6	8,5	28	13	28,0
11	6	8,5	29	13	28,0
12	7	13,0	30	14	31,0
13	7	13,0	31	14	31,0
14	7	13,0	32	14	31,0
15	8	16,5	33	15	33,0
16	8	16,5	34	16	34,0
17	8	16,5	35	18	35,0
18	8	16,5			

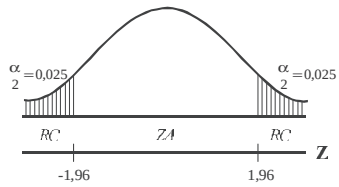
Como $\Sigma R_2 < \Sigma R_1$ (282,5 <348) operamos con la segunda fórmula de U así:

$U = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - R_2$ Reemplazando se tiene:

$U =$ 1 $+ \frac{1 \quad 1}{\quad} -$ $=$ 6 + $-$ $=$ 1

1) H_o: Todas las observaciones provienen de la misma población.
   H_a: Provienen de poblaciones diferentes.
2) $\propto = 0{,}05$
3) U = 71,5

Al nivel del 5% estas observaciones no provienen de una misma población.



MUESTRAS PEQUEÑAS

Cuando las dos muestras (n_1 y n_2) son menores de 8, (muy pequeñas), no se debe aplicar el método para *muestras pequeñas* que se explica a continuación, siendo necesario utilizar procedimientos y tablas diferentes a las anteriores.

Ahora, si una de las dos muestras tiene un tamaño mayor o igual a 9, se procede de la siguiente forma:

- Ordenamos todos los datos, de menor a mayor, sin tener en cuenta a qué muestra pertenecen.
- Se ordenan por rangos todas las observaciones, como se ha venido haciendo en las pruebas anteriores; por ejemplo 7 y 7, ocupan la posición 5 y 6 respectivamente, por lo tanto el rango será de 5,5.
- Se obtienen dos columnas para los rangos R_1 y R_2
- Se procede a la suma de cada columna de los rangos, con el fin de obtener las ΣR_1 y ΣR_2
- La suma de las dos columnas de rangos debe ser igual a: $\Sigma R_1 + \Sigma R_2 = \frac{n(n+1)}{2}$
- La medida U que se debe aplicar puede ser:

$$U = n_1 + n_2 + \frac{n_1 (n_1+1)}{2} - R_1 \qquad U = n_1 + n_2 + \frac{n_2 (n_2+1)}{2} - R_2$$

- Siempre se toma la fórmula que tenga la ΣR más pequeña.
- Finalmente utilizamos la Tabla XIV, localizando en ella a n_1 y n_2, para un determinado nivel de significación, teniendo en cuenta si la prueba es unilateral o bilateral.

Ejemplo: 2. Consideremos dos muestras pequeñas, donde $n_1 = 8$ y $n_2 = 11$, para observar la aplicación de la prueba U de Mann - Whitney, para muestras pequeñas, a fin de probar que ambas proceden de la misma población, al nivel del 5%.

Solución:

No.	PRUEBAS		RANGO	RANGO	CATEGORIZACIÓN					
	I	II	I	II	No.	Orden	Rango	No.	Orden	Rango
1	18	28	7,5	15,0	1	13	1,0	11	23	11,5
2	24	32	13,0	18,0	2	14	2,0	12	23	11,5
3	22	28	9,5	15,0	3	16	3,5	13	24	13,0
4	23	32	11,5	18,0	4	16	3,5	14	28	15,0
5	17	16	5,5	3,5	5	17	5,5	15	28	15,0
6	16	18	3,5	7,5	6	17	5,5	16	28	15,0

Continúa...

Continuación

7	14	17	2,0	5,5	7	18	7,5	17	32	18,0
8	20	13	9,5	1,0	8	18	7,5	18	32	18,0
9		23		11,5	9	22	9,5	19	32	18,0
10		28		15,0	10	22	9,5			
11		32		18,0						
	n_1=8	n_2=11	$\Sigma R_1 = 62$	$\Sigma R_2 = 1\ 2$						

Elegimos la sumatoria más pequeña, en este caso es $\Sigma R_1 = 62$, que nos permite calcular el valor de U, tal como se desarrolla a continuación:

$$U = n_1\, n_2 + \frac{n_1\,(n_1+1)}{2} - R_1 = 88 + \ 6 - 62 = 62$$

Observemos, que al utilizar la fórmula para R_2, el resultado es diferente, siendo:

$$U = n_1\, n \ + \frac{n \quad n\ +1}{\qquad} - R \ = \quad +66\ -1 \quad =$$

Calculadas las dos U, una para cada R, se elige la menor, en nuestro caso será U = 22

Utilicemos la Tabla XIV, localizando los valores para n_1 = 8 y n_2 =11, con un nivel de α = 0,05 para una *prueba bilateral*, obteniendo los valores críticos de U=19. El valor será significativo cuando el resultado aplicando la fórmula U, sea menor o igual al valor obtenido en la Tabla, en caso contrario no será significativo, como en este caso donde 22>19, por lo tanto, se acepta la hipótesis nula H_o , es decir, aceptamos que estas dos muestras provienen de una misma población, al nivel del 5%.

MUESTRAS MUY PEQUEÑAS

Hasta el momento, en cuanto se refiere a la Prueba U de Mann - Whitney, hemos visto dos casos:

- Cuando una de las n_1 y n_2 son mayores de 20, se consideran como muestras grandes y se aplica la prueba de normalidad Z.
- Cuando una de las n_1 y n_2 es igual o mayor que 9, consideradas como muestras pequeñas, utilizamos la Tabla XIV.

Ahora se tiene un tercer caso, cuando ninguna de las n_1 y n_2 son mayores a 8, consideradas a la vez como muestras muy pequeñas, para ello se utiliza la Tabla XIII y el procedimiento a seguir se explica mediante el desarrollo de un ejercicio tomado como modelo.

Ejemplo 3. Supongamos la realización de dos experimentos, para los cuales se obtuvieron 4 y 5 controles, respectivamente con los siguientes resultados o puntajes.

Puntaje A:	8	9	11	16	
Puntaje B:	7	10	12	6	13

Clasificamos los puntajes de menor a mayor, identificándolos de acuerdo a su respectiva muestra, A y B.

6	7	8	9	10	11	12	13	16
B	B	A	A	B	A	B	B	A

Ahora, contamos el número de veces que A precede a B. Observemos que la primera A correspondiente a la calificación 8, precede o antecede a tres veces B, (calificaciones 10, 12 y 13), luego la segunda A (calificación 9) precede también a las tres de las calificaciones (10, 12 y 13); y la tercera A (calificación 11) precede a dos B (calificaciones 12 y 13) y la última A no precede a ninguna calificación es decir 0.

De acuerdo a lo anterior se tiene que U = 3+3+2+0=8. Veamos qué sucede, si hubiéramos hecho lo contrario, donde B precede a A, habríamos encontrado la siguiente situación:

U' = 4+4+2+2+1+1 = 14, a esta suma, para diferenciarla de la anterior la simbolizamos por U'. Según la hipótesis nula planteada U y U' deben ser iguales, en caso contrario, de presentar una diferencia, procedemos a establecer si ella es significativa, al nivel de significación que se esté probando o docimando.

La hipótesis nula H_0 y la alternativa H_a se plantean de la siguiente manera:

H_0: Las distribuciones de frecuencias relativas de las poblaciones A y B son idénticas.

H_a: (*Prueba Bilateral*): La distribución de frecuencias relativas de una población está desplazada con relación a la distribución de la otra población.

H_a: (*Prueba Unilateral*): La distribución de frecuencias relativas de la población A está desplazada a la derecha de la distribución de frecuencias de la población B.

Procedemos a consultar la Tabla XIII que consta de 6 subtotales separados, una por cada valor correspondiente a n_2, que va desde n_2 =3 hasta n_2 =8, dado que n_1 siempre representará el grupo que tenga el menor número de observaciones, en este caso será n_1 = 4.

En el ejemplo conocemos: n_2 = 5, n_1 = 4 y U = 8, con lo cual consultamos en la Tabla (ver el tercer cuadro de la Tabla XIII), siendo el valor de p = 0,365 valor sombreado, es decir, correspondiente a la probabilidad de que U sea menor o igual a 8 cuando la *prueba es unilateral*; si la *prueba es bilateral*, se tomaría el doble de lo aparecido en la tabla, siendo el valor de p = 0,730.

PRUEBA H DE KRUSKAL Y WALLIS

En los capítulos correspondientes a distribuciones *paramétricas* consideramos el *análisis de varianza* (ver páginas 412 al 414), utilizada para probar si las medias poblacionales eran iguales bajo el supuesto de normalidad, con desviaciones estándar iguales para esas poblaciones.

El anterior procedimiento no es válido, ni recomendable, cuando se trata de *distribuciones no paramétricas*. Para estos casos, en el año 1952 *W.H. Kruskal* y *W.A. Wallis*, elaboraron una prueba que permite trabajar con rangos, sin exigencia en cuanto a los parámetros de la población. Actualmente se conoce como Prueba de *Kruskal - Wallis*, prueba H ó *Análisis de Varianza por Rangos*.

El procedimiento a seguir para el cálculo del valor para H, debe:

- ❑ Ser requisito indispensable que las poblaciones sean independientes.
- ❑ Combinar todos los valores de las muestras, como si fuera una sola.
- ❑ Ordenar de menor a mayor, es decir, categorizamos.
- ❑ Establecer los rangos, tal como se ha venido haciendo en las pruebas anteriores.
- ❑ Sumar los rangos para cada muestra.
- ❑ Aplicar la siguiente fórmula, para el cálculo de H:

$$H = \frac{1}{n\ n+1}\ \frac{R_1}{n_1} + \frac{R}{n} + \quad - \quad n+1$$

Es necesario que cada muestra tenga por lo menos 5 observaciones (es decir, que cada tamaño muestral sea mayor o igual a 5), de esta manera su distribución será muy cercana a la de *Chi cuadrado*, con K - 1 grados de libertad, donde K corresponderá al número de muestras tomadas para la prueba.

Ejemplo 1. Supongamos que el Presidente de una Corporación Financiera, está interesado en estudiar la rapidez de operación que se presenta, desde el momento en que se deposita y se retira el dinero de las cuentas. Para ello elabora un índice de permanencia del dinero, estableciendo un índice del 350 para una cuenta considerada extremadamente activa en el semestre; pero si sólo se hacen 10 retiros, en el mismo período, el índice es de 80. De acuerdo a la anterior escala (de 80 a 350), los índices obtenidos en cuatro (4) muestras, en igual número de sucursales, fueron:

SUCURSAL	ÍNDICES								
Caracas	120	116	80	206	305	206	180	150	
Chapinero	206	180	300	180	305				
Centro	100	80	206	305	310	300			
Niza	206	150	210	206	310	120	116	120	340

Utilizando un nivel de 5%, aplicar la prueba de Kruskal - Wallis, a fin de probar si hay una diferencia significativa en cuanto a los índices de movimiento de cuentas de ahorros personales, entre las cuatro (4) sucursales de esta corporación financiera.

Solución:

Lo *primero* que hacemos, es categorizar las observaciones como si fuera una sola muestra, obteniéndose el rango para cada uno de ellos, así:

No.:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ÍNDICE:	80	80	100	116	116	120	120	120	150	150	180	180	180	180
RANGO:	1,5	1,5	3,0	4,5	4,5	7,0	7,0	7,0	9,5	9,5	12,5	12,5	12,5	12,5

No.	15	16	17	18	19	20	21	22	23	24	25	26	27	28
ÍNDICE:	206	206	206	206	206	206	300	300	305	305	305	310	310	340
RANGO:	17,5	17,5	17,5	17,5	17,5	17,5	21,5	21,5	24	24	24	26,5	26,5	28

En *segundo* lugar, procedemos en cada muestra a colocar el rango a cada una de las observaciones, de esta manera se puede establecer la sumatoria de los rangos, en la respectiva muestra.

CARACAS	
Índice	*Rango*
120	7,0
116	,5
80	1,5
206	17,5
05	2 ,0
206	17,5
180	12,5
150	9,5
	$\Sigma R_1 = 9\ ,0$
	$n_1 = 8$

CHAPINERO	
Índice	*Rango*
6	1
1	1
	1
1	1
	ΣR =
	n =

CENTRO	
Índice	*Rango*
100	,0
80	1,5
206	17,5
05	2 ,0
10	26,5
00	21,5
	$\Sigma R = 9$
	$n = 6$

NIZA	
Índice	*Rango*
206	17,5
150	8,5
180	12,5
206	17,5
10	26,5
120	7,0
116	,5
120	7,0
0	28,0
	$\Sigma R = 1\ 0$
	$n = 9$

Reemplazamos los anteriores resultados, en la fórmula para calcular el valor de *H*, siendo:

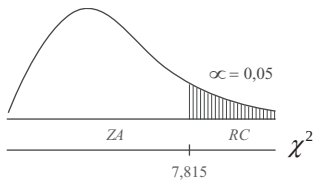
$n = n_1 + n_2 + n\ + n\ = 8 + 5 + 6 + 9 = 28$

$$H = \frac{12}{28(28+1)}\left[\frac{(9\ ,0)^2}{8} + \frac{(88,0)^2}{5} + \frac{(9\ ,0)^2}{6} + \frac{(1\ 0,0)^2}{9}\right] -\ (28+1)$$

$H = 0,01\ \ 778\ \ [1\ 10\ ,5 + 1\ 5\ \ 8,8 + 1\ \ \ 72,67 + 1\ 877,78] - 87$

$H = 0,01\ \ 778\ \ [6\ 00\ ,75] - 87 = 88,7\ \ -87 = 1,7$

Determinamos los grados de libertad:
$U = K - 1 = 4 - 1 = 3$. Recordemos que K corresponde al número de muestras, en este caso son 4.



Si el nivel de significación es α = 0,05, se tendrá en la Tabla para *Chi Cuadrado* (χ^2) un valor igual a 7,815.

La regla de decisión consiste en aceptar Ho, cuando $H < \chi^2_{0,05}$, tal como sucede en este caso donde 1,73 < 7,815.

H_0: No hay una diferencia significativa en cuanto a los índices de movimiento de cuentas de ahorro, entre las cuatro sucursales, al nivel del 5%.

H_a: Cuya conclusión sería, cuando $H > \chi^2_{0,05}$ las distribuciones de indicadores no son iguales, equivale a aceptar que existen diferencias significativas en cuanto a las puntuaciones.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

103. Consideremos que el ICFES realiza pruebas para el ingreso a la Universidad donde el puntaje máximo que puede obtener un egresado del bachillerato, es de 400 puntos. Una universidad dentro de sus aspirantes selecciona una muestra de 8 hombres y otra de 14 mujeres, con el fin de probar si existe una diferencia en su preparación de acuerdo al sexo. Probar el nivel del 5% que no hay diferencia.

HOMBRES:	310	280	320	180	260	280	310	280						
MUJERES:	180	280	260	340	370	320	280	250	240	310	260	340	300	240

104. En una fábrica se establecen dos procedimientos de entrenamiento para sus trabajadores. Se midió el tiempo (minutos) necesario para la realización de cierta operación, con los siguientes resultados:

PROCEDIMIENTO A:	18	25	10	14	17	22	16			
PROCEDIMIENTO B:	10	27	18	18	20	10	17	20	22	17

Pruebe que los dos procedimientos son igualmente efectivos, al nivel del 5%, utilizando la prueba de Mann-Whitney.

105. En una granja avícola, el administrador encuentra que en el mercado se ofrecen cuatro (4) productos que permiten alimentar sus animales, todos ellos de excelentes resultados en el aumento de peso (en onzas), utilizado durante un período determinado.

Alimento A:	12,8	13,4	16,0	14,2	12,8			
Alimento B:	16,1	14,2	13,0	13,4	12,8	16,0	12,6	12,1
Alimento C:		14,3	14,2	12,8	16,3	16,0	17,3	
Alimento D:		10,8	11,6	14,2	13,0	12,8		

Probar al nivel del 5%, utilizando la prueba H de Kruskal-Wallis, si los pesos promedios son iguales para los animales, sin importar la clase de alimento.

106. Mediante la aplicación de una escala de 0 a 30, en dos muestras de 18 trabajadores cada una, establecer el rendimiento en dos procesos de fabricación, con los siguientes resultados.

Proceso A:	16,3	16,1	18,3	19,2	18,7	18,9	16,1	18,7	17,6	19,5	16,1	19,2	18,3
Proceso B:	16,3	19,1	14,8	18,2	18,6	16,6	19,2	18,2	17,5	18,9	18,6	18,2	17,8

Proceso A:	16,3	16,1	18,3	19,2	18,7	18,9	16,1	18,7	17,6	19,5	16,1	19,2	18,3
Proceso B:	16,3	19,1	14,8	18,2	18,6	16,6	19,2	18,2	17,5	18,9	18,6	18,2	17,8

Existe evidencia, al nivel del 10%, que nos indique alguna diferencia entre las dos poblaciones, al ser sometidas a diferentes procesos. (a) Utilice la prueba U; (b) Considere que es aplicable la prueba del signo (c) Realice la prueba de rangos T de Wilcoxon, bajo el supuesto de que sea la prueba más indicada.

107. Con el fin de aumentar la productividad de la empresa, un industrial desea cambiar el método de entrenamiento con el que ha venido capacitando a sus trabajadores. Seleccionó una muestra de 20 trabajadores, determinó el número de unidades producidas con el método anterior y luego, un tiempo después, con el nuevo método, cuyos resultados se presentan a continuación:

TRABAJADOR:	1	2	3	4	5	6	7	8	9	10	11	12	13
Método A:	36	28	20	26	24	25	27	20	26	20	24	30	24
Método B:	40	30	20	24	20	30	27	20	28	25	24	24	28

TRABAJADOR:	14	15	16	17	18	19	20
Método A:	20	28	26	30	20	26	28
Método B:	19	30	30	30	24	24	24

Utilice primero la Prueba del signo y luego la Prueba de Wilcoxon, al nivel del 5%, a fin de demostrar que el nuevo método de entrenamiento arrojó mejores resultados que el método tradicional.

108. El hospital San Juan de Dios de la ciudad, tiene por costumbre contratar personal auxiliar de enfermería, únicamente en dos centros de formación de los que existen en el país, para ello prepara un formulario de evaluación para ser aplicado a sus aspirantes, a fin de examinar el grado de formación o conocimiento, aplicando una escala de 0 a 10, con la cual se obtuvieron los siguientes resultados:

CENTRO A:	86	94	65	72	83	94	90	72	62		
CENTRO B:	62	46	42	83	75	92	94	94	72	86	62

¿Al nivel del 5%, utilizando la prueba de U de Mann-Whitney, se podrá concluir que los dos centros ofrecen la misma calidad académica?

109. Varios accidentes automovilísticos menores ocurrieron en cruces de alto riesgo en la ciudad, a pesar de contar con semáforos. El departamento de tránsito dispuso la ubicación de agentes de tránsito, además de los semáforos, a fin de reducir los accidentes. Se eligieron al azar 14 cruces, y el número de accidentes durante el mes antes y después de la modificación fueron los siguientes:

CRUCES:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Antes de la Modificación	5	6	8	7	10	6	4	10	12	6	7	11	15	6
Después de la Modificación	3	2	10	7	4	6	4	6	5	10	2	5	3	8

¿Las modificaciones redujeron el número de accidentes en los cruces de alto riesgo? (nivel 5%0).

a. Resolverlo mediante la aplicación de la Prueba del signo.
b. Suponiendo que debe ser resuelto, aplicando la "***t***" de Student en observaciones apareadas (poblaciones dependientes).

110. Un analista financiero desea saber si ha habido o no un cambio significativo en las utilidades por acción de un período a otro entre las empresas más grandes del país. Una muestra de 12 empresas seleccionadas entre 150 arrojó los siguientes resultados:

EMPRESAS:	1	2	3	4	5	6	7	8	9	10	11	12
Año I:	4,12	12,85	2,03	3,39	4,10	-2,25	2,44	5,01	1,85	1,95	3,25	3,50
Año II:	4,79	13,20	-2,86	1,94	4,30	-2,50	3,80	5,20	2,15	3,07	2,80	3,50

Utilizando un nivel de significación del 5%, ¿Se puede afirmar que se ha presentado un cambio positivo en cuanto a las utilidades por acción entre los dos períodos?

111. Dentro de las informaciones que se dan sobre la preferencia por marcas de los automóviles para taxi que más se venden en el país, se dice que el 36% por Renault, 30% por Mazda, 20% por Chevrolet, 12% por Hyundai y 2% otras marcas. Una muestra reciente sobre las ventas de automóviles para taxi, reveló que 250 fueron Renault, 186 Mazda, 124 Chevrolet, 50 Hyundai y 6 otras marcas. Probar al nivel del 5% la hipótesis de que no ha habido cambio en las preferencias de compra según la marca.

112. ¿La prueba del signo corresponde a una distribución no paramétrica, sus observaciones se aparean y es apropiada para experimentos «antes y después»?

113. ¿Cuál es la finalidad cuando se realiza la prueba al coeficiente de correlación?

114. Un funcionario de una entidad que presta asesoría a pequeños establecimientos industriales ha realizado un estudio acerca de las diferencias entre la habilidad de los trabajadores para aprender el nuevo trabajo y la capacidad para efectuarlo sin supervisión. Después de varias semanas de observación, se ha obtenido la siguiente información, donde Y= producción mensual de unidades correctamente producidos y X = años de educación de cada uno de los 10 trabajadores.

X_i:	7	5	6	4	10	8	6	5	10	6	8
Y_i:	15	7	10	8	20	16	12	8	16	8	14

a) Calcule el Coeficiente de correlación lineal; b) ¿Puede rechazarse la hipótesis de que no existe Correlación lineal a un nivel de significación del 5%?

115. Trece estudiantes universitarios de sexo femenino y 13 masculino son clasificados por pares, según su edad y tiempo de experiencia en la conducción de automóviles. Se les aplica una prueba de conducción con los siguientes resultados.

PAR:	1	2	3	4	5	6	7	8	9	10	11	12	13
Mujer:	93	95	88	82	80	70	65	90	90	83	72	64	76
Hombre:	95	90	80	82	86	90	70	86	86	95	82	70	80

a) ¿Al nivel de 1% de significación, se puede decir que los hombres obtienen un mayor puntaje que las mujeres en la prueba de conducción?
b) Con la información que se tiene, establezca límites de confianza del 95% para la diferencia.

116. Un Editor está decidiendo cuál de los textos de estadística que se le han propuesto va a aceptar para su publicación. El uno tiene orientación técnica, el otro lo enfoca hacia la aplicación práctica. El Editor consulta a 50 instituciones de enseñanza superior para determinar en los tres tipos de instituciones sus preferencias. Los resultados fueron:

TIPO DE ENSEÑANZA	**TEXTO DE PREFERENCIA**	
	Orientación técnica	**Orientación práctica**
Intermedia	1	11
Tecnológica	9	11
Profesional	11	7

¿Al nivel del 5% se podrá concluir que las preferencias son idénticas en los tres tipos de institución?

117. La desviación típica del tiempo de duración de 120 pilas para radio transistor es de 150 horas. Establecer los límites de confianza del a) 95% y b) del 99% para la desviación típica poblacional de donde se extrajo la muestra.

118. En 400 lanzamientos de una moneda se observaron 230 caras y 170 sellos. Aplicando la prueba de Chi - Cuadrado, comprobar la hipótesis de que la moneda es correcta, al nivel de significación del 5%.

119. Una muestra realizada a 352 estudiantes correspondientes a dos cursos donde el mismo profesor desarrolla la asignatura de matemáticas, con el fin de determinar el número de alumnos que aprobaron y perdieron la asignatura, los resultados fueron:

CURSOS	APROBACIÓN	
	Si	No
A	144	34
B	128	46

Al nivel del 5% probar que no hay diferencia en cuanto a la aprobación de la materia.

120. Un coeficiente de Correlación obtenido mediante una muestra de tamaño 22 fue de 0,46 ¿Al nivel del 5%, se puede afirmar que el Coeficiente de correlación de la población de donde se obtuvo la muestra difiere de cero?

121. Un Coeficiente de correlación entre 52 familias en cuanto a los niveles de ingresos y gastos fue de 0,78 ¿Puede afirmarse al nivel del 5% que hay Correlación entre las variables?

122. Una muestra de 26 observaciones da como resultado un Coeficiente de correlación de 0,56 y se desea saber si el Coeficiente es significativo, al nivel del 1%.

123. ¿Qué significa el hecho de que una prueba al Coeficiente de correlación no sea significativa?

124. En una fábrica se desea implantar un programa de seguridad que permita disminuir el número de accidentes de trabajo (incluyendo aquéllos de poca importancia), se desarrolla una investigación por la Oficina de seguridad, durante una semana, con el fin de probar que los días de la semana no presentan diferencia significativa en la ocurrencia, al nivel del 5%.

Los resultados fueron:

DÍAS	Lunes	Martes	Miércoles	Jueves	Viernes	Total
Número de Accidentes	10	3	2	5	14	34

125. ¿Cuál es la finalidad de la aplicación de Chi - Cuadrado?

126. ¿Qué se espera concluir al realizar una prueba de independencia?

127. ¿Qué diferencia hay entre una prueba de independencia y una de homogeneidad?

128. ¿En qué consiste la prueba del signo? ¿En qué casos se emplea?

Síntesis de capítulo

Recordemos que la *prueba estadística* sólo es el procedimiento experimental que se sigue cuando se quiere saber si una hipótesis estadística es cierta.

Hay cientos de procedimientos de *pruebas estadísticas* a disposición del usuario, que abarcan una enorme variedad de aplicaciones. En el capítulo anterior se describieron las más importantes, las que generalmente aparecen en todos los libros de consulta sobre estadística, siendo las pruebas de: a) *una sola media muestral*; b) *de dos medias muestrales*; c) *de una proporción* y d) *de dos proporciones*. En todas ellas se hicieron los cálculos correspondientes, además, se establecieron los *límites de confianza*.

En este capítulo no se quiere dejar por fuera otras pruebas, no menos importantes, las que se presentan en forma breve, de tal manera que este libro cumpla con las exigencias de la mayoría de los programas académicos. Estas *pruebas* son:

Observaciones apareadas-Poblaciones dependientes. Resulta que en muchos casos el investigador puede preferir la realización de experimentos con una misma muestra, con el fin de comprobar si se producen cambios de actitud en dos situaciones distintas, que hemos llamado antes y después. En su aplicación se hace necesario tomar las precauciones correspondientes a fin de evitar la presencia de variables extrañas, que pudieran interferir en el experimento y con ello, se llegaría a conclusiones erróneas.

Prueba del signo. Dentro de las *pruebas no paramétricas* se le considera la más sencilla, razón por la cual se utiliza erróneamente. Como en la prueba anterior, se toman parejas de observaciones como lo indica su nombre, basándose en el signo de la diferencia para cada par. Su uso generalmente conduce a cometer *el error el tipo* II, es decir, aceptar algo falso.

Lo distintivo en una *prueba no paramétrica*, es que en ella no es necesario tener ningún conocimiento acerca del parámetro o valor estadístico de la población; por otra parte, se requieren más observaciones que las *paramétricas* a fin de reducir el error, que en la mayoría como ya se dijo es de *tipo* II.

Chi - Cuadrado o Ji - Cuadrado. También pertenece a las *pruebas no paramétricas*. Generalmente se utiliza en la realización de *pruebas de independencia*, cuando se toma una sola muestra que proviene de una sola población, con el fin de establecer si hay alguna relación en cuanto a los criterios de clasificación; y *la prueba de homogeneidad*, cuando se trabaja con dos o más muestras provenientes de dos o más poblaciones y se desea probar que provienen de una sola población por su supuesta homogeneidad. Una *tabla de contingencia* es una matriz, donde los elementos (frecuencias) son clasificados en columnas y filas. Si sólo se tienen dos columnas y dos filas nos referimos a una tabla de contingencia de «2 x 2». Si se trata de 4 columnas y dos filas será una tabla de «4 x 2».

Prueba del Coeficiente de correlación. Es la interpretación del grado de asociación o relación que puede haber entre dos variables; se obtiene con la aplicación del Coeficiente de correlación al cuadrado o Coeficiente de determinación R^2.

Generalmente se utiliza el C*oeficiente de correlación* ***r*** para realizar la prueba de significación. Una hipótesis referente a un valor del coeficiente de correlación en la población, ρ (la letra griega rho o ro), se prueba calculando la probabilidad de que un valor de ***r*** difiera de ρ en tanto como el valor de la muestra que pudiera surgir como resultado de sólo la variación del muestreo al azar. En este libro con fines de simplificación en la simbología hemos reemplazado a rho por **R**.

Las *pruebas de hipótesis con respecto a la varianza.* En muchos casos, este método tiene más importancia que la prueba aplicada a un promedio o a una proporción. Frecuentemente se dice que a un fabricante le resulta más adecuado conocer la variabilidad que presenta su producto, que el promedio del mismo. En este *proceso de prueba* se requiere el uso de la distribución *Chi - Cuadrado*.

Correlación de los rangos. Mide la intensidad de correlación entre dos conjuntos de ordenamientos y el grado de correspondencia entre ellos. Hay dos coeficientes de correlación de los rangos: el de Kendall (1938) y el de Spearman (1904).

La prueba T de Wilcoxon. Denominada también prueba a libre distribución para la homogeneidad de dos muestras, se basa en propiedades ordinales, y fue propuesta por Wilcoxon en 1945, luego desarrollada por otros autores, entre esos Mann-Whitney.

La prueba U de Mann - Whitney. Fueron quienes desarrollaron la teoría con una presentación activa, sobre la versión de la prueba inicialmente presentada por Wilcoxon de ahí que se reconozca a ésta con el nombre de los autores.

Prueba H de Kruskal-Wallis. Es una prueba de gran utilidad para decidir si K, muestras independientes son de poblaciones diferentes. La prueba en sí, determina si la decisión entre las sumas de rangos puede ser tan grande que exista la probabilidad de no provenir de la misma población de donde fueron tomadas las muestras.

10 CAPÍTULO

Números Índices

La rique a de un hombre no se encuentra en la cantidad de dinero que posee sino en la cantidad de su conocimiento y educación.

avier errera

CONTENIDO

- Términos, definiciones, aplicaciones y conceptos generales.
- Índices simples. Encadenamiento. Índices agregativos simples.
- Índices ponderados de precios y de cantidad. Aplicaciones prácticas.
- Índices de valor. Empalme de dos o más series.
- Usos de los números índices (Deflactación, Poder adquisitivo, Salario real, etc.)
- Índices de comercio exterior, Desvalorización y devaluación, etc.)
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Entender, aplicar y manejar correctamente cada uno de los términos y fórmulas para ser utilizados en el campo económico, educativo, social, salud, etc.
- Conocer la importancia que tiene la aplicación de los números índices, no solo como asignatura, sino en el ejercicio profesional.
- Entender y evitar el uso incorrecto, tanto en su elaboración como en la interpretación del mismo.
- Familiarizarse con el uso de indicadores, que frecuentemente aparecen en los medios de comunicación y que generan polémica en su interpretación.

GENERALIDADES

El *n mero índice* es una cifra relativa, expresada en términos porcentuales, que sirve para indicar las variaciones que sufre una variable con respecto a un valor de la misma, la cual es tomada como punto de referencia, denominada base.

Como se puede observar los *ndices* no miden, tan sólo sirven para indicar las variaciones en los precios, cantidades y valores de un período con respecto a otro.

Para algunos, un número índice es una cantidad que muestra por medio de sus variaciones, los cambios temporales o espaciales de una magnitud, que no es en si misma susceptible de medición directa o de observación directa en la práctica.

En una serie corta, el *período base* corresponderá al primer valor de la serie; en una serie larga debe seleccionarse un período que presente la mayor estabilidad, es decir, que se encuentre lo menos afectado por factores internos y/o externos. En algunas ocasiones la base dependerá de lo que se quiera presentar, por ejemplo, si examinamos una serie de precios de un artículo cualquiera, observamos que entre más lejana se encuentra la base, el índice para el último período será más grande.

Los rasgos más importantes en la construcción de un número índice son: su *cobertura*, el *período base* el sistema de *ponderación* y el método de *promediación* de las observaciones.

Según la composición, el índice puede ser: *Simp e o ompuesto*. A su vez los compuestos se clasifican en *agregati os* o de *promedios* y los promedios en: *aritméticos*, *geométricos*, *medianos*, etc., sin embargo los más utilizados en la vida diaria son los primeros.

ÍNDICES SIMPLES

El índice se obtiene dividiendo cada precio, cantidad o valor de un período (anual, mensual, semanal, etc.), por un precio, cantidad o valor de un período fijo, considerado base, multiplicando luego por cien.

$$I_o^t = \frac{X_t}{X_o} 100 \qquad I_o^t = \frac{P_t}{P_o} 100 \qquad I_o^t = \frac{q_t}{q_o} 100$$

X_t = valor, cantidad o precio del período que se investiga.
X_o = valor, cantidad o precio del período tomado como base.

Se puede calcular el índice manteniendo *fijo* o constante el período base, permaneciendo inalterado durante el período que es objeto de análisis. También se puede calcular con *base ariab e*, en este caso, la base es el período inmediatamente anterior, partiendo de un índice igual a 100,00, bajo el supuesto del desconocimiento de información anterior a ese período.

Nota: Es indispensable que usted conozca las diferencias que hay entre índices, porcentajes, progresiones, tasas y razones. Estos términos son explicados en forma amplia en el capítulo del SIL.

Ejemplo 1. Con los datos, de la serie 2006-2011, calcular los índices de *base fija* 2006 y 2011; luego el índice de *base ariab e* para esa misma serie:

AÑOS	y_i	a) (Base Fija) ÍNDICE 2006 = 100	a) (Base Fija) % VARIACIÓN	b) (Base Fija) ÍNDICE 2009 = 100	b) (Base Fija) % VARIACIÓN	c) (Base Variable) Base Variable	c) (Base Variable) % Variación
2006	20	100,00	0	50,00	-50,00	100,00	0
2007	28	140,00	+ 40,00	70,00	-30,00	140,00	+ 40,00
2008	24	120,00	+ 20,00	60,00	-40,00	85,71	-14,29
2009	40	200,00	+ 100,00	100,00	0	166,66	+ 66,66
2010	48	240,00	+ 140,00	120,00	+ 20,00	120,00	+ 20,00
2011	70	350,00	+ 250,00	175,00	+ 75,00	145,83	+ 45,83

En los índices con *base fija*, observemos el último período, mientras que en el primer ejemplo nos indica que hubo un aumento del 250 %, para el 2011 con base en el 2006; en el segundo caso (b) el aumento para el 2011 fue apenas del 75% pero con respecto al 2009.

Para una persona desprevenida, si la variable son precios, puede confundirse con estos resultados, si es un sindicato el que presenta el aumento del 250%, pero si la otra presentación es del gobierno con un aumento en los precios del 75%. Los dos resultados son correctos matemáticamente, pero el mensaje es diferente, dependiendo de la base que ha sido utilizada en cada uno de los indicadores.

En el caso del índice de *base ariab e* las variaciones se establecen con base en el período inmediatamente anterior.

Observemos los procedimientos de cálculo en las dos formas, según se tome la base:

a) Base fija:

$$I_{06}^{06} = \frac{Y_{06}}{Y_{06}} \ 100 = \frac{20}{20} 100 = 100{,}00 \qquad I_{06}^{09} = \frac{Y_{09}}{Y_{06}} \ 100 = \frac{40}{20} 100 = 200{,}00$$

$$I_{06}^{07} = \frac{Y_{07}}{Y_{06}} \ 100 = \frac{28}{20} 100 = 140{,}00 \qquad I_{06}^{10} = \frac{Y_{10}}{Y_{06}} \ 100 = \frac{48}{20} 100 = 240{,}00$$

$$I_{06}^{08} = \frac{Y_{08}}{Y_{06}} \ 100 = \frac{24}{20} 100 = 120{,}00 \qquad I_{06}^{11} = \frac{Y_{11}}{Y_{06}} \ 100 = \frac{70}{20} 100 = 350{,}00$$

Es más fácil de operar, si calculamos una constante $K = \frac{100}{20} = 5$ y la multiplicamos por 20; 28; 24; 40; 48 y 70 obteniendo el mismo resultado, pero en forma más rápida.

b) Base variable:

$$I_{06}^{06} = \frac{20}{20} 100 = 100{,}00 \qquad I_{08}^{09} = \frac{0}{2} \ 100 = 166{,}66$$

$$I_{06}^{07} = \frac{28}{20}\ 100 = 1\ 0{,}0 \qquad I_{09}^{10} = \frac{8}{0}\ 100 = 120{,}00$$

$$I_{07}^{08} = \frac{2}{28}\ 100 = 85{,}71 \qquad I_{10}^{11} = \frac{70}{8}\ 100 = 1\ 5{,}8$$

ENCADENAMIENTO DE ÍNDICES

Cuando se tienen índices de *base ariab e*, éstos se pueden transformar en *base fija*, mediante multiplicaciones sucesivas de los relativos para cada período o cada eslabón.

Ejemplo 2. Supongamos con la información del ejemplo 1, se quiere determinar la variación de *Y* entre los períodos 2008-2011

a) Consideremos la información original, que en la mayoría de los casos no está disponible

$$I_{08}^{11} = \frac{Y_{11}}{Y_{08}}\ 100 \Rightarrow I_{08}^{11} = \frac{70}{2}\ 100 = 291{,}67$$

b) En otros solamente se dispone de información en forma de índices.

$$I_{08}^{11} = \frac{I_{06}^{11}}{I_{06}^{08}} \Rightarrow I_{08}^{11} = \frac{50}{120}100 = 291{,}67$$ (La base de los índices fue 2006)

$$I_{08}^{11} = \frac{I_{09}^{11}}{I_{09}^{08}} \Rightarrow I_{08}^{11} = \frac{175}{60}\ 100 = 291{,}67$$ (La base de los índices fue 2009)

En los dos casos anteriores el resultado fue el mismo. Ahora hagámoslo con los resultados obtenidos con base variable, correspondiente al cuadro posterior.

$I_{08}^{11} = 100 \times R_{08}^{09} \times R_{09}^{10} \times R_{10}^{11} \times = 100 \times 1{,}6666 \times 1{,}2000 \times 1{,}45833 = 291{,}65$

Los resultados son iguales, la pequeña diferencia observada se debe a los decimales.
Ahora observemos en el cuadro siguiente el encadenamiento 2006-2011.

AÑOS	BASE VARIABLE	RELATIVOS	ÍNDICE ENCADENADO
2006	100,00	1,0000	100,00
2007	140,00	1,4000	140,00
2008	85,71	0,8571	120,00
2009	166,66	1,6666	200,00
2010	120,00	1,2000	239,98
2011	145,83	1,4583	349,96

$I_{06}^{06} = 100$
$I_{06}^{07} = 100 \times 1{,}40 = 140{,}00$
$I_{06}^{08} = 100 \times 1{,}40 \times 0{,}8571 = 119{,}99 \cong 120{,}00$
$I_{06}^{09} = 100 \times 1{,}40 \times 0{,}8571 \times 1{,}66666 = 199{,}98 \cong 200{,}00$
$I_{06}^{10} = 100 \times 1{,}40 \times 0{,}8571 \times 1{,}66666 \times 1{,}20 = 239{,}97$
$I_{06}^{11} = 100 \times 1{,}40 \times 0{,}8571 \times 1{,}66666 \times 1{,}20 \times 1{,}4583 = 349{,}96$

ÍNDICES AGREGATIVOS SIMPLES

En la mayoría de los casos, en forma errada, se calcula este índice sumando los precios, cantidades o valores de un período que está siendo investigado, dividiéndolo por la suma de los precios, cantidades o valores del período tomado como base. En algunos casos se considera improcedente este método, dado que no tiene en cuenta las variaciones que puede sufrir en mayor cuantía alguno de los elementos, de un período a otro.

$$I_o^t = \frac{\sum X_t}{\sum X_o} \times 100 \qquad I_o^t = \frac{\sum P_t}{\sum P_o} \times 100 \qquad I_o^t = \frac{\sum q_t}{\sum q_o} \times 100$$

Apliquemos la fórmula anterior a un caso específico de precios en un grupo de 5 artículos.

Ejemplo 3. Supongamos los precios de 5 artículos, para dos períodos 2009 - 2011

ARTÍCULOS	UNIDAD DE MEDIDA	PRECIOS		I_{09}^{11}
		2009	2011	
A	K/s.	3.200	3.200	100,00
B	Mts.	1.400	2.000	142,86
C	Doc.	800	6.000	750,00
D	Un.	5.600	5.200	92,86
E	Lbs.	11.200	12.000	107,14
Σ	-	22.200	28.400	1.192,86

En el cuadro anterior se ha calculado el índice de precios de la siguiente manera:

$$I_o^t = \frac{\sum P_t}{\sum P_o} 100 \qquad I_{09}^{11} = \frac{\sum P_{11}}{\sum P_{09}} 100 = \frac{28\ \ 00}{22\ 200} 100 = 127{,}9$$

Hay un aumento en los precios del 27,93% para el 2011 con respecto al 2009.

El anterior procedimiento no refleja la variación que sufre el artículo C pues de $800 pasa a $6.000, con un aumento de $\left(\frac{6\,000}{800} \times 100 = 750\right)$ del 650%. Para evitar este sesgo o error se procede a calcular el *índice agregati o simp e*, mediante el promedio de los índices simple para cada artículo, como lo muestra la última columna del cuadro.

$I_{09}^{11} = \frac{\sum I_{09}^{11}}{n} 100 = \frac{1\,192{,}86}{5} = 238{,}57$ el aumento es del 138,57% en vez del 27,93% obtenido por el primer método de cálculo.

La fórmula para el índice agregativo simple es: $I_o^t = \frac{\sum \left(\frac{X_t}{X_o}\right) 100}{n}$

ÍNDICES PONDERADOS O COMPUESTOS

Es aquél, en el cual los ítems o componentes están ponderados conforme a un sistema que refleja la importancia relativa de los mismos. Por lo general, casi todos los números índices están implícitamente

ponderados. Sin embargo, es corriente definirlo como ponderado sólo cuando en su expresión y cálculo, están explícitamente especificados los coeficientes de ponderación.

La *ponderación* es el peso o importancia que tiene un ítem o elemento dentro del conjunto.

Se cuenta con numerosas fórmulas para calcular estos índices, cuyo empleo y uso depende de su destino.

Se recomienda utilizar aquella fórmula que refleje en la forma más aproximada las variaciones que se deseen cuantificar.

Existen fórmulas para calcular índices de precios y de cantidad, siendo las más conocidas las de Laspeyres, Paasche, Fisher, Keynes, Marshall - Edgeworth, Drosbisch - Sidgwick y Walsh.

ÍNDICES PONDERADOS DE PRECIOS

En los índices de precios las ponderaciones, peso o importancia de cada elemento dentro del conjunto, están dadas por las cantidades y de acuerdo a la manera como éstas se tomen, dan origen a diferentes fórmulas, como se presentan a continuación:

Índice de Laspeyres

El índice de precios de Laspeyres, puede interpretarse entonces como la relación existente entre el valor que se obtiene al comparar, a los precios actuales, la misma cantidad de mercancías del precio base y el valor de tales artículos en el año base.

$$L^{I_G^t} = \frac{\sum p_t q_o}{\sum p_o q_o} 100$$

p_t = Precios del año que se investiga p_o = Precios del año base

$_o$ = Cantidades del año base $_t$ = Cantidades del año que se investiga.

Índice de Paasche

Permite determinar las variaciones en los precios para un período con relación a un período base, manteniendo constante como ponderaciones las cantidades del período que se investiga.

$$p^{I_G^t} = \frac{\sum p_t q_t}{\sum p_o q_t} 1$$

Índice de Fisher

Este índice se obtiene al calcular la media geométrica entre los índices de precios de Laspeyres y Paasche.

$$F^{I_G^t} = \sqrt{L^{I_G^t} \; p^{I_G^t}} = \sqrt{\frac{\sum p_t q_o}{\sum p_o q_o} \times \frac{\sum p_t q_t}{\sum p_o q_t}} 100$$

Índice de Sidg ic -Drobisch

Se obtiene promediando (media aritmética) los índices de precios de Laspeyres y Paasche.

$$S^{I_G^t} = \frac{L^{I_G^t} + P^{I_G^t}}{2}$$

Índice de Marshall-Edge orth

Indica las variaciones en los precios entre dos períodos, tomando como ponderaciones la suma de las cantidades del período base más las del período investigado.

$$M^{I_G^t} = \frac{\sum p_t \,(q_o + q_t)}{\sum p_o (q_o + q_t)} 100$$

Índice de Walsh

Como los índices anteriores, nos indica las variaciones que se manifiestan en los precios durante dos períodos, tomando como ponderación la raíz cuadrada del producto entre las cantidades del período base y el investigado.

$$W^{I_G^t} = \frac{\sum p_t \sqrt{q_o \; q_t}}{\sum p_o \sqrt{q_o \; q_t}} 100$$

Índice de Keynes

Toma como ponderaciones las cantidades más pequeñas, ya sean las del período base o las del investigado.

$$K^{I_G^t} = \frac{\sum p_t \; mínimo(q_o \; ; q_t)}{\sum p_o \; mínimo(q_o \; ; q_t)} 100$$

Ejemplo 4. Supongamos que se tienen los precios y las cantidades de cinco artículos para los períodos 2009 y 2011. Se desea calcular cada uno de los índices anteriores para determinar las variaciones de los precios para 2011 tomando como base 2009.

Tabla 12.1 Cálculo de índices Ponderados

ARTÍCULOS	UNIDAD DE PRECIO	2009		2011		$p_{09}\ _{09}$	$p_{11}\ _{11}$	$p_{11}\ _{09}$	$p_{09}\ _{11}$	$(\ _{09} + \ _{11})$	$p_{09}(\ _{09} + \ _{11})$	$p_{11}(\ _{09} + \ _{11})$
		Precios	Cantidad	Precios	Cantidad							
A	Lts.	820	3	1.000	3	2.460	3.000	3.000	2.460	6	4.920	6.000
B	Doc.	530	5	500	8	2.650	4.000	2.500	4.240	13	6.890	6.500
C	Mts.	1.120	10	1.400	8	11.200	11.200	14.000	8.960	18	20.160	25.200
D	Kls.	350	6	350	10	2.100	3.500	2.100	3.500	16	5.600	5.600
E	Un.	200	2	400	3	400	1.200	800	600	5	1.000	2.000
Σ	-	-	-	-	-	18.810	22.900	22.400	19.760	-	38.570	45.300

Calculemos los *índices de precios* de:

Laspeyres: $L^{I_{09}^{11}} = \frac{\sum p_{11}\ _{09}}{\sum p_{09}\ _{09}} 1 = \frac{\quad}{1\quad 1} 1 = 11$

Paasche: $P^{I_{09}^{11}} = \frac{\sum p_{11}\ _{11}}{\sum p_{09}\ _{11}} 1 = \frac{\quad}{1\quad 6} 1 = 11$

Fisher: $F^{I_{09}^{11}} = \sqrt{L^{I_{09}^{11}} \times P^{I_{09}^{11}}} = \sqrt{11 \qquad (11 \qquad)} = 11$

Sidgwick: $S^{I_{09}^{11}} = \frac{L^{I_{09}^{11}} + p^{I_{09}^{11}}}{} = \frac{11 \qquad + 11}{\qquad} = 11$

Marshall: $M^{I_{09}^{11}} = \frac{\sum p_{11}\ (\ _{09} + \ _{11})}{\sum p_{09}\ (\ _{09} + \ _{11})} 1 = \frac{\qquad}{\qquad} 1 = 11$

Tabla 12.2 Continuación del cálculo de los índices ponderados

$_{09}\ \ _{11}$	$\sqrt{\ _{09}\ \ _{11}}$	$p_{11}\sqrt{\ _{09}\ \ _{11}}$	$p_{09}\sqrt{\ _{09}\ \ _{11}}$	$mínimo\ _{09}\ \ _{11}$	$p_{11}\left(mínimo\ {}_{09};\ {}_{11}\right)$	$p_{09}\left(mínimo\ {}_{09};\ {}_{11}\right)$
9	3,000	3.000.00	2.460,00	3	3.000	2.460
40	6,324	3.162,00	3.351,72	5	2.500	2.650
80	8,944	12.521,60	10.017,28	8	11.200	8.960
60	7,746	2.711,10	2.711,10	6	2.100	2.100
6	2,449	979,60	489,80	2	800	400
-	-	22.374,30	19.029,90	-	19.600	16.570

Walsh: $W^{I_{09}^{11}} = \frac{\sum p_{11}\sqrt{q_{09}\ q_{11}}}{\sum p_{09}\sqrt{q_{09}\ q_{11}}} 100 = \frac{22.374{,}30}{19.029{,}90} 100 = 117{,}57$

Keynes: $K^{I_{09}^{11}} = \frac{\sum p_{11}\ mínimo\ (q_{09}; q_{11})}{\sum p_{09}\ mínimo\ (q_{09}; q_{11})} 100 = \frac{19.600}{16.570} 100 = 118{,}29$

ÍNDICES DE CANTIDAD

En los índices de precios, las ponderaciones son las cantidades y en los de cantidades las ponderaciones son los precios. Cuando las ponderaciones no se pueden obtener con facilidad durante varios períodos, se recomienda utilizar la fórmula de Laspeyres.

a) $L^{I_{o}^{t}} = \frac{\sum p_o\ q_t}{\sum p_o\ q_o} 100$

b) $P^{I_{o}^{t}} = \frac{\sum p_t\ q_t}{\sum p_t\ q_o} 100$

c) $F^{I_{o}^{t}} = \sqrt{L^{I_{o}^{t}}\ P^{I_{o}^{t}}} = \sqrt{\frac{\sum p_o\ q_t}{\sum p_o\ q_o} \times \frac{\sum p_t\ q_t}{\sum p_t\ q_o}} 100$

d) $S^{J_o^t} = \dfrac{L^{J_o^t} + P^{J_o^t}}{2}$

e) $M^{J_o^t} = \dfrac{\sum q_t (p_o + p_t)}{\sum q_o (p_o + p_t)} 100$

f) $W^{J_o^t} = \dfrac{\sum q_t \sqrt{p_o \; p_t}}{\sum q_o \sqrt{p_o \; p_t}} 100$

g) $K^{J_o^t} = \dfrac{\sum q_t \text{ mínimo } (p_o ; p_t)}{\sum q_o \text{ mínimo} (p_o ; p_t)} 100$

Ejemplo 5. Utilicemos la información obtenida anteriormente, para calcular los índices ponderados de cantidad, aplicando las diferentes fórmulas expuestas.

Laspeyres: $L^{J_{09}^{11}} = \dfrac{\sum p_{09} \;\; {}_{11}}{\sum p_{09} \;\; {}_{09}} 100 = \dfrac{19\,760}{18\,810} 100 = 105{,}05$

Paasche: $P^{J_{09}^{11}} = \dfrac{\sum p_{11} \;\; {}_{11}}{\sum p_{11} \;\; {}_{09}} 100 = \dfrac{22\,900}{22 \;\; 00} 100 = 102{,}2$

Fisher: $F^{J_{09}^{11}} = \sqrt{L^{J_{09}^{11}} \times P^{J_{09}^{11}}} = \sqrt{105{,}05 \times 102{,}2} = 10\;,6$

Sidgwick:

$$S^{J_{09}^{11}} = \frac{L^{J_{09}^{11}} + P^{J_{09}^{11}}}{2} = \frac{105{,}05 + 102{,}2}{2} = 10\;,6$$

Tabla 12.3 Cálculo de índices de cantidad

$(p_{09} + p_{11})$	${}_{09}(p_{09} + p_{11})$	${}_{11}(p_{09} + p_{11})$	$\sqrt{p_{09} \;\; p_{11}}$	${}_{09}\sqrt{p_{09} \;\; p_{11}}$	${}_{11}\sqrt{p_{09} \;\; p_{11}}$	Mínimo p_{09}; p_{11}	${}_{09}\left(\begin{matrix}mínimo\\ p_{09};\; p_{11}\end{matrix}\right)$	${}_{11}\left(\begin{matrix}mínimo\\ p_{09};\; p_{11}\end{matrix}\right)$
1.820	5.460	5.460	905,54	2.716,62	2.716,62	820	2.460	2.460
1.030	5.150	8.240	514,79	2.573,95	4.118,32	500	2.500	4.000
2.520	25.200	20.160	1.252,20	12.522,00	10.017,60	1.120	11.200	8.960
700	4.200	7.000	350,00	2.100,00	3.500,00	350	2.100	3.500
600	1.200	1.800	314,63	629,26	943,89	200	400	600
-	41.210	42.660	-	20.541,83	21.296,43	-	18.660	19.520

$W^{J_{09}^{11}} = \dfrac{21\,296,}{20\,5 \;\; 1{,}8} 100 = 10\;,67$ $\qquad$ $K^{J_{09}^{11}} = \dfrac{19\,520}{18\,660} \times 100 = 10\;,61$

ÍNDICES DE PROMEDIOS PONDERADOS RELATIVOS

En la práctica, es muy común utilizar los índices de Laspeyres y Paasche con ciertas modificaciones, siendo conocidos como *índices promedio ponderados re ati os*.

Los índices de promedios relativos de Laspeyres de precios y cantidad son:

$$L^{J_o^t} = \frac{\sum (p_o \; q_o) \; \dfrac{p_t}{p_o}}{\sum p_o q_o} \times 100$$

$$L^{J_o^t} = \frac{\sum (p_o \; q_o) \; \dfrac{q_t}{q_o}}{\sum p_o q_o} \times 100$$

Los índices de promedios relativos de Paasche de precios y cantidad son:

$$P^{I^{t}_{o}} = \frac{\sum (p_o\ q_t)\ \frac{p_t}{p_o}}{\sum p_o\, q_t} \times 100 \qquad P^{J^{t}_{o}} = \frac{\sum (p_t\ q_o)\ \frac{q_t}{q_o}}{\sum p_t\, q_o} \times 100$$

Tabla 12.4 Continuación del cálculo de los índices ponderados relativos

$\frac{p_{11}}{p_{09}}$	$\frac{\quad 11}{09}$	$(p_{09}\quad 09)\frac{p_{11}}{p_{09}}$	$(p_{09}\quad 09)\frac{\quad 11}{09}$	$(p_{09}\ p_{11})\frac{p_{11}}{p_{09}}$	$(p_{11}\quad 09)\frac{\quad 11}{09}$
1,21951	1,0000	3.000,00	2.460,00	3.000,00	3.000,00
0,94339	1,6000	2.499,99	4.240,00	3.999,99	4.000,00
1,25000	0,8000	14.000,00	8.960,00	11.200,00	11.200,00
1,00000	1,6666	2.100,00	3.500,00	3.500,00	3.500,00
2,00000	1,5000	800,00	600,00	1.200,00	1.200,00
-	-	22.399,99	19.760,00	22.899,99	22.900,00

Reemplazando los resultados obtenidos en el cuadro en las fórmulas anteriores, se tiene que: los índices de Laspeyres de precios y cantidad son:

$$L^{I^{11}_{09}} = \frac{\sum (p_{09}\quad 09)\ \frac{p_{11}}{p_{09}}}{\sum (p_{09}\quad 09)}\ 100 = \frac{22\quad 99{,}99}{18\ 810{,}00}\,100 = 119{,}09$$

$$L^{J^{11}_{09}} = \frac{\sum (p_{09}\quad 09)\ \frac{\quad 11}{09}}{\sum (p_{09}\quad 09)}\ 100 = \frac{19\ 760{,}00}{18\ 810{,}00}\,100 = 105{,}05$$

Los índices de Paasche serán:

$$P^{I^{11}_{09}} = \frac{\sum (p_{09}\quad 11)\ \frac{p_{11}}{p_{09}}}{\sum (p_{09}\quad 11)}\ 100 = \frac{22\ 899{,}99}{19\ 760{,}00}\,100 = 115{,}89$$

$$P^{J^{11}_{09}} = \frac{\sum (p_{11}\quad 09)\ \frac{\quad 11}{09}}{\sum (p_{11}\quad 09)}\ 100 = \frac{22\ 900}{22\quad 00}\,100 = 102{,}2$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. Con los siguientes datos sobre producción de cemento (Miles Ton. métricas) se pide:

2011 MESES	PRODUCCIÓN CEMENTO (Miles Ton. Métricas)
Enero	185
Febrero	178
Marzo	220

a) Calcular los índices simples, tomando como base el mes de enero.
b) Calcular los índices simples, tomando como base el mes de mayo.

ontin a

ontinuación

Abril	179
Mayo	199
Junio	175
Julio	216
Agosto	207
Septiembre	199
Octubre	208
Noviembre	218
Diciembre	213

c) Calcular los índices simples, con base variable.

2. Considerando las cosechas de ciertos productos agrícolas, determinar el índice agregativo simple para 2011 con base 2009, en las dos formas que usted conoce:

PRODUCTOS AGRÍCOLAS	**COSECHAS (Ciento de toneladas)**	
	2009	**2011**
A	11.158	13.044
B	1.196	1.357
C	1.111	1.326
D	1.460	1.840
E	859	997
F	1.106	870
A	41	659
H	6.686	7.978
B	204	202
Σ	23.821	28.273

3. Las cifras de ventas (en miles de millones de pesos) de unos grandes almacenes desde el año 2001 hasta el año 2011 son los siguientes:

AÑOS:	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
VENTAS*:	120	140	180	180	190	150	120	160	200	240	350

* Miles de millones de pesos

a) Hallar los índices simples de ventas tomando como base el año 2001.

b) Hallar el índice simple de ventas tomando como base el año de 2004.

4. El índice de cantidad de un grupo de artículos es igual a 200 si se usa la fórmula de Fisher y a 160 si se emplea la de Laspeyres. ¿Cuál es el valor del índice aplicando la fórmula de Paasehe?

5. Un índice para 2010 revela un aumento del 20% respecto del año anterior. En 2011 alcanzó a 174, es decir, presenta un incremento anual del 18%. Calcular los índices de 2009 y 2010.

6. Existen tres índices cuyas cifras son: para 2009, 107,83; para 2010, 108,12; para 2011, 104,26; es decir, entre 2008 y 2011 el índice eslabonado aumentó en un 19%. ¿Es cierto lo anterior?

7. Los números índices de la producción de un determinado bien de consumo desde 2005 hasta 2011 fueron los siguientes:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
ÍNDICES:	100,00	112,14	115,26	110,04	105,00	110,82	120,55

Sabiendo que en 2009 fueron producidas 3 T.m. de dicho bien. Hallar las cantidades producidas para los años 2005 y 2011.

8. En 2011 el precio de un artículo de consumo, bajó en un 29% con respecto al año anterior, pero en cambio, aumentó un 50% con respecto al año 2008. Hallar el número índice del precio de 2010, tomando como año base a 2008.

9. El índice de Laspeyres es 2/3 de Paasche y éste asciende a 130. ¿Cuál es el índice de Fisher?

10. Con la siguiente información determine:

DESTINACIÓN	PRÉSTAMOS AUTORIZADOS (miles de mill.)	CRÉDITOS EN NEGOCIACIÓN (miles de mill.)	SOLICITUDES PLAN DE EMERGENCIA (miles de mill.)	TOTAL (miles de millones)
Vivienda	25	50	175	250
Servicios Públicos	100	50	150	300
Desarrollo Industrial	65	60	-	125
Desarrollo Agrícola	175	-	-	175
Salubridad	-	150	-	150
Totales	**365**	**310**	**325**	**1.000**

1) Las asignaciones para salubridad de acuerdo al Plan Total en relación con lo asignado al Desarrollo Industrial, son aproximadamente:
a) Inferiores al 1,2% b) Inferiores en 16% e) Superiores en 1,2%
d) Superiores en 12% e) Igual o superior en 20%

2) Los préstamos autorizados para el renglón Vivienda, teniendo en cuenta el total programado para éste, corresponde aproximadamente:
a) 2 y 1/2% b) 5% e) 6 y 1/4% d) 10% e) 25%
c) Los préstamos autorizados para Desarrollo Agrícola y Vivienda, en relación al total de préstamos autorizados, corresponden aproximadamente al:
a) 15% b) 16% e) 20% d) 45% e) 54,79%

11. El índice de costo de construcción de casas de habitación (base 1996) registró las siguientes cifras:

AÑOS:	**1989**	**1991**	**1993**	**1995**	**1997**	**2000**	**2004**	**2011**
ICC*:	26,74	59,36	42,58	70,62	108,44	232,61	457,49	1.220,67

Si el costo de construcción de una casa en 1997 era de $6.200.000, ¿a cuánto ascenderá en 2011?
(* ICC = índice de costo de construcción)

12. En 2011 el precio medio de un bien disminuyó en un 25% con relación a 2005, pero se incrementó en un 50% con relación a 1997. Hallar el índice de precio para 2005 con base 1997.

13. Con la siguiente información, se pide calcular los índices para 2011 con base en 2008, aplicando las fórmulas siguientes:

ARTÍCULOS	UNIDAD	2008		2011	
		Precio	Cantidad	Precio	Cantidad
A	Kgs.	2.600	10	3.800	8
$	Kgs.	6.000	5	10.000	7
C	Lts.	1.000	2	4.000	5
D	Doc.	6.000	1	15.000	2
E	Kls.	3.600	2	2.000	1

a) Laspeyres, Paasche y Fisher para el índice de precios.
b) Keynes, Marshall, Sidgwick y Walsh para cantidades.
e) Laspeyres y Paasche de cantidad.

14. Con los siguientes índices eslabonados, se pide encadenarlos, tomando como período base el primer año de la serie.

AÑOS:	1	2	3	4	5
ÍNDICE ESLABONADO:	103,15	108,36	94,86	117,28	104,84

15. El índice nacional en costos de la construcción de vivienda para el período 2006 / 2011 es el siguiente:

AÑOS:	2006	2007	2008	2009	2010	2011
ÍNDICE 2001 = 100:	161,52	205,68	235,34	283,15	371,28	495,82

Si el valor promedio de una vivienda de 2 pisos, con 3 alcobas, sala, comedor, cocina y patio de ropas es de $31.000.000 para 2006, de acuerdo al incremento del índice, ¿Cuál será el valor en 2011?

16. Para los siguientes artículos:

ARTÍCULOS	UNIDAD	PRECIO PROMEDIO		CONSUMO (por mes)	
		2009	2011	2009	2011
A	Libra	2.450	2.605	6	8
B	Libra	426	750	2	3
C	Litro	472	740	80	65
D	Docena	1.275	2.583	3,50	2,50
E	Kilos	1.192	1.370	18	27

Se pide calcular:
a) Los índices ponderados de precios de Laspeyres, Paasche y Fisher, tomando como base 2009.
b) Los índices ponderados de cantidad de Laspeyres, Paasche y Fisher, tomando como base 2009.

17. Las ventas realizadas por un almacén de repuestos en el período 2005 / 2011 fueron las siguientes:

AÑOS :	2005	2006	2007	2008	2009	2010	2011
VENTAS:	12.320	14.563	18.624	27.632	24.830	29.316	32.514

* Millones de pesos

Se pide calcular:

a) Los relativos en cadena. (Índices eslabonados).

b) Encadene la serie obtenida en el punto (a), tomando como base 2005.
c) Obtenga el índice para la serie, tomando como base 2007.

18. El índice de precios calculado mediante la fórmula de Fisher es de 107,7. Además, se sabe que el índice de cantidad de Paasche es de 138,71 y el índice de precios de Paasche 136,06. Se pide calcular el índice de valor.

19. Dados los siguientes datos:

AÑOS:	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**
ÍNDICE 2000 = 100 :	182,63	204,28	210,34	214,36	286,49	322,24

Se pide calcular los índices, tomando como base 2006.

20. Que nos indica cada uno de los siguientes índices:
a) El índice de precios de Laspeyres.
b) El índice de cantidad de Laspeyres.
e) El índice de precios de Paasche.
d) El índice de cantidad de Paasche.

21. En 2009 la producción de una determinada firma aumentó en un 30% con respecto a 2008. En 2010 la producción bajó un 10% con respecto a 2009, pero en 2011 se superó en un 15% con respecto al año anterior. Hallar los números índices simples de cantidad, para el período 2009 / 2011; tomando como base el año 2008.

22. En una investigación nacional de morbilidad, se le hizo examen de "agudeza visual de lejos" a 3.400 niños de 8 a 12 años de edad y se obtuvieron los siguientes resultados:

Normales	3.209
Con diferencia leve	161
Con diferencia moderada	22
Con diferencia severa	8

Se pregunta por las tasas de cada uno de los niveles de agudeza visual por mil personas en la edad de 8 - 12 años.

23. Con los siguientes datos correspondientes a dos períodos, se pide calcular:

ARTÍCULOS	**2009**		**2011**	
	Precio	**Cantidad**	**Precio**	**Cantidad**
A	1.300	10	1.600	16
B	2.600	14	2.000	20
C	600	8	3.600	6
D	4.800	12	4.800	14

a) Los índices simples de precios para los dos períodos.
b) El índice agregativo simple de precios.
e) Los índices de Fisher de precios y de cantidad.

ÍNDICE DE VALOR

El índice está formado por la razón entre la sumatoria de valores (precio por cantidad) correspondiente al período investigado y la sumatoria de valores correspondientes al período base, al final multiplicado por cien.

Precio × Cantidad = Valor $\Rightarrow \sum$ PQ = Sumatoria de valores

$$IV_o^f = \frac{\sum V_f}{\sum V_o} 100 = \frac{\sum p_f q_f}{\sum p_o q_o} 100$$

Ejemplo 7. Utilicemos los ejercicios anteriores en donde se obtuvo:

Σp_{11} $_{11}$ = $_{11}$ = 22.900 y Σp_{09} $_{09}$ = $_{09}$ =18 810 con lo cual calculamos el índice de valor para los dos períodos I_{09}^{11} = ——— 1 = 1 1 . Sin embargo, este procedimiento puede ser reemplazado, utilizando los índices de Laspeyres y Paasche así:

$$I_{09}^{11} = {}^{I_{09}^{11}} \times P^{J_{09}^{11}} = 119\ 09\ (1\ 0223) = 121\ 745$$
$$I_{09}^{11} = {}^{J_{09}^{11}} \times P^{I_{09}^{11}} = 105\ 05\ (1\ 1589) = 121\ 742$$

Vale la pena observar en las anteriores operaciones, que los índices no pueden ser multiplicados entre sí, dado que serían multiplicados dos veces por cien, de ahí que uno de ellos debe ser expresado en valor relativo.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

24. Conocidos los índices de precios y de cantidad de Paasche, además, el índice de valor. ¿Cómo se pueden obtener los índices de precios y de cantidad de Laspeyres?

25. De un cierto producto tenemos los siguientes números índices simples de precios y de valor:

AÑOS	ÍNDICE DE PRECIOS (Base 2007 = 100)	ÍNDICE LE VALOR (Base 2003 = 100)
2007	100,00	210,00
2008	120,12	250,06
2009	130,24	280,31
2010	160,86	320,28
2011	180,08	350,34

Hallar los números índices de cantidad con base en 2007.

26. En enero de 2011 una fábrica pagó un total de $90.000.000 a 120 empleados en nómina. En junio del mismo año, la fábrica tenía 30 empleados más en nómina y pagó $22.500.000 más que en enero. Tomando el mes de enero como base, hallar:

a) El índice de empleo (cantidad relativa para junio).

b) El índice del costo de mano de obra (valor relativo).

e) Mediante la igualdad precio relativo por cantidad relativa = valor relativo. ¿Qué interpretación podría darse al precio relativo en este caso?

27. En 2011 el precio de un cierto bien de consumo aumentó en un 60% por encima del de 2008, mientras que su producción disminuyó en un 40 % ¿En qué porcentaje aumentó o disminuyó el índice de valor de dicho bien en 2011 respecto a 2008?

28. Una empresa espera aumentar sus ventas en el año próximo en un 50%. ¿En qué porcentaje deberá incrementar los precios para que el ingreso total se convierta en un 250%?

29. Un departamento de estadística le entrega a usted los siguientes índices, calculados para una serie de artículos.

$^{j}{}_{o}^{t} = 115{,}23 \qquad F^{J}{}_{o}^{t} = 125{,}32 \qquad {}^{I}{}_{o}^{t} = 130{,}65$ Calcular $I\ {}_{o}^{t}$

30. Una compañía espera incrementar sus ventas en un 34% en el próximo año. ¿En qué porcentaje deberá incrementar el precio de venta para que el ingreso total se triplique?

31. Un comerciante vende un artículo en $ 38.000 y busca una ganancia del 18,62 %. ¿En cuánto compró el artículo, para luego venderlo por ese precio?

EMPALME DE DOS O MÁS SERIES DE ÍNDICES

En numerosos casos los índices de una serie, *ámese índice de precios a consumidor* o de cualquier otra serie de precios, cantidad o valor, se suspende porque la información utilizada al conformar la base ha cambiado a través del tiempo, considerándose no representativa, dando lugar a la necesidad de elaborar una nueva serie de índices, con una nueva base. El índice deja de ser representativo cuando aparecen nuevos ítems, o desaparecen algunos otros que han venido siendo utilizados en el cálculo.

El método que permite realizar el cambio, para continuar la serie que ha sido suspendida o para conocer los índices anteriores a la nueva base, se conoce con el nombre de *empa me*. (Hacia arriba o hacia abajo).

Ejemplo 8. Con los siguientes datos correspondientes a dos series con distinta base, se pide empalmarlas, primero para los meses del 2011 (empalme hacia abajo) y luego para los meses y años anteriores. (Empalme hacia arriba).

El empalme hacia abajo lo hacemos mediante la aplicación de una regla de tres simple así:

106 21 → 1 644 06
115 86 → X

$$X = \frac{1.644{,}06\ (115{,}86)}{106{,}21} = 1.793{,}44$$

106 21 → 1 644 06
121 34 → X

$$X = \frac{1.644{,}06\ (121{,}34)}{106{,}21} = 1.878{,}26$$

AÑOS	MESES	ÍNDICE DE PRECIOS AL CONSUMIDOR PARA EMPLEADOS (Base jul. 97-jun. 98)	ÍNDICE DE PRECIOS AL CONSUMIDOR PARA EMPLEADOS Abril 1 10 = 100
2001	Marzo	233,78	15,10
	Agosto	240,42	15,53
2003	Febrero	433,56	28.01
2005	Octubre	536,53	34,66
2006	Abril	897,81	58,00
	Julio	929,60	60,05
	Diciembre	1.103,04	71,25
2010	Enero	1.460,08	94,32
	Junio	1.644,06	106,21
2011	Enero	1.793,44	115,86
	Febrero	1.878,26	121,34
	Marzo	2.344,81	151,48
	Abril	2.883,95	186,31
	Mayo	3.294,31	212,2

Sin embargo, es más fácil calcular una constante y multiplicarla por cada índice del período utilizado para el empalme.

$$K = \frac{1\,6\quad,06}{106{,}21} = 15,\ 79 \qquad \text{x} \quad \begin{cases} 115{,}86 = 1\ 79\ , \\ 121,\quad = 1\ 878{,}26 \end{cases} \quad \text{y así sucesivamente.}$$

En el empalme hacia arriba se invierten los cálculos así:

$$\begin{array}{l} 1.6\quad,06 \ \rightarrow\ 106{,}21 \\ 1.\ \ 60{,}08 \ \rightarrow\ \ x \end{array} \quad \text{ó la constante} \quad K = \frac{106{,}21}{1.6\quad,06} = 0{,}06\ \ 602265$$

$$K = 0{,}06\ \ 602265 \qquad \text{x} \quad \begin{cases} 1\quad 60{,}08 = 9\ ,\ 2 \\ \ \ 90\ ,0\ \ = 71\,25 \end{cases}$$

Nota: en la calculadora se realizan los cálculos utilizando todos los decimales.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

32. Con los siguientes datos:

AÑOS	MESES	ÍNDICE (1989 = 100)	ÍNDICE (mayo 10 = 100)
2001	Diciembre	2.382,68	
2003	Mayo	2.763,21	
2008	Junio	4.326,42	
2010	Enero	5.128,23	
	Abril	5.312,61	
	Septiembre	5.824,25	
	Diciembre	6.010,34	112,61
2011	Enero		114,52
	Febrero		118,63
	Marzo		120,11

a) Empalmar la serie obteniendo los índices para enero, febrero y marzo de 2011 con base 1989.
b) Obtener los índices para diciembre/01; mayo/03; junio/08 y enero, abril, septiembre de 2010 con base mayo 2003.

33. Dados los siguientes datos, empalmar las dos series de índices.

AÑOS:	2006	2007	2008	2009	2010	2011
ÍNDICE 2001 = 100	328,32	352,46	370,63	-	-	-
ÍNDICE 2008 = 100	-	-	100,00	105,61	108,89	111,23

34. Con la siguiente información, realice el empalme de dichas series.

AÑOS:	1	2	3	4	5	6
ÍNDICE A:	100,00	98,36	110,14	135,36		
ÍNDICE B:				100,00	108,12	96,84

ALGUNAS APLICACIONES DE LOS NÚMEROS ÍNDICES

Veamos, en forma bastante resumida, algunos de los indicadores más importantes, usados con frecuencia, y que reflejan no sólo la situación económica, sino otros aspectos diarios de la vida social y política de una región o país.

ÍNDICE DE PRECIOS AL CONSUMIDOR (IPC)

"EI IPC es un indicador estadístico, que permite establecer las variaciones en el gasto del consumo final promedio de los hogares, atribuido exclusivamente a los cambios en el nivel general de precios entre los períodos de tiempo determinados, de un conjunto representativo de bienes y servicios adquiridos por la población en referencia (hogares) para su propio consumo", según el DANE, Departamento Nacional de Estadística de Colombia.

Durante 1996 se adelantó el procesamiento, depuración y producción de resultados obtenidos en la encuesta de ingresos y gastos familiares realizados con anterioridad, con el fin de actualizar el IPC, cuya nueva base correspondió a diciembre de 1998.

La cobertura geográfica correspondió a 13 ciudades capitales de departamento, para un total de 20.000 establecimientos y 80.000 cotizaciones, siendo su recolección mensual. Los establecimientos que incluyó la muestra correspondió a los de comercio al por menor, viviendas en arrendamiento, educativos, servicios médicos, diversión, esparcimiento y financieros.

En este IPC se amplió la clasificación de los diferentes bienes y servicios, de 4 a 8. Ellos son:

1. Alimentos
2. Vivienda
3. Vestuario y calzado
4. Servicios médicos
5. Educación
6. Diversión, cultura y esparcimiento
7. Transporte y telecomunicaciones
8. Otros gastos

La canasta familiar fue constituida por 402 especificaciones, a las cuales se les hacía seguimiento de precios.

Es tan importante y variado el uso de este índice, que explicaremos a continuación algunos de ellos.

Poder adquisitivo (PA). Es denominado también como poder de compra y resulta de dividir la unidad monetaria, en nuestro caso el peso, por el IPC. Esta proporción se multiplica por 100.

$$PA = \left(1 \div IP\right)100$$

Si consideramos hipotéticamente que en diciembre de 2011 el IPC es 320,6 y su base es diciembre del 2005, el poder adquisitivo será igual a:

$$PA = \left(1 \div 320{,}6\right)100 = 0{,}3119$$

Lo anterior significa que un peso ($1) en diciembre de 2011 respecto al 2005 tendrá un valor real aproximadamente de 31 centavos, es decir, que el poder de compra entre estos dos períodos, ha disminuido en 69 centavos; por lo tanto, el consumidor necesitará más dinero para poder comprar la misma cantidad de artículos que adquirirla en diciembre de 2005.

Podríamos decir también que $1.000 en el 2011, tendrán un valor real de $310, siendo su poder de compra actual, aunque el billete aparentemente diga mil pesos.

El índice de poder adquisitivo se calcula con los índices de precios al consumidor que se tenga para los períodos comparados. En este caso, se tiene que para el mes de diciembre de 2005, el índice es de 100 y para diciembre de 2011 supuestamente es de 320,6.

$$IPA = \left(IP_{o}\ \ IP_{t}\right)100 \qquad IPA = \left(100{,}0 \div 320{,}6\right)100 = 31{,}19\%$$

Es decir, que el poder de compra es del 31,19% para el 2011 con respecto a diciembre del 2005.

El porcentaje de desvalori ación. Se obtiene aplicando la siguiente fórmula:

$$= 1 \quad \left[1 - \frac{IPC_{o}}{IPC_{t}}\right]$$

Podemos utilizar la información anterior para el cálculo, por lo tanto, se tendrá:

$$= 100\left[1 - \frac{100}{20,6}\right] = 68,81$$

La desvalorización del peso ha sido del 68,81%, lo cual significa que en ese porcentaje debería haber reajustado el salario durante el período, para que en el 2011 tenga un valor igual al que tenia en el período base.

> Por desvalorización se entiende la pérdida de poder de compra que tiene el dinero en relación con los precios internos de mercado.

En los tres casos de aplicación mencionados, se han tenido en cuenta los índices obtenidos en esos períodos.

Ejemplo 9. Supongamos que se tienen dos períodos, para los cuales se conoce los IPC con una misma base, siendo 396,48 y 480,53; correspondientes al mes de Enero/10 y Enero/11, la base es Mayo/05. Se pide calcular (a) Poder adquisitivo, (b) el índice de P.A. y (c) El porcentaje de desvalorización con base en Enero de 2010.

a) Primero cambiamos la base, para los IPC siendo Enero/10 = 100

   Índice Enero 2010 = $\frac{96,\ 8}{6,}\ 100 = 100$

   Luego, para enero de 2011 será: $\frac{80,5}{96,\ 8}\ 100 = 121,20$, hay un aumento del 21,20%

   El poder adquisitivo será: $PA = \frac{1}{121,20}\ 100 = 0,8251$, es decir, se registra una pérdida de: $1 - 0,8250 = 0,1749$ del 17,49%

b) El índice de poder adquisitivo será: $IPA = \frac{100}{121,20}\ 100 = 82,51$

c) El porcentaje de desvalorización $= 100\left[1 - \frac{96,\ 8}{80,5}\right] = 17,49\%$

Indexación. Consiste en vincular el valor de un capital o de una renta a la evolución de una variable de referencia, generalmente precios, producción, productividad. Numerosos profesionales, hablan de indización, que consiste en la acción de unir el valor de los salarios con la evolución de una variable de referencia, siendo de mayor frecuencia, el índice de precios al consumidor para preservar el poder de compra de los ingresos salariales, en períodos de alzas de precios o inflación.

Son términos muy comunes, especialmente para el economista, *deflactar*, *deflactación*, *deflactador, inde ar*. *Deflactar* es el proceso mediante el cual transformamos un valor o una serie dada en valores monetarios a *precios corrientes* (es decir, precios de mercado), a *precios constantes* de un período determinado, tomado como punto de referencia. Mediante la *deflactación* podemos determinar los cambios en el volumen real de valores de una serie.

Podemos observar, que la mayoria de las informaciones sobre valores correspondientes a inversión, producción, salarios, etc., dados en sus respectivos valores monetarios, a menudo están distorsionando la realidad, pues no muestran los cambios a nivel de precios corrientes.

Cada serie necesita de un índice deflactor propio, pues de lo contrario podríamos estar falseando la información. Generalmente se emplea el índice de precios al consumidor, para la deflactación o indexación de series monetarias.

Ejemplo 10. Hagamos un ejercicio hipotético sobre ventas (en miles de millones de pesos) en los últimos 6 años, a precios corrientes. Se desea deflactar utilizando el índice de Precios al Consumidor (valores arbitrarios).

Tabla 12.4 Índice de venta a precios corrientes y constantes

AÑOS	VENTAS (miles de millones)	IPC 2000 = 100	IPC 2006 = 100	VENTAS A PRECIOS CONSTANTE DE 2006	ÍNDICE DE VENTAS	
					Precios corrientes	Precios constantes
2006	38	225,61	100,00	38,00	100,00	100,00
2007	46	292,48	129,64	35,48	121,05	93,37
2008	54	346,56	153,61	35,15	142,11	92,50
2009	62	384,23	170,31	36,40	163,15	95,79
2010	78	436,58	193,51	40,31	205,26	106,08
2011	86	478,60	212,14	40,54	226,32	106,68

Procedimiento a seguir:

1. Cambiamos la base del índice de precios, (anteriormente 2000, ahora la nueva base es 2006). Calculamos la constante $K = \left(100 \div 225{,}61\right) = 0{,}443242764$ y la multiplicamos por $225\ 61 = 100$; por $292{,}48 = 129{,}64$; por $346{,}56 = 153{,}61$ etc.

2). Deflactamos, dividiendo las ventas de cada período por el respectivo índice de precios con base 2006

$$\frac{38}{100}\ 100 = 38 \qquad \frac{46}{129{,}64}\ 100 = 35{,}48 \qquad \frac{54}{153{,}61}\ 100 = 35{,}15$$

3). Observamos como han variado las ventas a precios corrientes y a precios constantes. Mientras que para 2011 las ventas (a precios corrientes) aumentaron en un 126,32%, a precios constantes o reales el aumento apenas fue del 6,68%, con respecto al 2006.

Salario nominal y salario real. Son dos términos de uso diario en los que interviene el IPC, ya que es utilizado para deflactar, es decir, *con ertir a ores o precios corrientes o de mercado a precios constantes* respecto a una base. En este caso, el salario nominal se puede transformar en salario real.

$$\text{Salario real} = (\text{Salario nominal} \div \text{IPC})\ 100$$

Si una persona, en Enero de 2010, devengaba un salario de $900 mil y en Enero de 2011 de $950 mil, se observa que el salario fue reajustado en 50 mil pesos, mejorando su salario nominal. Esta persona consultó el IPC para los mismos períodos: 264,94 y 320, 6, respectivamente, con base en diciembre de 2011.

Si cambiamos la base, al 2010 = 100,0; en 2011 será igual a (320,6 ÷ 264,94) 100 = 121,01, es decir, los precios se incrementaron en un 21,01% para el período de referencia. Luego, el salario real para 2011 será:

$$SR = (950.000 \div 121{,}01)\ 100 = 785.059{,}09$$

Este trabajador desmejoró su salario en $164.940,91 respecto al de 2010. Si se desea que esta persona devengue un salario real igual al de 2010 se obtendría así: $900.000 (1,2101) = $1.089.090, en vez de los $950.000 en que fue reajustado su salario.

TASA DE CAMBIO (TC)

Se usa para obtener algunos indicadores relacionados con las variaciones del tipo de cambio, es decir que refleje los cambios en la cantidad de dinero que se da por un dólar, o cualquiera otra unidad monetaria extranjera.

ndice de la tasa nominal de cam io

$_t$ = Tasa de cambio actual $\quad I \quad = (\ _t \quad _o)100$

$_o$ = Tasa de cambio que se toma como base.

Ejemplo 11. Se tiene que actualmente la tasa de cambio es $2.982,5 y se quiere determinar el índice de la tasa nominal, respecto a un período base cuyo valor era 2.782,6 por dólar.

Solución:

ITC = (2.982,5 ÷ 2.782,6)100 = 107,18

Quiere decir que la tasa de cambio ha aumentado durante ese período en un 7,18%. Directamente se hubiera podido calcular el porcentaje de variación, así:

$$\% \text{ de variación} = 100\left[\frac{T_t}{T_o} - 1\right] = 100\left[\left(2\,982\,5 \div 2\,782\,6\right) - 1\right] = 7\,18$$

Nota: Se puede determinar el índice real de cambio, del peso colombiano, si se deflacta por el índice de precios al productor.

Índice de devaluación (ID). Nos permite determinar la *pérdida de poder de compra de nuestro dinero en re ación a os precios de mercado e terno* es decir, el dólar o cualquiera otra moneda extranjera.

$$ID = \left(T_o \div T_t\right)1 \qquad ID = \left(2.782{,}6 \div 2.982{,}5\right)100 = 93{,}30\%$$

Índice de tasa de cambio real. Es la tasa de equilibrio creada por el Banco de la República en 1994 para establecer el precio del dólar y facilitar las operaciones de comercio exterior colombiano. Se calcula en forma mensual teniendo en cuenta el precio de las divisas de veinte de los países con mayores transacciones de Colombia con el exterior. Antes de 1994 existía una fórmula en la cual incluía 40 países.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

35. Deflactar la serie cronológica de las cantidades pagadas por las Compañías de Seguros, por siniestros ocurridos en un país X durante el período 2007-2011 tomando como deflactor el índice costo de vida.

AÑOS	SINIESTROS (miles de mill.)	ÍNDICE DE PRECIOS AL CONSUMIDOR (1995 = 100)
2007	318.511	643,12
2008	523.926	712,45
2009	670.718	807,78
2010	905.661	866,70
2011	1.036.129	876,93

36. Comprobar los siguientes puntos:
    a) Si el tipo de cambio sube de $2.662,0 a $2.871,0 por dólar ¿se dice que la devaluación ha sido del 9,63 %?
    b) Si la devaluación entre dos períodos fue de 9,83 %, siendo el tipo de cambio para el último período de $2.320,5 ¿Cuál fue el tipo de cambio para el período base?
    c) El índice de precios al consumidor sube en un 7,83%, ¿Se considera que el índice de poder adquisitivo baja en un 5,23%?

37. En los registros de inventario de equipo que lleva una compañía aparecen los siguientes Costos de adquisición (millones de pesos). Además se tienen los índices de precios del equipo en la industria metalúrgica.

AÑOS:	2006	2007	2008	2009	2010	2011
COSTOS (mill. $):	124,6	136,2	148,5	210,6	217,4	252,6
ÍNDICE 2001 = 100:	142,28	186,16	195,34	234,15	260,16	275,28

Obtenga el costo real de la maquinaria a precios del 2006.

38. Para el estudio del Producto Bruto se han tomado tres sectores y sus valores agregados (en miles de millones de pesos):

SECTORES	2008	2009	2010	2011
Agricultura	2.000	3.100	6.000	8.100
Servicios y otros	4.000	5.800	9.000	10.500
Industria	2.600	3.800	5.500	7.300

Por otra parte se dispone de los siguientes índices deflactores:

ÍNDICES	2008	2009	2010	2011
Índice de precios del productor	100,00	120,82	140,56	170,15
Índice de precios al consumidor	80,54	100,00	130,80	150,36
Índice de producción	110,63	125,32	135,42	140,42

    a) Calcule el producto bruto real, para los cuatro años, en unidades monetarias de 2008.
    b) Calcule el índice del producto bruto real con base 2008.

39. Los siguientes datos se refieren al monto total de los salarios nominales pagados por una empresa X (en miles de $) y a los índices del costo de vida
(Base 2001 = 100).

AÑOS	SALARIOS NOMINALES	ÍNDICE (2001 = 100)
2007	232.000	320,83
2008	235.500	325,12
2009	286.000	328,42
2010	338.000	350,63
2011	440.000	360,54

Se pide calcular:

a) Los salarios reales para cada año (base 2007).

b) El índice del poder adquisitivo para cada año (base 2007).

c) El poder adquisitivo con base 2001.

40. Con los siguientes datos obtenidos en el sector de la economía:

AÑOS	SALARIOS (miles mill.)	OBREROS (miles)	ÍNDICE DE PRECIOS AL CONSUMIDOR (2001 = 100)
2005	1.800	120	140,23
2006	2.060	180	148,11
2007	2.300	200	152,63
2008	3.800	300	160,21
2009	5.100	420	166,08
2010	5.800	450	168,26
2011	6.000	510	170,32

a) Calcular los salarios reales con respecto al 2005.
b) Calcular los salarios nominales per-cápita.
c) Calcular los índices de salarios reales, a precios de 2005.
d) Calcular los índices de salarios nominales per-cápita.
e) Calcular los salarios reales per-cápita, a precios de 2005.
f) Calcular los índices de salarios reales per-cápita, a precios de 2005.

41. Dados los siguientes datos sobre los salarios mensuales (miles de $) e índices de precios al consumidor con diferente base, encontrar si el salario real mensual es mayor o menor para 2011, que en 2006.

AÑOS	SALARIOS MENSUAL (miles de)	IPC (2001 = 100)	IPC (Mayo 08 = 100)
2006	400	142,39	
2007	500	160,51	
2008	1.100	280,32	
2009	1.300	420,16	110,18
2010	1.500	-	120,36
2011	1.900	-	140,56

42. Dados los salarios mensuales (miles de $) e índices de precios al consumidor, encontrar si el salario real mensual (miles de $) es mayor o menor en 2011 con relación al obtenido en 2004.

AÑOS	IPC (2001 = 100)	IPC Base Variable	SALARIOS MENSUALES (miles de)
2004	130,86	-	400
2005	146,32	-	600
2006	150,14	-	800
2007	170,78	100,30	1.100
2008	-	110,18	1.300
2009	-	90,36	1.800
2010	-	115,14	2.200
2011	-	112,26	2.500

43.

AÑOS :	2001	2002	2003	2004	2005	2006
IPC* :	464,53	532,98	607,31	759,42	963,53	1.136,04

AÑOS :	2007	2008	2009	2010	2011
IPC* :	1.430,41	1.849,32	2.179,28	2.376,42	2.516,83

*índice de Precios al Consumidor: 2001 - 2011

Con la serie de índices de precios al consumidor de 2001 al 2011, calcular el poder adquisitivo del peso para cada uno de los años, tomando como base 2001.

44. Cuando el índice de costo de vida sube en un 25%, ¿el índice del poder adquisitivo baja en un 20%?

45. Las informaciones del cuadro siguiente, corresponden al ingreso nacional de un país por rubros, en miles de millones de unidades monetarias.

RUBROS	2007	2008	2009	2010	2011
1. Salarios y prestaciones sociales	298,1	326,2	546,4	680,4	820,2
2. Arriendos	50,6	79,1	113,9	139,5	180,5
3. Intereses y utilidades	87,7	173,0	222,7	250,2	280,6

Además se tiene la siguiente información:

ÍNDICES / AÑOS:	2007	2008	2009	2010	2011
Índice precios consumidor 2011 = 100	24,51	38,06	63,28	81,42	100,00
Índice Precios del productor 2000 = 100	480,82	847,86	1.390,04	1.979,56	2.483,06
Índice precios vivienda 2009 = 100	49,84	83,67	100,00	115,62	120,46
Población en millones	7,8	8,0	8,2	8,4	8,7

a) Calcular el ingreso nacional real (total y per-cápita) a precios de 2007.
b) Calcular los índices de ingreso nacional real y per-cápita con base 2007.

46. Un empleado ganaba $632.000 mensuales en 2009. Hoy día percibe $972.000 con lo cual mejora su ingreso real en un 37%. Si el índice de precios para hoy (2011) es de 560,80 ¿Cuál era el de 2009?

47. El tipo de cambio de $2.829,8 por dólar significa una devaluación del peso de 37%. ¿Respecto a qué cotización?

48. La desvalorización monetaria en un país determinado aumenta cada año, durante el período 2005-2011 en un 10% constante, respecto al año anterior. Corregir la siguiente serie de valores de la depreciación monetaria.

AÑOS:	2005	2006	2007	2008	2009	2010	2011
Valores:	3.000	9.000	12.000	20.000	25.000	26.000	30.000

ÍNDICES DE COMERCIO EXTERIOR

En su elaboración se cuenta con la información necesaria, que se recopila mediante la utilización de los manifiestos de aduana. Se tienen dos índices: *de precios de cantidad*, tanto para exportación como para importación, con la diferencia de que al primero se le conoce como *ndice de a ores unitarios* (IVU) ya que no es precisamente un precio, sino el cuociente valor ÷ cantidad = p obteniéndose un valor unitario promedio de kilos, máquinas, telas, televisores, etc. Algo similar sucede con los índices de cantidad, denominados *ndices de uantum* (de cantidades o volumen físico) ya que en la mayoría de los casos, las mercancías, por razones arancelarias, se expresan en kilos; por tal razón es preferible utilizar el término *uantum* en vez de cantidad.

Algunos de los usos de estos índices, se emplean para calcular: la *re ación de precios de Intercambio* (RPI) y el *ndice de apacidad para Importar* (ICM).

Índice relación de precios de intercambio. El índice de Relación de Precios de Intercambio se obtiene, dividiendo el *ndice de a or unitario de e portación* (IVUX) por el *ndice de a or unitario de Importación* (IVUM) multiplicando el cuociente por cien.

$$IRPI = \frac{I\ UX}{I\ UM}\ 100$$

Este índice puede ser favorable para el país si es mayor del 100%, cuando los precios de los artículos que se exportan mejoran de precio en los mercados internacionales, mientras que el precio de los artículos que se importan se mantiene igual o el aumento es inferior a los precios de las exportaciones.

Índice de capacidad para importar. Se obtiene al multiplicar *e ndice de re ación de precios de intercambio por e ndice de uantum de e portaciones* **(**IQX). Es de anotar que uno de ellos debe estar dado en valores relativos, para que no quede multiplicado dos veces por cien.

$$ICM = (IRPI)\ (IQX)$$

Ejemplo 12. Supongamos que se dispone de los índices de valor unitario de importación y de exportación lo mismo que del índice quantum de exportación, para calcular los índices de Relación de Precios de Intercambio y los índices que muestren la capacidad para importar.

Tabla 12.6 Índice de relación de precios de intercambio y capacidad para importar

AÑOS	ÍNDICES (2001 = 100)		RELACIÓN NETA DE CAMBIO	ÍNDICE QUANTUM DE EXPORTACIÓN	ÍNDICE CAPACIDAD PARA IMPORTAR
	Exportación	Importación			
2005	126,41	110,52	114,38	105,82	121,04
2006	132,62	108,63	122,08	107,31	131,00
2007	120,48	109,42	110,11	109,22	120,26
2008	116,61	110,61	105,42	110,15	116,12
2009	128,24	111,20	115,32	110,83	127,81
2010	136,43	108,31	125,96	111,21	140,08
2011	144,81	114,56	126,41	111,86	141,40

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

49. Se conocen los índices de comercio exterior de un país: R = índice de relación de precios de intercambio.
Qx = índice de quantum de exportación.

$R_{06}^{07} = 110{,}18$ $R_{07}^{08} = 80{,}36$ $R_{08}^{09} = 120{,}14$ $R_{09}^{10} = 115{,}20$ $R_{10}^{11} = 116{,}18$

$_{09}^{06} = 90{,}36$ $_{09}^{07} = 120{,}12$ $_{09}^{08} = 90{,}34$ $_{09}^{09} = 100{,}00$ $_{09}^{10} = 110{,}21$ $_{09}^{11} = 80{,}36$

Se pide determinar el índice de capacidad para importar.

50. Un departamento de estadística entrega los siguientes índices de comercio exterior:

AÑOS	$_x$ Base variable	IVUX (2002 = 100)	IVUM (2006 = 100)
2006	100,00	124,35	100,00
2007	110,12	112,16	104,36
2008	80,36	118,14	103,16
2009	120,14	129,63	106,81
2010	115,16	132,35	105,32
2011	126,84	124,28	108,14

= Índice de quantum de exportación
IVUX = Índice de valor unitario de exportación
IVUM = Índice de valor unitario de importación
Se pide calcular el índice de relación precios de intercambio y la capacidad para importar con base 2006.

51. Con los siguientes datos:

AÑOS	ÍNDICE VALORES UNITARIOS		ÍNDICE QUANTUM DE EXPORTACIÓN (2002 = 100)
	Exportación (2002 = 100)	Importación (2002 = 100)	
2005	183,12	103,03	105,22
2006	143,23	108,62	109,34
2007	126,56	110,41	107,38
2008	148,14	106,28	124,81
2009	137,48	99,13	112,07
2010	139,16	108,64	120,46
2011	160,52	112,59	123,14

a) Calcular el índice de relación de precios de intercambio (2005 = 100).
b) Calcular el índice de capacidad para importar (2005 = 100).

ÍNDICES DE PRECIOS IMPLÍCITOS

La mayoría de las variables económicas se expresan en valores denominados a precios corrientes, que corresponden al efecto combinado de cambios en los precios y en las cantidades reales. Así mismo,

para determinar las modificaciones reales, se calculan los respectivos valores a precios constantes lo que implica aislar las variaciones en los precios. Este procedimiento requiere fijar los precios en un año determinado y evolucionar el valor en dicho período por las variaciones en las cantidades (volumen) a través del tiempo.

De las dos valoraciones anteriores, no es posible separar los precios absolutos ni las cantidades absolutas, sin embargo al efectuar la división entre el valor a precios corrientes por los constantes de un período determinado, es posible descomponer sus variaciones de valor en cambios de precios y cantidades. Ésta es la función de los números índices, que para el presente caso se denomina *ndice de precios Imp ícitos* al resultar cuantificado sólo el factor precio en términos relativos, con base en el período escogido en la serie a precios constantes.

Veamos parte del procedimiento que se sigue para obtener el *deflactor implícito del producto Interno ruto*.

Es necesario contar con la cantidad, precios y el valor de la producción del año que se investiga (n) y del año inmediatamente anterior ($n-1$) a precios corrientes, con el fin de obtener el valor de la producción a precios constantes del año anterior.

Se toma cada una de las cantidades (físicas) de los productos del año n y se multiplican por los precios unitarios del año $n-1$, así se obtiene el valor de la producción para cada uno de los productos involucrados en la muestra a precios constantes del año anterior. Para ello se requiere de la siguiente información:

Valor de la producción del año n (2011);	$Q_n \times P_n$	es decir, $_{11} \times P_{11}$
Valor de la producción del año $n-1$ (2010)	$Q_{n-1} \times P_{n-1}$	es decir, $Q_{10} \times P_{10}$
Valor de la producción del año $\frac{n}{n-1}\left(\frac{2011}{2010}\right)$	$Q_n \times P_{n-1}$	es decir, $_{11} \times P_{10}$

Ahora, aplicamos el *índice de aspe res* (para volumen) y los *índices de precios de Paasche* (para precios) de acuerdo con la metodología empleada, por algunos, en las cuentas nacionales. Estos índices, se calculan a nivel de cada sub-grupo.

$$\text{Índice de volumen} = \frac{\sum {}_{11}P_{10}}{\sum {}_{10}P_{10}} \qquad \text{Índice de precios} = \frac{\sum {}_{11}P_{11}}{\sum {}_{11}P_{10}} \times 100$$

Lo anterior se efectúa cuando el proceso es normal, es decir, cuando las variaciones de los precios se ajustan a la realidad, ya que se cuenta con información sobre el IPC y el IPP y las diferencias entre ellos no deben ser muy grandes. En el caso de que se tenga desconfianza sobre la variación presentada durante el período, se consulta el IPP y se hace evolucionar los precios del año $n-1$ en la misma proporción en que se aumentó el IPP; este sería el nuevo precio del producto que se está analizando.

Con los nuevos precios se consiguen las cantidades del año n respetando siempre el *a or*, luego se prosiguen calculando los índices de precios y volumen.

Ejemplo 13. Para explicar lo anterior, usamos valores arbitrarios de un producto, que se clasifica en 4 subgrupos.

SUB-GRUPOS	VALOR PRECIO - 2010	CANTIDAD -2010	PRECIO UNITARIO
A	2.392.338	4.413.637	542
B	3.528.972	7.352.025	480
C	13.681.533	25.083.019	545
D	2.875.359	4.940.728	582
Σ	22.478.202	41.789.409	2.149

VALOR PRECIO 11 10	VALOR PRECIO 2011	CANTIDAD 2011	PRECIO UNITARIO
2.729.248	3.139.254	5.039.417	623
2.740.563	3.787.173	5.709.507	663
13.786.548	18.216.593	25.275.548	721
3.680.098	4.951.689	6.323.511	783
22.936.457	30.094.709	42.347.983	2.790

Para calcular los *ndices de o umen*, *de precios* y *de a or* utilizamos la información de la tabla anterior, siendo:

Tabla 12.7 Símbolos y cálculos de precios y valor

2010			2011			
Valor precio	Cantidad	Precio unitario	Valor precio 2011 2010	Valor precio	Cantidad	Precio unitario
${}_{10}P_{10}$ 22.470.202	${}_{10}$	P_{10}	${}_{11}P_{10}$ 22.936.457	${}_{11}P_{11}$ 30.094.709	${}_{11}$	P_{11}

$$\text{Índice de volumen} = \frac{\sum {}_{11}P_{10}}{\sum {}_{10}P_{10}} \times 1 = \frac{6}{} 1 = 1$$

$$\text{Índice de precios} = \frac{\sum {}_{11}P_{11}}{\sum {}_{11}P_{10}} \times 1 = \frac{}{6} 1 = 1\ 1\ 1$$

$$\text{Índice de valor} = \frac{\sum {}_{11}P_{11}}{\sum {}_{10}P_{10}} \times 1 = \frac{}{} 1 = 1$$

Ahora, para obtener el *Producto Interno ruto* a precios constantes de un período determinado, se extrapolaron los valores del año base (PIB a costo de factores y a precios de mercado), de cada sector económico, por los respectivos índices de cantidades, generalmente referidos al volumen físico de producción, determinado mediante la aplicación de los precios del año base a las cantidades producidas en cada uno de los años del período. En algunos sectores, es posible individualizar cada uno de los artículos producidos y determinar el valor de la producción con respecto al período base; en otros no es posible (servicios), a éstos se les aplican otros indicadores, como representativos de la variación de la producción bruta. Se tendrán entonces deflactores de las exportaciones, de las importaciones, etc.

El Deflactor del PIB es el cuociente entre el PIB nominal (precios corrientes) de un determinado año y el PIB real (precios constantes) de ese año, indicando las variaciones que han experimentado los precios entre el año base y el año o años que se investigan.

Veamos el proceso para obtener el cálculo *deflactor implícito* del Producto Interno Bruto; para ello consideremos que en un país cualquiera se dispone del PIB a precios corrientes y constantes para los últimos 6 años y la base es 1995.

Tabla 12.8 Cálculo de índice deflactor del PIB

AÑOS	PIB Precios corrientes	PIB Precios constantes 95	ÍNDICE DEFLACTOR 1995 = 100	ÍNDICE DEFLACTOR 2006 = 100	% Variación
2006	26.240.771	950.694	2.760,17	100,00	0
2007	33.064.150	977.172	3.383,66	122,59	+22,59
2008	40.058.610	1.002.620	3.995,39	144,75	+44,75
2009	49.352.520	1.054.380	4.680,71	169,58	+69,58
2010	69.720.830	1.196.750	5.825,85	211,07	+111,07
2011	82.465.310	1.342.810	6.141,25	222,50	+122,50

$$ID_{06} = \frac{26\,2\;\;0\,771}{950\;69}\;100 = 2\,760{,}17 \qquad ID_{07} = \frac{6\;\;1}{1}\;1 = 66$$

$$ID_{08} = \frac{0\,058\,610}{1\,002\,620}\;100 = 995, \qquad ID_{09} = \frac{9\;\;52\,520}{1\,05\;\;80}\;100 = 680{,}71$$

Ahora procedemos a cambiar la base del *Índice Deflactor* de 1995 a la nueva base 2006.

$$ID_{06} = \frac{2\,760{,}17}{2\,760{,}17}\;100 = 100 \qquad ID_{07} = \frac{8\;,66}{2\,760{,}17}\;100 = 122{,}59$$

$$ID_{08} = \frac{995,\;9}{2\,760{,}17}\;100 = 1\;\;\;,75 \qquad ID_{09} = \frac{680{,}71}{2\,760{,}17}\;100 = 169{,}58\text{, etc..}$$

ÍNDICE DE PRODUCTIVIDAD (IP)

Corresponde a los cambios de eficiencia en la producción, obtenidos mediante la relación porcentual entre la producción y los insumos. Generalmente se calcula un *índice de producti idad* con un insumo: fuerza de trabajo, ya sea considerando el *n mero de personas ue aboran* o el *n mero de horas hombre*

I Prod = (Producción ÷ horas–hombre) 100

I Prod = (Producción ÷ número de trabajadores) 100

El Índice de producción se refiere a los cambios que sufre la producción, generalmente industrial, durante un período determinado, con respecto a otro período, considerado como base o punto de referencia.

Ejemplo 14. Con los datos que aparecen en el cuadro siguiente, se pide calcular el índice de producción y de productividad.

Tabla 12.9 Índice de producción, obreros y productividad

AÑOS	PRODUCCIÓN (miles Tons.)	NÚMERO OBREROS	ÍNDICE 2005 = 100		ÍNDICE DE PRODUCTIVIDAD 2005 = 100
			Producción	Obreros	
2005	350	136	100,00	100,00	100,00
2006	420	136	120,00	100,00	120,00
2007	580	150	165,71	110,29	150,25
2008	650	152	185,71	111,76	166,17
2009	810	175	231,43	128,68	179,85
2010	970	175	277,14	128,68	215,37
2011	1.150	184	328,57	135,29	242,86

Índice de producción $= \frac{100}{50} = 0{,}285761 \times \begin{cases} 50 = 100 \\ 20 = 120{,}00 \\ 580 = 165{,}71 \ etc \end{cases}$

Índice de obreros $= \frac{1}{1\ 6} = \times \begin{cases} 1\ 6 = 1 \\ 1\ 6 = 1 \\ 1 \quad = 11 \\ 1 \quad = 111\ \ 6 \ etc \end{cases}$

$= \text{———————} \ 100$

$I_{05}^{05} = P_{rod} = \frac{100}{100} \times 100 = 100$

$I_{05}^{06} = P_{rod} = \frac{120{,}00}{100} \times 100 = 120{,}00$

$I_{05}^{07} = P_{rod} = \frac{165{,}71}{110{,}29}\ 100 = 150{,}25$, etc...

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

52. Los índices de producción de un determinado bien de consumo, desde 2005 hasta 2011, fueron los siguientes:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
Índice	100,00	112,83	115,24	110,12	105,36	110,92	120,55

Sabiendo que en 2009 fueron producidas 300 T.m. de dicho bien, Hallar las cantidades producidas entre los años 2005 y 2011.

53. Con los siguientes datos, obtenga el índice de producción y de productividad.

AÑOS:	2005	2006	2007	2008	2009	2010	2011
P*:	1.420	1.630	1.580	1.710	1.812	1.750	1.800
PO**:	12.380	13.620	12.400	13.500	12.200	12.100	12.000

*P = Producción (miles de toneladas) * *PO = Promedio Obreros

54. Con los siguientes datos (miles millones de pesos), se pide calcular la producción industrial de un sector de la economía para los años 2007 al 2011.

AÑOS	PRODUCCIÓN INDUSTRIAL (precios de 2009)	ÍNDICE DE PRODUCCIÓN INDUSTRIAL (base 2007 = 100)
2007	5.000	100,00
2008	-	198,36
2009	-	120,42
2010	-	135,86
2011	-	150,14

ÍNDICE PRECIOS DEL PRODUCTOR (IPP)

Es un índice de gran importancia para economistas, empresarios y gobierno en general, ya que permite estudiar las variaciones generales en el nivel de precios, evaluar el desajuste entre la oferta y la demanda agregada y las relaciones entre los bienes individuales. Para ello, se toman los precios de un grupo de artículos que se venden en grandes cantidades en los mercados primarios, es decir, los precios pagados por el minorista, detallista o vendedor, al productor o mayorista. Indica, además, los cambios relativos de precios que los fabricantes pagan por la materia prima y por lo general, incluyen todos los principales productos que se utilizan en la industria.

> El IPP en nuestro país es elaborado por el Banco de la República, el cual lo define como *un indicador de a e o ución de os precios en e primer cana de comercia i ación de os bienes transados en a economía*.

Anteriormente se le denominaba índice de precios al por mayor, y su última base fue 1970, con una cobertura de 14 ciudades y una muestra constituida por 540 artículos. Para el índice total se incluían los bienes producidos y consumidos en el país, los importados y los exportados. A causa de las distorsiones que le fueron encontradas, a raíz de la bonanza cafetera, se suprimieron los bienes importados y exportados. Otro de los inconvenientes que ocasionaba este indicador, era la inclusión en el universo de observación a productores y distribuidores mayoristas, lo cual originaba la presentación de una gran variedad, con resultados de una diferencia acentuada en los componentes de los precios.

Debido a la apertura de 1990, se generaron distorsiones en el indicador. Para evitar el múltiple conteo se establecieron 4 índices: *bienes finales* y *consumidos naciona mente*; *bienes finales importados; bienes oficiales ofrecidos internamente*; y *demanda final* (incluye exportados), adicionalmente se calculan tres índices más: *E portados*; de *bienes destinados a a demanda intermedia* (materias primas); y de *oferta interna total* (IPP con actuales ponderaciones).

Dado que las cuentas nacionales tienen como nueva base 1994, se escogió este período como base,

para el cálculo de las ponderaciones. Por su parte, el mes de comparación de estos índices es junio de 2009 = 100.

ÍNDICE BURSÁTIL MUNDIAL DOW JONES

El índice Bursátil Mundial Dow Jones está formúlado con el propósito de mostrar el rendimiento de diferentes industrias en las bolsas mundiales. El índice se basa en el precio de las acciones de 2.600 empresas que se cotizan en 26 bolsas. Las bolsas incluyen algunas importantes, como las de Nueva York, Tokio y Londres, así como algunas más pequeñas, como las de Hong Kong, Estocolmo y Madrid. Hasta la fecha, la única bolsa latinoamericana que se incluye en el índice es la de ciudad de México. Para la creación del índice se tradujo el precio de las acciones a 31 de diciembre de 1991 a una base de 100 puntos. El índice sube o baja a medida que el precio en bolsa de las acciones que lo integran sube o baja en dólares, a la tasa de cambio del dia. Este índice hace referencia a 9 grupos que tienen en total 30 sub-grupos.

<table>
<tr><td>Materias básicas
Productos químicos
Productos forestales
Minería
Metales no ferrosos (todos)
Productos de papel
Metales preciosos
Acero</td><td>Consumo productos cíclicos
Publicidad
Aerolíneas
Fábrica de autos
Construcción de viviendas
Muebles y artículos para el hogar
Electrónica de consumo
Ropa</td></tr>
<tr><td>Consumo productos no cíclicos
Alimentos
Pesca
Farmacia
Tabaco</td><td>Industriales
Materiales de construcción
Transporte marítimo</td></tr>
<tr><td>Energía
Carbón
Petróleo</td><td rowspan="2">Tecnología
Computadoras
Medicina y biotecnología
Semiconductores, etc.
Programas informáticos</td></tr>
<tr><td rowspan="2">Finan as
Bancos (todos)
Seguros (todos)</td></tr>
<tr><td>Servicios públicos
Sistemas telefónicos</td></tr>
<tr><td></td><td>Independientes
Conglomerados</td></tr>
</table>

Refleja la internacionalización de los precios y las industrias, aunque las condiciones locales pueden afectar el precio de una acción en una empresa y una bolsa determinada.

INDICADORES DE DESEMPLEO

Para el cálculo de las tasas sobre *desemp eo*, *sub-emp eo* y *otros indicadores de la fuerza laboral*, se requiere en primer lugar la realización de una investigación denominada, en nuestro país, *Encuesta de hogares*.

La encuesta de hogares cubría 14 de las principales ciudades. En 7 ciudades la encuesta se realizá trimestralmente, semestralmente en dos y anualmente en 5, para determinar los cambios globales en los niveles de empleo de la población, a fin de suministrar la información básica necesaria para el diagnóstico de la fuerza laboral.

Los principales indicadores de empleo que se calculan son los siguientes:

a) Tasa de desempleo (TD). Se considera como la relación porcentual entre el número de personas desocupadas (D) y el número de personas que integran la población económicamente activa (PEA).

$$TD = (D \div PEA)\ 100$$

D = Personas con edades de más de 12 años que no trabajaron, hicieron o no diligencias para conseguir trabajo, o se hallaban disponibles para trabajar durante la semana de referencia.

PEA = Es el conjunto de personas mayores de 12 años, de cualquier sexo, que durante la semana anterior a la encuesta, ejercieron o buscaron ejercer una ocupación remunerada en la producción de bienes y servicios, y aquéllas que en su condición de ayudantes familiares trabajaron sin remuneración en la empresa del respectivo jefe de familia o pariente, por lo menos durante 15 horas semanales, de acuerdo con la definición establecida por el DANE.

b) Tasa de subempleo (TS), Es también una relación porcentual de la población que manifestó *Querer y poder trabajar más horas a la semana* (PS) y el número de personas que integran la población económicamente activa (PEA).

$$TD = (PS \div PEA)\ 100$$

c) Tasa bruta de participación (TBP). Es el cuociente entre el número de personas que componen la fuerza de trabajo (PEA) y el número de personas que integran la población total (PT).

$$TBP = (PEA \div PT)\ 100$$

d) Tasa global de participación (TCP). Es la relación porcentual entre la población económicamente activa (PEA) y el número de personas que integran la población en edad de trabajar (PET) (personas de doce años y más).

$$TGP = (PEA \div PET)\ 100$$

OTROS INDICADORES (OI)

Son numerosos los indicadores económicos, sociales etc., algunos de ellos conocidos como tasas, cuando la base con la cual se calcula es bastante grande.

Tasa de natalidad (TN). La tasa bruta de natalidad de un área, es el número de nacimientos realmente ocurridos durante un período de tiempo, dividido por la población estimada de esa área, expresándose en términos “por mil habitantes”.

Tasa de fertilidad (TF). Número de nacidos vivos en un período, expresado como proporción (generalmente por mil), de las mujeres potencialmente fértiles de la población respectiva.

Tasa de mortalidad (TM). Número de muertes en un período, dividido por la población expuesta a riesgo de muerte en ese período.

Tasa de incidencia (TI). Prácticamente es una razón que se obtiene al dividir el número de nuevos casos (de una enfermedad), por la población sometida a riesgo en un período.

Tasa negativa a responder (TNR). Proporción de individuos que a pesar de haber sido correctamente localizados, se niegan a responder.

Otras tasas (OT). *asa actuaria* ; *tasa de arrendamiento*; *tasa de cambio*; *tasa de captación*; *tasa de corrección monetaria*; *tasa de interés (acti a o pasi a)*; *tasa de redescuento*; *tasa dua de retorno*; *tasa preferencial*; *tasa rea* ; *tasa de accidenta idad* etc.

Algo similar ocurre con los índices.

Índice bursátil (IB). Evolución temporal de las cotizaciones de los valores, en relación con un valor tomado como base.

Índice de bolsa. Es el número relativo que indica diariamente la tendencia del mercado de las acciones inscritas y transcritas en dicha institución. Las más conocidas son: Dow Jones (bolsa de Nueva York); NIKKEY (bolsa de Tokio); FTSE 100 (bolsa de Londres); IGBC (bolsa de Colombia) y muchas más, dado que cada país tiene su propia bolsa para el mercado de sus acciones.

Se podrían mencionar ligeramente algunos otros índices: de actividad por rotación de inventarios; de actividad por velocidad de efectivo; de capital circulante; de rotación de cartera; de rotación de intereses; de rotación de inventarios de productos en proceso; de *rotación de obligaciones bancarias;* de *rotación de sobregiros*; *de so encia*; de *Morgan Stan e* y otros más, que no dejan de ser importantes como el *índice de concentración de INI* (1909); *e índice de Pareto*, etc.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

55. El valor de los activos para los últimos 6 años, de una compañía, son:

AÑOS:	2006	2007	2008	2009	2010	2011
Activos (miles):	850.000	1.000.700	1.370.000	1.720.600	2.120.300	2.850.320

Se pide: a) Un índice que muestre cómo ha evolucionado el valor de los activos con respecto a. 2006. b) Si para los mismos años el índice deflactor (IPC) fue de: 575,1 732,6 954,3 1.250,3 1.380,4 y 1.706,2 respectivamente. ¿Los activos han crecido en términos reales con respecto a 2006?

56. Una empresa espera aumentar los precios el año próximo en un 25% ¿En qué porcentaje deben aumentar las ventas para que el ingreso total sea de un 220%

57. Se tienen dos series de índices (con diferente base) para el IPC

AÑOS:	2006	2007	2008	2009	2010	2011
Índice 2002 = 100:	428,6	594,2	672,4			
Índice 2008 = 100:			100,0	135,6	162,8	201,4

Si el salario promedio que pagaba una empresa en el 2006 era $746.400 y para 2011 paga un promedio salarial de $1.110.100; (a) La situación laboral en la empresa mejoró en términos reales en 2011 con respecto al 2006?; b) ¿Cuál debe ser el salario promedio en 2011 para que sea igual al del 2006?

58. Conteste brevemente los siguientes puntos:
    a) ¿Cuál es la importancia y finalidad de los números índices?
    b) ¿Qué significa punto de referencia o período base?
    e) ¿ En qué consiste el índice de precios de Laspeyres?
    d) ¿Qué entiende por índice de cantidad de Paasche?
    e) ¿Para qué se hace el empalme entre dos o más series?
    f) ¿En qué consiste el encadenamiento?
    g) ¿Qué es deflactar y cómo se realiza?

59. Con las siguientes series de índices: a) deflactar la serie de inversión (miles millones de $) realizada por el gobierno en una determinada región del país, tomando como referencia a 1999; b) ¿Cuál debe ser la inversión real para 2011 a precios de 2008?; e) ¿Cuál debe ser la inversión (nominal) en el 2011, para que a precios reales sea igual a la de 2006?

AÑOS	IPC 1987 = 100	IPC Base variable	IPC 2007 = 100	INVERSIÓN (miles mill)
1993	285,3			25
1995	336,4			37
1997	520,6			42
1999	720,6			58
2001	810,4	106,4		65
2003		118,5		80
2005		107,3		105
2006		85,4		115
2008		114,3	116,4	130
2009			137,3	142
2010			155,4	160
2011			176,8	180

60. Explique brevemente qué nos indica cada uno de los siguientes índices:
    a) Índice de Dow Jones.
    b) Índice de Precios del Productor.
    c) Índice de Precios al Consumidor..
    d) Índice de Desvalorización.
    e) Índice de Devaluación.

61. Una persona que en el 2009 tenía un sueldo de $450.000 y el actual de $ 1.068.000, el cual representa un aumento del 34% de su salario real. Teniendo en cuenta que el actual índice de precios al consumidor es de 350,62, se pide determinar el índice para el 2009.

62. Comprobar los siguientes puntos:
    a) Si el tipo de cambio sube de $2.762,5 a $2.871,8, ¿se dice que la devaluación ha sido, para ese período, del 4,25%?
    b) El índice de precios al consumidor sube en un 31%, ¿se considera que el Índice de poder

adquisitivo baja en un 24,6%?

e) Si la devaluación entre dos períodos fue del 6,5l%, siendo el tipo de cambio para el último período de $2.376,8, ¿cuál fue el tipo de cambio para el período tomado como base?

63. Una compañía espera incrementar sus ventas en un 34% en el próximo año. ¿En qué porcentaje deberá incrementar el precio de venta para que el ingreso total se triplique?

64. Los índices de precios al consumidor fueron del 307,13 y del 386,82 durante un cierto período.
    a) Si una familia presupuesta $95.000 mensuales para pagos en servicios (agua y luz) al comienzo de ese período, ¿qué cantidad debería presupuestar para los mismos servicios al final del período?
    b) Suponga que el ingreso mensual familiar pasa de $720.000 a $800.000 durante ese período ¿Cuál es el ingreso mensual real durante ese período?

65. Considere el total de alumnos en su curso y determine las siguientes proporciones:
    a) Alumnos con promedio de calificación superior a 3,6
    b) Alumnos que ven en TV el noticiero de las 7 p.m..
    c) Alumnos que dejaron de asistir un día de la semana.
    d) Alumnos que tengan más de dos hermanos.
    e) Alumnos que midan más de 1,55 centímetros.
    f) Mujeres en el curso.

66. Calcular los porcentajes correspondientes con la información del punto anterior.

67. Encontrar las proporciones y porcentajes en los siguientes casos:
    a) Alumnos de su curso que están inscritos en algún equipo de fútbol.
    b) Número de ases en una baraja (naipes de 40 cartas).
    c) 10 alumnos ausentes en su curso.
    d) 5 votos favorables, en una elección realizada en su salón.

68. Un comerciante vendió 15 neveras en el mes de abril. En el mes de mayo, con motivo de la celebración del día de la madre, vendió 60. ¿Cuál fue el porcentaje de:
    a) disminución en las ventas en el mes de abril con relación a mayo?
    b) aumento en las ventas en el mes de mayo con relación a abril?

69. ¿Cuál es el porcentaje de ganancia en un artículo que un comerciante compró en $15.000 y lo vendió en $18.200?

70. En la vitrina de un almacén aparece al lado de un artículo un letrero que dice: “precio rebajado en un 200%”. Si el precio del artículo era $8.000 y su precio rebajado es $4.000; ¿cuál es el verdadero porcentaje de rebaja?

71. Si un tendero vende un artículo en $3.500 y busca una ganancia del 25%, ¿en cuánto compró el artículo para luego venderlo por ese precio?

72. Un fabricante vende 1.200 artículos en una semana. ¿Cuántos artículos deberá vender para obtener un incremento del 30%?

73. Suponga que una fábrica tiene 1.200 empleados, de los cuales 400 son mujeres y se quiere saber

cuál es:

a) la proporción de mujeres b) el porcentaje de hombres

c) la relación de mujeres a hombres d) la relación de hombres a mujeres

74. Cuál es la proporción de unidades defectuosas producidas por una máquina en un día, si la producción fue 7.500 unidades, de las cuales 320 fueron clasificadas como defectuosas?

75. Investigue y establezca cuáles han sido en su país, para los últimos 5 años, las tasas de:

a) Mortalidad b) Desempleo c) Nacimientos

d) Participación laboral e) Deserción escolar f) Analfabetismo

76. Investigue cómo se obtiene y qué significan las siguientes razones:

a) de liquidez, de solvencia o corriente;

b) de liquidez inmediata o prueba de fuego;

e) de alta fluidez;

d) de endeudamiento financiero;

e) de accionistas propietarios;

f) de rentabilidad del capital contable.

77. Suponga que se dispone de cierta información por sexo, sobre personas portadoras de un *irus*

SEXO	**ENFERMEDAD X**		**TOTAL**
	Portadores	**No portadores**	
Hombre	(a) 3.500	(c)	(n) 5.600.000
Mujeres	(b) 7.200	(d)	(m) 9.230.000
Total	(e) 10.700	(f)	(k) 14.830.000

De las siguientes relaciones, ¿cuáles considera que son razones y proporciones?

Nota: En el SIL *Sistema de Informacion en Línea* se incorporó un capítulo que debe ser consultado sobre Razones y Porcientos, titulado "Algunos elementos básicos de las matemáticas Capítulo 14.

$1\ \frac{(a)}{(b)}$ $\frac{(a)}{(e)}$ $\frac{(n)}{(k)}$ $\frac{(c)}{(d)}$ $5\ \frac{(c)}{(f)}$ $6\ \frac{(b)}{(m)}$ $7\ \frac{(n)}{(m)}$ $8\ \frac{(b)}{(e)}$ $9\ \frac{(d)}{(f)}$ $10\ \frac{(e)}{(k)}$

Síntesis de capítulo

Los números índices generalmente indican las variaciones o cambios que presenta una variable a través del tiempo, por esta razón se constituyen en series de tiempo. Su importancia para ser usados continuamente y en numerosos casos, se debe a la posibilidad de establecer comparaciones con datos presentados en diferentes unidades de medidas, de ahí su uso en el campo económico donde prevalecen los estudios con unidades heterogéneas.

Es una herramienta estadística fácil de manejar y entender, por esto varias entidades prefieren mostrar la información en forma de porcentajes, en vez de suministrarla de manera diferente. Es más simple el análisis del comportamiento de una serie cuando es expresado en términos relativos.

En el campo económico hay tantos indicadores que faltaría espacio para mencionarlos todos; algunos de estos índices son: los de costo de la construcción pesada y de vivienda; de las ventas del comercio al por menor; demanda de la energía eléctrica; sacrificio de ganado; de la oferta monetaria; medios de pago; colocaciones del sistema financiero, etc. Pero no todos están asociados al campo económico, sino que también son utilizados en campos muy variados, por ejemplo en la educación, para calcular los índices de escolaridad, de retención, de mortalidad académica; en el campo de la salud, justicia, transporte, etc.

11 CAPÍTULO

Series Cronológicas

"La ciencia es sólo un ideal. La de hoy corrige la de ayer y la de mañana la de hoy.

os rtega y asset

CONTENIDO

- Términos, definiciones, utilidad y aplicaciones. Conceptos generales.
- Componentes de una serie.
- Tendencia rectilínea, parabólica y exponencial. Cálculos y graficación.
- Método de los semi-promedios y Promedios móviles.
- Desestacionalización de una serie de tiempo.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Entender y manejar correctamente cada uno de los términos que deberán ser utilizados.
- Conocer la importancia que tiene la Estadística no sólo como asignatura, sino en el ejercicio profesional.
- Realizar pequeñas investigaciones estadísticas, aplicando los conocimientos recientemente adquiridos, en cada una de las etapas del proceso.
- Distinguir con claridad la diferencia entre una Monografía y una Encuesta,

GENERALIDADES

Las *series de tiempo* también se les denomina *series crono ógicas*

Datos atempora es: son aquéllos que se obtienen de investigaciones aisladas, no periódicas, es decir, se producen una vez, dos veces, etc, pero no se vuelven a producir.

Datos tempora es: son aquéllos que se van registrando a medida que se van produciendo, en un estricto orden cronológico. En la vida diaria nos encontramos con mucha información de este tipo: nacimientos, accidentes de tránsito o laborales, exportaciones e importaciones, ventas, presupuestos, etc.

Una *serie de tiempo* o *crono ógica* es un conjunto de observaciones ordenadas respecto a una característica cuantitativa de un fenómeno individual o colectivo, que se toman en diferentes períodos de tiempo (diario, semanal, mensual, etc...)

Las series de tiempo como el análisis de regresión, corresponden a distribuciones *bidimensiona es*, o *bi ariantes* es decir, se trabaja y analizan conjuntamente dos variables, salvo que en este caso una de ellas corresponde al *tiempo* que podría considerarse como la variable independiente o explicativa, y que se simboliza por X o t; la otra variable es la que se va a estimar, ya sea dentro de la serie (interpolar) o su comportamiento futuro (extrapolar), simbolizado por Y, y puede corresponder, como por ejemplo *producción e portaciones entas horas trabajadas, inversión, matrícula, defunciones, nacimientos, mo imiento de carga* etc. Se trata de una regresión unilateral, solo se podrá estimar a $\hat{Y}$ en función de X pero no al contrario.

Vale la pena mencionar que las proyecciones o la tendencia en el futuro, deben hacerse para períodos cortos de uno o dos años, máximo cinco años, bajo el supuesto de que las condiciones dadas en la serie van a seguir siendo iguales que en el presente; imposible que se mantenga en períodos largos, a fin de que no se produzcan diferencias entre lo esperado y su comportamiento real.

COMPONENTES DE UNA SERIE TEMPORAL

Los movimientos que presenta una *serie de tiempo*, son producidos por una variedad de factores de carácter económico, naturales o institucionales. El análisis de estas series consiste en descubrir y cuantificar dichas influencias, ya sean internas o externas, estableciendo la evolución que han tenido y su comportamiento futuro.

Generalmente, se han venido considerando cuatro posibles componentes: *tendencia estaciona idad ariaciones cíc icas* y *a eatorias*, de los cuales alguno o todos pueden estar afectando la serie de tiempo.

endencia Son aquellas variaciones suaves y constantes que se suceden en un período relativamente largo. El período debe ser largo como para establecer una línea de tendencia que sea significativa.

ariaciones estaciona es Generalmente ligadas a las estaciones del año, también corresponden a cambios periódicos que se repiten en lapsos de tiempo más cortos, como por ejemplo el movimiento de pasajeros urbanos en un dia, consumo de energía, asistencia a cines, discotecas, etc. Se puede determinar cuándo se inicia y termina el período de variación.

ariaciones cíc icas Son fluctuaciones a largo plazo, más o menos periódicas, que se repiten regularmente cada cierto número de años. Se refiere a las fluctuaciones cíclicas en la actividad económica (recuperación - auge - declinación - crisis). Es difícil reconocer la longitud del ciclo, ya que su período es bastante largo y difícil de precisar en algunos casos, alcanza hasta los 40 o más años, su longitud.

ariaciones a eatorias Son aquellos factores que se presentan en forma accidental siendo difíciles de predecir, tales como: huelgas, crisis políticas, inundaciones, sequías, etc.

Algunos autores consideran que la variable Y es el resultado de la suma de dichos factores, de tal manera que: $Y = T+VE+VC+VA$; para otros es el resultado del producto de los factores: $Y = T \times VE \times VC \times VA$. Sin embargo, la variable Y puede ser considerada como el resultado de la combinación de suma y producto de estos factores $Y = T(VC+VE+ VA)$.

La diferencia entre el *mode o aditi o* y el *mu tip icati o* es que, en el primero los componentes se consideran como residuos y se expresan en las unidades originales; en el segundo, por lo general, la tendencia se expresa en unidades originales y los otros componentes en términos porcentuales o relativos.

TENDENCIA

Movimiento a largo plazo de una serie ordenada cronológicamente, es el componente más utilizado para realizar estimaciones.

En el análisis de la *tendencia*, se considera importante iniciar con la graficación, con el fin de determinar la línea que mejor representa a esa poligonal. Puede decirse, que no es posible obtener en ningún caso una comprobación absoluta de la bondad de la adaptación de una línea, pues solamente el juicio personal y la experiencia del estadístico, son los elementos que pueden ponerse en juego en la elección del mejor ajuste.

Para la determinación de la *tendencia*, es decir, la obtención de una línea más sencilla que la poligonal que nos permita estimar el comportamiento futuro, aplicaremos el método de los *mínimos cuadrados* con el objeto de cuantificar los parámetros en la ecuación correspondiente: *recti ínea parabó ica e ponencia* o cualquier otra línea que consideremos representativa de la serie, o que mejor describa la tendencia.

MÉTODO ANALÍTICO

Consiste en determinar la ecuación matemática adecuada, que nos permita describir de la mejor forma posible la *tendencia*, a fin de poder obtener buenos estimativos. De ahí la importancia de graficar los datos originales, con la cual podemos visualizar, en la mayoría de los casos, la línea que nos va a describir la tendencia.

Tendencia lineal

Corresponde a la función más utilizada en el análisis de la tendencia, especialmente cuando la serie es corta y cuando la variable presenta variaciones bastante fuertes. La ecuación general se puede representar indistintamente por:

$\hat{Y} = b + c$ $\quad$ $\hat{Y} = \beta_1 X + \beta_0$ $\quad$ $\hat{Y} = \beta X + A$

Donde:

$\hat{Y}$ = Es el *a or estimado*, ya sea que se haya hecho una interpolación o una extrapolación.

b = β_1 = B Es la *pendiente*, que nos indica el crecimiento o decrecimiento por cada unidad de tiempo. Se conoce como *coeficiente angular*.

c = β_0 = A Es la intersección con la ordenada o eje de Y, en un plano cartesiano. Se conoce también como *coeficiente de posición*.

= Variable tiempo o *ariab e independiente* (días, meses, años, etc.).

Para el cálculo de los coeficientes **b** y **c**, se hace con un sistema de ecuaciones, operadas simultáneamente por ejemplo, a traves el método de los mínimos cuadrados. Veámoslo con un ejemplo:

Ejemplo 1. Con la siguiente serie (*continua* e *impar*), tomando como origen el primer período de la serie, estimar el valor $\hat{Y}$ para el año 2015

Solución:

AÑOS	$_i$
2005	24
2006	29
2007	36
2008	42
2009	56
2010	60
2011	67
Σ	314

x_i	x_i	$x_i\ y_i$
0	0	0
1	1	29
2	4	72
3	9	126
4	16	224
5	25	300
6	36	402
21	91	1.153

Sistema de ecuaciones:

$$(a)\quad \sum y_t = b\sum x + nc$$
$$(b)\quad \sum y_t x_t = b\sum x_t^2 + c\sum x_t$$

Se reemplaza:

$$(a)\quad 1 = 1b + c$$
$$(b)\quad 1\ 1 = 1b + 1c$$

Aplicando algunos de los métodos conocidos de: *igua ación*, *sustitución* o *e iminación*, obtenemos los valores para y ***c***. Utilizaremos el último, para ello multiplicamos la ecuación (a) por menos 3 y se le resta a la ecuación (b).

$$(b)\quad 1\ 15 = 91b + 21c$$
$$(a)\quad -9\ 2 = -6\ b - 21c$$
$$211 = 28b\quad 0 \qquad \Rightarrow\ b = \frac{211}{28} = 7{,}5 \qquad b = 7{,}54$$

Reemplazamos en la ecuación (a): $314 = 21\ (7{,}54) + 7c$

C = 22,24 $\qquad c = \dfrac{1\ -21\ (7{,}5\ )}{7} = 22{,}2$

Se tendrá que la ecuación de la recta será: $\hat{Y} = 7{,}54\ \mathrm{X} + 22{,}24$
Si se va a estimar $\hat{Y}$ para el año 2015, se tendrá que el valor de = 2015 – 2005 = 10 (se tomó el año 2015 por ser = 0)

De donde: $\hat{Y}_{15} = 7{,}54(10) + 22{,}24 = 97{,}64$

($\hat{Y}_{15} = 97{,}61$ trabajando con todos los decimales, usando la calculadora).

Otro procedimiento de cálculo. El procedimiento de cálculo se puede agilizar, utilizando fórmulas que permiten obtener resultados en forma directa.

$$b = \frac{n\sum x_t y_t - (\sum x_t)(\sum y_t)}{n\sum x_t^2 - (\sum x_t)^2} \qquad b = \frac{7\ (1\ 15\ ) - (21)(\ 1\ )}{7\ (91) - (21)^2} \cong 7{,}5 \quad (\qquad)$$

$$c = \frac{\sum y_t - b\sum x_t}{n} \qquad c = \frac{1\ -7{,}5\ \ (21)}{7} = 22{,}2 \quad (\qquad)$$

Varian a residual o no explicada (VR). Determina la dispersión o concentración de aquellos puntos, que no quedan sobre la recta. Ya sabemos, cómo en la regresión, si la VR es cero, todos los puntos quedan sobre la recta estimada, lo que equivale a decir, que no hay ninguno fuera de ella y el coeficiente de correlación es igual a uno.

Los símbolos utilizados son tantos, entre ellos: VR s^2 s_e^2 etc; así como la presentación de las fórmulas de cálculo:

1) $s_{y.x}^2 = \dfrac{\sum (y_t - Y_t)^2}{n}$ 2) $s_{y.x}^2 = \dfrac{\sum y_t^2 - c\sum y_t - b\sum x_t y_t}{n}$ 3) $s_{y.x} = s_{y.}\ (1 - r\ )$

Con los datos del Ejemplo 1, aplicaremos las anteriores fórmulas:

AÑOS	$_i$
2005	24
2006	29
2007	36
2008	42
2009	56
2010	60
2011	67
Σ	314

^2_i	$\hat{Y}_i$	$(\ _i - \hat{Y}_i)^2$
576	22,25	3,0625
841	29,79	0,6241
1.296	37,32	1,7424
1.764	44,86	8,1796
3.136	52,39	13,0321
3.600	59,93	0,0049
4.489	67,46	0,2116
15.702	314,00	26,8572

$\Sigma y_t = \Sigma Y_t$

(1) $VR = \dfrac{26{,}8572}{7} = \ ,8$

(2) $VR = \dfrac{15\,702 - 22{,}2\ (\ 1\ ) - 7{,}5\ (1\ 15\ )}{7} = \ ,57$

( ) $VR = 2\ \ 0{,}98(1 - 0{,}98\ \ ) = \ ,8$

Nota: por considerar que son muestras pequeñas, algunos, en vez de dividir por 7, lo hacen por 5, es decir trabajan con n - 2. Las diferencias que se observan en los resultados anteriores, se deben a los decimales que se omiten, pues hemos aproximado a dos decimales.

La desviación típica residual. Lo mismo que se dijo en el cálculo de la desviación típica de la media, en la regresión lineal, debe obtenerse mediante la raíz cuadrada de la varianza residual.

$$s_{y.x} = s_e = \sqrt{VR} \qquad s_{y.x} = \sqrt{\quad} = 1$$

LÍMITES DE CONFIANZA

El estimativo obtenido no se da como cifra exacta, es decir, el analista, espera que se cumpla la predicción, lo cual es difícil que en la realidad suceda, aún cuando se haya utilizado la línea adecuada a la serie analizada. La anterior razón, indica la necesidad de hacer la estimación por intervalos, o sea fijando límites dentro de los cuales esté el valor real, con una seguridad o confianza del 95% o cualquier otro porcentaje que establezca el analista; para ello se tendrá que:

$$Y \pm t\frac{s_{y.x}}{\sqrt{n}} = \begin{cases} Y_s\ (\qquad\qquad\qquad\qquad\ ) \\ Y_t\ (\qquad \text{inferior} \qquad\ ) \end{cases}$$

$\upsilon = n - 1 = 6$
$\propto = 0{,}05$
t = 2,447

$$Y_{15} = 97{,}6 \pm 2,\quad 7\ \frac{1{,}89}{\sqrt{7}} = 97{,}6 \pm 1{,}75 = \begin{cases} 97{,}6 + 1{,}75 = 99,\ 9 \\ 97{,}6 - 1{,}75 = 95{,}89 \end{cases}$$

De acuerdo con el comportamiento de la serie durante el período 2005- 2011, se considera que en el futuro no se alterará. Bajo estas condiciones se predice que en el 2015 el valor de Y (producción miles tons, ventas (miles mill. de $), etc.) será de 97,64, sin embargo, con una confianza el 95 % el valor esperado para ese año estará entre 95,89 y 99,39, por lo tanto se tendrá la posibilidad del 5% que no quede incluido dentro del intervalo. Como se verá enseguida, el *coeficiente de correlación* de 0,99, nos estará indicando que el método matemático lineal es muy bueno para esta proyección.

Coeficiente de correlación. Nos indica la bondad del ajuste lineal que se ha hecho. Denominado también como coeficiente de *corre ación de Pearson*, donde: $-1 \le r \le 1$; siendo positivo cuando la recta es ascendente y negativo cuando es descendente.

La fórmula más utilizada que aparece en la mayoría de los textos de estadística es:

$$r = \frac{n\Sigma x_t y_t - (\Sigma x_t)(\Sigma y_t)}{\sqrt{\left[n\Sigma x_t^2 - (\Sigma x_t)^2\right]\left[n\Sigma y_t^2 - (\Sigma y_t)^2\right]}}$$

$$r = \frac{7(1\ 15\ ) - (21)(\ 1\ )}{\sqrt{\left[7(91) - 21^2\right]\left[7(15\ 702) - \ 1\ ^2\right]}} = \frac{1\ \ 77}{\sqrt{(196)(11\ \ 18)}} = 0{,}99$$

Si r se eleva al cuadrado, se denomina *coeficiente de correlación al cuadrado* o *coeficiente de determinación* Estos coeficientes también se pueden calcular de la siguiente forma:

$R^2 = 1 - \frac{VR}{s_y^2} \quad \Rightarrow \quad R = 1 - \text{———} = \quad \Rightarrow \quad r =$

$s_y^2 = \frac{\Sigma y_t^2 - n\left(\frac{\Sigma y_t}{n}\right)^2}{n} \qquad s_y^2 = \frac{15\,702 - 7\left(\frac{1}{7}\right)^2}{7} = 2\ 0{,}98$ (Varianza de y)

Recordemos que: $0 \leq R^2 \leq 1$ y $-1 \leq r \leq 1$

Otro procedimiento para el cálculo de la tendencia lineal. En el Ejemplo 1, se consideró como origen, el primer año de la serie, siendo $\Sigma\ _1 = 0$.

Sin embargo, cualquier período de la serie, se puede tomar como origen siendo lo más práctico en una serie *continua e impar*, ubicarlo en el centro de la serie, donde $\Sigma\ _1 = 0$ y con ello agilizaremos, simplificando las operaciones. Veámoslo:

Ejemplo 2.

AÑOS	$_i$	$_i$	^2_i	$_i\ _i$	^2_i
2005	24	–3	9	–72	576
2006	29	–2	4	–58	841
2007	36	–1	1	–36	1.296
2008	42	0	0	–166/377	1.764
2009	56	1	1	56	3.136
2010	60	2	4	120	3.600
2011	67	3	9	201	4.489
Σ	314	0	28	211	15.702

$b = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \qquad b = \frac{211}{28} = 7{,}5$ (Exactamente igual al resultado dado en el Ejemplo 1)

$c = \frac{\Sigma y_i}{n} \qquad c = \frac{1}{7} = \ \ ,86$ El valor de C cambia, dependiendo del origen que se tome.

$\hat{Y}_{15} = 7{,}54 \ + 44{,}86 \ \Rightarrow \ \hat{Y}_{15} = 7{,}54\,(7) + 44{,}86 = 97{,}64$

$\left(\ = 2015 - 2008 = 7 \right)$

Así como éste y los demás resultados obtenidos con el ejemplo 2, serán exactamente iguales a los del ejemplo 1, sólo cambian los valores de algunas sumatorias.

Ahora, observemos una *serie continua con n mero par de períodos* En primer lugar, diremos que hay datos anuales que no se pueden presentar como la suma de valores mensuales, sino como el promedio mensual, que a su vez es anual, como por ejemplo, el precio de un artículo en cuatro meses diferentes en un año, siendo $ 5.000, $ 8.000, $ 6.000 y $ 10.000 respectivamente; en este caso el valor promedio mensual es de $ 7.250, al mismo tiempo es el promedio del valor del artículo para ese año.

Ejemplo 3.

AÑOS	$_i$	$_i$	^2_i	$_i\ _i$	^2_i
2004	56	–7	49	–392	3.136
2005	42	–5	25	–210	1.764
2006	70	–3	9	–210	4.900
2007	82	–1	1	–82	6.724
2008	41	1	1	41	1.681
2009	63	3	9	189	3.969
2010	110	5	25	550	12.100
2011	84	7	49	588	7.056
Σ	548	0	168	474	41.330

$b = \dfrac{\Sigma x_t y_t}{\Sigma x_t^2}$ $\quad b = \dfrac{7}{168} = 2{,}82$ $\quad \hat{Y} = 2{,}82\mathrm{x} + 68{,}5$

$c = \dfrac{\Sigma y_t}{n}$ $\quad c = \dfrac{5\ 8}{8} = 68{,}5$ $\quad \hat{Y}_{13} = 2{,}82\ (11) + 68{,}5 = 99{,}52$

En el cuadro anterior, en la columna de $_i$ se puede observar que hay dos períodos en el centro de la serie 2007 y 2008, suponiendo que la información está dada al primero de julio en cada uno de los años, se tendrá que en el primero de enero de 2008, estará localizado el origen donde = 0, y como toda información anual supuestamente estará dada al primero de julio, se tendrá para el 2008 un semestre y como cada año tiene dos semestres, se irá contando de dos en dos, así para el 2009 ⇒ = 3; 2010 ⇒ = 5 y así sucesivamente. Hacia arriba será lo mismo pero con signo negativo.

El valor de en la ecuación se calculará así: = 2013 − 2008 = 5 años × 2 = 10 semestres, luego se tendrá que = 10 + 1 = 11 semestres. El + 1 se debe al primer semestre de 2008. Procedamos de otra manera: = 2013 − 2011 = 2 años × 2 = 4 semestres + 7, que aparece en el 2011 = 11 semestres. Como en el caso anterior, calcularemos la *arian a residua* y el *error estándar* así:

$s^2_{y.x} = \dfrac{\Sigma y_t^2 - c\Sigma y_t - b\Sigma y_t x_t}{n}$ $\quad s^2_{y.x} = \dfrac{1\quad 0 - 68{,}5(5\ \ 8) - 2{,}82(\ 7\ )}{8} = \ 06{,}92$

$s^2_{y.x} = s^2_y\left(1 - r^2\right)$ $\quad s^2_{y.x} = \ 7\ \left(1 - 0{,}59\ \ 8^2\right) = \ 06{,}8$, dejamos que el estudiante aplique la fórmula

$s_{y.x} = VR = \dfrac{\Sigma\left(y_t - Y_t\right)}{n}$, y compare los resultados, los cuales deberán ser aproximadamente iguales.

Ahora obtenemos la desviación típica residual: $s_{y.x} = \sqrt{s^2_{y.x}} = \sqrt{\ 06{,}8\ } = 17{,}52$

Los límites de confianza del 95%, serán:

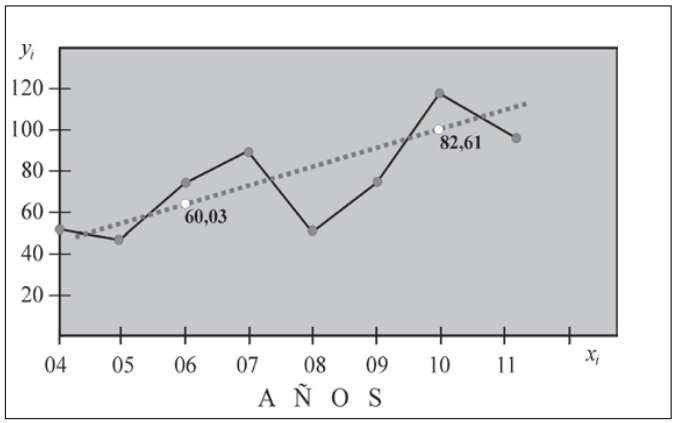
$Y \pm t\dfrac{s_{yx}}{\sqrt{n}} \Rightarrow 99{,}52 \pm 2,\quad 7\ \dfrac{17{,}52}{\sqrt{8}} \Rightarrow 99{,}52 \pm 15{,}16 = \begin{cases} 11\ ,68 \\ 8\ ,\ 6 \end{cases}$

$\upsilon = n - 2 = 6$ y $\propto = 0{,}05 \quad \Rightarrow t = 2{,}447 \qquad \upsilon = n - 1 = 7$ y $\propto = 0{,}05 \quad \Rightarrow t = 2{,}365$

Representación gráfica

Se le dan dos valores a $_i$ en la ecuación, por ejemplo: cuando = - 3 y = 5, los resultados serán para esos valores $\hat{Y} = 60{,}03$ y $\hat{Y} = 82{,}61$ respectivamente; tomamos un cuadrante, colocando en el

eje horizontal (abscisa) los años y en el eje vertical (ordenada) la variable *Y*. Establecemos los dos puntos de $\hat{Y}$ de 60,03 para 2006 y de 82,61 para 2010. Estos dos puntos los unimos y nos dará la recta que deseamos representar. Además, el coeficiente de correlación obtenido en el ejercicio es de $R^2 = 0{,}35$ y $r = 0{,}59$, lo cual nos indica que el ajuste rectilíneo no es el más indicado para esta serie, ya que ella se confeccionó con él a fin de mostrar variaciones grandes durante el período investigado.

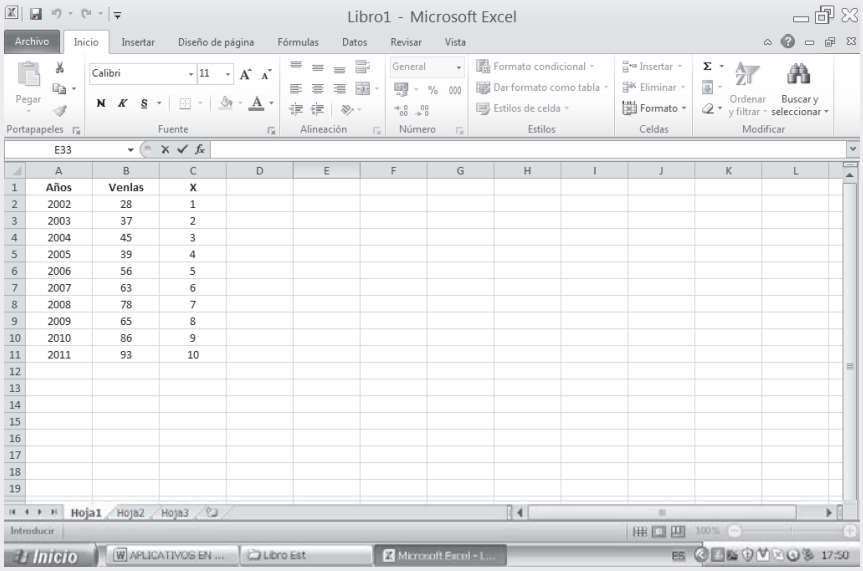


$$r = \frac{n\Sigma x_t y_t}{\sqrt{\left[n\Sigma x_t^2\right]\left[n\Sigma y_t^2 - (\Sigma y_t)^2\right]}} = \frac{8(\ 7\ )}{\sqrt{[8(168)]\left[8(\ 1\quad 0) - (5\ \ 8)^2\right]}} = 0{,}59$$

APLICACIÓN DEL PROGRAMA EXCEL EN SERIES DE TIEMPO O SERIES CRONOLÓGICAS

Ejemplo 3. Supongamos que hemos digitado en una *hoja de cálculo,* información acerca de la variable ( ), durante un período de 10 años ( ).

Figura No. 1. Información en la Hoja de Cálculo.



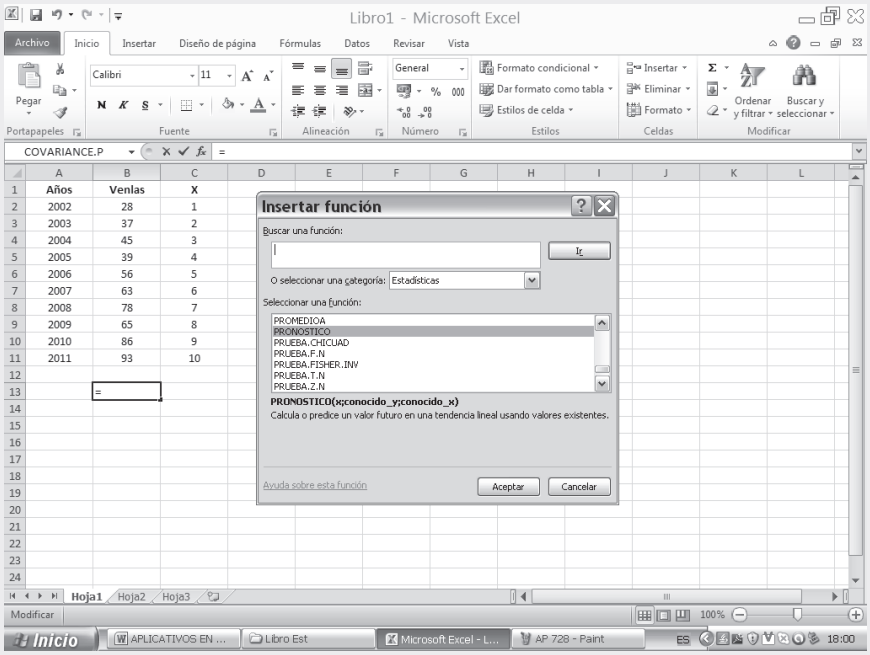
	A	B	C
1	Años	Venlas	X
2	2002	28	1
3	2003	37	2
4	2004	45	3
5	2005	39	4
6	2006	56	5
7	2007	63	6
8	2008	78	7
9	2009	65	8
10	2010	86	9
11	2011	93	10

En la SEGUNDA PARTE DEL SIL encontrará en conjunto toda la aplicación del EXCEL a la ESTADÍSTICA, ofreciendo una BASE DE DATOS para practicar.

Las VENTAS corresponden a la variable Y, la cual vamos a proyectar, es decir, estimar su comportamiento futuro. Esta variable esta dada en miles de millones de pesos y su rango será B2:B11. Se desea realizar la estimación de las ventas para el año 2013, así:

- Seleccionamos una celda como RANGO DE SALIDA, supongamos que el resultado de $\hat{Y}_{13}$ deberá aparecer en la celda B13.
- Hacemos CLIC en la función *fx*. En el cuadro de diálogo que aparece seleccionamos la opción PRONÓSTICO y luego CLIC en ACEPTAR.

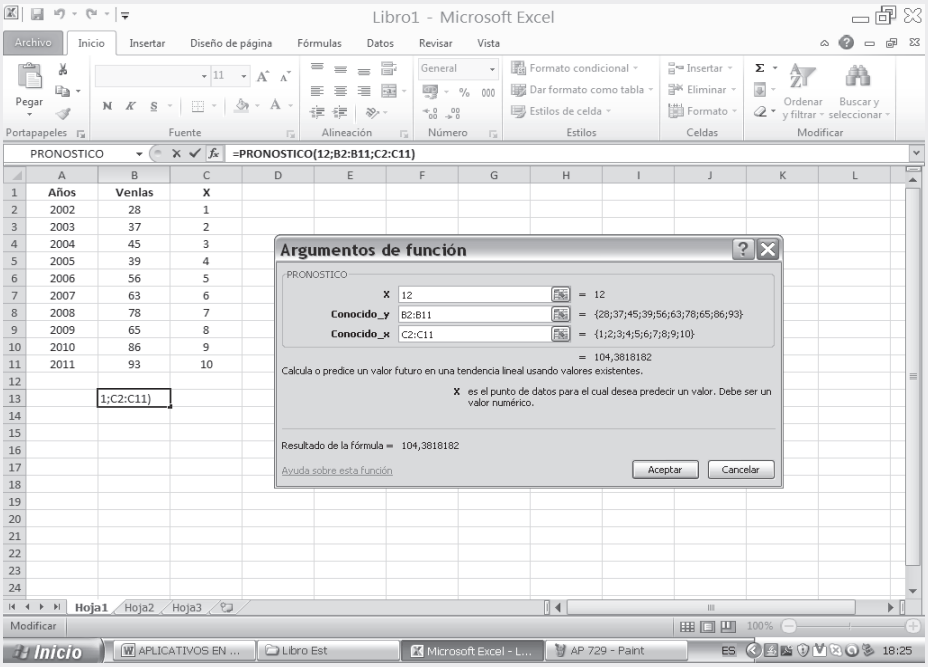
Figura No. 2. Insertar Función.



Observemos que en el nuevo cuadro de diálogo PRONÓSTICO, aparecen tres casillas las cuales debemos teclear de la siguiente forma:

- El campo X, corresponde a los años que se han establecido para el PRONÓSTICO. Contando a partir del año 2002 y teniendo en cuenta que nuestro pronóstico lo establecimos para el 2013, X corresponde a la cantidad de años que transcurren desde el 2002 hasta el 2013, es decir, X tomaría un valor de 12.
- En el segundo campo correspondiente a CONOCIDO _Y, seleccionamos el rango de datos, B2:B11.
- En el tercer campo CONOCIDO_X, seleccionamos el rango de datos C2:C11.
- Finalmente hacemos CLIC en ACEPTAR.

Figura No. 3. Argumentos de la Función.



	A	B	C
1	Años	Ventas	X
2	2002	28	1
3	2003	37	2
4	2004	45	3
5	2005	39	4
6	2006	56	5
7	2007	63	6
8	2008	78	7
9	2009	65	8
10	2010	86	9
11	2011	93	10
12			
13		1;C2:C11)	

Observemos que el resultado pronosticado para las ventas del año 2013 serán de 104.38 millones.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. Con los siguientes datos se pide:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
Producción:	360	383	337	390	406	459	480

a) Estimar la producción (Miles Tons) para el 2024 ajustando una recta (tomando como origen al 2005).

b) Estimar la producción (Miles Tons) para el 2009 ajustando una recta (trabajar con cambio de origen).

c) Estimar el mismo año y utilizar el mismo sistema del punto (b) pero agregando a la serie el año 2012 = 600.

(Nota: El cambio de origen, consiste en trasladarlo al centro de la serie)

2. Con los datos del ejercicio 1 y los resultados de (b), se pide:

a) Fijar límites de confianza del 95 %, para $\hat{Y}_{19}$

b) Calcular el coeficiente de correlación de Pearson.

3. Con los índices de precios anuales, se pide:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
Promedios:	128	135	148	145	152	161	162

a) Estimar el precio promedio para el 2013.

b) Obtener el error estándar de estimación y fijar límites de confianza del 95 %.

c) Calcular el coeficiente de correlación.

4. Se dan los siguientes datos:

AÑOS	$_i$	$\hat{Y}_i$
2007	$_1$	
2008	$_2$	
2009	550	580
2010	600	635
2011	750	

Se pide:

a) Estimar el valor de $\hat{Y}$ para el año 2020 mediante un ajuste rectilíneo.

b) Calcular el valor de $_1$ y $_2$.

5. De acuerdo con los siguientes datos.

AÑOS	$_i$	$\hat{Y}_i$
2007	50	
2008	$_2$	
2009	100	90
2010	$_4$	
2011	140	148

Se pide:

a) Estimar el valor de $\hat{Y}$ para el año 2021 mediante un ajuste rectilíneo.

b) Calcular el valor de $_2$ y $_4$.

6. Se conocen los siguientes datos:

AÑOS	$_i$	$\hat{Y}_i$
2007	200	0
2008	-	-
2009	-	-
2010	320	-
2011	380	-
Σ	-	3.260

Se sabe adicionalmente que la media de las observaciones es 280. Se pide reconstruir el cuadro anterior, además estimar el valor de $\hat{Y}$ para el año 2018.

7. Con los siguientes datos, sobre valores de las ventas en Miles Mill. $

AÑOS:	2005	2006	2007	2008	2009	2010	2011
$_i$:	42	64	82	110	96	84	124

Se pide:

a) Estimar 2007 y 2013.

b) Calcular el coeficiente de correlación.

8. Con los siguientes datos se pide:

AÑOS:	1991	1995	1999	2003	2007	2011
Producción:	12	30	89	114	140	200

Mediante un ajuste rectilíneo trabajando con cambio de origen, estimar la producción (miles Tons.) para 1993 y 2016.

9. El volumen de ventas que anualmente se hace en unos grandes almacenes desde 2000 hasta 2011, presenta una tendencia creciente según la ecuación.

$$\hat{Y} = 0{,}11 \quad + 1{,}52 \qquad R^2 = 0{,}87$$

Siendo Y el volumen de ventas (en millones) y la variable tiempo con origen en el punto medio entre 2000 y 2011 (unidad temporal: medio año). Se pide estimar el volumen de ventas para el 2018 y 2021

10. ¿Puede aceptarse la recta de tendencia secular $\hat{Y} = 2 \ +35 \qquad R^2 = 0{,}99 \qquad (\ = 0$ en 2006; unidad temporal: medio año) para predecir la producción de un bien para el año 2021?

11. La variable Y toma para cada unidad de tiempo, los siguientes valores: 10, 12, 13, 15, 20. Calcular el incremento medio de la variable Y por unidad de tiempo.

12. Se sabe que el desarrollo de la variable Y en el tiempo, tiene una tendencia lineal, pero de ella tan sólo se conocen los siguientes valores:

AÑOS:	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
$_i$:	8	-	-	16	14	-	-	20	-	22	22

Se puede estimar los valores desconocidos.

13. Con la siguiente información sobre el valor de las exportaciones, en millones de pesos, durante varios años:

$\Sigma \ _i = 0 \qquad \Sigma \ _i = 1.620 \qquad \Sigma \ _i^2 = 314.502$

$\Sigma \ _i^2 = 60 \qquad \Sigma \ _i \ _i = 1.148 \qquad = -2, \ en \ e \quad 2005$

Se pide:

a) Estimar el valor de las exportaciones en millones de pesos para el 2017, fijando límites de confianza del 95 %.

b) Determinar el coeficiente de correlación lineal.

14. Siendo:

$\Sigma \ _i = 0 \qquad \Sigma \ _i = 2.032 \qquad \Sigma \ _i^2 = 509.992$

$\Sigma \ _i^2 = 60 \qquad \Sigma \ _i \ _i = 1.723 \qquad n = 9$

El año base de la serie es 2008. Se pide:

a) Calcular el coeficiente de correlación lineal

b) Estimar el valor de Y para el 2015.

15. Los gastos personales (miles de millones de pesos) durante varios años, en una región cualquiera, fueron los siguientes:

AÑOS:	2007	2008	2009	2010	2011
Gastos personales:	230	285	310	325	360

Se pide:

a) Ajustar una recta, por el método de los mínimos cuadrados, determinando en qué año los gastos personales alcanzan los $602 miles de millones.

b) Estimar los gastos personales para el mes de septiembre de 2015.

16. Los gastos (cientos millones de pesos) de una Oficina Pública durante el período 2007–2011 fueron:

AÑOS:	2007	2008	2009	2010	2011
Gastos personales:	100	150	200	200	250

a) Supuesta una tendencia lineal; estime los gastos de la oficina para 2013.

b) Calcule el valor de r.

17. Se hizo una encuesta en Bogotá sobre las ventas de carros usados, en la década 2002 - 2011. Se ajustó una curva de tendencia para las ventas de cada año, obteniéndose la siguiente ecuación:

$\hat{Y} = 900 + 40$ Donde = 0 para el primero de enero de 2007.

(Si consideramos que está medida en unidades semestrales). Se pide predecir las ventas anuales para 2013.

18. Se ajustó una recta de tendencia, para los egresados de la facultad de Economía de esta universidad, durante seis años 2006 - 2011; por el método de los mínimos cuadrados.

a) La ecuación mostró que en 2006 el salario sería de $896.000 y de $ 1‘692.000 en 2011; encontrar la ecuación con origen en 2007 y estimar el promedio mensual de salario para los egresados en el 2013.

19. Con la siguiente información sobre el valor (millones de pesos) de las exportaciones.

$\Sigma_{\ i} = 0$ $\Sigma_{\ i} = 244$ $\Sigma_{\ i\ i} = 233$ $\Sigma_{\ i}^{\ 2} = 8.118$ $n = 9$ $\Sigma_{\ i}^{\ 2} = 60$

■ Donde = –2 se refiere al año 2005.

Se pide:

a) Estimar el valor (millones de pesos) de las exportaciones para el 2015, estableciendo límites de confianza del 95 %.

b) Hacer las anteriores estimaciones, trabajando con cambio de origen y cálcular el coeficiente de correlación lineal..

20. Las ventas anuales (miles millones de pesos) desde 2002 hasta 2011 arrojan la siguiente información:

$\Sigma_{\ i} = 45$ $\Sigma_{\ i} = 544$ $\Sigma_{\ i}^{\ 2} = 285$ $\Sigma_{\ i\ i} = 3.266$ $\Sigma_{\ i}^{\ 2} = 38.040$

a) Grafique la línea correspondiente a $\hat{Y}$.

b) Determine el coeficiente de correlación lineal.

21. Los pasajeros del interior del país llegados por vía aérea a la capital, durante 6 años han sido.

AÑOS	PASAJEROS
2006	723.000
2007	982.500
2008	1.236.420
2009	1.450.600
2010	1.636.250
2011	1.890.900

Suponiendo que la serie se ajusta a una recta

a) ¿Cuál es el número de pasajeros, por vía aérea, esperados para el 2016 y 2021?
b) Calcule el coeficiente de correlación.
c) Fije los límites de confianza del 95% para los estimados del punto a).

22. Los turistas extranjeros llegados a una ciudad *X* han sido, en los últimos cinco años:

AÑOS	PASAJEROS
2007	417.500
2008	572.800
2009	620.350
2010	679.260
2011	712.800

Se ha determinado que el 60 % de estos turistas requieren alojamiento en establecimientos hoteleros y que el 40 % requieren alojamiento en hoteles de cinco estrellas.

a) Determinar la demanda total de alojamiento en el 2017.
b) La demanda de alojamiento para hoteles de cinco estrellas en el 2017.

23. Los gastos de una empresa durante 6 años fueron:

AÑOS	GASTOS
2006	2.460
2007	2.700
2008	2.850
2009	2.950
2010	3.150
2011	3.300

Se pide:

a) Ajustar una recta por el método de los mínimos cuadrados, para estimar $\hat{Y}_{13}$.
b) Calcular el coeficiente de correlación.

24.

AÑOS:	1987	1991	1995	1999	2003	2007	2011
$_i$:	120	180	100	260	370	250	460

Estimar el valor de $\hat{Y}$ para el 2015, tomando como origen 1987.

25. Resolver el ejercicio anterior trabajando con cambio de origen de tal manera, que $\Sigma\ _i = 0$.

26. Calcule el coeficiente de correlación de Pearson, con los datos del ejercicio 25.

Estimaciones mensuales con base en datos anuales

Anteriormente se trabajó con períodos anuales, pero esto no impide que la serie corresponda a valores semanales, mensuales, trimestrales o información para cada 2 o 3 años, etc., de todas maneras el proceso será exactamente igual, con la diferencia de que X corresponderá a días, semanas, meses, trimestres, semestres, etc.

Ahora consideraremos el caso donde se tenga una serie con informaciones anuales y que sea necesario hacer estimaciones mensuales; es posible realizarlo siempre y cuando se siga el procedimiento adecuado, de la siguiente manera:

Al estimar valores mensuales con base en datos anuales, hay que considerar dos situaciones:

a) Datos anuales correspondientes a promedios de valores mensuales (ejemplo típico: precios).
b) Datos anuales correspondientes a suma de valores mensuales (ejemplo: producción).

El procedimiento a seguir en la estimación de valores mensuales, con base en datos anuales mediante el ajuste rectilíneo, se indica a continuación:

SERIE	PRECIOS		PRODUCCIÓN		FECHAS
X_i	b	c	b	c	Inicio
Semestral	Dividir por 6	Igual	Dividir por 72	Dividir por 12	1º de Enero
Anual	Dividir por 12	Igual	Dividir por 144	Dividir por 12	1º de Julio

Utilizaremos los ejemplos desarrollados anteriormente, correspondiente a series continuas par e impar, ya sea con cambio de origen ($\sum_i = 0$) o cuando es diferente a cero ($\sum_i \neq 0$).

Ejemplo 4. Consideremos la serie 2005 – 2011 del Ejemplo 1, para estimar los meses de septiembre/08; febrero/12, siendo Y_i la variable *producción*.

Solución

Se obtuvo la ecuación $\hat{Y} = 7{,}54 + 22{,}2$ para realizar estimaciones anuales. De acuerdo a la tabla anterior, transformamos la ecuación para realizar estimaciones mensuales:

$$Y = \frac{b}{1}\left(\frac{x}{1}\right) + \frac{c}{1} \quad \Rightarrow \quad Y = \frac{b}{1}x + \frac{c}{1}$$

$$Y = \frac{}{1}x + \frac{}{1} \quad \Rightarrow \quad Y = \quad x + 1$$

Si le damos a un valor de cero, se tendrá $\hat{Y} = 1{,}853$ para el 1º de Julio de 2005, pero la información se debe dar al 15 de cada mes; una forma sería: $\hat{Y} = 0{,}052\ (0{,}5) + 1{,}853 = 1{,}9055$ y así para todos los meses que se van a estimar, pero es mejor acondicionar la ecuación al quince de cada mes, agregándole al coeficiente de posición, la mitad (1/2 mes) del coeficiente angular.

$\hat{Y} = 0{,}052 + 1{,}9055 \rightarrow$ 15 de julio de 2007.

Ahora se tiene que Julio → = 0 Agosto → = 1 Sep/05 → = 2 Sep/06 → = 14
Sep/07 → = 26 y Sep/08 → = 38

$\hat{Y}_{Sep/08} = 0{,}052\ (38) + 1{,}9055 = 3{,}8815$ es el estimado de la producción (miles tons.)

Para la estimación de Feb/12 se tiene que Sep/05 → = 2 Oct/05 → = 3 Nov/05 → = 4 Dic/05 → = 5 Enero/06 → = 6 Febrero/06 → = 7, de aquí contamos de 12 en 12, número de meses por año: Feb/07 → = 19 Feb/08 = 31 Feb/09 → = 43 Feb/10 → = 55 Feb/11 → = 67 y Feb/12 → = 79.

$\hat{Y}_{\text{Feb/12}} = 0{,}052\ (79) + 1{,}879 = 5{,}987$

Ejemplo 5. Consideremos la ecuación obtenida, trabajando con los datos del ejemplo 2. Cuando la variable es considerada como *producción* y luego como *precios*:

So ución Ecuación para años: $\hat{Y} = 7{,}54 \quad + 44{,}86$

Producción - estimación mensual: $\hat{Y} = \frac{7{,}54}{144} \quad + \frac{44{,}86}{12} \Rightarrow \hat{Y} = 0{,}052 \quad + 3{,}738$

Cuando = 0, se tendrá $\hat{Y} = 3{,}738$ al 1º de julio/08 y al 15 de julio será: $\hat{Y} = 0{,}052 \quad + 3{,}764$

Al estimar Sep/08 se tendrá para agosto/08 → = 1; sep/08 → = 2.

$\hat{Y}_{\text{Sep/08}} = 0{,}052(2) + 3{,}764 = 3{,}868$ (miles Tons.)

$\hat{Y}_{\text{Feb/12}} = 0{,}052(43) + 3{,}764 = 6{,}00$ (miles de Tons.) se estima para feb/12.

Si usted estima $\hat{Y}_{12} = 7{,}54(4) + 44{,}86 = 75{,}02$ (Miles Tons.) y luego con la ecuación para estimaciones mensuales $\hat{Y} = 0{,}052x + 3{,}764$, se obtienen los valores para los doce meses (Enero, Feb....Dic) correspondiente al año 2012 y esos resultados los sumamos, debe dar aproximadamente 75,02 (miles Tons.), por lo tanto:

$$\Sigma \hat{Y}_{\text{mensuales}} = \hat{Y}_{\text{anual}}$$

En el caso de trabajar con la variable precios, se tendrá $\hat{Y} = 7{,}54 \quad + 44{,}86$ para estimaciones con años. Para estimativos mensuales sería:

$\hat{Y} = \frac{7{,}54}{12} \quad + \ 44{,}86 \rightarrow \hat{Y} = 0{,}6283 \quad + \ 44{,}86$ Cuando = 0 → $\hat{Y} = 44{,}86$

Corresponde al precio (promedio) para el primero de julio del 2008; se traslada al 15 de cada mes, siendo $\hat{Y} = 0{,}6283 \quad + 45{,}1742$. Se le suma a c la mitad de b.

Ejemplo 6. Consideremos el ejemplo 3, donde $\hat{Y} = 2{,}82 \quad + 68{,}5$ para realizar estimativos anuales. La ecuación para estimar meses, suponiendo que la variable Y representa precios será:

$\hat{Y} = \frac{b}{6}x + c \quad \hat{Y} = \frac{\ ,\ }{6}x + 6\ , \ \Rightarrow \hat{Y} = \ , \ x + 6\ ,$ si = 0

se tendrá que $\hat{Y} = 68{,}5$ al primero de enero de 2008, si lo trasladamos al 15 de enero, la ecuación será $\hat{Y} = 0{,}47 \quad + 68{,}735$, en ese punto = 0 y desde ahí contamos los meses necesarios para la estimación respectiva.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

27. Con los resultados obtenidos en el ejercicio 1, se pide: estimar la producción en los meses de Mayo de 2006; Septiembre de 2008; Octubre de 2010 y Diciembre de 2014. Se toma la ecuación cuyo origen es el año 2000, siendo (=0).

28. Con los datos y resultados del ejercicio 3, calcular los índices mensuales de precios (de Enero a Diciembre) para el 2009.

29. Con los datos del ejercicio 8, se pide estimar los meses de Marzo de 1996 y Abril de 2005. Recuerde que se trabajó con cambio de origen, es decir, cuando: $\Sigma_{i} = 0$

30. Con los datos del ejercicio 14, se pide estimar los meses de Mayo de 2006; Octubre de 2012, suponiendo que la variable corresponde a la suma de valores mensuales (Producción).

31. Con los datos del ejercicio 16; estimar los gastos de la oficina para Agosto de 2013.

32. Con los datos del ejercicio 20, estimar las ventas de Septiembre de 2009 y Octubre/2011 y 2013.

MÉTODO DE LOS SEMIPROMEDIOS

Este método, utilizado para determinar la tendencia no es muy recomendable, sin embargo se emplea cuando se requiere agilizar los cálculos y tener una aproximación al logrado a través de los mínimos cuadrados.

Consiste en dividir la serie en dos partes, en tal forma que cada una de ellas tenga un número impar de períodos. En algunos casos, pueden sobrar períodos, por lo tanto deberán ser localizados en la parte inicial de la serie o en el centro de la misma.

Ejemplo 7. Regresemos al Ejemplo 1 y realicemos la estimación para el 2015.

Solución:

AÑOS	i
2005	24
2006	29
2007	36
2008	42
2009	56
2010	60
2011	67

Semi suma	Semi promedio	Y_i
	→ — = 6	6
1	→ $\frac{1}{\quad}$ = 61	61

$$b = \frac{Y_t - Y_o}{2010 - 2006} \qquad b = \frac{61 \quad - \quad 6}{} =$$

1º Método

$\hat{Y}$= 7,8325 + 29,67 $\Rightarrow \hat{Y}_{15}$= 7,8325 (9)+ 29,67 = 100,16

a) = 2.015 − 2.006 = 9

b) $\hat{Y}$ = 7,8325x + 61,0 $\Rightarrow \hat{Y}_{15}$= 7,8325(5)+ 61,0 = 100,16

= 2.015 + 2010 = 5

Los resultados son exactamente iguales.

2º Método

Si 2005 → = 0 , para el 2006 → = 1 y en el 2010 → = 5 por lo tanto se tendrán dos ecuaciones, sabiendo que $\hat{Y} = b$ c.

(1) $61 = 5b + c$

(2) $29{,}67 = 1b + c$ $\Rightarrow$

$$61 = 5b + c$$
$$-29{,}67 = -1b - c$$
$$1, \quad = \quad b \quad 0$$

Despejamos:

$$\rightarrow \qquad b = \frac{1,}{} = 7{,}8 \;\; 25$$

Reemplazamos en una de las ecuaciones 29,67 = 1(7,8325) + c donde c = 29,67 - 7,8325 = 21,8375. La ecuación, con = 0 en el 2005 será: $\hat{Y}_{15}$= 7,8325(10)+ 21,8375. = 100,16 (Tambien por este procedimiento se obtiene el mismo resultado)

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

33. Con los siguientes datos, estimar $\hat{Y}_{18}$, utilizando el método de los semi-promedios.

AÑOS:	2005	2006	2007	2008	2009	2010	2011
$_i$:	200	360	490	500	620	780	910

34. Con los siguientes datos:

AÑOS:	1999	2001	2003	2005	2007	2009	2011
$_i$:	60	140	400	600	800	900	1.000

Se pide:

a) Por el método de los semi-promedios, estimar el valor de $\hat{Y}$ para el 2007, mediante una recta.

b) Con la anterior ecuación estimar los precios en los meses de Mayo 1998, Junio 2008.

35. Con los índices de precios anuales se pide:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
$_i$:	1.280	1.350	1.480	1.450	1.520	1.610	1.620

Trabajar utilizando el método de los semi-promedios para estimar el 2017.

36. Un almacén consolida sus ventas en períodos trimestrales.

AÑOS	TRIMESTRES	VENTAS (mill.)
2009	I	810
	II	1.200
	III	860
	IV	1.680

AÑOS	TRIMESTRES	VENTAS (mill.)
2010	I	900
	II	1.300
	III	850
	IV	1.600

AÑOS	TRIMESTRES	VENTAS (mill.)
2011	I	860
	II	1.100
	III	940
	IV	1.400

a) Representar gráficamente la anterior información.
b) Utilizando el método de los semi-promedios, estime el 2° semestre de 2013.
c) Con la ecuación anterior, dibuje la línea de tendencia lineal para estos datos.

37.

AÑOS:	1996	1999	2002	2005	2008	2011
$_i$:	200	360	460	580	720	800

Con la serie 1996–2011 se pide:

a) Mediante un ajuste parabólico, trabajando con cambio de origen ($\sum_i = 0$), estimar $\hat{Y}$ para el 2013.
b) Trabajando con el método de los semi-promedios, estimar a $\hat{Y}$ para el 2015.

38. Utilizando la siguiente serie, se pide estimar el valor de $\hat{Y}$ para el año 2015, mediante el método de los semi-promedios.

AÑOS:	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
$_i$:	200	380	520	600	640	720	580	660	940	800

TENDENCIA O SUAVIZACIÓN MEDIANTE PROMEDIOS MÓVILES

Los *tota es* y *promedios mó i es* permiten, en primer lugar, eliminar dos de los cuatro componentes que tiene la variable *Y*, éstos son: las *ariaciones estaciona es* y los *a eatorios*, sin embargo con frecuencia son utilizados como una aproximación a la *tendencia* de una variable temporal, permitiendo realizar estimaciones. También permite estimar la variación que está experimentando la serie, comparando un promedio móvil con otro del mismo período o períodos diferentes.

Este método fue sugerido por Thomson en 1947, considerado válido cuando la distribución de las tolerancias es simétrica. Algunos lo consideran como un método mecánico, generalmente utilizado para suavizar las irregularidades y fluctuaciones de una serie temporal.

Los totales y promedios móviles se pueden calcular utilizando un *n mero impar* de períodos: de 3 en 3; de 5 en 5; de 7 en 7; y así sucesivamente, o también un *n mero par* de períodos: de 2 en 2; de 4 en 4; de 6 en 6; y así sucesivamente.

Ejemplo 1. Con los siguientes datos calcular los totales y promedios móviles de “3 en 3” y de “5 en 5”.

So ución:

AÑOS	i	De 3 en 3		De 5 en 5	
		Total móvil (a)	Promedio móvil (a)	Total móvil (b)	Promedio móvil (b)
2001	28	—	—	—	—
2002	42	90	30	—	—
2003	20	120	40	220	44,0
2004	58	150	50	230	46,0
2005	72	168	56	244	48,8
2006	38	160	53,33	268	53,6
2007	56	138	46	290	58,0
2008	44	180	60	294	58,8
2009	80	200	66,66	290	58,0
2010	76	190	63,33	—	—
2011	34	—	—	—	—

Como se puede observar en la tabla anterior, consiste en tomar cierto número de períodos (impar o par) de tres, cinco, siete o dos, cuatro, seis, ocho, etc., que comprendan observaciones consecutivas, las que se suman y se dividen por el número de períodos tomados, obteniéndose la media aritmética del período. Este procedimiento debe ser constante y se va efectuando mediante el desplazamiento a lo largo de la serie, es decir, para el segundo promedio se omite el primer período pero se le agrega el siguiente, que no fue tomado en el promedio anterior. Veamos:

a) Se tomaron 3 períodos así: $\bar{y}_2 = \frac{y_1 + y_2 + y}{} = \frac{28 + \ 2 + 20}{} = \frac{90}{} = \ 0$

$\bar{y} \ = \frac{y \ + y \ + y}{} = \frac{\quad + \quad +}{} = \frac{1}{} =$

$\bar{y} \ = \frac{y \ + y \ + y}{} = \frac{\quad + \quad +}{} = \frac{1}{} = \quad etc$

b) Se tomaron 5 períodos así:

$\bar{y} \ = \frac{y_1 + y_2 + y \ + y \ + y_5}{5} = \frac{28 + \ 2 + 20 + 58 + 72}{5} =$

$\bar{y} \ = \frac{y_2 + y \ + y \ + y_5 + y_6}{5} = \frac{2 + 20 + 58 + 72 + \ 8}{5} = \ 6$, *etc.*

Se puede observar al calcular los promedios móviles de "3 en 3", se pierden dos períodos (2001 y 2011); al hacerlo de "5 en 5" se pierden dos arriba y dos abajo; y asi sucesivamente, de ahí que el procedimiento de seleccionar el número de períodos para calcular el promedio dependerá de la longitud de la serie.

Ejemplo 2. Con los mismos datos del ejemplo 1, calcular los promedios móviles de "2 en 2" y "4 en 4".

Solución:

		De 2 en 2			De 4 en 4		
AÑOS	y_i	**Total móvil (c)**	**Promedio móvil (c)**	**Promedio móvil (c)**	**Total móvil (d)**	**Promedio móvil (d)**	**Promedio móvil (d)**
2001	28						
		70	35				
2002	42			33			
		62	31		148	37	
2003	20			35			42,50
		78	39		192	48	
2004	58			52			47,50
		130	65		188	47	
2005	72			60			51,50
		110	55		224	56	
2006	38			51			54,25
		94	47		210	52,5	
2007	56			48,5			53,50
		100	50		218	54,5	
2008	44			56			59,25
		124	62		256	64	
2009	80			70			61,25
		156	78		234	58,5	
2010	76			66,5			
		110	55				
2011	34						

Si el número de observaciones es impar como en (a) y (b) del cuadro anterior, el promedio obtenido coincide con el período central; pero si es par como en (c) y (d), el promedio se coloca entre los dos valores centrales, procedimiento que en algunas ocasiones hace más engorroso el proceso, de ahí la costumbre del empleo de un número impar de observaciones.

c) Se tomaron períodos de dos en dos, así: $\bar{y}_{1,5} = \frac{y_1 + y_2}{2} = \frac{28 + 2}{2} = 5$

$\bar{y} = \frac{y + y}{} = \frac{+}{} = 1$ $\bar{y} = \frac{y + y}{} = \frac{+}{} =$ y así sucesivamente.

d) Se tomaron períodos de cuatro en cuatro, así: $\bar{y}_{2,5} = \frac{y_1 + y_2 + y + y}{} = \frac{28 + 2 + 20 + 58}{} = 7$

$\bar{y}_{,5} = \frac{y_2 + y + y + y_5}{} = \frac{2 + 20 + 58 + 72}{} = 8$, *etc*

En todos los casos, una vez obtenidas las medias o promedios móviles, puede repetirse el proceso, tratando de suavizar un poco más la serie, así:

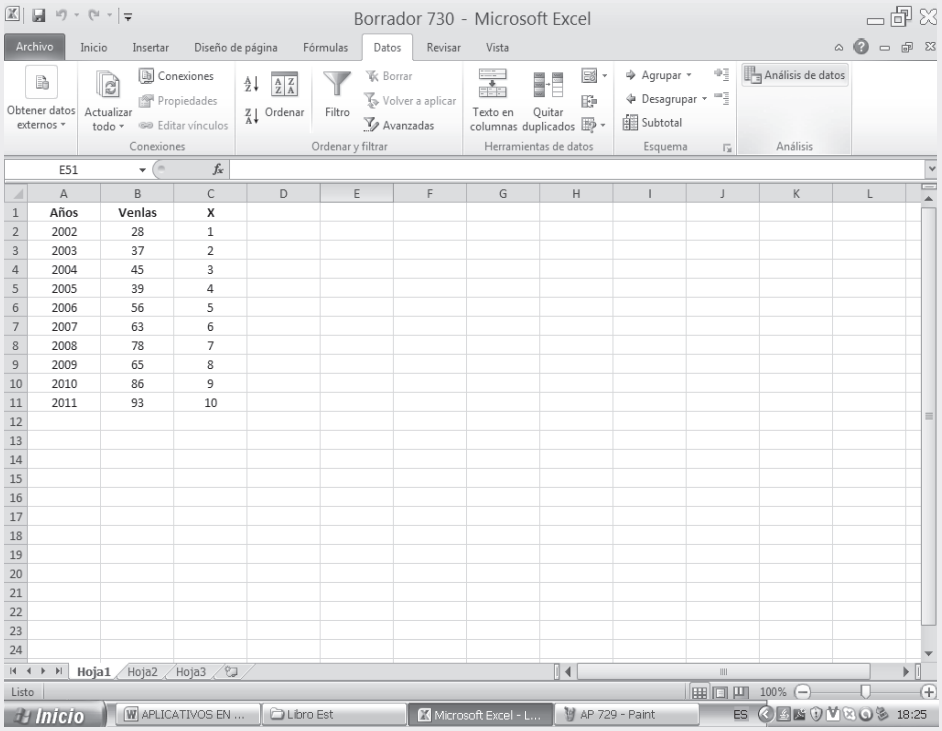
$$\bar{y} = \frac{\bar{y}_{2,5} + \bar{y}_{,5}}{2} = \frac{7 + 8}{2} = 2{,}5 \qquad \bar{y} = \frac{\bar{y}_{,5} + \bar{y}_{,5}}{2} = \frac{8 + 7}{2} = 7{,}5 \text{ } etc$$

La manera más fácil de proceder en el cálculo de los promedios móviles, consiste en calcular primero los *tota es mó i es*, es decir, los numeradores que luego hay que dividir por el número fijo. Algo que requiere cierta experiencia, es la determinación del número fijo, de tal manera que se obtenga una línea más suavizada que la original. Ya se ha dicho, que puede ser nuevamente suavizada utilizando los promedios móviles obtenidos, a los cuales se les vuelve aplicar el proceso.

APLICACIÓN DEL PROGRAMA DE EXCEL EN PROMEDIOS MÓVILES

Con los mismos datos, correspondientes a la serie de 10 años, con los que se realizó el ejercicio sobre tendencia lineal, podemos calcular los PROMEDIOS MÓVILES, en nuestro caso, de tres en tres años, de la siguiente manera:

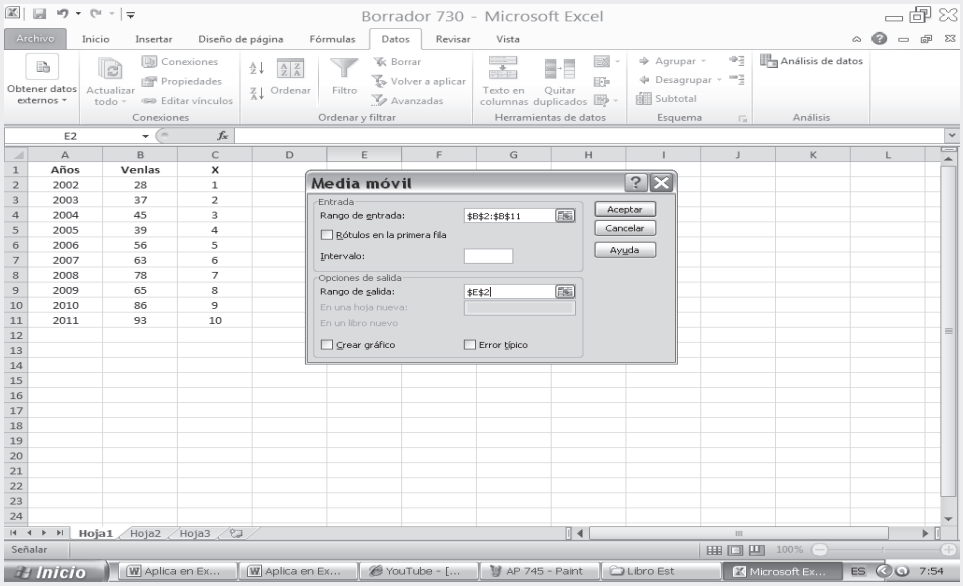
Figura No. 4. Análisis de datos.



- Seleccionamos la ficha DATOS, y luego hacemos CLIC en la opción ANÁLISIS DE DATOS.
- Aparece un cuadro de diálogo, correspondiente a ANÁLISIS DE DATOS, seleccionamos MEDIA MÓVIL y hacemos CLIC en ACEPTAR.
- En el campo RANGO DE ENTRADA tecleamos B2:B11; luego en el RANGO DE SALIDA señalamos una CELDA de la hoja de cálculo, en nuestro caso, será la celda E2.
- Finalmente hacemos CLIC en ACEPTAR, tal como aparece en las aplicaciones que se presentan a continuación:

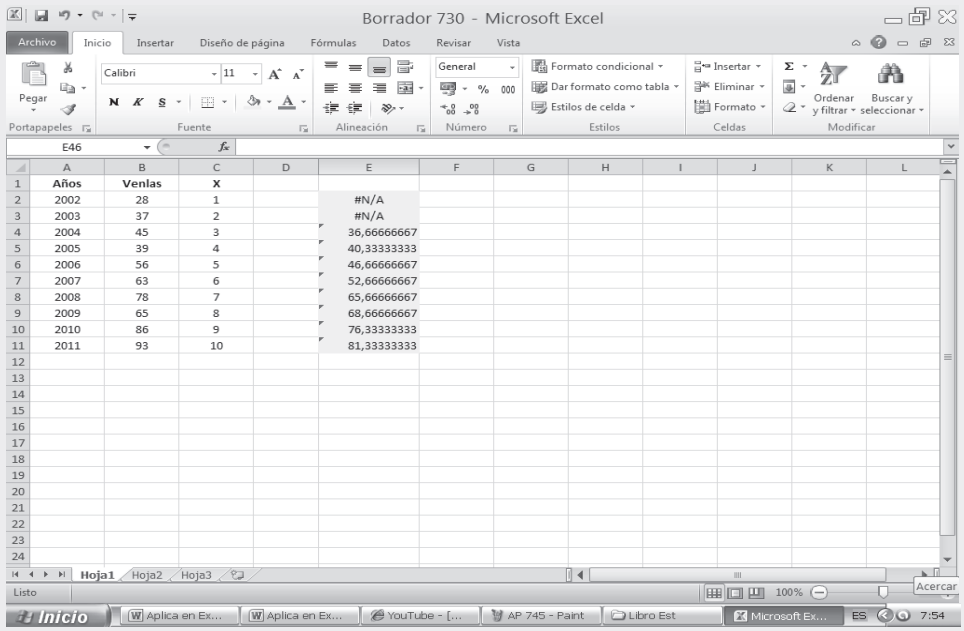
En la SEGUNDA PARTE DEL SIL en internet, encontrará en conjunto toda la aplicación del EXCEL a la ESTADÍSTICA, que igual a1 capítulo 13, ofrece una BASE DE DATOS para trabajar.

Figura No. 5. Cálculo de la Media Móvil



Como podemos observar en la columna E2 aparecen los promedios móviles de tres años.

Figura No. 6. Resultado cálculo de la Media Móvil.



TENDENCIA PARABÓLICA

La parábola por lo general se aplica en aquellos casos en que la línea poligonal presenta una parte ascendente y enseguida una descendente, o lo contrario; generalmente la línea es cóncava o puede ser convexa algunos la utilizan para proyecciones de utilidad, ingresos, etc., aunque su uso en la práctica es poco frecuente.

Es una ecuación de segundo grado. Las curvas de grado superior son poco usadas, por razón de su complejidad, ya que en algunos casos requieren de un cálculo muy engorroso. Sin embargo las de tercer grado suelen ser utilizadas para las funciones de producción.

La ecuación general de la parábola es: $\hat{Y} = ax_i^2 + bx_i + c$. El sistema de ecuaciones, para obtener los coeficientes es:

$$(1)\quad \Sigma y_t = a\Sigma x_t + b\Sigma x_t + nc$$
$$(2)\quad \Sigma x_t y_t = a\Sigma x_t + b\Sigma x_t + c\Sigma x_t$$
$$(3)\quad \Sigma x_t\, y_t = a\Sigma x_t + b\Sigma x_t + c\Sigma x_t$$

Ejemplo 1. Supongamos una serie cualquiera y realicemos un ajuste parabólico tomando como origen el primer año de la serie. Consideremos en primer lugar una serie continua.

AÑOS	$_i$	$_i$	^2_i	^3_i	^4_i	$_{i\ i}$	$^2_{i\ \ i}$	^2_i
2005	36	0	0	0	0	0	0	1.296
2006	28	1	1	1	1	28	28	784
2007	34	2	4	8	16	68	136	1.156
2008	52	3	9	27	81	156	468	2.704
2009	58	4	16	64	256	232	928	3.364
2010	62	5	25	125	625	310	1.550	3.844
2011	67	6	36	216	1.296	402	2.412	4.489
Σ	337	21	91	441	2.275	1.196	5.522	17.637

(1) 337 = 91a + 21b + 7c
(2) 1.196 = 441a + 91b + 21c Siguiendo el procedimiento de los mínimos cuadrados, se da:
(3) 5.522 = 2.275a + 441b + 91c

Multiplicamos por- 3 a la primera ecuación y la restamos de la (2)

(2) 1.196 = 441*a* + 91b + 21*c*
(1) − 1.011 = −273*a* − 63b − 21*c*
(4) 185 = 168*a* + 28b + 0

Multiplicamos por-13 a la primera ecuación y la restamos de la (3)

(3) 5.522 = 2.275*a* + 441b + 91*c*
(1) −4.381 = −1.183*a* − 273b − 91*c*
(5) 1.141 = 1.092*a* + 168b + 0

Ahora se tienen las ecuaciones (4) y (5).

(4) 185 = 168a + 28b
(5) 1.141 = 1.092a + 168b

Multiplicamos a la ecuación (4) por – 6 y la restamos de la (5)

(5) 1 1 1= 1 092a+168b

( )−1 110 = −1 008a − 168b

1= 8 a 0 $\Rightarrow$ $a = \frac{1}{8} = 0,\ 7$ a = 0,37

Reemplazamos en la ecuación (4).

(4) 1 =16 (,)+ b

1 =6 ,16+ b $\Rightarrow$ $b = \frac{1 \quad -6 \quad 16}{} =$ b = 4,39

Reemplazamos ahora en la ecuación (1).

$7 = 91(0,\ 7) + 21(\ ,\ 9) + 7c$

7 = ,67 + 92,19 + 7c $\Rightarrow$ $c = \frac{211,1}{7} =$ 0,16 c = 30,16

$\hat{Y} = ax^2 + bx + c$ $Y = 0,\ 7x^2 +\ ,\ 9x +\ 0,16$

El valor de $\hat{Y}$ para el año 2018 será:

= 2018 − 2005 = 13

$\hat{Y}_{18}$= 0,37(169)+ 4,39(13)+ 30,16 = 149,76

Ejemplo 2. Se puede trabajar con cambio de origen, dándole el mismo tratamiento que se le dio a la recta, es decir, que la sumatoria de la columna de X sea igual a cero.

AÑOS	$_i$	$_i$	^2_i	^3_i	^4_i	$_i\ _i$	$^2_i\ _i$	^2_i	$\hat{Y}_i$
2007	36	–2	4	–8	16	–72	144	1.296	33,88
2008	28	–1	1	–1	1	–28	28	784	31,67
2009	34	0	0	0	0	0	0	1.156	35,73
2010	52	1	1	1	1	52	52	2.704	46,03
2011	60	2	4	8	16	120	240	3.600	62,69
Σ	210	0	10	0	34	72	464	9.540	210,00

(1) $210 = 10a + 5c$

(2) $72 = 10b$

( ) 6 = $a + 10c$

$\Sigma y_t = a\Sigma x_t^2 + nc$

$\Sigma x_t y_t = b\Sigma x_t^2$

$\Sigma x_t^2 y_t = a\Sigma x_t + c\Sigma x_t^2$

Cuando $\Sigma x_i = 0$

$b = 7,2$

b = 7,2

Multiplicamos la primera ecuación por menos 2, y la restamos de la (3)

( ) 6 = $a + 10c$

(1) − 20 = −20a − 10c

= 1 a $\Rightarrow$ $a = \frac{}{1} =$,1 a = 3,14

Reemplazamos en la ecuación (1).

$1 = 1\ (\ 1\ ) + c \qquad \Rightarrow \qquad c = \dfrac{1\ -\ 1}{} = \qquad$ c = 35,72

$Y = \ 1\ x\ +\ \ x +$

Supongamos que se va a estimar el año 2017, teniendo en cuenta la información anterior, tenemos que:

$$\hat{Y}_{17} = 3{,}14(64) + 7{,}2(8) + 35{,}72 = 200{,}96 + 57{,}6 + 35{,}72 = 294{,}28$$
$$= 2017 - 2009 = 8$$

Cuando se trabaja con cambio de origen y la suma de es igual a 0, se pueden presentar fórmulas simplificadas para obtener directamente los parámetros *a*, *b* y *c*.

$$a = \frac{n\Sigma x_t^2 y_t - \Sigma x_t^2 (\Sigma y_t)}{n\Sigma x_t - (\Sigma x_t^2)^2} \Rightarrow a = \frac{5(\ 6\ ) - (10)(210)}{5(\ ) - (10)^2} = \ ,1$$

$$b = \frac{\Sigma x_t y_t}{\Sigma x_t^2} \Rightarrow b = \frac{72}{10} = 7{,}2$$

$$c = \frac{\Sigma y_t - a\Sigma x_t^2}{n} \Rightarrow c = \frac{210 - \ ,1\ (10)}{5} = \ 5{,}72$$

Ahora se puede determinar el valor de $\hat{Y}$ mediante un ajuste parabólico:

$Y = ax^2 + bx + c \qquad Y = \ ,1\ x^2 + 7{,}2x + \ 5{,}72$

La misma ecuación obtenida anteriormente.

Ejemplo 3. Se va a trabajar con cambio de origen, de tal manera que $\Sigma\ _1 = 0$ cuando la serie es continua y tiene además un número par de períodos.

So ución:

AÑOS	$_i$	$_i$	^2_i	^3_i	^4_i	$_i\ _i$	$^2_i\ _i$	^2_i	$\hat{Y}_i$
2006	28	–5	25	–125	625	-140	700	784	27,20
2007	34	–3	9	–27	81	-102	306	1.156	37,74
2008	52	–1	1	– 1	1	-52	52	2.704	47,72
2009	58	1	1	1	1	58	58	3.364	57,14
2010	62	3	9	27	81	186	558	3.844	66,00
2011	76	5	25	–125	625	380	1.900	5.776	74,30
Σ	310	0	70	0	1.414	330	3.574	17.628	310,10

$$b = \frac{\Sigma x_i y_i}{\Sigma x_i^2} \Rightarrow b = \frac{0}{70} = \ ,71$$ Despejamos los valores para *a* y *c*

(1) $\Sigma y_t = a\Sigma x_t^2 + nc \qquad$ (1) $\quad 10 = \quad 70a + 6c$

(2) $\Sigma x_t^2 y_i = a\Sigma x_t + c\Sigma x_t^2 \qquad$ (2) $\quad 57\ = 1\ \ 1\ \ a + 70c$

Multiplicamos la primera ecuación por 20,2 y se la restamos a la segunda ecuación con el fin de eliminar *a*

$$b = 4,71$$

$$c = 52,5$$

$$\begin{array}{lrrrrrr} (1) & 6\,262 = & 1 & 1 & a + & 121,2c \\ (2) & -\ 57 = & -1 & 1 & a - & 70c \\ \hline (a) & 2\,688 = & & & & 51,2c \end{array} \quad \Rightarrow \quad c = \frac{2\,688}{51,2} = 52,5$$

Reemplazamos en la primera ecuación, para despejar el valor de *a*

$$(1) \quad 10 = 70a + 6(52,5) \quad \Rightarrow \quad a = \frac{10 - 6(52,5)}{70} = -0,07$$

$$a = -0,07$$

La ecuación de la parábola será: $\hat{Y} = -0,07\ ^2 + 4,71\ + 52,5$

$= 2016 - 2009 = 7 \times 2 = 14 + 1 = 15$ semestres

El valor estimado de $\hat{Y}$ para el año 2016 será: $\hat{Y}_{16} = -0,07(225) + 4,71(15) + 52,5 = 107,4$

LÍMITES DE CONFIANZA

Se requiere en primer lugar calcular la varianza residual, que se puede obtener mediante la aplicación de una de las siguientes fórmulas:

$$(1) \quad s^2_{y.x} = VR = \frac{\Sigma\left(y_t - Y_t\right)^2}{n} \qquad (2) \quad s^2_{y.x} = \frac{\Sigma y_t^2 - c\Sigma y_t - b\Sigma y_t x_t - a\Sigma y_t x_t^2}{n}$$

No debemos olvidar que muchos la calculan, dividiendo por *n*–2, en vez de *n*.

Ejemplo 4. Consideremos los datos del ejemplo 1, para calcular la varianza residual y fijar límites de confianza del 95 %.

Solución:

AÑOS	y_i	$\hat{Y}_i$	$y_i - \hat{Y}_i$
2005	36	30,16	34,1056
2006	28	34,92	47,8864
2007	34	40,42	41,2164
2008	52	46,66	28,5156
2009	58	53,65	18,9225
2010	62	61,36	0,4096
2011	67	69,83	8,0089
Σ	337	337,00	179,0650

$$(1) \quad s^2_{y.x} = \frac{179,065}{7} = 25,58$$

$$(2) \quad s^2_{y.x} = \frac{17\,6\ 7 - \ 0,16(\ \ 7) - \ ,\ 9(1\ 196) - 0,\ 7(5\ 522)}{7} = 25,6$$

La desviación típica residual será: $s_{y.x} = \sqrt{s^2_{y.x}} \Rightarrow s_{y.x} = \sqrt{25{,}58} = 5{,}06$

Los límites de confianza del 95 % serán:

$$Y_1 \pm t\,\frac{s_{y.x}}{\sqrt{n}} \Rightarrow 1\ 9{,}76 \pm 2,\quad 7\ \frac{5{,}06}{\sqrt{7}} = 1\ 9{,}76 \pm\ ,68 = \begin{cases} 15\ , \\ 1\ 5{,}08 \end{cases}$$

Se hubiera podido utilizar t = 2,571 con $\upsilon = n - 2 = 7 - 2 = 5$

Coeficiente de correlación

La fórmula que más se aplica en la tendencia parabólica es:

$$r^2 = R^2 = 1 - \frac{s^2_{y.x}}{s^2_y} \Rightarrow r = \sqrt{R^2}$$

Ejemplo 5. Con los datos del ejemplo calcular el coeficiente de correlación parabólico.

Solución:

Calculamos la varianza de *Y* siendo $s^2_y = \frac{17\,6\ \ 7 - 7(\ \ 7/7)^2}{7} = 201{,}8$ y con el valor obtenido de la varianza residual, en este ejercicio $S^2 = 25{,}58$ se podrá calcular el coeficiente de correlación al cuadrado o coeficiente de determinación, así:

$$R^2 = 1 - \frac{s^2_{y.x}}{s^2_y} \qquad R\ = 1 - \frac{\qquad}{1} = \qquad \text{y} \qquad R = \sqrt{\qquad} =$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

39. Con los siguientes datos:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
$_i$:	234	171	147	124	140	144	206

Se pide:

a) Ajustar una parábola, tomando como origen a 2005 y estimar el valor de $\hat{Y}$ para 2015.
b) Ajustar una parábola, trabajando con cambio de origen y estimar el valor para 2020.

40. Con los datos del ejercicio anterior, se pide calcular:
a) El coeficiente de correlación parabólica.
b) El error estándar de estimación.
c) Fijar los límites de confianza del 95 % para $\hat{Y}_{2020}$

41. Con los siguientes datos de una serie discontinua, se pide estimar $\hat{Y}_{16}$, mediante el ajuste parabólico.

AÑOS:	1986	1994	1999	2006	2011
$_i$:	70	36	24	60	110

42. Con los siguientes datos, de una serie continua y par, se pide:

AÑOS:	2004	2005	2006	2007	2008	2009	2010	2011
$_i$:	56	42	70	82	41	63	110	84

a) Estimar $\hat{Y}$ para el 2013.
b) Calcular la desviación típica residual.
c) Calcular el coeficiente de correlación parabólico y elaborar la gráfica respectiva.

43. Con los siguientes datos:

AÑOS:	1996	1999	2002	2005	2008	2011
$_i$:	20	36	42	62	84	74

Se pide:

a) Mediante un ajuste parabólico, trabajando con cambio de origen ($\sum$ $_i = 0$) estimar el valor de $\hat{Y}$ para el 2013.
b) Trabajando por el método de los semipromedios, estimar $\hat{Y}$ para 2013, mediante un ajuste rectilíneo.

44. Con los siguientes datos: c = 351,1 b = 13,188 ; a = 0,311; $_i$ se da por semestres, con origen en el primero de Enero de 2006.

Se pide estimar el valor $\hat{Y}$ de para 2016.

45. Con los siguientes datos:

$\Sigma x_t = 0$ $\Sigma y_t = 20$ $\Sigma x_i =$ $\Sigma x_i = 1$ $\Sigma y_i = 1\ 6$

$\Sigma x_t =$ $\Sigma x_t^2 y_t = 8$ $n =$ $\Sigma x_t y_t =$; (Año base 2009)

Mediante un ajuste parabólico:

a) Estimar al valor de $\hat{Y}$ cuando = 20, además, fijar límites de confianza del 95%.
b) Calcular el coeficiente de correlación al cuadrado.
c) Con los anteriores datos haga un ajuste rectilíneo, dándole a el valor de 20 y compare el resultado con el obtenido en el punto (a).

46. Con los siguientes datos:

AÑOS:	2003	2004	2005	2006	2007	2008	2009	2010	2011
$_i$:	20	26	42	58	78	63	42	36	14

Se pide:

a) Estimar, mediante un ajuste parabólico, el valor de $\hat{Y}$ para el 2015.
b) Al valor estimado, calcule el error estándar de estimación y fije límites de confianza del 95 %.
c) Calcule el coeficiente de correlación parabólico.
d) Haga la representación gráfica para los datos originales y los estimados.

47. Con los siguientes datos:

AÑOS:	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
$_i$:	36	48	72	84	63	52	48	36	56	72
$_i$:	-9	-7	-5	-3	-1	1	3	5	7	9

a) Estime a $\hat{Y}_{2016}$ y fije límites de confianza del 95 % para el valor estimado, utilizando el ajuste parabólico.
b) Calcule el coeficiente de correlación y haga la gráfica correspondiente.

48.

MESES	$_i$
Enero	13
Febrero	17
Marzo	38
Abri1	50
Mayo	40
Junio	20
Ju1io	10

a) Mediante un ajuste parabólico estime el valor de $\hat{Y}$ para el mes de Diciembre.
b) Fije límites de confianza al 95%
c) Calcule el coeficiente de correlación.
d) Elabore la gráfica correspondiente.

TENDENCIA EXPONENCIAL

En la práctica son numerosas las series que presentan un crecimiento geométrico, considerándose como ejemplo típico la población de una ciudad, región o país y aquellas variables que guardan relación con el comportamiento de la población, como por ejemplo: ventas, salarios, ingreso nacional, etc., todos ellos pueden ser descritos por una ecuación denominada como *ínea e ponencia*

La ecuación general está dada en la siguiente forma: $\hat{Y} = cb$

La ecuación general es $= cb$, función muy utilizada en matemática financiera, siendo: c_1 = capital inicial; = capital final o monto; $b = (1 + r)$ donde r = tasa de crecimiento geométrico o interés compuesto mensual, anual, etc., y finalmente = tiempo, siendo:

$$P_f = P_i\,(1 + r)^n \qquad c_f = c_i\,(1 + r)$$

En la estimación de una población se tiene $P_f = P_0\,(1 + r\,)^n$, donde P_f = población final o estimada; P_0 = población base o inicial; r = tasa vegetativa y n = tiempo.

La ecuación general, se puede expresar en forma logarítmica como: $\log \hat{Y} = \log c + \log b$

Las ecuaciones normales para el cálculo de los parámetros **c** y **b**, utilizando el método de los mínimos cuadrados, será:

$$(1)\ \Sigma \log\ _i = n \log c + (\Sigma\ _i)\log b$$
$$(2)\ \Sigma\ _i \log\ _i = (\Sigma\ _i)\log c + (\Sigma\ _i^{\,2})\log b$$

Las anteriores ecuaciones se utilizan cuando se toma como origen al primer período (años, meses o días) de la serie o cuando la sumatoria de $_i$, es diferente a 0 ($\sum\ _i \neq 0$). Si la sumatoria de $_i$ es igual a 0, los parámetros se obtienen directamente en la siguiente forma:

$$\log c = \frac{\Sigma \log y_t}{n} \qquad \log b = \frac{\Sigma x_t \log y_t}{\Sigma x_t^2}$$

Ejemplo 1. Aplicaremos en primer lugar este último método, en una serie continua y con un número impar de años, comprendida entre 2005/2011, tal como aparece a continuación:

Solución:

AÑOS	$_i$	log $_i$	$_i$	$_i$ log $_i$	$_i^2$
2005	5.453	3,73664	-3	–11,20992	9
2006	7.558	3,87841	-2	–7,75682	4
2007	9.190	3,96332	-1	–3,96332	1
2008	10.584	4,02465	0	0	0
2009	16.645	4,22128	1	4,22128	1
2010	25.314	4,40336	2	8,80672	4
2011	40.419	4,60658	3	13,81974	9
Σ	115.163	28,83424	0	3,91768	28

(1) $c = \frac{28{,}8\ \ 2}{7} =\ \ ,119176$ $\quad \log c = \frac{\Sigma \log y_t}{n} \quad \rightarrow\ c = 13.157{,}61$

(2) $b = \frac{,91768}{28} = 0{,}1\ \ 9917$ $\quad \log b = \frac{\Sigma x_t \log y_t}{\Sigma x_t^2} \quad \rightarrow\ b = 1,\ \ 801$

Los valores estimados se obtendrán por cualquiera de las dos fórmulas:

a) $\log \hat{Y} = 4{,}119176 + 0{,}139917\ (\ \ )$

b) $\hat{Y} = 13.157(1{,}3801)$ (los resultados deben ser iguales, si se trabaja con todos los decimales)

Si se desea estimar el valor de $\hat{Y}$ para el 2018 se tendrá:

a) $\log \hat{Y}_{18} = 4{,}119176 + 0{,}139917(10) = 4{,}119176 + 1{,}39917 = 5{,}518346$

Si se obtiene el antilogaritmo, será: $\hat{Y}_{18} = 329.872{,}41$ $\quad = 2018 - 2008 = 10$

b) $\hat{Y}_{18} = 13.157(1{,}3801)^{10} = 329.808{,}95$

La *tasa de crecimiento* para esta serie se obtiene de la siguiente manera: primero se determina el antilogaritmo para el $\log b = 0{,}139917 \quad \rightarrow \quad b$ = antilog de 0,139817 = 1,3801.

En segundo lugar, se sabe que $b = 1 + r$, siendo r la tasa de crecimiento geométrico.
Si $b = 1 + r = 1{,}3801$, se tendrá que $r = 0{,}3801$ ó $r = 38{,}01\%$ o $380{,}1‰$ ó de acuerdo como se considere debe presentarse.

Ejemplo 2. Hagamos el estimativo para $\hat{Y}_{18}$, con la serie que se ha venido trabajando, correspondiente al período 2005-2011, siendo continua, para $\Sigma\ _1 \neq 0$, es decir, tomando como origen el primer período de la serie:

Solución:

AÑOS	y_i	$_i$	^2_i	log $_i$	$_i$log $_i$	log $\hat{Y}_i$	$\hat{Y}_i$	(log $_i$)^2
2005	36	0	0	1,55630	0	1,47629	29,94	2,42208
2006	28	1	1	1,44716	1,44716	1,53813	34,52	2,09427
2007	34	2	4	1,53148	3,06296	1,59999	39,81	2,34543
2008	52	3	9	1,71600	5,14800	1,66183	45,90	2,94466
2009	58	4	16	1,76343	7,05372	1,72368	52,93	3,10967
2010	62	5	25	1,79239	8,96195	1,78553	61,03	3,21267
2011	67	6	36	1,82607	10,95645	1,84738	70,37	3,33453
Σ	337	21	91	11,63283	36,63024	11,63283	334,50	19,46331

En el cuadro anterior observamos que Σ $_i \neq \Sigma \hat{Y}_i$ y que Σ log $_i$ = Σ log $\hat{Y}_i$.

Ahora procedemos a calcular los **log b** y **log c** siguiendo el procedimiento de los mínimos cuadrados, para ello reemplazamos los valores que aparecen en la tabla, en el sistema de ecuaciones normales.

(1) 11,63283 = 7 log c +21 log b
(2) 36,63024 =21 log c + 91 log b

Multiplicamos la ecuación (1) por –3 y lo restamos de la ecuación (2).

(2) 36,63024 = 21 log *c* + 91 log *b*
(1) <u>– 34,89849 = –21 log *c* – 63 log *b*</u>
1,73175 = 0 28 log *b*

Siendo: log b = $\frac{1,73175}{28}$ = 0,06185 log *b* = 0,06185

b = antilog 0,06185 = 1,1530. La tasa de crecimiento será: r = 15,30% ó 153‰.

Reemplazamos el valor del **log b**, en la primera ecuación, para determinar el **log c**.

(1) 11,63283 = 7 log c + 21(0,06185) log *c* = $\frac{11,63283 - 21(0,06185)}{7}$ = 1,47629

La función logarítmica será: log $\hat{Y}$ = 1,47629 + (0,06185). Si se desea estimar el año 2018 se tendrá que:

log $\hat{Y}_{18}$ = 1,47629 + 13 (0,06185) = 2,2803; Siendo: $\hat{Y}$ = antilog de 2,2803
$\hat{Y}_{18}$ = 190,68 Cuando = 2018 – 2005 = 13

Límites de confianza. Primero que todo se debe calcular la *arian a residua* y luego la *des iación típica residua* , a fin de obtener el error estándar de estimación. Las fórmulas más usadas son:

(1) $s^2_{x \quad y} = VR = \frac{\Sigma(\quad y_t - \quad Y_t)^2}{n}$

(2) $s^2_{x \quad y} = \frac{\Sigma(\quad y_t)^2 - (\quad c)(\Sigma \quad y_t) - (\quad b)(\Sigma x_t \quad y_t)}{n}$

Ejemplo 3. Con los datos y cálculos realizados en el Ejemplo 2, calcularemos las medidas anteriores.

Solución:

AÑOS	$\log y_i$	$\log \hat{Y}_i$	$(\log y_i - \log \hat{Y}_i)^2$
2005	1,55630	1,47629	0,00640
2006	1,44716	1,53813	0,00828
2007	1,53148	1,59999	0,00469
2008	1,71600	1,66183	0,00293
2009	1,76343	1,72368	0,00158
2010	1,79239	1,78553	0,00005
2011	1,82607	1,84738	0,00045
Σ	11,63283	11,63283	0,02438

$$s^2_{y\cdot} = \frac{\Sigma(\quad y_t)^2 - n\ \overline{\quad y_t}^2}{n} \qquad s^2_{y\cdot} = \frac{19,\ 6\quad 1 - 7(1{,}6618\ \ )^2}{7} = 0{,}01879$$

$$\overline{\quad y_t} = \frac{\Sigma\quad y_t}{n} = \frac{11{,}6\ \ 28}{7} = 1{,}6618 \qquad \text{(varianza y media logarítmica)}$$

$$(1)\ \ s^2_{x\quad y\cdot} = \frac{0{,}02\quad 8}{7} = 0{,}00\quad 8$$

$$s^2_{x\quad y\cdot} = \frac{19,\ 6\quad 1 - 1,\ 7629(11{,}6\ \ 28\ \ ) - 0{,}06185(\ 6{,}6\ \ 02\ \ )}{7} = 0{,}00\quad 7$$

La diferencia observada es muy pequeña, comparando los resultados anteriores. También se puede hacer el cálculo de la varianza residual, de la siguiente manera:

$$(\ )\ \ s^2_{x\quad y\cdot} = s^2_{\quad y\cdot}(1-R)^2$$

Siendo: $s^2_{x\quad y\cdot} = 0{,}01878(1 - 0{,}81\ \ 5\ \ ) = 0{,}00\quad 8$

Coeficiente de correlación. Al igual que la tendencia rectilínea y parabólica, este coeficiente se puede calcular aplicando diferentes fórmulas, cuyos resultados son aproximadamente iguales.

$$(1)\ \ r^2 = R^2 = 1 - \frac{s^2_{x\quad y\cdot}}{s^2_{\quad y\cdot}}$$

$$(2)\ \ r = \frac{n\Sigma x_t\quad y_t - \sum_{\ i}(\sum \log\ _i)}{\sqrt{\left[n\Sigma x_t^2 - (\Sigma x_t)^2\right]\left[n\Sigma(\quad y_t)^2 - (\Sigma\quad y_t)^2\right]}}$$

Ejemplo 4. Con los datos y resultados del ejemplo 3, calcular el coeficiente de correlación.

Solución:

$$(1)\ \ R^2 = 1 - \frac{0{,}00\quad 8}{0{,}01878} = 0{,}81\ \ 5 \qquad r = \sqrt{0{,}81\ \ 5} = 0{,}90$$

$$(2)\ r = \frac{7\ (\ 6{,}6\ \ 02\ ) - (21)(11{,}6\ \ 28\ )}{\sqrt{\left[7(91) - (21)^2\right]\left[7(19,\ \ 6\ \ \ ) - (11{,}6\ \ 28\ )^2\right]}} = 0{,}90$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

49. Los últimos censos de población efectuados en un país latinoamericano arrojaron los siguientes resultados:

AÑOS:	1956	1969	1979	1999	2011
POBLACIÓN (miles):	3.221	3.715	4.287	5.024	6.933

a) Se pide estimar la población de dicho país para el año 2015.
b) Determinar la tasa de crecimiento anual.
c) El error estándar de estimación.
d) El coeficiente de correlación.
e) Entre qué valores estará la población para el 2017 con una probabilidad del 95%.

50. La población de un país en 2001 era de 12 millones; en 2011 fue de 21 millones.
Se desea determinar:
a) La tasa de crecimiento.
b) La población estimada para el 2017.

51. Si las importaciones (M) de una mercancía se duplican en un período de 4 años, ¿cuál es la tasa media acumulativa de incremento anual?

52. Un capital de 20 millones de pesos. ¿En cuántos años, a una tasa del 7% se duplicará?

53. Con la siguiente información:

AÑOS:	2005	2006	2007	2008	2009	2010	2011
IPRA*:	100,0	102,0	105,1	109,3	113,1	116,5	120,3

* Índice Producción Real Agrícola.

Se pide:

a) Establecer la ecuación para el ajuste exponencial.
b) La tasa de crecimiento acumulativo anual.

54. Si en el 2005 el valor de $\hat{Y}$ es 120 y en el 2011 es de 460. Se pide, con esos dos períodos, determinar la tasa de crecimiento geométrico, además estimar el valor de $\hat{Y}$ para el 2015.

55. De acuerdo con el plan de desarrollo de un país, se comprueba que el producto real del sector agrícola muestra los siguientes crecimientos porcentuales anuales en el período 2005/2011.

AÑOS:	2005	2006	2007	2008	2009	2010	2011
CRECIMIENTO BASE %:	-	2,0	3,6	4,2	3,8	3,4	3,5
ÍNDICE:	100,0	102,0	105,6	110,1	114,3	118,2	122,3

En una serie encadenada, teniendo en cuenta los crecimientos porcentuales en cada año:

a) Determine la tasa anual promedio de crecimiento (geométrico) para el período considerado.
b) Estime el índice para 2013 manteniendo constante la base 2005.

56. Un capital de $50.000 colocado a una tasa de interés compuesto del 1,75% mensual.
a) A los cuántos años se triplica el capital?
b) Si se coloca ese capital ($50.000) el 1° de Enero de 2005 a la misma tasa de interés, ¿cuál será el capital al 31 de Diciembre de 2011?

57. Con los siguientes datos:
$\Sigma\ _{i} = 105 \quad \Sigma\ _{i}^{2} = 1.015 \quad \Sigma \log\ _{i} = 80,33100 \quad \Sigma\ _{i} \log\ _{i} = 583,45830$
$n = 15 \quad = 0$ para 1996 (primer año de la serie).
Se pide:
a) Determinar la tasa de crecimiento anual.
b) Estimar el valor de $\hat{Y}$, para 2013 mediante el ajuste exponencial.

58. Dados los siguientes datos sobre la producción de un empresa, en millones de unidades.

AÑOS:	2006	2007	2008	2009	2010	2011
PRODUCCIÓN:	150	140	180	220	190	380

Se pide estimar la producción para el año 2016 y calcular los coeficientes de correlación, mediante el:
a) el ajuste exponencial. b) el ajuste rectilíneo y el coeficiente de correlación.

59. Con los siguientes datos, correspondientes al P.N.B. de un país cualquiera.
$\Sigma\ _{i} = 0 \quad \Sigma \log\ _{i} = 28,744073 \quad \Sigma\ _{i} \log\ _{i} = 1,708451 \quad \Sigma\ _{i}^{2} = 60 \quad n = 9$
Si se sabe que el origen está localizado en el año 2007, se pide determinar mediante un ajuste exponencial.
a) La tasa de crecimiento anual. b) El valor estimado de $\hat{Y}$, para el 2015.

60. Un capital puesto a una tasa de interés anual del 24,8%, al cabo de 4 años se transformó en $2.580.000. ¿Cuál fue el valor del capital inicial?

61. Un capital de $150.000 se triplica al cabo de 8 años. a) ¿Cuál es la tasa acumulativa anual?
b) Si el capital se colocó el primero de enero de 2006, ¿Cuál será el capital a 31 de diciembre 2011, si se mantiene dicha tasa?
c) ¿Cuál es la tasa acumulativa mensual, con el enunciado de este ejercicio?

62. Siendo: $n = 14 \quad \Sigma \log\ _{i} = 18,43915 \quad \Sigma\ _{i}^{2} = 819 \quad \Sigma(\log\ _{i})^{2} = 36,87826$
$\Sigma\ _{i} = 91; \quad \Sigma\ _{i} \log\ _{i} = 129,55270$ Para 1997 tomando como origen $= 0$.
a) Se pide estimar $\hat{Y}_{13}$.
b) Calcular el coeficiente de correlación exponencial.

OTROS MÉTODOS UTILIZADOS EN EL ANÁLISIS DE LA TENDENCIA

En forma breve se exponen otros procedimientos utilizados en el análisis de series, dependiendo del tipo de variable y análisis que se disponga.

CURVA EXPONENCIAL MODIFICADA

Es una curva análoga a la *e ponencia* , la cual se modifica agregando una constante a, que sirve simplemente para indicarnos el desplazamiento de la curva hacia arriba o hacia abajo en una cantidad constante, usada especialmente en aquellos casos en que los valores de la serie tienen una marcada tendencia hacia un límite superior o inferior.

La ecuación es: $\hat{Y} = a + cb$

CURVA LOGÍSTICA

Esta curva está dada por la ecuación: $\hat{Y} = \frac{1}{a + cb}$

Como se puede observar, corresponde al *recíproco de la exponencial modificada*, donde los valores a, b y c se obtienen de la misma manera que en la *exponencial modificada*, la que, a su vez, muestra una tasa de crecimiento o de decrecimiento relativamente grande al comienzo, pero que disminuye al final de la serie.

CURVA DE GOMPERTZ

Es también una *cur a de tendencia*, como las anteriores, que se expresa mediante la fórmula de $\hat{Y} = a + c^b$ y cuyo nombre corresponde a Benjamin Gompertz, quien la utilizó en estudios sobre tablas de mortalidad. Su forma de cálculo se simplifica empleando logaritmos, siendo: $\log \quad = \log a + (b \quad) \log c$

MÉTODO DE LOS COMPONENTES

No es propiamente una línea de tendencia, pero si es un método que nos permite calcular el valor para los períodos siguientes, en especial para estimar la población de una ciudad, región o país. Este método se obtiene tomando la poblacional inicial o base, correspondiente al último censo (P_o), a este valor se le suman los nacidos vivos en el período (N); menos las defunciones en el período (D) y más o menos las migraciones (M).

$$P_1 = P_0 + N - D \pm M$$

MÉTODO GRÁFICO

Denominado también de *mano a ada* o método a *mano ibre* Establecido para describir las relaciones en una serie de datos ordenados en el tiempo o en el espacio. En algunos casos, la tendencia se establece aplicando el método a mano libre o mano alzada, sin empleo matemático, trazando una línea (recta, parabólica, exponencial, etc.) que pase por o cerca de los puntos que representan gráficamente las observaciones. Generalmente se trata de que la línea pase por el centro de la (línea) poligonal o datos observados.

El método en un principio parece muy sencillo, pero la bondad del resultado depende de la persona que lo aplique, pues se requiere experiencia y conocimiento del comportamiento de la variable, para lograr la proyección al período deseado. La crítica que se le hace a este procedimiento es su carácter subjetivo, dado que el resultado o sea el trazo de la línea, depende de la persona que lo haga; sin embargo se dice, que si es bien aplicado su resultado puede ser mejor que los otros métodos matemáticos que hemos denominado *métodos ana íticos*

EJERCICIOS MISCELÁNEOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

63. Dada la siguiente información, correspondiente a una serie de tiempo.

AÑOS:	2006	2007	2008	2009	2010	2011
$_i$:	70	56	74	80	96	114

El valor para $\hat{Y}$ en el 2016, mediante un ajuste rectilíneo será:
a) 135,4 b) 138,6 c) 144,2 d) 151,6 e) 155,8

64. Si se dispone de los siguientes datos y mediante un ajuste rectilíneo.

AÑOS	y_i	$\hat{Y}_i$
2006	18	15
2011	72	75

Se podrá decir que el coeficiente angular (crecimiento anual) es de:
a) 12 b) 16,8 c) 20,4 d) 22,5 e) 25,6

65. Con los siguientes datos, correspondientes a las ventas trimestrales de un almacén:

AÑOS	**MESES**	**VENTAS (millones)**
2009	Octubre	75
2010	Enero	92
	Abril	63
	Julio	107
	Octubre	94
2011	Enero	130
	Abril	68
	Julio	162
	Octubre	134

Se pide:

a) Mediante un ajuste rectilíneo, estimar el valor de las ventas correspondientes al primer trimestre de 2013.
b) Mediante el ajuste parabólico, trabajando con cambio de origen $(\Sigma\ _i = 0)$, estimar los meses de Septiembre de 2010 y Diciembre de 2011.

66. Con la ecuación obtenida para el ajuste rectilíneo y el parabólico, junto con los datos reales de las ventas, elaborar la gráfica correspondiente (las tres líneas en una sola gráfica). (trabajar con el ejercicio N° 65).

67. Utilizar el método de los semipromedios (en el ejercicio No. 65) para estimar los meses de Septiembre de 2010 y Diciembre de 2011.

68. Un capital de 300 millones de $. ¿En cuántos años se duplicaría, si se tiene:
    a) una tasa de interés al 2% mensual?
    b) una tasa de interés anual del 32%?

69. Con los siguientes datos, se pide determinar por el método que considere conveniente:

AÑOS	y_i	$\hat{Y}_i$
2006	120	110
2011	480	516

    a) La tasa de crecimiento aritmético.
    b) La tasa de crecimiento geométrico.

70. Con la información del ejercicio anterior, estime el valor de $\hat{Y}$ para el 2015, mediante un ajuste rectilíneo y luego, mediante un ajuste logarítmico.

71. Dé respuesta en forma abreviada a cada uno de los siguientes interrogantes.
    a) ¿Qué utilidad práctica le ve al análisis de una serie de tiempo?
    b) ¿Qué entiende por: tendencia; variaciones estacionales; movimientos cíclicos; variaciones aleatorias?
    c) ¿En qué difieren los modelos de adición y de multiplicación en las series de tiempo?
    d) ¿En qué consiste el método de mano alzada?
    e) ¿Qué opina del método de los semipromedios?
    f) ¿En qué consiste la suavización de una línea?
    g) ¿En qué consiste la desestacionalización de una serie?

72. En series de tiempo al calcular el coeficiente de correlación, ¿Cómo interpreta los siguientes resultados? r = 1 r = 0,9 r = 0,20 r = -1

73. En un ejercicio, respecto a una serie de tiempo, se obtuvo que, r = 1,32, ¿cómo lo interpretaría?

74. Un Instituto de educación superior, lleva un registro periódico sobre la tasa de mortalidad académica.

AÑOS:	2005	2006	2007	2008	2009	2010	2011
TASA %:	11,3	2,5	8,3	7,6	18,9	12,6	18,3

    a) Trace los datos anteriores en una gráfica.
    b) Haga un ajuste rectilíneo y trace la recta en la gráfica anterior.
    c) Haga un ajuste parabólico y trace la línea en la misma gráfica solicitada anteriormente.

75. Con los datos del ejercicio 74, se pide que:
    a) Haga un pronóstico para el año 2013 con los dos ajustes.
    b) Fije los límites de confianza del 95% en ambos casos.
    c) Obtenga el coeficiente de correlación y determine cuál de los dos ajustes aceptaría o recomendaría.

76. a) Con los datos del ejercicio anterior ajuste una línea exponencial, estime la tasa para el 2013 y fije límites del 95% para el valor estimado.
    b) Calcule el coeficiente de correlación ¿Qué opina respecto de los resultados del ejercicio anterior?

VARIACIONES ESTACIONALES

En los componentes de una serie, se mencionaban las *ariaciones estaciona es*, cuya duración puede ser de un año (de acuerdo a las estaciones del año), o períodos más cortos de un mes o un día; en estos últimos generalmente se consideraba como ejemplo, el movimiento de pasajeros en el transporte urbano donde encontramos horas pico de mucha congestión y mínimas a otras horas del día; algo similar se observa con el consumo de energía en las 24 horas del día. En los supermercados vemos como aumentan las ventas en los días de pago (quincenal o mensual); en las tabernas los días finales de la semana o en los aeropuertos y terminales de transportes en las temporadas de vacaciones.

Es costumbre, en el análisis de una *Serie de tiempo*, en cuanto al segundo componente, expresar las fluctuaciones estacionales utilizando índices, mediante los cuales pueden ser aplicados fácilmente a cualquier serie similar.

CÁLCULO DE ÍNDICES ESTACIONALES

Supongamos que en la terminal de transporte, de una ciudad, de un país cualquiera, el movimiento de pasajeros intermunicipales, en miles, es como sigue:

Número de pasajeros transportados durante 2009-2011.

MESES	2009	2010	2011	SUMA	PROMEDIO	(IVE) ÍNDICE VARIACIÓN ESTACIONAL
Enero	796	680	948	2.424	808	160,54
Febrero	712	548	552	1.812	604	120,00
Marzo	696	632	640	1.968	656	130,34
Abri1	448	512	504	1.464	488	96,96
Mayo	408	384	416	1.208	404	80,27
Junio	368	336	372	1.076	360	71,52
Ju1io	400	388	448	1.236	412	81,86
Agosto	364	376	392	1.132	376	74,71
Septiembre	452	432	508	1.392	464	92,19
Octubre	432	412	472	1.316	440	87,42
Noviembre	420	416	464	1.300	432	85,43
Diciembre	588	564	640	1.792	596	118,42
TOTAL	**6.084**	**5.680**	**6.356**	**18.120**	**6.040**	**1.199 12**
					503 3	

El procedimiento a seguir es:

- Sumamos la información para el mes de cada período y obtenemos la columna suma.
- Calculamos el promedio para cada mes, por ejemplo en Enero será 2.424/3 = 808.
- Sumamos el valor correspondiente a la columna promedio, dividimos entre 12 y nos da 503,3, es decir, 6.040 ÷ 12 = 503,3.

- Dividimos cada uno de los promedios mensuales por el promedio total (503,3) y el resultado lo multiplicamos por cien, así obtenemos el Índice de variación estacional.
- La suma del índice de *ariación estaciona* debe ser exactamente igual a 1.200; como se puede ver el resultado de nuestro ejercicio fue 1.199,12 y la diferencia se debe a las aproximaciones.

Es necesario tener en cuenta que en la mayoría de los casos los índices resultan influenciados por la *endencia*, en especial cuando en ella se refleja la inflación (cuando la serie corresponde a valores o precios) *que vale la pena tener presente;* por tal razón se debe corregir la serie por medio de la deflactación, como se *indicó* en el capítulo anterior sobre *n meros índices*

ÍNDICE ESTACIONAL MÉTODO DE RAZÓN RESPECTO A LA TENDENCIA

Se aconseja, antes de construir los *ndices estaciona es mensua es*, eliminar las variaciones dadas por la *endencia*.

Veamos con el siguiente ejercicio, uno de los procesos que generalmente se sigue, en cuanto a los precios de un determinado artículo.

Precios de un determinado artículo durante los meses de 2007 al 2011

MESES	2007	2008	2009	2010	2011
Enero	316	340	366	384	412
Febrero	300	320	342	376	392
Marzo	358	380	394	410	438
Abril	348	386	410	424	470
Mayo	370	404	424	450	476
Junio	378	406	414	444	476
Julio	358	382	410	442	482
Agosto	366	398	420	436	460
Septiembre	362	378	384	426	454
Octubre	376	412	430	452	502
Noviembre	384	418	430	434	504
Diciembre	458	482	502	554	612
TOTAL	**4.374**	**4.706**	**4.926**	**5.232**	**5.678**

a) Trabajando con los totales para cada año podemos determinar la *tendencia*, procedimiento más ágil que haciéndolo con todos los meses.

CÁLCULO DE LA TENDENCIA

AÑOS	$_i$	$_i$	^2_i	$_i$ $_i$
2007	4.374	0	0	0
2008	4.706	1	1	4.706
2009	4.926	2	4	9.852
2010	5.232	3	9	15.696
2011	5.678	4	16	22.712
Σ	24.916	10	30	52.966

$$b = \frac{5(52\,966) - (10)(2\;\;916)}{5(\;0) - 10^2} = \;\;1\;, \qquad c = \frac{2\;\;916 - \;\;1\;,\;(10)}{5} = \;\;\;56,$$

$$Y = \;\;1\;,\; x + \;\;\;56,$$

Nota: Los valores anuales corresponden a acumulados mensuales.

b) La ecuación anterior utilizada para hacer estimaciones anuales, la transformamos en otra ecuación que emplearemos para estimar valores mensuales. Como se observa, los valores para y de cada año son el resultado de la sumatoria de los 12 meses. El cambio se obtiene dividiendo a por 144 y a $\boldsymbol{c}$ por 12, de acuerdo a la tabla anterior.

$$Y = \frac{1}{1}x + \frac{6}{1} \qquad Y = \;\;1\;\;6\;\;\;x + \;\;6$$

Nota: los años corresponden a acumulados mensuales.

En la ecuación anterior, el origen estará situado el 10 de julio de 2007, siendo la unidad de tiempo = 11 meses y el precio mensual ( ). Decimos que la información se da al primero de Julio por ser la mitad del año, así que por comodidad trasladamos el origen al 15 de Enero del mismo año, siendo la ecuación resultante:

$\hat{Y}$ = 2,17639(-5,5) + 363,03333 = 351,06 valor que será el nuevo coeficiente, $\boldsymbol{c}$ = 351,06318 cuando: = 0 al 15 de Enero de 2007.
La ecuación resultante será: $\hat{Y}$ = 2,17639 + 351,06318.

c) Calculamos los valores estimados para cada mes, utilizando la ecuación anterior, haciendo que = 0 para enero de 2007; = 1 Febrero del mismo año; = 2 Marzo y asi sucesivamente. Estos valores aparecen en la tabla siguiente (por comodidad trabajamos con un decimal).

Cálculo de los valores estimados

MESES	2007	2008	2009	2010	2011
Enero	351,0	377,2	403,2	429,4	455,6
Febrero	353,2	379,4	405,4	431,6	457,8
Marzo	355,4	381,6	407,6	433,8	459,8
Abril	357,6	383,8	409,8	436,0	462,0
Mayo	359,8	385,8	412,0	438,2	464,2

ontin a

ontinuación

Junio	362,0	388,0	414,2	440,2	466,4
Julio	364,2	390,2	416,4	442,4	468,6
Agosto	366,2	392,4	418,6	444,6	470,8
Septiembre	368,4	394,6	420,8	446,8	473,0
Octubre	370,6	396,8	422,8	449,0	475,2
Noviembre	372,8	399,0	425,0	451,2	477,2
Diciembre	375,0	401,2	427,2	453,4	479,4

d) El paso a seguir, consiste en dividir los valores reales para cada mes por el respectivo valor de la Tendencia, cuyos resultados aparecen a continuación.

Índices para cada mes

MESES	2007	2008	2009	2010	2011
Enero	0,9003	0,9014	0,9077	0,8443	0,9043
Febrero	0,8494	0,8434	0,8436	0,8712	0,8563
Marzo	1,0073	0,9958	0,9660	0,9451	0,9526
Abril	0,9732	1,0057	1,0050	0,9725	1,0173
Mayo	1,0283	1,0472	1,0291	1,0269	1,0254
Junio	1,0442	1,0464	0,9995	1,0086	1,0206
Julio	0,9830	0,9790	0,9846	0,9991	1,0286
Agosto	0,9995	1,0143	1,0033	0,8807	0,9771
Septiembre	0,9826	0,9579	0,9125	0,9534	0,9598
Octubre	1,0146	1,0383	1,0170	1,0067	1,0564
Noviembre	1,0300	1,0476	1,0118	0,9618	1,0562
Diciembre	1,2213	1,2014	1,1751	1,2219	1,2766

e) En la penúltima columna aparece el *promedio mensua* para los cinco años, así que para ello sumamos las 5 cantidades correspondientes al mes de Enero, siendo iguales a 4,508; luego dividimos entre 5, con resultados de 0,9016.

PROMEDIO MENSUAL	ÍNDICE ESTACIONAL
0,9016	90,25
0,8528	85,37
0,9735	94,75
0,9938	99,48
1,0314	103,25
1,0239	102,05
0,9949	99,59

ontin a

ontinuación

0,9950	99,60
0,9532	95,42
1,0266	102,77
1,0215	102,26
1,2193	122,06
11,9875	1.200,00

- La suma de la penúltima columna nos dará 11,9875 y la dividimos entre 12: 11,9875/ 12 = 0,998958.
- Luego cada *promedio mensua* lo dividimos por 0,998958 y lo multiplicamos por 100; nos resulta el Índice tal como aparece en la última columna.

Si se trabaja la última columna, con todos los decimales la suma debe ser igual a 1.200,00

ÍNDICES ESTACIONALES
MÉTODO DE LAS RAZONES AL PROMEDIO MÓVIL

Consideremos los precios bimestrales para un determinado artículo, durante los últimos diez años.

Precios de un artículo para el período 2002 al 2011

BIMESTRES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	390	390	412	422	458	526	556	534	578	468
2	304	308	330	354	408	458	456	424	450	410
3	312	318	340	376	446	478	444	450	460	426
4	354	370	386	430	492	538	542	544	522	516
5	412	430	454	488	558	610	614	596	568	492
6	460	488	498	532	616	650	646	652	580	684

Generalmente se consideran los doce meses del año para calcular los *ndices estaciona es* o si se desea desestacionalizar una serie, sin embargo con fines de simplificación se ha dado la información por bimestre, pero el procedimiento a seguir con 6 o 12 períodos es casi el mismo.

Los pasos requeridos al utilizar el Método de las razones al promedio (media) móvil, son:

a) Lo primero que hacemos consiste en calcular los Totales Móviles tomando de seis en seis bimestres (se podría, en el caso de tener los 12 meses, de doce en doce). Como seis o doce son números par, cada total móvil se debe ubicar en la mitad de los dos períodos (bimestres o meses).

   Siendo 390 + 304 +312 + 354 + 412 + 460 = 2.232 este valor se sitúa entre el 3° y 4° bimestre; luego será: 304 + 312 + 354 + 412 + 460 + 390 = 2.232, situándolo entre el 4° y 5° bimestre y así sucesivamente, quitando el primer período a la suma y agregando el siguiente valor al último de la serie anterior, o sea 354 + 412 + 460 + 390 + 308 = 2.236.

Totales móviles

BIMESTRES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1→	–	2.258	2.386	2.534	2.824	3.174	3.258	3.212	3.258	2.968
2→	–	2.276	2.410	2.568	2.894	3.226	3.262	3.194	3.230	2.892
3→	2.232	2.304	2.420	2.602	2.978	3.260	3.258	3.200	3.158	2.996
4→	2.232	2.326	2.430	2.638	3.046	3.290	3.236	3.244	3.048	–
5→	2.236	2.348	2.454	2.692	3.096	3.288	3.204	3.270	3.008	–
6→	2.242	2.370	2.490	2.762	3.128	3.254	3.210	3.280	2.974	–

b) Luego procedemos a centrar la información, es decir, que e *tota* corresponda a cada período o bimestre (puede ser mensual), para ello se obtienen unos nuevos *tota es mó i es* con dos bimestres (o dos meses), proceso conocido como Totales Móviles Centrados. El proceso será: 2.232 + 2.232 = 4.464 y lo situamos en el 4° bimestre, luego 2.232 + 2.236 = 4.468 y corresponderá al 5° bimestre; 2.236 + 2.242 = 4.478 correspondiente al 6° bimestre y así sucesivamente.

Totales móviles centrados

BIMESTRES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	–	4.500	4.756	5.024	5.586	6.302	6.512	6.422	6.538	5.942
2	–	4.534	4.796	5.102	5.718	6.400	6.520	6.406	6.488	5.860
3	–	4.580	4.830	5.170	5.872	6.486	6.520	6.394	6.388	5.888
4	4.464	4.630	4.850	5.220	6.024	6.550	6.494	6.444	6.206	–
5	4.468	4.674	4.884	5.330	6.142	6.578	6.440	6.514	6.056	–
6	4.478	4.718	4.944	5.454	6.224	6.542	6.414	6.550	5.982	–

c) Seguimos con el cálculo de los *promedios* (medias) *mó i es*, dividiendo cada total entre 12, ya que cada uno de los totales es el acumulado de dos sumandos de la tabla, y cada sumando es el resultado de seis sumandos o seis períodos. Si se hubiese trabajado con los doce meses dividiríamos cada total por 24. El procedimiento seguido es: 4.464 ÷ 12 = 372; 4.468 ÷ 12 = 372 y así sucesivamente; se podrá observar que se han despreciado los decimales, por dos motivos: espacio y agilización de las operaciones.

Promedios (medias) para cada período

BIMESTRES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	–	374	396	418	466	526	542	536	544	496
2	–	378	400	426	476	534	544	534	540	488
3	–	382	402	430	490	540	544	532	532	490
4	372	386	404	434	502	546	542	536	518	–

5	372	390	406	444	512	548	536	542	504	–
6	374	394	412	454	518	546	534	546	498	–

Con este procedimiento se logró eliminar las *ariaciones estaciona es*, dado que el período de la media móvil presenta coincidencia con la periodicidad de esas variaciones. Por otra parte, vale la pena recordar que los *promedios* o medias *mó i es* son una forma de suavizar una serie, por tanto suponemos que se han eliminado aquellas variaciones consideradas como aleatorias o accidentales.

d) Ahora se calcularán los *porcentajes* o *ra ones* a los promedios o medias móviles presentados en el cuadro anterior, mediante una operación sencilla, dividiendo a cada uno de los precios reales, que anteceden al ordinal (precios de un artículo para el período 2002 - 2011) por el respectivo promedio móvil, (promedios para cada período) y se obtienen los cuocientes o relativos para cada período, los cuales multiplicados por 100, los convertimos en porcentajes.

Ejemplos: (354 ÷ 372) 100 = 95,16 ó 95,2; (412 ÷ 372) 100 = 110,75 ó 110,8; (460 ÷ 374) 100 = 122,99 o 123,0 y así sucesivamente. Por comodidad trabajaremos con un decimal.

Índices estacionales

BIMESTRES	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	–	104,3	104,0	101,0	98,3	100,0	102,6	99,6	106,3	94,4
2	–	81,5	82,5	83,1	85,7	85,8	83,8	79,4	83,3	84,0
3	–	83,2	84,6	87,4	91,0	88,5	81,6	84,6	86,5	86,9
4	95,2	85,9	95,5	99,1	98,0	98,5	100,0	101,5	97,4	–
5	110,8	110,3	111,8	109,9	109,0	111,3	114,6	110,0	112,7	–
6	123,0	129,9	120,9	117,2	118,9	119,0	121,0	119,4	116,5	–

Los porcentajes obtenidos en el cuadro anterior son considerados *índices estaciona es*, cuando las variaciones para los períodos correspondientes a una "estación" son cambiantes o evolutivas en el transcurso de los años estudiados; en muchos casos estos cambios se deben a la variación en los hábitos de consumo.

En caso contrario, es decir, cuando son estables para esos períodos, los *índices de ariación estacio-na* se pueden tomar como promedio para cada período, en este caso, para cada bimestre.

Índice estacional

BIMESTRE	1	2	3	4	5	6
ÍNDICE	101,2	83,3	86,1	98,0	111,3	120,1

AJUSTE ESTACIONAL

Se procedió anteriormente al cálculo del *ndice estaciona* para cada período y año. Ahora procedemos a calcular la variable pero desestacionalizada, cuyo proceso es muy sencillo, dividiendo las cantidades, precios o valores reales, es decir, los valores observados, por los respectivos índices estacionales.

Consideremos los dos ejemplos anteriores, ventas y precios, para desestacionalizarlos. En el primer ejemplo (ventas) se tomará en consideración el año de 2011; para ello el primer mes desestacionalizado será: 412 ÷ 0,9025 = 456,51; para Febrero será: 392 ÷ 0,8537 = 459,18; Marzo 438 ÷ 0,9475 = 462,27 y así sucesivamente. Veamos los resultados en la tabla siguiente:

Ventas desestacionalizadas para los meses de 2011

MESES	VENTAS REALES	ÍNDICE ESTACIONAL	VENTAS DESESTACIONALIZADAS
Enero	412	90,25	456,51
Febrero	392	85,37	459,18
Marzo	438	94,75	462,27
Abril	470	99,48	472,46
Mayo	476	103,25	461,02
Junio	476	102,05	466,44
Julio	482	99,59	483,98
Agosto	460	99,60	461,85
Septiembre	454	95,42	457,79
Octubre	502	102,77	488,47
Noviembre	504	102,26	492,86
Diciembre	612	122,06	501,39

Precios desestacionalizados para los bimestres - 2011

BIMESTRE	PRECIOS	ÍNDICE ESTACIONAL	PRECIOS DESESTACIONALIZADOS
1	468	101,2	462,46
2	410	83,3	492,20
3	426	86,1	494,77
4	516	98,0	526,53
5	492	111,3	442,05
6	684	120,1	569,53

$$\frac{468}{1,012} = 462,46$$

$$\frac{410}{0,833} = 492,20\ etc$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

77. Con los siguientes datos, correspondientes a miles de unidades vendidas por mes, durante cinco años.

MESES	2007	2008	2009	2010	2011
Enero	520	610	716	750	805
Febrero	410	520	620	690	754
Marzo	300	480	600	650	712
Abril	380	500	450	520	610
Mayo	260	320	410	480	680
Junio	390	390	380	410	520
Julio	480	494	490	560	610
Agosto	400	450	500	520	540
Septiembre	280	320	412	470	510
Octubre	250	300	400	430	680
Noviembre	360	412	570	694	784
Diciembre	760	814	850	910	1.020

Se pide:

a) Calcular los Índices estacionales aplicando las tres formas desarrolladas, en el capítulo.

b) Desestacionalizar la serie.

78. La información siguiente corresponde al precio de un artículo en los últimos cinco años. La serie se da en bimestres

BIMESTRES	2007	2008	2009	2010	2011
Primero	$350	410	605	750	960
Segundo	$220	380	520	700	910
Tercero	$200	360	480	650	870
Cuarto	$360	420	590	720	980
Quinto	$450	580	760	870	1.080
Sexto	$620	750	880	1.020	1.120

a) Calcular los Índices por los tres procedimientos que se vieron.

b) Desestacionalizar la serie.

Síntesis de capítulo

En la vida diaria, se va registrando numerosa información a medida que se produce, formando series en el tiempo. El análisis de estas series trata de describir su comportamiento pasado y predecir su comportamiento futuro, de acuerdo con los datos registrados a través del tiempo.

Parte del análisis de estas series consiste en descomponer la variable en sus cuatro factores: *tendencia ariaciones estaciona es cíc icas* y *a eatorias* Los componentes más importantes en el análisis de la serie son la *tendencia* y la *ariación estaciona* .

La *tendencia* describe mediante una línea *recta parabó ica e ponencia* o de cualquier otro tipo, buscando que sea la mejor línea que represente al conjunto de datos. Por lo tanto la línea de tendencia podrá ser ascendente o descendente.

Los pronósticos o estimaciones futuras deben hacerse a corto plazo, ya que se considera que el comportamiento de los datos será igual al del período de predicción. Como máximo sería a 10 años, lo aconsejable uno o dos períodos.

Es importante conocer e identificar las fluctuaciones estacionales a fin de poder eliminarlas en aquellos casos en que sea posible y necesario.

La *tendencia* la podemos determinar por el método de los mínimos cuadrados; también, utilizando el método de los promedios móviles, mediante el procedimiento denominado "mano alzada" o aplicando los promedios móviles, siendo esta última una forma de alisar o suavizar las fluctuaciones hasta llegar a conformar una línea.

Como en la regresión que se verá en el Capítulo siguiente, se podrá determinar el "error en la estimación", que permite fijar límites dentro de los cuales estará el valor real con cierto grado de confiabilidad. También se podrá calcular el coeficiente de correlación, que determina el grado o bondad del ajuste de estimación.

12

CAPÍTULO

Regresión y Correlación

Si un hombre comien a con certe as acabará con dudas pero si comien a con dudas acabará con certe as.

Francis acón

CONTENIDO

- Términos, fórmulas, cálculos, definiciones, y conceptos generales.
- Regresión Lineal, Parabólica y Exponencial simple. Límites de confianza.
- Regresión Lineal, Parabólica y Exponencial ponderada.
- Regresión y correlación múltiple. Correlación por Rangos (Spearman).
- Pruebas de los Coeficientes de regresión y correlación.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Conocer, entender y aplicar correctamente cada uno de los métodos de Regresión y correlación.
- Valorar la importancia que tiene la Regresión en la toma de decisiones. No sólo como asignatura, sino en el ejercicio profesional.
- Calcular e interpretar los resultados, los errores de estimación, límites de confianza y Coeficientes de regresión y correlación.
- Distinguir el uso de la correlación de Spearman y la de Pearson.

ASPECTOS GENERALES

En el desarrollo de la mayoría de los capítulos que antecedieron al presente a excepción del capítulo anterior, nos hemos venido refiriendo o analizando únicamente una variable, es decir, se realiza una muestra aleatoria con la cual se obtiene información para una o varias características, que permiten hacer inferencias acerca de cada uno de los parámetros en forma independiente, denominadas: *distribuciones unidimensiona es* o *datos uni ariantes*, dado que se analiza cada variable por separado.

En este capítulo se tendrán dos casos. En la primera parte se estudiará la relación que puede haber entre dos variables, analizadas símultáneamente, por tal razón se les clasifica como *distribuciones bidimensiona es* o datos *bi ariantes*; en la segunda parte se estarán relacionando más de dos variables símultáneamente y nos referimos a distribuciones *p uridimensiona es* o *mu ti ariantes*.

REGRESIÓN Y CORRELACIÓN

Son numerosos los ejemplos que se pueden dar al trabajar con dos variables que están relacionadas: producción y consumo; ingresos y gastos; horas trabajadas y accidentes de trabajo; horas de estudio y rendimiento académico, etc. El ejemplo más común que lo practicamos diariamente: peso y estatura. Pero no basta con disponer de la información que ha sido recolectada, es necesario establecer si en realidad existe relación entre ellas, procedimiento que denominaremos como *aná isis de regresión corre ación*, para poder estimar una de ellas en función de la otra, que suponemos conocida.

La relación que puede existir entre dos variables las podemos clasificar de la siguiente forma:

- *Dependencia causa uni atera* . Esta relación se da cuando una de las variables influye en la otra, pero no al contrario.
- *Interdependencia*. Se presenta cuando la influencia entre las dos variables es recíproca. Sería un caso de *dependencia bi atera* .
- *Dependencia indirecta*. Dos variables pueden mostrar una correlación a través de una *tercera ariab e* que influye en ellas.
- *oncordancia*. Se presenta en *dos ariab es independientes* a las cuales se les determina la correlación que puede existir.
- *o ariación casua* . Cuando la correlación que se presenta entre las dos variables es totalmente *casua* o *accidenta*

La mayor parte de este capítulo hará referencia a casos de dependencia bilateral o interdependencia.

Regresión: el término de regresión fue originariamente utilizado por Galton para indicar ciertas relaciones en la teoría de la herencia biológica aunque con posterioridad ha llegado a significar el método estadístico desarrollado para investigar tales relaciones.

El *aná isis de regresión* da lugar a una ecuación matemática que nos permite describir la relación existente entre dos variables. Es decir, obtener una línea "ideal" conocida como línea de regresión, que nos describa la relación o dependencia entre dos variables.

Esta línea o función matemática, en el caso de una sola variable independiente o explicativa puede ser expresada, a través de una:

Recta o función lineal: $\hat{Y} = b + c$

Parábola de segundo grado: $\hat{Y} = a^2 + b + c$

Función exponencial: $\hat{Y} = cb$

Función potencia: $\hat{Y} = c^b$

Con estas cuatro funciones se puede resolver la mayoría de las situaciones que se nos presentan en la vida diaria, sin embargo, vale la pena tener presente que existen muchos más tipos de funciones, que deben ser aplicadas en casos especiales.

El *análisis de correlación* nos describe el grado o fuerza con que se produce esta relación, para ello utilizaremos una medida conocida como *coeficiente de correlación* o *correlación de Pearson*.

Así como *el análisis de correlación* permite medir la fuerza de asociación entre dos variables, el *análisis de regresión* permite la predicción o sea la estimación de un valor o promedio de una variable denominada dependiente, con base en un valor o promedio supuestamente conocido para la otra variable, denominada independiente.

Hay que tener claridad, que el análisis de regresión, además de explicar la relación entre dos variables, de causa y efecto, nos indica si la relación matemática puede ser *lineal parabólica exponencial* etc., además, nos permite estimar los valores de una variable, suponiendo conocido un valor de la otra variable.

Antes de iniciar el *análisis de regresión* recordemos algunos cálculos de *medidas de posición* y de *dispersión*, indispensable para el desarrollo de este capítulo y que fueron vistas en los primeros capítulos.

Llamaremos $x_1, x_2, x_3, \ldots\ldots x_n$ los valores observados en la primera variable X y $y_1, y_2, y_3, \ldots y_n$ para la segunda variable Y. Se tendrá pares de observaciones, de tal manera que para x_i, se tendrá un valor y_1; x_2 corresponderá a y_2, y así sucesivamente, de esta forma se tendrá n pares de observaciones.

Tabla 12.1

x_i	y_i
x_1	y_1
x_2	y_2
x_3	y_3
x_4	y_4
.	.
.	.
.	.
x_n	y_n

- Para cada variable se podrán calcular las medias aritméticas, las varianzas y sus desviaciones típicas, además, se tendrá una nueva medida de dispersión, denominada *covarianza*.

- Fórmulas:

Medias:	$\bar{x} = \frac{\sum x_i}{n}$	$\bar{y} = \frac{\sum y_i}{n}$
Varianzas:	$s_x^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n}$	$s_y^2 = \frac{\sum y_i^2 - n\bar{y}^2}{n}$
Desviaciones típicas:	$s_x = \sqrt{s_x^2}$	$s_y = \sqrt{s_y^2}$

Ejemplo 1. Supongamos la realización de una investigación por muestreo aleatorio a 10 empresas, que producen un determinado producto, en cuanto al valor de la producción (miles mill $ anuales) y el costo del mismo (miles mill de $). Los resultados fueron:

Tabla 12.2 Cálculo de las Medias, Varianzas y Covarianza en la Regresión.

PRODUCCIÓN (miles mill)	COSTO (miles mill)
$_i$	$_i$
10	3
18	5
12	4
16	5
22	8
36	12
30	10
32	14
26	12
12	3

Calculemos las medidas (media, varianza, desviación típica y varianza para el ejemplo 1).

$_i$	$_i$	^2_i	^2_i	$_i$ $_i$	($_i$ – –)	($_i$ – –)
10	3	100	9	30	-11,4	-4,6
18	5	324	25	90	-3,4	-2,6
12	4	144	16	48	-9,4	-3,6
16	5	256	25	80	-5,4	-2,6
22	8	484	64	176	0,6	0,4
36	12	1.296	144	432	14,6	4,4
30	10	900	100	300	8,6	2,4
32	14	1.024	196	448	10,6	6,4
26	12	676	144	312	4,6	4,4
12	3	144	9	36	-9,4	-4,6
214	76	5.348	732	1.952	0	0

$$\bar{x} = \frac{\sum x_i}{n} = \frac{214}{10} = 21{,}4 \qquad s_x^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n} = \frac{5.348 - 10(21{,}4)^2}{10} = 76{,}84 \qquad s_x = 8{,}76$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{76}{10} = 7{,}6 \qquad s_y^2 = \frac{\sum y_i^2 - n\bar{y}^2}{n} = \frac{732 - 10(7{,}6)^2}{10} = 15{,}44 \qquad s_y = 3{,}93$$

a co arian a: se simboliza por *co ; m* o *m* . Es una medida de dispersión, a diferencia de las varianzas que siempre dan resultados positivos, ésta puede ser negativa o positiva.

Se define como "la media del producto de as diferencias entre os a ores de a ariab e su media aritmética nos determina a ariabi idad conjunta de X y *Y*:

$$o \qquad \frac{\sum z_x z_y}{n} = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{n}$$

$$Cov = \frac{325{,}6}{10} = 32{,}56$$

$(x_i - \bar{x})\ (y_i - \bar{y})$	$(x_t - \bar{x})^2$
52,44	129,96
8,84	11,56
33,84	88,36
14,04	29,16
0,24	0,36
64,24	213,16
20,64	73,96
67,84	112,36
20,24	21,16
43,24	88,36
325,60	768,40

También se puede calcular de una manera más sencilla y más rápida:

$$Cov = \frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y} \qquad Cov = \frac{1.952}{10} - (21{,}4)(7{,}6) = 32{,}56$$

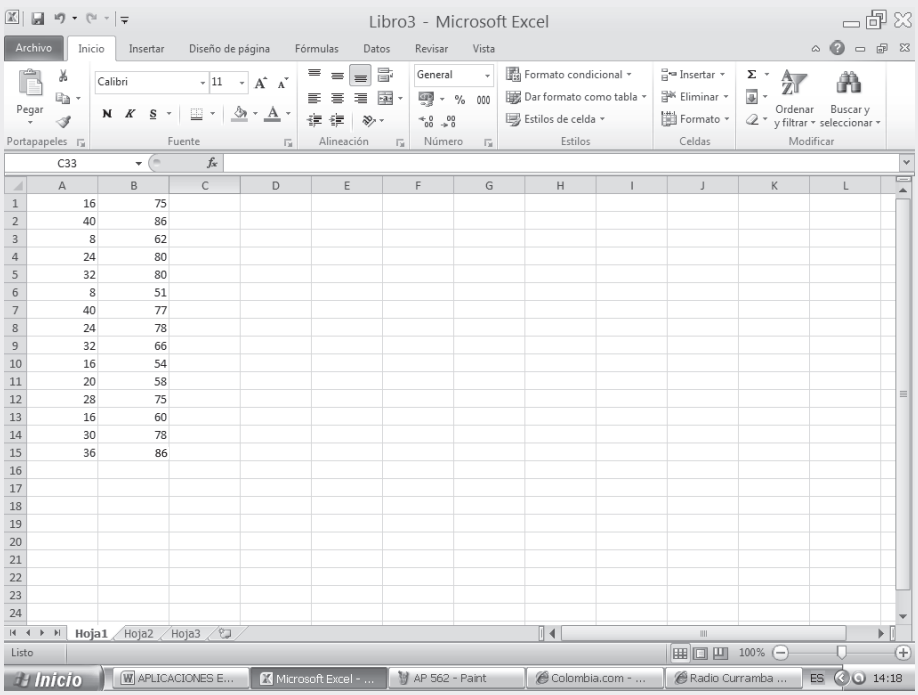
Hay quienes utilizan la *co arian a* como norma para medir la *dependencia inea* pero no es lo más indicado, dado que no existe ninguna unidad de medida estandarizada. Se dice, que si la *co arian a* es positiva nos estará indicando una pendiente positiva, es decir, en el caso de una recta ésta sería ascendente, y si por el contrario es negativa la línea sería descendente. Esta presentación puede ser visualizada. Si la mayoría de los puntos se sitúan en las secciones inferior izquierda y superior derecha, la *co arian a* es positiva; si por el contrario, los puntos se sitúan en la parte inferior derecha y superior izquierda, es indicativo de una *co arian a* negativa, tal como se puede observar en un *diagrama de dispersión*

CÁLCULO DE LA COVARIANZA UTILIZANDO EXCEL

Complemento virtual SIL

Ejemplo 2. En una hoja de cálculo, tecleamos quince (15) pares de observaciones para dos variables, cuyos datos fueron seleccionados en forma arbitraria, sólo para mostrar cómo se obtiene la COVARIANZA, aplicando Excel.

Figura No. 1. Muestra de Datos.

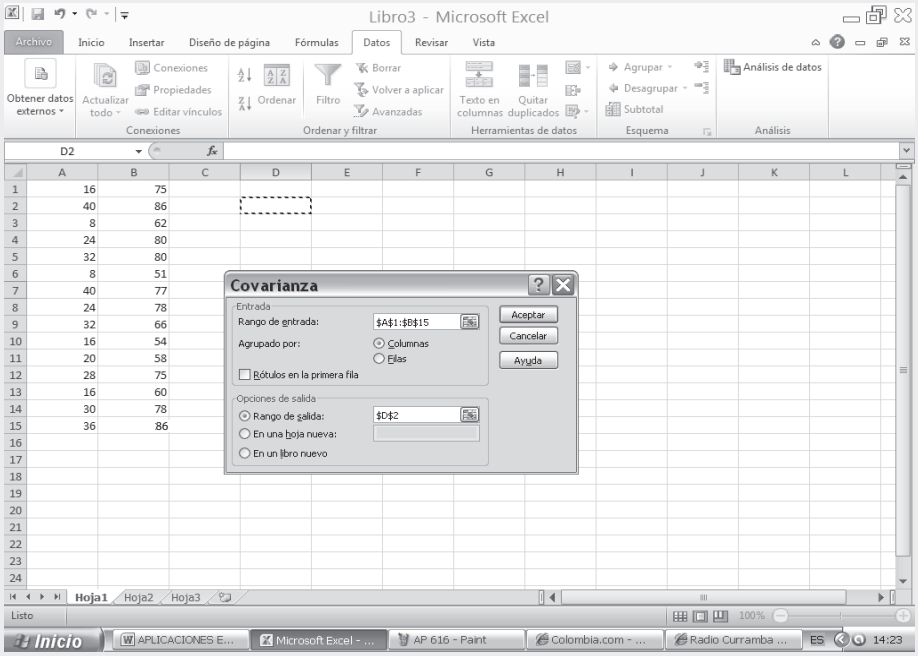


	A	B
1	16	75
2	40	86
3	8	62
4	24	80
5	32	80
6	8	51
7	40	77
8	24	78
9	32	66
10	16	54
11	20	58
12	28	75
13	16	60
14	30	78
15	36	86

- Seleccionamos la ficha DATOS y hacemos CLIC en la opción ANÁLISIS DE DATOS.
- En la lista que ofrece el menú ANÁLISIS DE DATOS, seleccionamos la opción COVARIANZA. Aparece el siguiente cuadro de diálogo:
- RANGO DE ENTRADA, que en nuestro caso será A1:B15.
- RANGO DE SALIDA, seleccionamos la celda D2.

En la SEGUNDA PARTE DEL CD que acompaña al libro encontrará en conjunto toda la aplicación del EXCEL a la ESTADÍSTICA, que igual a1 capítulo 13, ofrece una BASE DE DATOS para trabajar

Figura No. 2. Covarianza.



- Finalmente hacemos CLICK en ACEPTAR, obteniendo los siguientes resultados:

Figura No. 3. Cuadro de Resultados.

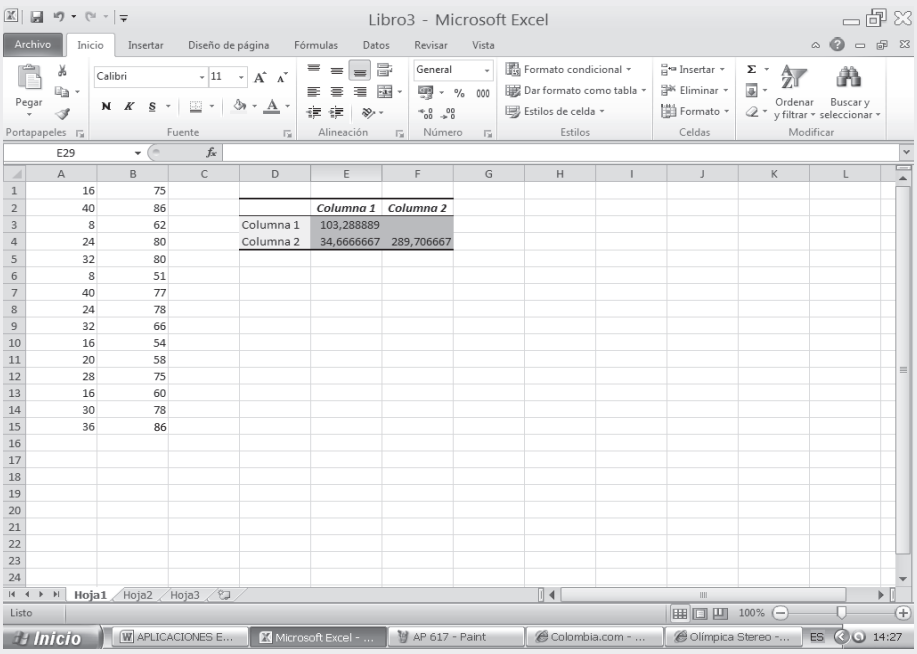


DIAGRAMA DE DISPERSIÓN

Los datos de la distribución bidimensional, anteriormente dada en la tabla No. 12.2, puede ser representada gráficamente en un par de ejes coordenados, considerando al eje de las abscisas para la primera variable ($_i$) y al eje vertical o de las ordenadas, para los valores de la segunda variable ($_i$).

En un *p ano cartesiano*, se presentan tantos puntos como pares de observaciones se tengan, correspondiendo cada punto a un par de observaciones; a esta representación gráfica se le denomina indistintamente: *diagrama de desparramiento de dispersión de esparcimiento o nube de puntos*.

El problema consiste ahora en unir varios puntos de ese *conjunto o nube* de puntos, mediante una línea ya sea *recti íneo parabó ico e ponencia* o cualquier otro tipo que represente al conjunto. En algunos casos esos puntos estarán condensados alrededor de ella, en otros presentará dispersión pequeña o grande; en este último caso se pueden producir grandes errores, como consecuencia de que el ajuste realizado no sea el más indicado. El tipo de línea que se relaciona, dependerá de la forma que asuma el conjunto de puntos al hacer la respectiva gráfica, siendo ésta una de las maneras de visualizar mejor; también de acuerdo a la experiencia del estadístico se puede determinar, cuál debe ser mejor ajuste, ya sea mediante el *método gráfico* (*mano a ada*) o el *ana ítico*.

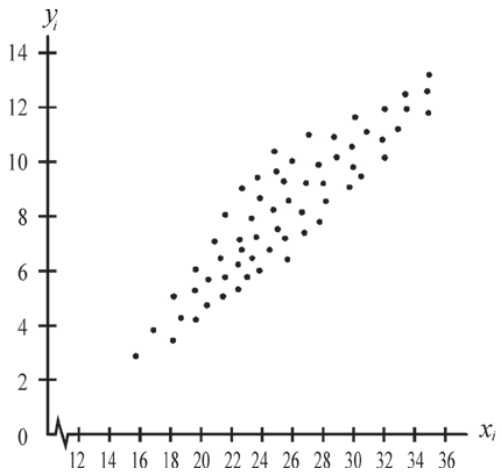
Lo malo de el *método gráfico*, es el del ser muy subjetivo, por lo tanto, el resultado dependerá de la persona que haga el análisis.

Una vez elaborada la gráfica surgen los siguientes interrogantes:

¿Qué línea se va a dibujar?, Recta?, Curvilínea?, Exponencial?... El criterio técnico y la experiencia serán de buena ayuda, pero se deben tener en cuenta algunos aspectos:

Gráfica 12.1 Nube de puntos o Diagrama de dispersión.

- Observar c*uidadosamente* la forma que toma el conjunto de puntos.
- La línea debe reflejar la mejor posible tendencia de los puntos en la gráfica.
- La línea debe representar al conjunto de puntos, por lo tanto debe pasar por el centro de esa nube. La línea debe ser, hasta donde sea posible, la más sencilla.



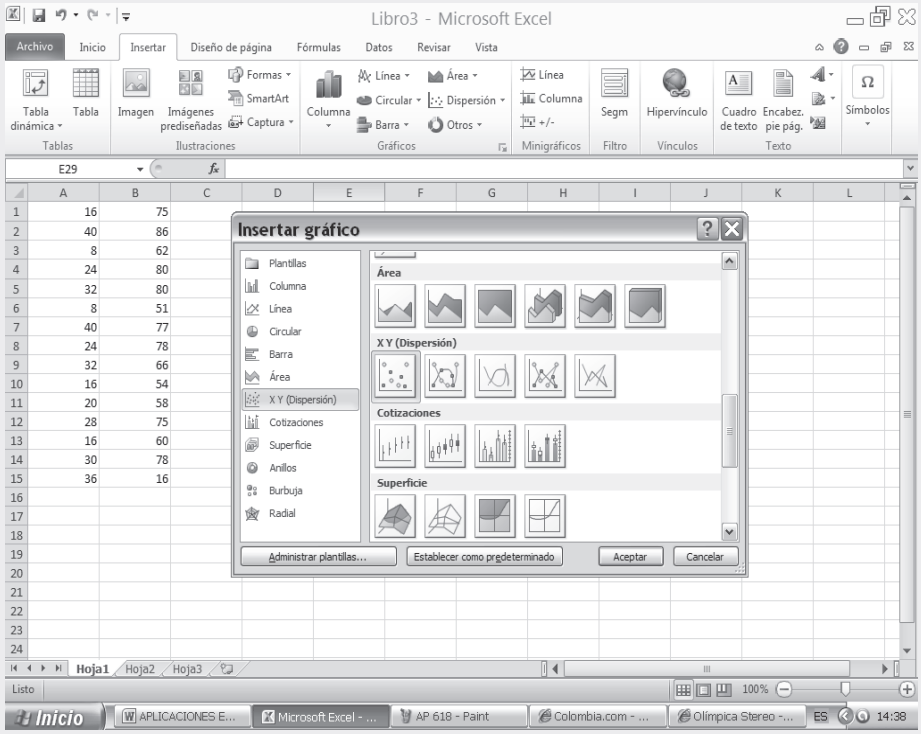
APLICACIÓN DE EXCEL EN LA ELABORACIÓN

DIAGRAMA DE DISPERSIÓN

Tal como procedimos en la elaboración de GRÁFICAS (ver página 101), se procederá a realizar la GRÁFICA DE DISPERSIÓN, utilizando para ello la información de los 15 pares de observaciones, con las cuales calcularemos la COVARIANZA y el COEFICIENTE DE CORRELACIÓN (ver páginas 576 y 586). Recordando los pasos utilizados anteriormente en la elaboración de Gráficas, se tiene:

- Activamos las CELDAS que contienen la información de las dos variables incluidos TÍTULOS y RÓTULOS
- Procedemos a hacer CLICK sobre la pestaña del grupo Gráficos en el menú INSERTAR y seleccionamos el TIPO DE GRÁFICO, en este caso, el de DISPERSIÓN, seleccionando el primero (1) dentro de las cinco (5) posibilidades que ofrece Excel.

Figura No. 4. Insertar Gráfico.



- Observemos el gráfico denominado DIAGRAMA DE DISPERSIÓN o nube de puntos, el cual permite visualizar el tipo de línea que más ajusta a ese conjunto de puntos.

Figura No. 5. Resultados Gráfico de Dispersión.

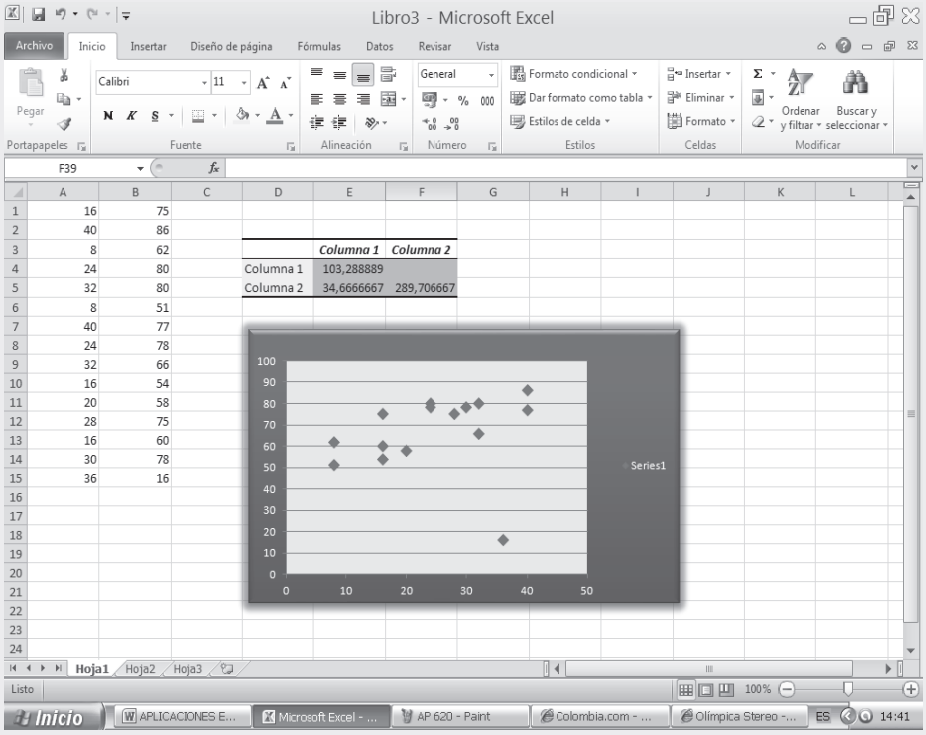
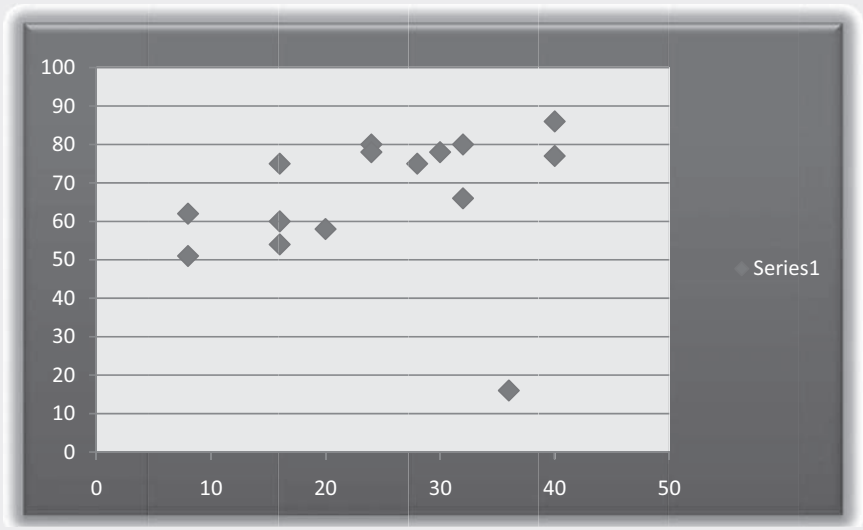


Gráfico 2.Gráfico de Dispersión XY.



REGRESIÓN Y CORRELACIÓN LINEAL

Generalmente en el análisis de *regresión* se utiliza la *ínea recta* por su simplicidad en el cálculo matemático, además, en muchos casos de la vida real nos proporciona aproximaciones suficientes para ser aceptada. El modelo matemático que describe una relación lineal cuando se estima el valor de ***y*** en función de ***x*** está representado indistintamente por:

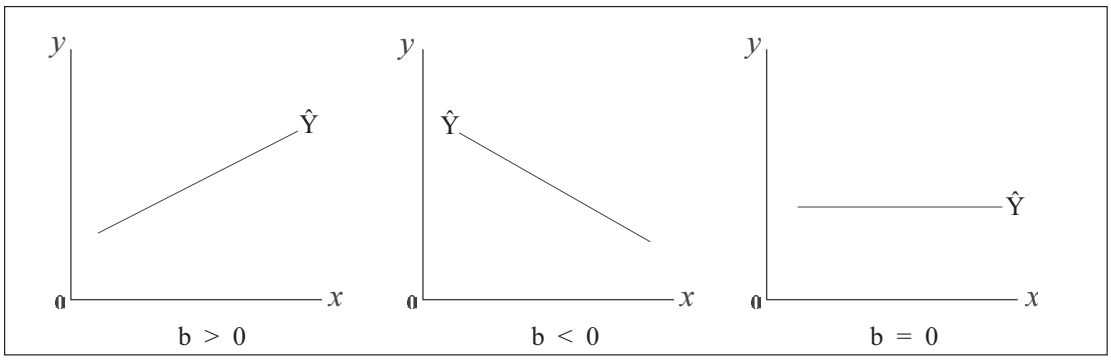
$$\hat{Y} = bx + c \quad \text{ó} \quad \hat{Y} = Bx + c \quad \text{ó} \quad \hat{Y} = \beta_1 x + \beta_0 \quad \text{ó} \quad \hat{Y} = \beta x + A$$

En las anteriores ecuaciones encontramos cuatro componentes:

- $\hat{Y}$ es la variable que se va a estimar en función de otra variable (X) supuestamente conocida. Se le denomina también como *ariab e dependiente*, explicada o predictando.
- es la variable cuyo valor que supuestamente se conoce. Se le denomina *ariab e independiente*, *predictor* o *e p icati a*.
- $b = B = \beta_1$ es la *pendiente*, la que nos determina el ángulo de inclinación de la recta. Denominada también *coeficiente angular*, que nos permite cuantificar la cantidad que aumenta o decrece $\hat{Y}$, por cada valor que toma la variable independiente (X) o explicativa.

El coeficiente angular puede ser representado de las siguientes formas:

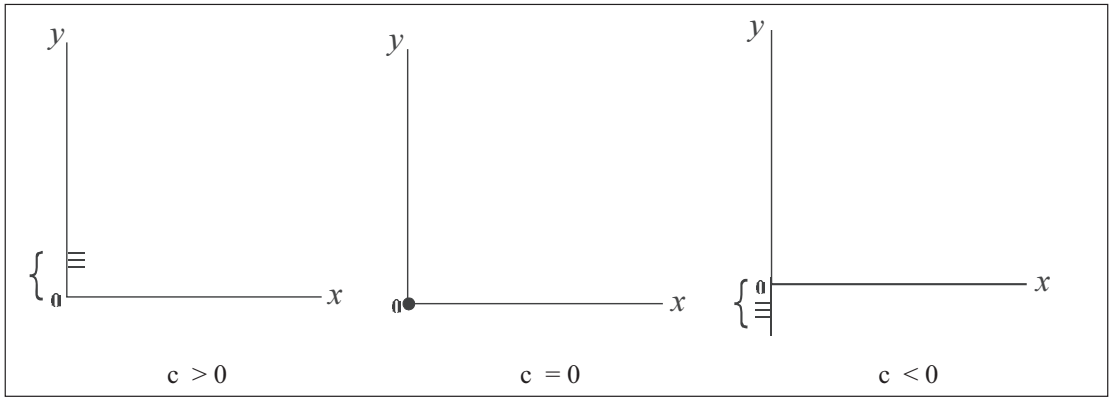
Figura 12.2 Representación del coeficiente angular o pendiente



Si *b* es un valor mayor que 0, es decir positivo, nos indicará que la recta es ascendente; si *b* es menor que 0 la recta será descendente y, si *b* es igual a 0 será una paralela al eje horizontal.

- $A =$ $= \beta_0$ son símbolos que se utilizan para indicar el *coeficiente de posición* u origen en la ordenada. Se dice que es un punto localizado en el eje de la ordenada, factor constante que se incluye en la ecuación, siendo igual a $\hat{Y}$ cuando $= 0$. El coeficiente de posición puede ser mayor, menor o igual a 0, como se puede observar en las siguientes figuras.

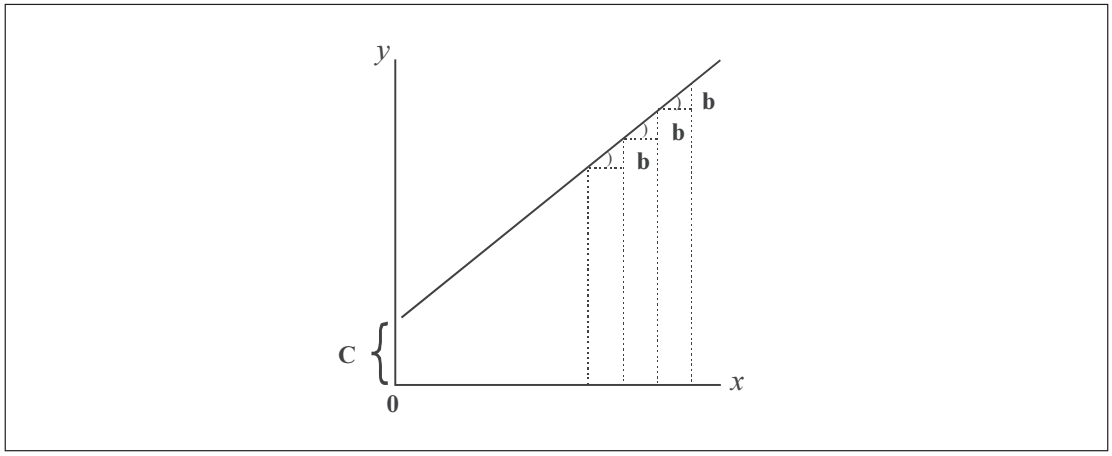
Figura 12.3 Ubicación del coeficiente de posición



En el primer caso será un punto por encima del origen, en el segundo pasará por el origen y en el tercero estará por debajo del origen.

Ajuste de una recta

Figura 12.4 Ajuste de una recta



En el desarrollo de este capítulo, utilizaremos letras minúsculas *b* y *c* para simbolizar los coeficientes de regresión, dado que en general las letras mayúsculas y o las letras del alfabeto griego β_1 y β_0 sirven para simbolizar *parámetros*, es decir, coeficientes de regresión poblacionales y no muestrales, como es el objetivo de este tema.

El problema ahora consiste en calcular los coeficientes de regresión muestrales, que son estimadores de los parámetros o coeficientes de regresión poblacional. En la práctica, la dificultad radica en determinar el criterio que nos sirva para definir el "*mejor ajuste*", es decir, la mejor línea que represente al conjunto de puntos.

Esa es la razón de hallar los valores de *b* y *c* de modo que $\hat{Y}_i$ sea lo más cercano posible a los valores observados de $_i$, para ello el procedimiento de cálculo más indicado es *e método de os mínimos cuadrados*.

Método de los mínimos cuadrados

Es un método preferido de estimación, por el cual las cantidades a estimar son determinadas, minimizando la suma de los cuadrados de las diferencias entre los valores observados ($_i$) y los valores estimados $\hat{Y}_i$

$$\sum e_i^2 = \sum (\ _i - \hat{Y}_I)^2 = mínimo$$

e es considerado como el *error* que se puede cometer en la estimación, el cual debe ser mínimo frente a otros modelos, para que proporcione una buena representación de esos puntos por la línea de regresión establecida. Es de importancia o de preferencia en dos casos:

- (a) Cuando se buscan estimadores lineales con varianza mínima (ver teorema de Gauss Markoff).
- (b) Cuando el modelo implica errores normalmente distribuidos y la estimación por mínimos cuadrados llega a ser equivalente a la estimación por máxima verosimilitud.

Procedimiento de cálculo de los coeficientes de regresión b y c (β_1 y β_0)

De acuerdo con la ecuación general de la recta $Y = b$ c, en la cual tenemos como incógnitas (b y c), requiere para su solución de dos ecuaciones normales; para ello obtenemos la primera ecuación anteponiendo a cada término el signo de la sumatoria. La segunda ecuación se logra multiplicando la ecuación general por , es decir, $= b^{\ 2} + c$ y luego se les antepone el signo de la sumatoria.

En el desarrollo de las dos ecuaciones, es conveniente recordar las propiedades de la sumatoria.

Reemplacemos en las dos ecuaciones, utilizando la información obtenida en el ejemplo 1:

(1) $\sum y_t = b\sum x_t + nc$
(2) $\sum y_t x_t = b\sum x_t^2 + c\sum x_t$

ó

(1) $\sum y_i = \beta_1 \sum x_i + n\beta_0$
(2) $\sum y_i x_i = \beta_1 \sum x_i^2 + \beta_0 \sum x_i$

Mediante algunos de los métodos que usted conoce (*sustitución*, *igua ación* o *e iminación*), encontrará los valores correspondientes a b y c.

$$\sum y_t = 76 \quad \sum x_t = 214 \quad \sum x_t^2 = 5.348 \quad \sum x_t y_t = 1.952$$

(1) 76 = 214b + 10c
(2) 1.952 = 5.348b + 214c

Procedemos a multiplicar la ecuación (1) por -21,4 y luego se la restamos a la ecuación (2), de esta manera:

(2) 1.952 = 5.348b + 214c
(1) -1.626,4 = - 4.579,6b - 214c
325,6 = 768,4b

Despejamos el valor de $b = \dfrac{325,6}{768,4} = 0,4237$

Ahora, conociendo el valor de b o β_1, reemplazamos en ecuación (1), así:

(1) 76 = 214 (0,4237) + 10c

76 = 90,6718 + 10c

Siendo $c = \dfrac{76 - 90{,}6718}{10} = -1{,}4672$

NOTA: con el uso de la calculadora los resultados son b = 0,4237 y c = –1,4679, eso se debe al trabajar con todos los decimales.

Veamos otras fórmulas que nos permitan operar más rápido y sencillo, que trabajando con un sistema de ecuaciones normales tal y como lo acabamos de realizar.

$$b = \frac{Cov}{s_x^2} \Rightarrow \quad b = \frac{32{,}56}{76{,}84} = 0{,}4237$$

$$b = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \Rightarrow \quad b = \frac{10(1.952) - (214)(76)}{10(5.348) - (214)^2} = \frac{3.256}{7.684} = 0{,}4237$$

$$b = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^2 - n\bar{x}^2} \Rightarrow \quad b = \frac{1.952 - 10(21{,}4)(7{,}6)}{5.348 - 10(21{,}4)^2} = \frac{325{,}6}{768{,}4} = 0{,}4237$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \Rightarrow \quad b = \frac{325{,}60}{768{,}40} = 0{,}4237$$

Como se puede observar con la aplicación de las fórmulas anteriores, se obtienen exactamente los mismos resultados para b ó β_1, siendo de 0,4237. Ahora veremos como se abrevia el cálculo de c ó β_0.

$$c = \frac{\sum y_t - b\sum x_t}{n} \Rightarrow \quad c = \frac{76 - 0{,}4237(214)}{10} = -1{,}4672$$

$$c = \bar{y} - b\bar{x} \Rightarrow \quad c = 7{,}6 - 0{,}4237(21{,}4) = -1{,}4672$$

Predicción o estimación puntual

Se podrá decir que una de las finalidades, al establecer la función que represente adecuadamente al conjunto de pares de observaciones que se han tomado para el análisis, es el de poder realizar *interpo aciones*, ya que no es indicado para hacer *e trapo aciones*. Lo anterior significa, que si se va a predecir o estimar el valor de $\hat{Y}$, se requiere que el valor de (variable independiente) deba tomar un valor que se encuentre en el rango de dicha variable y no un valor que sea superior o inferior a los límites que tenga el rango poblacional. Por ejemplo en el ejercicio 1, se señala una producción, que en la muestra va desde 10 hasta 36 (miles de millones de pesos), lo mismo en cuanto a la variable costo va de 3 a 14 (miles de millones de pesos), siendo posible, que en algún momento estos rangos sean ligeramente superiores en la población, esto hace que al elegir el valor que toma la variable independiente para hacer la estimación, éste deberá estar en el rango poblacional, y nunca por fuera de él, pues dejaría de ser una *interpo ación* y pasaría a una *e trapob ación* Se puede decir, además, que no es suficiente lo anterior, sino también deben tenerse en cuenta las condiciones con las cuales se

definió la unidad o el elemento de la investigación, a fin de conocer la población de la cual se extrajo la muestra aleatoria.

Obtenidos los valores para los coeficientes b y c, se reemplazan en la ecuación general de la recta, así: $\hat{Y} = 0{,}4237 \quad - 1{,}4672$

Ejemplo 2. ¿Cuál sería el costo (estimado) en miles de millones de $, para una fábrica que desea tener una producción por un valor de $24 (miles de millones)?

$\hat{Y} = b \ + c \Rightarrow \hat{Y} = 0{,}4237(24) - 1{,}4672 = 8{,}7$ miles de millones de pesos

Se estiman los costos en 8,7 miles de millones de pesos, para una producción de 24 miles de millones de pesos.

Esta misma estimación la obtenemos mediante la ecuación $\hat{Y} = b(\quad - \bar{}) + \bar{}$

Veámoslo: $\hat{Y} = 0{,}4237(24 - 21{,}4) + 7{,}6 = 8{,}7$ miles de millones

ANÁLISIS DE REGRESIÓN

Nuevamente consideramos los 15 pares de observaciones (ejemplo 2 de la página 577) lo utilizaremos para el *aná isis de regresión*, empleando para ello EXCEL, de la siguiente manera:

- Seleccionamos la ficha DATOS, y luego hacemos CLIC en la opción ANÁLISIS DE DATOS.

Figura No. 4. Insertar Gráfico.

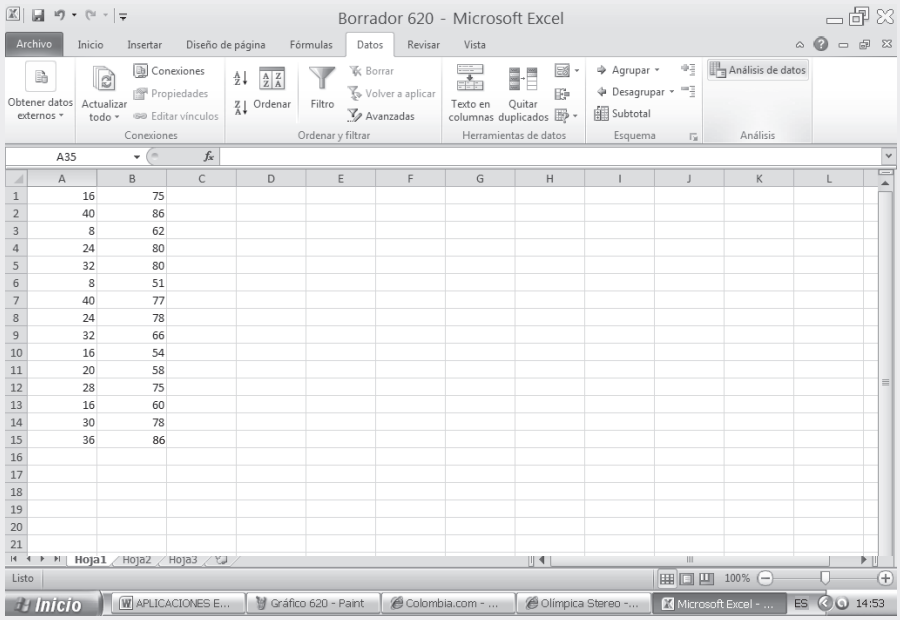
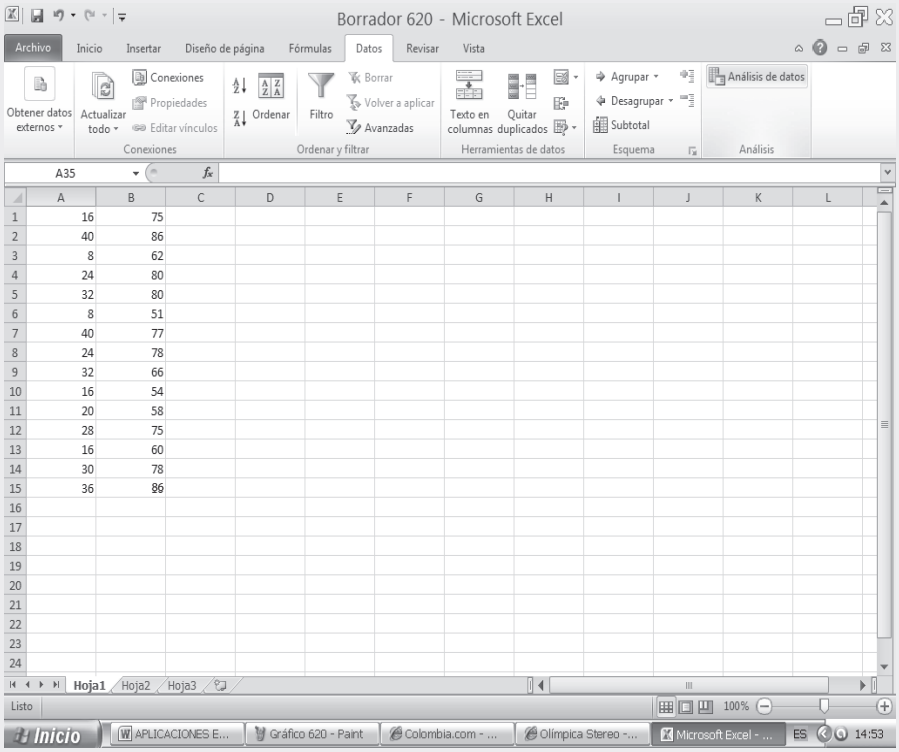


Figura No. 6. Análisis de Datos.



	A	B
1	16	75
2	40	86
3	8	62
4	24	80
5	32	80
6	8	51
7	40	77
8	24	78
9	32	66
10	16	54
11	20	58
12	28	75
13	16	60
14	30	78
15	36	86

- En el cuadro de diálogo que aparece, seleccionamos la opción REGRESIÓN y tecleamos lo siguiente:
- RANGO Y DE ENTRADA, seleccionamos B1:B15.
- RANGO X DE ENTRADA, seleccionamos A1:A15
- Finalmente se debe fijar el RANGO DE SALIDA, en este caso será la celda D2, luego hacemos CLIC en aceptar con lo cual obtenemos la información del ANÁLISIS DE REGRESIÓN LINEAL SIMPLE.
- El anterior procedimiento permite calcular los COEFICIENTES DE REGRESIÓN, $ß_o$ y $ß_1$, además el COEFICIENTE DE CORRELACIÓN y DE DETERMINACIÓN DE PEARSON, así como las VARIANZAS DE LAS VARIABLES, X y Y.

Coeficiente correlación

$$r = 0{,}77 \quad R^2 = 0{,}5975 \quad n = 15$$

$$c = \beta_o = 50{,}157; \quad \beta_1 = b = 0{,}8476; \quad \hat{Y} = 0{,}8476x + 50{,}157; \quad s^2_{yx} = 49{,}977; \quad s_{xy} = 7{,}069$$

En la SEGUNDA PARTE DEL SIL, encontrará en conjunto toda la aplicación del EXCEL a la ESTADÍSTICA, que igual a1 capítulo 13, ofrece una BASE DE DATOS para trabajar.

Figura No. 7. Regresión.

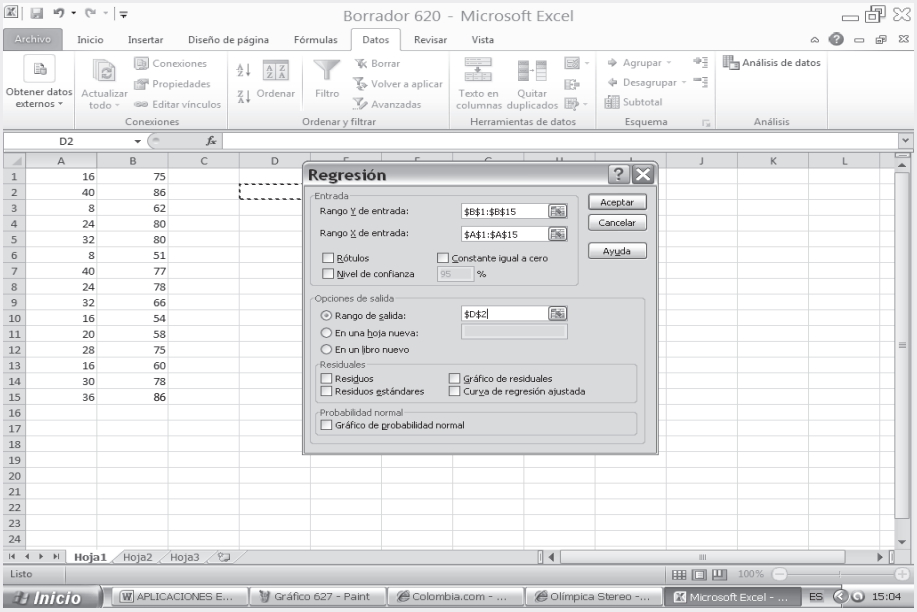
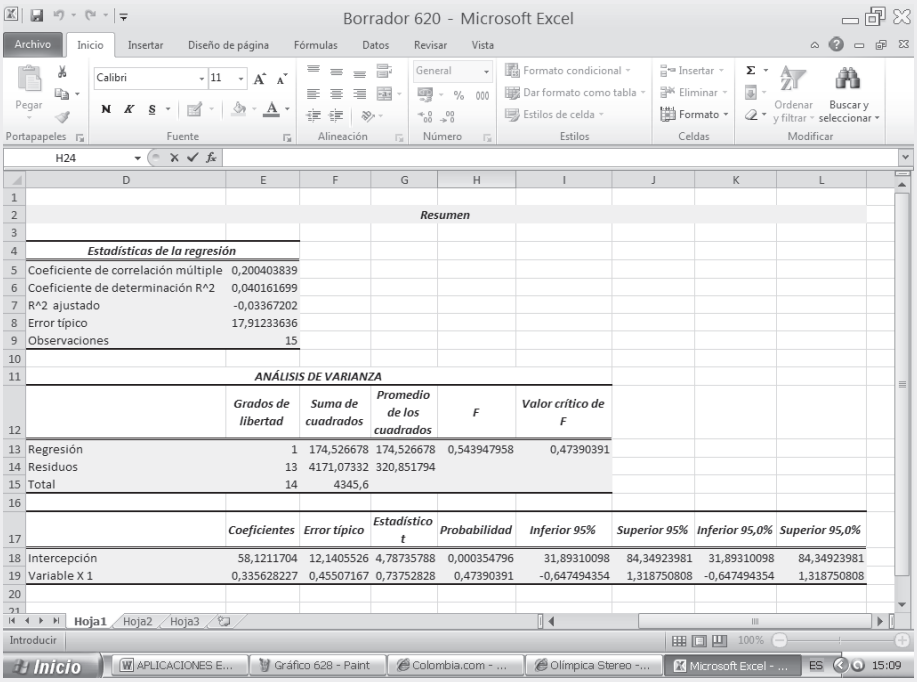


Figura No. 8. Cuadro de Resultados.



Varian a residual y error estándar de estimación $\Rightarrow$ VR = S_y^2

La *arian a residua arian a no e p icada* o la *suma de os cuadrados de error* son los nombres que se emplean para identificar a esta medida de dispersión, que permite calcular el grado de variación o de dispersión que presenta el conjunto u observaciones que no quedaron explicados por la recta de regresión. Es decir, la dispersión de aquellos valores que no quedan dentro de la recta establecida para el conjunto de observaciones.

Se define como *a media aritmética de cuadrado de as diferencias entre os a ores obser ados os a ores estimados*. Las fórmulas que se emplean para su cálculo son varias, entre las cuales tenemos:

(1) $S^2_{y\cdot x} = \dfrac{\sum (y_t - \hat{Y}_t)^2}{n}$ (2) $S^2_{y\cdot x} = S^2_{y\cdot} - \dfrac{m^2_{xy\cdot}}{S^2_x}$ (3) $S^2_{y\cdot x} = S^2_{y\cdot} - 2bm_{xy\cdot} + b^2 S^2_x$

(4) $S^2_{y\cdot x} = \dfrac{\sum y_t^2 - c\sum y_t - b\sum y_t x_t}{n}$ (5) $S^2_{y\cdot x} = S^2_{y\cdot}(1 - R^2)$

Veamos ahora la aplicación de cada una de estas fórmulas:

x_i	y_i	$\hat{Y}_i$	$y_i - \hat{Y}_i$	$(y_i - \hat{Y}_i)$
10	3	2,77	0,23	0,0529
18	5	6,16	-1,16	1,3456
12	4	3,62	0,38	0,1444
16	5	5,31	-0,31	0,0961
22	8	7,85	0,15	0,0225
36	12	13,79	-1,79	3,2041
30	10	11,24	-1,24	1,5376
32	14	12,09	1,91	3,6481
26	12	9,55	2,45	6,0025
12	3	3,62	-0,62	0,3844
214	76	76,00	0,00	16,4382

1) $S^2_{y\cdot x} = \dfrac{16,4382}{10} = 1,64$

2) $S^2_{y\cdot x} = 15,44 - \dfrac{32,56^2}{76,84} = 1,64$

3) $S^2_{y\cdot x} = 15,44 - 2(0,4237)(35,56) + 0,4237^2 + (76,84) = 1,64$

4) $S^2_{y\cdot x} = \dfrac{732 + 1,4672(76) - 0,4237(1.952)}{10} = \dfrac{16,4448}{10} = 1,64$

5) $S^2_{y\cdot x} = 15,44(1 - 0,8938) = 1,64$

Observemos los resultados anteriores de la VR (*arian a residua o no e p icada*), son iguales, salvo en la ultima que nos dio una pequeña diferencia que no debe preocuparnos, pues generalmente trabajamos con dos decimales, lo cual altera un poco el resultado.

Error estándar de estimación

Es una expresión dada a la desviación estándar de los valores observados respecto a una línea de regresión. Se podría decir que es una estimación de la variación que probablemente acompañará a las predicciones realizadas por medio de la ecuación de regresión.

Se define como la raíz cuadrada de la varianza residual. $s_{y.x} = \sqrt{s^2_{y.x}}$

En el ejercicio que hemos venido desarrollando será igual a $s = \sqrt{1{,}64} = 1{,}28$

Intervalo de predicción o de estimación

El *inter a o de predicción* o *límites de confianza* para un valor estimado ($\hat{Y}$), consiste en establecer dos puntos para el estimador, dentro del cual debe estar contenido el parámetro o valor estadístico de la población, con cierto grado de confianza o de seguridad, que será establecido por el analista o estadístico. Se consideran dos formas de calcular estos intervalos, según se vaya a estimar (a) un *a or de a ariab e* y (b) *un a or promedio*. La diferencia en cada caso dependerá de cómo se considere a (*X*), ya sea un valor de la variable o como un valor promedio para esa misma variable. Veamos cada cálculo del estimador en forma separada, con los mismos datos del problema anterior.

En el *primer caso*, cuando a la variable independiente (*X*) se le da un *a or* que debe estar dentro del recorrido de la variable poblacional, y se desea estimar la variable dependiente ($\hat{Y}$), la fórmula que se debe aplicar en este caso, será:

$$\hat{Y} \pm t_{n-2}\ S_{y.x} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x_t^2 - \frac{(\sum x_t)^2}{n}}}$$

Suponiendo que se quieren establecer los límites de confianza del 95%, en el ejercicio en que se desea estimar cuál debe ser el costo en función de la producción, cuando = 24 (valores dados en miles de millones de pesos), para las 10 empresas seleccionadas en la muestra.

$$\hat{Y} = 8{,}7 \pm 2{,}306(1{,}28) \sqrt{1 + \frac{1}{10} + \frac{(24 - 21{,}4)^2}{5.348 - \frac{(214)^2}{10}}} \rightarrow \hat{Y} = 8{,}7 \pm 3{,}11 \ = \begin{cases} \hat{Y}_s = 11{,}81 \\ \hat{Y}_t = 5{,}59 \end{cases} \text{(miles de millones de pesos)}$$

Dentro de esos valores (11,81 y 5,59), deberá estar el valor verdadero, con una confianza de que así sea, del 95%; nos queda un margen del 5% que esté por encima o por debajo de esos valores. Observe que el error que podemos cometer en la estimación es relativamente grande siendo de 3,11, resultado lógico cuando la muestra seleccionada, como en este caso, es muy pequeña, obteniéndose un intervalo de confianza muy grande que nos hace desconfiar un poco del procedimiento utilizado. El error se disminuye a medida que aumentamos el tamaño de la muestra.

El *segundo caso*, cuando el valor de se establece como *promedio*. En el ejercicio que estamos desarrollando (estimar costo en función de la producción) se aplicará la siguiente fórmula:

$$\hat{Y} \pm t_{n-2}\, S_{y.x} \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}}$$

Reemplazando con los datos obtenidos en el ejercicio respectivo se tendrá que:

$$\hat{Y} = 8{,}7 \pm 2{,}306\,(1{,}28) \sqrt{\frac{1}{10} + \frac{(24-21{,}4)^2}{5.348 - \frac{(214)^2}{10}}} \rightarrow \hat{Y} = 8{,}7 \pm 0{,}97 = \begin{cases} 9{,}67 \\ 7{,}73 \end{cases} \text{(miles de millones de pesos)}$$

Dentro de esos límites (7,73 y 9,67) estará el valor real (costos), con una seguridad del 95%

Varian a explicada = VE

Hasta el momento se han calculado las varianzas para la variable y (S^2 y S^2), denominadas: *arian a tota de* y *arian a tota de* . También se han calculado los valores para la *arian a no e p icada* o *arian a residua* (S^2). Ahora nos ocuparemos del cálculo de otra varianza, de gran utilidad en el análisis de regresión, conocida como *arian a e p icada*, simbolizada por S_a^2 .

a arian a e p icada es una medida que nos permite calcular el grado de dispersión de aquellos valores que quedaron explicados por la recta de regresión, es decir, la dispersión de aquellos puntos en la gráfica que se confunden con los puntos de la recta o valores estimados.

Se define como la *media aritmética, de os cuadrados de as diferencias entre os a ores estimados a media aritmética de esa ariab e*. Su fórmula es:

$$S^2_{a\hat{y}} = \frac{\sum (\hat{Y}_t - \bar{y})^2}{n}$$

Es necesario saber que la varianza total debe ser igual a la suma de la Varianza residual más la Varianza explicada: VT = VR+ VE . De acuerdo con dicha igualdad se tendrá que VE = VT– VR . Procedemos a calcular esta medida de dispersión en el ejercicio que hemos venido desarrollando:

x_i	y_i	$\hat{Y}_i$	$\hat{Y}_i - \bar{y}$	$(\hat{Y}_i - \bar{y})^2$
10	3	2,77	– 4,83	23,3289
18	5	6,16	– 1,44	2,0736
12	4	3,62	– 3,98	15,8404
16	5	5,31	– 2,29	5,2441
22	8	7,85	0,25	0,0625
36	12	13,79	6,19	38,3161
30	10	11,24	3,64	13,2496
32	14	12,09	4,49	20,1601
26	12	9,55	1,95	3,8025
12	3	3,62	– 3,98	15,8404
214	76	76,00	0	137,9182

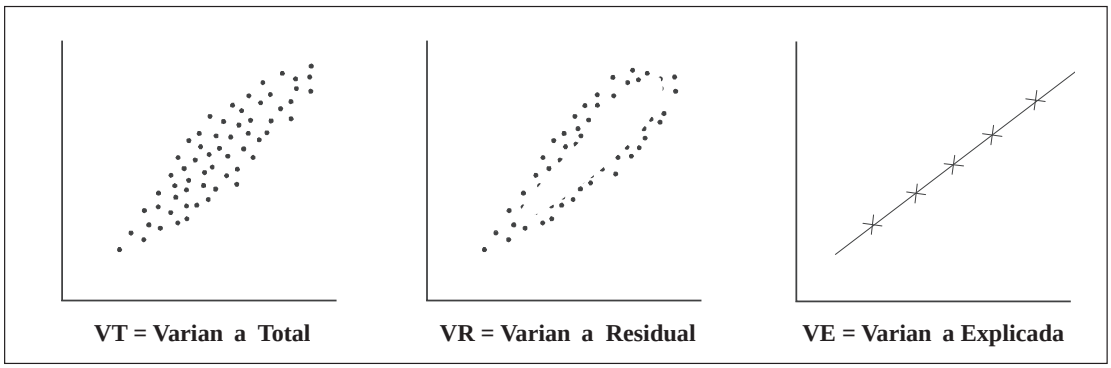
Reemplacemos los resultados anteriores en las fórmulas que se dan para calcular la varianza explicada:

$$(1)\ S^2_{\hat{y}} = \frac{\sum(\hat{Y}_t - \bar{y})^2}{n} = \frac{137{,}9182}{10} = 13{,}79$$

$$(2)\ VE = VT - VR$$
$$S^2_{\hat{y}} = S^2_y - S^2_{y.x}$$
$$S^2_{\hat{y}} = 15{,}44 - 1{,}64 = 13{,}80$$

Observemos que los resultados obtenidos, mediante las dos fórmulas señaladas, son iguales.

La interpretación que se les da a las varianzas *tota* , *residua* y *e p icada* es: la primera determina la dispersión de todos los puntos o pares de observaciones representados en la gráfica o *diagrama de dispersión*; en la segunda, la *residua* se reduce a determinar la dispersión en los puntos o pares de observaciones que no quedan sobre la línea de regresión, elaborada con los valores estimados; y en la tercera, la varianza *e p icada* sólo toma en cuenta, los puntos u observaciones que quedan ubicados sobre la recta de regresión.



VT = Varian a Total **VR = Varian a Residual** **VE = Varian a Explicada**

Coeficiente de correlación o de Pearson $\Rightarrow$ R = r

Comentábamos al iniciar el tema sobre análisis de regresión y correlación, lo que se entendía por *regresión* y por *corre ación*. En este último decíamos que nos determinaba el grado de relación que puede haber entre las dos variables. Este grado de correlación lo obtenemos mediante el cálculo del *Coeficiente de corre ación*

El coeficiente de correlación es una medida de interdependencia de dos variables aleatorias.

En *primer ugar* nos referiremos al cálculo del *coeficiente de correlación al cuadrado* denominado también como *coeficiente de determinación*, simbolizado por R^2 Las fórmulas para el cálculo de este coeficiente son variadas, obteniéndose un resultado igual por cualquiera de los métodos que se utilice. Veámoslo:

(1) $R^2 = \dfrac{\text{Covarianza}^2}{S^2_x(S^2_y)}$ $\qquad R^2 = \dfrac{m^2_{xy}}{S^2_x\ S^2_y} = \dfrac{32{,}56^2}{(76{,}84)(15{,}44)} = 0{,}8936 \cong 0{,}89$

(2) $R^2 = \dfrac{VE}{VT}$ $\qquad R^2 = \dfrac{S^2_{\hat{y}}}{S^2_y} = \dfrac{13{,}79}{15{,}44} = 0{,}8931 \cong 0{,}89$

(3) $R^2 = 1 - \dfrac{VR}{VT}$ $\qquad R^2 = 1 - \dfrac{1{,}64}{15{,}44} = 0{,}8938 \cong 0{,}89$

El *coeficiente de correlación al cuadrado* debe ser un valor tal, que cumpla con la siguiente condición: $0 \leq R^2 \leq 1$. Cuando el coeficiente de correlación al cuadrado es igual a 1, decimos que hay una *corre ación perfecta*, ya que los valores observados son exactamente iguales a los estimados, en otras palabras, los puntos en una gráfica (nube de puntos) se confunden con los de la recta; se dice en este caso, que la *arian a e p icada* es igual a la *arian a tota* . A medida que el coeficiente de correlación al cuadrado disminuye, se aleja de uno, se dice también, que la recta representa cada vez menos a ese conjunto de observaciones. Un $R^2 \leq 0{,}30$ nos estará indicando que no hay correlación entre las variables.

De acuerdo con la formula presentada en al página anterior (2) podemos decir, que el *coeficiente de corre ación a cuadrado* o *coeficiente de determinación* mide la proporción de la varianza que queda explicada por la ecuación de regresión, que describe la relación establecida entre las dos variables. Se podría decir, en otras palabras, que indica el porcentaje de las variaciones de la variable dependiente, atribuible a la influencia de la variable independiente.

En *segundo ugar* nos podemos referir simplemente al *coeficiente de correlación*, en este caso de una recta, el cual se simboliza por *r* ó *R*. Se define como la *raíz cuadrada del coeficiente de determinación*. El signo del coeficiente dependerá del signo que tenga la covarianza o los coeficientes angulares (estos últimos deben tener el mismo signo y el producto debe ser menor o igual a uno).

Su cálculo se puede realizar mediante la aplicación de diferentes fórmulas, obteniéndose exactamente el mismo resultado. Siendo:

(1) $$r = \frac{Covarianza}{S_x \; S_y} \qquad r = \frac{32{,}56}{(8{,}76)\;(3{,}93)} \cong 0{,}9458 = 0{,}95$$

(2) $$r = \sqrt{R^2} \qquad r = \sqrt{0{,}8938} \cong 0{,}9454 = 0{,}95$$

(3) $$r = \frac{n\Sigma y_i x_i - (\Sigma x_i)\;(\Sigma y_i)}{\sqrt{\left[n\Sigma x_i^2 - (\Sigma x_i)^2\right]\left[n\Sigma y_i^2 - (\Sigma y_i)^2\right]}}$$

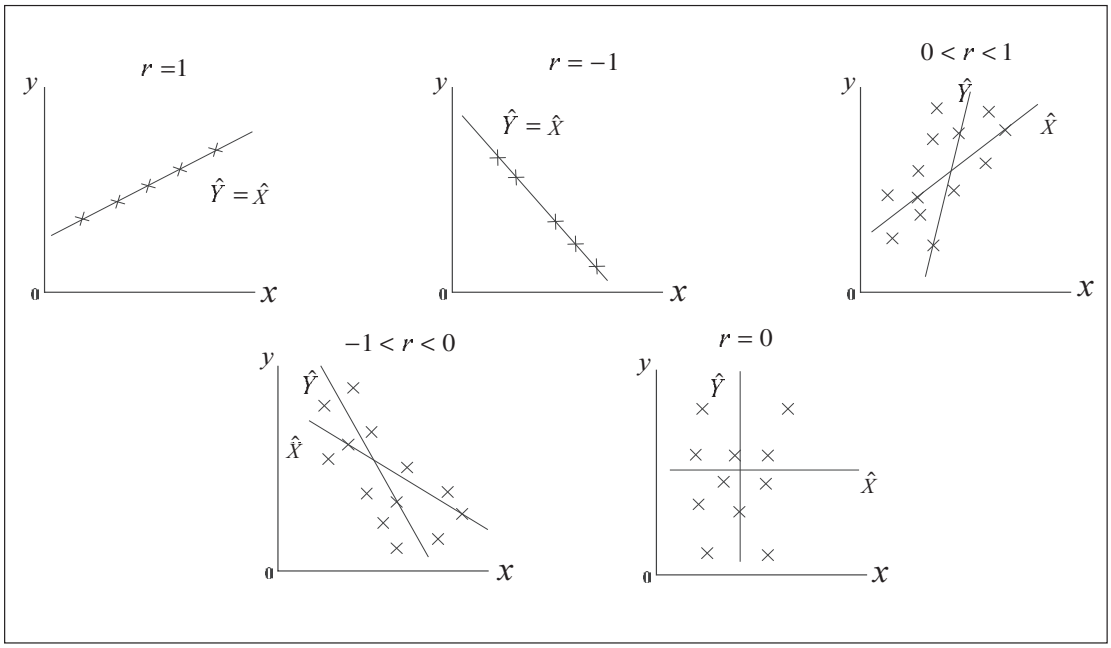
$$r = \frac{10(1.952) - (214)(76)}{\sqrt{\left[10(5.348) - (214)^2\right]\left[10(732) - (76)^2\right]}} = 0{,}9452 \cong 0{,}95$$

(4) $$r = \frac{\Sigma y_i x_i - n\bar{x}\,\bar{y}}{\sqrt{\left[\Sigma x_i^2 - n\bar{x}^2\right]\left[\Sigma y_i^2 - n\bar{y}^2\right]}} \Rightarrow r = \frac{1.952 - 10(21{,}4)(7{,}6)}{\sqrt{\left[5.348 - 10(21{,}4)^2\right]\left[732 - 10(7{,}6)^2\right]}} \cong 0{,}95$$

El coeficiente de correlación debe ser un valor comprendido entre –1 y 1, por lo tanto se dirá que: $-1 \leq r \leq 1$. *a re ación será perfecta* cuando *r* es igual a 1 o –1; no hay ninguna relación cuando *r* = 0, ó menor a 0,30.

Sabiendo que el coeficiente de correlación será un valor $-1 \leq r \leq 1$, se representará, así:

Figura 12.5 Comportamiento del coeficiente de correlación



El *grado de corre ación* lo podemos clasificar tentativamente, evitando un tanto la rigidez de sus límites.

- Correlación *perfecta*, cuando $r = 1$ r = − 1
- Correlación *e ce ente*, cuando r es mayor de 0,90 y menor de 1 $(-1 < r < -0{,}90)$
- Correlación *aceptab e*, cuando r se encuentra entre 0,80 y 0,90 $(-0{,}9 < r < -0{,}8)$
- Correlación *regu ar*, cuando r se encuentra entre 0,60 y 0,80 $(-0{,}8 < r < -0{,}6)$
- Correlación *mínima*, cuando r se encuentra entre 0,30 y 0,60 $(-0{,}6 < r < -0{,}3)$
- *No ha* correlación para r menor de 0,30 y mayor a 0 $(-0{,}3 < r < 0)$

No debemos confiarnos del resultado obtenido, en especial cuando es bueno o muy bueno, mientras no estemos seguros de que las variables tienen una relación de causa y efecto, pues podemos obtener una buena correlación entre dos variables, lo cual puede deberse a la casualidad; por ejemplo un coeficiente de 0,86 entre las variables: asistencia diaria a clases (de lunes a viernes) y el precio del café en esos mismos días.

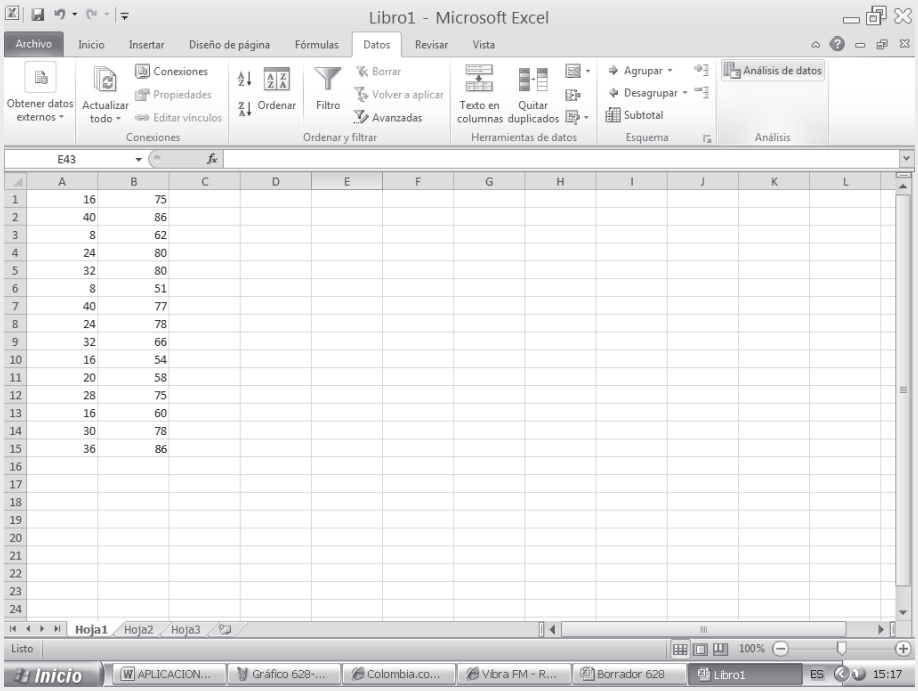
Como complemento al análisis de regresión, se puede aplicar la *prueba de hipótesis al coeficiente de regresión* y al *coeficiente de correlación*.

CÁLCULO DEL COEFICIENTE DE CORRELACIÓN CON APLICACIÓN DE EXCEL

Con los mismos datos de la tabla (ver página 577), que aparece en la hoja de cálculo, procedemos a calcular el COEFICIENTE DE CORRELACIÓN (r) de la siguiente manera:

- Con los mismos datos de la tabla (ver página 616), que aparece en la hoja de cálculo, procedemos a calcular el COEFICIENTE DE CORRELACIÓN (r) de la siguiente manera:
- En la ficha DATOS seleccionamos la opción ANÁLISIS DE DATOS.

Figura No. 9. Análisis de Datos.



	A	B
1	16	75
2	40	86
3	8	62
4	24	80
5	32	80
6	8	51
7	40	77
8	24	78
9	32	66
10	16	54
11	20	58
12	28	75
13	16	60
14	30	78
15	36	86

- Seleccionamos de la lista que nos ofrece el cuadro de diálogo ANÁLISIS DE DATOS, la opción COEFICIENTE DE CORRELACIÓN y hacemos CLIC en ACEPTAR.
- Tecleamos en el campo RANGO DE ENTRADA siendo A1:B15; luego en el RANGO DE SALIDA señalamos una CELDA de la hoja de cálculo, en nuestro caso, será la celda D3.
- Finalmente hacemos CLIC en ACEPTAR tal como aparece en las aplicaciones que se presentan a continuación:

Figura No. 10. Cálculo del Coeficiente de Correlación.

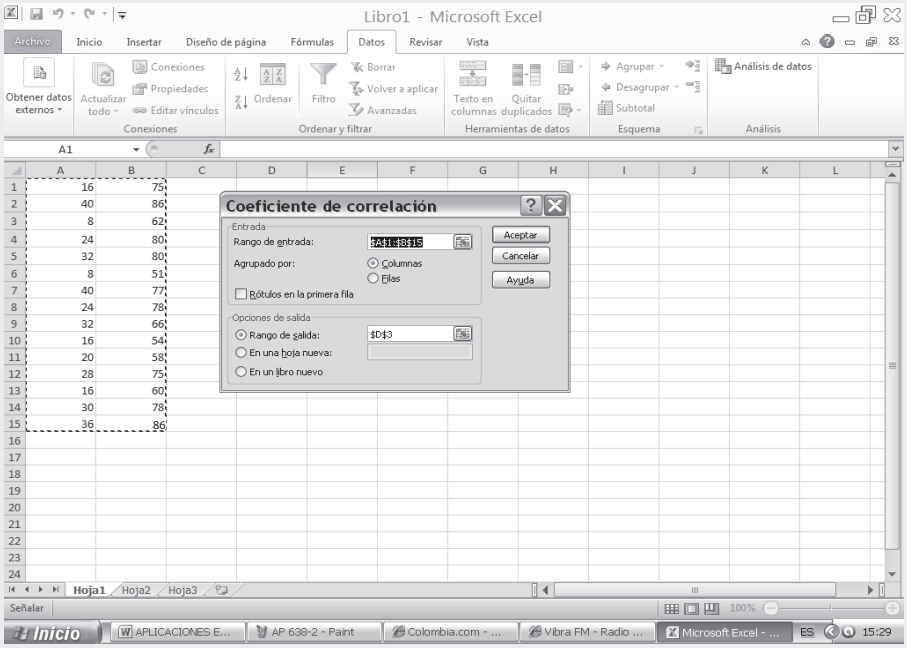
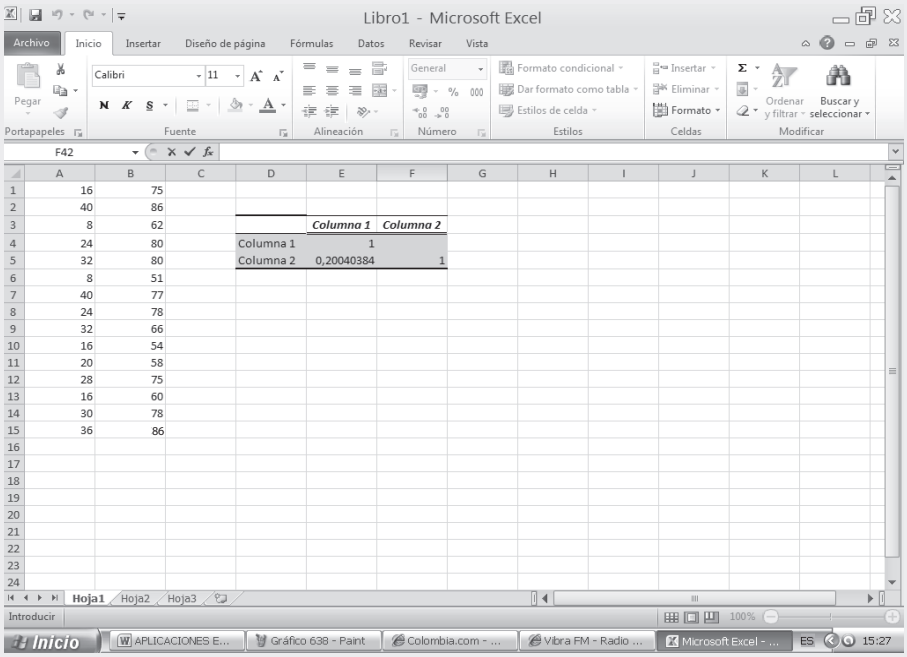
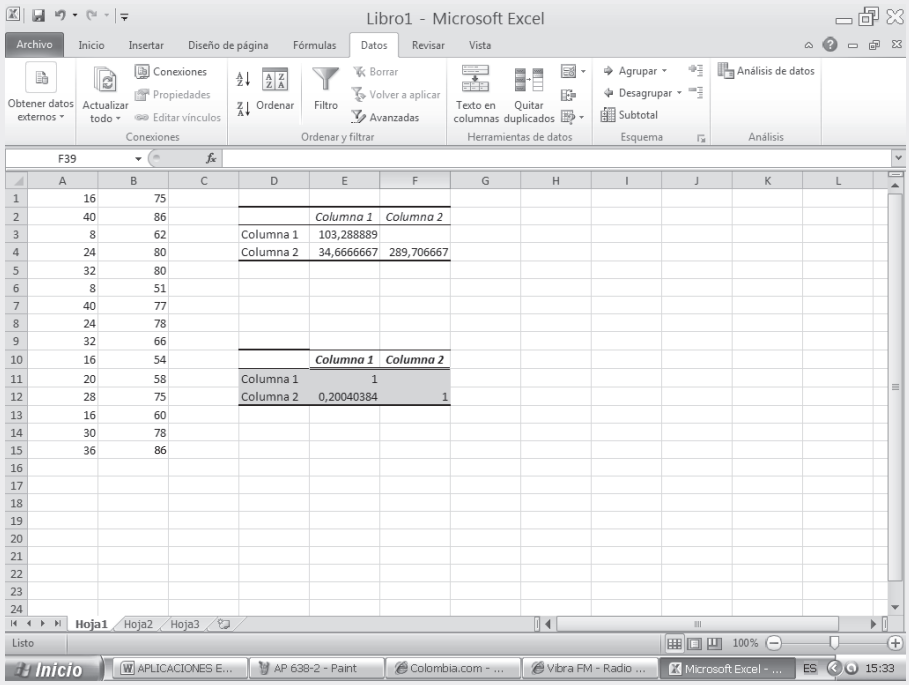


Figura No. 11. Cuadro de Resultados.



- Tomando los resultados del ejercicio anterior (ver página 577) junto con los resultados del ejercicio que estamos desarrollando, es posible observar lo siguiente:

Figura No. 12. Comparación Cuadro de Resultados de los Dos Ejercicios Anteriores.



- El coeficiente de correlación o coeficiente de correlación de Pearson es:

$r = 0{,}20 \qquad o \quad = 34{,}67 \qquad S^2 = 103{,}29 \qquad S^2 = 289{,}71$

PRUEBA DEL COEFICIENTE ANGULAR O COEFICIENTE DE REGRESIÓN β ó b

La pendiente estimada b se halla sujeta a la variabilidad muestral, por lo tanto, en algunos casos puede ser un mal estimador y por consiguiente, la ecuación de regresión no tiene validez como elemento de predicción. También se debe tener en cuenta el tamaño de la muestra para utilizar la prueba de "Z" o de "t" en este último caso se debe considerar n - 2 grados de libertad, por ser dos variables, las que se constituyen en medidas muy generales de significación y confianza.

Siendo b, la *pendiente muestra* y β, la *pendiente pob aciona* , se puede considerar como hipótesis nula (H) cuando el coeficiente angular es 0, frente a la alternativa que sea mayor, menor, o diferente de cero, pudiendo decir, que la recta de regresión es útil para estimar la variable dependiente ($\hat{Y}$). Siendo la variante estadística de "t" igual a:

$$t_{n-2} = \frac{b_1 - \beta_1}{S_{b_1}} \qquad \text{donde:} \qquad S_{b1} = S_{yx}\sqrt{\frac{1}{\Sigma(x_t - \bar{x})^2}}$$

S_{b_1} , es el *error estándar del coeficiente de regresión b*, es decir, una medida de la cantidad de error muestral que puede haber en el coeficiente b de la ecuación de regresión de la recta, en este caso, se tendrá que:

$$S_{b_1} = 1{,}28\sqrt{\frac{1}{768{,}4}} = 0{,}046$$

Ahora teniendo que:

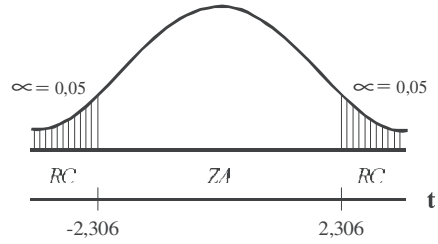
a) $H_0 : \beta_1 = 0$ (No hay relación)

$H_a : \beta_1 \neq 0$ (Si hay relación)

b) $\propto = 0{,}05$

c) $b = 0{,}4237$

$$t = \frac{0{,}4237 - 0}{0{,}046} = 9{,}21; \quad \left.\begin{matrix} v = 10 - 2 = 8 \\ \propto = 0{,}05 \end{matrix}\right\} t = 2{,}306$$



d) Cuando t es menor que t_{n-2} se acepta la hipótesis nula, en caso contrario, cuando es mayor, se acepta la hipótesis alternativa. En este ejemplo aceptamos H_a, por lo cual podemos llegar a la conclusión de que la recta de regresión constituye una ayuda para estimar la variable dependiente. Observe que en este ejercicio se está rechazando la hipótesis nula H_0, y aceptando la hipótesis alternativa H_a.

En conclusión se podrá decir, que si se rechaza la hipótesis nula se debe a que hay relación lineal. En este caso concreto, hay relación entre las variables producción y costo, llegaríamos a la conclusión de que existe una relación significativa.

Un segundo procedimiento para realizar la prueba del coeficiente angular, a fin de probar la existencia de una relación lineal entre estas dos variables sometidas al análisis, se hace estableciendo un *intervalo de confianza* para β_1, determinando si el valor hipotético dado $H_0 : \beta_1 = 0$ se encuentra incluido dentro del intervalo. Si esto sucede llegaríamos a la conclusión de que no hay relación, si por el contrario el valor de β_1 cae por fuera de estos límites, significa que hay cierta relación entre las variables.

$$b_1 \pm t_{n-2}\ S_{b_1} = 0{,}4237 \pm 2{,}306(0{,}046) = 0{,}4237 \pm 0{,}1061 = \begin{cases} 0{,}5298 \cong 0{,}53 \\ 0{,}3176 \cong 0{,}32 \end{cases}$$

Ambos límites son positivos, por lo tanto no admite o no incluye el cero, es decir, la hipótesis nula no cae dentro del intervalo y se estará aceptando la alternativa, es decir, que sí puede haber una relación de causa y efecto entre estas dos variables.

PRUEBA DEL COEFICIENTE DE CORRELACIÓN r

Además de las dos pruebas anteriores, existe otro procedimiento más para establecer si hay relación lineal, siendo ésta la *prueba al coeficiente de correlación* obtenida en la muestra. Para ello se considera que $\rho = 0$ (letra griega que se lee rho), es decir, que el coeficiente de correlación en la población es igual a cero, frente a la alternativa de que sea diferente; para ello se tendrá que:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0{,}95}{\sqrt{\frac{1-0{,}89}{10-2}}} = 8{,}10$$

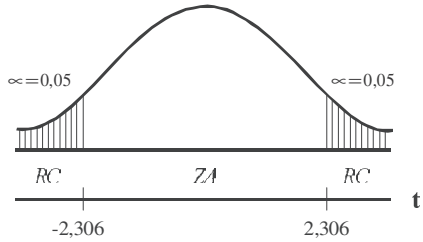
Nota: Se puede ampliar estos conocimientos consultando el Capítulo 9

La prueba se realiza, al igual que las anteriores, cuyo procedimiento a seguir será:

a) $H_0 : \rho = 0$ (No hay correlación)
$H_a : \rho \neq 0$ (Si hay correlación)

b) $\propto = 0{,}05$

c) $t = 2{,}306$



d) Se observa que 8,10 se sitúa en la zona de rechazo, es decir, que hay correlación entre las variables de la población, lo que equivale a concluir, que al nivel de significación del 5% la correlación en la población es diferente a cero.

Suposiciones en el análisis de la regresión

Son las suposiciones que se deben hacer en el análisis de regresión, entre las principales se tienen: (a) *Norma idad*, (b) *Homoscedasticidad* y (c) *Independencia de error*.

a) *Norma idad*: se supone que los valores de Y tengan distribución normal en cada valor de X.

b) *Homoscedasticidad*: es otro supuesto, en el cual la variación en torno a la recta de regresión es constante para todos los valores de X, no importando que el valor que toma sea alto o bajo; de todas formas la variación se supone es la misma.

c) *Independencia de error*: establece que el error estándar de estimación es independiente para cada valor de X.

¿QUÉ PASA SI EN VEZ DE ESTIMAR Ŷ , SE DESEA ESTIMAR Xˆ ?

En la mayoría de los casos, se aplica uno de los dos procedimientos, que presentamos a continuación:

Primer procedimiento: consiste en cambiar el orden de las variables, veamos que en el ejemplo 1, producción era variable $_i$ y costo la variable $_i$, ahora, costo será la variable $_i$ y producción la variable $_i$, una vez realizado el cambio se procede igual tal como se realizó con los datos del ejemplo 1.

COSTO (miles mill)	PRODUCCIÓN (miles mill)
$_i$	$_i$
3	10
5	18
4	12
5	16
8	22
12	36
10	30
14	32
12	26
3	12

Para abreviar no se hará la tabla de cálculo, sólo colocaremos los resultados de las columnas, cuyos datos serán utilizados para ser reemplazados en las diferentes fórmulas ya vistas:

$n = 10$ $\quad \Sigma x_i = 76$

$\Sigma x_t^2 = 732$ $\quad \Sigma y_i = 214$

$\Sigma y_t^2 = 5.348$ $\quad \Sigma x_i y_i = 1.952$

$\bar{x} = 7{,}6$ $\quad \bar{y} = 21{,}4$

- Cálculo de b y c $\boxed{\overline{\beta_1}\ \text{y}\ \overline{\beta_0}}$

$$\begin{array}{l}(1)\quad 214 = 76b + 10c \\ (2)\ 1.952 = 732b + 76c\end{array} \Rightarrow \begin{array}{rl} (2) & 1.952 = 732b + 76c \\ (1) & -1.626{,}4 = -577{,}6b - 76c \\ \hline & 325{,}6 = 154{,}4b \quad 0\end{array}$$

$$b = \frac{325{,}6}{154{,}4} = 2{,}1088$$

Reemplazamos en (1): $214 = 76(2{,}1088)+10c \Rightarrow \dfrac{214 - 160{,}27}{10} = c = 5{,}373$

- $\boxed{b = \dfrac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2}} \Rightarrow b = \dfrac{10(1.952) - (76)(214)}{10(732) - (76)^2} = 2{,}1088$
- $\boxed{b = \dfrac{n\Sigma x_i y_i - n\,\bar{x}\,\bar{y}}{\Sigma x_i^2 - n\,\bar{x}^2}} \Rightarrow b = \dfrac{1.952 - 10(7{,}6)(21{,}4)}{732 - 10(7{,}6)^2} = 2{,}1088$
- $\boxed{b = \dfrac{Cov}{S_x^2}} \Rightarrow b = \dfrac{32{,}56}{15{,}44} = 2{,}1088$
- $\boxed{c = \dfrac{\Sigma y_i - b\Sigma x_i}{n}} \Rightarrow c = \dfrac{214 - 2{,}1088(76)}{10} = 5{,}373$
- $\boxed{c = \bar{y} - b\bar{x}} \Rightarrow c = 21{,}4 - 2{,}1088(76) = 5{,}373$

Ejemplo 2. Con la información anterior (ejemplo 1) estime el valor de la producción (miles de millones de pesos), cuando supuestamente los costos de producción son de 13 miles de millones de pesos, fijando límites de confianza. Calcule además el coeficiente de correlación de Pearson.

Solución:

Primero realizamos la estimación de *Y*

$\hat{Y} = 2{,}1088 \ + 5{,}373 \Rightarrow \ \hat{Y} = 2{,}1088(13) + 5{,}373 \cong 32{,}79$ (miles mill $) ó

$\hat{Y} = 2{,}1088\ (13 - 7{,}6) + \ 21{,}4 \cong 32{,}79 \Rightarrow \ \boxed{\hat{Y} = b(x - \bar{x}) + \bar{y}}$

Ahora procedemos a calcular la varianza residual y la desviación residual o no explicada:

$\boxed{S_{y.x}^2 = VR = S_y^2(1 - r^2)} \Rightarrow \quad S_{y.x}^2 = 76{,}84(1 - 0{,}8936) \cong 8{,}18$

$\boxed{S_{y.x}^2 = S_y^2 - \dfrac{Cov^2}{S_x^2}} \Rightarrow \quad S_{y.x}^2 = 76{,}84 - \dfrac{32{,}56^2}{15{,}44} \cong 8{,}18$

$\boxed{S_{y.x}^2 = \dfrac{\Sigma y_t^2 - c\Sigma y_t - b\Sigma y_t x_t}{n}} \Rightarrow \quad S_{y.x}^2 = \dfrac{5.348 - 5{,}373(214) - 2{,}1088(1.952)}{10} = 8{,}18$

Usted podrá calcular la $VR = \dfrac{\Sigma(y_t - \hat{Y}_t)^2}{n} = 8{,}18$ (aproximadamente)

La desviación residual será: $S_{y.x} = \sqrt{S^2_{y.x}} = \sqrt{8{,}18} = 2{,}86$

Los límites de confianza se obtendrán así:

$$\hat{Y} = 32{,}79 \pm 3{,}306\ (2{,}86) \sqrt{1 + \frac{1}{10} + \frac{(13-7{,}6)^2}{732 - \frac{(76)^2}{10}}} = 32{,}79 \pm 10{,}73 = \begin{cases} 43{,}52 \\ 22{,}06 \end{cases}$$

(Valor)

$$\text{ó} \quad \hat{Y} = 32{,}79 \pm 2{,}306\ (2{,}86) \sqrt{\frac{1}{10} + \frac{(13-7{,}6)^2}{732 - \frac{(76)^2}{10}}} = 32{,}79 \pm 3{,}54 = \begin{cases} 36{,}33 \\ 29{,}25 \end{cases}$$

(Promedio)

Los límites, cuando es un *a or* o un *promedio* son 22,06 y 43,52; 29,25 y 36,33 miles millones respectivamente, dentro de los cuales estará el parámetro (valor poblacional) con una seguridad o confianza del 95%.

Ahora calcularemos el coeficiente de correlación lineal o de Pearson, mediante algunas de las fórmulas ya vistas:

$$r = \frac{n\Sigma x_t y_t - (\Sigma x_t)(\Sigma y_t)}{\sqrt{\left[n\Sigma x_t^{\,2} - (\Sigma x_t)^2\right]\left[n\Sigma y_t^{\,2} - (\Sigma y_t)^2\right]}}$$

$$r = \frac{10(1.952) - (76)(214)}{\sqrt{\left[10(732) - (76)^2\right]\left[10(5.348) - (214)^2\right]}} \cong 0{,}95$$

$$r^2 = 1 - \frac{VR}{VT} = 1 - \frac{8{,}18}{76{,}84} = 0{,}8936 \ \Rightarrow\ r = \sqrt{0{,}8936} \cong 0{,}95$$

$$r^2 = \frac{VE}{VT} = \frac{68{,}66}{76{,}84} = 0{,}8936 \ \Rightarrow\ r = 0{,}95$$

$$VE = S^2_{y.} - S^2_{y.x} = 76{,}84 - 8{,}18 = 68{,}66 = S^2_{\hat{y}.}$$

r = 0,95 es un buen coeficiente de correlación, indicándonos que existe una buena relación (causa y efecto) entre las dos variables.

La prueba de hipótesis sobre el coeficiente de regresión (b) y el coeficiente de correlación (r), tendrá un procedimiento de cálculo exactamente igual como se hizo anteriormente, antes de intercalar las variables.

E segundo procedimiento, se realiza tal como funciona generalmente el cálculo de la regresión, en cualquiera de las calculadoras que habitualmente manejamos, pero para ello, utilizamos de ayuda una nueva símbología que espero no cause dificultades al estudiante, pues la primera impresión que se siente, es de complejidad, lo cual no es cierto.

USO DE LA CALCULADORA

En primer lugar planteamos las dos ecuaciones, dependiendo de la variable que se desea estimar:

$\hat{Y} = b_{y.x}\, x + c_{y.x}$

$\hat{Y} = b_{y.x}(x - \bar{x}) + \bar{y}$

Cuando se desea estimar la variable $\hat{Y}$ en función de la variable . En nuestro caso *producción* en función del *costo*.

$\hat{X} = b_{xy.}\, y + c_{xy.}$

$\hat{X} = b_{xy.}(y - \bar{y}) + \bar{x}$

Cuando se desea estimar la variable $\hat{X}$ en función de la variable . En nuestro caso *costo* en función de *producción*.

Al calcular b , b , c , c , se deberá resolver mediante uno de los sistemas de ecuaciones normales, que a continuación se presenta:

$$(1)\ \Sigma y_i = b_{y.x}\, \Sigma x_i + n\, c_{y.x}$$
$$(2)\ \Sigma y_i x_i = b_{y.x}\, \Sigma x_i^2 + c_{y.x}\, \Sigma x_i$$

ó

$$(1)\ \Sigma x_i = b_{xy.}\, \Sigma y_i + n\, c_{xy.}$$
$$(2)\ \Sigma x_i y_i = b_{xy.}\, \Sigma y_i^2 + c_{xy.}\, \Sigma y_i$$

También las fórmulas, que ya fueron aplicadas anteriormente, para hallar el *coeficiente angular*:

$$b_{y.x} = \frac{n\,\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \quad \text{ó} \quad b_{xy.} = \frac{n\,\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma y_i^2 - (\Sigma y_i)^2}$$

$$b_{y.x} = \frac{\Sigma x_i y_i - n\,\bar{x}\,\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} \quad \text{ó} \quad b_{xy.} = \frac{\Sigma x_i y_i - n\,\bar{x}\,\bar{y}}{\Sigma y_i^2 - n\bar{y}^2}$$

$$b_{y.x} = \frac{Cov}{S_x^2} \quad \text{ó} \quad b_{xy.} = \frac{Cov}{S_y^2}$$

$$b_{y.x} = \frac{\Sigma(x_i - \bar{x})\,(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} \quad \text{ó} \quad b_{xy.} = \frac{\Sigma(x_i - \bar{x})\,(y_i - \bar{y})}{\Sigma(y_i - \bar{y})^2}$$

Para el *coeficiente de posición* serán:

$$c_{y.x} = \frac{\Sigma y_i - b_{y.x}\,\Sigma x_i}{n} \quad \text{ó} \quad c_{xy.} = \frac{\Sigma x_i - b_{xy.}\,\Sigma y_i}{n}$$

$$c_{y.x} = \bar{y} - b_{y.x}\,\bar{x} \quad \text{ó} \quad c_{xy.} = \bar{x} - b_{xy.}\,\bar{y}$$

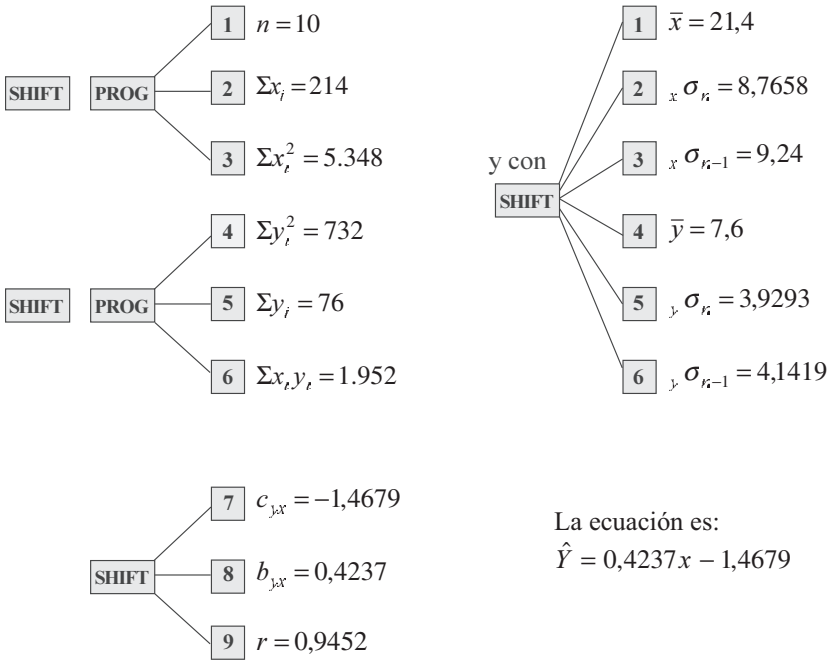
Estos valores se obtienen en la calculadora de la siguiente forma:

CASIO fx - 5.000F: primero **MODE** ÷ y aparece en pantalla LR (*regresión inea*); luego borramos por si acaso hay información así: **SHIFT** **DEL** **EXE**

Se procede a entrar la información así: **10** **SHIFT** **(** **3** **DT** ; luego **18** **SHIFT** **(** **5** **DT** y así sucesivamente. Es de anotar que antes de oprimir **DATA** ó **DT** en la pantalla deberá aparecer así **10** **3** **18** **5** ....

La tecla **SHIFT** **(** equivalente a , (coma).

Una vez digitada la información, se tendrá:



La ecuación es:
$\hat{Y} = 0{,}4237x - 1{,}4679$

Digitamos el valor de = 24 y luego ▢ y aparecerá en pantalla 24 $\hat{Y}$ le damos **EXE** y aparece el valor estimado 8,70 (miles de millones de pesos) como *costo*. Si en cambio digitamos 13 como costo y luego se oprime la tecla **X** y **ENTER** en pantalla primero aparece 13$\hat{X}$ y luego con **ENTER** se obtiene 34,14 (miles de millones de pesos), resultados exactamente iguales a los ya obtenidos para el mismo ejercicio, ahora mucho más rápido.

En cuanto a la *co arian a* que es igual para las dos ecuaciones se obtendrá así: **SHIFT** **2** **x²** **x** **SHIFT** **8** **EXE** igual a 32,56, siendo igual al resultado obtenido por el proceso anterior. Para el cálculo de la *arian a residua* o *no e p icada* de S^2 y S^2, se obtendrán así:

$S^2 = S^2(1 + r^2) \Rightarrow$

SHIFT **5** **x²** **(** **1** **–** **SHIFT** **9** **x²** **)** **EXE**

$S^2 = S^2(1 + r^2) \Rightarrow$

SHIFT **2** **x²** **(** **1** **–** **SHIFT** **9** **x²** **)** **EXE**

En el primero el resultado será de 1,64 y en el segundo es de 8,18.

Para determinar el coeficiente de correlación lineal (r) se deben oprimir las teclas **SHIFT** **9** **EXE** cuyo valor obtenido será de 0,9452 ≅ 0,95. Queda por determinar b , para ello partimos de que b (b) = R^2 luego $b = \frac{R^2}{b_{y.x}} = \frac{0{,}9452^2}{0{,}4237} = 2{,}108$ y la ecuación a utilizar será $\hat{X} = 2{,}1088\,(13 - 7{,}6) + 21{,}4 = 32{,}78$, ya que no se puede hallar con la calculadora en forma directa, sino mediante la aplicación de las fórmulas que se presentaron anteriormente.

■ En las calculadoras de otras marcas y modelos se procede así:

La fx-350 MS de CASIO se oprime **MODE** **REG** y luego se selecciona REG oprimiendo **3**; enseguida se selecciona con la tecla **1** Lin (lineal) procediendo a entrar la información: **10** ▢ **3** **DT** y así sucesivamente.

La salida de los datos, se hará: con SHIFT [1] aparece $\Sigma\ _i^2$; con [2] $\Sigma\ _i$ y con la tecla [3] el valor de n. Observe la flecha (→) en la pantalla, oprima con El REPLAY → aparece $\Sigma\ _i^2, \Sigma\ _i, \Sigma\ _i\ _i$, operando con las teclas [1] [2] [3] luego con [SHIFT] [2] en el primer pantallazo aparece $\sigma_{n_}\ \sigma_{n-1}$ cuyos valores se obtienen con las teclas [1] [2] [3] respectivamente; con REPLAY → aparece $\sigma_n\ \sigma_{n-1}$ y las mismas teclas anteriores son operadas; se oprime nuevamente REPLAY aparece A que equivale a C o β_0; $= \beta_1 = b$ y finalmente r (*Coeficiente de correlación*); se procede con REPLAY → aparece ˆ y ˆ.

Algo más sobre varian as

- $\frac{VE}{VT}100 = \frac{S_{\hat{y}}^2}{S_y^2}100 = \frac{S_{\hat{x}}^2}{S_x^2}100$ Es el porcentaje de la varianza total, que queda explicada por la recta de regresión.
- $\frac{VR}{VT}100 = \frac{S_{y.x}^2}{S_y^2}100 = \frac{S_{xy}^2}{S_x^2}100$ Porcentaje de la varianza total que no queda explicada por la recta de regresión.
- $1 = \frac{VR}{VT} + \frac{VE}{VT} \rightarrow$
- $V_{[x+y]} = S_x^2 + S_y^2 + 2cov$
- $V_{[x-y]} = S_x^2 + S_y^2 - 2cov$

Ejemplo 3. De una distribución de dos variables se conocen los siguientes datos:

$r = 0{,}90 \qquad S\ = 1{,}20 \qquad S\ = 2{,}10 \qquad \bar{\ } = 5 \qquad \bar{\ } = 10$

a) Obtener el valor de $\hat{X}$ cuando toma el valor de 40 y el de $\hat{Y}$ cuando se le da el valor de 30.
b) Calcular la varianza explicada de en función de

Solución:

$$r = \frac{m_{xy.}}{S_x\ S_y} \quad \Rightarrow \quad r(S_x)(S_y) = m_{xy.} \quad \Rightarrow \quad 0{,}9(1{,}2)\ (2{,}1) = m_{xy.} = 2{,}268$$

$$b_{xy.} = \frac{m_{xy.}}{S_y^2} = \frac{2{,}268}{4{,}41} = 0{,}5143 \cong 0{,}51$$

Siendo: $\hat{X} = b_{xy.}(y - \bar{y}) + \bar{x}$ Cuando $y = 40$ se tendrá que:

$$\hat{X} = 0{,}51(40 - 10) + 5 = 0{,}51(30) + 5 = 20{,}30$$

$$b_{yx} = \frac{m_{xy.}}{S_x^2} = \frac{2{,}26}{1{,}44} = 1{,}56 \qquad \hat{Y} = b_{yx}(x - \bar{x}) + \bar{y}$$

Cuando $x = 30 \quad \Rightarrow \quad \hat{Y} = 1{,}56\ (30 - 5) + 10 = 39 + 10 = 49$

b) La *arian a residua* y la *e p icada* de se obtendran de la siguiente forma:

$S_{yx}^2 = S_y^2 - \frac{m_{xy.}}{S_x^2} \qquad S_{yx}^2 = 4{,}41 - \frac{5{,}10}{1{,}44} = 4{,}41 - 3{,}54 = 0{,}87 \qquad (0{,}868333)^* \qquad (VR)$

$S_{\hat{y}}^2 = S_y^2 - S_{yx}^2 \qquad S_{\hat{y}}^2 = 4{,}41 - 0{,}87 = 3{,}54 \qquad (3{,}541666)^* \qquad (VE)$

* Resultados usando la calculadora

Ejemplo 4. El coeficiente de correlación lineal entre X y Y es 0,60. ¿Qué porcentaje de la variación total queda sin explicar por la ecuación de regresión?

Solución:

$$R^2 = 1 - \frac{S^2_{y.x}}{S^2_y} \qquad 0{,}36 = 1 - \frac{S^2_{y.x}}{S^2_y} \Rightarrow \frac{S^2_{y.x}}{S^2_y} = 1 - 0{,}36 = 0{,}64 = 64\%$$

El 64% de la varianza total queda sin explicar, por la recta de regresión.

Ejemplo 5. Si en una distribución bidimensional (dos variables) se tiene que: $[\quad] = 10{,}3$; $[\ - \ ] = 8{,}1$. ¿Cuál es la covarianza?

Solución:

a) $[\quad] = [X] + [Y] + 2m$ Multiplicamos la ecuación b) por -1 y la restamos de a)

b) $[\ - \ ] = [X] + [Y] - 2m$

a) $10{,}3 = [X] + [Y] + 2m$

b) $-8{,}1 = -[X] - [Y] + 2m$

$2{,}2 = 0 \quad 0 \quad 4m \qquad \Rightarrow \quad m = \frac{2{,}2}{4} = 0{,}55$

Ejemplo 6. Cinco niños de 2, 4, 6, 7 y 8 años pesan, respectivamente, 15, 19, 25, 38 y 34. Se pide hallar:

a) La recta de regresión mínima cuadrática del peso con respecto a la edad
b) El coeficiente de determinación correspondiente.

* Algunos de los resultados fueron obtenidos mediante el uso de la calculadora .

Solución:

a)

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
2	15	30	4	225
4	19	76	16	361
6	25	150	36	625
7	38	266	49	1.444
8	34	272	64	1.156
27	131	794	169	3.811

$$\bar{x} = \frac{27}{5} = 5{,}4 \qquad \bar{y} = \frac{131}{5} = 26{,}2$$

$$m_{xy} = \frac{\Sigma x_i y_i}{n} - \bar{x}\,\bar{y} \qquad m_{xy} = \frac{794}{5} - (5{,}4)\ (26{,}2) = 17{,}32$$

$$b_{y.x} = \frac{m_{xy}}{S^2_x} = \frac{17{,}32}{4{,}64} = 3{,}7328 \cong 3{,}73 \qquad S^2_x = \frac{\Sigma x_i^2}{n} - \bar{x}^2 = \frac{169}{5} - 5{,}4^2 = 4{,}64$$

$$\hat{Y} = 3{,}73(x - 5{,}4) + 26{,}2$$

b) $$R^2 = b_{xy}\ b_{y.x} = \frac{m^2_{xy}}{S^2_x\ S^2_y} \qquad S^2_y = \frac{\Sigma y_i^2}{n} - \bar{y}^2 = \frac{3.811}{5} - 26{,}2^2 = 75{,}76$$

$$b_{xy} = \frac{17,32}{75,76} = 0,2286 \cong 0,23 \qquad R^2 = (3,73)\ (0,23) = \frac{299,98}{(4,64)\ (75,76)} = 0,85$$

(Hay una pequefla diferencia en los resultados obtenidos, por las aproximaciones hechas)

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. Con los siguientes datos:

   $\Sigma_{\ i} = 110 \quad n = 5 \quad \Sigma_{\ i\ i} = 1.868 \quad \Sigma_{\ i}^{\ 2} = 3.156 \quad \Sigma_{\ i} = 60 \quad \Sigma_{\ i}^{\ 2} = 1.138$

   Se pide calcular:

   a) El error estándar de estimación *Y*.
   b) La varianza explicada de *Y*
   c) Estime el valor de $\hat{Y}$ cuando = 80, con un margen de confianza del 95%.

2. Dos variables tienen las siguientes rectas de regresión mínimo cuadrática:
   $8X + 2Y = 1 \qquad 16X + 9Y = 1.$
   Calcular el coeficiente de correlación al cuadrado.

3. La línea de regresión de fruta agusanada (*Y*) y el tamaño de la producción del árbol, en centenas de frutas (X) es $\hat{Y} = 64 - 1,4$
   Si se sabe que $\bar{\ } = 45 \quad s\ = 7,2 \quad s\ = 10,5$ encontrar:
   a) $\bar{\ }$ b) Coeficiente de correlación c) La línea de regresión de $\hat{X}$ sobre

4. De una encuesta sobre presupuestos familiares quincenales, se publicaron los siguientes datos:

   Desviación típica de los gastos quincenales = 2.000 cientos de pesos

   Desviación típica de los ingresos quincenales = 3.000 cientos de pesos

   Coeficiente de correlación entre ingresos y gastos = 0,85

   Con base en estos datos, se pide calcular la desviación típica de los ahorros, entendiendo por tales las diferencias entre ingresos y gastos quincenales.

5. Dados: $r = 0,8 \quad s\ = 6 \quad Y = 7 + 4$ se pide calcular $s\ $.

6. Se ha efectuado un estudio donde se relacionan los puntajes de aptitud con la productividad en una industria. Después de tres meses de entrenamiento del personal, sus postulantes, elegidos al azar, obtuvieron los seis pares de puntajes de aptitud y productividad que se indican a continuación:

A. PUNTAJE DE APTITUD:	9	17	20	19	20	23
B. PRODUCTIVIDAD:	23	35	29	33	43	32

   a) ¿Cuál es la productividad esperada de un trabajador, cuyo puntaje de aptitud fue de 16?
   b) Obtener el coeficiente de correlación.
   c) El error estándar de estimación.
   d) Fije con una probabilidad del 95%, la productividad estimada en el punto a).

7. De una distribución bidimensional de frecuencias se sabe, que la recta de regresión mínima cuadrática de la segunda variable sobre la primera es $\hat{Y} = 73 - 0{,}6$; la media de la primera variable es 8, la desviación típica de dicha variable es 2 y la varianza de la segunda es 9. Calcular el coeficiente de correlación rectilínea y la media de la segunda variable.

8. Con los siguientes datos:

   $\Sigma_i = 14.225 \quad \Sigma_i = 287 \quad \Sigma_i = 273 \quad \Sigma_i^2 = 14.911 \quad \Sigma_i^2 = 13.599 \quad n = 8$

   a) Calcule el error estándar de estimación de la recta con base en .

   b) Calcule la varianza explicada de la variable *Y*.

   c) Calcule el coeficiente de correlación rectilíneo.

   d) Obtenga la recta de regresión mínima cuadrática de $\hat{Y}$ con base en .

9. En 50 matrimonios, se analizó la situación económica por las entradas diarias (miles de $) del esposo *X* y de la esposa *Y* encontrándose los siguientes estadígrafos:

   $\bar{} = 6{,}6$ miles de pesos; $\bar{} = 20$ miles de pesos $\quad S^2 = 30 \quad S^2 = 60 \quad m = 8{,}4$

   En las condiciones anteriores, ¿qué entradas diarias le corresponden a la esposa de un señor que gana en un día $28.000,oo?

10. De una tabla de correlación se han calculado los siguientes datos:

    $b_{xy} = -0{,}2 \quad R^2 = 0{,}9 \quad S_x^2 = 8 \quad \bar{x} = 2$

    Se pide calcular la varianza de la suma de dos variables.

11. Si $S_{y.x} = 3 \quad S_y = 5$ Hallar el valor de *r*.

12. Conteste si son ciertas o falsas las siguientes relaciones:

    a) $m_{xy} = 10 \quad S_x = 5 \quad S_y = 8 \quad S_{a_y}^2 = 4$

    b) $b_{y.x} = 0{,}95 \quad b_{xy} = 1{,}24$

    c) $b_{xy} = -0{,}7 \quad b_{y.x} = 0{,}9$

    d) $m_{xy} = -10 \quad r = 0{,}8$

    e) $V_{[x-y]} = -14 \quad V_{[x]} = 8 \quad V_{[y]} = 12 \quad m_{xy} = 17$

13. Determine si las siguientes relaciones son posibles o no y justifique su respuesta.

    a) $S_{y.x}^2 = 19 \quad R^2 = 0{,}81 \quad S_y = 10$

    b) $\hat{Y} = 4 + 7x \quad \bar{x} = 10 \quad \bar{y} = 64 \quad r = 0{,}8$

    c) $b_{y.x} = 0{,}3 \quad b_{xy} = -3 \quad R^2 = -0{,}9$

    d) $m_{xy} = 60 \quad S_x^2 = S_y^2 = 50 \quad b_{y.x} = 1{,}2$

14. De un análisis de regresión rectilínea, se tomaron los siguientes datos a partir de 20 elementos de una muestra:

$M_{[x_i^2]} = 81 = \dfrac{\Sigma x_i^2}{n}$ $M_{[x-y]} > 0$ $M_{[x+y]} = 7$

$m_{xy} = 50$ $M_{[y^2]} = 49 = \dfrac{\Sigma y_i^2}{n}$ $\Sigma x_i y_i = 1.240$

Calcular el error de estimación de $\hat{Y}$ con base en .

15. Hallar la covarianza si: $V_{[x+y]} = 9{,}2$ y $V_{[x-y]} = 14{,}8$

16. Para dos variables que tienen los siguientes datos:

$S_y = 8$ $S_x = 10$ $M_{[y]} = 13 = \bar{y}$ $M_{[x]} = 16 = \bar{x}$ $r = 0{,}6$

Calcular S^2 y la ecuación de $\hat{Y}$ con base en .

17. Conteste cierto o falso a los siguientes puntos:
    a) Un coeficiente de correlación igual a –1 indica que la línea de regresión no es la que mejor refleja las relaciones entre las dos variables.
    b) La recta de regresión pasa por el punto donde se cortan las medias aritméticas.
    c) El producto de las desviaciones típicas es menor que la covarianza.
    d) Si *r* es mayor que uno, la recta ajustada explica más que suficiente las relaciones entre las dos variables.
    e) El coeficiente de correlación se da en valores menores de –1 y mayores de 1.

18. En una tabla de frecuencias correspondiente a 20 parejas se obtuvo la siguiente información:

$\Sigma x_i = 174$ $\Sigma y_i^2 = 5.904$ $\Sigma y_i = 340$ $\Sigma x_i y_i = 3.062$ $\Sigma x_i^2 = 1.626$

Se pide estimar el valor de $\hat{Y}$ cuando = 30; (promedio), con una confianza del 95%.

19. Las variables *X* y *Y* están correlacionadas. De las observaciones numéricas correspondientes se ha obtenido: $S_x = 3$ $S_y^2 = 25$ $m_{xy}^2 = 400$

Hallar el coeficiente de correlación lineal entre dichas variables y discutir el resultado.

20. Siendo: $b_{yx} = 0{,}40$ $r = 0{,}94$ $S_x^2 = 18$ Se pide:
    a) Determinar el porcentaje de la varianza que queda sin explicar, en la regresión de $\hat{Y}$ en función de .
    b) La varianza de la suma de las dos variables *X* y *Y*.

21. Con los siguientes datos, correspondientes a 20 parejas de observaciones.

$\Sigma x_i = 174$ $\Sigma y_i = 340$ $\Sigma x_i^2 = 1.626$ $\Sigma y_i^2 = 5.904$ $\Sigma x_i y_i = 3.062$ $n = 20$

    a) Se pide estimar el valor de $\hat{Y}$ cuando = 10, aplicando la regresión rectilínea.
    b) Calcular el coeficiente correlación lineal.

22. De un análisis de regresión se obtuvo que: $b_{yx} = 0{,}2$ $r = 0{,}9$ $V_{[x]} = 20$ Hallar la $V_{[x+y]}$

23. Si la $V_{[x]} = 56{,}09$ $V_{[y]} = 64{,}8$ $M_{[xy]} = 11.223{,}5$ $M_{[y^2]} = 4.567{,}21$ $M_{[x^2]} = 27.812{,}25$.

Hallar el valor de $\hat{X}$ cuando toma el valor de 20.

Nota: $V_{[x]} = S_x^2 \quad V_{[y]} = S_y^2 \quad M_{[xy]} = \frac{\Sigma x_t y_t}{n} \quad M_{[y^2]} = \frac{\Sigma y_t^2}{n} \quad M_{[x^2]} = \frac{\Sigma x_t^2}{n}$

24. De dos variables, *X* y *Y*, relacionadas entre sí, se sabe que:

$\bar{x} = 8 \quad \bar{y} = 21 \quad S_x = 2 \quad S_y^2 = 18 \quad m_{xy} = 3{,}6$

a) Hallar la ecuación de la recta mínima cuadrática de $\hat{Y}$ en función de .

b) El coeficiente de correlación.

25. Durante varios años se recopilaron datos sobre ventas (X) y gastos de publicidad (*Y*), expresados en millones de $.

$\bar{x} = 28 \quad \bar{y} = 1{,}2 \quad S_x^2 = 8 \quad S_y^2 = 2 \quad Cov = 4$

En los años siguientes, los datos de publicidad se aumentaron en un 15%, lo que originó un incremento en las ventas. Calcular e interpretar:

a) El coeficiente de correlación de Pearson entre las dos variables, antes y después del incremento

b) ¿Cuál será el valor esperado en las ventas, si gastan $3.000.000 en publicidad antes y después del incremento?

Nota: al incrementar cada valor de la variable Y en un 15%, se modifica la media, la varianza y la covarianza, aplicando las propiedades del producto así:

$\bar{y} = 1{,}15(\bar{y}) = 1{,}15(1{,}2) = 1{,}38$

$S_y^2 = 1{,}15^2(S_y^2) = 1{,}15^2(2) = 2{,}645$

$Cov = 1{,}15(Cov) = 1{,}15(4) = 4{,}6$

26. En un curso de Estadística para 20 alumnos se dieron los siguientes resultados, correspondientes a las notas de parciales ( ) y el examen final ( )

$\bar{x} = 70 \quad \bar{y} = 75 \quad \Sigma(\hat{Y}_i - \bar{y})^2 = 115{,}2 \quad S_x = 4 \quad S_y^2 = 64$

a) ¿Qué porcentaje de la varianza total de queda sin explicar por la recta de regresión?
b) Un estudiante obtuvo una calificación de 62 en el examen parcial. ¿Cuál es nuestra predicción para la calificación que obtendrá en el examen final?
c) Cuál será la varianza total de la suma de las variables (*X* + *Y*)

27. Se conoce:

a) $\Sigma(\hat{Y}_i - \bar{y})^2 = 810$
b) $n = 30$
c) La suma de los cuadrados de las variaciones que no queda explicada por la recta de regresión de *Y* en función de es de 290.

Se pide:

a) Calcular el coeficiente de correlación rectilínea
b) Siendo $\hat{Y} = 120$; fijar los límites de confianza del 90% para dicho estimativo.

28. El 31 de julio de 2010 se tomaron los datos de los empleados de una empresa, siendo *X* = salarios en cientos de pesos; Y = tiempo de servicio en años; $\bar{x} = 3.820 \quad \hat{Y} = 4 \quad S_y^2 = S_x^2 = 2.500$ y Cov = 300.

Si el 31 de diciembre de 2010, el número de empleados en la empresa era el mismo y los salarios fueron aumentados en el 20% en agosto del mismo año, ¿cuál será el coeficiente de correlación de las dos variables a 31 de diciembre?

29. Al observar el número de sucursales (X) y los costos mensuales (Y) en comunicación telefónica con la casa central, en millones de $, para 10 empresas se encontró:

x_i :	3	5	3	2	4	1	5	2	6	3
y_i :	2	3	5	4	6	2	5	1	3	5

Apoyándose en la información anterior, ajuste la recta que muestre la relación de los costos en función del número de sucursales y estime los costos de una empresa con 10 sucursales.

30. Diga si los siguientes puntos son ciertos o falsos y por qué.

a) En una serie bidimensional de datos es posible que; $S^2_{(xy)} = 28{,}6 \quad S^2_y = 28{,}2$

b) Al estudiar las relaciones entre *X* y *Y* es posible que: $\hat{Y} = 5 + 4{,}8 \quad \bar{\ } = 5 \quad \bar{\ } = 20$

c) $Cov = 80 \quad S^2_x = S^2_y = 50 \quad b_{yx} = 1{,}2$

d) $S^2_{(xy)} = 19 \quad R^2 = 0{,}81 \quad S^2_y = 10$

e) $b_{xy} = 0{,}3 \quad b_{yx} = -30 \quad R^2 = -0{,}9$

31. Se desea buscar la relación existente entre las ventas (*Y*) (en millones de $) y las utilidades (*X*) de una compañía con 50 sucursales. Se dispone de la siguiente información.

$$\Sigma x_i = 100 \qquad \frac{\Sigma x_t y_t}{n} = 30 \qquad \Sigma y_t^2 = 5.390{,}5 \qquad \Sigma y_i = 500 \qquad r = 0{,}8$$

Halle e interprete:

a) $X = b_{xy} y + C_{xy}$ b) S^2_{xy} y S^2_x

c) Las utilidades de una compañía con 20 millones pesos en ventas.

32. Considere dos variables, una continua (*X*) y otra discreta (*Y*). Se pide ordenar los datos en una tabla de frecuencia bidimensional. Considere $n = 30$ para obtener el valor de la amplitud del intervalo ( ) en la variable *X*.

x_i:	2	9	7	6	5	8	7	9	3	9	6	5	6	6	8	9	7	8	6	7	8	3	8	7	9	5	7	6	4	8
y_i:	2	3	1	3	2	1	1	3	2	3	4	2	3	2	3	4	3	3	1	3	4	3	4	3	4	3	4	2	3	3

Se pide:

a) Estimar el valor de $\hat{Y}$ cuando = 6.

b) Fijar límites de confianza de $\hat{Y}$, con una P = 95%.

c) Determinar el coeficiente de correlación.

33. De un análisis de regresión se obtuvo que: $b = 0{,}2 \quad r = 0{,}9$ y $S^2 = 20$ hallar la varianza de la suma de las dos variables (+) .

34. Con los siguientes datos: $\Sigma_i = 45 \quad \Sigma_i = 15 \quad \Sigma_i^2 = 495 \quad \Sigma_i^2 = 55 \quad \Sigma_i{}_i = 165 \quad n = 5$

Se pide

a) Calcular el coeficiente de correlación
b) Estimar el valor de $\hat{Y}$ cuando = 6, y establecer los límites de confianza del 95%
c) Determinar el porcentaje de la varianza que quedó sin explicar por la recta de regresión.

35. Con los ingresos y gastos mensuales de 10 familias (en cientos de $)

INGRESO x_i :	6.500	7.000	8.000	10.000	11.000	7.500	8.000	9.000	10.800	8.300
GASTOS y_i :	6.300	6.700	7.000	8.200	9.000	7.300	7.900	8.500	8.600	7.700

a) Dibujar el diagrama de dispersión o nube de puntos.
b) Encontrar la línea de regresión de $\hat{X}$ (ingresos) en función de (gastos), dibujarla en el diagrama.
c) Calcular los ingresos de una familia que tiene un gasto mensual de 12.400 (cientos de pesos).

36. Siendo $\bar{x} = 30$ $S_x^2 = 25$ $r = 0{,}80$ $\bar{y} = 45$ $S_y^2 = 8$

a) Estimar el valor de $\hat{X}$ cuando = 28

b) Determinar la varianza residual.

37. Diga si las siguientes relaciones son verdaderas o falsas. ¿Por qué?

a) $\hat{Y} = 0{,}932x + 6{,}84$ $\bar{x} = 104$ $\bar{y} = 105$

b) $\Sigma(y_i - \hat{Y}_i)^2 = 24$ $\Sigma(\hat{Y}_i - \bar{y})^2 = 16$ $n = 5$ $S_y^2 = 18$

c) $m_{xy} = 4$ $\hat{Y} = \bar{y} + 4(x - \bar{x})$ $S_x^2 = 9$

d) $\hat{Y} = 0{,}2x - 4$ $S_x^2 = 50$ $\bar{x} = 30$ $\Sigma x_t y_t = 50n$

e) $S_x^2 = 16$ $r = 0{,}3$ $S_{yx} = 6$ $\hat{Y} = 2x + 5$

38. Una compañía de transporte analiza una muestra de 10 embarques por camión y registra la distancia en kilómetros, además el tiempo de entrega de la mercancía medido por días, obteniéndose la siguiente información:

$\Sigma x_t = 7.620$ $\Sigma y_t = 28{,}5$ $\Sigma x_t y_t = 26.370$ $\Sigma x_t^2 = 7.104.300$ $\Sigma y_t^2 = 99{,}75$

a) La compañía informa que si el kilometraje es de 1.500, el tiempo necesario para la entrega es mayor de tres días. ¿Está usted de acuerdo?
b) ¿Qué porcentaje de información es explicada por la recta de regresión?

39. De un grupo de 50 individuos, del cual se examinan dos variables, se tiene la siguiente información:

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 241$ $\Sigma(\hat{Y}_t - \bar{y})^2 = 196{,}34$ $\Sigma(x_t - \bar{x})^2 = 310$

$\Sigma(y_t - \bar{y})^2 = 218{,}15$ $\bar{x} = 20$ $\bar{y} = 30$

Se pide:

a) Estimar el valor de $\hat{Y}$, cuando = 30, fijando límites del 92,8%
b) ¿Qué porcentaje de la varianza total queda sin explicar por la recta de regresión de en ?

40. Suponga que se está estudiando el efecto de cierta fricción constante y uniforme sobre láminas metálicas del mismo grosor y de la misma aleación. Se han efectuado 8 observaciones cuyos resultados son:

X = Tiempo de fricción Y = Grosor de la lámina en mm.

:	3	5	6	8	9	10	12	15
:	9,5	9,0	8,8	8,5	8,2	8,0	7,6	7,0

Se pide mediante el cálculo de la ecuación de regresión:

a) ¿Cuál fue el grosor original de las láminas?

b) ¿Cuántas horas tardará la fricción en destruir una lámina?

c) ¿Tiene algún sentido pronosticar el grosor de una lámina después de 70 horas de fricción?

41. La solubilidad en el agua de cierto producto orgánico se hace más rápida con el auxilio de un agente químico. Según presumimos de las siguientes observaciones efectuadas en 6 muestras de 20 gramos de producto y diversas dosis del agente químico:

AGENTE (gramos):	3	2	1	3	5	4
SOLUBILIDAD (segundos):	7	9	10	8	5	6

a) Calcule el coeficiente de correlación.

b) ¿Cuál es la rapidez de solución de 10 gramos del producto en ausencia del agente químico?

c) ¿Qué proporción del agente deberá contener el producto para que la solución sea instantánea?

42. Se ha observado, que para predecir la demanda (consumo) de combustible para calefacción, resulta ser más preciso el pronóstico a largo plazo de las temperaturas y el uso, de la relación temperatura - consumo, que el tratar de pronosticar directamente la demanda analizando las ventas de combustibles. Un distribuidor mantiene un registro de ventas mensuales de combustibles y de temperaturas máximas en esos meses. A continuación aparecen los datos de 8 de estos meses seleccionados al azar.

VOLUMEN VENTAS:	16,2	17,4	12,3	35,9	42,1	26,4	19,0	10,1
PROMEDIO DE TEMPERATURAS MÁXIMAS (grados centígrados):	18,3	21,2	17,6	23,4	25,2	20,4	21,6	18,6

Se pide:

a) Grafique los puntos (observaciones reales) y las rectas de $\hat{Y}$ y $\hat{X}$

b) ¿Qué porcentaje de la variación en las ventas mensuales se explica por la temperatura máxima promedio?

c) Encuentre el intervalo de confianza del 95% para las ventas mensuales esperadas (medias) en aquellos meses en que el promedio de la temperatura máxima sea de 25 grados centígrados.

43. El dueño de una taberna encontró una correlación de 0,80 entre el tiempo de permanencia de los compradores en el establecimiento (X) y el valor de la cuenta cancelada (Y).

PERMANENCIA (minutos):	$\bar{} = 25$	$s = 7{,}5$
VALOR DE LA CUENTA (miles de $):	$\bar{} = 60$	$s = 14{,}4$

a) Para un tiempo de permanencia de 40 minutos, estimar el valor de la cuenta (en miles de $)

b) ¿Cuál es el porcentaje de la varianza total que queda sin explicar por la recta de regresión?

44. Cuando el coeficiente de correlación es 0,90, dibujar la gráfica de las líneas de regresión ($\hat{Y}$ y $\hat{X}$); sabiendo que:

$\bar{} = 30 \qquad s = 5 \qquad \bar{} = 45 \qquad s^2 = 64$

45. Un gerente de producción, en asocio con el jefe de personal, hizo un estudio al personal empleado en el proceso de manufactura. Encontró que después de que estos empleados habían permanecido en el trabajo por un año, eran capaces de obtener una medida de actuación de dichas personas, que reflejase adecuadamente su eficiencia. De modo que diseñó una prueba de selección para predecir su eficiencia eventual y obtuvo una correlación de 0,65 en la medida de eficiencia (Y). La media de la prueba es 50; $s = 6 \quad \bar{} = 100 \quad s = 10$. Contestar las siguientes preguntas basadas en esos hechos:

Si un empleado obtiene una calificación de 40 en la prueba de selección.

a) ¿Cuál es su calificación predicha de eficiencia?

b) ¿Qué porcentaje de la varianza total queda sin explicar por la recta de regresión?

46. Responda VERDADERO cuando el enunciado sea válido. En caso contrario, la palabra en negrilla deberá sustituirse por otra que sea válida.

a) Un **diagrama de frecuencias** consiste en la representación de todos los pares ordenados de datos observados en un sistema de ejes coordenados.

b) El **coeficiente de variación**, es la medida que nos determina el grado de correlación entre dos variables.

c) Un coeficiente de correlación **cero** significa que las dos variables se encuentran perfectamente correlacionadas.

d) Una **varian a no explicada** igual a **cero** significa que la varianza total es igual a la varianza explicada.

e) El coeficiente de correlación lineal varía entre **cero** y + **1**.

f) El valor que se predice corresponde a la variable **dependiente**.

47. En una regresión lineal simple para la cual se tienen 30 observaciones, se halla que $R^2 = 0{,}70$ y $S_a^2. = 1{,}2$. Halle S^2 y S^2

48. En un análisis de regresión simple y de correlación basado en 150 observaciones, se halla que $r = 0{,}85$ y $S^2 = 16$.

a) Determinar la varianza explicada de *Y*.

b) Además, la proporción de la variación no explicada respecto a la variación total.

49. En un análisis de regresión con 20 observaciones, la variación explicada es de 56,2 y la total 66,3. Calcular la desviación residual para la ecuación de regresión $\hat{Y} = b + c$.

50. En una regresión lineal simple, el valor estimado del coeficiente de regresión $b = 2{,}5$, con $S_b = 0{,}8$

a) Establecer un intervalo de confianza del 95% para el verdadero parámetro de β , si n = 25

b) Probar al nivel del 5% el coeficiente de regresión.

51. Supóngase que al analizar la relación entre 30 pares de observaciones, se encuentra que la suma del cuadrado de las diferencias entre los valores de la variable y su media aritmética es de 220, $(\Sigma(\ _i - \bar{\ })^2 = 220)$; y que el cuadrado de las diferencias entre los valores estimados y su media es de 38,4; $\left(\sum (\hat{Y}_i - \bar{\ })^2 = 38{,}4\right)$. La pendiente de la recta es positiva.

    a) Calcule el coeficiente de determinación.

    b) ¿Qué porcentaje de la varianza ha sido explicada por la recta de regresión?

    c) ¿Puede rechazarse la hipótesis de que no existe correlación lineal a un nivel de significación del 5%?

52. El departamento de pinturas en una cadena de almacenes tiene 2 promociones en el año, una en enero y la otra en octubre. El precio de la pintura se reduce del 10 al 30% durante estas realizaciones. La administración desea aumentar ese porcentaje en Enero. Aunque no va a tener utilidades si aumenta el porcentaje, el gerente de mercados espera generar ventas de artículos auxiliares, tales como: brochas, rodillos, cintas protectoras y diluyentes de pinturas que le ayudaría a aumentar las utilidades. Un nuevo empleado del departamento de asesoría en mercados está deseoso de impresionar al Comité de fijación de política de precios. Uno de ellos dice "nada molesta más a un cliente que asistir a una realización y encontrar que el artículo anunciado se agotó". Además se debe tener una provisión de tal manera que los costos de inventario puedan contrarrestar las utilidades de la realización.

PORCIENTO DE DESCUENTO:	10	12	14	16	18	20	22	24	26	28	30
PORCIENTO DE LA VARIACIÓN DE LAS VENTAS:	3,43	3,75	4,52	5,13	5,94	6,35	7,99	8,23	9,46	10,35	12,00

    a) Si la empresa considera que tiene unos inventarios que permiten el aumento de las ventas en un 18%, ¿qué porcentaje de descuento podrá dar?

    b) ¿Qué porcentaje de la varianza total no queda explicada por la recta de regresión?

    c) ¿Qué grado de correlación hay entre estas variables, la aceptaría usted?

    d) Fije un intervalo de confianza al 95% para el valor estimado del punto (a).

53. Supongamos que una cadena de supermercados financia un estudio sobre gastos mensuales en alimentos, de familias de 4 miembros. La investigación se limitó a familias con ingresos netos entre $688.000 y $820.000, con lo cual se obtuvo la siguiente ecuación: $\hat{Y}$ = -18.000 + 0,85 .
    = gastos mensuales estimados en alimentos; = ingresos.
    Ademas suponga que la ecuación proporciona un buen ajuste

    a) Estime los gastos en alimentos en un mes, para una familia de 4 miembros, con un ingreso de $700.000

    b) Uno de los directivos de la compañía se preocupa por el hecho de que la ecuación aparentemente indica que para una familia que tiene un ingreso de $12.000 no gastaría nada en alimentos. ¿Cuál sería su respuesta?

    Explique brevemente que sucedería si la ecuación anterior se utiliza para cada uno de los siguientes casos:

    c) Estimación de los gastos en alimentos de familias de cinco miembros, que tienen un ingreso de $598.000

d) Estimación de los gastos en alimentos de familias de cuatro miembros, con ingresos de $450.000

54. Suponga que Y = nota promedio en el examen de ingreso (escala de 0 - 100) y X = miles de pesos en gastos semanales, que incurre en el primer año de estudio en esa universidad, sabiendo que la recta de regresión Y sobre X está expresada por la ecuación $\hat{Y} = 50 + 0{,}3$.

a) ¿Cuál es la interpretación del coeficiente de regresión igual a 0,3.

b) ¿Qué valor de $\hat{Y}$ se podría estimar si = 40.000?

55. En una regresión lineal de $\hat{Y}$ sobre X, basada en 20 observaciones, el valor del coeficiente de determinación es de 0,64.

a) ¿Indica este resultado que existe una correlación significativa entre X y Y a un nivel del 5%, para una prueba bilateral?

b) ¿Cuál es la proporción de la variación que queda sin explicar por esta recta de regresión?

56. El analista de una fábrica de automóviles quiere desarrollar un modelo estadístico para predecir el tiempo de entrega (los días entre la compra del automóvil y la entrega real del mismo) de automóviles nuevos de fabricación especial. El cree que hay relación lineal entre las opciones del automóvil ordenado y su tiempo de entrega. Se selecciona una muestra aleatoria de 16 automóviles comprados, con los resultados que se presentan a continuación:

No. DE OPCIONES ORDENADAS:	3	4	4	7	7	8	9	11
TIEMPO DE ENTREGA (EN DÍAS):	25	32	26	38	34	41	39	46
No. DE OPCIONES ORDENADAS:	12	12	14	16	17	20	23	25
TIEMPO DE ENTREGA (EN DÍAS):	44	51	53	58	61	64	66	70

a) ¿Se puede predecir cuántos días se necesitarían para la entrega de un automóvil, si el analista considera que se puede presentar el caso de un automóvil con 24 opciones?

b) Estime un intervalo de confianza del 95% del tiempo promedio de entrega para todos los automóviles comprados con 15 opciones.

c) Pruebe al nivel del 5%, la existencia de una asociación real entre el número de opciones y el tiempo de entrega. (Probar si hay correlación).

d) Pruebe al nivel del 5% que hay una relación significativa entre las variables X y Y. (Relación lineal significativa).

57. Cinco empresas arrojan las siguientes cifras: X = utilidad bruta; Y = costo mano de obra (millones de pesos semanales).

$_i$:	4	9	8	10	12
$_i$:	3	8	8	6	4

Calcule por el método de los mínimos cuadrados, la utilidad bruta esperada si el costo de la mano de obra es de $10 millones semanales.

58. Con los siguientes datos: $\Sigma\ _i = 15$ $\Sigma\ _i^2 = 55$ $\Sigma\ _i = 45$ $\Sigma\ _i^2 = 495$ $\Sigma\ _i\ _i = 165$ $n = 5$

a) Establecer la ecuación de en función de .

b) Calcular el coeficiente de correlación de Pearson

c) Calcular el % de la varianza que queda explicada por la recta de regresión.

59. La décima estación de policía de la ciudad desea determinar el tiempo de respuesta para llegar al sitio solicitado. La presencia depende linealmente de la distancia que hay del sitio a la estación. Para 35 llamadas, los datos son:

$\Sigma\ _i = 234$ $\Sigma\ _i^2 = 1.796$ $\Sigma\ _i = 831$ $\Sigma\ _i^2 = 21.037$ $\Sigma\ _i\ _i = 5.890$

a) Estimar $\hat{Y}$ en función de X.

b) Calcular el coeficiente de correlación.

60. Con los siguientes datos de 10 pares de observaciones:

$_i$:	12	22	36	20	26	34	18	20	25	30
$_i$:	5	8	12	10	14	16	10	13	18	19

a) Se puede estimar el valor de $\hat{X}$ cuando = 15.

b) Calcular el coeficiente de determinación.

Regresión lineal ponderada

Cuando se tiene un número grande de pares de observaciones, digamos 50, 100 o más, no hay problema si aplicamos EXCEL, pero si no contamos con esa facilidad o la información se da en cuadro de doble entrada, no tenemos más remedio que hacerlo manualmente en forma simplificada. Veamos un ejemplo de *producción* y *costo*, pero no de 10 pares de observaciones, sino de 80 establecimientos, cuya información se presenta a continuación.

Ejemplo 1.

Tabla 12.2 Tabla de cálculo

PRODUCCIÓN (miles mill)	COSTO (miles mill)	No. ESTABLE TOS
$_i$	$_i$	n_i
10	3	10
18	5	8
12	4	12
16	5	5
22	8	15
36	12	6
30	10	4
32	14	10
26	12	8
12	3	2
Σ		80 = n

$_i n_i$	$_i n_i$	$_i^2 n_i$	$_i^2 n_i$	$_i\ _i n_i$
30	100	90	1.000	300
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
617	1.686	5.961	41.172	15.476
$_i f_i$	$_i f_i$	$_i^2 f_i$	$_i^2 f_i$	$_i\ _i f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} \qquad \bar{y} = \frac{617}{80} = 7{,}7125 \cong 7{,}71$$

$$\bar{x} = \frac{\Sigma x_t n_t}{n} \qquad \bar{x} = \frac{1.686}{80} = 21{,}075$$

$$s_x^2 = \frac{\Sigma x_t^2 n_t - n\bar{x}^2}{n} \qquad s_x^2 = \frac{41.172 - 80\,(21{,}075)^2}{80} = 70{,}4943$$

$$s_y^2 = \frac{\Sigma y_t^2 n_t - n\bar{y}^2}{n} \qquad s_y^2 = \frac{5.961 - 80\,(7{,}7125)^2}{80} = 15{,}0298$$

$$Cov = \frac{n\,\Sigma x_i\, y_i n_i - (\Sigma x_i\, n_i)\,(\Sigma y_i\, n_i)}{n} \qquad Cov = \frac{80\,(15.476) - (1.686)(617)}{80} = 30{,}9091$$

$$Cov = \frac{\Sigma x_i\, y_i n_i}{n} - \bar{x}\,\bar{y} \qquad Cov = \frac{15.476}{80} - 21{,}075(7{,}7125) = 30{,}9091$$

$$b = \frac{Cov}{s_x^2} \qquad b = \frac{30{,}9091}{70{,}4943} = 0{,}4384 \Rightarrow \text{Con calculadora da el mismo valor}$$

$$b = \frac{n\,\Sigma x_i\, y_i\, n_i - (\Sigma x_i\, n_i)\,(\Sigma y_i\, n_i)}{n\Sigma x_i^2\, n_i - (\Sigma x_i\, n_i)^2} \qquad b = \frac{80(15.476) - (1.686)(617)}{80(41.172) - (1.686)^2} \cong 0{,}4384$$

$$b = \frac{\Sigma x_i\, y_i\, n_i - n\bar{x}\,\bar{y}}{\Sigma x_i^2\, n_i - n\bar{x}^2} \qquad b = \frac{15.476 - 80\,(7{,}7125)\,(21{,}075)}{41.172 - 80\,(21{,}075)^2} \cong 0{,}4384$$

Mediante un sistema de ecuaciones normales:

$$\begin{array}{l} (1)\quad \Sigma y_t\, n_t = b\Sigma x_t\, n_t + nc \\ (2)\quad \Sigma y_t\, x_t\, n_t = b\Sigma x_t^2\, n_t + c\Sigma x_t\, n_t \end{array} \Rightarrow \begin{array}{l} (1)\quad 617 = 1.686b + 80c \\ (2)\quad 15.476 = 41.172b + 1.686c \end{array}$$

El cálculo de $b = \beta_1$ y $c = \beta_0$ lo podemos calcular de varias maneras.

donde $b = 0{,}4384$ y $c = -1{,}52807$ (resultados con la calculadora)

En forma directa el valor de c se obtendrá así:

$$c = \frac{\Sigma y_t\, n_t - b\Sigma x_t\, n_t}{n} \qquad c = \frac{617 - 0{,}4384\,(1.686)}{80} \cong -1{,}52806$$

$$c = \bar{y} - b\bar{x} \qquad c = 7{,}7125 - 0{,}4384\,(21{,}075) \cong -1{,}52806$$

En la regresión lineal simple si queremos estimar los costos de una empresa, presuntamente su producción es de 24 miles de millones de pesos, lo hacemos con los datos de las 80 empresas.

$\hat{Y} = b + c \Rightarrow \hat{Y} = 0{,}4384\,(24) - 1{,}52806 = 8{,}99$ miles de millones de \$, sería el costo estimativo para esa empresa.

Ahora fijamos los límites de confianza para ese estimativo; en primer lugar hay necesidad de calcular la varianza residual, por alguno de los siguientes métodos:

$$s^2_{y.x} = \frac{\Sigma (y_t - \hat{y}_t)^2 n_t}{n} \quad \Rightarrow$$ No lo realizamos por el espacio y tiempo que exige la aplicación de esta fórmula, pero el lector podrá hacerlo.

$$s^2_{y.x} = \frac{\Sigma y_t^2 n_t - c\Sigma y_t n_t - b\Sigma y_t x_t n_t}{n}$$

$$s^2_{y.x} = \frac{5.961 + 1{,}52806\,(617) - 0{,}4384\,(15.476)}{80} \cong 1{,}49$$

$$s^2_{y.x} = s^2_{y.} - \frac{Cov^2}{s^2_x} \qquad s^2_{y.x} = 15{,}030 - \frac{30{,}91^2}{70{,}49} \cong 1{,}48 \qquad s_{y.x} = \sqrt{1{,}48} = 1{,}22$$

$$s^2_{y.x} = s^2_{y.}\,(1 - r^2) \qquad s^2_{y.x} = 15{,}030\,(1 - 0{,}90) \cong 1{,}50$$ (Mediante la calculadora es 1,48)

$$\hat{Y} \pm z\,\frac{s_{y.x}}{\sqrt{n}} \quad \Rightarrow \quad 8{,}99 \pm 1{,}96\,\frac{1{,}22}{\sqrt{80}} = 8{,}99 \pm 0{,}27 = \begin{cases} 9{,}26 \\ 8{,}72 \end{cases}$$ Miles mill $

Algunos aplican las siguientes fórmulas para la estimación del valor del promedio respectivamente:

$$\hat{Y} \pm z\, s_{y.x} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}} \quad \text{ó} \quad \hat{Y} \pm z\, s_{y.x} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma x_t^2 - \frac{(\Sigma x_t)^2}{n}}}$$

Recuerde que = 24 en el ejercicio que se acaba de desarrollar. Nos falta calcular el coeficiente de correlación de *Pearson* siendo:

$$r = \frac{n\,\Sigma x_i\, y_i\, n_i - (\Sigma x_i\, n_i)\,(\Sigma y_i\, n_i)}{\sqrt{\left[n\Sigma x_t^2 n_t - (\Sigma x_t\, n_t)^2\right]\left[n\Sigma y_t^2 n_t - (\Sigma y_t\, n_t)^2\right]}}$$ También se puede obtener mediante

$$r = \frac{\Sigma x_i\, y_i\, n_i - n\bar{x}\bar{y}}{\sqrt{\left[\Sigma x_t^2 n_t - n\bar{x}^2\right]\left[\Sigma y_t^2 n_t - n\bar{y}^2\right]}}$$

$$r = \frac{15.476 - 80\,(21{,}075)\,(7{,}7125)}{\sqrt{\left[41.172 - 80(21{,}075)^2\right]\left[5.961 - 80(7{,}7125)^2\right]}} \cong 0{,}95 \quad \text{y} \quad R^2 \cong 0{,}9$$

Ejemplo 2. Cuando el número par de observaciones es grande, parece aconsejable su agrupación mediante tablas de frecuencias o cuadros de doble entrada. Esta última forma de presentación, permite observar que el uso de momentos bidimensionales, por la simbología empleada, antes de simplificar la tarea del alumno, dificulta el cálculo de los estimativos de X y de Y.

Hagamos un ejercicio con 50 pares de observaciones:

$_i$:	8	2	4	6	8	6	6	6	8	6	4	2	6	6	8	6	2
$_i$:	2	0	1	1	3	2	3	1	0	3	1	2	2	1	3	3	0

$_i$:	6	4	6	2	6	8	6	6	4	2	8	4	8	6	8	6	2
$_i$:	2	0	0	2	3	2	2	2	0	2	3	1	1	2	2	3	0

$_i$:	4	6	4	6	4	2	4	6	4	4	8	4	2	4	6	6
$_i$:	2	0	0	3	1	2	1	2	2	0	1	2	2	1	2	1

En la determinación de los valores que toma la variable, discreta o continua, se sigue el mismo procedimiento que en la unidimensional. En este ejercicio toma valores de: 2, 4, 6 y 8; la variable de: 0, 1, 2 y 3, que pueden ser colocados, indistintamente, en las casillas de la parte inferior de la tabla o en las casillas del vertical.

Tabla 12. 3 Cuadro en doble entrada

3	–	–	6	3
2	5	3	8	3
1	–	6	4	2
0	3	4	2	1
$_i$ / $_i$	2	4	6	8

Tabla 12. 4 Datos y cálculo con frecuencia

$_i$	$_i$	n_i	$_i n_i$	$_i n_i$	$_i^2 n_i$	$_i^2 n_i$	$_i \, _i n_i$
2	0	3	6	0	12	0	0
2	2	5	10	10	20	20	20
4	0	4	16	0	64	0	0
4	1	6	24	6	96	6	24
4	2	3	12	6	48	12	24
6	0	2	12	0	72	0	0
6	1	4	24	4	144	4	24
6	2	8	48	16	288	32	96
6	3	6	36	18	216	54	108
8	0	1	8	0	64	0	0
8	1	2	16	2	128	2	16
8	2	3	24	6	192	12	48
8	3	3	24	9	192	27	72
Σ		50	260	77	1.536	169	432
$_i$	$_i$	n	$_i f_i$	$_i n_i$	$_i^2 f_i$	$_i^2 f_i$	$_i \, _i f_i$

$$\bar{x} = \frac{\Sigma x_t n_t}{n} = \frac{260}{50} = 5{,}2$$

$$\bar{y} = \frac{\Sigma y_t n_t}{n} = \frac{77}{50} = 1{,}54$$

Se puede estimar $\hat{Y}$ $\hat{X}$ como en a recta uti i ando a simbo ogía que aparece a continuación. Si queremos simplificar las operaciones podemos cambiar de posición por para ue siempre estimemos a $\hat{Y}$.

$$s_x^2 = \frac{\Sigma x_t^2 n_t}{n} - \bar{x}^2 = \frac{1.536}{50} - 5{,}2^2 = 30{,}72 - 27{,}04 = 3{,}68$$

$$s_y^2 = \frac{\Sigma y_t^2 n_t}{n} - \bar{y}^2 = \frac{169}{50} - 1{,}54^2 = 3{,}38 - 2{,}37 = 1{,}01 \qquad (1{,}0084)^*$$

$$m_{xy} = \frac{\Sigma x_t y_t n_t}{n} - \bar{x}\,\bar{y} = \frac{432}{50} - 5{,}2(1{,}54) = 8{,}64 - 8{,}01 = 0{,}63 \qquad (0{,}632)^*$$

* Son valores que arroja la calculadora.

Para el cálculo de los parámetros de las rectas de regresión se procede como en la regresión simple:

$$b_{yx} = \frac{m_{xy}}{s_x^2} = \frac{0{,}63}{3{,}68} = 0{,}17 \qquad (0{,}171739)^*$$

$$b_{xy} = \frac{m_{xy}}{s_y^2} = \frac{0{,}63}{1{,}01} = 0{,}62 \qquad (0{,}626735)^*$$

$$c_{yx} = \bar{y} - b_{yx}\bar{x} = 1{,}54 - (0{,}17)\,(5{,}2) = 1{,}54 - 0{,}88 = 0{,}66 \qquad (0{,}6469565)^*$$

$$c_{xy} = \bar{x} - b_{xy}\bar{y} = 5{,}2 - (0{,}62)\,(1{,}54) = 5{,}2 - 0{,}95 = 4{,}25$$

* la cifra aparece con más decimales cuando usamos la calculadora.

Las ecuaciones serán: $\hat{Y} = 0{,}17 \quad + 0{,}66$ y $\hat{X} = 0{,}62 \quad + 4{,}25$

Para el cálculo de la varianza residual se pueden aplicar las fórmulas siguientes:

1) $$S_{yx}^2 = \frac{\Sigma (y_t - \hat{Y}_t)^2 n_t}{n} \qquad S_{xy}^2 = \frac{\Sigma (x_t - \hat{X}_t)^2 n_t}{n}$$

2) $$S_{yx}^2 = \frac{\Sigma y_t^2 n_t - c_{yx} \Sigma y_t n_t - b_{yx} \Sigma x_t y_t n_t}{n} \qquad S_{xy}^2 = \frac{\Sigma x_t^2 n_t - c_{xy} \Sigma x_t n_t - b_{xy} \Sigma x_t y_t n_t}{n}$$

3) $$S_{yx}^2 = S_y^2 - \frac{m_{xy}^2}{s_x^2} \qquad S_{xy}^2 = S_x^2 - \frac{m_{xy}^2}{s_y^2}$$

Aplicaremos las fórmulas (2) y (3) para el cálculo de la varianza residual de la recta de regresión de *X* en función de *Y*.

$$(2)\ S^2_{xy} = \frac{1.536 - 4{,}25\ (260) - 0{,}62\ (432)}{50} = 3{,}26$$

$$(3)\ S^2_{xy} = 3{,}68 - \frac{0{,}63^2}{1{,}01} = 3{,}68 - 0{,}40 = 3{,}28$$

En la aplicación de las anteriores fórmulas se presentan pequeñas diferencias, debido a las aproximaciones. Con la varianza residual, se puede calcular la desviación residual, los límites de confianza y el coeficiente de correlación.

$$S_{xy} = + \sqrt{S^2_{xy}} \qquad S_{xy} = \sqrt{3{,}26} = 1{,}81$$

Los límites de confianza con una P = 95,5% para $\hat{X}$ cuando toma el valor de 4.

$$\hat{X} = b_{xy}\ (y - \bar{y}) + \bar{x}$$

$$\hat{X} = 0{,}62\ (4 - 1{,}54) + 5{,}2 = 0{,}62\ (2{,}46) + 5{,}2 = 1.53 + 5{,}2 = 6{,}73$$

$$\hat{X} = 6{,}73 \pm 2(1{,}81) \sqrt{\frac{1}{50} + \frac{(4 - 1{,}54)^2}{1.536 - \frac{(260)^2}{50}}} = 6{,}73 \pm 0{,}83 = \begin{cases} 7{,}56 \\ 5{,}90 \end{cases}$$

El coeficiente de correlación lineal se podrá calcular por diferentes fórmulas.

$$(1)\ r = \sqrt{b_{yx}\ b_{xy}} \qquad (2)\ r = \frac{m_{xy}}{S_x\ S_y} \qquad (3)\ r = \sqrt{1 - \frac{S^2_{yx}}{S^2_y}} = \sqrt{1 - \frac{S^2_{xy}}{S^2_x}}$$

Aplicaremos la fórmula (3).

Con la fórmula (2) dará aproximadamente: $r = \sqrt{1 - \frac{3{,}28}{3{,}68}} = \sqrt{1 - 0{,}89} = 0{,}33$

Otra forma de obtener los parámetros **b** y **c**, trabajando con un sistema de ecuaciones, es como sigue:

Para $\hat{Y} = b + c$ las ecuaciones son:

$$(1)\ \Sigma y_t\, n_t = b_{yx}\ \Sigma x_t\, n_t + c_{yx}\ \Sigma n_t$$
$$(2)\ \Sigma x_i\, y_i\, n_i = b_{yx}\ \Sigma x_i^2\, n_i + c_{yx}\ \Sigma x_i\, n_i$$

Para $\hat{X} = b + c$ las ecuaciones son:

$$(1)\ \Sigma x_t\, n_t = b_{xy}\ \Sigma y_t\, n_t + c_{xy}\ \Sigma n_t$$
$$(2)\ \Sigma x_i\, y_i\, n_i = b_{xy}\ \Sigma y_i^2\, n_i + c_{xy}\ \Sigma y_i\, n_i$$

Obtengamos los valores de b y c

(1) $260 = 77b + 50\ c$ Anulamos a c para ello multiplicamos la primera ecuación por - 1,54

(2) $432 = 169b + 77c$

$$\begin{array}{r} 432{,}0 = 169{,}00 b_{xy.} + 77 c_{xy.} \\ -400{,}4 = 118{,}58 b_{xy.} - 77 c_{xy.} \\ \hline 31{,}6 = 50{,}42\ b_{xy.} \quad 0 \end{array} \Rightarrow \quad b_{xy.} = \frac{31{,}6}{50{,}42} = 0{,}63$$ Reemplazamos en le primera ecuación

$$(1)\ 260 = 77\,(0{,}63) + 50 c_{xy.} \quad \Rightarrow \quad c_{xy.} = \frac{260 - 48{,}51}{50} = 4{,}23$$ aproximadamente

USO DE LA CALCULADORA

Todos estos cálculos pueden ser agilizados mediante el uso de calculadoras que muestren en pantalla los siguientes comandos, oprimiendo MODE, y luego la sigla *R*, o *RE* , símbolos equivalentes a *regresión inea* .

El proceso de entrada de los datos se hace de la siguiente manera [2] [,] [0] [;] [3] [DT] en la pantalla debe aparecer así: [2 , 0 ; 3] [2 , 2 ; 5] [4 , 0 ; 4] y así sucesivamente, la salida de los datos es exactamente igual como se hizo en el ejercicio anterior, salvo que $\sum_{\ i}$ corresponde a $\sum_{\ i} n_i$, lo mismo en las demás operaciones, todas ellas multiplicadas por n_i *o* f_i

Ejemplo 3. De una población de 1.200 parejas que han contraído matrimonio se extraen, al azar, 50 parejas de edades, $_i$ simboliza las edades de los hombres y $_i$ las mujeres. Los datos son:

Tabla 12. 5 Cálculo para aparejas de observaciones moderadas

$_i$	$_i$	n_i	$_i n_i$	$_i n_i$	$_i^2 n_i$	$_i^2 n_i$	$_i\ _i n_i$	$\hat{Y}_i$
29	23	6	174	138	5.046	3.174	4.002	25,86
35	34	14	490	476	17.150	16.184	16.660	30,95
29	26	20	580	520	16.820	13.520	15.080	25,86
38	30	8	304	240	11.552	7.200	9.120	33,50
40	35	2	80	70	3.200	2.450	2.800	35,20
Σ		50	1.628	1.444	53.768	42.528	47.662	–
$_i$	$_i$	f_i	$_i f_i$	$_i f_i$	$_i^2 f_i$	$_i^2 f_i$	$_i\ _i f_i$	–

Calculemos los coeficientes angulares (b) y de posición (c) cuando se va a estimar $\hat{Y}$.

(1) $\Sigma y_i x_i = b\Sigma x_i n_i + nc$ (1) $1.444 = 1.628b + 50c$

(2) $\Sigma x_t y_t\ n_t = b\Sigma x_t^2 n_t + c\Sigma x_t n_t$ (2) $47.662 = 53.768b + 1.628c$

Multiplicamos a la ecuación (1) por 32,56 y se lo restamos a la ecuación (2)

$$\begin{array}{r} (1) \quad 47.662 = 53.768b + 1.628c \\ (2) - 47.016{,}64 = -53.007{,}68b - 1.628c \\ \hline 645{,}36 = 760{,}32b \end{array} \Rightarrow \quad b = \frac{645{,}36}{760{,}32} = 0{,}8488 \cong 0{,}85$$

Reemplazamos a b en la ecuación (1)

$$(1)\ 1{,}444 = 1.628\ (0{,}8488) + 50c \quad \Rightarrow \quad c = \frac{1.444 - 1.628\,(0{,}8488)}{50} = 1{,}243$$

También se hubiesen podido obtener estos *coeficientes* (b y c) mediante la aplicación de otras fórmulas. Veámoslas

$$\bar{y} = \frac{\Sigma y_t n_t}{n} \qquad \bar{y} = \frac{1.444}{50} = 28{,}88 \qquad \bar{x} = \frac{\Sigma x_t n_t}{n} \qquad \bar{x} = \frac{1.628}{50} = 32{,}56$$

$$S_{y.}^2 = \frac{\Sigma y_t^2 n_t - n\bar{y}^2}{n} \qquad S_{y.}^2 = \frac{42.528 - 50(28{,}88)^2}{50} = 16{,}51$$

(Corregida $s_{y.}^2 = 16{,}84$)

$$S_x^2 = \frac{\Sigma x_t^2 n_t - n\bar{x}^2}{n} \qquad S_x^2 = \frac{53.768 - 50(32{,}56)^2}{50} = 15{,}21$$

(Corregida $S_x^2 = 15{,}52$)

$$m_{xy.} = \frac{\Sigma x_t y_t n_t}{n} - \bar{x}\,\bar{y} \qquad m_{xy.} = \frac{47.662}{50} - 28{,}88(32{,}56) = 12{,}91$$

$$b = \frac{n\Sigma x_i y_i n_i - (\Sigma x_i n_i)(\Sigma y_i n_i)}{n\Sigma x_i^2 n_i - (\Sigma x_i n_i)^2} \qquad b = \frac{50(47.662) - 1.628(1.444)}{50(53.768) - (1.628)^2} = 0{,}85$$

$$b = \frac{\Sigma x_i y_i n_i - n\bar{x}\,\bar{y}}{\Sigma x_i^2 n_i - n\bar{x}^2} \qquad b = \frac{47.662 - 50(32{,}56)(28{,}88)}{53.768 - 50(32{,}56)^2} = 0{,}85$$

$$b = \frac{m_{xy.}}{s_x^2} \qquad b = \frac{12{,}91}{15{,}21} = 0{,}85$$

$$c = \bar{y} - b\bar{x} \qquad c = 28{,}88 - 0{,}85(32{,}56) = 1{,}21$$

$$c = \frac{\Sigma y_t n_t - b\Sigma x_t n_t}{n} \qquad c = \frac{1.444 - 0{,}85(1.628)}{50} = 1{,}21$$

Si le damos a el valor de 25 se tendrá que $\hat{Y}$ será igual a:

$\hat{Y} = 0{,}85 \quad +1{,}21$ $\qquad$ $\hat{Y} = 0{,}85\ (\quad - 32{,}56) + 28{,}88$

$\hat{Y} = 0{,}85\ (25) + 1{,}21 = 22{,}46$ $\qquad$ ó $\qquad$ $\hat{Y} = 0{,}85\ (25 - 32{,}56) + 28{,}88 = 22{,}46$

La diferencia que se puede presentar en estos resultados, se debe a las aproximaciones que se han hecho al trabajar con dos decimales.

En cuanto a la *arian a residua* , se obtendrá aplicando las siguientes fórmulas:

$$S^2_{y.x} = \frac{\Sigma (y_t - \hat{Y})^2 n_t}{n} \qquad S^2_{y.x} = \frac{277{,}50}{50} = 5{,}55$$

$$S^2_{y.x} = S^2_{y.}(1 - R^2) \qquad S^2_{y.x} = 16{,}51\,(1 - 0{,}66) = 5{,}61$$

$$S^2_{y.x} = \frac{\Sigma y_t^2 n_t - c\Sigma y_t n_t - b\Sigma y_t x_t n_t}{n} \qquad S^2_{y.x} = \frac{42.528 - 1{,}21(1.444) - 0{,}85(47.662)}{50} = 5{,}36$$

$$S^2_{y.x} = S^2_{y.} - \frac{m^2_{xy.}}{S^2_x} \qquad S^2_{y.x} = 16{,}51 - \frac{12{,}91^2}{15{,}21} = 5{,}55$$

$$s^2_{y.x} = s^2_{y.} - 2b\ m_{xy.} + b^2 S^2_x \qquad S^2_{y.x} = 16{,}51 - 2(0{,}85)(12{,}91) + 0{,}85^2(15{,}21) = 5{,}55$$

El *error estándar de estimación* será igual: $S_{y.x} = \sqrt{S^2_{y.x}}$ $\qquad S_{y.x} = \sqrt{5{,}55} \cong 2{,}36$

En cuanto al *coeficiente de correlación*, al igual que en la regresión lineal se tendrá que:

$$R^2 = \frac{m^2_{xy.}}{S^2_x\ S^2_{y.}} \qquad R^2 = \frac{12{,}91^2}{(15{,}21)(16{,}51)} = 0{,}66$$

$$R^2 = 1 - \frac{S^2_{y.x}}{S^2_{y.}} \qquad R^2 = 1 - \frac{5{,}55}{16{,}51} = 0{,}66$$

$$R^2 = \frac{\left[n\Sigma x_i\, y_i\, n_i - (\Sigma x_i\, n_i)\,(\Sigma y_i\, n_i)\right]^2}{\left[n(\Sigma x_t^2\, n_t) - (\Sigma x_t\, n_t)^2\right]\left[n\Sigma y_t^2\, n_t - (\Sigma y_t\, n_t)^2\right]}$$

$$R^2 = \frac{\left[50(47.662) - 1.628(1.444)\right]^2}{\left[50(53.768) - (1.628)^2\right]\left[50(42.528) - (1.444)^2\right]} = 0{,}66 \qquad R = \sqrt{0{,}66} = 0{,}81$$

Los límites del 95% de confianza para $\hat{Y}$, cuando $\quad = 25$ como promedio.

$$\hat{Y} = \hat{y} \pm Z\ S_{y.x} \sqrt{\frac{1}{n} + \frac{(x_t - \bar{x})^2}{\Sigma x_t^2 n_t - \frac{(\Sigma x_t\, n_t)^2}{n}}}$$

$$\hat{Y} = 22{,}46\ \pm\ 1{,}96\ (2{,}12) \sqrt{\frac{1}{50} + \frac{(25 - 32{,}56)^2}{53.768 - \frac{(1.628)^2}{50}}} = 22{,}46 \pm 1{,}28\ =\ \begin{cases} 23{,}74 \\ 21{,}18 \end{cases}$$

Se observa, tal y como lo mencionamos en la regresión líneal símple, que si el tamaño de la muestra se hace grande, el error será menor.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

61. Con los siguientes datos:

$_i$	2	2	4	4	6	6	8	8	8
$_i$	1	2	1	3	2	3	1	2	3
n_i	3	2	5	3	5	5	1	3	3

Se pide:

a) Estimar el valor de $\hat{X}$ cuando = 14.
b) Determinar el coeficiente de correlación.
c) Calcular las respectivas varianzas residual y explicada.
d) Fijar límites de confianza del 95% para la estimación obtenida en el punto (a).

62. Utilizar los siguientes datos sobre salarios semanales ( ) y gastos semanales ( ) (ambos en miles de $) de 30 familias.

SALARIO SEMANAL	GASTO SEMANAL			
	35	40	60	70
120	–	1	4	5
100	–	6	2	3
80	2	2	1	–
60	4	–	–	–

Se pide:

a) Determinar el coeficiente de correlación.
b) Estimar el salario semanal de una familia que tiene gastos de $50.000 semanales.
c) Fijar límites de confianza del 95% para la estimación del ingreso.
d) Calcular la varianza explicada y el porcentaje de la varianza total que queda sin explicar por la recta de regresión.

63. Una fábrica clasifica a sus empleados, 30 en total, de acuerdo con los años de antigüedad y edad, así:

AÑOS ANTIG EDAD	EDAD		
	20 1-28	28 1-36	36 1-44
2	3	2	1
4	4	5	3
6	1	3	5
8	-	1	2

a) Determinar cuál debe ser la edad del empleado con 5 años de antigüedad en la fábrica.
b) Determine el coeficiente de correlación al cuadrado.

64. Los siguientes datos del semestre pasado, representan el número de días de incapacidad por enfermedad o accidente de trabajo (*Y*) y los años de antigüedad en la compañía (*X*). Se tomó una muestra de 30 empleados que fueron incapacitados.

$_i$:	7	8	2	3	5	3	7
$_i$:	2	1	5	6	4	9	2
n_i:	2	3	6	5	4	7	3

a) De acuerdo a la información dada, se desea estimar el número de días de incapacidad para un trabajador con una antigüedad de 6 años.
b) Además determine el coeficiente de correlación.

65. Con los siguientes datos $n = 50$ $\Sigma_i n_i = 374$ $\Sigma_i^2 n_i = 4.410$
$\Sigma_i^2 n_i = 19.736$ $\Sigma_i n_i = 9.282$ $\Sigma_i n_i = 800$

Con esta información se pide estimar $\hat{Y}$, cuando = 30, fijando límites de confianza del 95%.

66. Con el punto anterior calcular el coeficiente de correlación de *Pearson*

67. Con los siguientes datos:

$_i$:	3	5	10	16	26	
$_i$:	1	2	4	8	12	
n_i:	3	7	10	8	2	= 30

Se pide calcular:

a) $\hat{Y}$ cuando = 14; b) El coeficiente de determinación

REGRESIÓN PARABÓLICA SIMPLE

Anteriormente se explicó el concepto y la aplicación que tiene la *regresión inea* (simple y ponderada), además de anotar la importancia al ser usada para señalar la tendencia en las series de tiempo, que será desarrollada en el capítulo siguiente; la *regresión parabó ica* también es considerada de gran utilidad en el estudio de muchos fenómenos que explicamos en forma sencilla, permitiendo ser comprendida por la mayoría de las personas, pues sigue siendo una función fácil de aplicar.

Ahora nos dedicaremos a desarrollar y explicar la llamada *regresión parabó ica simp e* aplicada cuando aquellos fenómenos que se observan en un *diagrama de dispersión*, presentan una concentración de puntos inicialmente ascendente y en seguida descendente (puede darse lo contrario). Esta regresión parabólica es utilizada, en gran parte, por los economistas en el análisis de funciones de utilidad, ingresos, etc.

La ecuación general de la parabólica es $\hat{Y} = a^{2} + b + c$

COSTO (miles mill.)	PRODUCCIÓN (miles mill.)
23	32
36	48
48	57

ontin a

Siguiendo el procedimiento de los mínimos cuadrados, calcularemos *a*, *b* y *c* para los siguientes 10 pares de observaciones de datos tomados en forma arbitraria, pues nos interesa mostrar los procedimientos, fórmulas y cálculos que se deben realizar en su aplicación.

ontinuación

56	70
70	86
38	46
30	40
26	36
20	32
10	26
$_i$	$_i$

$_i$	^2_i	^3_i	^4_i	$_i$	^2_i	$_i\ _i$	$^2_{i\ i}$
23	529	12.167	279.841	32	1.024	736	16.298
36	1.296	46.656	1.679.616	48	2.304	1.728	62.208
48	2.304	110.592	5.308.416	57	3.249	2.736	131.328
56	3.136	175.616	9.834.496	70	4.900	3.920	219.520
70	4.900	343.000	24.010.000	86	7.396	6.020	421.400
38	1.444	54.872	2.085.136	46	2.116	1.748	66.424
30	900	27.000	810.000	40	1.600	1.200	36.000
26	676	17.576	456.976	36	1.296	936	24.336
20	400	8.000	160.000	32	1.024	640	12.800
10	100	1.000	10.000	26	676	260	2.600
357	15.685	746.479	44.634.481	473	25.585	19.924	992.914

Las ecuaciones que serán consideradas para hallar: a, b, c y los respectivos reemplazos en las ecuaciones, se dan a continuación:

Ecuaciones normales

$$(1)\quad \Sigma y_t = a\Sigma x_t^2 + b\Sigma x_t + nc$$
$$(2)\quad \Sigma x_t y_t = a\Sigma x^3 + b\Sigma x_t^2 + c\Sigma x_t$$
$$(3)\quad \Sigma x_t^2 y_t = a\Sigma x^4 + b\Sigma x_t^3 + c\Sigma x_t^2$$

Reemplazando se tiene:

$$(1)\quad 473 = 15.685a + 357b + 10c$$
$$(2)\quad 19.924 = 746.479a + 15.685b + 357c$$
$$(3)\quad 992.914 = 44.634.481a + 746.479b + 15.685c$$

Comenzamos por eliminar a c para ello trabajamos con las ecuaciones (1) y (2) multiplicando la ecuación (1) por -3,57 y la resultante se la restamos a la (2).

$$(2)\quad 19.924 = 746.479a + 15.685b + 357c$$
$$(1)\quad -16.886,1 = -559.954,5a - 12.744,9b - 357c$$
$$(4)\quad 3.037,9 = 186.524,5a + 2.940,1b \quad 0$$

Mediante la operación anterior obtuvimos la ecuación señalada como (4).

Ahora, procedemos con las ecuaciones (1) y (3), multiplicando la primera por 1.568,5 y se la restamos a la (3).

$$
\begin{array}{lrrrr}
(3) & 992.914 & = & 44.634.481a & 746.479b & + 15.685c \\
(1) & -741.900,5 & = & -24.601.922,5a & - 559.954,5b & - 15.685c \\
\hline
(5) & 251.013,5 & = & 20.032.558,5a & 186.524,5b & 0
\end{array}
$$

Se obtuvo la ecuación (5). Procedemos a eliminar b para ello multiplicamos a la ecuación (4) por 63,4415496 y se la restamos a la ecuación (5).

$$
\begin{array}{lrrrr}
(5) & 251.013,50 = & 20.032.558,50\,a & +186.524,5b \\
(4) & -192.729,08 = & 11.833.403,32\,a & -186.524,5b \\
\hline
 & 58.284,42 = & 8.199.155,18\,a & 0
\end{array}
$$

$$a = \frac{58.284,42}{8.199.155,18} = 0,0071 \qquad \boxed{a = 0,0071}$$

Reemplazamos en la ecuación (4) y despejamos b

$$(4)\ 3.037,9 = 186.524,5\ (0,0071) + 2.940,1b$$

$$b = \frac{3.037,9 - 186.524,5(0,0071)}{2.940,1} = 0,5828 \qquad \boxed{b = 0,5828}$$

Reemplazamos los valores de a y b en la ecuación (1) para obtener el valor de c

$$(1)\ 473 = 15.685(0,0071) + 357\ (0,5828) + 10c$$

$$c = \frac{473 - 15.685(0,0071) - 357(0,5828)}{10} = 15,3577 \qquad \boxed{c = 15,3577}$$

La ecuación general nos queda:

$$\boxed{\hat{Y} = ax^2 + bx + c} \Rightarrow \hat{Y} = 0,0071x^2 + 0,5828x + 15,3577$$

Ahora, si queremos estimar el valor de la producción ( ) en función del costo, considerando a = 20 miles de millones de pesos se tendrá:

$\hat{Y} = 0,0071\ (20^2) + 0,5828\ (20) + 15,3577 = 29,86$ millones de pesos, será el valor estimado para esa empresa.

El resultado, se le denomina *estimador puntua* , pero lo más aconsejable es calcular el *estimador por inter a os*, fijando un nivel de seguridad o confianza, por ejemplo del 95%, lo cual requiere en primer lugar calcular la *arian a residua* y luego la *des iación típica residua* veámoslo:

$\boxed{S^2_{y.x} = \frac{\Sigma(y_t - \hat{Y})^2_t}{n}}$ es la fórmula más utilizada y por ende la más engorrosa para calcular, de ahí que procedamos a utilizar otras fórmulas que sean más rápidas y simplificadas.

$\boxed{S^2_{y.x} = \frac{\Sigma y_t^2 - c\Sigma y_t - b\Sigma y_t x_t - a\Sigma y_t x_t^2}{n}}$ reemplazando se tiene:

$$S^2_{y.x} = \frac{25.585 - 15,3577\,(473) - 0,5828(19.924) - 0,0071(992.914)}{10} = 34,06$$

$$S_{y.x} = \sqrt{S^2_{y.x}} \quad \Rightarrow \quad S_{y.x} = \sqrt{34{,}06} = 5{,}84$$

Nos quedan por calcular los límites con una confianza del 95%

$$\hat{Y} \pm t \frac{S_{y.x}}{\sqrt{n}} \Rightarrow 31{,}76 \pm 2{,}262 \frac{5{,}84}{\sqrt{10}} = 31{,}76 \pm 4{,}18 = \begin{cases} 35{,}94 \\ 27{,}58 \end{cases}$$

$$\left.\begin{aligned} \upsilon = n-1 = 10-1 = 9 \\ \propto = 0{,}05 \end{aligned}\right\} \; t_{0{,}05} = 2{,}262$$

Se observan límites bastante exagerados, quizás por ser *n* tan pequeña; para estar más seguros sobre el uso o no de la regresión parabólica, calculamos el coeficiente de correlación:

$$R^2 = 1 - \frac{VR}{VT} \Rightarrow R^2 = 1 - \frac{S^2_{y.x}}{S^2_y} \quad \Rightarrow \quad R^2 = 1 - \frac{34{,}06}{321{,}21} = 0{,}89$$

$$S^2_y = VT = \frac{\Sigma y_t^2 - n\bar{y}^2}{n} \Rightarrow S^2_y = \frac{25.585 - 10\left(\frac{473}{10}\right)^2}{10} = 321{,}21$$

Por lo tanto $r = R = \sqrt{R^2} \Rightarrow r = \sqrt{0{,}89} = 0{,}94$ Considerado regular la relación.

Ejemplo 1. A pesar de que en el ejercicio de las 5 parejas, en las que se analiza las edades de quienes contrajeron matrimonio, no corresponde a un ajuste parabólico, de acuerdo con la observación en el diagrama de dispersión o nube de puntos, será utilizado como ejemplo, únicamente con el fin de comparar los resultados que se pueden obtener mediante las diferentes líneas o ajustes que frecuentemente hacemos:

Tabla 12.6 Datos y cálculos en ajuste parabólicos

i	i	^2_i	^3_i	^4_i	$i\ i$	$^2_i\ i$	$\hat{Y}_i$
29	23	841	24.389	707.281	667	19.343	24,65
35	34	1.225	42.875	1.500.625	1.190	41.650	32,00
29	26	841	24.389	707.281	754	21.866	24,65
38	30	1.444	54.872	2.085.136	1.140	43.320	33,34
40	35	1.600	64.000	2.560.000	1.400	56.000	33,36
171	148	5.951	210.525	7.560.323	5.151	182.179	148,00

Calculamos *a*, *b* y *c* utilizando el método de los mínimos cuadrados

(1) $148 = 5.951a + 171b + 5c$

(2) $5.151 = 210.525a + 5.951b + 171c$

(3) $182.179 = 7.560.323a + 210.525b + 5.951c$

Eliminamos a *c*, trabajando con las ecuaciones (1) y (2), para ello, multiplicamos a la (1) por -34,2 y restamos a la ecuación (2).

$$(2) \quad 5.115,0 = \quad 210.525,0a + 5.951,0b + 171c$$
$$(1) \quad -5.061,6 = -203.524,2a - 5.848,2b - 171c$$
$$(4) \qquad 89,4 = \quad 7.000,8a + 102,8b$$

Ahora se procede con las ecuaciones (1) y (3), para ello eliminamos a c, multiplicando la ecuación (1) por -1.190,2 y se la restamos a la (3)

$$(3) \ 182.179,0 = 7.560.323,0a + 210.525,0b + 5.951c$$
$$(1) \ 176.149,6 = 7.082.880,2a - 203.524,2b - 5.951c$$
$$(5) \quad 6.029,4 = \quad 477.442,8a + \quad 7.000,8b$$

Ahora procedemos a multiplicar a la ecuación (4) por 7.000,8 y a la ecuación (5) por -102,8 y así eliminamos al coeficiente (b) y obtenemos el valor de a.

$$(4) \quad 625.871,52 = \quad 49.011.200,64a + 719.682,24b$$
$$(5) -619.822,32 = -49.081.119,84a - 719.682,24b$$
$$6.049,20 = \qquad -69.919,20a \qquad 0$$

$$a = -\frac{6.049,2}{69.919,2} = -0,0865 \qquad a = -0,0865$$

Reemplazamos en la ecuación (4), por lo tanto se obtendrá el valor de b.

$$(4)\ 89,4 = 7.000,8(-0,0865) + 102,8b \quad \Rightarrow \quad b = \frac{89,4 + 605,57}{102,8} = 6,76$$

Si reemplazamos en la ecuación (1) se tendrá el valor de c.

(1) $148 = 5.951\ (-0,0865) + 171(6,76) + 5\ c$ de donde c será igual a:

$$c = \frac{148 + 514,76 - 1.155,96}{5} = -98,64 \qquad c = -98,64$$

La ecuación para la parábola será: $\hat{Y} = -0,0865\ ^2 + 6,76\ - 98,64$

Cuando = 29 se tendrá con la ecuación anterior que $\hat{Y} = 24,65$

Cuando = 35 se tendrá que $\hat{Y} = 31,997$ y así sucesivamente (*er tab a No 12*)
Con los datos de la siguiente tabla se haran otros cálculos de importancia.

Tabla 12.7 Estimativos y desviaciones

$_i$	$_i$	$\hat{Y}_i$	$_i - \hat{Y}_i$	$(_i - \hat{Y}_i)^2$	$_i^2$
29	23	24,65	-1,65	2,7225	529
35	34	32,00	2,00	4,0000	1.156
29	26	24,65	1,35	1,8225	676
38	30	33,34	-3,34	11,1556	900
40	35	33,36	1,64	2,6896	1.225
171	148	148,00	0	22,3902	4.486

Varian a residual o no explicada

El procedimiento generalizado de cálculo es exactamente igual al utilizado para la recta de regresión, es decir:

$$S^2_{y.x} = \frac{\Sigma (y_t - \hat{Y}_t)^2}{n} \qquad S^2_{y.x} = \frac{22{,}3902}{5} = 4{,}48$$

$$S^2_{y.x} = \frac{\Sigma (y_i - \hat{Y}_t)^2}{n-2} \qquad S^2_{y.x} = \frac{22{,}3902}{3} = 7{,}46 \text{ (Corregida)}$$

También se puede calcular en forma directa, mediante la aplicación de la siguiente fórmula:

$$S^2_{y.x} = \frac{\Sigma y_t^2 - c\Sigma y_t - b\Sigma y_t x_t - a\Sigma y_t x_t^2}{n}$$

Reemplazando se tiene que:

$$S^2_{y.x} = \frac{4.486 - (-98{,}64)(148) - 6{,}76\ (5.151) - (-0{,}0865)(182.179)}{5}$$

$$S^2_{y.x} = \frac{22{,}4435}{5} = 4{,}49$$

ó $\qquad S^2_{y.x} = \frac{22{,}4435}{n-2} = \frac{22{,}4435}{3} = 7{,}48$ (corregida)

Error estándar de estimación

Ya se vio en la recta de regresión, que el cálculo del error estándar de estimación se obtiene mediante la raíz cuadrada a la varianza residual.

$S_{y.x} = \sqrt{4{,}49} = 2{,}12$ para $n = 5$

$S_{y.x} = \sqrt{7{,}48} = 2{,}73$ para $n - 2$

Límites de confianza para $\hat{Y}$

Cuando el estimador se calcula con base en un valor para , es decir, la variable independiente, se tendrá que:

$$\hat{Y} \pm t_{n-2}\ S_{y.x} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma x_t^2 - \frac{(\Sigma x_t)^2}{n}}}$$ así, cuando = 29, se tendrá que:

$$\hat{Y} = 24{,}65 \pm 3{,}182(2{,}73) \sqrt{1 + \frac{1}{5} + \frac{(29 - 34{,}2)^2}{5.951 - \frac{(171)^2}{5}}} = 24{,}65 \pm 10{,}5 = \begin{cases} 35{,}15 \\ 14{,}15 \end{cases}$$

Cuando el valor de se toma como promedio, se fijan los límites de confianza así:

$$\hat{Y} = 24{,}65 \pm 3{,}182\,(2{,}73)\sqrt{\frac{1}{5}+\frac{(29-34{,}2)^2}{5.951-\frac{(171)^2}{5}}} = 24{,}65 \pm 5{,}91 = \begin{cases} 30{,}56 \\ 18{,}74 \end{cases}$$

La fórmula que se aplicó corresponde a $\hat{Y} \pm t_{n-2}\, S_{y.x}\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{\Sigma x_t^2-\frac{(\Sigma x_t)^2}{n}}}$

Coeficiente de correlación parabólico (R^2)

El proceso más fácil para calcular el coeficiente de determinación parabólico o coeficiente de correlación al cuadrado será:

$$R^2 = 1-\frac{VR}{VT} \quad \Rightarrow \quad R^2 = 1-\frac{4{,}49}{26{,}3} = 1-0{,}17 = 0{,}83$$

$$S_y^2 = \frac{\Sigma\,(y_i-\bar{y})^2}{n-1} = \frac{105{,}2}{4} = 26{,}3$$

r = es el coeficiente de correlación parabólico. $r = \sqrt{0{,}83} = 0{,}91$

Veamos ahora cómo se procede a trabajar con la regresión Parabólica (simple), utilizando la simbología que nos permite en un solo ejercicio estimar el valor de $\hat{Y}$ y el de $\hat{X}$, *sin necesidad de in ertir e orden de as ariab es* procedimiento que sólo permite hacer los cálculos siempre estimando la variable que identificamos como $\hat{Y}$.

Regresión parabólica simple (empleando otra simbología)

En la regresión rectilínea, simple y ponderada, se trabaja con un sistema de ecuaciones con dos incógnitas, que corresponden a los parámetros b y c – en la regresión de "2 en 1" b y c en la regresión de "1 en 2". Para la regresión parabólica, simple y ponderada, se tendrán, tres ecuaciones con tres incógnitas: a, b y c en la regresión parabólica de "2 en 1". Las ecuaciones en la regresión parabólica simple son:

$\hat{Y} = a_{y.x}\,x^2 + b_{y.x}\,x + c_{y.x}$ (ecuación general)

Sistema de ecuaciones en la regresión parabólica simple de "2 en 1"

$$\begin{aligned}
(1)\,\Sigma y_t &= a_{y.x}\Sigma x_t^2 + b_{y.x}\Sigma x_t + nc_{y.x} \\
(2)\,\Sigma y_t x_t &= a_{y.x}\Sigma x_t^3 + b_{y.x}\Sigma x_t^2 + c_{y.x}\Sigma x_t \\
(3)\,\Sigma y_t x_t^2 &= a_{y.x}\Sigma x_t^4 + b_{y.x}\Sigma x_t^3 + c_{y.x}\Sigma x_t^2
\end{aligned}$$

En la regresión parabólica de "1 en 2", la ecuación general, es:

$\hat{X} = a_{x.y}\,y^2 + b_{x.y}\,y + c_{x.y}$

Las ecuaciones para obtener los parámetros: a b y c

$$(1)\,\Sigma x_t = a_{x.y}\Sigma y_t^2 + b_{x.y}\Sigma y_t + nc_{x.y}$$

$$(2)\ \Sigma x_t y_t = a_{xy}\Sigma y_t^3 + b_{xy}\Sigma y_t^2 + c_{xy}\Sigma y_t$$
$$(3)\ \Sigma x_t y_t^2 = a_{xy}\Sigma y_t^4 + b_{xy}\Sigma y_t^3 + c_{xy}\Sigma y_t^2$$

Consideremos un ejercicio pequeño para calcular los parámetros en la regresión parabólica de "1 en 2"

Tabla 12.8 Datos y cálculos en ajuste parabólico

i	i	i^2	i^3	i^4	$i\ i$	$i\ i^2$	i^2
2	1	1	1	1	2	2	4
6	3	9	27	81	18	54	36
8	4	16	64	256	32	128	64
10	6	36	216	1.296	60	360	100
14	6	36	216	1.296	84	504	196
40	20	98	524	2.930	196	1.048	400

Calculamos los parámetros con base en la información de la tabla anterior:

$$(1)\quad 40 = 98\,a_{xy} + 20\,b_{xy} + 5\,c_{xy}$$
$$(2)\quad 196 = 524\,a_{xy} + 98\,b_{xy} + 20\,c_{xy}$$
$$(3)\quad 1.048 = 2.930\,a_{xy} + 524\,b_{xy} + 98\,c_{xy}$$

Trabajamos con las dos primeras ecuaciones, multiplicando la primera por -4 y se le restan a la ecuación (2)

$$(2)\quad 196 = 524\,a_{xy} + 98\,b_{xy} + 20\,c_{xy}$$
$$(1)\quad -160 = -392\,a_{xy} - 80\,b_{xy} - 20\,c_{xy}$$
$$(4)\quad 36 = 132\,a_{xy} + 18\,b_{xy} \qquad 0$$

Luego trabajamos con la primera ecuación (multiplicada por -19,6) y la tercera ecuación

$$(3)\ 1.048 = 2.930\,a_{xy} + 524\,b_{xy} + 98\,c_{xy}$$
$$(1) - 784 = -1.920{,}8\,a_{xy} - 392\,b_{xy} - 98\,c_{xy}$$
$$(5)\quad 264 = 1.009{,}2\,a_{xy} + 132\,b_{xy} \qquad 0$$

Ahora procedemos con las ecuaciones (4) y (5) y multiplicamos la (4) por 132 y la (5) por 18.

$$(5)\quad 4.752 = 18.165{,}6\,a_{xy} + 2.376\,b_{xy}$$
$$(4)\quad -4.752 = -17.424\,a_{xy} - 2.376\,b_{xy}$$
$$0 = 741{,}6\,a_{xy} \qquad 0$$

$$a_{xy} = 0$$

Este resultado se debe a que los datos se ajustan a una regresión rectilínea y no a una parábola. Con la ecuación (4) determinamos el valor del parámetro *b*

$$(4)\ 36 = 132(0) + 18\,b_{xy} \quad \Rightarrow \quad 36 = 18\,b_{xy} \quad \Rightarrow \quad b_{xy} = \frac{36}{18} = 2$$

$$b_{xy} = 2$$

Ahora procedemos con la ecuación (1)

(1) $40 = 98\,(0) + 20(2) + 5c_{xy} \quad \Rightarrow \quad \dfrac{40-40}{5} = c_{xy} = 0$ $\boxed{C_{xy} = 0}$

La *ecuación genera* $\hat{X} = a \;\; {}_i^2 + b \;\; + c \;\; \Rightarrow \hat{X} = (0) \;\; {}_i^2 + 2 \;\; {}_i + 0$

Cuando $= 40,$ $\hat{X}$ es igual a 80 $\hat{X} = (0)\,(160) + 2(40) + 0 = 80.$

El cálculo de la *arian a residua* se puede obtener con la aplicación de las siguientes fórmulas:

$$S^2_{y.x} = \frac{\Sigma(y_t - \hat{Y}_t)^2}{n} \qquad S^2_{x.y} = \frac{\Sigma(x_t - \hat{X}_t)^2}{n}$$

También se puede aplicar y parece más fácil su cálculo.

$$S^2_{y.x} = \frac{\Sigma y_t^2 - c_{y.x}\,\Sigma y_t - b_{y.x}\,\Sigma x_t y_t - a_{y.x}\,\Sigma x_t^2 y_t}{n}$$

$$S^2_{x.y} = \frac{\Sigma x_t^2 - c_{x.y}\,\Sigma x_t - b_{x.y}\,\Sigma x_t y_t - a_{x.y}\,\Sigma y_t^2 x_t}{n}$$

Calculemos la varianza residual aplicando la última fórmula.

$$s^2_{x.y} = \frac{400 - (0)\,(40) - 2\,(196) - (0)\,(1.048)}{5} = \frac{400-392}{5} = \frac{8}{5} = 1{,}6$$

El *coeficiente de correlación parabólica*, se obtendrá en la siguiente forma:

$$r = \sqrt{1 - \frac{S^2_{x.y}}{S^2_x}} \qquad r = \sqrt{1 - \frac{S^2_{y.x}}{S^2_y}} \qquad r = \sqrt{1 - \frac{1{,}6}{16}} = \sqrt{1-0{,}1} = \sqrt{0{,}90} = 0{,}95$$

Siendo: $S^2_x = \dfrac{\Sigma x_t^2}{n} - \bar{x}^2 = \dfrac{400}{5} - \left(\dfrac{40}{5}\right)^2 = 16 \qquad S^2_x = 16 \;\Rightarrow\; S_x = \sqrt{16} = 4$

REGRESIÓN PARABÓLICA PONDERADA

Se presentarán las fórmulas generales de las ecuaciones de regresión parabólica y las del sistema de ecuaciones normales para el cálculo de los parámetros, la varianza residual y el coeficiente de correlación al cuadrado.

Ecuación general: $\hat{Y} = a_{y.x}\,x^2 + b_{y.x}\,x + c_{y.x}$ $\hat{X} = a_{x.y}\,y^2 + b_{x.y}\,y + c_{x.y}$

Sistema de ecuaciones para la regresión de $\hat{Y}$ en función de

$$\Sigma y_i n_i = a_{y.x}\Sigma x_i^2 n_i + b_{y.x}\Sigma x_i n_i + c_{y.x}\Sigma n_i$$

$$\Sigma x_i y_i n_i = a_{y.x}\Sigma x_i^3 n_i + b_{y.x}\Sigma x_i^2 n_i + c_{y.x}\Sigma x_i n_i$$

$$\Sigma x_i^2 y_i n_i = a_{y.x}\Sigma x_i^4 n_i + b_{y.x}\Sigma x_i^3 n_i + c_{y.x}\Sigma x_i^2 n_i$$

Las ecuaciones en la regresión de $\hat{X}$ en función de serán:

$$\Sigma x_i n_i = a_{xy} \Sigma y_i^2 n_i + b_{xy} \Sigma y_i n_i + c_{xy} \Sigma n_i$$

$$\Sigma y_i x_i n_i = a_{xy} \Sigma y_i^3 n_i + b_{xy} \Sigma y_i^2 n_i + c_{xy} \Sigma y_i n_i$$

$$\Sigma y_i^2 x_i n_i = a_{xy} \Sigma y_i^4 n_i + b_{xy} \Sigma y_i^3 n_i + c_{xy} \Sigma y_i^2 n_i$$

Las varianzas residuales se expresan así:

$$S_{yx}^2 = \frac{\Sigma y_t^2 n_t - c_{yx} \Sigma y_t n_t - b_{yx} \Sigma x_t y_t n_t - a_{yx} \Sigma x_t^2 y_t n_t}{n}$$

$$S_{xy}^2 = \frac{\Sigma x_t^2 n_t - c_{xy} \Sigma x_t n_t - b_{xy} \Sigma x_t y_t n_t - a_{xy} \Sigma y_t^2 x_t n_t}{n}$$

El coeficiente de correlación al cuadrado o coeficiente de determinación será igual a:

$$R^2 = 1 - \frac{S_{yx}^2}{S_y^2} = 1 - \frac{S_{xy}^2}{S_x^2}$$

Ejemplo 1. Con los siguientes datos, se pide determinar las ecuaciones de regresión parabólica ponderada y el coeficiente de correlación.

$_i$	$_i$	n_i
2	1	2
2	1	8
4	3	5
6	1	4
6	2	3
6	3	2
7	2	4
7	4	2
–	–	30

$_i n_i$	$_i n_i$	$_i \, _i n_i$	$_i^2 n_i$	$_i^2 n_i$
4	2	4	2	8
16	8	16	8	32
20	15	60	45	80
24	4	24	4	144
18	6	36	12	108
12	6	36	18	72
28	8	56	16	196
14	8	56	32	98
136	57	288	137	738

$_i^3 n_i$	$_i^4 n_i$	$_i \, _i^2 n_i$	$_i^3 n_i$	$_i^4 n_i$	$_i \, _i^2 n_i$
16	32	8	2	2	4
64	128	32	8	8	16
320	1.280	240	135	405	180
864	5.184	144	4	4	24
648	3.888	216	24	48	72
432	2.592	216	54	162	108
1.372	9.604	392	32	64	112
686	4.802	392	128	512	224
4.402	27.510	1.640	387	1.205	740

$$\hat{Y} = a_{yx} x^2 + b_{yx} x + c_{yx}$$

$$1)\ \Sigma y_i n_i = a_{y.x}\Sigma x_i^2 n_i + b_{y.x}\Sigma x_i n_i + n c_{y.x}$$
$$2)\ \Sigma y_i x_i n_i = a_{y.x}\Sigma x_i^3 n_i + b_{y.x}\Sigma x_i^2 n_i + c_{y.x}\Sigma x_i n_i$$
$$3)\ \Sigma y_i x_i^2 n_i = a_{y.x}\Sigma x_i^4 n_i + b_{y.x}\Sigma x_i^3 n_i + c_{y.x}\Sigma x_i^2 n_i$$

Reemplazando se tiene que:

$$1)\ 57 = 738a_{y.x} + 136b_{y.x} + 30c_{y.x}$$
$$2)\ 288 = 4.402a_{y.x} + 738b_{y.x} + 136c_{y.x}$$
$$3)\ 1.640 = 27.510a_{y.x} + 4.402b_{y.x} + 738c_{y.x}$$

$$\hat{X} = a_{xy.}y^2 + b_{xy.}y + c_{xy.}$$

$$1)\ \Sigma x_i n_i = a_{xy.}\Sigma y_i^2 n_i + b_{xy.}\Sigma y_i n_i + n c_{xy.}$$
$$2)\ \Sigma x_i y_i n_i = a_{xy.}\Sigma y_i^3 n_i + b_{xy.}\Sigma y_i^2 n_i + c_{xy.}\Sigma y_i n_i$$
$$3)\ \Sigma x_i y_i^2 n_i = a_{xy.}\Sigma y_i^4 n_i + b_{xy.}\Sigma y_i^3 n_i + c_{xy.}\Sigma y_i^2 n_i$$

Reemplazando en las ecuaciones que se tienen para estimar a $\hat{X}$, se observa:

$$1)\ 136 = 137a_{xy.} + 57b_{xy.} + 30c_{xy.}$$
$$2)\ 288 = 387a_{xy.} + 137b_{xy.} + 57c_{xy.}$$
$$3)\ 740 = 1.205a_{xy.} + 387b_{xy.} + 137c_{xy.}$$

Hagamos el proceso para hallar a y c para obtener la ecuación de $\hat{Y}$, inicialmente tomamos las ecuaciones (1) y (2), las multiplicamos por 136 y 30 respectivamente a fin de eliminar a c restando de la ecuación (2) la ecuación (1).

$$2)\quad 8.640 = 132.060a_{y.x} + 22.140b_{y.x} + 4.080c_{y.x}$$
$$1)\quad -7.752 = -100.368a_{y.x} - 18.496b_{y.x} - 4.080c_{y.x}$$
$$4)\quad 888 = 31.692a_{y.x} + 3.644b_{y.x} \quad 0$$

Ahora procedemos con las ecuaciones (1) y (3), para lo cual multiplicamos la ecuación (1) por 24,6 y luego se la restamos de la ecuación (3)

$$3)\quad 1.640,0 = 27.510,0a_{y.x} \quad 4402,0b_{y.x} \quad 738c_{y.x}$$
$$1)\quad 1.402,2 = -18.154,8a_{y.x} - 3.345,6b_{y.x} - 738c_{y.x}$$
$$237,8 = 9.355,2a_{y.x} \quad 1.056,4b_{y.x} \quad 0$$

Se procede a trabajar con las ecuaciones (4) y (5) a fin de determinar el valor para a . se multiplica la ecuación (4) por 1.056,4 y la ecuación (5) por 3.644.

$$5)\quad 866,543,2 = 34.090.348,8a_{y.x} + 3.849.521,6b_{y.x}$$
$$4)\quad -938.083,2 = -33.479.428,8a_{y.x} - 3.849.521,6b_{y.x}$$
$$-71.540 = 610.920a_{y.x} \quad 0$$

$$a_{y.x} = -0,12$$

$a_{y.x} = \dfrac{-71.540}{610.920} = -0,12$ Reemplazamos en la ecuación (4) con lo cual obtenemos que:

(4) $888 = 31.692\ (-0,12) + 3.644 b_{y.x} \quad \Rightarrow \quad b_{y.x} = \dfrac{888 + 3.803,04}{3.644} = 1,29$

Ahora reemplazamos en la ecuación (1) $b_{y.x} = 1,29$

(1) $57 = 738(-0,12) + 136(1,29) + 30c_{y.x}$

$57 = -88,56 + 175,44 + 30c_{y.x} \Rightarrow \quad c_{y.x} = \dfrac{57 + 88,56 - 175,44}{30} = -0,996$

La ecuación será $\hat{Y} = -0,12\ ^2 + 1,29\ - 0,996$ $c_{y.x} = -0,996$

En cuanto al cálculo de la varianza residual, se procederá así:

$$s^2_{y.x} = \frac{137 + 0,996(57) - 1,29(288) + 0,12(1.640)}{30} = 0,64$$

La *des iación residua* será: $S = \sqrt{0,64} = 0,80$

La *arian a de* será: $S^2 = \dfrac{137 - 30\left(\dfrac{57}{30}\right)^2}{30} = 0,96$

El coeficiente de determinación: $R^2 = 1 - \dfrac{S^2_{y.x}}{S^2_{y.}} = 1 - \dfrac{0,80}{0,96} = 0,17$

Se deja de ejercicio al lector, establecer la ecuación de regresión parabólica de $\hat{X}$ en función de , lo mismo, el cálculo del coeficiente de correlación para esa estimación. Observe que el resultado del coeficiente de correlación para la estimación de $\hat{X}$ en función de , debe ser igual al obtenido, cuando se estimó a $\hat{Y}$ en función de .

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

68. Con la siguiente información:

$\Sigma x_t = 174$	$\Sigma y_t = 340$	$\Sigma x_t^2 = 1.626$
$\Sigma x_t^3 = 16.110$	$\Sigma x_t^4 = 167.130$	$\Sigma x_i\, y_i = 3.062$
$\Sigma x_t^2\, y_t = 29.418$	$\Sigma y_t^2 = 5.904$	$n = 20$

Se pide:

a) Estimar el valor de $\hat{Y}$ cuando = 10 mediante la regresión parabólica

b) Determinar el coeficiente de correlación parabólica y compararlo con el de la recta.

69. Un fabricante quiere establecer si hay una relación parabólica entre las ausencias al trabajo (X) (número de permisos al mes) y la edad del trabajador (Y). Para ello selecciona una muestra de 10 trabajadores, con la siguiente información:

$_i$:	28	32	46	24	28	36	42	37	51	42
$_i$:	5	8	4	7	10	4	3	4	3	4

a) Estimar la edad para una persona que solicita 6 permisos en el mes.

b) Calcular el coeficiente de correlación parabólica.

c) Fije los límites de confianza del 95%.

70. En un estudio sobre transporte urbano se quiere predecir, mediante un ajuste parabólico, la capacidad que debe tener un vehículo de servicio (X) de acuerdo al número de pasajeros que en determinado momento debe llevar (Y). Se tomaron 10 vehículos de diferente capacidad de pasajeros.

$_i$:	12	20	12	20	40	12	20	40	12	20
$_i$:	8	12	5	10	14	8	11	16	6	5

a) Efectuar la estimación de la capacidad de un vehículo que lleva 13 pasajeros.

b) Calcular el coeficiente de correlación parabólico.

REGRESIÓN Y CORRELACIÓN EXPONENCIAL Y LOGARÍTMICA SIMPLE

Cuando las variables estudiadas presentan un crecimiento o decrecimiento *aritmético*, *a regresión inea* es la más adecuada, pero si hay un crecimiento o decrecimiento geométrico, se debe adoptar la *regresión e ponencia* .

La *función e ponencia* $\hat{Y} = cb$ se puede convertir en una función lineal cuando trabajamos con logaritmos, ya sean neperianos o con base 10, dando como resultado una función logarítmica.

$$\log \hat{Y} = \log c + (\ ) \log b$$

Para la representación gráfica se debe utilizar *pape semi ogarítmico* cuando la variable *X*, localizada en el eje horizontal o abscisa, se presente en forma de progresión aritmética, mientras que en la ordenada, donde se ubica la variable *Y*, se expresa en forma logarítmica. Si ambas variables tienen crecimiento geométrico, la representación gráfica se hace en *pape ogarítmico*.

Veamos el cálculo y la aplicación de este proceso de análisis de *Regresión e ponencia* , en un ejercicio con datos arbitrarios, así:

Tabla 12.9 Datos y cálculos del ajuste exponencial

X_i	Y_i	X_i^2	$\log Y_i$	$X_i \log Y_i$	$(\log Y_i)^2$	$\log \hat{Y}_i$	$\hat{Y}_i$
2	3	4	0,47712	0,95424	0,22764	0,60768	4,05
4	6	16	0,77815	3,11261	0,60552	0,84063	6,93
5	12	25	1,07918	5,39591	1,16463	0,95710	9,06
7	24	49	1,38021	9,66147	1,90498	1,19005	15,50
12	45	124	1,65321	19,83855	2,73312	1,77241	59,21
30	90	238	5,36787	38,96278	6,63588	5,36787	94,75

Nota: se debe observar que $\Sigma Y_i \neq \Sigma \hat{Y}_i$, por el contrario $\Sigma \log Y_i = \Sigma \log \hat{Y}_i$

Las ecuaciones normales son:

$$(1)\ \Sigma \log Y_i = n \log c + (\Sigma X_i) \log b$$
$$(2)\ \Sigma X_i \log Y_i = (\Sigma X_i) \log c + (\Sigma X_i^2) \log b$$

Reemplazando se tiene que:

$(1)\ 5{,}36787 = 5 \log c + 30 \log b$

$(2)\ 38{,}96278 = 30 \log c + 238 \log b$

Eliminamos el término **log c**, multiplicando la primera ecuación por -6 y lo restamos a la segunda ecuación, así:

$$\begin{array}{rl} (2) & 38{,}96278 = 30\log c + 238\log b \\ (1) & -32{,}20722 = -30\log c - 180\log b \\ \hline & 6{,}75556 = 0 + 58\log b \end{array} \quad \Rightarrow \quad \log b = \frac{6{,}75556}{58} = 0{,}11647$$

Ahora, reemplazaremos en la primera ecuación y se tendrá que:

$$(1)\ 5{,}36787 = 5\log c + 30(0{,}11647) \Rightarrow \log c = \frac{5{,}36787 - 30\,(0{,}11647)}{5} = 0{,}37475$$

Reemplazamos en la ecuación general, nos quedará así: $\log \hat{Y}_i = 0{,}37475 + (0{,}11647)$

Ahora si deseamos estimar a $\hat{Y}$ cuando = 10, se tendrá que:

$\log \hat{Y} = 0{,}37475 + (10)\,(0{,}11647) = 1{,}53947$ Ahora sí obtenemos el antilogaritmo se tendrá que:

$\hat{Y} =$ antilog *de* $1{,}53945 = 34{,}63$ $\hat{Y} = 34{,}63$

Cálculo de b y c directamente

Se podrá obtener el *og b* y *og c* empleando fórmulas que permiten calcularlos directamente, así:

$$\log b = \frac{n\Sigma x_i \log y_i - (\Sigma x_i)(\Sigma \log y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad \log b = \frac{5(38{,}96278) - 30\ (5{,}36787)}{5(238) - (30)^2} = 0{,}11647$$

$$\log b = \frac{\Sigma x_i \log y_i - n(\bar{x})(\overline{\log y})}{\Sigma x_i^2 - n\bar{x}^2} \qquad \log b = \frac{38{,}96278 - 5(6)(1{,}07357)}{238 - 5(6)^2} = 0{,}11647$$

$$\log c = \frac{\Sigma \log y_t - \log b(\Sigma x_t)}{n} \qquad \log c = \frac{5{,}36788 - (0{,}11647)(30)}{5} = 0{,}37475$$

$$\log c = \overline{\log y_t} - \log b(\bar{x}) \qquad \log c = 1{,}07357 - (0{,}11647)\ (6) = 0{,}37475$$

La media de los logaritmos de $_i$ es: $\overline{\log y} = \frac{\Sigma \log y_i}{n}$ $\overline{\log y} = \frac{5{,}36787}{5} = 1{,}07357$

Por otra parte, pueden haber pequeñas diferencias en los resultados, aplicando diferentes fórmulas, dependiendo de los decimales que se empleen en los cálculos, especialmente cuando utilizamos calculadoras.

b = antilogaritmo del logaritmo de *b*

b = antilogaritmo de 0,11647 ⇒ $b = 1{,}30758$

Varian a residual y error estándar de estimación

Comencemos aplicando la fórmula que se utiliza con mayor frecuencia en los textos de estadística, siendo:

$$s^2_{x\log y} = \frac{\Sigma(\log y_t - \log \hat{Y}_t)^2}{n} \quad \Rightarrow \quad S^2_{x\log y} = \frac{0{,}0874684}{5} = 0{,}01749368$$

La varianza residual la calculamos mediante las siguientes operaciones

Tabla 12.10 Estimativos y desviaciones - Exponencial

log $_i$	log $\hat{Y}_i$	log $_i$ – log $\hat{Y}_i$	(log $_i$ – log $\hat{Y}_i$)²
0,47712	0,60768	-0,13056	0,0170459
0,77815	0,84063	-0,06248	0,0039038
1,07918	0,95710	0,12208	0,0161493
1,38021	1,19005	0,19016	0,0361608
1,65321	1,77241	-0,11920	0,0142086
5,36787	5,36787	0	0,0874684

Puede observarse que $\Sigma \log \ _i = \Sigma \log \hat{Y}_i$, es decir que la diferencia es igual a cero.

Además $\Sigma \ _i \neq \Sigma \hat{Y}_i$, se puede decir que: $\Sigma \ _i - \Sigma \hat{Y}_i \neq 0$

También se puede calcular la varianza residual, mediante las siguientes fórmulas:

$$S^2_{x \log y} = S^2_{\log y}(1 - R^2)$$ Siendo $S^2_{\log} = 0{,}17461(1 - 0{,}9012) = 0{,}017251$

$$S^2_{x \log y} = \frac{\Sigma(\log y_t)^2 - \log c(\Sigma \log y_t) - \log b(\Sigma x_t \log y_t)}{n}$$ reemplazando se tendrá que:

$$S^2_{x \log y} = \frac{6{,}63588 - (0{,}37473)\ (5{,}36787) - 0{,}11647\ (38{,}96278)}{5} = 0{,}017277$$

Es de anotar que la varianza de la variable se obtiene así:

$$S^2_{\log y} = \frac{\Sigma(\log y_t)^2 - n\overline{\log y}^2}{n} \quad \Rightarrow \quad S^2_{\log y} = \frac{6{,}63588 - 5(1{,}07357)^2}{5} = 0{,}17462$$

Coeficiente de correlación al cuadrado

Se puede obtener por diferentes fórmulas, entre otras se tiene:

$$R^2 = 1 - \frac{S^2_{x \log y}}{S^2_{\log y}} \qquad R^2 = 1 - \frac{0{,}017251}{0{,}17461} = 0{,}9012$$

Siendo el Coeficiente de correlación $r = \sqrt{0{,}9012} = 0{,}9493 \cong 0{,}95$

También lo podemos obtener mediante la siguiente fórmula:

$$r = \frac{n \Sigma x_t \log y_t - (\Sigma x_t)(\Sigma \log y_t)}{\sqrt{\left[n\Sigma x_t^2 - (\Sigma x_t)^2\right]\left[n\Sigma(\log y_t)^2 - (\Sigma \log y_t)^2\right]}}$$

$$r = \frac{5(38{,}96278) - (30)(5{,}36787)}{\sqrt{\left[5(238) - (30)^2\right]\left[5(6{,}63588) - (5{,}36787)^2\right]}} = 0{,}9493 = 0{,}95$$

$$R^2 = 0{,}9012$$

Algunos autores denominan al *coeficiente de correlación* (de cualquier ajuste) como *coeficiente de corre ación de Pearson*.

REGRESIÓN Y CORRELACIÓN MÚLTIPLE

Cuando se trabaja con más de dos variables independientes, relacionadas entre sí, el análisis de regresión recibe el nombre de *regresión m tip e*. El grado de relación que puede haber entre estas variables es cuantificado, mediante el cálculo del *coeficiente de correlación múltiple*.

En *estadística* con fines de explicación y aplicación, por lo general se trabaja con dos variables independientes, esto permite la realización de operaciones menos engorrosas que trabajando con más de

dos, lo que obligaría a la aplicación del álgebra matricial o al manejo de algunos paquetes estadísticos que contienen programas de regresión, entre otros SPSS, TSP, SAS, MICROSTAT y otros más, que no son muy comunes para los lectores de estos *apuntes*.

La expresión de la función utilizada en la regresión múltiple es variada, como se puede ver a continuación:

$$(1)\quad \hat{Y} = a + b_1 X_1 + b_2 X_2 + b_3 X_3 \ ...$$
$$(2)\quad \hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ ...$$
$$(3)\quad \hat{X}_1 = b_{1.23} + b_{12.3} x_2 + b_{13.2} x_3 \ ...$$

la (3) será la simbología que utilizaremos en esta unidad.

A los parámetros b_1 b_2 b_3 ó β_1 β_2 β_3 o $b_{12.3}$ $b_{13.2}$ se les denomina *coeficientes de regresión*. El coeficiente b , β , b , mide la cantidad (+/-) que aumenta o disminuye la variable dependiente $\hat{Y}$ ó $\hat{X}_1$, cuando la primera variable independiente X aumenta en una unidad, manteniendo constante los otros coeficientes; lo mismo puede decirse para los demás coeficientes de regresión, siempre que las demás variables independientes en la ecuación permanezcan constantes. Los *coeficientes de regresión* miden las elasticidades de la variable dependiente con respecto a la variable.

Ejemplo 1. Veremos cómo se opera utilizando este sistema de ecuaciones para el cálculo de los parámetros, tomando la ecuación general para $\hat{Y}$, como ejercicio de aplicación.

Tabla 12. 11 Datos y cálculo en la regresión múltiple

$_i$	$_i$	$_2$
2	3	5
3	5	8
5	8	12
9	10	15
11	14	20
30	40	60

$_i$ $_i$	$_i^2$	$_1$ $_2$	$_1$ $_2$	$_2^2$	$\hat{Y}_i$	$_i^2$
6	9	15	10	25	1,27	4
15	25	40	24	64	3,54	9
40	64	96	60	144	6,00	25
90	100	150	135	225	8,27	81
154	196	280	220	400	10,92	121
305	394	581	449	858	30,00	240

$$\hat{Y} = a + b_1 x_1 + b_2 x_2$$

$$(1)\quad \Sigma y_i = n a + b_1 \Sigma x_1 + b_2 \Sigma x_2 \qquad 30 = 5a + 40 b_1 + 60 b_2$$
$$(2)\ \Sigma y_t x_1 = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 \qquad 305 = 40a + 394b + 581 b_2$$
$$(3)\ \Sigma y_t x_2 = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 \qquad 449 = 60\, a + 581 b + 858\, b_2$$

Multiplicamos la ecuación (2) por -8 y se la restamos a la (1)

$$(2)\quad 305 = 40a + 394 b_1 + 581 b_2$$
$$(1)\quad -240 = -40a - 320 b_1 - 480 b_2$$
$$(4)\quad 65 = 0 + 74 b_1 + 101 b_2$$

Procedemos a multiplicar por -12 la ecuación (1) y se la restamos a la ecuación (3).

$$(3)\quad 449 = 60a + 581b_1 + 858b_2$$
$$(1)\quad -360 = -60a - 480b_1 - 720b_2$$
$$(5)\quad 89 = 0 + 101b_1 + 138b_2$$

Con la ecuaciones (4) y (5) procedemos a eliminar b_1

$$(5)\quad 6.586 = 7.474b_1 + 10.212b_2$$
$$(4)\quad -6.565 = -7.474b_1 - 10.201b_2$$
$$21 = 0 \qquad 11b_2 \quad \Rightarrow \quad b_2 = \frac{21}{11} = 1{,}909$$

(Se ha multiplicado la ecuación (5) por 74)
(Y multiplicamos la ecuacion (4) por -101)

Reemplazamos en la ecuación (4) y se tiene que:

$$(4)\ 65 = 74b_1 + 101(1{,}909) \quad \Rightarrow \quad b_1 = \frac{65 - 1{,}909(101)}{74} = -1{,}727$$

Conociendo $b_1 = -1{,}727$ y $b_2 = 1{,}909$, reemplazamos en la ecuación (1)
$30 = 5a + 40\,(-1{,}727) + 60\,(1{,}909)$

$$30 = 5a - 69{,}08 + 114{,}54 \Rightarrow \quad a = \frac{-15{,}46}{5} = -3{,}092$$

Los valores de los coeficientes de regresión son: $\beta_1 = b_1 = -1{,}727 \quad \beta_2 = b_2 = 1{,}909$ y $a = -3{,}092$

La ecuación general nos queda: $\hat{Y} = -3{,}092 - 1{,}727\ _1 + 1{,}909\ _2$

Si le damos valores arbitrarios a $_1 = 10$ y $_2 = 15$, se tendrá que
$\hat{Y} = -3{,}092 - 1{,}727\,(10) + 1{,}909\,(15) = 8{,}273$

La varianza residual será: $S^2_{y.x_1x_2} = \dfrac{\Sigma(y_t - \hat{Y}_t)^2}{n} \Rightarrow S^2_{y.x_1x_2} = 0{,}418$

y su desviación es: $S_{y.x_1x_2} = \sqrt{0{,}418} = 0{,}65$

El coeficiente de *corre ación m tip e* de en función de $_1$ y $_2$ se obtiene en forma simplificada mediante la fórmula:

$$R_{y.x_1x_2} \Rightarrow R = \sqrt{1 - \frac{S^2_{y.x_1x_2}}{S^2_y}} \quad \Rightarrow \quad R = \sqrt{1 - \frac{0{,}42}{12}} = 0{,}98 \qquad R^2 = 0{,}965 \cong 0{,}97$$

Los límites de confianza serán:

$$\hat{Y} \pm t\frac{S_{y.x_1x_2}}{\sqrt{n}} \quad \Rightarrow \quad 8{,}273 \pm 4{,}303\left[\frac{0{,}65}{\sqrt{5}}\right] \Rightarrow 8{,}273 \pm 1{,}25 = \begin{cases} 9{,}47 \\ 7{,}02 \end{cases}$$

Algunos trabajan con $\upsilon = n - 1 = 5 - 1 = 4$ grados de libertad otros, con: $\upsilon = n - 3 = 5 - 3 = 2$, en este caso $t = 4{,}303$ a un nivel de significación del 5%.

CÁLCULO DE ESTIMATIVOS APLICANDO OTRA SIMBOLOGÍA

Veamos como algunos trabajan, en primer lugar, cambiando de posición a Y_i dependiendo de la variable que se desea estimar:

(1) | Y_i | X_1 | X_2 | (2) | X_1 | Y_i | X_2 | (3) | X_1 | X_2 | Y_1 |

En estos casos, se estima $\hat{Y}$ en función de las otras dos o más variables, donde siempre se estará utilizando la misma ecuación general y el procedimiento de cálculo de los coeficientes de regresión se hará de la misma manera.

Al principio del tema sobre *regresión corre ación m tip e* se presentó la ecuación general de tres formas diferentes, ahora de ellas seleccionamos la tercera, considerándola con X_3.

Veamos el proceso de cálculo trabajando con la simbología de la ecuación (3)

Las variables las simbolizamos por $_1$ $_2$ $_3$ ..... $_n$ y los coeficientes de regresión de $\hat{X}_1$ por $b_{1.23}$, $b_{12.3}$, $b_{13.2}$ que son constantes, denominados también parámetros, cuando son aplicadas en poblaciones. Para los de $\hat{X}_2$ son: $b_{2.13}$ $b_{21.3}$ $b_{23\ 1}$ y para $\hat{X}_3$ son $b_{3.12}$ $b_{31.2}$ $b_{32.1}$.

$_1$ = Primera variable $_2$ = Segunda variable $_3$ = Tercera variable.

$b_{1.23}$= *Coeficiente de regresión parcial* de $_1$ (coeficiente angular) sobre $_2$ manteniéndose constante $_3$ (coeficiente angular).

$b_{13.2}$ = *Coeficiente de regresión parcial* de $_1$ sobre $_3$ manteniéndose constante $_2$ (coeficiente angular).

$b_{1.23}$= *Coeficiente de posición* de la primera variable en función de la segunda y tercera variable.

La ecuación $\hat{X}_1 = b_{1.23} + b_{12.3}\, x_2 + b_{13.2}\, x_3$ se denomina *ecuación de regresión inea* de $\hat{X}_1$ sobre $_2$ y $_3$. Las ecuaciones generales para $\hat{X}_2$ y $\hat{X}_3$ son:

$$\hat{X}_2 = b_{2.13} + b_{21.3}\, x_1 + b_{23.1}\, x_3 \qquad \hat{X}_3 = b_{3.12} + b_{31.2}\, x_1 + b_{32.1}\, x_2$$

Para obtener los *coeficientes de regresión parciales*, hay necesidad de resolver un sistema de ecuaciones normales. $\hat{X}_1 = b_{1.23} + b_{12.3}\, x_2 + b_{13.2}\, x_3$

$$\begin{aligned}
&a)\quad \Sigma x_1 = n b_{1.23} + b_{12.3}\Sigma x_2 + b_{13.2}\Sigma x_3\\
&b)\ \Sigma x_1 x_2 = b_{1.23}\Sigma x_2 + b_{12.3}\Sigma x_2^2 + b_{13.2}\Sigma x_3 x_2\\
&c)\ \Sigma x_1 x_3 = b_{1.23}\Sigma x_3 + b_{12.3}\Sigma x_2 x_3 + b_{13.2}\Sigma x_3^2
\end{aligned}$$

Para la regresión lineal de $\hat{X}_2$ en función de $_1$ y $_3$ se tiene que: $\hat{X}_2 = b_{2.13} + b_{21.3}\, x_1 + b_{23.1}\, x_3$

Las ecuaciones para hallar los parámetros son:

$$\begin{aligned}
&a)\quad \Sigma x_2 = n b_{2.13} + b_{21.3}\Sigma x_1 + b_{23.1}\Sigma x_3\\
&b)\ \Sigma x_2 x_1 = b_{2.13}\Sigma x_1 + b_{21.3}\Sigma x_1^2 + b_{23.1}\Sigma x_3 x_1\\
&c)\ \Sigma x_2 x_3 = b_{2.13}\Sigma x_3 + b_{21.3}\Sigma x_1 x_3 + b_{23.1}\Sigma x_3^2
\end{aligned}$$

En la regresión lineal de $\hat{X}_3$ en función de $_1$ y $_2$ se tiene: $\hat{X}_3 = b_{3.12} + b_{31.2}\,x_1 + b_{32.1}\,x_2$

Las ecuaciones para hallar los parámetros, son:

$$\begin{aligned}
&a)\quad \Sigma x_3 = n b_{3.12} + b_{31.2}\,\Sigma x_1 + b_{32.1}\,\Sigma x_2\\
&b)\,\Sigma x_3 x_1 = b_{3.12}\,\Sigma x_1 + b_{31.2}\,\Sigma x_1^2 + b_{32.1}\,\Sigma x_2 x_1\\
&c)\,\Sigma x_3 x_2 = b_{3.12}\Sigma x_2 + b_{31.2}\,\Sigma x_1 x_2 + b_{32.1}\,\Sigma x_2^2
\end{aligned}$$

Se puede simplificar la simbología usada en los tres casos anteriores, de la siguiente forma:

$x_1 = X$	$b_{1.23}$	$b_{2.13}$	$b_{3.12}$	$\rightarrow$ *por a*
$x_2 = Y$	$b_{12.3}$	$b_{21.3}$	$b_{31.2}$	$\rightarrow$ *por b*
$x_2 = Y$	$b_{13.2}$	$b_{23.1}$	$b_{32.1}$	$\rightarrow$ *por c*

Proceso de cálculo de los coeficientes de regresión mediante la primera simbología sigue siendo igual al aplicado en el ejercicio anterior:

$$\begin{aligned}
&a)\quad 30 = 5b_{1.23} + 40\,b_{12.3} + 60b_{13.2}\\
&b)\quad 305 = 40b_{1.23} + 394\,b_{12.3} + 581b_{13.2}\\
&c)\quad 449 = 60b_{1.23} + 581b_{12.3} + 858b_{13.2}
\end{aligned}$$

Multiplicamos la ecuación *a*) por -8 y la restamos a *b*)

$$\begin{aligned}
&b)\quad 305 = 40b_{1.23} + 394\,b_{12.3} + 581b_{13.2}\\
&a)\quad -240 = -40b_{1.23} - 320\,b_{12.3} - 480b_{13.2}\\
\hline
&d)\quad 65 = 0 + 74b_{12.3} + 101b_{13.2}
\end{aligned}$$

Multiplicamos por -12 la ecuación *a*) y la restamos a la *c*)

$$\begin{aligned}
&c)\quad 449 = 60b_{1.23} + 581b_{12.3} + 858b_{13.2}\\
&a)\quad -360 = -60b_{1.23} - 480\,b_{12.3} - 720b_{13.2}\\
\hline
&e)\quad 89 = 0 + 101b_{12.3} + 138b_{13.2}
\end{aligned}$$

Con las ecuaciones *d*) y *e*) podemos eliminar a $b_{12.3}$ para determinar $b_{13.2}$

$$\begin{aligned}
&6.586 = 7.474\,b_{1.23} + 10.212\,b_{13.2} \qquad \text{(multiplicamos por 101)}\\
&-6.565 = -7.474\,b_{1.23} + 10.201b_{13.2} \qquad \text{(multiplicamos por -79)}\\
\hline
&21 = 0 + 11b_{13.2} \quad \Rightarrow \quad b_{13.2} = \frac{21}{11} = 1{,}909
\end{aligned}$$

Reemplazamos en la ecuación d) y tenemos que:

$$65 = 74\,b_{12.3} + 101(1{,}909)$$
$$65 = 74b_{12.3} + 192{,}809 \;\Rightarrow\; -127{,}809 = 74b_{12.3} \;\Rightarrow\; b_{12.3} = -\frac{127{,}81}{74} = -1{,}727$$

Teniendo $b_{13.2} = 1{,}909$ y $b_{12.3} = -\,1{,}727$ reemplazamos en la ecuación (a):

$$30 = 5\,b_{1.23} + 40\,(-1{,}727) + 60\,(1{,}909)$$
$$30 = 5b_{1.23} - 69{,}8 + 114{,}54 \;\Rightarrow\; 99{,}08 - 114{,}54 = 5b_{1.23} \;\Rightarrow\; b_{1.23} = -\frac{15{,}43}{5} = -3{,}092$$

La *ecuación genera* nos queda: $\hat{X}_1 = -3{,}092 - 1{,}727\ {}_2 + 1{,}909\ {}_3$

Si le damos arbitrariamente valores a ${}_2 = 10$ y ${}_3 = 15$, se tendrá que:

$\hat{X}_1 = -3{,}092 - 1{,}727\,(10) + 1{,}909\,(15) = 8{,}273$

La *des iación típica residua* de $\hat{X}_1$ sobre ${}_2$ y ${}_3$ se calcula mediante la fórmula:

$$S_{1.23} = +\sqrt{\frac{\Sigma\left(x_1 - \hat{X}_1\right)^2}{n}} \qquad S_{1.23} = \sqrt{\frac{2{,}09}{5}} = \sqrt{0{,}418} = 0{,}65$$

$S^2_{1.23} = 0{,}65^2 = 0{,}42$ (varianza residual)

El *coeficiente de correlación múltiple* de $\hat{X}_1$ sobre ${}_2$ y ${}_3$ se obtiene:

$$R_{1.23} = \sqrt{1 - \frac{S^2_{1.23}}{S^2_1}} \qquad S^2_1 = \frac{\Sigma\left(x_1 - \bar{x}_1\right)^2}{n} \text{ (varianza)} \qquad S^2_1 = \frac{60}{5} = 12$$

$$R_{1.23} = \sqrt{1 - \frac{0{,}42}{12}} = \sqrt{1 - 0{,}035} = \sqrt{0{,}965} = 0{,}98$$

OTRO PROCEDIMIENTO APLICANDO EL MÉTODO DE JORDAN - GAUSS

Mediante la solución de un sistema de ecuaciones lineales, utilizando el método de Jordan - Gauss, así:

a	$5X_1 + 40X_2 + 60X_3$	30
b	$40X_1 + 394X_2 + 581X_3$	305
c	$60X_1 + 581X_2 + 858X_3$	449
$a_1 = \frac{a}{5}$	$X_1 + 8X_2 + 12X_3$	6
$b_1 = b - 40a_1$	$0 + 74X_2 + 101X_3$	65
$c_1 = c - 60a_1$	$0 + 101X_2 + 138X_3$	89
$a_2 = a_1 - 8b_2$	$X_1 + 0 + 1{,}0808X_3$	-1.0272
$b_2 = b_1 / 74$	$0 \quad X_2 + 1{,}3649X_3$	0,8784
$c_2 = c_1 - 101b_2$	$0 \quad 0 + 0{,}1451X_3$	0,2816
$a_3 = a_2 - 1{,}0808c_3$	$X_1 \quad 0 \quad 0$	-3,1247
$b_3 = b_2 - 1{,}3649c_3$	$0 \quad X_2 \quad 0$	-1,7705
$c_3 = c_2 / 0{,}1451$	$0 \quad 0 \quad X_3$	1,9407

Procedimiento

- Dividimos la primera ecuación (*a*) por 5 y luego con esta ecuación eliminamos a X_1 en las ecuaciones *b* y *c*
- El segundo paso consiste en dividir la ecuación b_1 por 74 y luego con ella eliminamos a X_2 en las ecuaciones a_1 y c_1 .
- Luego, en el tercer paso se trabaja con la tercera ecuación c_3 la cual se obtiene, dividiendo a la ecuación c_2 por 0,1451, la que a su vez nos permitirá eliminar a X_3 en las ecuaciones a_2 y b_2.

Se tendrá que $X_1 = b_{1.23} = -3{,}1251$ $X_2 = b_{12.3} = -1{,}7705$ y $X_3 = b_{13.2} = 1{,}9407$ siendo: $\hat{X}_1 = -3{,}125 - 1{,}7705\ _2 + 1{,}9407\ _3$

Supongamos que desea estimar el valor para $\hat{X}_1$ cuando $_2 = 9$ y $_3 = 14$.

Reemplazando se tiene que: $\hat{X}_1 = -3{,}1251 - 1{,}7705\,(9) + 1{,}9407\,(14) = 8{,}11$

En cada una de las estimaciones de $\hat{X}_1$ $\hat{X}_2$ y $\hat{X}_3$, se puede hallar el *coeficiente de correlación m tip e* y la *arian a residua* para cada caso, siendo:

$$R_{1.23} = \sqrt{1 - \frac{S^2_{1.23}}{S^2_1}} \qquad R_{2.13} = \sqrt{1 - \frac{S^2_{2.13}}{S^2_2}} \qquad R_{3.12} = \sqrt{1 - \frac{S^2_{3.12}}{S^2_3}}$$

$$S^2_{1.23} = \frac{\Sigma(x_1 - \hat{X}_1)^2}{n} \qquad S^2_{2.13} = \frac{\Sigma(x_2 - \hat{X}_2)^2}{n} \qquad S^2_{3.12} = \frac{\Sigma(x_3 - \hat{X}_3)^2}{n}$$

Se ha dicho que el *error estándar de estimación* o error típico de estimación se obtiene mediante la raíz cuadrada de la varianza residual. Este error se puede calcular utilizando los *coeficientes de correlación parcia* r_{12} r_{13} r_{23}. Se debe observar que $r_{12} = r_{21}$ $r_{13} = r_{31}$ $r_{23} = r_{32}$ con el fin de no repetir operaciones.

Se tendrá que:

$$S_{1.23} = S_1 \sqrt{\frac{1 - r^2_{12} - r^2_{13} - r^2_{23} + 2r_{12}\,r_{13}\,r_{23}}{1 - r^2_{23}}}$$

$$S^2_{2.13} = S_2 \sqrt{\frac{1 - r^2_{21} - r^2_{23} - r^2_{13} + 2r_{21}\,r_{23}\,r_{13}}{1 - r^2_{13}}}$$

$$S^2_{3.12} = S_3 \sqrt{\frac{1 - r^2_{31} - r^2_{32} - r^2_{12} + 2r_{31}\,r_{32}\,r_{12}}{1 - r^2_{12}}}$$

Los *coeficientes de correlación parcial* se calculan así:

$$r_{21} = r_{12} = \frac{n\Sigma x_1 x_2 - (\Sigma x_1)(\Sigma x_2)}{\sqrt{\left[n\Sigma x_1^2 - (\Sigma x_1)^2\right]\left[n\Sigma x_2^2 - (\Sigma x_2)^2\right]}}$$

$$r_{21} = r_{12} = \frac{5(305) - (30)(40)}{\sqrt{\left[5(240) - (30)^2\right]\left[5(394) - (40)^2\right]}} = 0{,}975 \qquad r_{12}^2 = 0{,}95$$

$$r_{31} = r_{13} = \frac{n\Sigma x_1 x_3 - (\Sigma x_1)(\Sigma x_3)}{\sqrt{\left[n\Sigma x_1^2 - (\Sigma x_1)^2\right]\left[n\Sigma x_3^2 - (\Sigma x_3)^2\right]}}$$

$$r_{31} = r_{13} = \frac{5(449) - (30)(60)}{\sqrt{\left[5(240) - (30)^2\right]\left[5(858) - (60)^2\right]}} = 0{,}978 \qquad r_{13}^2 = 0{,}956$$

$$r_{32} = r_{23} = \frac{n\Sigma x_2 x_3 - (\Sigma x_2)(\Sigma x_3)}{\sqrt{\left[n\Sigma x_2^2 - (\Sigma x_2)^2\right]\left[n\Sigma x_3^2 - (\Sigma x_3)^2\right]}}$$

$$r_{32} = r_{23} = \frac{5(581) - (40)(60)}{\sqrt{\left[5(394) - (40)^2\right]\left[5(858) - (60)^2\right]}} = 0{,}999 \qquad r_{23}^2 = 0{,}998$$

Ahora se podrán calcular los *errores estándar* o *típico* para cada estimador, así:

$$S_{1.23} = 3{,}46\sqrt{\frac{1 - 0{,}95 - 0{,}956 - 0{,}998 + 2(0{,}975)(0{,}978)(0{,}999)}{1 - 0{,}998}} = 2{,}67$$

$$S_{1.23}^2 = 7{,}14$$

$$S_{2.13} = 3{,}85\sqrt{\frac{1 - 0{,}95 - 0{,}998 - 0{,}956 + 2(0{,}975)(0{,}999)(0{,}978)}{1 - 0{,}956}} = 0{,}63$$

$$S_{2.13}^2 = 0{,}40$$

$$S_{3.12} = 5{,}25\sqrt{\frac{1 - 0{,}956 - 0{,}998 - 0{,}95 + 2(0{,}978)(0{,}999)(0{,}975)}{1 - 0{,}95}} = 0{,}81$$

$$S_{3.12}^2 = 0{,}66$$

Las *des iaciones típicas* respectivamente son:

$$S_1 = \sqrt{\frac{\Sigma x_1^2 - n\bar{x}_1^2}{n}} = \sqrt{\frac{240 - 5(6)^2}{5}} = 3{,}46$$

$$S_2 = \sqrt{\frac{\Sigma x_2^2 - n\bar{x}_2^2}{n}} = \sqrt{\frac{394 - 5(8)^2}{5}} = 3{,}85$$

$$S_3 = \sqrt{\frac{\Sigma x_3^2 - n\bar{x}_3^2}{n}} = \sqrt{\frac{858 - 5(12)^2}{5}} = 5{,}25$$

Los coeficientes de *corre ación m tip e inea* , pueden también ser calculados mediante la aplicación de las fórmulas siguientes.

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2\, r_{12}\, r_{13}\, r_{23}}{1 - r_{23}^2}}$$

$$R_{1.23} = \sqrt{\frac{0{,}95 + 0{,}956 - 2(0{,}975)(0{,}978)(0{,}999)}{1 - 0{,}998}} = 0{,}64$$

$$R_{2.13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2\, r_{21}\, r_{23}\, r_{13}}{1 - r_{13}^2}}$$

$$R_{2.13} = \sqrt{\frac{0{,}95 + 0{,}998 - 2(0{,}975)(0{,}999)(0{,}978)}{1 - 0{,}956}} = 0{,}99$$

$$R_{3.12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2\, r_{31}\, r_{32}\, r_{12}}{1 - r_{12}^2}}$$

$$R_{3.12} = \sqrt{\frac{0{,}956 + 0{,}998 - 2(0{,}978)(0{,}999)(0{,}975)}{1 - 0{,}95}} = 0{,}99$$

Recordemos que si el coeficiente de correlación se acerca a 1, se tendrá una mejor relación lineal entre las variables, pero si se aproxima a 0 nos estará indicando que no hay relación lineal entre las variables, pero es posible que exista entre ellos una relación no lineal.

Como observación a los resultados anteriores de $R_{1,23}$ $R_{2,13}$ $R_{3,12}$ éstos no son necesariamente iguales, tal como se puede comprobar en el ejercicio que se acaba de desarrollar.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

71. Una empresa considera que lo más conveniente para sus actividades, es lograr una participación del 24% en el mercado y una utilidad no menor de $345.000.000.

    La información obtenida durante diez años sobre: la participación en el mercado, el porcentaje de descuento que se hace para lograr aumento en las ventas y la utilidad obtenida, sin necesidad de hacer reajustes excesivos en el precio, es la siguiente.

PARTICIPACIÓN EN EL MERCADO %	DESCUENTO CONCEDIDO %	UTILIDAD (en millones de)
5	20	270
9	18	250
12	16	280
8	10	260
16	14	310
18	16	330
19	16	350
20	17	320
18	17	360
27	20	330

Se pide: ¿Cuál debe ser el nivel de descuento por conceder, teniendo en cuenta el nivel de participación en el mercado y la utilidad deseada, si la participación en el mercado es del 35% y el descuento que se concede es del 24%?, ¿qué utilidad esperaría?

72. Con los siguientes datos correspondientes a 6 períodos:

X_1 = Consumo de bienes industriales

X_2 = Ingreso disponible

X_3 = Importaciones de bienes de consumo

(La información está dada en unidades monetarias constantes).

$\Sigma x_1 = 308$ $\Sigma x_1 x_2 = 18.555$ $\Sigma x_1 x_3 = 4.264$

$\Sigma x_2 = 363$ $\Sigma x_2^2 = 22.157$ $\Sigma x_3^2 = 1.145$

$\Sigma x_3 = 81$ $\Sigma x_2 x_3 = 4.994$ $\Sigma x_1^2 = 16.152$ $n = 6$

a) Estimar el valor de $\hat{X}_1$ cuando $x_2 = 70$ y $x_3 = 23$
b) Estimar el valor para $\hat{X}_2$ cuando $x_1 = 48$ y $x_3 = 10$
c) Determinar los coeficientes de correlación parciales.

73. Siendo:

$\Sigma x_1 = 129$ $\Sigma x_1 x_2 = 1.996$ $\Sigma x_1 x_3 = 201,98$

$\Sigma x_2 = 79$ $\Sigma x_2^2 = 1.279$ $\Sigma x_3^2 = 14,1881$

$\Sigma x_3 = 7,23$ $\Sigma x_3 x_2 = 129,39$ $n = 6$

X_1 = Precio de bienes agrícolas X_2= Costo unitario de producción X_3 = Expectativas de inflación (%)

Se pide estimar el precio de los bienes agrícolas, cuando el costo de producción es de 40 y la expectativa de inflación del 8%.

74. Con la siguiente información:

$\Sigma x_1 = 3.247,4$ $\Sigma x_2 x_1 = 98.060,1$ $\Sigma x_2 = 604$ $n = 15$

$\Sigma x_3 = 95$ $\Sigma x_3 x_1 = 18.057,0$ $\Sigma x_2^2 = 30.308$

$\Sigma x_3^2 = 725$ $\Sigma x_1^2 = 939.175{,}68$ $\Sigma x_2 x_3 = 3.833$

a) Establezca la ecuación de regresión para estimar $\hat{X}_2$
b) Obtener el valor estimado para $\hat{X}_2$ cuando $x_1 = 210$ $x_3 = 5$

75. Considere un estudio realizado para examinar el papel que juega la televisión en la vida de un grupo de personas de edades superiores a los 30 años e inferior a los 50, teniendo en cuenta los años de escolaridad.

Individuo:	1	2	3	4	5	6	7	8	9	10
Horas TV.:	0,5	0,5	0,9	1,6	2,5	2,8	3,0	3,2	3,3	3,7
Edad:	43	36	39	41	34	46	35	32	49	38
Escolaridad:	14	16	10	12	8	12	6	6	4	12

Estime el número de horas TV., para una persona de 40 años de edad y 11 años de escolaridad. Determine el coeficiente de correlación múltiple.

76. Con los siguientes datos:

$_i$	$_i$	$_2$
5	2	4
7	4	6
10	7	9
16	10	14
22	12	17

Se pide:
a) Estimar $\hat{Y}$
cuando $_1 = 11$ y $_2 = 16$
b) Determinar el coeficiente de correlación de $\hat{Y}$ con base en $_1$ y $_2$

COEFICIENTE DE CORRELACIÓN POR RANGOS (VER CAP. 9 PG. 460)

Los coeficientes de *corre ación por rangos* más utilizados son los de Pearson y Spearman, este último de gran aplicación en la investigación de mercados, especialmente cuando se desean investigar preferencias, actitudes o atributos, por medio de algún sistema de calificación que va en orden ascendente de graduación. Puede utilizarse en los fenómenos cuantitativos o variables, sin que éstos tengan significado numérico.

La fórmula aplicada es $r_s = 1 - \dfrac{6\Sigma d_t^2}{n^3 - n}$

Nota: en el *capítu o* , página 460 se hace referencia a esta prueba *no paramétrica*, donde el alumno podrá complementar sus conocimientos.

- d_i = representa la diferencia entre los rangos pareados de cada par de variantes $_i$ y $_i$.
- n = es el número de rangos o sea el número de variantes apareadas.

El procedimiento que se sigue en el cálculo del coeficiente, consiste en colocar en orden: primero, segundo, tercero, etc., con respecto al valor que toma cada una de las variables y que denominamos rango y luego se establece el cuadrado de las diferencias entre los dos rangos que usaremos para el cálculo del coeficiente de correlación.

Ejemplo 1. Supongamos que una fábrica ordena a sus 15 obreros, de acuerdo con la eficiencia, dos trabajos. Se desea calcular el coeficiente de correlación por rangos entre estos datos.

TRABAJADOR	RANGO PARA EL TRABAJO I	RANGO PARA EL TRABAJO II	d_i	d_i
A	1	15	-14	196
B	2	14	-12	144
C	3	13	-10	100
D	4	3	1	1
E	5	4	1	1
F	6	12	-6	36
G	7	1	6	36
H	8	11	-3	9
I	9	2	7	49
J	10	7	3	9
K	11	8	3	9
L	12	5	7	49
M	13	6	7	49
N	14	10	4	16
O	15	9	6	36
–	–	–	–	740

$$r_s = 1 - \frac{6\,(740)}{15^3 - 15} \qquad r_s = 1 - 1{,}32 = -0{,}32$$

Siendo $-1 \le r_s \le 1$ se puede afirmar que no hay ninguna relación clara entre los dos rangos.

Ejemplo 2. Con los siguientes datos, correspondientes a 13 estudiantes aspirantes a ingresar a la universidad, calcular el coeficiente de Spearman.

ESTUDIANTE :	1	2	3	4	5	6	7	8	9	10	11	12	13
PROMEDIO Bto :	6,7	6,1	9,2	6,1	6,7	9,2	7,0	6,1	6,7	7,5	7,5	9,6	7,5
PUNTAJE ICFES:	333	300	286	290	286	300	326	348	326	300	384	300	286

Solución:

Primero ordenamos las calificaciones de menor a mayor (también se puede de mayor a menor) y luego establecemos el rango para cada calificación.

No.	x_i	RANGO x_i	y_i	RANGO y_i	x_i	y_i	RANGO y_i	RANGO x_i	$d_i = x_i - y_i$	d_i^2
1	6,1	2	286	2	6,7	333	5	11	-6,0	36,00
2	6,1	2	286	2	6,1	300	2	6,5	-4,5	20,25
3	6,1	2	286	2	9,2	286	11,5	2	9,5	90,25
4	6,7	5	290	4	6,1	290	2	4	-2,0	4,00
5	6,7	5	300	6,5	6,7	286	5	2	3,0	9,00
6	6,7	5	300	6,5	9,2	300	11,5	6,5	5,0	25,00
7	7,0	7	300	6,5	7,0	326	7	9,5	-2,5	6,25

ontin a

ontinuación

8	7,5	9	300	6,5	6,1	348	2	12	-10,0	100,00
9	7,5	9	326	9,5	6,7	326	5	9,5	-4,5	20,25
10	7,5	9	326	9,5	7,5	300	9	6,5	2,5	6,25
11	9,2	11,5	333	11	7,5	384	9	13	-4,0	16,00
12	9,6	11,5	348	12	9,6	300	13	6,5	6,5	42,25
13	9,6	13	384	13	7,5	286	9	2	7,0	49,00
					Σ		-	-	0	424,50

Al frente de cada valor de y se coloca el respectivo rango, luego se establecen las diferencias, se elevan al cuadrado y finalmente se suma esta columna.

$$r_s = 1 - \frac{6\, d_t^2}{n^3 - n} \qquad r_s = 1 - \frac{6\,(424{,}5)}{13^2 - 13} = -0{,}166$$

No hay correlación alguna entre las calificaciones obtenidas en el bachillerato, con las pruebas del ICFES, para ingreso a la Universidad.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

77. Una firma de compraventa de automóviles desea determinar la relación que existe entre el puntaje de una prueba de aptitud y el número de automóviles vendidos por cada vendedor en su primer año de trabajo. Una muestra aleatoria de 15 vendedores da la siguiente información.

VENDEDOR	PUNTAJE PRUEBA	No. DE AUTOMÓVILES
1	72	34
2	88	42
3	70	32
4	72	44
5	87	29
6	71	32
7	85	41
8	89	46
9	93	50
10	98	51
11	93	53
12	96	32
13	86	39
14	82	43
15	88	37

Calcular el coeficiente de correlación de Spearman, como medida del grado de relación entre el puntaje en la prueba y el número de automóviles vendidos.

78. Para medir la reacción de un grupo de consumidores respecto a la presentación de un producto, se tomaron 12 personas y se les interrogó en cuanto al concepto que les merecía la presentación tradicional de ese artículo; luego se les presentó el mismo artículo en su nuevo empaque. Las mediciones se efectuaron utilizando una escala de 20 puntos, obteniéndose los siguientes resultados, para cada una de las 12 personas.

TRADICIONAL:	6	10	14	17	6	6	11	14	10	17	11	13
NUEVO:	10	13	18	12	9	8	12	18	17	10	18	20

Calcular el coeficiente de correlación de Spearman.

79. Los datos siguientes corresponden a los costos y ventas, en millones de pesos, de 14 droguerías, en una semana:

DROGUERÍA	COSTOS	VENTAS
1	13	21
2	12	17
3	16	22
4	15	16
5	14	18
6	22	35
7	23	34
8	17	20
9	24	31
10	20	24
11	21	25
12	18	25
13	15	24
14	22	31

a) Calcule el coeficiente de correlación de *Pearson por rangos de Spearman*

b) Compare los resultados. ¿Cuál cree usted que sea el indicado?

80. Una oficina gubernamental clasifica 14 diferentes marcas de un articulo X de acuerdo a la calidad general de ellos. La tabla da el rango asignado a cada marca, así como también los precios promedios de cada una (en miles de pesos).

MARCA:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
RANGO DE CALIDAD:	11	1	2	5	3,5	3,5	6	9	12	13,5	9	7	13,5	9
PRECIO PROMEDIO:	45	60	53	52	60	60	20	52	70	68	70	52	68	45

Calcular el coeficiente de correlación por rangos, para medir el grado de correlación entre la calidad y el precio.

81. Supongamos que se desea determinar si las calificaciones obtenidas por los obreros en cierto test, se correlacionan con su rendimiento en el trabajo, expresado por el número de unidades producidas en un período dado.

CALIFICACIÓN TEST:	65	80	76	75	80	78	83	84	85	90	65	83	76	76	83	65
UNIDADES PRODUCIDAS:	30	25	35	40	38	42	48	50	55	45	25	35	38	35	25	42

Calcular el coeficiente de correlación por Rangos - Método de Spearman.

Nota: a partir del siguiente ejercicio, se plantean situaciones donde el alumno debe identificar a qué tema pertenece y dar la solución correspondiente.

82. Al observar los resultados del proceso de sedimentación o formación de costras en el interior de tubos de gran diámetro, que acarrean aguas negras, se construyó la siguiente tabla, con el objeto de establecer la comparación (relación) entre los años de servicio y el espesor de las costras formadas en su interior:

AÑOS DE SERVICIO	$_i$:	4	2	3	6	5	1	7
ESPESOR DE COSTRAS (mm)	$_i$:	10	6	8	15	11	5	16

De acuerdo con los datos anteriores, se pide:

a) Calcular el coeficiente de correlación.

b) Se dice que un tubo está totalmente fuera de servicio por acción del espesor de la costra cuando alcanza los 22 mm. ¿A los cuántos años queda fuera de servicio el tubo?

83. Explique brevemente:

a) ¿La diferencia entre el análisis de regresión y el de correlación?. ¿Cuándo debe aplicarse cada uno?

b) ¿Qué interpretación le da usted a la varianza residual y a la varianza explicada?

c) ¿Qué pasa si la varianza explicada es igual a la varianza total?

84. Un funcionario de una entidad, que generalmente, presta asesoría a pequeños establecimientos donde se reconstruyen pequeños motores a gasolina para bombas hidráulicas, ha realizado un estudio acerca de las diferencias entre la habilidad de los trabajadores para aprender el nuevo trabajo y su capacidad de realizarlo sin supervisión. Después de varias semanas, ha obtenido la siguiente información, donde Y = producción mensual de motores correctamente reconstruidos y X = años de educación en cada uno de los 10 trabajadores.

$_i$:	7	5	6	4	10	8	6	5	10	6	8
y_i :	15	7	10	8	20	16	12	8	16	8	14

a) Estime la producción mensual promedio de motores correctamente reconstruidos por un trabajador que presenta en su hoja de vida 9 años de estudio.

b) Fije los límites de confianza del 95%, para dicho promedio.

c) Interprete el coeficiente de regresión y el coeficiente de correlación

85. El departamento de personal de una empresa industrial desea estudiar la relación que hay entre varias variables obtenidas en una investigación. Estas variables son: salario semanal (miles de $), tiempo de servicio (meses) y edad de sus empleados (años). La información correspondiente a una muestra de 15 de sus empleados.

EMPLEADO:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1 5
SALARIO:	61,9	73,6	65,8	86,3	50,8	54,2	66,2	70,8	79,0	54,2	86,3	73,6	64,4	51,9	52,0
TIEMPO SERVICIO:	330	569	375	113	215	343	252	265	352	8 7	220	160	4 2	30	130
EDAD:	46	59	52	40	46	54	39	41	46	30	4 8	35	30	30	36

1. Se pide mediante una regresión lineal simple:
   a) Estimar el salario semanal para una persona con 200 meses de servicio.
   b) Estimar el salario semanal para una persona con una edad de 38 años.
   c) En los dos casos, cómo interpretar el coeficiente de correlación.

2. Mediante la regresión lineal múltiple, se pide:
   a) ¿Cuál deber ser el salario semanal para un trabajador de la empresa que tiene 260 meses de servicio y 53 años de edad?
   b) Interprete los dos coeficientes de regresión (b_1 y b_2) simbolizados también como: $b_{12.3}$ y $b_{13.2}$
   c) Calcule el coeficiente de determinación múltiple ($R^2_{1.23}$) e interprete su resultado.
   d) Calcule el error estándar de estimación múltiple para el ejercicio (a).
   e) Establezca una estimación de intervalo con el 95% de confianza para el salario semanal (miles de $) promedio de un empleado que tiene 10 años de servicio en la empresa y 47 años de edad.
   f) Calcule los coeficientes de correlación total $r_{1.23}$, $r_{2.13}$ y $r_{3.12}$, interprete los resultados.

EJERCICIOS MISCELÁNEOS

(CUBRE TODOS LOS TEMAS DEL CAPÍTULO)

86. Marque con una *X* la respuesta correcta en cada punto:

A. El coeficiente de correlación
   a) Es una medida que describe la intensidad de la relación de dos variables.
   b) Es una medida de posibilidad asociada a un intervalo de confianza.
   c) Es una medida de asociación entre dos atributos.
   d) Se usa para designar el cuadrado del coeficiente de correlación.
   e) Es el coeficiente de una variable independiente en una ecuación de regresión.

B. El error estándar de estimación.
   a) Es el cuadrado de las diferencias entre los valores reales y estimados.
   b) Es el promedio de los cuadrados de las diferencias entre los valores estimados y la media.
   c) Es la raíz cuadrada de la media de los cuadrados de las diferencias entre los valores reales y estimados, multiplicados por el coeficiente de correlación.
   d) Es la raíz cuadrada de la media de los cuadrados de las diferencias entre los valores reales y estimados.

87. En un análisis de dos variables, qué significado e importancia tienen para usted los resultados correspondientes a:
   a) Varianza residual; b) Varianza explicada; c) Error estándar de estimación d) Coeficiente de correlación; e) Límites de confianza para el estimativo.

88. Defina de la manera más simple:
   a) Regresión b) Correlación
   c) Nube de puntos d) Método de los mínimos cuadrados

89. Dos ecuaciones de regresión, cuyas ecuaciones son: $\hat{Y} = 2{,}8 \quad + 10$ y $\hat{X} = 0{,}3 \quad + 20$. ¿Cuál es el valor del coeficiente de correlación lineal?

90. La siguiente información corresponde: X = edad de máquina en años y Y = gastos de mantenimiento durante los dos últimos años (en diez miles de $) Utilizando los datos de cuatro máquinas, los cuales fueron: $\Sigma \; _i = 12$; $\Sigma \; _i = 24$; $\Sigma \; _i \; _i = 78$; $\Sigma \; _i^2 = 44$. Calcular:

a) El valor de c b) $\hat{Y} = b \quad + c$ c) $\hat{Y}$ para $= 6$

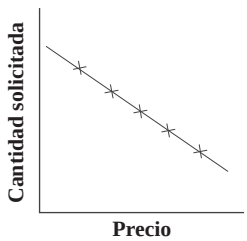
91. Dadas las siguientes variables X = gastos de mantenimiento del local (cientos de $) y Y = miles de artículos vendidos en un día por almacén.

$\Sigma \; _i = 194 \quad \Sigma \; _i^2 = 323.200 \quad ^- = 9.360 \quad s^2 = 7.694 \quad r = 0{,}9 \quad n = 20$

a) Estime el modelo de regresión lineal para $\hat{Y}$ e interprete el valor de b (β)

b) ¿Cuántos artículos se venderán en un día si se gastan $940.000 en mantenimiento?

92. Observe la gráfica y responda los siguientes puntos:



1. En este problema específico, el investigador intenta pronosticar:
   a) La cantidad solicitada (demandada) con base en el precio
   b) El precio con base a la cantidad solicitada
   c) Tanto el precio como la cantidad solicitada
   d) Ninguna de las opciones es correcta

2. Cualquier pronóstico con base en estas nociones serían:
   a) De ninguna o poca utilidad
   b) Con un error tan grande que no satisface
   c) Sin error
   d) Ninguna de las opciones es correcta

93. ¿Cuál es el coeficiente de correlación entre dos variables si:
   a) ¿Una de las variables es constante?
   b) ¿El valor de una de las variables excede en 50 unidades al de la otra variable?
   c) ¿La variación no explicada es el doble de la explicada?

94. Suponga que al analizar la relación entre 30 observaciones de dos variables, se encuentra que la suma de los cuadrados de las diferencias entre los valores de la variable Y y su media aritmética es 220; y que la suma de los cuadrados de las diferencias entre los valores estimados y su media es de 38,4. La pendiente es positiva.

   a) ¿Cuál es el coeficiente de determinación?

   b) ¿Qué porcentaje de la variación total ha sido explicada?

   c) ¿Puede rechazarse la hipótesis de que no existe correlación lineal a un nivel de significación del 5%?

95. En una regresión de sobre basada en 20 observaciones, el valor del coeficiente de determinación es 0,36.

a) ¿Indica este resultado que existe una correlación significativa entre y a un nivel de significación del 5% siendo la prueba bilateral?

b) ¿Cuál es la proporción de la variación que queda sin explicar por esta recta de regresión?

96. Se dispone de la siguiente información sobre dos variables relacionadas entre sí, obtenidas mediante una muestra aleatoria de 50 observaciones.

x_i:	2	6	9	14	16	23
y_i:	5	13	20	25	35	52
n_i:	18	12	8	6	4	2

Resultados de algunas operaciones

$\Sigma x_i n_i = 374$ $\Sigma x_t^2 n_t = 4.410$

$\Sigma x_t y_t n_t = 9.288$ $\Sigma y_t n_t = 800$

$\Sigma y_t^2 n_t = 19.736$ $n = 50$

Con la información anterior, se pide estimar $\hat{Y}$ cuando = 30 fijando límites de confianza del 95%

97. El personal técnico operativo que solicita ingreso a una empresa manufacturera, es sometido a un proceso de evaluación (conocimientos, experiencia, etc.), utilizando una escala de valoración entre 0 - 20; dos meses después de su ingreso realiza un curso de capacitación avanzada, calificada de 0 - 5, tal como lo muestra la tabla adjunta de 14 funcionarios seleccionados al azar:

EMPLEADO:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
EVALUACIÓN:	12	10	13	8	7	11	11	14	12	14	9	10	10	13
CAPACITACIÓN:	3,9	4,2	4,2	3,8	3,8	3,8	4,3	4,4	4,5	4,4	3,9	3,7	4,3	3,6

Determinar la relación existente entre la evaluación y las calificaciones que recibieron en la capacitación. ¿Cuál es el coeficiente de correlación de rangos? (Método de Spearman)

98. Una firma comercial tiene actualmente 180 puntos de ventas, con deseos de ampliar su cobertura a 300 puntos en total. Para decidir, desarrolla una encuesta a 15 de los puntos de ventas seleccionados al azar. Entre las 20 características investigadas, se encuentran: ventas diarias (mill $); mts^2 de área; Número de vehículos que pueden parquear, información que se presenta a continuación.

PUNTOS:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
VENTA DIARIA: (mill. $)	146	162	148	172	165	148	150	140	170	162	184	174	165	170	190
METROS^2:	70	83	62	56	80	70	65	72	63	80	82	64	79	90	80
N° VEHICULOS:	12	18	20	16	15	12	26	20	18	23	26	20	30	25	22

Operaciones realizadas:

$\Sigma x_i = 2.446$ $\Sigma x_2 = 1.096$ $\Sigma x_3 = 303$ $n = 15$

$\Sigma x_t^2 = 401.778$ $\Sigma x_2^2 = 81.408$ $\Sigma x_3^2 = 6.507$

$\Sigma x_i x_2 = 179.293$ $\Sigma x_2 x_3 = 22.406$ $\Sigma x_1 x_3 = 49.761$

Con la información anterior, aplicando simbología y procedimientos por usted seleccionados, se pide:

a) Estimar el espacio requerido, si consideramos ventas diarias de $180 millones de pesos y un número promedio de 25 vehículos estacionados.

b) Calcular el coeficiente de correlación total o múltiple.

OPTATIVO: Con la anterior información, realizar la prueba al coeficiente de regresión, de la 3° variable, cuando se estima la 2° variable.

Síntesis de capítulo

El capítulo se dedica al análisis de *regresión corre ación*. Son dos aspectos que se deben considerar cuando se trabaja simultáneamente con *dos o más ariab es*. Con el primero se establece mediante un *mode o matemático* que nos permita representar el conjunto de observaciones a través de una *ínea recta parabó ica e ponencia* o de cualquier otro tipo. Una vez determinada la ecuación, se podrá estimar un valor o promedio de una de las variables en función de otra o más variables y, mediante la *corre ación*, técnicas que nos van a permitir establecer la fuerza, la intensidad de relación (correlación) entre esas variables.

La forma de visualizar mejor la información sobre la relación que puede haber entre dos variables, se obtiene mediante la elaboración de una gráfica denominada "*nube de puntos"* o *diagrama de dispersión* solamente cuando se trabaja con dos variables en forma simultanea

Dentro del desarrollo del tema, se calcula y explica una serie de medidas de gran importancia en el análisis de relación, tales como *coeficientes de regresión, arian a residua* , *arian a e p icada*, *error estándar de estimación*, *obtención de estimati os* y *límites de confianza*. Finalmente la aplicación de *pruebas de significación* tanto para el coeficiente de regresión, como para el de correlación.

13 CAPÍTULO

Algunos Métodos de Muestreo

Lo que hay que averiguar no son las respuestas correctas sino las preguntas correctas.

Peter Druc er

CONTENIDO

- Aplicaciones, términos, simbología, definiciones, conceptos generales.
- Muestreo aleatorio y no aleatorio.
- Muestreo aleatorio simple. Tamaño muestral. Estimaciones de promedios y totales. Límites de confianza. Error de muestreo.
- Muestreo estratificado. Asignación igual, proporcional y óptima. Tamaño de la muestra. Estimativos. Límites de confianza.
- Muestreo Sistemático de Conglomerados, Bietápico, Múltiple y Mixto.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Entender y manejar correctamente cada uno de los términos y símbolos que deberán ser utilizados en las investigaciones por muestreo.
- Conocer la importancia que tiene el muestreo no sólo como asignatura, en la investigación, sino también en el ejercicio profesional.
- Realizar pequeñas y medianas investigaciones, aplicando los conocimientos recientemente adquiridos, en cada una de las etapas del proceso.
- Saber interpretar los resultados obtenidos en las investigaciones.

GENERALIDADES

En la vida diaria nos encontramos con muchos ejemplos de aplicación del muestreo: exámenes de laboratorio ordenados por el médico; en la televisión y la radio, se dan resultados de encuestas de opinión; antes de comprar algún artículo comestible, se nos ofrece una pequeña porción para que decidamos su compra; en la casa reclamamos si está frío o salado el alimento, sólo con la prueba de una pequeña porción, por lo tanto no hay necesidad de probar la totalidad del producto para determinar el comportamiento de la característica.

Lo anteriormente expuesto y otros factores que veremos en el transcurso de esta unidad, hacen indispensable realizar una *in estigación parcia* y no total para llegar a conocer el comportamiento de un grupo.

Hemos mencionado la palabra *muestra* para describir una parte de la población y *muestreo* son las técnicas o herramientas utilizadas para la realización de una muestra. En la aplicación de estas técnicas es necesario conocer y manejar algunos términos indispensables en el desarrollo de una investigación, tal como se expondrán a continuación.

POBLACIÓN O UNIVERSO

Es un conjunto de unidades o e ementos ue presentan una característica com n; también se le considera como un conjunto de medidas. Si la característica observada ha sido medida, recibe el nombre de variable continua; si, por el contrario, tan sólo se hace el recuento se le denomina atributo o puede ser una variable discreta.

Considerar la población como conjunto de unidades o elementos, debe entenderse como un grupo de personas, familias, establecimientos, manzanas, barrios, objetos, etc., pero en realidad es un conjunto de medidas obtenidas de las características estudiadas. Por ejemplo, al considerar un grupo de personas de acuerdo a la finalidad de la investigación, se podrá estudiar su nivel educativo, sexo, ocupación, edad, preferencias, hábitos o cualquiera otra característica. Para un conjunto de personas que presentan las mismas características, se buscará información cuantificada, algunas veces mediante el recuento de unidades. En este caso particular, número de personas con determinado nivel educativo; número de hombres y/o mujeres; cuántas de ellas se encuentran trabajando, etc.; en otros casos la característica será medida (estatura, peso, ingresos, etc.).

No es correcto, entonces, considerar la *pob ación* o *uni erso* como si forzosamente estuviese constituida por la totalidad de habitantes de una zona geográfica o por un grupo más pequeño, para investigar características tales como estado civil, sexo, ocupación, posición económica, siempre y cuando se trate de seres humanos; se puede conformar otro tipo de población o universo diferente a personas, como establecimientos comerciales, industriales, servicios públicos, ubicados dentro de la zona o cobertura geográfica que se desea estudiar. Se pueden constituir poblaciones de animales, objetos, artículos producidos por una fábrica, etc.

UNIDAD Y ELEMENTO

La *unidad* es divisible, por ejemplo una familia se puede subdividir en personas que la conforman; una universidad se divide en facultades, a su vez puede considerarse como unidad, ya que se subdivide en

cursos y éstos en estudiantes. En cambio el *e emento* es indivisible, una persona, una cosa, etc, sigue siendo la forma más elemental de presentación, sin embargo, generalmente se habla en forma indistinta de elemento o unidad como una sola cosa.

La *Unidad* hace referencia a una persona, una familia, una vivienda, una manzana, un barrio, un establecimiento, una tarjeta o ficha académica, etc., y se denomina *unidad e ementa* o *e emento* cuando con ella obtenemos la información necesaria. Es el individuo, entidad u objeto, del cual deseamos observar todas o algunas de sus características para ser medidas o contadas. La unidad o elemento debe ser *c ara*, es decir de fácil identificación por todos los que intervengan en el proceso de la investigación; debe ser *adecuada* al objeto de la investigación, además, *mensurab e* y *comparab e*.

En la encuesta sobre la industria manufacturera, se define la unidad elemental estadística o de enumeración, como *a ue a unidad económica bajo un propietario o contro nico ue se dedica e c usi a o principa mente a una c ase de acti idad industria en una ubicación nica*.

En la encuesta de hogares, la unidad de selección final es la vivienda, en tanto que la unidad de observación está constituida por el hogar o grupos familiares que habitan en las unidades de vivienda, seleccionadas en la encuesta. En los hogares se establece distinción entre las diferentes unidades de consumo existentes, clasificando las familias en *principa es*, *secundarias* y *miembros sup ementarios*.

Una *unidad* o *e emento* de consumo podría ser una persona que vive sola, que atiende sus propios gastos de mantenimiento y que no depende económicamente de otra o de grupo familiar, o bien, varias personas que viven bajo el mismo techo relacionadas por factores monetarios, es decir, que aportan sus ingresos a un fondo común, que sirve para los gastos de la familia y para el consumo personal de cada uno de los miembros. La *fami ia principa* se compone en el caso más simple, por el jefe del hogar, su cónyuge y todos los hijos solteros que no reciben ingresos o, si reciben, realizan sus gastos del fondo familiar común.

La *fami ia secundaria* está compuesta por personas que no forman parte de la principal, pero comparten entre sí los gastos de alimentación. Habrá, por lo tanto, tantas familias secundarias como grupos adicionales de personas que compartan los gastos de alimentación entre sí.

Se considera también como *miembros sup ementarios* a las personas que comparten, al menos, los gastos de alimentos con una de las familias (principal o secundaria) aunque no comparten otro tipo de gastos o ingresos.

En conclusión, la unidad o unidades así como el elemento a considerar en la investigación deben ser bien *definidas*, *adecuadas*, *mensurab es* y *comparab e*s.

Unidades de Selección: son los diferentes grupos o subgrupos que pueden formarse de la población total o universo. Deben distinguirse las unidades parciales y finales de selección, las unidades de observación y las unidades o elementos de análisis.

Unidades Parciales de Selección: son aquéllas que se obtienen, como consecuencia del proceso de selección de las unidades que conformarán la muestra, constituyéndose en subdivisiones de la población

a través de la cual se llega a la unidad final de selección. Ejemplo, una ciudad para efectos de una investigación puede ser dividida en zonas o barrios. Estos, a su vez, pueden subdividirse en manzanas y estas últimas, en viviendas. La unidad final puede ser la familia, pero en cada una de !as subdivisiones se debe efectuar una selección, generalmente aleatoria.

Unidad final de selección: tiene características definidas de permanencia y puede ser identificada con facilidad en el transcurrir del tiempo. En el ejemplo anterior la unidad final de selección puede ser la familia, la vivienda, etc.

Unidades de observación: son aquéllas que, en conjunto, conforman la población o universo, dada una característica común.

Unidad de análisis: son aquellos elementos o unidades sobre las cuales se concentra el estudio, quienes suministran la información que luego va a ser analizada a fin de obtener conclusiones.

En la mayoría de los casos se trabaja con unidades de observación, como sucede en el *muestreo a eatorio simp e*. Sin embargo hay casos, como en la encuesta de hogares, donde se consideran varias etapas de selección (*muestreo mu tietápico*), que implica unidades de selección, tales como:

a. Selección de barrios.
b. Selección de manzanas.
c. Selección de viviendas.
d. Selección de familias.
e. La unidad de observación puede ser el jefe del hogar de quien debe obtenerse la información.
f. La unidad de análisis podría ser la persona, si el objeto de la investigación es por ejemplo el ingreso per-cápita.

POBLACIÓN FINITA E INFINITA

Dependiendo del número de *unidades* o *e ementos de obser ación*, la población puede ser considerada como *finita* o *infinita*.

Población infinita, conformada por un indeterminado número de unidades, tal es el número de peces en un río o el número de granos de arroz recolectados en una región. El comportamiento de una población demasiado grande, aún siendo finita, tiende a ser considerada como una población infinita al calcular el tamaño de la muestra, por ejemplo, el número de tachuelas producidas por las fábricas situadas en los países latinoamericanos.

Al determinar el t*ama o de a muestra*, en una población finita pero demasiado grande, el resultado no varía sustancialmente al establecido para la población infinita.

La *población finita* es aquélla constituida por un determinado o limitado número de elementos o unidades y en la mayoría de los casos, considerada como relativamente pequeña.

CARACTERÍSTICAS

Lo investigado en una unidad o elemento son sus características, las cuales se clasifican en *cua itati as* y *cuantitati as*.

Características Cualitativas: son *atributos*, susceptibles de ser expresados mediante palabras, (ocupación, profesión, actividad económica, nacionalidad, marcas, puertos de embarque, sexo, estado civil, propietario de vivienda, etc.). Sólo se pueden contar, no son mensurables.

Características Cuantitativas: son *ariab es* que se expresan numéricamente (ingresos, ventas, consumos, edades, estaturas, etc.).

Ejemplos de algunas unidades cuyas características (cualitativas y cuantitativas) pueden ser estudiadas, acorde con la unidad objeto de investigación:

UNIDAD	CARACTERÍSTICAS A ESTUDIAR	
Familia	Número de personas que la componen	Niveles de escolaridad
	Número de personas que trabajan	Ingresos y gastos
	Ocupación	Propiedad de la vivienda
	Número de personas que estudian	Datos generales de la vivienda
Establecimiento industrial	Producción bruta	Horas de producción
	Empleo	Consumo de energía
	Sueldos y salarios	Actividad económica
Hospital Pacientes según	Edad	Clase de consulta
	Sexo	Número de hospitalizaciones
	Estado civil	Clase de tratamiento
	Residencia	Ocupación

Las *características* se describen mediante utilización de *promedios*, *proporciones*, *ra ones* etc. Tales medidas reciben el nombre de *parámetros*, cuando son aplicadas a las características de las unidades en la población y *estimadores* en la muestra. Son resultados de estas mediciones el promedio de duración de hospitalización, el porcentaje de mujeres atendidas por maternidad, la relación mujeres/hombres, etc.

INVESTIGACIÓN TOTAL Y PARCIAL

Dependiendo del objetivo u objetivos que se persigan, puede realizarse una *in estigación e haustia* o una *in estigación parcia*

Exhaustiva: En este caso, se deben observar todas las unidades que constituyen la población o universo, objeto de la investigación. La enumeración total de una población en un tiempo dado, recibe el nombre de censo (censos de población, de la industria manufacturera, del sector agropecuario, etc.). El censo requiere de una organización enorme y la ocupación de un gran número de personas en cada una de las diferentes etapas de la investigación.

Algunos factores negativos del censo son:

- Demasiado *tiempo* requerido para realizarlo, lo cual impide obtener información antes de que se torne obsoleta.
- El *costo*, tanto durante la etapa de planeamiento, como en la de recolección, sistematización y publicación es bastante elevado, lo que impide que se realice con mayor frecuencia.
- Es muy frecuente que se presente el *error sistemático*, como consecuencia del volumen de unidades numeradas y la cantidad de características investigadas; el *error* aumenta a medida que se van efectuando las diferentes operaciones requeridas por el censo.

Un *censo* no puede realizarse cuando el elemento investigado se *destru e*. (control de calidad de un producto). Por ejemplo, al tratar de investigar la duración del total de las bombillas producidas por una fábrica, se tendría como resultado la inexistencia de unidades para la venta, por la destrucción misma del elemento.

En algunos casos no podrá obtener información especializada; por ejemplo, someter a todos los habitantes del país a un examen médico, cuando se quieren elaborar estadísticas sobre salud.

Sin embargo, el censo es necesario, cuando se necesiten datos locales para cada subdivisión geográfica del país y se desee una cobertura del 100%.

En resumen, un *censo* no se podrá realizar, cuando:

- La población sea infinita.
- Aún siendo finita, sea demasiado numerosa.
- El tiempo sea mayor al requerido.
- El costo de la investigación es superior a los recursos disponibles.
- La observación implica destrucción de los elementos (resistencia de materiales).
- La característica es homogénea, tal como en la prueba de alimentos, exámenes de sangre, orina; etc.

Las anteriores consideraciones, obligan a buscar un método más práctico para obtener información, en especial para poblaciones grandes. Este método consiste en realizar un estudio *parcia* de la población denominada *muestra* y tiene las ventajas de:

- Reducir costos y tiempo.
- Ser aplicable en poblaciones infinitas o muy grandes.
- Servir como forma práctica para no destruir la totalidad de !os elementos de la población. Sin embargo, la decisión entre *censo* y *muestra* depende, en la mayoría de los casos, de la finalidad del estudio.

Muestra o Investigación Parcial: se define como un conjunto de medidas pertenecientes a una parte de la población.

También se considera una parte de la población o subconjunto de elementos, que resultan de la aplicación de algún proceso, generalmente de selección aleatoria, con el objeto de investigar todas o parte de las características de estos elementos.

La selección de las unidades debe ser *a eatoria*, lo que equivale a decir probabilística, con el fin de obtener una muestra representativa de la población, respecto de todas o algunas de sus características, para las cuales el valor del parámetro es desconocido. En caso contrario se habla de una muestra *no a eatoria*.

MUESTREO NO ALEATORIO

Una *muestra no es a eatoria* cuando los elementos son elegidos por medio de métodos no aleatorios. Este procedimiento está sujeto a errores, ya que se confía al juicio subjetivo de seres humanos, como en el caso de selección dirigida (no aleatoria), donde las unidades son tomadas en forma caprichosa y en la mayoría de las veces por conveniencia; tal sucede en las encuestas de opinión que realizan los locutores deportivos, cuando se juega un partido de fútbol o en las entrevistas para televisión, radio o prensa, hechas en las principales vías de la ciudad, con el ánimo de conocer la opinión sobre algún asunto de interés para la comunidad.

La *muestra dirigida* está constituida por una parte de la población, que el investigador considera representativa del universo, como en el caso de seleccionar diez (10) familias ubicadas en un barrio de la ciudad, para que opinen sobre la seguridad o medidas de vigilancia en la zona. También pudo haberse hecho un llamado a las familias del barrio, para que se presenten voluntariamente a dar su opinión; en este caso el proceso se denomina *muestreo de o untarios*.

Si la característica es *homogénea*, la representatividad de la muestra puede ser satisfactoria; por ejemplo en un restaurante, una cucharada de sopa o un sorbo de vino; en un almacén, un trozo de tela o un par de zapatos, son representativos del total de sopa preparada, del vino para la venta o de la tela o calzado que dispone el establecimiento. Si se desea practicar examen de sangre a un paciente, el resultado será el mismo aunque la muestra varíe de 1 cm^3 a un litro.

Hay otros métodos de muestreo no aleatorio como: el *opinático* la unidad se selecciona en opinión del investigador; el *intenciona* , hay intención al seleccionar; *por cuotas*, en este caso se le dice al entrevistador cuántos hombres, mujeres, cuántos casados, solteros se deben entrevistar, es decir, se le fijan cuotas: de todas maneras es un muestreo sin validez, por esa razón se le llama *muestreo errático*, generalmente conduce a errores.

MUESTREO ALEATORIO

Una *muestra es a eatoria* cuando los elementos que constituyen la población o universo investigado, tienen la misma posibilidad de ser seleccionados.

Existen numerosos procedimientos en la aplicación del *muestreo a eatorio*. Los más importantes son:

1. Muestreo Aleatorio Simple: con reposición de las unidades, corresponde a una *distribución bino-*

mia ; sin reposición, a una *distribución hipergeométrica*. También se le conoce como *muestreo irrestricto*

El método de *muestreo a eatorio simp e* es recomendable, en especial, cuando la población no es numerosa y las unidades se concentran en un área pequeña; por otra parte, la característica no debe tener gran variabilidad, porque implicaría un tamaño muestral muy amplio lo que, a su vez, incrementaría costos y tiempo; por último, la población debe facilitar su enumeración para que permita la aplicación de este método.

- Se requiere un listado de unidades de la población, lo cual no es fácil, en muchos casos (marco).
- Si el área es amplia, es probable que haya necesidad de traslado a lugares lejanos, para investigar unas pocas unidades, dificultando su aplicación (costo y tiempo).
- No existe garantía de que todas las unidades queden representadas en la muestra, cuando la característica tiene una gran variabilidad.
- Si el coeficiente de variación es mayor del 30%, el tamaño de la muestra se va haciendo más grande.

Las anteriores observaciones conllevan a fijar tres inconvenientes que presenta el método de *muestreo a eatorio simp e*.

2. Muestreo Aleatorio Estratificado: se le denomina también como *muestreo a eatorio restringido* Este procedimiento implica una división de la población en grupos, denominados *estratos* en tal forma que el elemento presenta una característica tan definida que sólo le permite pertenecer a un único estrato. Por lo tanto, para que la división por grupos sea efectiva, los factores de estratificación deberán guardar estrecha relación con las características que se investigan y con el objetivo o finalidad del estudio. De esta manera se logra una mayor precisión en los resultados. Supongamos que se desea hacer una investigación sobre transporte urbano en una determinada ciudad, con el fin de establecer los costos de operación por pasajero. En este caso, una muestra aleatoria simple de 300 vehículos no nos da garantías de que los diferentes tipos de vehículos, buses, busetas, micros y colectivos queden bien representados en la muestra, si se tiene en cuenta que pueden haber grandes diferencias, no sólo en cuanto a capacidad del vehículo, sino en su recorrido. En este caso, es preferible establecer estratos o grupos mediante la clasificación según el tipo de vehículo.

En el *muestreo aleatorio estratificado*, como en el *muestreo a eatorio simp e*, se requiere una lista detallada de las unidades que constituyen la población a fin de elaborar los estratos o subpoblaciones, pero a pesar de lo dispendioso que puede resultar este procedimiento, hay garantía de que los elementos de la población queden bien representadas en la muestra.

Este método es más eficiente que el *muestreo a eatorio simp e*, pese al grado de heterogeneidad y dispersión que puede presentar la característica, con la ventaja de que al formar grupos más o menos homogéneos, la muestra resultante es más pequeña y representativa.

Mediante la selección aleatoria de los elementos, en cada uno de los estratos, se conformará la muestra. Dependiendo de la distribución o escogencia de los tamaños muestrales para cada estrato, podrá aplicarse alguno de estos tres procedimientos:

- **Afijación Igual o Asignación Igual:** se da cuando los elementos quedan asignados o repartidos por igual en cada estrato muestral.

- **Afijación Proporcional o Asignación Proporcional**: los elementos se distribuyen en los estratos muestrales, en la misma proporción en que se distribuyen los elementos en la población.
- **Afijación Óptima**: cuando el tamaño, tanto de la muestra general como para cada uno de los estratos muestrales, depende del grado de variabilidad de la característica en cada estrato y del costo mínimo para una precisión dada.

Se debe tener en cuenta que: si el *coeficiente de variación* ($S\ \bar{}$)100 es menor o igual al 30%, debe realizarse el *muestreo a eatorio simp e*; si por el contrario es mayor a dicho valor, es decir el tamaño de la muestra crece nos veremos obligados a aplicar el *muestreo aleatorio estratificado*, con el cual disminuiremos tiempo, costos y además, logramos una mayor precisión en la estimación.

3. Muestreo sistemático: más que un método de muestreo, es considerado como un método de selección, al que algunos denominan *métodos de se ección a inter a os regu ares*. Se aplica cuando la característica a investigar se encuentra ordenada de acuerdo al valor, tiempo, cantidad, etc.

Prácticamente sobre los tres métodos anteriores, (*muestreo a eatorio simp e*, *muestreo a eatorio estratificado* y *muestreo sistemático*) se basa este capítulo, sin embargo es necesario que el lector sepa que existen otros métodos, no estudiados con amplitud, dado lo extenso para ser presentados.

4 Muestreo Doble: denominado también *bifásico* porque se efectúa en dos fases. Es aplicado de preferencia, cuando no existe información auxiliar que permita conocer los tamaños poblacionales de los estratos, ni la identificación de las unidades; en este caso, primero se realiza una muestra aleatoria simple, generalmente grande, en forma rápida y sencilla para conocer en forma muy general algunas de las características objeto del estudio; luego se procede a una segunda muestra extraída de la anterior, la que podría considerarse como submuestra con la cual se hacen las estimaciones. Si se utilizan más de dos fases, se habla de *muestreo m tip e* o *po ifásico*.

5 Muestreo por conglomerados Áreas o Etapas: en los métodos anteriormente expuestos es necesario contar con un listado de unidades finales o unidades elementales de la población. El *muestreo por cong omerados* se utiliza cuando no existe listado, o bien las unidades están demasiado dispersas y no permiten llegar directamente a las fuentes de información.

Esto nos obliga a sustituir las unidades físicas o elementales, por grupos de unidades, que llamaremos *cong omerados*; generalmente son superficies o áreas en las que se ha dividido el espacio ocupado por la población. Supongamos que se desea realizar estudios a familias en una ciudad. Cada unidad o familia se constituye en un conglomerado, por tal razón se aplica el *muestreo a eatorio monoetápico*, es decir, se realiza la investigación en una sola etapa.

Si iniciamos con una selección de manzanas y de cada una de estas manzanas, se selecciona otra muestra correspondiente a familias, el método será *bietápico*. Obsérvese que hay una modificación con respecto al anterior método, ya que los elementos no forman parte de los conglomerados, sino que son una muestra o sub-muestra de cada uno de ellos.

Si hacemos una selección de barrios, luego de manzanas y por último de familias, el método será *trietápico*. *E muestreo po ietápico o mu tietápico* es una generalización y consta de más de tres etapas.

Los *cong omerados* son unidades que contienen unidades o elementos, es de anotar, que en la aplicación del *muestreo por etapas* la unidad cambia en cada una de ellas, por lo tanto se requiere de varios marcos o listados, ya que se tienen varias poblaciones.

6. Sub-muestras Interpenetrantes: en este método las submuestras, cada una con un número de unidades, son seleccionadas aleatoriamente de la población. Este método es utilizado en general para medir la concordancia entre los resultados obtenidos de muestras sucesivas.

7. Métodos Mixtos: son el resultado de la combinación de algunos o de todos los métodos anteriormente mencionados.

MARCO

En la selección de unidades se requiere de un listado, que en algunos casos debe prepararse, en otros, se actualiza. Este listado debe corresponder al total de unidades o elementos que integran la población a investigar o a seleccionar que, a su vez, constituyen el *marco de referencia* o *marco muestra* correspondiente a la población objetivo. Por ejemplo, la nómina de pago, ya sea para un grupo de trabajadores en una fábrica o de empleados en una empresa, podría ser el marco; otro ejemplo podría ser el listado de matrícula de los estudiantes en una universidad.

Sin embargo, el *marco* no siempre estará representado por un listado, podrá ser también un mapa o fotografías aéreas. En el *marco* no deben aparecer unidades falsas; ejemplo: en el listado de matrículas no pueden figurar aquellos estudiantes que se han retirado por diferentes causas, pero que se alcanzaron a matricular.

En el *muestreo por etapas m tip es*, debe existir un marco para cada etapa. El marco en cada etapa, debe ser lo suficientemente detallado en cuanto a la identificación de sus unidades, de tal manera que si una de ellas es incluida en la muestra, pueda ser ubicada e investigada rápidamente, sin que ello implique especificar la información que debe suministrar cada unidad.

Rara vez los marcos son perfectos y lo común es que sean inexactos, inadecuadamente descritos, desactualizados o con unidades duplicadas; por tal motivo, deben ser revisados y actualizados de manera que permitan una fiabilidad razonable por cuanto la muestra se basa en ellos.

FALSEAMIENTO DEL ESQUEMA DE SELECCIÓN

El buen resultado de una investigación por muestreo, no obedece solamente al empleo de un acertado método (la fijación del tamaño, el método de selección y la estimación), sino que se requieren otras etapas en la investigación, complementarias al buen éxito del trabajo; algunas de ellas son: la *e aboración de formu ario*, *instrucciones*, *entrenamiento de persona entre istador* y la necesidad de una buena dosis de coordinación en el trabajo de campo.

Por ejemplo, el entrevistador juega un papel muy importante en la recolección de información y en el resultado final. De allí la necesidad de hacer comprender al personal, ciertos conceptos básicos de muestreo para que no se introduzcan vicios. Dentro de estos conceptos son de gran importancia la claridad sobre la unidad de observación; valga la pena mencionar la encuesta de hogares, donde se debe conocer el segmento (unidad de muestreo) por medio de mapas o croquis; el entrevistador debe saber

determinar las viviendas que son seleccionables, (unidades de enumeración) y en cada vivienda seleccionada debe saber determinar a qué personas va a pedir información (unidades de observación). Debe existir una estrecha comunicación y coordinación entre el *persona técnico* y el *persona de campo* o recolectores. Fallas que se producen durante el trabajo de campo, como la localización y la entrevista de las unidades seleccionadas, ocasionan *fa seamiento en e es uema de se ección*. Veamos una de las tantas fallas que pueden ocurrir en el trabajo de campo y que se refieren a casos de unidades no entrevistadas, que algunos las clasifican en categorías, cada una de las cuales produce un efecto negativo en los resultados de la encuesta y requieren de una acción para ser corregida.

- El *informante se rehusa* a ser entrevistado quizás porque considera que la encuesta no tiene importancia, sospecha que el entrevistador sea un farsante, motivos políticos o religiosos, alega estar muy ocupado o estar cansado de participar en innumerables investigaciones que hacen tanto entidades públicas como privadas. Muchos de estos pretextos pueden eliminarse mediante la publicidad, carnet de identificación, acordando la hora de entrevista a conveniencia del informante y en especial, motivando al entrevistado sobre la importancia de la investigación.
- El *informante no puede ser entre istado*, tal es el caso de que no esté, el establecimiento o la vivienda se encuentren cerrados en el momento de hacer la entrevista, incapacidad física o mental del informante, etc. Algunos inconvenientes pueden eliminarse mediante nueva visita.
- El *informante no debe ser entre istado*. El establecimiento fue cerrado definitivamente por liquidación, la dirección de la unidad se incluyó erróneamente en el marco de la muestra, la casa está ocupada por personas que no son elegibles, sólo hay un lote o la edificación fue demolida, etc. También puede suceder que la unidad localizada no corresponde al objeto de la investigación.

¿Qué se puede hacer, en especial, en estos últimos casos?

Algunos entrevistadores toman la decisión de sustituir la unidad no encuestada, por la vecina, lo cual tiene como consecuencia una probabilidad de cero para la que debió ser investigada, mientras que la unidad vecina tendrá una probabilidad mayor que las otras unidades.

De acuerdo con lo anterior, se considera que en teoría, no debe existir sustitución de unidades, para no falsear el esquema de selección; por tanto, a las unidades de las cuales se obtuvo información se les considera como *dominio de estudio*, ya que las no investigadas dejarán de pertenecer al grupo de nuestro interés.

SUSTITUCIÓN DE UNIDADES

El concepto universalmente aceptado, es el de no sustituir la unidad que no haya informado; en la práctica se acepta la aplicación de algunos *métodos de sustitución*, sin que sus resultados sean objetados. Algunos de ellos se presentan a continuación:

- Se hace *sustitución*, cuando en un segmento se encuentra, por ejemplo, una vivienda no entrevistada; en este caso, se toma la vivienda siguiente de acuerdo al orden de visita o de selección y la información obtenida se duplica. Este proceso se basa en el supuesto de que,

dentro de un mismo segmento, las familias que ocupan viviendas contiguas tienen una alta probabilidad de poseer características económicas o sociales similares.

- Otro método de *sustitución*, consiste en determinar el número de unidades no entrevistadas, ya sea en la muestra, en el segmento, estrato, etc., luego se selecciona una muestra aleatoria dentro de las que han sido encuestadas, equivalente al número de las que no respondieron y se duplica la información.
- El más común, consiste en tomar una muestra mayor a la establecida, mediante fórmula, conocida como tamaño óptimo. El número mayor depende del conocimiento que se tenga sobre las unidades que no tendrán información.
- Seleccionar al azar un número igual a las unidades que no informaron, de la población que no fue seleccionada en la muestra.

A pesar de los esfuerzos para reducir la proporción de casos no entrevistados, siempre quedará un residuo en todas las encuestas. En este residuo, deben preocuparnos aquellas unidades que debieron haberse entrevistado, tales como: nadie en casa, negativas, etc.; por tal razón se deben establecer medidas para reducir la resistencia a conceder las entrevistas. Los avisos en la prensa, radio y televisión, son ayudas para hacer consciente al público de la importancia de la investigación, también es usual el envío de cartas de introducción, con anticipación a la visita del entrevistador; otro método, a través de los mismos entrevistadores, consiste en distribuir material ilustrativo sobre la encuesta, antes de que la persona sea entrevistada.

DOMINIO DE ESTUDIO

Se dijo anteriormente que en el proceso de selección no debe haber sustitución de unidades muestrales; si por algún motivo la unidad seleccionada no puede localizarse y no se sustituye, se va a tener un número de elementos o unidades, (con información) inferior a la del tamaño óptimo. Si realizamos el análisis con estas unidades, corresponderá al denominado *dominio de estudio* y los resultados no se consideran confiables por haber trabajado con un número inferior al tamaño óptimo. Por ejemplo, inicialmente se estableció una muestra de 500 establecimientos, de los cuales, por diversas razones, 80 no respondieron. Deberá considerarse a los 420 como tamaño de la muestra en lugar de los 500 calculados. Las 80 unidades, disminuidas por quienes no informaron, son un subconjunto de la población original que deberían ser tomadas en cuenta, con el fin de ser válida la estimación realizada.

También se denomina como *dominio de estudio*, cuando se analiza parte de los elementos en la muestra que presentan algunas características especiales; por ejemplo, se realiza una investigación por muestreo en un barrio de la ciudad y entre las características que se investigan está el de ser propietario o no de vehículo; pero luego, en el análisis de la información tan solo nos interesa estudiar el ingreso de las familias propietarias de vehículo; en este caso se constituye una subpoblación de familias con ingresos, las cuales, además requieren ser propietarias de vehículo.

MÉTODOS DE SELECCIÓN

Existen diferentes métodos de selección al azar, tales como:

Selección con reempla o: una vez determinado el tamaño de la muestra, se deberá establecer cómo se hará la *se ección de as unidades*. Inicialmente la selección puede hacerse con o sin repetición. En el

primer caso el elemento podrá ser seleccionado más de una vez, entonces se habla de una *muestra a eatoria con reemp a o*; en el segundo, el elemento seleccionado no podrá ser tomado en cuenta nuevamente.

El método de *se ección con reemp a o*, es aplicado generalmente en el muestreo por conglomerados y en el submuestreo, siendo importante anotar que las diferencias en precisión entre una selección con reemplazo y otra sin el, usualmente son de escasa importancia; de ahí que en el desarrollo de los diferentes métodos de muestreo que se estudiarán en esta unidad, se utilizará la selección sin repetición, es decir, cada unidad sólo podrá seleccionarse una vez.

Selección por sorteo: consiste en identificar, a cada una de las unidades de la población mediante números consecutivos; se pueden utilizar tarjetas o balotas que lleven el número; luego se introducen en un recipiente, se mezclan y se extraen de una en una, hasta completar el tamaño de la muestra.

Selección a intervalos regulares o sistemático: este método consiste en determinar un *inter a o* dividiendo el tamaño poblacional por el tamaño muestral. Dentro de este intervalo, establecido con las primeras unidades de la población, se selecciona aleatoriamente una unidad, luego se le va sumando el valor del intervalo al número resultante; por ejemplo: si el intervalo es de 10 y el número seleccionado (entre 01 y 10) es 8, los siguientes números serán 18, 28, ..., etc. Este método se denomina *se ección sistemática*.

Uso de las tablas de números aleatorios: es común la utilización de tablas que muestran conjuntos de dígitos agrupados, tanto en sentido vertical como horizontal, de acuerdo a una selección aleatoria con repetición de una población de 10 dígitos (0 a 9), construida de tal manera que permita garantizar la aleatoriedad en la selección de las unidades. La manera como se elabora una tabla, en la forma más elemental, consiste en numerar diez balotas, tarjetas o fichas de 0 a 9, se introducen en una urna y luego se sacan una a una, teniendo en cuenta que al extraer una, se lee, se anota el número y se regresa a la urna, formando cifras de dos o mas dígitos.

Observemos la Tabla VI, que aparece al final del libro, su importancia está en el uso; se podrá leer en cualquier orden, en columnas hacia arriba o hacia abajo, en filas hacia la derecha o hacia la izquierda. Las cifras podrán ser dos o más dígitos, dependiendo del tamaño poblacional; por ejemplo, si la población no es superior a 999 serán cifras entre 001 y 999. Si la población es de 400 elementos, cualquier número aleatorio seleccionado que sea mayor a 400 no se tomará en cuenta, sólo se seleccionarán aquellos números comprendidos entre 001 y 400.

En resumen, para la utilización de las tablas se debe proceder así:

- Se escoge al azar, una página cualquiera de las tablas.
- Se determina un punto de arranque al azar, es decir, sin mirar la página en el punto donde se coloque el dedo.
- La población deberá estar enumerada desde 1 hasta 9; desde 01 hasta 99; desde 001 hasta 999 y así sucesivamente, de acuerdo al tamaño de la población.
- Se sigue una misma dirección de selección, en columnas, filas, hacia arriba, abajo, a la derecha o izquierda.
- Se debe omitir aquel número que no corresponda a la enumeración de la población y continuar hasta obtener los números deseados.

- En casos especiales, cuando no se tenga una lista completa de las unidades, es preferible utilizar otras técnicas de muestreo.

USO DE LA CALCULADORA:

la calculadora CASIO Fx 3.600P de bolsillo o cualquiera otra, permite la selección aleatoria. Su uso es sencillo, solo requiere oprimir primero la tecla **INV** en otras **SHIFT**, luego la tecla **•** que aparece identificada como **RAN**, apareciendo en pantalla por ejemplo la cifra 0.759. Este procedimiento no es aplicable en poblaciones mayores a 999, por lo tanto serán utilizados calculadoras que permiten obtener un mayor número de dígitos. Si se trata de poblaciones de dos dígitos, se deben utilizar los dos últimos o los dos primeros dígitos que aparecen en pantalla. No se tiene en cuenta (para ningún caso) el 0, que antecede al dígito.

APLICACIÓN DE EXCEL PARA SELECCIÓN ALEATORIA

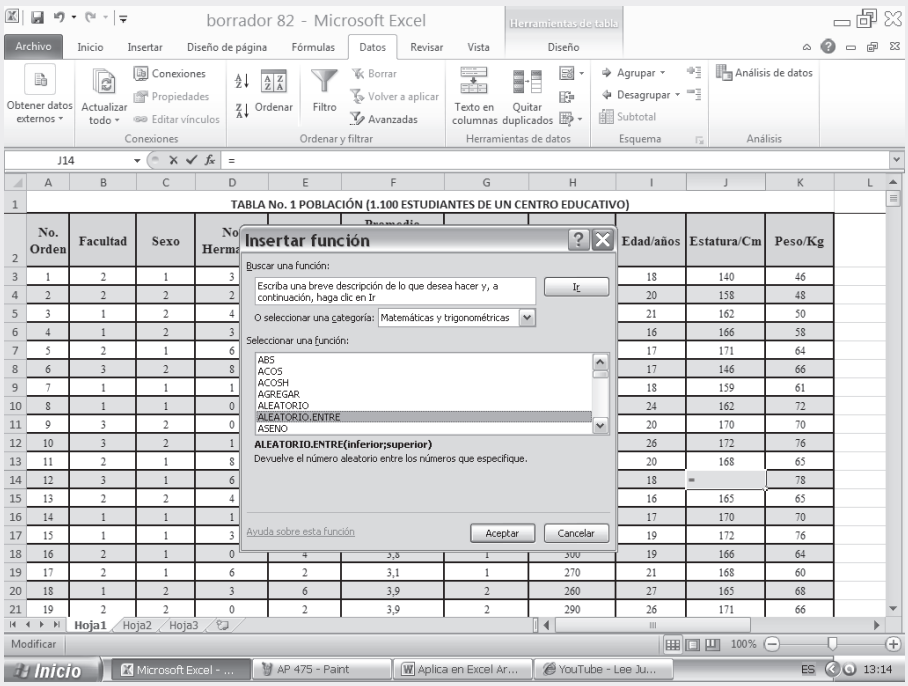
Complemento virtual SIL

A. PRIMER MÉTODO DE SELECCIÓN ALEATORIA PARA 50 DATOS.

Trabajando con fx(función).

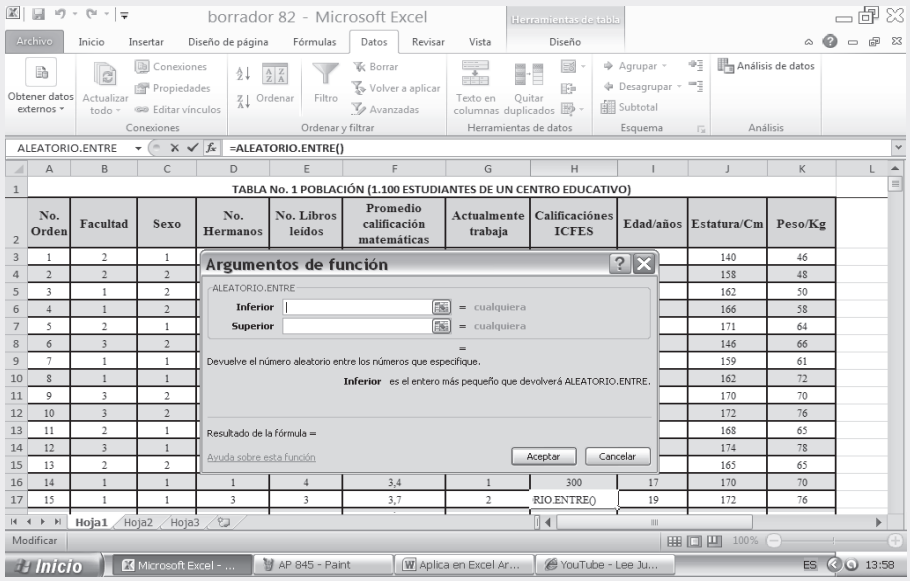
- En una hoja de cálculo seleccionamos la opción función fx, con el cual aparece un cuadro de diálogo correspondiente a INSERTAR FUNCIÓN.
- Observemos dentro del cuadro de diálogo, el campo CATEGORÍA DE LA FUNCIÓN, en el cual seleccionamos la opción MATEMÁTICAS Y TRIGONOMÉTRICAS y en el espacio correspondiente a SELECCIONAR FUNCIÓN, señalamos la función ALEATORIO ENTRE.

Figura No.1. Insertar Función.



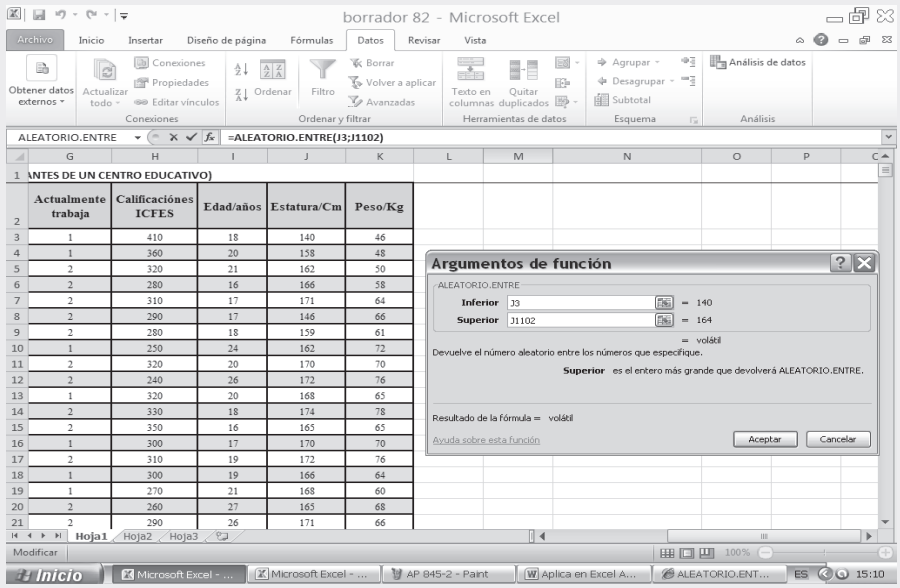
- Al hacer CLIC en ACEPTAR, nos aparecerá la siguiente figura:

Figura No. 2. Aleatorio entre (AB)



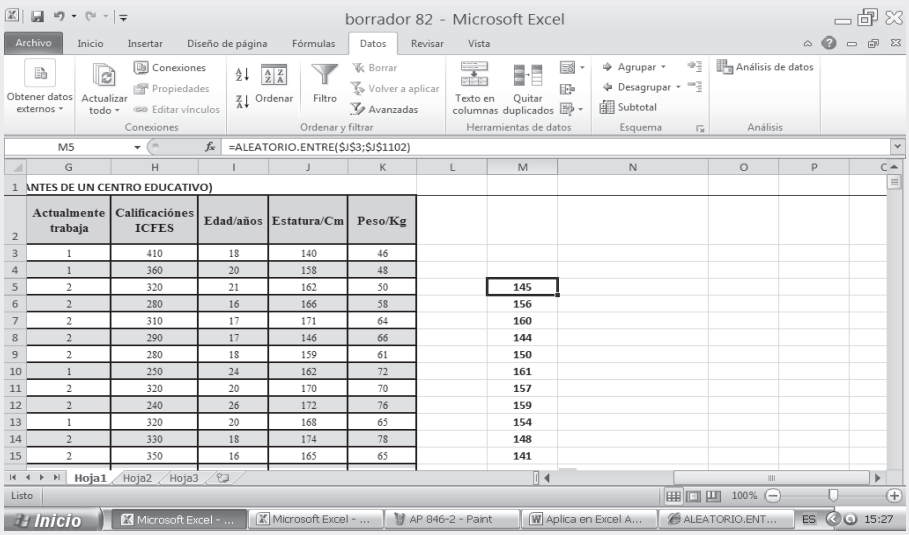
- Para la selección de una muestra aleatoria, vamos a tener en cuenta la información que se presenta en la Tabla No. 1 (ver SIL). Si por ejemplo vamos a evaluar la estatura de 1.100 estudiantes de un colegio, cuyos resultados se observan en la celda J, consideremos como el valor INFERIOR, la celda J3 y como el valor SUPERIOR, la celda J1102 que podría seguir nuestro número aleatorio. Bajo este intervalo, esta fórmula nos arrojará un número al azar entre estos valores. Es importante saber que los dos números tanto el máximo como el mínimo pueden aparecer en el resultado.

Figura No.3. Cálculos Aleatorios Entre.



- Dado que este resultado solamente arroja un número aleatorio entre los 1.100 estudiantes, y queremos una muestra significativa de mínimo 50 datos, seleccionamos la casilla donde se generó el resultado y desplegamos hasta conseguir un intervalo continuo de 50 datos, tal como se presenta en la figura siguiente:

Figura No. 4. Resultados Muestra Aleatoria.

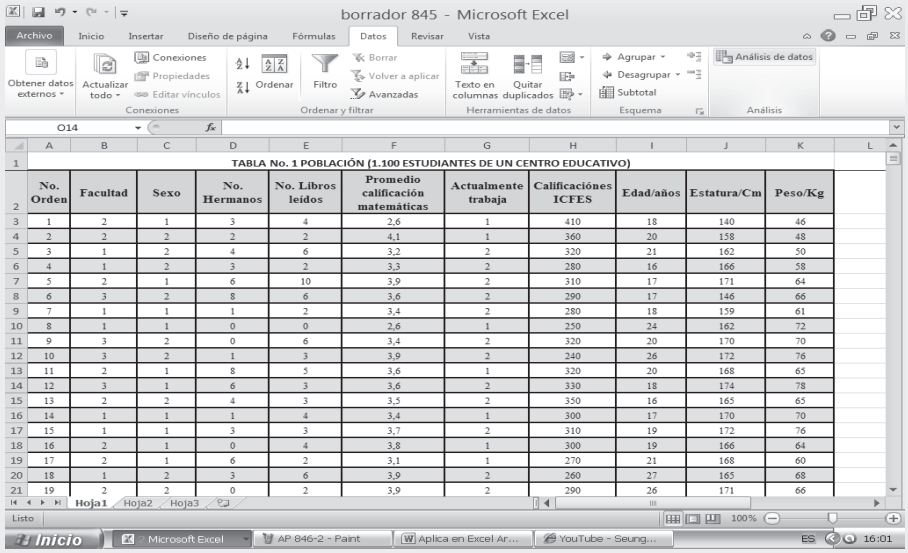


B. SEGUNDO MÉTODO DE SELECCIÓN ALEATORIA.

Para efectos de realizar la selección aleatoria, nos ubicamos en la hoja electrónica de cálculo y procedemos de la siguiente manera:

- Hacemos CLIC en la ficha DATOS y seleccionamos la opción ANÁLISIS DE DATOS.

Figura No.5. Análisis de Datos



- Al presionar esta tecla aparece el siguiente cuadro de diálogo sobre el cual seleccionamos la opción GENERACIÓN DE NÚMEROS ALEATORIOS tal como se presenta en la figura siguiente:

Figura No.6. Insertar Función.

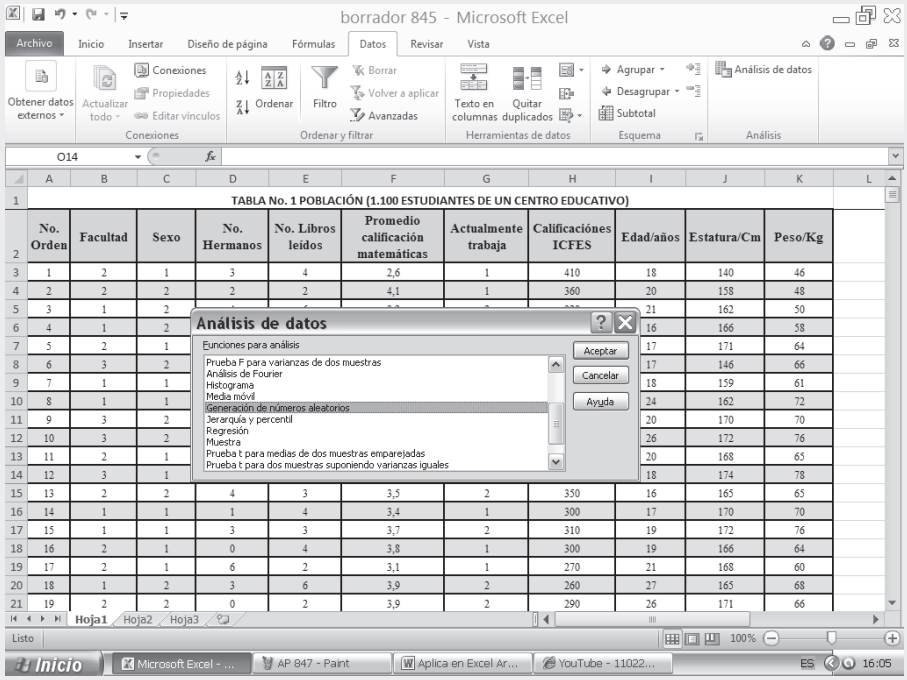


TABLA No. 1 POBLACIÓN (1.100 ESTUDIANTES DE UN CENTRO EDUCATIVO)

No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
3	1	2						21	162	50
4	1	2						16	166	58
5	2	1						17	171	64
6	3	2						17	146	66
7	1	1						18	159	61
8	1	1						24	162	72
9	3	2						20	170	70
10	3	2						26	172	76
11	2	1						20	168	65
12	3	1						18	174	78
13	2	2	4	3	3,5	2	350	16	165	65
14	1	1	1	4	3,4	1	300	17	170	70
15	1	1	3	3	3,7	2	310	19	172	76
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
18	1	2	3	6	3,9	2	260	27	165	68
19	2	2	0	2	3,9	2	290	26	171	66

En el cuadro de diálogo que aparece, digitamos lo siguiente:

- NÚMERO DE VARIABLES, si se requiere de todas las columnas o variables correspondientes a la información dada en la Tabla No. 1, las cuales son, por ejemplo 10, entonces el valor que debe colocarse en la primera casilla es 10, y la tabla de salida tendría el mismo número de columnas correspondiente. Para este ejemplo tendremos en cuenta solamente una variable correspondiente a ESTATURA, por tanto digitaremos 1 en este campo.
- CANTIDAD DE NÚMEROS ALEATORIOS, en este campo se digitará el número aleatorio requerido, en nuestro caso será 50.
- DISTRIBUCIÓN, en esta celda se presentan varias opciones (uniforme, normal, Bernoulli, binomial Poisson, frecuencia relativa, y discreta) para nuestro ejemplo seleccionaremos la opción UNIFORME.
- Una vez escogida la opción UNIFORME debe aparecer un espacio que indica PARÁMETROS, en esta casilla tenemos en cuenta los datos utilizados en el ejemplo anterior, sólo que esta vez tecleamos el número 140 y 164 correspondientes a los valores de la primera y última observación.
- INICIAR CON; en ésta se fija un número cualquiera con el cual se inicia la selección.
- Finalmente en OPCIONES DE SALIDA activamos la opción RANGO DE SALIDA y digitamos L3 que corresponde a la celda donde se presentarán los resultados.

Figura No.7. Generación de Números Aleatorios.

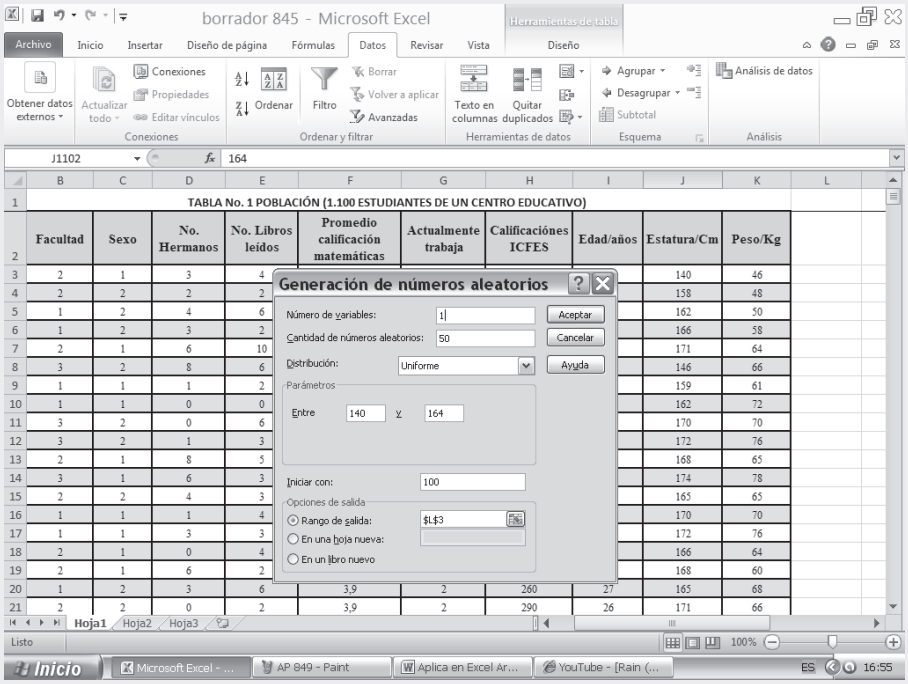
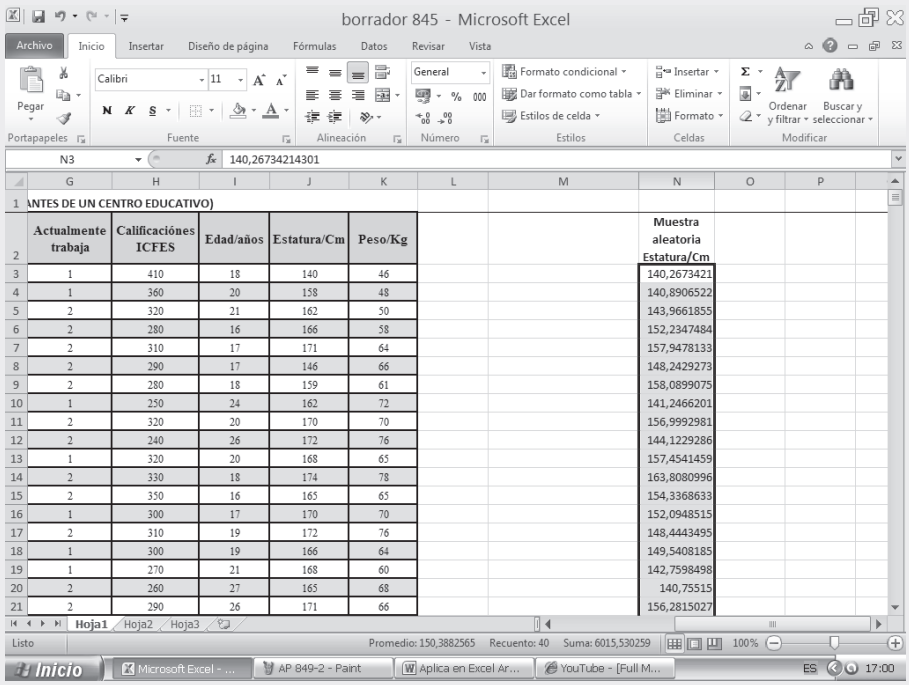


Figura No.8. Resultado Generación de Números Aleatorios.

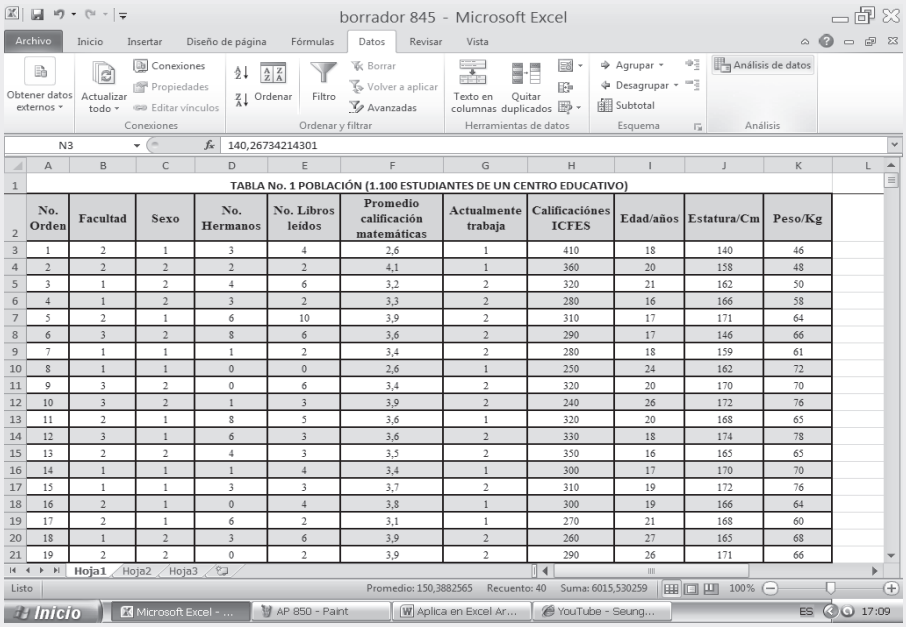


Nota: Cuando ya se hayan introducido los datos que deseamos estimar, se puede elegir entre diferentes opciones de salida según las necesidades y el informe que se vaya a presentar. No obstante debemos agregarle las leyendas a cada una de las variables introducidas.

C. TERCER MÉTODO DE SELECCIÓN ALEATORIA.

- ❑ Para el desarrollo de este nuevo método, nos ubicamos nuevamente en la hoja de cálculo y realizamos las siguientes operaciones:
- ❑ Hacemos CLIC en la ficha DATOS y seleccionamos la opción ANÁLISIS DE DATOS

Figura No.9. Análisis de Datos.

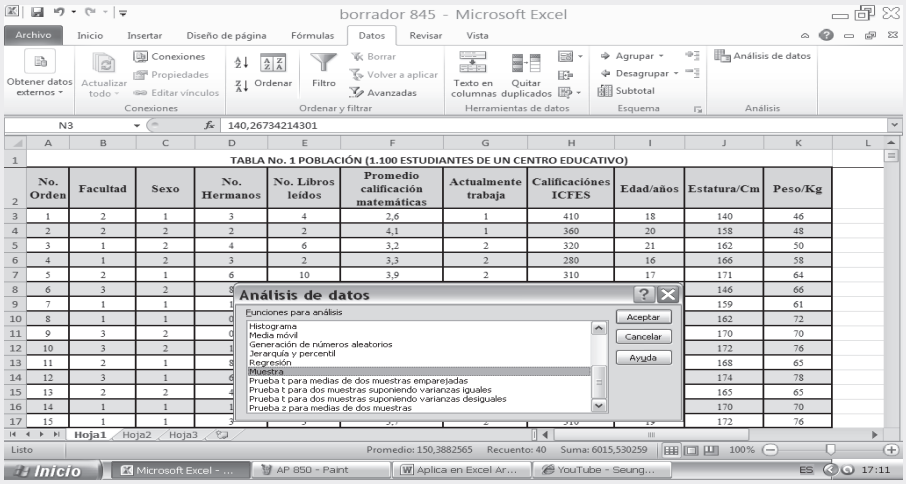


No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
3	1	2	4	6	3,2	2	320	21	162	50
4	1	2	3	2	3,3	2	280	16	166	58
5	2	1	6	10	3,9	2	310	17	171	64
6	3	2	8	6	3,6	2	290	17	146	66
7	1	1	1	2	3,4	2	280	18	159	61
8	1	1	0	0	2,6	1	250	24	162	72
9	3	2	0	6	3,4	2	320	20	170	70
10	3	2	1	3	3,9	2	240	26	172	76
11	2	1	8	5	3,6	1	320	20	168	65
12	3	1	6	3	3,6	2	330	18	174	78
13	2	2	4	3	3,5	2	350	16	165	65
14	1	1	1	4	3,4	1	300	17	170	70
15	1	1	3	3	3,7	2	310	19	172	76
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
18	1	2	3	6	3,9	2	260	27	165	68
19	2	2	0	2	3,9	2	290	26	171	66

- ❑ Esta vez nos trasladamos a la opción MUESTRA, hacemos CLIC en ACEPTAR y aparece la siguiente figura:

- Esta vez nos trasladamos a la opción MUESTRA, hacemos CLIC en ACEPTAR y aparece la siguiente figura:

Figura No.10. Insertar Función.



Una vez elegida la función, se deben incluir los datos de la muestra como lo presenta el siguiente cuadro de diálogo.

- RANGO DE ENTRADA, es importante saber que este método sólo admite un rango de entrada, que para el caso, será el rango J2:J1102 correspondiente a la ESTATURA.
- Activamos la Opción RÓTULO.
- Escogemos el método ALEATORIO digitando, en la opción NÚMERO DE MUESTRAS, el valor de 50 que corresponde al valor de la muestra que estamos utilizando.
- Finalmente seleccionamos alguna de las OPCIONES DE SALIDA que nos presenta, en nuestro caso activamos la opción RANGO DE SALIDA digitando la celda M3.

Figura No.11. Datos de Muestra. (Pantallazo adicional)

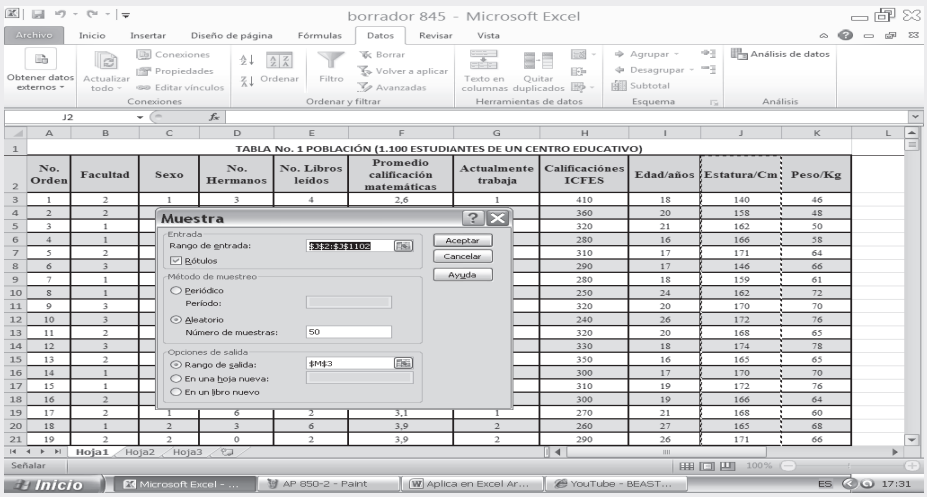
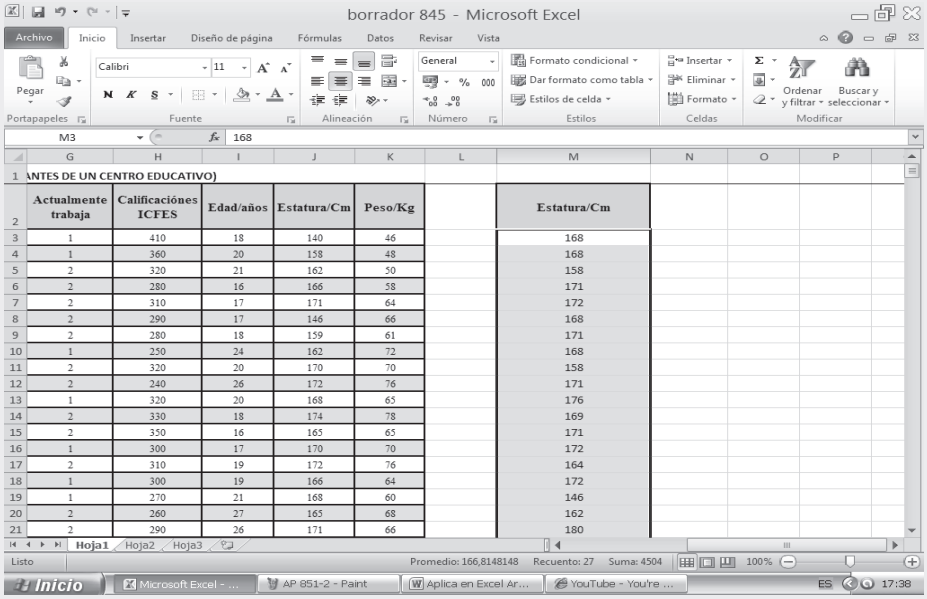


Figura No.12. Resultados de la Muestra. (Pantallazo adicional)



	G	H	I	J	K	L	M	N	O	P
1	ANTES DE UN CENTRO EDUCATIVO)									
2	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg		Estatura/Cm			
3	1	410	18	140	46		168			
4	1	360	20	158	48		168			
5	2	320	21	162	50		158			
6	2	280	16	166	58		171			
7	2	310	17	171	64		172			
8	2	290	17	146	66		168			
9	2	280	18	159	61		171			
10	1	250	24	162	72		168			
11	2	320	20	170	70		158			
12	2	240	26	172	76		171			
13	1	320	20	168	65		176			
14	2	330	18	174	78		169			
15	2	350	16	165	65		171			
16	1	300	17	170	70		172			
17	2	310	19	172	76		164			
18	1	300	19	166	64		172			
19	1	270	21	168	60		146			
20	2	260	27	165	68		162			
21	2	290	26	171	66		180			

***Nota*:** una vez usted obtenga el rango de entrada, recuerde que por este método sólo es posible tomar solamente un rango, para mayor manipulación de la muestra debe ordenarlo en forma ascendente y luego ir a la tabla base donde están los 1.080 elementos que constituyen la población. En la segunda parte del CD que acompaña al libro encontrará en conjunto toda la aplicación del EXCEL a la estadística, que al igual al capítulo 13, ofrece una base de datos para trabajar.

OBJETO DEL MUESTREO ALEATORIO

El objeto del *muestreo a eatorio* es seleccionar una parte representativa de la población, con el fin de obtener estimadores de los parámetros. Para lograrlo existen varios métodos y en cada uno de ellos se puede establecer:

a. El grado de precisión requerido en los estimadores.

b. Tamaño de la muestra.

c. Costo y tiempo.

Estimadores: el estimador es una medida que describe una determinada característica de la muestra, obtenida mediante la aplicación de promedios, proporciones, tasas, razones, varianzas, desviaciones típicas, etc.

Se les llama *estimadores*, ya que se les considera como el mejor valor que sustituye a los valores numéricos poblacionales, generalmente desconocidos, a los que se llega por medio de datos proporcionados por la muestra, en tanto que un *parámetro* describe una determinada característica de las unidades en la población.

Los *estimati os* obtenidos a través de una muestra pueden considerarse como el comportamiento de esa característica en toda la población. La estimación de los *parámetros* se hace a partir de los estadígrafos calculados en una muestra, con los cuales se obtienen conclusiones, que permiten, a la vez, la realización de *pruebas de hipótesis*, proceso denominado *inferencia estadística*.

Si para cada parámetro desconocido se calcula una cifra única, la estimación se denomina *puntua* o *estimación de punto*

Si se establece un intervalo o rango de valores dentro de los cuales se espera encontrar el parámetro, con cierto grado de confianza, se le denomina estimación de intervalos. Los límites que definen el intervalo se conocen como *límites de confianza*.

Un buen estimador debe reunir cuatro condiciones o propiedades:

- *Insesgado*: se dice que un estimador es insesgado, si el valor esperado del mismo es igual al parámetro de la población que estima.
- *Eficiente*: se refiere a lo cerca que se encuentre el valor estimado del parámetro.
- *onsistente*: se obtiene cuando el tamaño de la muestra se incrementa en tal forma que la varianza disminuya, siendo menor la diferencia entre el valor real y el estimado.
- *Suficiente*: es un estimador que utiliza toda la información que posee una muestra sobre el parámetro que se estima.

Tamaño de la muestra: una de las decisiones más importantes que se debe tomar en una investigación por muestreo, consiste en la determinación del *tama o óptimo*, simbolizado por *n*. Se dice que una muestra debe ser pequeña, con el fin de que el costo de la investigación sea lo suficientemente bajo, y bastante grande para que el error del muestreo sea admisible. Algunos creen que el tamaño de la muestra crece indefinidamente a medida que aumenta el tamaño poblacional, simbolizado por N, pero no es cierto, ya que existe un punto en el cual el tamaño de la muestra permanece constante, así el tamaño poblacional aumente.

En el cálculo del *tama o óptimo*, se deben tener en cuenta, los siguientes componentes:

E Error de Muestreo: es la diferencia que puede haber entre el valor poblacional (parámetro) y la estimación de la misma (estimación puntual o estadígrafo), obtenida por medio de una muestra aleatoria, observada en una de las tantas muestras posibles de una población dada. La totalidad de estos errores, genera la distribución de muestreo empleada para estimar el valor poblacional.

$$\pm E = \bar{} - \mu$$
$$\mu = \bar{} \pm E$$

$\bar{}$ = media muestral
μ = media poblacional
E = error de muestreo

En muchos casos, el error será determinado por el investigador; por ejemplo, si en un sector de la economía se requiere estimar el promedio de salarios, y se sabe por observaciones aisladas o por experiencia, que el ingreso promedio para ese sector es aproximadamente, de $780.000, el error respecto al parámetro podría ser: $780.000 ± E. De acuerdo con la finalidad de la investigación podría considerarse, para este caso, que un error de ± $60.000 o cualquier otro valor es más que suficiente, dado el grado de

variabilidad que supuestamente tengan los salarios en dicho sector. Sin embargo, este procedimiento requiere, además de cierta experiencia, un conocimiento previo sobre el comportamiento de la característica en la población que se estudia, de ahí que es mucho más práctico determinar el error como porcentaje, en la mayoría de los casos no mayor del 10%, aplicado a la media aritmética, generalmente obtenida en una encuesta preliminar:

$$E = (\%)\ (\bar{\ })$$

El porcentaje podría considerarse en algunos casos como un complemento del nivel de confianza, así que una confianza del 95% daría supuestamente un error del 5%, pero no siempre será el complemento, ya que este porcentaje podrá aumentarse o disminuirse dependiendo del grado de precisión con que se desee hacer la estimación.

Pero hay otro tipo de *error* que no tiene nada que ver con el muestreo y que no se puede medir. Estos errores denominados *ajenos a muestreo*, *no muestra es* o *sistemáticos*, se consideran como el resultado de instrumentos de medición incorrectos, cuestionarios mal definidos, errores cometidos por el entrevistador al efectuar las preguntas o al interpretar las respuestas, preguntas vagas o ambiguas; en otros casos, como consecuencia de la influencia negativa del entrevistador, del mal diseño del formulario. También podría ser el resultado en los procesos de crítica, codificación, tabulación, en los cuales se pueden cometer errores ajenos al muestreo.

a arian a: Del grado de variabilidad, depende en gran parte el tamaño de la muestra. Por ejemplo, el médico que realiza un análisis de sangre, el panadero que se come un pan, el cocinero que paladea una cucharada de sopa; cada uno de ellos ha hecho uso de una muestra. Obsérvese que un centímetro de sangre, un pan o una cucharada de sopa, puede determinar el comportamiento de la totalidad (sangre del paciente, total de panes producidos en una jornada, total de sopa preparada para el almuerzo) debido al grado de homogeneidad que presenta la característica observada; cualquier muestra que se tome de estas poblaciones se espera que dé un resultado igual, representativo del total.

Si todas las características presentaran un alto grado de homogeneidad, sería muy sencilla la investigación por muestreo, pero esa situación no se da en la mayoría de las poblaciones que se investigan; por lo tanto el tamaño de la muestra deberá aumentarse de acuerdo al grado de variabilidad, de tal forma que queden representados en ella los valores extremos que toma la variable.

En el cálculo del tamaño de la muestra, lo ideal sería que la varianza se obtuviera en la población investigada, pero en la mayoría de los casos esto no es posible, de ahí que exista la necesidad de estimarla mediante:

- *Encuestas pre iminares*, *pi oto* o *pretest*, sistema que consiste en la realización de una pequeña encuesta, cuyo tamaño por lo general se toma en forma arbitraria, fijándose un porcentaje que debe ser aplicado al tamaño poblacional. Esta varianza, resultante de la encuesta, se considera un buen sustituto al de la población.
- El empleo de la *arian a* obtenida en censos o investigaciones similares, realizadas con anterioridad.
- Otras veces son el resultado de conjeturas que se hacen sobre la población.

Las *encuestas pre iminare*s permiten, no sólo estimar la varianza de las características de aquellas variables que tienen que ver con el objetivo de la investigación, sino que facilitan su conocimiento, familiarizan al entrevistador al aplicar el formulario, determinan las fallas del mismo, además, se puede conocer la reacción del entrevistado, el tiempo necesario para cada entrevista y, como se dijo anteriormente, permite el cálculo del error necesario para fijar el tamaño óptimo de la muestra.

De igual modo la *encuesta pre iminar* permite hasta cierto punto, probar gran parte de los mecanismos técnicos de muestreo, establecer costos, localizar unidades, comprobar si la medición es práctica, etc.

La confianza: Se ha mencionado que el grado de precisión depende de alguna manera, del tamaño de la muestra y éste, a su vez, de factores tales como el error, la varianza y la confianza. Esta última es fijada por el investigador quien, por lo general, trabaja con una Z = 2, valor correspondiente a un nivel de confianza del 95,5%; pero es posible que sean otros los valores de Z, tales como: Z = 2,57 (confianza del 99%) y Z=1,64 (90% de confiabilidad). Lo más práctico es trabajar con Z = 2 ó Z = 1,96, pero de todas maneras la decisión sigue dependiendo del investigador, de la finalidad y del grado de precisión que se desee.

ama o de a pob ación: Es el último componente de la fórmula, para calcular el tamaño óptimo de la muestra, cuando se trata de poblaciones finitas.

EJERCICIOS PARA RESOLVER

Nota: En el capítulo 7 (pág. 303) se trató este tema, con un gran número de ejercicios.

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. Decir si es cierto o falso a los siguientes puntos:
   a) Población estadística, es un conjunto de elementos que presentan una característica común. b) Una población finita es aquélla que está formada por un limitado número de elementos que presentan una característica (cualitativa o cuantitativa), común. c) La muestra aleatoria puede no ser representativa del universo o población de la que se extrajo. d) El método de selección al azar, no está basado en el cálculo de las probabilidades. e) La muestra, es obtenida con el fin de investigar, a partir del conocimiento de sus características particulares, las propiedades de toda la población. f) Una de las ventajas del muestreo, es el de mayor control de la integralidad y calidad de la enumeración. g) No es necesario utilizar la misma unidad de selección para la aplicación del muestreo, que para la tabulación y análisis.

2. Decir si es cierto o falso a los siguientes puntos:
   a) A medida que aumenta el tamaño de la muestra, disminuye el error. b) El intervalo comprendido entre los límites de confianza se llama intervalo de confianza. c) El método de selección de la muestra, debe ser completamente independiente de las características que se desean examinar. d) Una de las ventajas de la selección al azar, es la poca necesidad del conocimiento previo de las características del universo. e) Por la falta de actualización de la lista o registro de las unidades a ser investigadas, la muestra puede resultar sesgada. f) La expresión $\sqrt{\frac{N-n}{N-1}}$ se conoce como «*Factor de corrección para poblaciones finitas*». g) El error no se da en las mismas unidades de medición de la variable principal del estudio.

3. Decir si es cierto o falso a los siguientes puntos:
   a) La selección de informantes por parte de los entrevistados, expone el riesgo de introducir propensiones o vicios. b) El tamaño de la muestra necesaria para alcanzar una cierta confiabilidad, no aumenta indefinidamente cuando crece el número de elementos en la población. c) El muestreo por proporciones se utiliza en especial, cuando se desea estimar las unidades en la población que poseen alguna característica cualitativa o atributo. d) El término Afijación óptima, a menudo, se usa para métodos estratificados en los que se consideran tanto la variación como el costo de la misma. e) El muestreo sistemático, consiste en seleccionar uno a uno los elementos de la muestra en un orden indeterminado. f) Cuando la fracción de muestreo no es la misma en cada estrato, se habla de afijación proporcional. g) La idea básica de la estratificación, es la posibilidad de dividir una población heterogénea o de gran variabilidad, en sub-poblaciones, cada una de las cuales es homogénea en cuanto a su composición interna.

4. Determinar en los ejercicios que se dan a continuación, la clase de muestra (probabilística o no probabilística) y el método (aleatorio simple, estratificado, por conglomerados, sistemático, por conveniencia o cuotas).
   a) Población: listado de nombres que aparecen en el directorio telefónico. Técnica de muestreo: seleccionar un nombre por cada cincuenta. b) Población: estudiantes matriculados en nuestra universidad, en el presente semestre. Técnica de muestreo: ir a la cafetería de la universidad y seleccionar un estudiante cada tercera mesa y entrevistarlo, o por cada tercer estudiante, en forma ordenada, seleccionar uno. c) Población: alumnos matriculados en nuestra universidad, en el presente período académico. Técnica de muestreo: elegir los primeros 80 alumnos que llegan a la Universidad el día miércoles de una semana cualquiera. d) Población: nombres de 1.000 familias con sus respectivas direcciones, anotadas en papelitos de igual tamaño e introducirlas en una caja. Técnica de muestreo: una persona extrae 100 papelitos, uno tras otro, removiéndolos en cada extracción. e) Población: alumnos matriculados en la universidad, en el período académico actual. Técnica de muestreo: durante una semana, entrevista a 50 alumnos por cada facultad.

5. En cada una de las situaciones que se presentan a continuación determinar la clase (probabilística o no probabilística) y tipo (al azar simple, sistemático, estratificado, por cuotas, conglomerados, etapas, áreas, voluntario o por conveniencia) de muestreo.
   a) Población: hojas de registro de pasajeros en el Hotel Hilton de Bogotá. Técnica de muestreo: tomar una hoja de registro a intervalos iguales. b) Población: estudiantes matriculados en una universidad en el presente semestre académico. Técnica de muestreo: con base en los listados dados por la Oficina de Registro Académico, tomar muestras de alumnos por cada facultad.
   c) Población: los compradores en un supermercado durante una semana. Técnica de muestreo: se informa por los diferentes medios de comunicación, para que las personas que deseen colaborar con una investigación, se dirijan a una determinada oficina a suministrar la información requerida.

6. Diga cuál de estos ejemplos corresponde al Muestreo aleatorio.
   a) Si de las jaulas de un criadero de conejos se toma una muestra de 10 animales, donde siempre el investigador saca aquellos animales que alcance con la mano. b) Con una población pequeña pero heterogénea, el investigador inspecciona la totalidad de ella y selecciona una muestra de unidades típicas, es decir, unidades que a su parecer están cercanas al promedio de la población. c) Una muestra de frutas (mora), de un recipiente bastante grande, tomadas a 15 o 20 centímetros de la parte superior del recipiente. d) Una muestra de la cual el informante va a ser voluntario, cuando la información o la medición es desagradable o penosa para la persona que está siendo

investigada. e) Nombres y apellidos de mil personas colocados en cápsulas idénticas en un mismo recipiente. Una persona con los ojos vendados extrae 100 cápsulas y después de cada extracción se revuelven las cápsulas en el recipiente.

7. El movimiento en el mercado de cierto producto alimenticio será estimado, registrando las compras almacenadas del producto durante algunas semanas seleccionadas del año. Analice las ventajas y desventajas de una selección sistemática de las semanas para este estudio.

8. La universidad desea hacer una investigación sobre rendimiento académico y condiciones socio-económicas de los estudiantes. Como sería el diseño del método de muestreo, si le pidieran hacerlo mediante el:
   a) Muestreo aleatorio simple. b) Muestreo por etapas. c) Muestreo voluntario. d) Muestreo estratificado.
   Explique cómo sería el procedimiento a seguir en cada método, si se sabe que hay 3 jornadas, 10 semestres o niveles y 8 facultades en la universidad. Usted puede fijar condiciones para que se pueda aplicar cada método, como por ejemplo: el tamaño de la población, de la muestra, características a estudiar, etc.

9. Suponga, que una gran tienda de ventas al detal tiene sus cuentas por cobrar listadas por departamento. La empresa desea usar el muestreo para estimar en un día específico el total de las cuentas por cobrar. Analice los méritos relativos del muestreo aleatorio estratificado, del muestreo por conglomerados, del muestreo sistemático y del muestreo por conglomerados en dos etapas. ¿Qué información adicional la gustaría tener sobre estas cuentas, antes de seleccionar el diseño de muestreo?

10. El Directorio de una ciudad data de 4 años atrás y tiene las direcciones a lo largo de cada calle, con los nombres de las personas que viven en cada dirección. ¿Cuáles son las deficiencias de este marco, si se realiza una encuesta por muestreo? ¿Durante los trabajos de campo pueden ser corregidas tales deficiencias? Al usar el directorio, ¿extraería usted una lista de direcciones o una lista de personas?

11. Analice los méritos relativos al utilizar entrevistas personales, entrevistas por teléfono y cuestionarios por correo como métodos de recolección de datos para cada una de las situaciones siguientes:
   a) Un ejecutivo de televisión quiere estimar la proporción de televidentes que, en el país, están viendo el canal 11 a cierta hora. b) El periódico El Espectador desea encuestar las actitudes del público hacia el tipo de noticias que publica el diario. c) El alcalde de la ciudad, está interesado en conocer la reacción de los ciudadanos a las tarifas de los servicios públicos. d) La Secretaría de Salud de un municipio, quiere estimar la proporción de perros que han sido vacunados contra la rabia, el año pasado.

12. Explique como obtendría una estimación, sobre la proporción de señoras que usan un nuevo detergente en el mercado, considerando solamente la ciudad donde usted vive.

13. Una empresa de investigación de mercados, ideó un plan de muestreo para estimar las ventas semanales de un cereal de la marca A en determinada región del país. La empresa decidió tomar muestras de ciudades dentro del área y luego en supermercados dentro de esas ciudades. La medición de interés es el número de cajas vendidas del cereal en una semana específica. Se tomó muestra de cinco (5) ciudades de las 20 en el área. ¿Es este un problema de muestreo aleatorio simple?;

¿muestreo por conglomerados en dos etapas?; ¿ó muestreo aleatorio estratificado de asignación proporcional? Analice los méritos relativos de estas tres opciones y diseñe un plan de muestreo para cada caso.

14. Un perito de seguridad está interesado en estimar la proporción de llantas de automóvil en mal estado. ¿Deberá usar automóviles individuales o en grupos, tales como estacionamientos, como unidades de muestreo? ¿Qué marca podrá usar?

15. Una empresa de investigación de mercados estima la proporción de clientes potenciales que prefieren cierta marca de lápiz labial, mediante la selección aleatoria de 100 mujeres que llegan a su punto de venta, en un centro comercial. De las 100 mujeres seleccionadas, 65 de ellas manifestaron tener preferencia por la marca A,
    (a) ¿Cómo estimaría usted la proporción verdadera de mujeres que prefieren la Marca A, con un límite de confianza cualquiera para el error de estimación?; (b) ¿Cuál es la población objetivo de este estudio?; (c) ¿Seleccionó la empresa de investigación de mercados, una muestra irrestrictamente aleatoria?; ¿De qué manera cree usted que la realizó?; (d) ¿Qué opinión tiene usted de la aplicación, en este caso, de una selección sistemática?

16. El gerente de ventas de una empresa quiere medir la relación entre las ventas y los costos de publicidad mensuales. ¿Qué parámetros le sugeriría estimar? ¿Por qué? ¿Qué datos debe recolectar el gerente?

17. Responda en forma breve cada uno de los siguientes interrogantes:
    a) Cuáles son las diferencias que puede haber entre: a_1) Muestreo sistemático o muestreo aleatorio simple; a_2) Muestreo de criterio, por cuotas y de conveniencia; a_3) Muestreo estratificado y por conglomerados. b) ¿En qué casos a un investigador le interesaría obtener conclusiones mediante la realización de una muestra y no mediante la investigación total? c) Explique brevemente el objetivo del muestreo. d) Explique los métodos de sustitución que más se utilizan en el muestreo.

18. En una encuesta sobre salud de las personas en una ciudad, se parte de un mapa, se identifican los barrios, se enumeran y se efectúa la selección sistemática, luego en cada barrio se lista y enumeran las manzanas, se hace una nueva selección al azar; se conforma un listado de viviendas para cada manzana, para luego hacer otra selección aleatoria. Para cada vivienda seleccionada se llena un cuestionario por cada persona o miembro. Se pide: a) Identificar la población o poblaciones que presenta este estudio. b) Identificar el marco (o los marcos) que se debe elaborar. c) ¿Cuáles son las unidades?; d) ¿Qué método de muestreo debe aplicarse en este caso? e) Proporcione cuatro (4) características a estudiar; f) Defina cuatro parámetros a estimar.

19. Un grupo de consumidores quiere determinar el sentir público hacia cierto producto. En la ciudad donde se va a realizar el estudio por muestreo, se tienen cuarenta manzanas, con 10 casas cada una.
    a) Formule un método de dos etapas para realizar un muestreo aleatorio, seleccionando primero (10) diez manzanas y después, en cada una de las manzanas, dos (2) casas. b) Si el objetivo es utilizar un plan de una etapa, ¿cómo procedería usted?

20. Una corporación lista a sus empleados por grupos de ingresos (alfabéticamente dentro de cada grupo) desde el menor hasta el más alto ingreso. Si el objetivo es estimar el ingreso promedio por

empleado, ¿qué método deberá utilizar en una investigación por muestreo: sistemático, estratificado o irrestricto aleatorio? Analice las ventajas y desventajas de cada uno de estos métodos.

21. Un auditor se enfrenta a una larga lista de cuentas por cobrar de una empresa. El auditor debe verificar las cantidades con base en un 10% de esas cuentas, y estimar la diferencia promedio entre los valores revisados y los asentados en el libro. a) Suponga que las cuentas están ordenadas cronológicamente, teniendo las más antiguas una tendencia a tomar valores pequeños. Para seleccionar la muestra, ¿usaría usted un muestreo de diseño sistemático o un muestreo irrestricto aleatorio? b) Suponga que las cuentas están colocadas en forma desordenada para la selección de una muestra, ¿usaría usted un diseño de muestreo sistemático o un muestreo irrestricto aleatorio? c) Suponga que las cuentas están agrupadas por departamentos y que dentro de estos, están listados en orden cronológico. Nuevamente las cuentas antiguas tienden a tomar valores más pequeños. Para seleccionar la muestra. ¿usaría usted un muestreo de diseño sistemático o un muestreo irrestricto aleatorio?

TAMAÑO DE LA MUESTRA - M.A.S.

Nota: Los siguientes ejercicios, complementan a los presentados en el Capítulo 7, en el que se explica con más detalle el proceso de cálculo, y se dan las soluciones en el SIL, en el mismo capítulo.

22. Para efectos de planeación económica en la Sabana de Bogotá, es necesario estimar entre 2.200 hatos: a) El número de vacas lecheras por hato, con un error del 8% y una confianza del 95,5%; b) El rendimiento medio de leche por hato, con un error del 8% y una confianza del 95,5%; c) La proporción de hatos con un rendimiento medio de leche superior a los 250 litros, un error del 8% y confianza del 95,5%. Una encuesta preliminar de tamaño 40, arrojó las siguientes estimaciones: a) Promedio de vacas por hato, 46 y desviación típica, 40. b) Rendimiento medio de leche por hato, 345 litros y varianza, 9.700 litros. c) 60% de los hatos con rendimiento medio de leche superior a 250 litros. d) ¿Qué tamaño de muestra recomendaría usted?

23. Si en el ejercicio anterior se desea estimar el número total de vacas lecheras en los 2.200 hatos, con un error de 8.000 vacas y una confianza del 95,5%. ¿Qué tamaño de muestra se requeriría?

24. Se requiere determinar el tamaño de la muestra para el estudio de gastos semestrales, en productos de tocador, en una zona de Bogotá. Para estimar la desviación típica de la población, el investigador realiza una encuesta preliminar en 80 familias de dicha zona, encontrando una desviación típica de $250. a) ¿Cuál es el valor de n si se desea que el error estándar de la media inferida de la población no exceda de ± 240, a un nivel de confianza del 95%? b) ¿Cuál es el tamaño de la muestra si se sabe que en el barrio investigado viven 1.200 familias?

25. Una empresa de transporte urbano permite que ciertos gastos de sus afiliados (gasolina, lubricantes y lavado) se haga mediante la utilización de la tarjeta de crédito expedida para pago en las bombas de gasolina locales. La empresa ha expedido 10.050 tarjetas de crédito. Se desea hacer una investigación sobre la utilización de la tarjeta y otras características.
Para ello se realizó una encuesta preliminar de 90 tarjetas y se encontró que 63 de ellas fueron utilizadas para pagar servicios durante el mes de referencia. Por otra parte, el total de los gastos cancelados con tarjetas fueron de $2.390.000 y la desviación típica de $4.000. Se desea determinar el tamaño de la muestra, con un error del 2% y una confianza del 95%, para estimar: a) la proporción

de afiliados que utilizan la tarjeta; b) el gasto promedio mensual cancelado con tarjeta; c) El total del gasto mensual, cancelado con tarjeta; d) De los tamaños anteriormente obtenidos ¿Cuál aconsejaría usted?

26. Decir si son ciertas o falsas las siguientes aseveraciones: a) Se dice que en todo proceso de selección debe haber sustitución de las unidades que no se han podido entrevistar. b) El parámetro describe una determinada característica de las unidades que conforman la población. c) El muestreo sistemático consiste en seleccionar uno a uno los elementos de la muestra en un orden indeterminado. d) A medida que aumenta el error, manteniendo constante la varianza, se aumenta el tamaño de la muestra. e) Una muestra aleatoria es aquélla en la cual ciertas unidades tienen mayor posibilidad que otras de ser seleccionadas.

DISEÑO DE MUESTREO

El *dise o* y *e tama o* de la muestra van a determinar la cantidad de información necesaria respecto a los objetivos de la encuesta. Recordemos que el *objeti o de muestreo* es contar con el mayor número de unidades y la mayor cantidad de información, pero con el menor costo posible.

A veces se cree, que con mucha información se logran buenas estimaciones; sin embargo, generalmente ocasionan despilfarro de dinero, pérdida inútil de tiempo y complicaciones en la organización y análisis de los datos. Otras veces se peca por poca información, que hace temer por el éxito de la investigación. Los dos casos se debe a que la muestra no fue correctamente diseñada.

Para el *dise o de a muestra* y *cá cu o de tama o*, se requiere que se hayan adelantado algunas etapas, tales como:

a. *P anteamiento de prob ema:* nos permite determinar la necesidad, o no, de adelantar una investigación que permita recoger o utilizar información que proporcione un mayor conocimiento del problema, aclarar algunas inquietudes y tener más herramientas de juicio para la toma de decisiones. Supongamos que se tiene planteado como ejercicio de aplicación: el problema de la mala calidad de la leche, la baja nutrición de los niños, el precio elevado de este producto y su distribución.
b. Una vez determinada la necesidad de realizar la investigación, debe *estab ecerse e objeti o*, o *os objeti os* de una manera clara y precisa, los cuales no deben perderse de vista, y sí ser tenidos en cuenta a medida que se vaya avanzando, tanto en el diseño como en la instrumentación de la encuesta. En nuestro caso, el ejercicio que desarrollaremos nos permitirá aplicar algunos de los métodos de muestreo, la realización de estimaciones e interpretación de las mismas. Consideraremos que el objetivo principal del estudio será investigar el consumo de carne en uno de los barrios de la ciudad; como objetivos secundarios nos interesa conocer: las variaciones en el consumo; composición y niveles de ingreso de la familia; además, se busca información sobre la calidad, precios, consumo de pescado y pollo, distribución, entre otros.
c. Debemos *determinar a pob ación objeti o*, que debe estar claramente definida y plenamente identificable al iniciar el proceso de selección y la entrevista. Implica identificar la unidad o el elemento, la cual debe ser: *c ara*, *mensurab e*, *adecuada* y *comparab e*. En nuestro problema, la unidad corresponde a familias que viven en el barrio objeto de investigación. Observemos que la familia es, en nuestro caso, *a unidad de se ección*

d. *Formu amos as hipótesis*, ya sea para afirmar o rechazar. Anotaremos como ejemplo algunas de ellas, al considerar la mala calidad de la leche.
   - Hay tendencia a consumir leche de marca.
   - Se observan variaciones en las cantidades compradas, frente a variaciones de precios y marcas.
   - La posesión de neveras influye en los hábitos de compra.
   - El consumo de leche aumenta con la población escolar.
   - El consumo de leche por persona es bajo, en comparación con el mínimo nutricional recomendable.

e. Se elabora el *marco de referencia*. Recordemos que es un *istado de todas as unidades* que constituyen la población objetivo, y puede corresponder a un mapa o croquis que indique la localización de las unidades. En nuestro ejercicio con el cual desarrollaremos la teoría, el marco estará constituido por 355 familias, que además de estar completamente localizadas se encuentran numeradas desde 001 hasta 305. (Ver Tabla 13.1 en el SIL)

f. Se establece el *método de reco ección*, teniendo en cuenta las ventajas y desventajas que presenta cada unos de ellos, que nos permita hacer una buena selección. Algunos de estos métodos son: la *entre ista*, el *correo*, la *entrega persona* del *cuestionario*, el *te éfono* y el *pane* . Para nuestro caso, se ha considerado la entrevista como el método de recolección más indicado.

g. Se seleccionará el *método de muestreo a eatorio*, dependiendo de la variabilidad que presente la característica principal que tiene que ver con el objetivo de la investigación. Aplicaremos el *Muestreo A eatorio Simp e*

h. A esta altura se han desarrollado otras etapas tales como: la elaboración del instrumento de recolección; la selección y preparación del personal; se ha hecho examen de la documentación, en especial para determinar si este tipo de estudio ya fue realizado y si existe alguna metodología que se haya aplicado en una investigación similar con buenos resultados; además, se cuenta con los recursos necesarios que permitan el buen desarrollo y terminación del trabajo.

i. En el diseño y determinación del tamaño de la muestra se hicieron algunas consideraciones que se deben tener en cuenta en una investigación, siendo una de ellas el grado de homogeneidad o heterogeneidad que presenten las características que se van a investigar, pues nos permite tomar decisiones respecto al método de muestreo que debe corresponder a la investigación que se piensa realizar.

FORMULARIO O CUESTIONARIO

Se diseñó un instrumento con el fin de recoger la información necesaria, de acuerdo a los objetivos establecidos en la investigación. El cual debió ser sometido a prueba, realizando una encuesta preliminar o piloto, cuyo tamaño fue elegido por el investigador teniendo en cuenta factores tales como: población, objetivo, costos, tiempo y recursos humanos.

En el caso de una población de 355 familias, ubicadas en un determinado sector de la ciudad, el tamaño de la muestra preliminar será: n = (%)(N) o sea n = 0,04(355) = 14 familias. Recordemos que el porcentaje es arbitrario y lo determina el investigador, pero nunca puede ser mayor del 10%.

Presentamos a continuación una parte del formulario y seguidamente el marco o listado de las 355 familias (incluimos información de algunas características que luego serán recolectadas, tanto para la muestra piloto como la definitiva con la cual se harán los estimativos). **Ver información en el SIL capítulo 13 (Tabla 13.1).**

Realiza el departamento de
investigación y mercadeo **No. 001**

**ENCUESTA SOBRE INGRESO Y CONSUMO DE CARNE
DE LAS FAMILIAS EN EL BARRIO FUTURO / 2006**

Dirección _____ Ciudad _____ Fecha _____
Teléfono _____

I. GENERALIDADES

Nombre del informante: _____

Dirección: _____ Teléfono: _____

Ocupación: _____ Edad: _____

II. …….....

1. Ingreso mensual: _____
2. Propiedad de la vivienda ☐ 2.1 Propietario
☐ 2.2 Inquilino

3. Consumo diario de carne:_____ gramos etc...

Si se fuera a hacer una investigación total, la información sobre algunas características: ingreso (miles de $); propiedad vivienda; total personas por sexo y ocupación; y consumo de carne diaria (en gramos) se tendría que utilizar toda la información de las 355 familias tal como aparece en el SIL, (Capítulo 13).

APLICACIÓN DE MÉTODOS Y TÉCNICAS DE MUESTREO

MUESTREO ALEATORIO SIMPLE (MAS)

Tamaño de la muestra

En la práctica es muy frecuente que no se conozca la varianza de la característica en la población (σ^2); en tales casos se debe recurrir a censos, a investigaciones similares realizadas con anterioridad o a investigaciones preliminares, denominadas *encuestas pi oto*. Este último procedimiento es el que más se emplea para determinar el tamaño de la muestra, partiendo del supuesto de que no existe información sobre la población. En primer lugar, se elabora un listado de las familias a investigar que conforman el marco muestral, tal como aparece en el cuadro 13.1 (*er capítu o 1 en SI*). Después de haber identificado la población a la cual se va a investigar, definida claramente la unidad de investigación y las características que tienen que ver con el estudio, se procede a establecer el tamaño óptimo para la muestra.

Para diseñar la muestra es indispensable contar con un marco de referencia, es decir, la lista, mapa u otra especificación de las unidades, que resulta de la información previamente disponible, respecto a la población sobre la cual se basan los esquemas particulares de muestreo. En nuestro caso, el marco estará constituido por una lista de 355 familias. La familia o unidad de investigación se ha definido como el conjunto de personas, (del barrio El Futuro), que viven bajo un mismo techo, ligadas por un lazo de consanguinidad y dependencia económica. Una vez identificada y enumerada la población, se procede a calcular el tamaño de la muestra, mediante algunas de las fórmulas que se presentan a continuación:

$$(1)\ n = \frac{n_o}{1 + \frac{n_o}{N}} \quad \Rightarrow \quad \text{donde} \quad n_o = \frac{Z^2S^2}{E^2} \qquad (2)\ n = \frac{NZ^2S^2}{NE^2 + Z^2S^2}$$

$$(3)\ n = \frac{S^2}{\left(\frac{E}{Z}\right)^2 + \frac{S^2}{N}}$$

Como no se conocen las varianzas poblacionales, de una o algunas de las características que tienen que ver con el objetivo principal de la investigación, se procede a estimarlas mediante una *encuesta pi-oto* o *pre iminar siendo*:

$$n_{pilot} = 4\%\ (n) \qquad n_{pilot} = 0{,}04\ (355) = 14 \text{ familias}$$

El porcentaje del 4% es arbitrario, corresponde a una decisión del investigador, en algunos casos depende: del tamaño poblacional, tiempo y costos, nunca podrá ser mayor al 10%, el cual se considera grande. Una vez establecido el valor de n piloto, se utiliza cualquiera de los métodos de selección aleatoria ya presentados en este capítulo. Es necesario recordar que hasta el momento se está utilizando el M.A.S. (Muestreo Aleatorio Simple), donde las unidades o los elementos se seleccionan sin reposición, todas con igual probabilidad de ser elegidas.

Luego se realiza la selección, ya sea mediante el uso de Calculadoras, Tablas de números aleatorios o del Programa EXCEL, las 14 familias que formarán la *muestra pi oto*, (la población está numerada desde 001 hasta 355), sin tener en cuenta números superiores o repetidos. Supongamos que se obtuvieron los siguientes números aleatorios con los cuales elaboramos la tabla:

	335	004	289	128	097	271	253
	348	202	349	197	018	020	207

Información obtenida mediante la *Encuesta Pre iminar* o *Pi oto*

No. Orden	Números Aleatorios	Ingresos (Miles $)	Vivienda Propia	No. total de personas	M	F	Trabajando	Consumo diario de carne (grs.)
1	335	2.420	si	3	1	2	1	760
2	004	1.060	si	3	1	2	1	765
3	289	2.860	no	3	1	2	2	592
4	128	2.320	no	3	1	2	2	520
5	097	1.120	si	3	1	2	1	610
6	271	2.500	si	5	1	4	2	636
7	253	1.930	si	2	0	2	2	650
8	348	1.020	si	3	2	1	1	605

ontin a

ontinuación

9	202	1.960	si	2	1	1	1	642
10	349	1.110	no	2	0	2	1	520
11	197	970	no	2	1	1	1	508
12	018	1.790	no	4	2	2	1	767
13	020	2.250	no	3	1	2	2	842
14	207	1.750	si	2	0	2	1	684

Para el cálculo del tamaño muestral, por lo general se selecciona la característica principal, en nuestro caso podría ser el *consumo* o el *ingreso*. Sin embargo se seleccionan los dos para comparar y explicar sus resultados, a esto le añadimos la característica cualitativa (propiedad vivienda), para que veamos como se trabaja con atributos.

Con la anterior información obtenida de la encuesta preliminar, se determina la varianza y se calcula el tamaño óptimo de la muestra, dependiendo de las características que hayamos seleccionado. Veamos los resultados, cuando trabajamos con un error de muestreo del 5% para las variables (*ingreso consumo*) y del 8% para el atributo (*propiedad de a i ienda*). La confianza en los tres casos será del 95,5% (Z=2).

a. *Ingreso promedio (mi es de pesos)* $\bar{x} = \dfrac{\Sigma x_i}{n} = \dfrac{25.060}{14} = 1.790$ ($ 1'790.000)

E = 0,05 ($\bar{x}$) = 0,05 (1.790) = 89,5 (ochenta y nueve mil quinientos pesos)

$$s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1} \qquad s^2 = \frac{50.155.400 - 14(1.790)^2}{14-1} = 407.538,46$$

$$n_o = \frac{Z^2S^2}{E^2} \Rightarrow \qquad n_o = \frac{Z^2S^2}{E^2} = \frac{2^2(407.538,46)}{89,5^2} \cong 203,5$$

$$n = \frac{n_o}{1 + \dfrac{n_o}{N}} \qquad n = \frac{203,5}{1 + \dfrac{203,5}{355}} \cong 130 \text{ familias}$$

$$n = \frac{NZ^2S^2}{NE^2 + Z^2S^2} \qquad n = \frac{355(2^2)(407.538,46)}{355(89,5^2) + 2^2(407.538,46)} = 130 \text{ familias}$$

También se puede calcular mediante:

$$n = \frac{S^2}{\left(\dfrac{E}{Z}\right)^2 + \dfrac{S^2}{N}} \qquad n = \frac{407.538,46}{\left(\dfrac{89,5}{2}\right)^2 + \left(\dfrac{407.538,46}{355}\right)} = 130 \text{ familias}$$

Algunos utilizan la fórmula con corrección, ya que se considera que la varianza obtenida mediante una encuesta preliminar, cuando ésta es muy pequeña, se hace poco representativa, es decir, que el resultado de *n* es menor que el obtenido con varianza poblacional.

Procedemos al cálculo de *n* con corrección:

$$n_0 = \left[\frac{Z^2S^2}{E^2}\right]\left[1 + \frac{2}{n_1}\right] \qquad n = \left[\frac{2^2(407.538,46)}{89,5^2}\right]\left[1 + \frac{2}{14}\right] = 232,58$$

$$n = \frac{232,58}{1 + \frac{232,58}{355}} \cong 141 \text{ familias}$$

El tamaño de la muestra obtenida al ser corregida, va a ser un poco mayor en comparación que cuando no es corregida. Si fuera necesario calcular el total de ingresos de las 355 familias, el tamaño de la muestra arrojaría los mismos resultados anteriores, a pesar de que la fórmula es ligeramente modificada.

$$n_o = \frac{Z^2N^2S^2}{E^2} \qquad E = N(0,05)(\bar{x}) = 355(0,05)(1.790) = 31.772,25$$

$$n_o = \frac{2^2(355^2)(407.538,46)}{31.772,25^2} = 203,5 \quad \Rightarrow \quad n = \frac{203,5}{1 + \frac{203,5}{355}} = 130 \text{ familias}$$

b. *onsumo promedio de carne diaria por fami ia*

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{9.101}{14} = 650,07$$

$$S^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n - 1} = \frac{6.054.547 - 14(650,07)^2}{14 - 1} \cong 10.636,38$$

$$S = 103,12$$

$$E = 0,05(\bar{x}) = 0,05(650,07) = 32,50 \qquad E^2 = 1.056,48$$

$$n_o = \frac{Z^2S^2}{E^2} = \frac{2^2(10.636,38)}{1.056,48} \cong 40,27$$

$$n = \frac{n_o}{1 + \frac{n_o}{N}} = \frac{40,27}{1 + \frac{40,27}{355}} \cong 37 \text{ familias}$$

También se puede calcular mediante la fórmula:

$$n = \frac{NZ^2S^2}{NE^2 + Z^2S^2} \qquad n = \frac{355(2^2)(10.636,38)}{355(32,50^2) + 2^2(10.636,38)} = 37 \text{ familias}$$

Otra forma de calcular el tamaño óptimo, es así:

$$n = \frac{S^2}{\left(\frac{E}{Z}\right)^2 + \frac{S^2}{N}} \qquad \text{siendo: } n = \frac{10.636,38}{\left(\frac{32,5}{2}\right)^2 + \frac{10.636,38}{355}} = 37 \text{ familias}$$

Cuando hay cierto grado de homogeneidad en la característica investigada, el tamaño tiende a ser pequeño, tal es el caso observado con el consumo de carne diaria donde se obtuvo 37 familias. El grado de homogeneidad se establece cuando el *coeficiente de variación* calculado es menor del 30%; en estos casos es recomendable la aplicación del *Muestreo A eatorio Simp e*, en caso contrario es aconsejable el *muestro aleatorio estratificado*, siendo en nuestro caso, la variable ingreso.

$$CV = \frac{S}{\bar{x}}\ 100 \qquad CV = \frac{103,12}{650,07}\ 100 = 15,86\% \qquad CV = \frac{615.17}{1.790}\ 100 = 34,36\%$$

(Consumo de carne) (Ingresos)

Con los anteriores resultados hay dos posibilidades, o se trabaja con una muestra grande (130 familias) o se aplica el muestreo aleatorio estratificado como se hará más adelante, cuando calculemos el tamaño óptimo por ese método, con lo cual se obtiene una mejor estimación con las mismas características seleccionadas.

c. *Proporción de fami ias con i ienda propia*

$$P = \frac{\Sigma a_i}{n} = \frac{8}{14} \cong 0,57 \qquad s_p^2 = p = 0,57(0,43) = 0,2451 \qquad E = 8\% = 0,08$$

$$n_o = \frac{Z^2PQ}{E^2} = \frac{Z^2S_P^{\ 2}}{E^2} \qquad n_o = \frac{2^2(0,2451)}{0,08^2} \cong 153,18$$

$$n = \frac{n_o}{1 + \frac{n_o}{N}} = \frac{153,18}{1 + \frac{153,18}{355}} \cong 107 \text{ familias}$$

También se puede aplicar mediante la siguiente formúla:

$$n = \frac{NZ^2PQ}{NE^2 + Z^2PQ} = \frac{355(2^2)(0,57)(0,43)}{355(0,08^2) + 2^2(0,57)(0,43)} \cong 107 \text{ familias}$$

Si en alguna ocasión seleccionamos más de una característica, para calcular el tamaño óptimo, nos decidimos por aquél cuyo resultado sea el más alto, en este caso, sería 130 familias, si incluimos la característica vivienda.

Cálculo de algunos estimativos

Se ha dicho que la estimación de un *parámetro* se hace a través de los resultados obtenidos con la muestra. Si sólo se utiliza una cifra única para las características, tales como: promedio, proporción o razón, etc., se le denomina *estimación puntua*

Si se desea determinar unos *Límites de confianza* dentro de los cuales debe estar probablemente el parámetro, es denominado como estimación por intervalo o límites de confianza. Se podrá hablar de una estimación total, cuando se hace referencia a la totalidad de la característica en la población.

El proceso a seguir una vez determinado el tamaño de la muestra, consiste en la realización de la verdadera encuesta para la obtención de la información; luego se continúa con el proceso de tabulación o

recuento, lo que permitirá la elaboración de cuadros y el cálculo de los diferentes estadígrafos o medidas (promedios, razones, varianza, etc.), denominados *estimadores de os parámetros*

Vale la pena observar este proceso de estimación (puntual y por intervalos) desarrollando el ejercicio de las 355 familias que se tomaron para el cálculo de *n*. Primero se calculan los promedios para datos sin agrupar, es decir, datos originales y luego (cuando así se requiera) para datos agrupados (tablas de frecuencias). Supongamos, que el tamaño óptimo de la muestra obtenida por cualquiera de los métodos anteriores y características utilizadas es de 28 familias en vez de las 130 familias establecidas, (valor arbitrario que no tiene nada que ver con los resultados anteriores), tomados únicamente con la intención de facilitar la aplicación de la «t» de Student, válido para muestras pequeñas ($n \leq 30$), casos poco usuales en la práctica, pues casi siempre se trabaja con muestras grandes ($n > 30$) donde utilizamos Z. La muestra, como en todos los procesos, va a ser aleatoria y el método aplicado será el de muestreo aleatorio simple, donde las probabilidades de selección de cada unidad se van a hacer constantes y *a se ección a a ser sin repetición*

Nuevamente se utiliza la calculadora o el programa EXCEL para obtener los 28 números aleatorios, sin repetición:

170	045	186	217	076	213	132	087	006	163	312	111	248	018
096	224	303	196	230	286	032	168	318	009	144	250	036	095

Información obtenida con la muestra de 28 familias. Datos sin agrupar.

No. ORDEN	NÚMEROS ALEATORIOS	INGRESOS (Miles)	VIVIENDA PROPIA	No. TOTAL DE PERSONAS	M	F	TRABAJANDO	CONSUMO DIARIO DE CARNE (grs.)
1	170	2.700	si	3	1	2	2	812
2	045	1.920	si	2	1	1	2	630
3	186	880	no	3	1	2	1	510
4	217	860	si	5	2	3	1	506
5	076	1.840	no	5	1	4	2	620
6	213	1.860	no	3	2	1	1	716
7	132	1.260	si	3	2	1	1	520
8	087	2.060	si	5	1	4	2	780
9	006	1.700	no	2	1	1	2	895
10	163	960	no	3	2	1	1	520
11	312	1.890	no	4	3	1	2	576
12	111	1.110	si	3	1	2	1	620
13	248	1.980	si	3	1	2	2	586
14	018	1.790	no	5	1	4	2	760
15	096	1.000	no	2	0	2	2	650
16	224	1.520	no	6	2	4	2	666
17	303	3.460	si	4	1	3	3	718
18	196	950	no	3	1	2	1	515
19	230	980	si	2	1	1	1	560
20	286	1.950	si	5	1	4	2	710
21	032	960	si	2	1	1	1	782

ontin a

ontinuación

No. ORDEN	NÚMEROS ALEATORIOS	INGRESOS (Miles)	VIVIENDA PROPIA	No. TOTAL DE PERSONAS	M	F	TRABAJANDO	CONSUMO DIARIO DE CARNE (grs.)
22	168	1.640	no	5	2	3	2	720
23	318	1.760	si	2	1	1	1	543
24	009	2.820	no	4	2	2	3	742
25	144	1.880	no	3	1	2	2	560
26	250	2.060	no	2	1	1	2	616
27	036	1.450	no	2	1	1	2	580
28	095	1.010	no	2	1	1	1	630
Σ		46.250	12 (si)	93	36	57	47	18.043

Los límites para los estimadores se fijarán con una confianza del 95%. En caso de que la muestra sea grande ($n > 30$) se utiliza = 1,96. En este caso, siendo $n \leq 30$, aplicamos la "t" de Student; en primer lugar se establecen los grados de libertad υ (letra griega nu) igual a n-1 o sea $\upsilon = 28–1=27$, siendo $\alpha = 0{,}05$ complemento del 95%, tomado como nivel de confianza (la suma será del 100%); por lo tanto el valor de "t" para este ejercicio será igual a 2,052.

Veamos ahora los *estimati os puntua es* y de *inter a os* para algunas de las características que aparecen en el cuadro anterior.

ESTIMACIÓN DE PROMEDIOS Y TOTALES

(1) *Promedio de Ingresos por fami ia*: Los estimativos de promedio y total de los ingresos (miles de $) por familia, se calculan de la siguiente manera:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{46.250}{28} \cong 1.651{,}78 \quad \Rightarrow \quad \$1.651.780{,}oo$$

$$s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1} \qquad s^2 = \frac{87.416.300 - 28(1.651{,}78)^2}{28-1} \cong 408.212{,}57$$

$$S = \sqrt{408.212{,}57} = 638{,}91$$

El *estimati o puntua o promedio* es de $1.651,79, equivalente a un *ingreso promedio estimado* para las 355 familias de $1.651.785,71

La estimación con intervalos de un 95%, será:

$$\hat{\bar{X}} = \bar{x} \pm t \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \qquad o \qquad \hat{\bar{X}} = \bar{x} \pm t \frac{s}{\sqrt{n}} \sqrt{1-f}$$

f = fracción de muestreo; $f = \frac{n}{N} = \frac{28}{355} = 0{,}078$; Aplicada como factor de *corrección para poblaciones finitas,* cuando la fracción sea mayor del 5%.

$$\hat{\bar{X}} = 1.651{,}79 \pm 2{,}052 \left[\frac{638{,}91}{\sqrt{28}}\right] \sqrt{1-0{,}078} = \begin{cases} 1.889{,}69 = \$1.889.691{,}23 \\ 1.413{,}89 = \$1.413.888{,}77 \end{cases}$$

Estos resultados indican que el ingreso promedio poblacional deberá estar aproximadamente entre $1.413,89 y $1.889,69 (miles de $); con una probabilidad o confianza del 95%; sin embargo queda un margen de error del 5%, es decir, el promedio verdadero puede ser superior o inferior a esos valores. Se sabe que el valor de la *media pob aciona* o *a or erdadero*, es de $1.782 o $1.782.000 el cual se encuentra dentro de esos límites de confiabilidad. De los estimativos anteriores es preferible deducir que el promedio poblacional deberá estar dentro de esos límites con cierto nivel de confianza, que considerar a $ 1.651,79 como si fuera el promedio verdadero de la población.

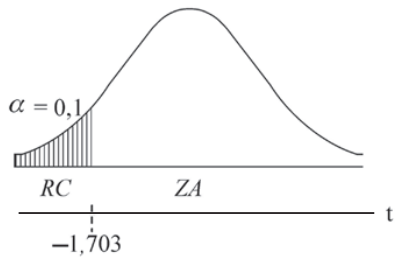
Por otra parte, en el proceso de una investigación por muestreo, es común la *formu ación de hipótesis* que deben ser comprobadas. Por ejemplo, el investigador de las 355 familias puede considerar como hipótesis que el ingreso de esas familias es de $1.750.000 ante la alternativa de que sea diferente, es decir, que puede ser mayor o menor a $1.750.000. En este caso el procedimiento que se sigue para contrastar esta hipótesis es:

1) Ho: $\mu = 1.750$ *(Hipótesis nula)*
   Ha: $\mu = 1.750$ *(Hipótesis alternativa)*
2) $\alpha = 0{,}05$ *(nivel de significación)*
3) $\hat{\mu}_s = \hat{\bar{x}}_s = 1.889{,}69$
   $\hat{\mu}_i = \hat{\bar{x}}_i = 1.413{,}89$ *(Límites de confianza)*

(4) Como la *hipótesis nu a* (*Ho*: $\mu = 1.750$) se ubica dentro de los límites de confianza, se acepta lo indicado por el investigador, es decir, que el ingreso promedio de las familias puede ser considerado como de $1.750 ó $1.750.000.

Obsérvese en el ejercicio puesto a consideración dado su tamaño y una información supuesta, se pudo conocer la media poblacional ($\mu = 1.782$) lo cual, en la práctica no sucede con frecuencia; además se está afirmando que el ingreso es de $1.750 cuando en realidad fue $1.782; por tanto se ha cometido un error de tipo II. Si se quiere probar que el grupo familiar investigado tenga un nivel de ingresos inferior a 1.750 (miles de $), no se deben utilizar los límites de confianza, para constatar la hipótesis; el procedimiento a seguir es:

1) Ho: $\mu = 1.750$
   Ha: $\mu < 1.750$
2) $\alpha = 0{,}05$
3) $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{1.652 - 1.750}{638{,}9/\sqrt{28}} = -0{,}81$



$\upsilon = n - 1 = 27$
$\alpha = 0{,}10$ } $t = -1{,}703$

RC = Región Crítica o de rechazo
ZA = Zona de Aceptación
t = –1,703 punto crítico

Al nivel del 5%, se estará aceptando que el nivel de ingresos es de $1.750.000 y se estaría cometiendo un error de tipo II, ya que estamos aceptando un ingreso de $1.750.000, cuando en realidad es de $1.782 (miles de $) (Estamos aceptando algo falso).

La *estimación de tota de ingresos* para las 355, se procede mediante la aplicación de la siguiente fórmula:

$$\hat{\bar{X}}_{s_i} = N\,\bar{x} \pm t\,\frac{Ns}{\sqrt{n}}\sqrt{1-f}$$

$$\hat{\bar{X}}_{s_i} = 355(1.651{,}79) \pm 2{,}052(355)\left[\frac{638{,}91}{\sqrt{28}}\right]\sqrt{1-0{,}078}$$

$$\hat{\bar{X}}_{s_i} = 586.385{,}45 \pm 84.454{,}94 \begin{cases} \$670.840{,}39 \\ \$501.930{,}51 \end{cases} \text{(miles \$)}$$

El *ingreso tota* para las 355 familias se estima en $ 586.386.450, pero el valor verdadero puede estar entre $670.840.390 y $501.930.510 con una seguridad del 95%.

Promedio de personas por fami ia

Para determinar los límites de confianza del 95%, se debe calcular la media y la desviación estándar.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{93}{28} = 3{,}3214 \qquad s = \sqrt{\frac{\Sigma x_i - n\bar{x}^2}{n-1}} = \sqrt{\frac{351 - 28(3{,}3214)^2}{27}} = 1{,}2488 \cong 1{,}25$$

$$s = 1{,}25$$

$$\hat{\bar{X}}_{s_i} = \bar{x} \pm t\,\frac{s}{\sqrt{n}}\sqrt{1-f}$$

$$\hat{\bar{X}}_{s_I} = 3{,}32 \pm 2{,}052\left[\frac{1{,}25}{\sqrt{28}}\right]\sqrt{1-0{,}078} = \begin{cases} 3{,}79 \\ 2{,}85 \end{cases}$$

Promedio de personas por familia.

El total de personas para las 355 familias, será obtenido así:

$$\hat{X}_{s_I} = N\bar{x} \pm t\,\frac{Ns}{\sqrt{n}}\sqrt{1-f}$$

$$\hat{X}_{s_I} = 355(3{,}32) \pm 2{,}052\left[\frac{(355)(1{,}25)}{\sqrt{28}}\right]\sqrt{1-0{,}078} =$$

$$\hat{X} = 1.178{,}6 \pm 165{,}24 = \begin{cases} 1.343{,}84 \\ 1.013{,}36 \end{cases}$$

Total de personas

Se puede afirmar que, aproximadamente, el número total de personas para las 355 familias es de 1.179; sin embargo, con una confianza del 95%, el total poblacional deberá estar entre 1.013 y 1.344. Comparado el total poblacional (1.164) con el estimador puntual, se aprecia una pequeña diferencia y, como en los casos anteriores, el valor se encuentra incluido dentro de los límites de confianza establecidos.

ESTIMACIÓN DE PROPORCIONES Y TOTALES

Proporción de fami ias con i ienda: Para el cálculo de la proporción y la varianza se procede así:

$$p = \frac{\Sigma a_i}{n} = \frac{12}{28} = 0{,}4286 \text{ ó } 42{,}86\% \qquad s_p^2 = pq = 0{,}4286(0{,}5714) = 0{,}2449$$

El 42,86% o 43% es el estimador puntual, es decir, se estima que el 42,86% de las familias son propietarias de vivienda.

Para obtener los *Límites de confianza,* se aplica la siguiente fórmula:

$$\hat{P}_{s_i} = p \pm t\sqrt{\frac{pq}{n-1}}\sqrt{1-f}$$

$$\hat{P}_{s_i} = 0{,}4286 \pm 2{,}052\sqrt{\frac{0{,}2449}{27}}\sqrt{1-0{,}078} \begin{cases} \nearrow 0{,}6163 = 61{,}63\% \\ \searrow 0{,}2409 = 24{,}09\% \end{cases}$$

Analicemos los anteriores resultados: en primer lugar, el estimador puntual fue de 42,86% y el verdadero (poblacional) es de 48,73%, el cual se encuentra ubicado dentro de los Límites de confianza del 95%.

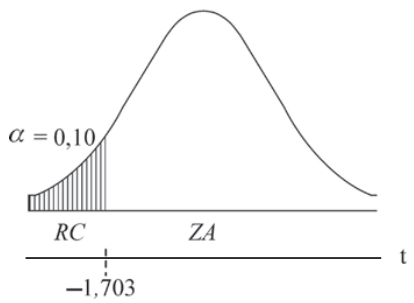
En la formulación de hipótesis, se procede en forma similar al promedio. Supongamos que el investigador asegura que el 45% de las familias son propietarias de las viviendas.

1. *Ho*: P = 0,45
   Ha: P = 0,45
2. α = 0,05

3. Como 0,45 está dentro de los límites de confianza (24,09% y 61,63%), se acepta la hipótesis nula, es decir, que el 45% de las familias son propietarias. Ahora, si observamos la proporción verdadera (48,73%), notamos que se ha cometido un error de tipo II (aceptar algo falso). En una prueba unilateral, por ejemplo, si el investigador considera que el porcentaje puede ser inferior al 45%; en este caso se deberá proceder de la siguiente manera:

1. *Ho*: P = 0,45
   Ha: P < 0,45
2. α = 0,05
3. $t = \dfrac{p - P}{\sqrt{\dfrac{pq}{n-1}}} = \dfrac{0{,}4286 - 0{,}45}{\sqrt{\dfrac{0{,}2449}{27}}} = -0{,}22$

$\upsilon = n - 1 = 27$
$\alpha = 0{,}10$ $\Big\}\ t = -1{,}703$



El valor t = -0,22 se ubica en la zona de aceptación, por lo tanto, al nivel del 5%, no existe suficiente razón para considerar que el porcentaje de familias propietarias sea inferior al 45%. Aquí también se comete un error de tipo II, dado que el promedio verdadero es superior al 45%.

ota de fami ias con i ienda

Para la *estimación de tota es*, se aplica la siguiente fórmula:

$$\hat{A}_{s_i} = Np \pm tN\sqrt{\frac{pq}{n-1}}\sqrt{1-f}$$

$$\hat{A}_{s_i} = 355(0{,}4286) \pm 2{,}052(355)\sqrt{\frac{0{,}2449}{27}}\sqrt{1-0{,}078} = \begin{cases} 218{,}77 \\ 85{,}54 \end{cases} \quad \hat{A} = 355(0{,}4286) = 152 \text{ familias}$$

El estimado puntual (total) es de 152 familias con vivienda propia; y el estimado por intervalos, el número verdadero de propietarios de vivienda, deberá estar entre 86 y 219 familias, con una seguridad del 95%.

ESTIMACIÓN DE PROPORCIONES Y TOTALES EN CONGLOMERADOS (una etapa)

En el cálculo del *estimador de a proporción de tota* de mujeres por familia, se le debe dar el tratamiento de *cong omerados desigua es*, ya que cada familia se constituye en un conglomerado y el número de personas que lo conforman es diferente en cada una de ellas.

Para e estimado de a proporción de mujeres se tendrá:

$$p = \frac{\Sigma a_i}{\Sigma m_i} \qquad p = \frac{\text{número de mujeres}}{\text{total de personas}} \qquad p = \frac{57}{93} = 0{,}6129 = 61{,}29\%$$

$$\hat{P}_{s_i} = p \pm t\sqrt{\left[\frac{1-f}{n\overline{m}^2}\right]\left[\frac{\Sigma a_i^2 - 2p\Sigma a_i m_i + p^2\Sigma m_i^2}{n-1}\right]}$$

$$\hat{P}_{s_i} = 0{,}6129 \pm 2{,}052\sqrt{\left[\frac{1-\frac{28}{355}}{28(3{,}32)^2}\right]\left[\frac{151 - 2(0{,}6129)(223) + (0{,}6129)^2(351)}{28-1}\right]}$$

$$\overline{m} = \frac{\Sigma m_i}{n} = \frac{57}{93} = 3{,}32 \qquad \hat{p}_{s_i} = 0{,}6129 \pm 2{,}052(0{,}032) = \begin{cases} 0{,}6785 = 67{,}85\% \\ 0{,}5473 = 54{,}73\% \end{cases}$$

El estimador puntual de 61,29% corresponde al porcentaje de mujeres, por tanto el 38,71% serán hombres. En la población (355) se obtuvo que el 57,47% eran mujeres y el 42,53% hombres. Por otra parte, la proporción de mujeres (en la población) queda incluida dentro de los límites de confianza del 95%, establecida con la muestra.

Para el planteamiento de la hipótesis (nula y alternativa) se sigue el mismo procedimiento realizado para la proporción; siendo innecesario repetir el proceso.

En e estimati o de tota es, se seguirá el proceso visto para la proporción:

$$\hat{A}_{s_i} = Mp \pm tM\sqrt{\left[\frac{1-f}{n\overline{m}^2}\right]\left[\frac{\Sigma a_i^2 - 2p\Sigma a_i m_i + p^2\Sigma m_i^2}{n-1}\right]}$$

M = total de personas en los 355 conglomerados.

El total de personas se estima de la siguiente manera:

M = 355(3,32) = 1.178,16 = 1.179 (El total verdadero es de 1.164 personas)

$$\hat{A}s_i = 1.179(0,6129) \pm 2,052(1.179)\,(0,032) \begin{cases} \nearrow 800,03 = 800 \\ \searrow 645,19 = 645 \end{cases}$$

Total de mujeres en las 355 familias

El total verdadero para la población de 355 familias es de 669 mujeres en relación a las 723 obtenidas a través de la muestra; además, observemos que ese total, como en todos los casos anteriores, se encuentra dentro de los límites de confianza.

ESTIMATIVOS DE RAZONES COMO MÉTODO INDIRECTO PARA ESTIMAR PROMEDIOS Y TOTALES

En el caso de una Razón: mujeres/hombres, se pueden hacer estimativos indirectos sobre el promedio y total de mujeres en las 355 familias. Veamos como se realiza este proceso:

Estimado de la Ra ón: mujeres/hombres

$$\hat{R} = \frac{\Sigma y_i}{\Sigma x_i} \Rightarrow \hat{R} = \frac{y_i}{x_i} = \frac{mujeres}{hombres} \qquad \hat{R} = \frac{\Sigma y_i}{\Sigma x_i} = \frac{57}{36} = 1,58$$

Lo anterior indica, que aproximadamente por cada tres mujeres hay dos hombres. Este valor corresponde a un estimador puntual; en la población, el valor de la razón es de 1,35.

Para los límites de confianza en una razón se emplea la siguiente fórmula:

$$\hat{R} = \left(\frac{\bar{y}}{\bar{x}}\right) \pm t\sqrt{\left[\frac{1-f}{n\bar{x}^2}\right]\left[\frac{\Sigma y_i^2 - 2\hat{R}\Sigma y_i x_i + R^2\Sigma x_i^2}{n-1}\right]}$$

$$\hat{R}_{\hat{I}}^{s} = 1,58 \pm 2,052\sqrt{\left[1-\frac{28}{355}\right]\left[\frac{151 - 2(1,58)(72) + (1,58)^2(56)}{28(1,29)^2}\right]\left[\frac{}{28-1}\right]}$$

Se puede trabajar con $\bar{} = 1,2857$ ó $\bar{X} = 1,39$, pero es usual emplear la media aritmética obtenida a través de la muestra. 2,02

$$\hat{R}_{\hat{I}}^{s} = 1,58 \pm 2,052(0,2152) = \begin{cases} \nearrow 2,02 \\ \searrow 1,14 \end{cases}$$

La media o el promedio obtenido en forma indirecta a través de una razón se calculará de la siguiente manera:

$$\hat{\bar{Y}}_{\hat{R}} = \left(\frac{\bar{y}}{\bar{x}}\right)\bar{X} \pm t\sqrt{\left[\frac{1-f}{n}\right]\left[\frac{\Sigma y_i^2 - 2\hat{R}\Sigma y_i x_i + \hat{R}^2\Sigma x_i^2}{n-1}\right]}$$

Reemplazando, siendo $\bar{x} = 1,29$ será:

$$\hat{\bar{Y}}_{\hat{R}} = (1{,}58)(1{,}29) \pm 2{,}052\sqrt{\frac{151 - 2(1{,}58)(72) + (1{,}58)^2(56)}{28-1}}\sqrt{1-\frac{\frac{28}{355}}{28}}$$

$$\hat{\bar{Y}}_{\hat{R}} = 2{,}04 \pm 0{,}57 \begin{cases} \nearrow 2{,}61 \\ \searrow 1{,}47 \end{cases}$$ Es el promedio de mujeres por familia.

Se podrá obtener el total de mujeres en las 355 familias, usando el método indirecto a través de una razón:

$$\hat{Y}_R = N\left(\frac{\bar{y}}{\bar{x}}\right)\bar{X} \pm tN\sqrt{\left[\frac{1-f}{n}\right]\left[\frac{\Sigma y_i^2 - 2\hat{R}\,\Sigma y_i x_i + \hat{R}^2\Sigma x_i^2}{n-1}\right]}$$

$$\hat{Y}_R = 355(2{,}04) \pm 355(0{,}57) = 724 \pm 200 \begin{cases} \nearrow 924 \\ \searrow 524 \end{cases}$$ Mujeres

ESTIMACIÓN DE PROMEDIOS Y TOTALES MEDIANTE LA REGRESIÓN

La estimación de *promedios tota es*, utilizando la *regresión inea* y una *ra ón*, son considerados como *métodos indirectos de estimación*. El término *regresión* fue inicialmente empleado por Galtón, con el fin de iniciar ciertas relaciones en la teoría de herencia biológica (padre e hijo), aunque con posterioridad el término se hizo extensivo al método estadístico que trata de cuantificar la relación existente entre dos o más variables.

La *regresión* se aplica a variables que tengan una relación lógica, es decir que haya una relación de causalidad entre las variables. Cuando se ha comprobado la existencia de causalidad directa entre dos o más variables, es necesario determinar la función matemática que represente adecuadamente dicha relación. Con las informaciones obtenidas y mediante la construcción de una gráfica, se podrá decidir si la función adecuada es la de una *recta parábo a e ponencia potencia* etc. Una vez determinada la función, se calculan los parámetros.

En la *regresión inea simp e* sólo intervienen dos variables, una dependiente o predictando y otra independiente o predictor.

La *regresión* aplicada como técnica de muestreo, es un método indirecto para realizar estimaciones, de promedios y totales, lo mismo a partir de una razón (R). La importancia de estos métodos radica en el mejoramiento de los procedimientos de estimación. A menudo se emplea la *regresión inea* entre dos variables, por ser la más sencilla de realizar aunque no siempre la más apropiada para un conjunto de observaciones. Por tal razón se recomienda realizar un trabajo preliminar a fin de determinar si el modelo lineal es el más aconsejable. El método más sencillo, para determinar qué tipo de línea se ajusta mejor a ese conjunto de datos, es la graficación; permite visualizar si la relación lineal es apropiada; en caso contrario, se debe buscar un modelo no lineal. Una vez efectuadas las estimaciones, es necesario determinar el grado de correspondencia o dependencia que exista entre las dos variables, mediante el cálculo de un índice abstracto denominado *coeficiente de correlación al cuadrado,* simbolizado por R^2 y cuyo valor oscila entre 0 y 1. $(0 \leq R^2 \leq 1)$

La ecuación de la recta Y en X se expresa de la siguiente manera, para las variables poblacionales: $(\hat{Y} = c + b$, siendo Y la característica cuantitativa o variable objeto de estudio, denominada *ariab e*

dependiente o *predictando*; es una variable auxiliar supuestamente conocida y relacionada con ***y***, denominada *ariab e independiente o predictor*.

es el *coeficiente angular* e indica la relación entre la variación de Y y la variación de . También se le conoce con el nombre de coeficiente de regresión de Y en X o pendiente de la recta.

c es el *coeficiente de posición,* denominado también origen de la ordenada y corresponde al segmento interceptado sobre el eje de la Y (ordenada) en un plano cartesiano.

Cuando se estima el valor promedio Y, se obtendrá mediante las ecuaciones:

$$\hat{\bar{Y}}_{RL} = c + b\bar{X} \qquad \text{o} \qquad \hat{\bar{Y}}_{RL} = \bar{y} + b(\bar{X} - \bar{x})$$

Siendo $\hat{\bar{Y}}$ el estimador del promedio poblacional de la variable; $\bar{X}$ es el promedio en la población de la segunda variable, supuestamente conocida; ***y*** y ***x*** son promedios obtenidos a través de la muestra.

También se puede calcular al estimador del total poblacional, siendo igual a: $\hat{Y} = N\bar{y}_{RL}$.

El *coeficiente angular* tanto para el promedio como para el total, se puede obtener por diferentes métodos, cuyos resultados obtenidos son iguales:

$$b = \frac{\Sigma Z_x Z_y}{\Sigma Z_x^2} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} \qquad b = \frac{m_{xy}}{s_x^2} = \frac{\text{Covarianza}}{\text{Varianza de } x}$$

$$m_{xy} = \frac{\Sigma x_i y_i}{n} - \bar{x}\,\bar{y} \text{ (Covarianza)} \qquad \text{Siendo: } s_x^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n}$$

(Varianza de la variable)

También se puede calcular *b* directamente

$$b = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \qquad \text{ó} \qquad b = \frac{\Sigma x_i y_i - n\bar{x}\,\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}$$

En aquellas ecuaciones donde sea necesario calcular el coeficiente de posición ***c***, para la estimación del promedio se podrá obtener por cualquiera de las siguientes fórmulas:

$$c = \bar{y} - \bar{x}\,b \qquad c = \frac{\Sigma y_i - b\Sigma x_i}{n}$$

También los valores de *b* y *c* se podrán hallar por el *método de os mínimos cuadrados*, para ello se utilizará un sistema de ecuaciones normales:

(1) $\Sigma yi = nc + b\Sigma_{\ i}$ (2) $\Sigma_{\ i\ i} = c\Sigma_{\ i} + b\Sigma_{\ i}^{\ 2}$

En las estimaciones de totales, el *coeficiente de posición* ***c*** será igual al producto del tamaño poblacional por el coeficiente de posición en la estimación del promedio: (N*c*)

Veamos la aplicación de estas fórmulas con los datos obtenidos a través de la encuesta por muestreo, realizada en una población de 355 familias y cuyo tamaño muestral se calculó en 28.

Ejemplo:

Supongamos que se va a estimar el consumo promedio de carne (Y), en relación a los niveles de ingreso (*x*) de las familias.

Continúa

x_i	y_i	x_i	y_i
2.700	812	1.000	650
1,920	630	1.520	666
880	510	3.460	718
860	506	950	515
1.840	620	980	560
1.860	716	1.950	710
1.260	520	960	782
2.060	780	1.640	720
1.700	895	1.760	543
960	520	2.820	742
1.890	576	1.880	560
1.110	620	2.060	616
1.980	586	1.450	580
1.790	760	1.010	630

$$\Sigma x_i = 46.250$$
$$\Sigma y_i = 18.043$$
$$\Sigma x_i y_i = 30.686.590$$
$$\Sigma x_i^2 = 87.416.300$$
$$\Sigma y_i^2 = 11.923.031$$
$$n = 28$$
$$\bar{x} = \frac{46.250}{28} = 1.651,78$$
$$\bar{y} = \frac{18.043}{28} = 644.39$$

Un primer método de cálculo de los *coeficientes de regresión*, sería trabajando con un *sistema de ecuaciones normales*:

(1) $\Sigma y_i = nc + b\Sigma x_i$

(2) $\Sigma x_i y_i = c\Sigma x_i + b\Sigma x_i^2$

Reemplazando:

(1) $18.043 = 28c + 46.250\,b$

(2) $30.686.590 = 46.250c + 87.416.300b$

Multiplicamos la segunda ecuación por 28 y a la primera por 46.250 para eliminar c:

$$\begin{array}{rrrrr} (2) & 859.224.520 & = & 1.295.000c + 2.447.656.400b \\ (1) & -\ 834.488.750 & = & -\ 1.295.000c - 2.139.062.500b \\ \hline & 24.735.770 & = & 0 \qquad 308.593.900\,b \end{array}$$

$$b = \frac{24.735.770}{308.593.900} = 0,080156 \cong 0,0802$$

Ahora reemplazamos en la ecuación (1)

(1) $18.043 = 28c + 46.250\ (0,0802)$

$$c = \frac{18.043 - 46.250\ (0,0802)}{28} = 511,9196 \cong 511,92$$

Otros procedimientos de cálculo

$$b = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} \quad \Rightarrow \quad b = \frac{28(30.686.590) - (46.250)(18.043)}{28(87.416.300) - (46.250)^2} = 0,0802 \text{ (aprox.)}$$

también:

$$b = \frac{\Sigma x_i y_i - n\bar{x}\,\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} \quad \Rightarrow \quad b = \frac{30.686.590 - 28(1.651,79)(644,39)}{87.416.300 - 28(1.651,79)^2} =$$

$$= \ 0,0802 \text{ (aprox.)}$$

$$b = \frac{cov}{S_x^2} \qquad cov = \frac{\Sigma x_i y_i}{n} - \bar{x}\,\bar{y} \ \Rightarrow \ cov = \frac{30.686.590}{28} - (1.651,79)(644,39) =$$

$$= \ 31.552,68 \text{ (aprox.)}$$

$$S_x^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n} \ \Rightarrow \ S_x^2 = \frac{87.416.300 - 28(1.651,79)^2}{28} \ = \ 393.600,51$$

$$S_y^2 = \frac{\Sigma y_i^2 - n\bar{y}^2}{n} \ \Rightarrow \ S_y^2 = \frac{11.923.031 - 28(644,39)^2}{28} \ = \ 10.584,06$$

Siendo $\hat{\bar{Y}} = bx + c$ $\Rightarrow$ $\hat{\bar{Y}} = 0,0802(1.700) + 511,92 = 648,26$ (gramos de consumo de carne)

También: $\hat{\bar{Y}} = b(x - \bar{x}) + \bar{y}$ $\Rightarrow$ $\hat{\bar{Y}} = 0,0802(1.700 - 1.652) + 644,39 = 648,24$ grs.

Para fijar los límites de confianza del 95%, se debe calcular la *Varianza residual*:

$$\hat{\bar{Y}} = \bar{y}_{RL} \pm t\sqrt{\left[\frac{1-f}{n}\right]\left[\ S_x^2\,(1 - r^2)\ \right]}$$

$$\hat{\bar{Y}} = 648,26 \pm 2,052\ \sqrt{\left[\frac{1 - \frac{28}{355}}{28}\right]\left[10.584,06(1 - 0,24)\right]}$$

$$\hat{\bar{Y}} = 648,26 \pm 33,38 \begin{cases} \nearrow 681,64 \text{ gramos} \\ \searrow 614,88 \text{ gramos} \end{cases}$$

El resultado de 648,26 gramos debe ser el consumo promedio de carne diaria, para familias con ingresos de $1.700 (miles)

El coeficiente de *correlación al cuadrado* o *coeficiente de determinación* de Pearson (R^2) ó (r^2), se calcula así:

$$r^2 = \frac{[n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)]^2}{[n\Sigma x_i^2 - (\Sigma x_i)^2][n\Sigma y_i^2 - (\Sigma y_i)^2]}$$

$$r^2 = \frac{[28(30.686.590) - (46.250)(18.043)]^2}{[28(87.416.300) - (46.250)^2][28(11.923.031) - (18.043)^2]} = 0,24$$

También se puede obtener aplicando la siguiente fórmula:

$$r^2 = R^2 = \frac{cov^2}{S_x^2 S_y^2} \quad \Rightarrow \quad r^2 = \frac{(31.547,32)^2}{(393.613,72)(10.584,06)} = 0,24$$

Algo similar ocurre con el cálculo de los límites de confianza, cuando se requiere calcular el valor de la *Varianza residual*

$$\hat{\bar{Y}} = \bar{y}_{RL} \pm t \frac{s_{yx}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

$$s_{yx}^2 = \frac{\Sigma y_i^2 - c\Sigma y_i - b\Sigma x_i y_i}{n} \qquad \text{o} \qquad s_{yx}^2 = s_y^2 - 2c\text{ov}(b) + b^2 s_x^2$$

Todos estos resultados se obtienen directamente con las calculadoras que tengan el programa LR (Regresión Lineal) o aplicando el programa EXCEL. Sin embargo, los resultados difieren un poco generalmente por los decimales, cuando utilizamos algunas fórmulas que son comunes en el muestreo, tales como:

$$s_x^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1} = \frac{\Sigma (x_i - \bar{x})^2}{n-1} \qquad s_y^2 = \frac{\Sigma y_i^2 - n\bar{y}^2}{n-1} = \frac{\Sigma (y_i - \bar{y})^2}{n-1}$$

$s_x^2 = 408.178,31$ $\qquad$ $s_y^2 = 10.976,07$

La *varianza residual* muchos la calculan dividiendo por n - 2 es decir:

$$s_{yx}^2 = \frac{\Sigma y_i^2 - c\Sigma y_i - b\Sigma x_i y_i}{n-2}$$

Lo anterior hace que los cálculos que hicimos varíen, dependiendo de las correcciones que utilicemos, su aplicación dependerá del criterio que tenga la persona que realiza la investigación.

Si usted requiere estimar el total de consumo de carne diaria (en gramos) para las 355 familias, se hará una pequeña modificación, sabemos que:

$$\hat{Y} = N\bar{y}_{RL} \pm tN\sqrt{s_{\bar{y}}^2}$$

$$\hat{Y} = 355(648,26) \pm 2,062(355)\sqrt{280,71} = 230.132,3 \pm 12.264,39 \begin{cases} \nearrow 242.396,69 \\ \searrow 217.867,91 \end{cases}$$

ESTIMACIÓN DE PROMEDIOS, PROPORCIONES Y TOTALES EN DOMINIO DE ESTUDIO

Tal como se mencionó anteriormente, el *dominio de estudio*, se presenta por un mal diseño de la muestra o por errores en el análisis

Recordemos algunos estimativos de proporciones, promedios y totales, con la información obtenida en la muestra de 28 familias, tomada de la población de 355 familias. El grupo o dominio a analizar son las *familias propietarias de vivienda, sin tener en cuenta el ingreso de aquéllas que no son propietarias, el cual se constituye en error al hacer el análisis*. Los datos para esta condición son los siguientes:

Ejemplo 1:

NÚMERO DE ORDEN	NÚMERO ALEATORIO	INGRESOS (miles $)	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE gr.
			Total	M	F	Trabajando	
1	170	2.700	3	1	2	2	812
2	045	1.920	2	1	1	2	630
3	217	860	5	2	3	1	506
4	132	1.260	3	2	1	1	520
5	087	2.060	5	1	4	2	780
6	111	1.110	3	1	2	2	620
7	248	1.980	3	1	2	2	586
8	303	3.460	4	1	3	3	718
9	230	980	2	1	1	1	560
10	286	1.950	5	1	4	2	710
11	032	960	2	1	1	1	732
12	318	1.760	2	1	1	1	543

Se pide establecer los límites de confianza del 95% para los siguientes estimadores:

a) Ingreso promedio de las familias propietarias de vivienda.

b) Ingreso total de propietarios, si se sabe que existen 173 familias en el barrio investigado.

c) Ingreso total de propietarios, si no se conoce el número de éstos en la población.

d) El promedio de personas correspondiente a familias propietarias de vivienda.

e) El total de personas en las familias poseedoras de vivienda, si no se conoce el número de propietarios en la población.

f) El porcentaje de familias propietarias de vivienda y consumo diario de carne superior a 600 gramos.

Solución:

a) Ingreso promedio de las familias propietarias:

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{21.000}{12} = 1.750 \qquad s = \sqrt{\frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{43.453.400 - 12(1.750)^2}{12-1}} = 780{,}64$$

Si la confianza es del 95% se tendrá que t = 2,201. $\hat{\bar{X}}_{s_i} = \bar{x} \pm t \frac{s}{\sqrt{n}} \sqrt{1-f}$

$f = \frac{n}{N}$ Como no conocemos el número de familias propietarias de vivienda, se tendrá como *fracción de muestreo*: $f = \frac{n}{N} = 28/355 = 0{,}079$

$$\hat{\bar{X}}_{\substack{s\\ i}} = 1.750 \pm 2,201 \frac{780,64}{\sqrt{12}} \sqrt{1-0,079}$$

$$\hat{\bar{X}}_{\substack{s\\ i}} = 1.750 \pm 476,00 \begin{cases} \nearrow 2.226,00 = \$2.226.000 \\ \searrow 1.274,00 = \$\ 1.274.000 \end{cases}$$

b) Ingreso total de las familias propietarias, sabiendo que en la población hay 173 familias.

$N_1 = 173$ $\quad \hat{X}_{\substack{s\\ i}} = N_1 \bar{x} \pm t \frac{N_1 s}{\sqrt{n_1}} \sqrt{1-f}$ $\quad n_1 = 12$ $\quad f = \frac{n_1}{N_1}$

$$\hat{X}_{\substack{s\\ i}} = 173(1.750) \pm 2,201 \left[\frac{173(780,64)}{\sqrt{12}}\right]\left[\sqrt{1-\frac{12}{173}}\right]$$

$$\hat{X} = 302.750 \pm 82.778,24 \begin{cases} \nearrow 385.528,24 \\ \searrow 219.971,76 \end{cases} \text{ (miles de \$)}$$

c) Total de ingresos de las familias propietarias, cuando no se conocen las familias propietarias en la población.

En este caso, no puede ser utilizado $N = 355$, ya que el valor representa la totalidad de familias propietarias y no propietarias. Para determinar el total de ingresos se debe proceder de la siguiente manera:

$$\hat{X} = N\left(\frac{\Sigma x_i}{n}\right) = 355\left(\frac{21.000}{28}\right) = 266.250 \text{ (miles de \$)}$$

Σx_i = Suma de los ingresos de las familias propietarias de vivienda, en la muestra ($n = 12$)

En el cálculo de la varianza, se sigue un proceso similar al utilizado para el total.

$$s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n-1} = \frac{43.453.400 - 28\left(\frac{21.000}{28}\right)^2}{28-1} = 1.026.051,85$$

$$s = \sqrt{1.026.051,85} = 1.012,94$$

Los límites de confianza del 95% para estimado del total serán:

$$\hat{X}_{\substack{s\\ i}} = N\left(\frac{\Sigma x_i}{n}\right) \pm t \frac{N\sqrt{\frac{\Sigma x_i^2 - n\left(\frac{\Sigma x_i}{n}\right)^2}{n-1}}}{\sqrt{n}} \sqrt{1-\frac{n}{N}}$$

$$\left.\begin{array}{l} \upsilon = n-1 = 28-1 \\ \alpha = 0,05 \end{array}\right\} t = 2,048$$

$$\hat{X}_{\substack{s\\ i}} = 355\left(\frac{21.000}{28}\right) \pm 2,048 \left[\frac{355(1.012,94)}{\sqrt{28}}\right] \sqrt{1-\frac{28}{355}}$$

$$\hat{X}_{s_I} = 266.250 \pm 133.574,24 \begin{cases} \nearrow 399.824,24 \\ \searrow 132.675,76 \end{cases} \quad \text{(miles de \$)}$$

Sería más fácil el procedimiento, si primero estimamos el total de familias propietarias, así:

Proporción de propietarios = 12/28 = 0,4286; luego, el número de total de propietarios

$N = 0,4286(355) = 152$ (Total propietarios de vivienda)

$$\hat{X}_s = N\bar{x} \pm tN \frac{s}{\sqrt{n}} \sqrt{1-\frac{n}{N}}$$

$$\hat{X} = 152(1.750) \pm 2,201(152) \frac{780,64}{\sqrt{12}} \sqrt{1 - \frac{12}{152}}$$

$$\hat{X} = 266.000 \pm 72.354,58 \begin{cases} \nearrow 338.354,58 \\ \searrow 193.645,42 \end{cases} \quad \text{(miles de \$)}$$

d) Promedio de personas de familias propietarias de vivienda.

La *media aritmética*: $\bar{x} = \frac{\Sigma x_j}{n}$ $\quad \bar{x} = \frac{39}{12} = 3,25$

La *desviación típica* o estándar:

$$s = \sqrt{\frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{143 - 12(3,25)^2}{12-1}} = 1,22$$

Los límites de confianza del 95% $\quad \upsilon = n - 1 = 12 - 1 = 11$ $\quad t = 2,201$
$\alpha = 0,05$

$$\hat{X}_{s_I} = \bar{x} \pm t \frac{s}{\sqrt{n}} \sqrt{1-f}$$

$$\hat{X}_{s_I} = 3,25 \pm 2,201 \frac{1,22}{\sqrt{12}} \sqrt{1 - \frac{28}{355}} = 3,25 \pm 0,74 \begin{cases} \nearrow 3,99 \\ \searrow 2,51 \end{cases} \quad \text{(Personas por familia)}$$

e) Para estimar el *total de personas, en las familias propietarias de vivienda*, se presentan dos situaciones; en primer lugar, cuando se conoce el número de familias propietarias ($N_1 = 173$).

$$\hat{X}_{s_i} = N_1\bar{x} \pm tN_1 \frac{s}{\sqrt{n_1}} \sqrt{1-f}$$

$$\hat{X}_{s_i} = 173(3,25) \pm 2,201(173) \frac{1,22}{\sqrt{12}} \sqrt{1 - \frac{12}{173}} = 562,25 \pm 129,36 \begin{cases} \nearrow 691,61 = 692 \\ \searrow 432,89 = 433 \end{cases}$$

Número de personas en las familias propietarias de vivienda.

En segundo lugar, lo que más interesa, cuando no se conoce el número de propietarios, es decir, cuando el tamaño poblacional del subgrupo o dominio estudiado es desconocido.

$$\hat{X}_{\substack{s\\1}} = 355\left(\frac{39}{28}\right) \pm 2{,}048(355)\ \frac{1{,}81}{\sqrt{28}}\ \sqrt{1-\frac{28}{355}}$$

$$\hat{X}_{\substack{s\\1}} = 494{,}46 \pm 238{,}68 \begin{cases} \nearrow\ 733{,}14 \Rightarrow (733) \\ \searrow\ 225{,}78 \Rightarrow (256) \end{cases} \text{ Personas}$$

$$s = \sqrt{\frac{143 - 28(39/28)^2}{28-1}} = 1{,}81$$

f) En cuanto hace referencia al *número de familias propietarias, con un consumo diario de carne superior a los 600 gramos*, en la muestra (n = 12) encontramos 6 familias. Si, en primer lugar, consideramos las 173 familias propietarias de vivienda, se tendrá:

$$\hat{A}_{\substack{s\\i}} = N_1 p \pm tN_1\sqrt{\frac{pq}{n-1}}\sqrt{1-f}$$

$$\hat{A}_{\substack{s\\i}} = 173\left(\frac{6}{12}\right) \pm 2{,}201(173)\sqrt{\frac{\left(\frac{6}{12}\right)\left(\frac{6}{12}\right)}{12-1}}\sqrt{1-\frac{12}{173}} = 86{,}5 \pm 55{,}38 \begin{cases} \nearrow\ 141{,}88 = 142 \\ \searrow\ 31{,}12 = 31 \end{cases} \text{ Familias}$$

Cuando no se conoce el número de familias propietarias de vivienda, en la estimación de una proporción, se procede exactamente igual al del promedio.

$$\hat{A} = Np = N\left(\frac{\Sigma a_i}{n}\right) = 355\left(\frac{6}{28}\right) = 76{,}07$$

$$s_p = \sqrt{pq} = \sqrt{\left(\frac{6}{28}\right)\left(\frac{22}{28}\right)} = \sqrt{(0{,}21)(0{,}79)} = 0{,}407$$

$$\hat{A}_{\substack{s\\i}} = Np \pm tN\sqrt{\frac{pq}{n-1}}\sqrt{1-f}$$

$$\hat{A}_{\substack{s\\i}} = 355\left(\frac{6}{28}\right) \pm 2{,}048(355)\sqrt{\frac{(0{,}21)(0{,}79)}{28-1}}\sqrt{1-\frac{12}{173}} \quad \text{ó}$$

$$\hat{A}_{\substack{s\\i}} = 76{,}07 \pm 54{,}98 \begin{cases} \nearrow\ 131{,}05 = 131 \\ \searrow\ 21{,}09 = 21 \end{cases} \Big\} \text{familias}$$

El *estimador puntual*, cuando se conoce la población del subgrupo, será 87 familias de las 173, cuyo consumo de carne diario es superior a los 600 grs.

Cuando no se conoce la población del subgrupo (dominio), el número de familias, con dicho consumo se estima en 76 familias.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

27. El censo de establecimientos comerciales realizado en un país, reveló que había 3.720 clasificados como medianos. De esa población se tomó una muestra aleatoria de 30, obteniéndose la siguiente información:

No.	PERSONAL EMPLEADO	SEXO		VENTA SEMANAL (Millones $)	PROPIEDAD LOCAL	ACTIVIDAD COMERCIAL	PERSONAL AFILIADO AL SEGURO
		H	M				
1	4	2	2	37,8	si	alm.telas	3
2	2	2	0	22,3	no	fama*	0
3	3	1	2	28,4	si	tienda	1
4	5	2	3	49,6	no	almacén	5
5	2	0	2	19,6	no	tienda	0
6	3	0	3	27,4	si	tienda	1
7	6	2	4	42,6	si	panadería	4
8	8	2	6	51,4	no	almacén	6
9	2	2	0	23,6	no	tienda	0
10	5	3	2	42,5	si	tienda	3
11	6	3	3	51,4	no	almacén	5
12	2	1	1	30,6	no	tienda	0
13	4	2	2	34,5	si	papelería	2
14	6	2	4	48,6	si	almacén	4
15	3	1	2	31,4	no	tienda	1
16	2	0	2	18,6	no	droguería	0
17	2	1	1	19,7	si	fama*	2
18	3	1	2	27,6	si	licores	0
19	4	1	3	38,4	si	tienda	0
20	2	2	0	19,3	no	tienda	1
21	4	1	3	41,4	no	almacén	3
22	6	2	4	53,6	si	almacén	3
23	2	1	1	42,5	si	tienda	2
24	3	2	1	20,6	si	licores	3
25	2	0	2	32,8	si	tienda	0
26	3	1	2	31,6	no	droguería	0
27	4	1	3	42,5	si	papelería	2
28	5	2	3	53,4	no	almacén	4
29	2	1	1	23,6	si	fama*	2
30	3	2	1	32,5	no	tienda	2

* Venta de carne.

Estimar el total de ventas semanales (miles de $) por empleado en los establecimientos cuya actividad aparece como almacén.

28. Con los datos del ejercicio 27, estimar la proporción y total de establecimientos que aparecen con actividad comercial TIENDA y que sus gestores sean propietarios del local.

29. Con los datos del ejercicio 27, estimar el total de hombres que trabajan en los establecimientos comerciales que funcionan en locales propios.

30. Con los datos del ejercicio 27, estimar el promedio y el total de ventas en millones de pesos por empleado.

31. Con los datos del ejercicio 27, estimar la proporción y el total de personas afiliadas al seguro de salud y pensión.

32. Con los datos del ejercicio 27, estimar el promedio y el total de mujeres empleadas en los establecimientos comerciales.

33. Mediante la Regresión lineal, con los datos del ejercicio 27, estimar el promedio y el total de ventas semanales por empleado, sabiendo que el promedio de empleados por establecimiento es de 3,8 ($\overline{X}$= 3,8), fijando límites del 95%.

34. Estimar el total de ventas semanales (millones de $), para los establecimientos que tienen como actividad comercial TIENDA, fijando límites del 95%.

35. De acuerdo al censo levantado en 2011 en una ciudad del país, se reveló que existían aproximadamente 15.000 viviendas. De esta población se tomó una muestra aleatoria de 35 viviendas, obteniéndose información relativa a la clasificación según: a) propiedad de la vivienda; b) número de personas clasificadas por sexo; c) que hayan ido o no al médico durante el último año; d) que hayan ido o no al odontólogo durante el último año. La información es la siguiente:

NO. DE ORDEN	PROPIETARIO VIVIENDA	NÚMERO PERSONAS	NÚMERO		VISITARON MÉDICO		VISIT. ODONTÓLOGO	
			Hombres	Mujeres	Si	No	Si	No
1	si	5	1	4	5	0	1	4
2	no	6	3	3	0	6	0	6
3	si	3	1	2	1	2	1	2
4	si	3	1	2	3	0	2	1
5	si	2	1	1	0	2	0	2
6	no	3	1	2	0	3	0	3
7	si	3	1	2	0	3	1	2
8	no	3	1	2	0	3	1	2
9	no	4	2	2	0	4	1	3
10	no	4	3	1	0	4	0	4
11	si	3	2	1	2	1	1	2
12	no	2	1	1	0	2	0	2
13	no	7	3	4	4	3	2	5
14	si	4	3	1	4	0	1	3
15	no	3	2	1	1	2	0	3
16	si	4	3	1	4	0	1	3
17	no	4	1	3	0	4	4	0
18	n o	4	1	3	2	2	1	3
19	no	3	2	1	1	2	1	2
20	no	3	1	2	3	0	0	3
21	si	4	1	3	4	0	1	3
22	no	3	2	1	0	3	0	3
23	sí	3	1	2	0	3	1	2
24	no	1	0	1	0	1	0	1
25	no	2	1	1	2	0	2	0
26	sí	4	2	2	2	2	1	3
27	si	3	1	2	1	2	1	2
28	si	4	3	1	2	2	1	3
29	no	2	0	2	1	1	0	2
30	no	4	2	2	1	3	2	2
31	si	3	1	2	2	1	0	3
32	no	5	4	1	3	2	3	2
33	si	1	1	0	0	1	1	0
34	si	2	0	2	0	2	1	1
35	no	3	3	0	1	2	2	1
Σ	-	117	56	61	49	70	34	83

Se pide calcular los estimados puntuales y fijar límites de confianza del 95%.
a) Proporción de mujeres.
b) Promedio de personas.
c) Total de propietarios de vivienda.
d) Número de personas que visitan al odontólogo.

36. Con los datos del ejercicio 35, estimar:
a) Razón entre las personas que visitan al médico (y_i) y aquéllos que visitan al odontólogo. (x_i)
b) Proporción de personas que visitan al médico.
c) Proporción de personas que visitan al odontólogo.

37. Se realiza una muestra de 10 fusibles, de una cierta marca, para determinar el punto medio de ruptura. Los puntos medios en amperios fueron:

24	18	14	26	20	22	19	22	17	20

a) Determine el estimador puntual.
b) En el punto (a), fije los límites de confianza del 95%.
c) Al nivel del 1%, ¿podría afirmarse que estos resultados son inferiores a los señalados por la empresa, de 22 amperios el punto medio de ruptura?
d) Si la población de fusibles está constituida por 5.000 unidades, sabiendo que su promedio de ruptura es de 20 amperios. ¿qué tipo de error se comete?

38. De la zona de una ciudad con 10.000 familias, se extrajo una muestra aleatoria de 40. El número de personas registradas por familia en la muestra es de:

4	2	4	3	3	4	7	2	4	2	1	3	3	4	3	3	4	4	6	5
3	4	4	3	3	3	2	3	6	5	3	3	4	5	6	3	2	2	1	4

a) Hacer la estimación puntual para el total de personas.
b) Calcular la probabilidad de que dicha estimación esté comprendida dentro de ± 10% del valor verdadero (Nota: se pide el valor de Z).
c) Resolver los puntos a) y b) elaborando una tabla de frecuencias (variable discreta).

39. Al efectuar el inventario de un almacén, se pretende aplicar el muestreo para reducir el trabajo; por tal motivo se contabiliza el valor de los artículos de cada uno de los 36 estantes (N = 36) que tiene el local. Se sabe que la varianza poblacional (σ^2) es aproximadamente 134,53. Además la estimación del valor total obtenido a partir de una muestra debe ser correcta dentro de ±2, con una confianza del 95% . Un asesor sostiene que una muestra aleatoria simple de 12 (n) estantes satisface los requerimientos. ¿Está de acuerdo con esta estimación?

40. Una universidad tiene 7.000 estudiantes matriculados, de los cuales se selecciona una muestra de 700; 480 de ellos menores de 20 años. Obtenga el estimado puntual del total de estudiantes menores de 20 años y luego fije los límites de confianza del 99%.

41. Se tiene la siguiente población de 20 valores

Y_i =	6	8	3	5	10	6	12	8	6	7	2	6	8	7	10	12	5	8	2	4

Información adicional: $\Sigma Y_i = 135 \quad \Sigma Y_i^2 = 1.069$

a) Determine el tamaño de la muestra, si se desea un error de 2 y una confianza del 95%.

b) Utilizando la calculadora o la tabla de números aleatorios, seleccione una muestra de acuerdo al tamaño anterior y estime tanto el promedio como el total, fijando límites de confianza del 95%, en un muestreo aleatorio simple.

42. De una lista de 468 colegios se seleccionó una muestra aleatoria simple de 100, de los cuales resultaron 54 públicos y 46 privados. La siguiente es la información obtenida de los colegios de la muestra donde Y y X se refieren respectivamente, a un número de estudiantes y número de profesores.

COLEGIOS	n	y_i	y_i^2	x_i	x_i^2	$x_i y_i$
Públicos	54	31.281	29.881.219	2.024	111.090	1.729.349
Privados	46	13.707	6.366.785	1.075	33.119	431.041

Se pide, en cada tipo de colegio, estimar la razón entre el número de estudiantes y el número de profesores, estableciendo límites de confianza del 95%.

43. De una lista de 10.000 empleados se obtuvo una M.A.S. de 300, de los cuales 120 demostraron no tener aptitudes para ser promotores de ventas, Determine la proporción y el total de empleados que tiene aptitudes de promotores de ventas y establezca el intervalo de confianza del 90%.

44. Una asociación de odontólogos, realizó una investigación relativa a las piezas dentarias a 300 niños de cierto pueblo. Se seleccionó una muestra aleatoria simple de 40 niños, para determinar el número de piezas cariadas en cada niño. Los resultados fueron:

No. DE PIEZAS CARIADAS:	0	1	2	3	4	5	6	7	8	9	10
No. de niños:	12	6	8	2	2	3	2	1	1	2	1
Varones:	8	3	5	1	0	2	1	1	1	1	0
Niñas:	4	3	3	1	2	1	1	0	0	1	1

a) Estime el total de piezas cariadas en los niños del pueblo.

b) La proporción de niñas sin piezas cariadas.

c) El total de niños (varones) con piezas cariadas.

45. Con los resultados del ejercicio 44 se pide:

a) En las tres características anteriores, fije límites de confianza del 99%.

b) Considere que la muestra de 40 niños, se realizó como encuesta preliminar (piloto) y determine los tamaños muestrales para el promedio de piezas cariadas por niño, con un error de ± 2, y la proporción de niños (varones) sin caries, con un error del 8%. En ambos casos la confianza es del 95%.

46. En cuatro barrios de la ciudad, conformados por 48 manzanas, se eligieron aleatoriamente 20 familias. A cada familia seleccionada, se le preguntó por el número de miembros y posesión de vehículo. Las respuestas fueron:

FAMILIAS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
y_i :	3	5	6	5	5	3	4	3	5	3
x_i :	1	2	2	3	2	2	1	2	4	2

FAMILIAS	1	2	3	4	5	6	7	8	9	10
y_i :	3	5	6	5	5	3	4	3	5	3
x_i :	1	2	2	3	2	2	1	2	4	2

N = 2.000 familias y_i = Número de personas por familia x_i = Número de vehículos

Se pide:

a) Fijar límites de confianza del 95% para la estimación del número de personas por vehículo, siendo $\overline{X} = 2$.

b) Si el número de vehículos particulares en esos cuatro barrios es de 3.500, ¿será igual el total de personas en dicha zona?

c) ¿Cuál es el promedio estimado de personas por familia?

47. Con la siguiente información, correspondiente a una encuesta realizada a un grupo de 25 empleados, mediante el muestreo aleatorio simple, en una empresa importadora, sobre niveles de ingresos y gastos semanales en alimentación (miles de $), se pide estimar la razón entre ingresos y gastos semanales con una confianza del 99%.

$\Sigma x_i = 405$ $\Sigma y_i = 183$ $\Sigma x_i^2 = 22.687$ $N = 300$ $\Sigma y_i^2 = 4.157$ $\Sigma x_i y_i = 9.612$

48. Un empresario se propone probar la hipótesis de que las normas de llenado en botellas de un litro, de cierta marca de aceite, tiene como media 30 cm^3. La muestra examinada revela las siguientes normas (cm^3):

20	25	35	25	5	30	20	40	15	35	20	25

a) Establezca el intervalo de confianza del 95% de la media obtenida.

b) ¿Es correcta la hipótesis del investigador, a un nivel de significación del 5%?

c) Si la media (verdadera) de norma es de 28 cm^3, ¿qué tipo de error se comete?

49. Se ha medido el contenido de nicotina de 36 cigarrillos de una determinada marca. A continuación se resumen los resultados:

x_i = Contenido de nicotina de un cigarrillo, medido en miligramos

$\overline{x}$ = 576 miligramos $\Sigma(x_i - \overline{x})^2 = 315$ miligramos

Obtenga un intervalo de confianza del 99%, para estimar el contenido promedio de nicotina.

50. En una zona residencial con 3.000 viviendas se requiere estimar el porcentaje de habitadas por: a) su propietario; b) familias que poseen dos automóviles. Se desea que los errores de muestreo de esos porcentajes estimados no superen el 2% y 1% respectivamente. Se presume que el porcentaje de propietarios de vivienda en dicha zona debe estar entre el 45 y el 65% y el de propietarios de automóviles entre el 5 y 10%. ¿Qué tamaño de muestra se requiere, para lograr ambos propósitos?

51. En un barrio de la ciudad viven 150 familias. Una muestra aleatoria de 15 familias suministra información relativa al ingreso familiar mensual y a su condición de propietarios o arrendatarios de vivienda.

FAMILIAS	INGRESOS (millones de $)	PROPIETARIOS
1	5,0	si
2	3,5	si
3	4,0	si
4	2,5	no

Continúa...

Continuación

5	3,5	si
6	4,0	si
7	2,0	no
8	2,5	no
9	1,5	no
10	3,0	no
11	4,2	si
12	2,8	no
13	3,6	si
14	1,8	no
15	2,6	no

En cada uno de los siguientes casos obtener el estimador puntual y fijar límites de confianza del 95%.
a) Ingreso promedio de las familias del barrio.
b) Ingreso total de las familias del barrio
c) Ingreso promedio de los propietarios.

52. Con los datos del ejercicio 51 se pide calcular:
a) Ingreso total de los propietarios, si se sabe que existen 50 propietarios.
b) Ingreso total de los propietarios, si no se dispone de la información del número de propietarios.
c) Porcentaje de arrendatarios.

53. Con los datos del ejercicio 51 se pide calcular:
a) Número de arrendatarios.
b) Porcentaje de arrendatarios con ingresos superiores a 2,5 (mill. de $)
c) Número de arrendatarios con ingresos superiores a 2,5 (mill. de $)
d) Porcentaje de familias con ingresos superiores a 3,7 (mill. de $).

54. Se realizó una encuesta a cierto grupo de personas para estimar el ingreso total del grupo y el ingreso promedio (miles de $) de las personas. Se usó una estimación por el método de la razón con viariable auxiliar X_i (ingreso de hace 5 años). Los datos son:

x_i :	900	800	600	500	500	800	600
y_i :	2.500	2.000	2.700	1.800	3.000	2.500	2.000

$X = 210.000 \qquad N = 400$

Se pide estimar $\hat{Y}$; $\hat{\bar{Y}}$ y $\hat{R}$ fijando límites del 95% para los estimadores.

55. Se realiza una encuesta a 17 familias en cuanto a: tamaño, ingreso mensual, gastos en alimentos quincenal (incluyendo servicios), y si tienen suscripción a algún periódico de la ciudad. Los resultados fueron:

FAMILIAS	NÚMERO DE PERSONAS	INGRESOS (miles $)	GASTOS EN ALIMENTOS INCLUIDOS SERVICIOS (miles $)	SUSCRIPCIÓN PERIÓDICO
1	2	2.620	820	si
2	3	1.840	960	no
3	3	1.930	1.010	no
4	5	1.450	760	no

Continúa...

Continuación

5	4	2.620	1.220	si
6	7	3.010	1.130	si
7	2	2.420	1.040	no
8	4	2.620	1.230	no
9	2	2.210	1.020	si
10	5	2.060	1.020	no
11	3	2.420	1.130	no
12	6	2.030	840	no
13	4	1.860	760	si
14	4	1.520	780	no
15	2	1.790	790	no
16	5	1.940	830	no
17	3	2.820	1.140	si

Se requiere estimar para una población de 1.200 familias:

a) Promedio de personas por familia.
b) Gasto en alimentos por persona.
c) Proporción de familias con suscripción al periódico
(En todos los casos fijar límites de confianza del 95%)

56. Se realiza una investigación por muestreo en una industria, relacionando los puntajes de aptitud con la productividad después de tres meses de entrenamiento del personal. Con los postulantes elegidos al azar se obtuvieron 10 pares de puntajes de aptitud y productividad, tal como se indica a continuación:

APTITUD	:	9	12	17	20	18	20	24	23	20	25
PRODUCTIVIDAD	:	23	28	30	32	30	36	38	32	34	36

Se pide:

a) Estimar el promedio de productividad si consideramos $\bar{x} = 20$, con límites del 95%. b) Estimar el total de productividad, con límites del 95%, sabiendo que N = 500, mediante la recta de regresión. c) Calcular el coeficiente de correlación lineal.

57. Con los siguientes datos obtenidos a través de una muestra de 20 unidades, se obtuvo la siguiente información:

$\Sigma x_i = 174$ $\Sigma y_i = 340$ $\Sigma x_i^2 = 1.626$ $n = 20$ $\Sigma y_i^2 = 5.904$ $\Sigma x_i y_i = 3.062$

a) Estimar $\hat{\bar{Y}}_{RL}$, fijando límites de confianza del 90%, siendo $\bar{x} = 12$
b) Estimar $\hat{Y}_{RL}$, fijando límites de confianza del 90%, siendo N = 300

58. Los siguientes datos fueron obtenidos a través de una encuesta por el método de muestreo aleatorio simple, cuyo tamaño fue de 30.

x_i:	2	2	4	6	6	6	7	7
y_i:	1	2	3	1	2	3	2	4
n_i	2	8	5	4	3	2	4	2

a) Estimar por el método indirecto de la regresión lineal (ponderada) el promedio para la variable Y, considerando $\bar{x} = 5{,}8$.
b) Estimar por el método indirecto de la regresión lineal (ponderada) el total para la variable Y, considerando N = 500.

59. Con la información del ejercicio No. 56 estimar $\hat{R}$, $\hat{\bar{Y}}_R$, $\hat{Y}_R$ fijando límites de confianza del 90% considerando Y = Productividad y X = Aptitud.

60. Hallar $\hat{R}$, $\hat{\bar{Y}}_R$, $\hat{Y}_R$ y sus límites de confianza del 95% tomando la información de los ejercicios 57 y 58.

61. Suponga que el administrador de un pequeño negocio, mantiene en depósito 2.000 artículos y desea elaborar el inventario trimestral con el fin de preparar un informe (balance y estado de ingresos). Considera innecesario y dispendioso el conteo físico de cada unidad para todos los 2.000 artículos y decide obtener una muestra de 30. De cada uno de los 30 artículos cuenta el número de unidades en existencia y luego determina el costo por artículo, multiplicando el número de unidades existentes por el costo unitario. Sus resultados fueron:

ARTÍCULO	:	1	2	3	4	5	6	7	8	9	10
COSTO (miles $)	:	4.320	800	1.200	3.650	2.150	300	580	7.200	6.000	5.100
ARTÍCULO	:	11	12	13	14	15	16	17	18	19	20
COSTO (miles $)	:	6.840	1.150	1.820	1.200	1.300	840	260	380	4.120	6.150
ARTÍCULO	:	21	22	23	24	25	26	27	28	29	30
COSTO (miles $)	:	420	1.100	1.260	840	7.200	640	1.150	2.100	2.160	480

Estimar el costo total de los 2.000 artículos (inventario total) fijando límites del 95%.

62. Un administrador considera que el costo del inventario anterior (ver ejercicio 61), correspondiente al 31 de marzo de 2010 está correlacionado con el costo del inventario a 31 de diciembre de 2006. Para ello dispone de la información para los 30 artículos de acuerdo a los registros de inventario a 31 de diciembre del 2011, así:

ARTÍCULO	:	1	2	3	4	5	6	7	8	9	10
COSTO (miles $)	:	3.600	960	1.510	2.150	2.800	1.000	420	360	4.600	4.800
ARTÍCULO	:	11	12	13	14	15	16	17	18	19	20
COSTO (miles $)	:	3.600	960	1.160	1.000	1.150	920	120	240	2.630	6.240
ARTÍCULO	:	21	22	23	24	25	26	27	28	29	30
COSTO (miles $)	:	310	860	1.150	920	6.130	480	530	1.620	1.510	210

Si suponemos que el costo promedio (inventario a 31 de diciembre) fue de $2.728,6 miles. Se puede mediante la regresión, estimar el promedio del costo del inventario para el 31 de marzo de 2011 por artículo. Nivel de confianza, 95%.

63. En una zona comercial de la ciudad se estudió el número de personas que visitan a cada uno de los almacenes y cuántos de ellos compran, con lo cual se pudo determinar el valor promedio de compra por visitante. Con base en los resultados anteriores se desarrolló un programa de publicidad y presentación de los artículos y atención al visitante, posteriormente se volvió a calcular el promedio de ventas por visitante, concluyendo que el programa desarrollado fue eficaz.

De acuerdo con el resultado anterior a la muestra, el número de visitantes para estos almacenes fue:

ALMACÉN :	1	2	3	4	5	6	7	8	9	10
VISITANTES POR HORA :	18	26	30	41	23	42	28	31	46	42
ALMACÉN :	11	12	13	14	15	16	17	18	19	20
VISITANTES POR HORA :	40	36	25	28	36	35	38	42	40	39

Ha pasado un determinado tiempo, desde el último estudio, el cual no se volvió a realizar por ser demasiado costoso, obligando al diseño de una muestra aleatoria simple de 20 almacenes, para determinar el número de visitantes por hora (en una semana de observación), con el siguiente resultado:

ALMACÉN :	1	2	3	4	5	6	7	8	9	10
VISITANTES POR HORA :	21	36	42	56	28	60	34	42	58	36
ALMACÉN :	11	12	13	14	15	16	17	18	19	20
VISITANTES POR HORA :	48	54	36	42	63	51	70	68	56	54

Si el número de visitantes (promedio) para los 2.000 almacenes (total) es de 37,6. Se pide estimar:

a) El total de visitantes por hora, sin utilizar los datos de la segunda encuesta.

b) El total de visitantes por hora, utilizando los datos de la segunda encuesta. Aplicar el metodo de la regresión.

64. Con los datos del ejercicio anterior, se pide estimar:
a) La proporción de almacenes con un número de visitantes por hora, superior a 50 personas. (Segunda encuesta).
b) El total de almacenes que reciben a más de 50 personas por hora. (Segunda encuesta).
Nota: en cada uno de los estimadores fijar límites de confianza del 95%.

65. En una ciudad cualquiera se ha determinado, que el número de establecimientos comerciales es de 9.200, con un máximo de 3 empleados por establecimiento. Se quiere saber si el local donde funciona es propio o arrendado y si tiene uso de baños exclusivo o se comparte con el de alguna vivienda vecina. Se investigaron 800 establecimientos de los cuales se obtuvo la siguiente información:

CON BAÑO	LOCAL PROPIO	LOCAL ARRENDADO
SI	241	409
NO	39	111

Con los anteriores datos se pide el estimador puntual y los límites de confianza del 95%, así:

a) Establecimientos con local arrendado, y el porcentaje con baño propio.

b) Total de establecimientos en local arrendado y sin baño propio.

c) Se sabe que el total de establecimientos con local arrendado es de 6.400, obtener una nueva estimación del número de ellos, con local y baño propio.

66. Con la siguiente muestra de 12 establecimientos comerciales pequeños (con menos de 8 empleados):

ESTABLECIMIENTO :	1	2	3	4	5	6	7	8	9	10	11	12
No. EMPLEADOS :	7	6	5	4	7	6	5	4	7	7	5	4

EMPLEAD. CON MENOS DE 5 AÑOS DE SERVICIO :	3	2	3	0	2	2	1	3	4	3	2	3
SALARIO PROMEDIO (miles de $) :	664,8	972,5	870,6	1.902,0	1.830,0	985,0	980,0	774,6	865,3	768,4	986,2	848,0

Sabiendo que el total de establecimientos en la población es de 2.000 y se desea establecer límites de confianza del 95%, se pide estimar:

a) El total de establecimientos con más de 5 empleados.
b) El total de empleados con más o igual a 5 años de servicio.

67. Con los datos del ejercicio anterior se pide estimar:

a) El total de salario que pagan los dos mil establecimientos.

b) El salario promedio que debe pagar un establecimiento que tiene en promedio cinco empleados.

68. Un auditor detecta que cierta empresa regularmente exagera las cantidades (millones de $) de los inventarios, a causa de los registros en los retiros. El auditor quiere estimar la cantidad total exagerada sobre 1.000 artículos listados, obteniendo las cantidades exactas (revisadas) del inventario, mediante una muestra aleatoria simple de artículos. Comparando estas cifras exactas con las cantidades registradas obtenemos la exageración (diferencia). Con los datos de la tabla siguiente, utilice el total de cantidad exagerada sobre los 1.000 tipos de artículos y establezca los límites de confianza del 95%.

ARTÍCULO NO. :	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CANT. REVISADA :	1.750	2.950	680	740	1.280	2.410	3.620	720	590	1.120	1.180	2.100	2.400	2.100
CANT. REGISTRADA:	2.100	3.050	910	820	1.400	2.500	3.840	800	820	1.400	1.240	2.300	2.600	2.300

69. Con la información del ejercicio No. 68, mediante la regresión lineal simple, estime la cantidad total registrada para los 1.000 tipos de artículos con base en un promedio de cantidad revisada de $1.500 millones, fijar límites de confianza del 95%. Para ello utilice la información del ejercicio anterior.

70. Con los siguientes datos:

$\Sigma x_i = 1.031,7$ $\quad \Sigma x_i^2 = 106.818,03$ $\quad N = 500$ $\quad \Sigma x_i y_i = 128.762,51$

$\Sigma y_i = 1.250,6$ $\quad \Sigma y_i^2 = 156.749,88$ $\quad n = 10$ $\quad \overline{X} = 105$

a) Estime mediante la razón (método indirecto) el promedio y total para Y.

b) Estime mediante el método directo (promedio) el total para Y.

71. Con los siguientes datos correspondientes a 28 familias sobre sus ingresos semanales, obtenidos a través de una encuesta y un censo realizado con anterioridad:

Información total y parcial de 28 familias.

NÚMERO DE ORDEN	NÚMERO ALEATORIO	INGRESO MENSUAL (miles $)	
		CENSO X_i	MUESTRA y_i
1	170	1.640	2.700
2	045	1.310	1.920
3	186	520	880
4	217	580	860
5	076	1.230	1.840
6	213	1.320	1.860
7	132	840	1.260
8	087	1.390	2.060
9	006	1.200	1.700
10	163	760	960
11	312	1.450	1.890
12	111	730	1.110
13	248	1.420	1.980
14	018	1.160	1.790
15	096	750	1.000
16	224	920	1.520
17	303	2.140	3.460
18	196	720	950
19	230	660	980
20	286	1.440	1.950
21	032	680	960
22	186	1.250	1.640
23	318	1.160	1.760
24	009	2.400	2.820
25	144	1.400	1.880
26	250	1.350	2.060
27	036	1.060	1.450
28	095	730	1.010

ΣX_i^2 = Se pide:

a) Estimar el valor de $\hat{R}$ y fijar los límites del 95%

b) Estimar el valor de $\hat{\bar{Y}}_R$ y fijar los límites del 95%

b) Estimar el valor de $\hat{Y}_R$ y fijar los límites del 95%

MUESTREO ALEATORIO ESTRATIFICADO MAE

El *Muestreo Aleatorio Estratificado*, denominado también *muestreo aleatorio restringido*, es un método que permite una selección más eficiente que el obtenido mediante el *muestreo aleatorio simple*, en especial, cuando la característica que se investiga es de gran variabilidad, lo cual, implica un tamaño muestral relativamente grande, en comparación al del *muestreo aleatorio estratificado.*

En una *muestra aleatoria estratificada,* la población a investigar se divide en grupos relativamente homogéneos con relación a la característica en estudio. Estos grupos se denominan *estratos*, donde la unidad o el elemento investigado presenta una característica tal, que sólo le permite pertenecer a un estrato.

Es necesario entender que un *estrato* es una subpoblación y, como tal, cada uno se constituye en *dominio de estudio*.

Los *estratos* pueden o no estar compuestos del mismo número de unidades, por tal razón la fracción de muestreo ($f = n/N$) puede variar de un estrato a otro.

Al conformar los estratos, lo primero que se debe considerar, siempre que sea posible, es la característica de interés principal, que guarde relación con los objetivos de la investigación, para lo cual deben elaborarse estratos que generen la mayor homogeneidad posible. Las investigaciones de sondeo o las encuestas preliminares ayudan, no sólo a identificar las características de la población, sino que permiten una buena estratificación.

En resumen, se puede decir que la estratificación logra:

- Una *mayor precisión* en los estimadores, a pesar de ser menor el número de unidades seleccionadas, en comparación al muestreo aleatorio simple.
- Garantizar la *representatividad* y una *reducción del error* de la muestra, por agrupación de aquellos elementos que se parecen más por la característica que se investiga, que por pertenecer a la población.
- Las características de la población total quedan representadas; además, los *errores de muestreo* que resultan de una muestra aleatoria simple, serán mayores que los de la muestra aleatoria estratificada del mismo tamaño.
- Allegar información detallada para un grupo o estrato específico, ya que se logra una muestra independiente de los demás estratos. De cada uno se obtienen promedios y varianzas separados y mediante la ponderación, se calcula el valor del estimador del promedio y de la varianza.
- Dividir la población en grupos o estratos más o menos homogéneos en cuanto a su composición interna, pero heterogéneos si se comparan los estratos entre sí.
- Ventajas administrativas.

Un plan de estratificación es óptimo, cuando se minimiza la desviación estándar del estrato. Mientras más estratos se establezcan, tanto más homogéneos serán. Sin embargo, para algunos autores, hay incrementos en los costos en consideración al número de elementos que conlleva la muestra en cada estrato. Por otra parte, la estratificación será efectiva, dependiendo de la estrecha relación que guarden los factores de estratificación con las características que se investigan.

Las estimaciones podrán ser más precisas en la medida del tamaño de la muestra por estrato y, en especial, por la forma en que se realiza la selección de las unidades de la muestra en cada uno de ellos. Para su cálculo, se utilizan varios métodos:

- *Muestras de igual tamaño*, denominada también *afijación igual*, consistiendo en seleccionar el mismo número de unidades para cada estrato muestral.
- *Muestras de asignación o afijación proporcional*, donde los tamaños muestrales en cada estrato, tendrán un número de unidades en forma proporcional a las de los estratos poblacionales.
- *Muestras de asignación o afijación óptima*, cuyo tamaño estará determinado teniendo en cuenta los costos y el grado de variabilidad, en tal forma que el error de estimación sea mínimo para un costo total dado.

Simbología

En el *Muestreo Aleatorio Estratificado* se utilizan símbolos especiales tanto para la población, como para la muestra.

N = Total de unidades que constituyen la población objetivo.

N_h = Total de unidades que contiene cada estrato poblacional.

h = Identifica el estrato, donde ***h*** toma valores de 1 2 3 hasta L estratos.

N_1 N_2 N_3 etc., serán los *tamaños poblacionales* en los estratos 1 2 3 etc.

$$\Sigma N_h = N = N_1 + N_2 + N_3 + \ldots N_L = N$$

$\overline{Y}_h$ = Media aritmética poblacional para cada estrato

$$\overline{Y}_h = \frac{\Sigma Y_{hi}}{N_h} \Rightarrow \quad \overline{Y}_1 = \frac{\Sigma Y_{1i}}{N_1} \qquad \overline{Y}_2 = \frac{\Sigma Y_{2i}}{N_2} \qquad \overline{Y}_3 = \frac{\Sigma Y_{3i}}{N_3}$$

$\overline{Y}_{st}$ = Media aritmética poblacional estratificada ponderada.

$$\overline{Y}_{st} = \frac{\Sigma \overline{Y}_h N_h}{N} \qquad \overline{Y}_{st} = \Sigma \overline{Y}_h W_h$$

W_h = Proporción de elementos en cada estrato

$$W_h = \frac{N_h}{N} \Rightarrow \quad W_1 = \frac{N_1}{N} \qquad W_2 = \frac{N_2}{N} \qquad W_3 = \frac{N_3}{N}$$

$$\Sigma W_h = W_1 + W_2 + W_2 + \ldots W_L = 1$$

S_h^2 = Varianza poblacional en cada estrato $\qquad S_h^2 = \frac{\Sigma Y_{hi}^2 - N_h \overline{Y}_h^2}{N_h - 1}$

n = Número de unidades que contiene la muestra total.
n_h = Número de unidades que contiene la muestra en cada estrato muestral.

$\Sigma n_h = n = n_1 + n_2 + n_3 + \ldots$

$\bar{y}_h$ = Media aritmética muestral para cada estrato.

$$\bar{y}_h = \frac{\Sigma y_{hi}}{n_h} \Rightarrow \bar{y}_1 = \frac{\Sigma y_{1i}}{n_1} \qquad \bar{y}_2 = \frac{\Sigma y_{2i}}{n_2} \qquad \bar{y}_3 = \frac{\Sigma y_{3i}}{n_3}$$

$\bar{y}_{st}$ = Media aritmética muestral estratificada.

$$\bar{y}_{st} = \frac{\Sigma N_h \bar{y}_h}{N} \quad \text{o} \quad \bar{y}_{st} = \Sigma \bar{y}_h W_h$$

S_h^2 = Varianza muestral en cada estrato $\qquad s_h^2 = \dfrac{\Sigma y_{hi}^2 - n_h \bar{y}_{hi}^2}{n_h - 1}$

Dependiendo de la manera como se distribuyen los elementos dentro de los estratos maestrales, se ha dicho que se consideren tres métodos, los que serán tratados en forma independiente.

a) *Asignación o afijación igual;*
b) *Asignación o afijación proporcional*
c) *Asignación o afijación óptima.*

ASIGNACIÓN IGUAL

Se ha dicho que este método implica, que los tamaños muestrales en cada estrato deben ser iguales.
$n_1 = n_2 = n_3 \ldots$ etc.

Antes de determinar el tamaño de la muestra, estratificamos la población de las 355 familias que, hasta ahora, se ha venido analizando.

En primer lugar, se hace necesario identificar la característica principal que se tendrá en cuenta al conformar los grupos o estratos, en este caso, serán los ingresos de las 355 familias. Los niveles de ingresos para los diferentes estratos fueron fijados en forma caprichosa, (en este ejercicio) de tal manera que permitan la explicación de algunos resultados, no esperados.

Estrato I : *Ingresos mensuales menores a $ 1.650.000.*
Estrato II : *Ingresos mensuales entre $ 1.650.000 y $ 2.500.000.*
Estrato III : *Ingresos mensuales superiores a $ 2.500.000.*

En la Tabla 13.2 que se encuentra en el Sistema de Información en Línea SIL, presentamos la población de las 355 familias clasificadas en los tres estratos anteriormente establecidos.

Cálculo de los tamaños muestrales

Veamos la aplicación de las fórmulas. En primer lugar, si consideramos que no existe forma alguna de obtener las varianzas poblacionales para cada estrato, hay necesidad de realizar una encuesta preliminar. Supongamos, arbitrariamente, que el tamaño de la encuesta preliminar es del 4% del tamaño poblacional, tal como lo hicimos en el *Muestreo Aleatorio Simple,* sabiendo que en la mayoría de los casos son decisiones que debe tomar el investigador.

$n_{\text{preliminar}} = 0{,}04(355) = 14{,}2 \cong 14$ familias

Aplicando el método de asignación igual, donde $n_1 = n_2 = n_3$ se tendrá:

$$n_h = \frac{n_{\text{preliminar}}}{L} = \frac{14}{3} = 4{,}67 \cong 5 \qquad \text{L = número de estratos = 3}$$

$n_1 = 5 \qquad n_2 = 5 \qquad n_3 = 5 \qquad n_{preliminar} = 15$

Con los anteriores tamaños establecidos en la encuesta preliminar, seleccionamos 5 unidades en cada uno de los estratos poblacionales, utilizando la tabla de números aleatorios o la calculadora.

Estrato I (Ingresos menores a 1.650 miles $)

No. DE ORDEN	No. ALEATORIO	INGRESO MENSUAL (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE EN grs
				Total	Hombres	Mujeres	Trabajan	
1	136	1.460	si	3	2	1	1	508
2	008	790	si	2	0	2	1	680
3	052	1.560	no	4	3	1	1	720
4	036	790	no	1	0	1	1	580
5	111	1.450	si	2	1	1	1	632

Estrato II (Ingresos entre 1.650 y 2.500 miles $)

No. DE ORDEN	No. ALEATORIO	INGRESO MENSUAL (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE EN grs
				Total	Hombres	Mujeres	Trabajan	
1	028	2.050	si	3	2	1	2	630
2	126	2.050	si	3	2	1	2	630
3	096	1.930	si	2	0	2	2	650
4	056	2.080	no	3	1	2	2	638
5	088	1.960	no	3	1	2	2	584

Estrato III (Ingresos superiores a 2.500 miles $)

No. DE ORDEN	No. ALEATORIO	INGRESO MENSUAL (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE EN grs
				Total	Hombres	Mujeres	Trabajan	
1	23	2.730	si	3	1	2	2	816
2	55	2.960	no	2	1	1	2	555
3	09	3.950	no	5	2	3	3	680
4	18	3.230	si	7	2	5	3	744
5	46	2.936	si	3	1	2	2	686

A continuación, calculamos la varianza y la media del ingreso por familia en cada uno de los estratos.

$\bar{x}_1 = 1.210$ $\bar{x}_2 = 2.014$ $\bar{x}_3 = 3.161,2$ $s_1 = 385,8$ $s_2 = 65,0$ $s_3 = 475,41$

$s_1^2 = 148.850$ $s_2^2 = 4.330$ $s_3^2 = 226.017,2$ $n_1 = 5$ $n_2 = 5$ $n_3 = 5$

La media aritmética estratificada será igual a:

$\bar{x}_{st} = \Sigma W_h \bar{x}_h = 0,46\ (1.210) + 0,37\ (2.014) + 0,17\ (3.161,2) = 1.839,18$ (miles de $)

Supongamos que el investigador establece de acuerdo a su experiencia, que el error (E) debe ser igual al 5% de $\bar{x}_{st}$ y la confianza del 95%, es decir, $z = 1,96$

$$V = \left(\frac{E}{z}\right)^2 = \left(\frac{91,96}{1,96}\right)^2 = 2.201,33 \qquad E = 0,05(1.839,18) = 91,96$$

El *número de unidades* en cada estrato poblacional es:

Estrato I: $N_1 = 162$ Estrato II: $N_2 = 132$ Estrato III: $N_3 = 61$

$\Sigma N_h = N_1 + N_2 + N_3 = 162 + 132 + 61 = 355$ $\Sigma N_h = N = 355$

La *proporción de unidades* en cada uno de los estratos poblacionales es:

$$W_1 = \frac{N_1}{N} = \frac{162}{355} = 0,4563 \cong 0,46 \text{ ó } 46\% \qquad W_2 = \frac{N_2}{N} = \frac{132}{355} = 0,3718 \cong 0,37 \text{ ó } 37\%$$

$$W_3 = \frac{N_3}{N} = \frac{61}{355} = 0,1718 \cong 0,17 \quad \text{ó} \quad 17\%$$

Para determinar el tamaño de la muestra en una población estratificada, cuando utilizamos el *método de asignación o afijación igual,* se deben aplicar las siguientes fórmulas de acuerdo con la característica, sea cualitativa o cuantitativa que se tome como referencia.

a. Estimación del *promedio de ingresos* por familia:

$n_o = \dfrac{\Sigma W_h^2 S_h^2}{V}$ $n_o = \dfrac{1}{V}\Sigma W_h^2 S_h$ Siendo: $V = \dfrac{E^2}{z^2} = \left(\dfrac{E}{z}\right)^2$ $V = \left(\dfrac{91,96}{1,96}\right)^2 = 2.201,33$

y $n = \dfrac{n_o}{1 + \dfrac{n_o}{N}}$ También se puede calcular mediante: $n = \dfrac{NZ^2\Sigma W_h S_h^2}{NE^2 + Z^2\Sigma W_h S_h^2}$

La confianza es considerada del 95%, o sea $z = 1,96$

$n_o = \dfrac{1}{V}\ \Sigma W_h S_h^2$ $n_o = \left[\dfrac{1}{2.201,33}\right]$ [0,46(148.850)+0,37(4.330)+0,17(226.017,2)]=49,29

$$n = \frac{n_o}{1 + \dfrac{n_o}{N}} = \frac{49,29}{1 + \dfrac{49,29}{355}} = 44 \text{ familias}$$ Si: $n = \dfrac{NZ^2\Sigma W_h S_h^2}{NE^2 + Z^2\Sigma W_h S_h^2}$

$$n = \frac{355(1,96)^2[0,46(148.850) + 0,37(4.330) + 0,17(226.017,2)]}{355(91,96)^2 + 1,96^2[0,46(148.850) + 0,37(4.330) + 0,17(226.017,2)]} = 44 \text{ familias}$$

y los tamaños muestrales en cada estrato serán: $n_h = \frac{n}{L}$

$n_h = \frac{44}{3} = 15$ $\quad n_1 = 15$ $\quad n_2 = 15$ $\quad n_3 = 15$ $\quad n = n_1 + n_2 + n_3 = 45$

b. En la estimación de *proporciones*, antes que todo se debe familiarizar con la simbología a utilizar, tanto en la aplicación de las diferentes medidas muestrales como en los parámetros.

N = Tamaño de la población.

$A_h = \Sigma A_{h_i}$ = Número de unidades que presenta la característica, en cada estrato poblacional.

P_h = Proporción de elementos que presenta la característica en cada estrato poblacional.

$P_{h_i} = \frac{\Sigma A_{h_i}}{N}$ $\quad P_1 = \frac{\Sigma A_{1i}}{N_1}$ $\quad P_2 = \frac{\Sigma A_{2i}}{N_2}$ $\quad P_3 = \frac{\Sigma A_{3i}}{N_3}$

$S^2_{P_h}$ = Varianza en los estratos poblacionales: $S^2_{P_h} = P_h Q_h$ $\quad S^2_{P_1} = P_1 Q_1$ $\quad S^2_{P_2} = P_2 Q_2$ $\quad S^3_{P_3} = P_3 Q_3$

n = Tamaño de la muestra.

Σa_{hi} = Número de unidades que presenta la característicaen el estrato muestral.

$p_h = \frac{\Sigma a_{hi}}{n}$ $\quad p_1 = \frac{\Sigma a_{1i}}{n_1} = \frac{a_1}{n_1}$ $\quad p_2 = \frac{\Sigma a_{2i}}{n_2} = \frac{a_2}{n_2}$ $\quad p_3 = \frac{\Sigma a_{3i}}{n_3} = \frac{a_3}{n_3}$

P_{st} = Proporción poblacional estratificada: $p_{st} = \frac{\Sigma N_h P_h}{N}$ ó $P_{st} = \Sigma W_h P_h$

P_{st} = Proporción muestral estratificada $P_{st} = \frac{\Sigma N_h p_h}{N}$ ó $p_{st} = \Sigma W_h p_h$

Dentro del esquema que se ha venido siguiendo, en primer lugar se deben calcular los tamaños muestrales en cada estrato, para realizar estimativos de proporciones y totales, cuando se conocen y desconocen las varianzas poblacionales.

Con la información obtenida sobre familias propietarias de la vivienda, a través de la encuesta preliminar a 15 familias, se tendrá:

$p_1 = \frac{3}{5} = 0,6$ $\quad p_2 = \frac{3}{5} = 0,6$ $\quad p_3 = \frac{3}{5} = 0,6$

(Es coincidencial el hecho que las proporciones sean iguales para los tres estratos).

Además se sabe que:

$W_1 = 0,46$ $\quad W_2 = 0,37$ $\quad W_3 = 0,17$ $\quad q_1 = 0,4$ $\quad q_2 = 0,4$ $\quad q_3 = 0,4$
$z = 1,96$ $\quad E = 10\%$ $\quad N = 355$ $\quad N_1 = 162$ $\quad N_2 = 132$ $\quad N_3 = 61$

Estos datos permiten trabajar con las fórmulas anteriormente vistas. En primer lugar veamos su aplicación, para obtener el valor de n en los *estimativos de una proporción*, con un error del 10% y una confianza del 95%.

$$n_o = \frac{\Sigma W_h P_h Q_h}{(E/Z)^2} \qquad \text{Siendo } n = \frac{n_o}{1 + \frac{n_o}{N}}$$

$$n_o = \frac{1}{\left(\frac{0,10}{1,96}\right)^2}[0,46(0,6)(0,4) + 0,37(0,6)(0,4) + 0,17(0,6)(0,4)] = 92,20 \qquad n = \frac{92,20}{1+\frac{92,20}{355}} \cong 74 \text{ familias}$$

Ahora si consideramos a $p_1 = p_2 = p_3 = 0,50$, el tamaño de la muestra será el máximo valor posible; además, con un error del 10% y confianza del 95%, se tendrá:

$$n_o = \frac{1}{0,0026}\ [0,46(0,5)(0,5) + 0,37(0,5)(0,5) + 0,17(0,5)(0,5)] = 96,15 \qquad n = \frac{96,15}{1+\frac{96,15}{355}} \cong 76 \text{ familias}$$

También podemos calcular n mediante la fórmula: $$n = \frac{NZ^2 W_h P_h Q_h}{NE^2 + Z^2 \Sigma W_h P_h Q_h}$$

Veámoslo:

$$n = \frac{355(1,96)^2[0,46(0,6)(0,4) + 0,37(0,6)(0,4) + 0,17(0,6)(0,4)]}{355(0,10)^2 + 1,96^2[0,46(0,6)(0,4) + 0,37(0,6)(0,4) + 0,17(0,6)(0,4)]} = 76 \text{ familias}$$

El resultado debe ser exactamente igual aplicando cualquiera de las fórmulas anteriores.

Cálculo de estimativos y fijación de límites

En primer lugar consideraremos que el tamaño de la muestra (n) es igual a 36, en lugar de los resultados anteriores obtenidos, a fin de agilizar y simplificar operaciones. Teniendo en cuenta la característica (variable) ingresos, una confianza del 95%, procederemos a calcular el estimador del promedio y fijar límites de confianza.

Siendo $n = 36$, por el método de afijación igual, se tendrán los siguientes tamaños maestrales para cada uno de los estratos:

$$n_h = \frac{n}{L} = \frac{36}{3} = 12 \quad \Rightarrow \quad n_1 = 12 \qquad n_2 = 12 \qquad n_3 = 12 \qquad n = 36$$

Determinados los tamaños muestrales, se hará la selección de las unidades en cada uno de los estratos, con el fin de recoger información correspondiente cada una de las familias seleccionadas.

MUESTRA ASIGNACIÓN IGUAL

Estrato I (Ingresos inferiores a 1.650 miles $)

NÚMERO DE ORDEN	NÚMERO ALEATORIO	INGRESO MENSUAL (miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE EN Grs.
				Total	Hombres	Mujeres	Trabajan	
1	151	1.230	no	3	2	1	1	605
2	008	790	si	2	0	2	1	680
3	033	790	si	3	2	1	1	600
4	122	1.260	no	4	2	2	1	520
5	076	1.260	no	3	1	2	1	516
6	082	860	no	2	1	1	1	510
7	131	790	no	3	2	1	1	509
8	096	1.130	si	6	2	4	1	700
9	048	890	si	2	1	1	1	720
10	055	1.020	no	1	1	0	1	530
11	018	860	no	1	0	1	1	510
12	076	1.260	no	3	1	2	1	516

Estrato II (Ingresos entre 1.650 y 2.500 miles $)

NÚMERO DE ORDEN	NÚMERO ALEATORIO	INGRESO MENSUAL (miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE EN Grs.
				Total	Hombres	Mujeres	Trabajan	
1	020	1.880	no	3	1	2	2	880
2	048	1.940	si	4	3	1	1	536
3	111	2.050	si	4	2	2	2	734
4	124	2.360	si	4	2	2	2	716
5	076	2.140	si	4	2	2	2	716
6	001	1.860	si	5	3	2	2	789
7	032	1.860	si	5	2	3	2	757
8	086	1.720	no	5	3	2	2	584
9	116	1.730	no	3	1	2	2	612
10	094	1.990	no	4	2	2	2	532
11	035	1.930	si	4	2	2	2	520
12	022	1.870	no	2	2	0	2	520

Estrato III (Ingresos superiores a 2.550 miles $)

NÚMERO DE ORDEN	NÚMERO ALEATORIO	INGRESO MENSUAL (miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DIARIO DE CARNE EN Grs.
				Total	Hombres	Mujeres	Trabajan	
1	25	2.680	si	2	1	1	2	760
2	51	2.880	si	2	1	1	2	724
3	58	2.860	si	3	1	2	3	510
4	36	3.890	no	3	1	2	3	802
5	08	2.620	si	5	2	3	3	620
6	05	3.130	no	5	2	3	3	868
7	11	2.890	no	4	2	2	2	770
8	29	2.870	si	3	1	2	2	555
9	37	2.970	no	3	1	2	2	794
10	44	3.860	si	3	1	2	3	830
11	19	4.960	si	6	2	4	4	700
12	50	2.980	no	3	2	1	2	600

Ahora recordemos que la *selección*, se realiza mediante la aplicación de EXCEL, de las Tablas de números aleatorios o la Calculadora, para la cual se toman 12 unidades en cada uno de los estratos, cuyos números estén comprendidos entre 001 y 162; lo mismo se hará para el estrato II entre 001 y 132 y para el estrato III, entre 01 y 61. Se ha acordado hacer la selección sin repetición de las unidades.

Con las anteriores Tablas, se pueden obtener diferentes estimadores tales como: promedios, proporciones, razones y proporciones en conglomerados, fijando límites de confianza para cada estimador puntual.

Estimación de promedios y totales

a. Consideremos el promedio y el total de ingresos de las familias, calculando en primer lugar los promedios para cada uno de los estratos muestrales.

$$\bar{x}_1 = \frac{\Sigma x_{1i}}{n_1} = \frac{12.140}{12} = 1.011{,}67 \qquad \bar{x}_2 = \frac{\Sigma x_{2i}}{n_2} = \frac{23.330}{12} = 1.944{,}17$$

$$\bar{x}_3 = \frac{\Sigma x_{3i}}{n_3} = \frac{38.590}{12} = 3.215{,}83$$

Las *desviaciones típicas muestrales* para cada estrato serán:

$$s_1 = \sqrt{\frac{\Sigma x_{1i}^2 - n_1 \bar{x}_1^2}{n_1 - 1}} = \sqrt{\frac{12.736.600 - 12\,(1.011{,}67)^2}{12 - 1}} = 203{,}35 \qquad s_1^2 = 41.353{,}25$$

(Nota: Se trabajó en la calculadora con el programa de estadística).

$$s_2 = \sqrt{\frac{\Sigma x_i^2 - n_2 \bar{x}_2^2}{n_2 - 1}} = \sqrt{\frac{45.702.100 - 12\,(1.944,17)^2}{12 - 1}} = 176,98 \qquad s_2^2 = 31.321,47$$

$$s_3 = \sqrt{\frac{\Sigma x_{3i}^2 - n_3 \bar{x}_3^2}{n_3 - 1}} = \sqrt{\frac{129.241.300 - 12\,(3.215,83)^2}{12 - 1}} = 683,74 \qquad s_3^2 = 467.504,45$$

La *Media muestral estratificada* será igual a:

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h = 0,46(1.011,67) + 0,37(1.944,17) + 0,17(3.215,83) = 1.731,4 \text{ ó } \$\,1.731.402,2$$

La fórmula para fijar los límites de confianza para el promedio, es:

$$\hat{\bar{X}}_{st} = \bar{x}_{st} \pm t\sqrt{\left[\frac{1}{N^2}\ \Sigma N_h\,(N_h - n_h)\,\frac{s_h^2}{n_h}\right]}$$

$$\hat{\bar{X}}_{st} = 1.731,4 \pm 2,035\ \sqrt{\frac{1}{355^2}\left[162(162\text{-}12)\,\frac{41.353,25}{12} + 132(132\text{-}12)\,\frac{31.321,47}{12} + 61(61\text{-}12)\,\frac{467.504,45}{12}\right]}$$

$$\hat{\bar{X}}_{st_{\substack{s\\i}}} = 1.731,4 \pm 89,09 \begin{cases} \nearrow\ 1.820,49 = \$\,1.820.490 \\ \searrow\ 1.642,31 = \$\,1.642.310 \end{cases}$$

Al fijar los *límites de confianza* del 95% para el promedio, el valor de t, será:

$\upsilon = n_1 + n_2 + n_3 - 3 = 12 + 12 + 12 - 3 = 33$

Para $\upsilon = 33$ y $\alpha = 0,05$, se tendrá que $t = 2,035$

En los *estimativos de totales* para establecer los límites de confianza del 95%, se procederá de la siguiente manera:

$$\hat{X}_{st_{\substack{s\\j}}} = N\bar{x}_{st} \pm t\sqrt{\Sigma N_h\,(N_h - n_h)\frac{s_h^2}{n_h}}$$

El *estimador de la varianza* con el cual se obtiene el error de estimación, se puede calcular de la siguiente manera:

$$V_{[\bar{x}_{st}]} = \Sigma N_h\,(N_h - n_h)\,\frac{s_h^2}{n_h}$$

$$\hat{}_{[\bar{x}_s]} = 162(162-12)\,\frac{41.353,25}{12} + 132(132-12)\,\frac{31.321,47}{12} + 61(61-12)\,\frac{467.504,45}{12} = 241.532.238,5$$

$$\hat{}_{s\,si} = 355(1.731,4) \pm 2,035\sqrt{241.532.238,5} \begin{cases} \nearrow\ 646.273,56 \cong \$\,646.273.560,7 \\ \searrow\ 583.020,44 \cong \$583.020.439,3 \end{cases}$$

Con las anteriores tablas se pueden obtener los diferentes estimadores de: *promedios, proporciones, razones y proporciones en conglomerados*, fijando límites de confianza para cada estimador puntual.

El *estimador de la Varianza* con el fin de obtener el error de estimación, se puede calcular mediante la fórmula:

$$\hat{V}_{[\bar{x}_{st}]} = \left[\frac{1-f}{n}\right][\Sigma W_h s_h^2] \quad \Rightarrow \text{ siendo: } \hat{\bar{X}}_{st} = \bar{x}_{st} \pm t\sqrt{\left[\frac{1-f}{n}\right][\Sigma W_h s_h^2]}$$

Veamos cómo se calcula, tomando los anteriores resultados muestrales:

$$\hat{V}_{[\bar{x}_{st}]} = \left[\frac{1-\frac{36}{355}}{36}\right][0{,}46(41.353{,}25) + 0{,}37(31.321{,}47) + 0{,}17(467.504{,}45)] = 2.747{,}87$$

$$\hat{\bar{X}}_{st} = 1.731{,}4 \pm 2{,}035\sqrt{2.747{,}87} \begin{cases} \nearrow 1.838{,}08 \\ \searrow 1.624{,}72 \end{cases} \text{ (miles de \$)}$$

La diferencia observada es mínima con respecto al estimador de X, resulta al ser calculadas con todos los decimales, ya que los resultados deberán ser exactamente iguales. Algo similar sucede con la siguiente fórmula, al calcular el *estimador de la Varianza*.

$$\hat{V}_{[\bar{x}_{st}]} = \sum\left[\frac{W_h^2 s_h^2}{n_h}(1-f_h)\right] \quad \Rightarrow \text{ siendo: } \hat{\bar{X}}_{st} = \bar{x}_{st} \pm t\sqrt{\sum\left[\frac{W_h^2 s_h^2}{n_h}(1-f_h)\right]}$$

$$\hat{V}_{[\bar{x}_{st}]} = \left[\frac{0{,}46^2(41.353{,}25)}{12}\right]\left[1-\frac{12}{162}\right] + \left[\frac{0{,}37^2(31.321{,}47)}{12}\right]\left[1-\frac{12}{132}\right] + \left[\frac{0{,}17^2(467.504{,}45)}{12}\right]\left[1-\frac{12}{61}\right] = 1.904{,}44$$

b. Para el cálculo del *promedio* y *total de personas por familia*, el procedimiento es igual al utilizado para los ingresos.

$\bar{x}_1 = 2{,}75$ $\bar{x}_2 = 3{,}92$ $\bar{x}_3 = 3{,}5$ $s_1^2 = 1{,}84$ $s_2^2 = 0{,}81$ $s_3^2 = 1{,}55$

$\bar{x}_{st} = 0{,}46(2{,}75) + 0{,}37(3{,}92) + 0{,}17(3{,}5) = 3{,}31$ personas por familia.

$$\hat{V}_{[\bar{x}_{st}]} = \frac{1}{355^2}\left[162\,(162\text{-}12)\,\frac{1{,}84}{12} + 132(132\text{-}12)\,\frac{0{,}81}{12} + 61(61\text{-}12)\,\frac{1{,}55}{12}\right] = 0{,}041$$

$$\hat{\bar{X}}_{st_i} = 3{,}31 \pm 2{,}035\sqrt{0{,}041} \begin{cases} \nearrow 3{,}72 \cong 4 \\ \searrow 2{,}89 \cong 3 \end{cases} \text{ (Promedio de personas)}$$

$$\hat{X}_{st} = 355(3{,}31) \pm 355(0{,}41) \begin{cases} \nearrow 1.320{,}06 = 1.321 \\ \searrow 1.029{,}5 = 1.030 \end{cases} \text{ (Total de personas)}$$

c. Si desea estimar y fijar límites de confianza del 95%, para el *consumo promedio* y *total de carne* diario de las familias, se tendrá:

$\bar{x}_1 = 576{,}33$ $\bar{x}_2 = 658{,}0$ $\bar{x}_3 = 711{,}08$ $s_1^2 = 6.726{,}97$ $s_2^2 = 15.017{,}27$ $s_3^2 = 13.226{,}45$

$\bar{x}_{st} = 0{,}46(576{,}33) + 0{,}37(658) + 0{,}17(711{,}08) = 629{,}46$ gramos de carne diaria.

$$V_{[\bar{x}_{st}]} = \frac{1}{355^2}\left[162\,(162\text{-}12)\frac{6.726{,}97}{12} + 132(132\text{-}12)\,\frac{15.017{,}27}{12} + 61(61\text{-}12)\frac{13.226{,}45}{12}\right] = 291{,}52$$

$$\hat{\bar{X}}_{\substack{st_s \\ i}} = 629{,}46 \pm 2{,}035\sqrt{291{,}52} = \begin{cases} 664{,}21 \\ 594{,}71 \end{cases}$$ (Promedio diario de consumo de carne)

$$\hat{X}_{\substack{st_s \\ i}} = 355(629{,}46) \pm 355(34{,}74) = \begin{cases} 235.791{,}0 \\ 211.125{,}6 \end{cases}$$ (Total de consumo para las 355 familias)

Estimación de una proporción y total

A través de los resultados obtenidos en la encuesta correspondiente a la muestra de 36 unidades, se calcularán las proporciones muestrales en cada estrato, para la proporción y total de familias propietarias de vivienda.

$$p_1 = \frac{\Sigma a_{1i}}{n_1} = \frac{4}{12} \cong 0{,}33 \qquad q_1 = 0{,}67 \qquad p_2 = \frac{\Sigma a_{2i}}{n_2} = \frac{7}{12} \cong 0{,}58 \qquad q_2 = 0{,}42$$

$$p_3 = \frac{\Sigma a_{3i}}{n_3} = \frac{7}{12} \cong 0{,}58 \qquad q_3 = 0{,}42 \qquad \text{Siendo:} \qquad p_{st} = \frac{\Sigma N_h p_h}{N} = \Sigma W_h\, p_h$$

p_{st} = 0,46(0,33)+0,37(0,58)+0,17(0,58) = 0,465 o 46,5% son propietarios.

Para el *estimador de la varianza en la proporción*, su cálculo es similar al utilizado para la variable; veamos el primer procedimiento:

$$\hat{V}_{[p_{st}]} = \frac{1}{N^2}\left[\Sigma N_h(N_h - n_h)\frac{p_h q_h}{n_h}\right] \qquad \text{Siendo:} \quad \hat{P}_{st} = p_{st} \pm t\sqrt{\frac{1}{N^2}\left[\Sigma N_h(N_h - n_h)\frac{p_h q_h}{n_h}\right]}$$

$$\hat{V}_{[p_{st}]} = \frac{1}{355^2}\left[162(162\text{-}12)\frac{0{,}33(0{,}67)}{12} + 132(132\text{-}12)\frac{0{,}58(0{,}42)}{12} + 61(61\text{-}12)\frac{0{,}58(0{,}42)}{12}\right] = \frac{829{,}9562}{126.025} = 0{,}0066$$

También podemos obtener este mismo resultado, mediante un *segundo procedimiento*, como se puede observar a continuación:

$$\hat{V}_{[p_{st}]} = \left[\frac{1-f}{n}\right][\Sigma W_h p_h q_h] \qquad \text{Siendo:} \quad \hat{P}_{st} = p_{st} \pm t\left[\frac{1-f}{n}\right][\Sigma W_h p_h q_h]$$

$$\hat{V}_{[p_{st}]} = \left[\frac{1 - \frac{36}{355}}{36}\right][0{,}46(0{,}33)(0{,}67) + 0{,}37(0{,}58)(0{,}42) + 0{,}17(0{,}58)(0{,}42)] = 0{,}0058$$

y por el *tercer procedimiento*, para el cálculo del estimador de la varianza se tendrá:

$$\hat{V}_{[p_{st}]} = \Sigma\left(\frac{W_h^2 p_h q_h}{n_h}\right)(1 - f_h) \qquad \text{Siendo:} \quad \hat{P}_{st} = p_{st} \pm t\sqrt{\Sigma\left[\frac{W_h^2 p_h q_h}{n_h}\right][1 - f_h]}$$

$$\hat{V}_{[p_{st}]} = \left[\frac{0{,}46^2(0{,}33)(0{,}67)}{12}\right]\left[1 - \frac{12}{162}\right] + \left[\frac{0{,}37^2(0{,}58)(0{,}42)}{12}\right]\left[1 - \frac{12}{132}\right] + \left[\frac{0{,}17^2(0{,}58)(0{,}42)}{12}\right]\left[1 - \frac{12}{61}\right] = 0{,}0066$$

Los *límites de confianza* del 95% para la *proporción de familias con vivienda* serán:

$$\hat{P}_{\substack{st_s \\ i}} = p_{st} \pm t\sqrt{\hat{V}_{[p_{st}]}} \qquad \hat{P}_{\substack{st_s \\ i}} = 0{,}465 \pm 2{,}035\sqrt{0{,}0066} \begin{cases} \nearrow 63{,}03\% \\ \searrow 29{,}97\% \end{cases}$$

Proporción de familias con vivienda

y los *límites de confianza* del 95% para el *total de familias con vivienda* serán:

$$\hat{A}_{\substack{st_s \\ i}} = Np_{st} \pm tN\sqrt{\hat{V}_{[p_{st}]}} \qquad \hat{A}_{\substack{st_s \\ i}} = 355(0{,}465) \pm 355(0{,}165) \begin{cases} \nearrow 224 \\ \searrow 107 \end{cases}$$

Total de familias con vivienda

Estimación para la razón, promedio y el total

El *estimador de la razón en el muestreo aleatorio estratificado* se puede efectuar de dos maneras: en el primer caso, se le denomina estimación mediante la **razón separada**, ya que ésta se obtiene en forma separada para cada estrato.

$$\hat{R}_h = \frac{\bar{y}_h}{\bar{x}_h} \quad \text{(Razón para cada estrato)} \qquad \hat{R}_1 = \frac{\bar{y}_1}{\bar{x}_1} \qquad \hat{R}_2 = \frac{\bar{y}_2}{\bar{x}_2} \qquad \hat{R}_3 = \frac{\bar{y}_3}{\bar{x}_3}$$

a. La *razón separada* para una población estratificada será obtenida mediante la fórmula:

$$\hat{R}_{s_{st}} = \Sigma W_h\left(\frac{\bar{y}_h}{\bar{x}_h}\right) = \Sigma \hat{R}_h W_h$$

b. En el segundo caso, se le denomina ***razón combinada***, siendo:

$$\hat{R}_{c_{st}} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \qquad \bar{x}_{st} = \Sigma W_h \bar{x}_h \qquad \bar{y}_{st} = \Sigma W_h \bar{y}_h$$

El *estimador de la varianza*, para cada uno de los casos, para calcular luego el error de estimación

$$\hat{V}_{[R_{s_{st}}]} = \Sigma\left[\frac{1-f_h}{n_h \bar{x}_h^2}\right][s_{y_h}^2 - 2\hat{R}_h \rho_h s_{y_h} s_{x_h} + \hat{R}_h^2 s_{x_h}^2] \quad \text{(Para la razón separada)}$$

$$\hat{V}_{[R_{c_{st}}]} = \Sigma\left[\frac{1-f_h}{n_h \bar{x}_h^2}\right][s_{y_h}^2 - 2\hat{R}_c \rho_h s_{x_h} s_{y_h} + \hat{R}_c s_{x_h}^2] \quad \text{(Para la razón combinada)}$$

Para la aplicación de estas fórmulas, consideraremos los datos obtenidos a través de la muestra de 36 familias y como características, para la estimación de la razón, la relación que hay, entre hombres (x) y mujeres (y).

Las *medias aritméticas,* para hombres y mujeres en estos estratos serán:

$$\bar{x}_1 = \frac{\Sigma x_{1i}}{n_1} = \frac{15}{12} = 1{,}25 \qquad \bar{x}_2 = \frac{\Sigma x_{2i}}{n_2} = \frac{25}{12} = 2{,}08 \qquad \bar{x}_3 = \frac{\Sigma x_{3i}}{n_3} = \frac{17}{12} = 1{,}42$$

$$\bar{y}_1 = \frac{18}{12} = 1{,}5 \qquad \bar{y}_2 = \frac{22}{12} = 1{,}83 \qquad \bar{y}_3 = \frac{25}{12} = 2{,}08$$

$\hat{R}_h$ = *Razón estimada en cada estrato* = $\dfrac{\text{Mujeres}}{\text{Hombres}}$

$$\hat{R}_1=\frac{\overline{y}_1}{\overline{x}_1}=\frac{1,5}{1,25}=1,2 \qquad \hat{R}_2=\frac{\overline{y}_2}{\overline{x}_2}=\frac{1,83}{2,08}\cong 0,88 \qquad \hat{R}_3=\frac{\overline{y}_3}{\overline{x}_3}=\frac{2,08}{1,42}=1,46$$

El cálculo de la ***Razón separada*** en este ejercicio, será:

$$\hat{R}_{s_{st}}=\Sigma W_h\hat{R}_h \qquad \hat{R}_{s_{st}}=0,46(1,2)+0,37(0,88)+0,17(1,46)=1,13$$

La relación obtenida nos permite indicar que por cada 7 hombres, hay 8 mujeres

$\hat{R}_{c_h}$ = *Razón combinada*

$$\hat{R}_{c_{st}}=\frac{\overline{y}_{st}}{\overline{x}_{st}}=\frac{1,7207}{1,586}=1,085$$

$$\overline{y}_{st}=0,46(1,5)+0,37(1,83)+0,17(2,08)=1,7207$$

$$\overline{x}_{st}=0,46(1,25)+0,37(2,08)+0,17(1,42)=1,586$$

En el cálculo del *estimador de la varianza* de los dos casos anteriores, se requieren algunas operaciones preliminares:

Las *desviaciones típicas* y las *varianzas* de la variable ***x***:

$$s_{x_1}=0,75 \qquad s_{x_2}=0,67 \qquad s_{x_3}=0,51 \qquad s^2_{x_1}=0,57 \qquad s^2_{x_2}=0,45 \qquad s^2_{x_3}=0,27$$

(Nota: Se trabajó en la calculadora con el programa de estadística)

Las *desviaciones típicas* y las *varianzas* de la variable *y*:

$$s_{y_1}=1 \qquad s_{y_2}=0,72 \qquad s_{y_3}=0,90 \qquad s^2_{y_1}=1 \qquad s^2_{y_2}=0,52 \qquad s^2_{y_3}=0,81$$

El valor de ρ (rho) se obtiene de la siguiente manera:

$$\rho_h=\frac{d_{xh}}{2d_{y_h}} \qquad \rho_1=\frac{d_{x_1}}{2d_{y_1}} \qquad \rho_2=\frac{d_{x_2}}{2d_{y_2}} \qquad \rho_3=\frac{d_{x_3}}{2d_{y_3}}$$

Los valores de los c*oeficientes de variación* de d_{xh} y d_{yh} son calculados así:

$$d_{x_1}=\frac{s_{x1}}{\overline{x}_1}=\frac{0,75}{1,25}=0,6 \qquad d_{x_2}=\frac{s_{x2}}{\overline{x}_2}=\frac{0,67}{2,08}=0,32 \qquad d_{x_3}=\frac{s_{x3}}{\overline{x}_3}=\frac{0,51}{1,42}=0,36$$

$$d_{y_1}=\frac{s_{y1}}{\overline{y}_1}=\frac{1}{1,5}=0,67 \qquad d_{y_2}=\frac{s_{y2}}{\overline{y}_2}=\frac{0,72}{1,83}=0,39 \qquad d_{y_3}=\frac{s_{y3}}{\overline{y}_3}=\frac{0,9}{2,08}=0,43$$

$$\rho_1=\frac{d_{x1}}{2d_{y1}}=\frac{0,6}{2(0,67)}=0,44 \qquad \rho_2=\frac{d_{x2}}{2d_{y2}}=\frac{0,32}{2(0,39)}=0,41 \qquad \rho_3=\frac{d_{x3}}{2d_{y3}}=\frac{0,36}{2(0,43)}=0,42$$

Las *fracciones de muestreo* para cada estrato serán:

$$f_1=\frac{n_1}{N_1}=\frac{12}{162}=0,074 \qquad f_2=\frac{n_2}{N_2}=\frac{12}{132}=0,09 \qquad f_3=\frac{n_3}{N_3}=\frac{12}{61}=0,20$$

Con la información anterior se podrá hacer el cálculo de los *estimadores de la Varianza*.

a. El *estimador de la varianza* o *error de estimación* al cuadrado para el método de la *Razón Separada*

$$\hat{V}_{\hat{R}s} = \frac{1}{12(1,25)^2}\left(\frac{0,074}{1}\right) 1 \quad 2(1,2)(0,44)(0,75)(1) + 1,2^2(0,57) +$$

$$\frac{1}{12(2,08)^2}\left(\frac{0,09}{1}\right) 0,52 \quad 2(0,9)(0,41)(0,67)(0,72) + 0,9^2(0,45) +$$

$$\frac{1}{12(1,42)^2}\left(\frac{0,20}{1}\right) 0,81 \quad 2(1,46)(0,42)(0,51)(0,9) + 1,46^2(0,27) =$$

$$\hat{V}_{Rs} = 0,04938(1,0108) + 0,0175(0,5285) + (0,0331)0,82261 = 0,0864$$

$$\hat{R}_{s_{st_{s_i}}} = \hat{R}_s \pm t\sqrt{\hat{V}_{\hat{R}s}} \qquad \hat{R}_{s_{st_{s_i}}} = 1,13 \quad 2,035\sqrt{0,0864} = 1,13 \quad 060 \begin{cases} 1,73 \\ 0,53 \end{cases}$$

b. El *estimador de la varianza* o *error de estimación al cuadrado* en el método de la Razón Combinada:

$$\hat{V}_{\hat{R}c} = \frac{1}{12(1,25)^2}\left(\frac{0,074}{1}\right) 1 \quad 2(1,085)(0,45)(0,75)(1) + 1,085^2(0,57) +$$

$$\frac{1}{12(1,25)^2}\left(\frac{0,09}{1}\right) 0,52 \quad 2(1,085)(0,41)(0,67)(0,72) + 1,085^2(0,45) +$$

$$\frac{1}{12(1,42)^2}\left(\frac{0,20}{1}\right) 0,81 \quad 2(1,085)(0,42)(0,51)(0,9) + 1,085^2(0,27) =$$

$$\hat{V}_{\hat{R}c} = 0,04938(0,9386) + 0,0175(0,6206) + 0,0331(0,7095) = 0,0807$$

$$\hat{R}_{c_{st_{s_i}}} = \hat{R}_c \pm t\sqrt{\hat{V}_{Rc}} \qquad \hat{R}_{c_{st_{s_i}}} = 1,085 \quad 2,035\sqrt{0,0807} \begin{cases} 1,66 \\ 0,51 \end{cases}$$

A partir del *estimador de la razón* en el Muestreo aleatorio estratificado así como en el Muestreo aleatorio simple, se pueden hacer estimativos de *promedios* y de *totales*, tanto con la Razón Separada como la Combinada:

a. $\hat{\bar{Y}}_{R_s} = \Sigma W_h \hat{R}_h \bar{X}_h$ $\quad \hat{Y}_{R_s} = N\Sigma W_h \hat{R}_h \bar{X}_h$ $\qquad$ b. $\hat{\bar{Y}}_{R_c} = \hat{R}_c \bar{X}_{st}$ $\quad \hat{Y}_{RC} = N R_c \bar{X}_{st}$

Recordemos que: $= N\bar{}$ $\qquad$ $_h = N_h\bar{}_h$ (Total)

Ahora supongamos que se tiene información sobre el valor de la media aritmética poblacional por cada estrato y para el total,

$$\hat{\bar{Y}}_{R_s} = 0,46(1,5)(1,2) + 0,37(0,9)(1,8) + 0,17(1,46)(2) = 1,924 = 1,9 \text{ Promedio de mujeres por amilia}$$

Los *límites de confianza* del 95%, para el estimador del Promedio de mujeres por familia:

$$\hat{\bar{Y}}_{R_{s_s}} = \Sigma W_h \hat{R}_h \bar{X}_h \pm t \sqrt{\Sigma \frac{1}{n_h} f_h \quad s^2_{yh} \quad 2\hat{R}_h \rho_h s_{yh} s_{xh} + \hat{R}^2_h s^2_{xh}}$$

$$\hat{\bar{V}}_{\bar{y}_{R_S}} = 1{,}9 \pm 2{,}035 \sqrt{\begin{array}{l} \frac{1}{12}\ 0{,}074 \quad 1 \quad 2(1{,}20)(0{,}44)(0{,}75)(1) + 1{,}2^2(0{,}57) + \frac{1}{12}\ 0{,}09 + \\ 0{,}52 \quad 2(0{,}9)(0{,}41)(0{,}67)(0{,}72) + 0{,}9^2(0{,}45) + \\ \frac{1}{12}\ 0{,}20 \quad 0{,}81 \quad 2(1{,}46)(0{,}42)(0{,}51)(0{,}9) + 1{,}46^2(0{,}27) \end{array}}$$

$$\hat{\bar{Y}}_{Rss_j} = 1{,}9 \pm 2{,}035 \sqrt{0{,}1685} \begin{array}{l} \nearrow 2{,}74 \\ \searrow 1{,}06 \end{array} \text{ Promedio de mujeres por amilia.}$$

Como conocemos el promedio estimado en la Razón Separada, se podrá obtener el estimador por intervalos para el total de mujeres en las 355 familias.

$$\hat{Y}_{R_{s_s}} = (\Sigma W_h \hat{R}_h \bar{X}_h) N \pm tN \sqrt{\frac{1}{n_h} f_h \quad s^2_{yh} \quad 2\hat{R}_h \rho_h s_{yh} s_{xh} + \hat{R}^2_h s^2_{xh}}$$

$$\hat{Y}_{Rss_i} = 1{,}9(355) \pm 2{,}035(355) \sqrt{0{,}1685} \begin{array}{l} \nearrow 971{,}04 = 971 \\ \searrow 371{,}95 = 372 \end{array} \text{ Total mujeres en las 355 amilias.}$$

En la *razón combinada*, el promedio y total se calculan de la siguiente manera:

$$\hat{\bar{Y}}_{R_c} = \hat{R}_c \bar{X}_{st} = 1{,}085(1{,}7) = 1{,}85 \text{ Promedio de mujeres por amilia.}$$

Para el total se tendrá que:

$$\hat{Y}_{R_c} = \hat{R}_c \bar{X}_{st} N = 1{,}085(1{,}7)(355) = 654{,}80 = 655 \text{ Total de mujeres en las 355 amilias.}$$

Los *límites de confianza* del 95%, tanto para el Promedio como para el Total, se deberá en primer lugar calcular la *Varianza estimada*:

$$\hat{\bar{Y}}_{R_{c_{st}}} = \left(\frac{\bar{y}_{st}}{\bar{x}_{st}}\right) \pm t \sqrt{\Sigma \frac{1}{n} f_h \quad s^2_{yh} \quad \hat{2}R_c \rho_h s_{x_h} s_{yh} + R^2_c s^2_{x_h}}$$

Límites de confianza del 95% para el promedio de mujeres por familia.

$$\hat{\bar{Y}}_{\bar{R}_{c_{st}}} = 1{,}85 \pm 2{,}035 \sqrt{\begin{array}{l} \frac{1}{12}\ 0{,}074 \quad 1 \quad 2(1{,}085)(0{,}44)(0{,}75)(1) + 1{,}085^2(0{,}57) + \frac{1}{12}\ 0{,}09 \\ 0{,}52 \quad 2(1{,}085)(0{,}41)(0{,}67)(0{,}72) + 1{,}085^2(0{,}45) + \\ \frac{1}{12}\ 0{,}20 \quad 0{,}36 \quad 2(1{,}085)(0{,}42)(0{,}51)(0{,}9) + 1{,}085^2(0{,}27) \end{array}}$$

$$\hat{\bar{Y}}_{Rc_s} = 1{,}85 \pm 2{,}035\sqrt{0{,}0807} \begin{cases} \nearrow 2{,}42 \\ \searrow 1{,}27 \end{cases}$$

Los *límites de confianza* del 95 para el Total de mujeres en las 355 amilias

$$\hat{\bar{Y}}_{Rc_s} = 1{,}85(355) \pm 2{,}035(355)\sqrt{0{,}0807} = 656{,}75 \quad 205{,}22 \begin{cases} \nearrow 881{,}97 \cong 882 \\ \searrow 451{,}53 \cong 452 \end{cases}$$ Total mujeres

ESTIMACIÓN DE PROMEDIOS Y TOTALES MEDIANTE LA REGRESIÓN LINEAL SIMPLE

Dentro de los diferentes métodos de estimación que se han venido estudiando, también se utiliza la *regresión lineal*, al igual que la *razón* y se les considera como *métodos indirectos* para estimar *promedios y totales.*

En el caso de la *regresión lineal*, se usan dos formas para calcular el estimador del promedio y del total, mediante la *regresión lineal separada y regresión lineal combinada*.

a. En la *regresión lineal separada*, la ecuación general para el estimador del promedio será:

$$\bar{Y}_{RL_h} = \bar{y}_h + b_h(\bar{X}_h - \bar{x}_h) \qquad \bar{y}_h = \frac{\Sigma y_{hi}}{n_h} \qquad \bar{x}_h = \frac{\Sigma x_{hi}}{n_h}$$

RL = Regresión lineal

b_h = es el coeficiente angular o pendiente para cada estrato muestral

$$b_h = \frac{\Sigma(y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)}{\Sigma(x_{hi} - \bar{x}_h)^2} \qquad \text{Tambi n:} \quad b_h = \frac{n\Sigma x_{hi}y_{hi} - (\Sigma x_{hi})(\Sigma y_{hi})}{n\Sigma x_{hi}^2 - (\Sigma x_{hi})^2}$$

Con la información obtenida a través de una Muestra estratificada (*asignación igual*) de 36 familias, estimar el consumo promedio de carne diaria (en gramos) en función del nivel de ingresos, así:

Estrato I

$\Sigma x_1 = 12.140$ $\Sigma x_1^2 = 12.736.600$ $\Sigma y_1 = 6.916$ $\Sigma y_1^2 = 4.059.918$ $b_1 = 0{,}07496$

$s_{x_1}^2 = 41.360{,}61$ $\Sigma x_1 y_1 = 6.962.580$ $\bar{x}_1 = 1.011{,}67$ $r_1 = 0{,}19$ $\bar{y}_1 = 576{,}33$ $s_{y_1}^2 = 6.726{,}97$

Estrato II

$\Sigma x_2 = 23.330$ $\Sigma x_2^2 = 45.702.100$ $\Sigma y_2 = 7.896$ $\Sigma y_2^2 = 5.360.758$ $b_2 = 0{,}12556$

$s_{x_2}^2 = 31.335{,}61$ $\Sigma x_2 y_2 = 15.394.420$ $\bar{x}_2 = 1.944{,}17$ $r_2 = 0{,}18$ $\bar{y}_2 = 658$ $s_{y_2}^2 = 15.017{,}27$

Estrato III

$\Sigma x_3 = 38.590$ $\Sigma x_3^2 = 129.241.300$ $\Sigma y_3 = 8.533$ $\Sigma y_3^2 = 6.213.165$ $b_3 = 0{,}04686$

$s_{x_3}^2 = 467.481{,}06$ $\Sigma x_3 y_3 = 27.681.670$ $\bar{x}_3 = 3.215{,}83$ $r_3 = 0{,}28$ $\bar{y}_3 = 711{,}08$ $s_{y_3}^2 = 13.226{,}45$

Calculamos las b_h: b_1, b_2 y b_3 mediante fórmulas que anteriormente fueron presentadas:

$$b_1 = \frac{12(6.962.580) \quad (12.140)(6.916)}{12(12.736.600) \quad (12.140)^2} = \quad 0{,}07496$$

$$b_2 = \frac{12(15.394.420) \quad (23.330)(7.896)}{12(45.702.100) \quad (23.330)^2} = 0{,}12556$$

$$b_3 = \frac{12(27.681.670) \quad (38.590)(8.533)}{12(129.241.300) \quad (38.590)^2} = 0{,}04686$$

Para efectuar los estimativos de Promedio, mediante la *Regresión lineal*, se debe conocer o suponer conocida la Media poblacional de $\overline{X}_h$ (medias) para cada estrato, que serán dadas en este ejercicio así:

$\overline{X}_1 = 1.080$ $\overline{X}_2 = 1.850$ $\overline{X}_3 = 3.220$

Con estos valores se podrán determinar los Promedios estimados de $\hat{Y}$:

$\overline{y}_{RL_1} = 576{,}33 \quad 0{,}07496(1.080 \quad 1.011{,}67) = 571{,}21$

$\overline{y}_{RL_2} = 658{,}0 + 0{,}12556(1.850 \quad 1.944{,}17) = 646{,}17$

$\overline{y}_{RL_3} = 711{,}08 + 0{,}04686(3.220 \quad 3.215{,}83) = 711{,}27$

La *Media ponderada* en la *Regresión separada* será igual a:

$$\hat{\overline{y}}_{RL_S} = \Sigma W_h \, \overline{y}_{RL_h} \qquad \overline{y}_{RL_S} = 0{,}46(571{,}21) + 0{,}37(646{,}17) + 0{,}17(711{,}27) = 622{,}76$$

Calculemos ahora el *Estimador de la Varianza*, para luego obtener el *error*.

$$\hat{V}_{y_{RL_S}} = \Sigma \frac{W_h^2(1 \quad f_h)}{n_h} \quad s_{y_h}^2 \quad 2b_h m_{x_h y_h} + b_h^2 s_{x_h}^2$$

Cálculo de las *Covarianzas*: (Nota: se trabajó en la calculadora con el programa de estadística).

$$m_{x_1 y_1} = \frac{\Sigma x_1 y_1}{n_1} \quad \overline{x}_1 \overline{y}_1 = \frac{6.962.580}{12} \quad (1.011{,}67)(576{,}33) = \quad 2.840{,}77$$

$$m_{x_2 y_2} = \frac{\Sigma x_2 y_2}{n_2} \quad \overline{x}_2 \overline{y}_2 = \frac{15.394.420}{12} \quad (1.944{,}17)(658) = 3.604{,}47$$

$$m_{x_3 y_3} = \frac{\Sigma x_3 y_3}{n_3} \quad \overline{x}_3 \overline{y}_3 = \frac{27.681.670}{12} \quad (3.215{,}83)(711{,}08) = 20.093{,}44$$

También se puede calcular la *Covarianza para cada estrato* de la siguiente manera:

$$Cov_{x_h y_h} = m_{x_h y_h} = \frac{n_h \Sigma x_{hi} y_{hi} \quad (\Sigma x_{hi})(\Sigma y_{hi})}{n_h^2} \qquad \acute{o} \qquad m_{x_h y_h} = \frac{\Sigma x_{hi} y_{hi} \quad n_h(\overline{x}_h)(\overline{y}_h)}{n_h}$$

Obtenidas las *Covarianzas* se podrá determinar el *estimador de la varianza*; para calcular el error.

$$\hat{V}_{y_{RL_s}} = \frac{0{,}46^2\left(1 - \frac{12}{355}\right)}{12}(6.726{,}97) \quad 2(\ 0{,}07496)(\ 2.840{,}77) + (\ 0{,}07496)^2(41.360{,}61) +$$

$$\frac{0{,}37^2\left(1 - \frac{12}{355}\right)}{12}(15.017{,}27) \quad 2(0{,}12556)(3.604{,}47) + 0{,}12556^2(31.335{,}61) +$$

$$\frac{0{,}17^2\left(1 - \frac{12}{355}\right)}{12}(13.226{,}45) \quad 2(0{,}04686)(20.093{,}44) + 0{,}04686^2(467.481{,}06) = 291{,}02$$

Los *límites de confianza* del 95% para el estimador del promedio de consumo diario serán:

$$\hat{\bar{Y}}_{RL_{S_{S_i}}} = \bar{y}_{RL_S} \quad t\sqrt{\hat{V}_{y_{RL_s}}}$$

$$\hat{\bar{Y}}_{RL_{S_{S_i}}} = 622{,}78 \quad 2{,}035\sqrt{291{,}02} \begin{cases} \nearrow 657{,}50 \\ \searrow 588{,}06 \end{cases}$$ Consumo diario de carne en gramos

b. En la *regresión lineal combinada*, el proceso a seguir, para el estimador del promedio será:

$$\hat{\bar{Y}}_{RL_C} = \bar{y}_{st} + b_c(\bar{X}_{st} - \bar{x}_{st})$$

Calculamos en primer lugar las *medias aritméticas*, tanto de **X** como de **Y**:

$$\bar{y}_{st} = \Sigma W_h \bar{y}_h = 0{,}46(576{,}33) + 0{,}37(658) + 0{,}17(711{,}08) = 629{,}46$$

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h = 0{,}46(1.011{,}67) + 0{,}37(1.944{,}17) + 0{,}17(3.215{,}83) = 1.731{,}40$$

Suponiendo que $\bar{X}_{st} = 1.729$ (poblacional) Se tendrá:

$$\bar{y}_{RL_c} = 629{,}46 + 0{,}01994(1.729 \quad 1.731{,}40) = 629{,}41$$

El *coeficiente angular combinado* se calculó de la siguiente manera:

$$b_c = \Sigma W_h b_h = 0{,}46(\ 0{,}07496) + 0{,}37(0{,}12556) + 0{,}17(0{,}04686) = 0{,}01994$$

Estimador de laVarianza o Error de estimación al cuadrado en la Regresión *lineal combinada*:

$$\hat{V}_{\bar{y}_{RL_c}} = \Sigma \quad \frac{W_h^2(1 \ f_h)}{n_h} \quad s_{y_h}^2 \quad 2b_c m_{x_h y_h} + b_c^2 s_{x_h}^2$$

$$\hat{V}_{y_{RL_s}} = \frac{0{,}46^2\left(1 - \frac{12}{162}\right)}{12}(6.726{,}97) \quad 2(0{,}01994)(\ 2.840{,}77) + 0{,}01994^2(41.360{,}61) +$$

$$\frac{0{,}37^2\left(1 - \frac{12}{132}\right)}{12} \quad (15.017{,}27) \quad 2(0{,}01994)(3.604{,}47) + 0{,}01994^2(31.335{,}65) +$$

$$\frac{0{,}17^2\left(1 - \frac{12}{61}\right)}{12} \quad (13.226{,}45) \quad 2(0{,}1994)(20.093{,}44) + 0{,}1994^2(467.481{,}06) = 290{,}42$$

Los *límites de confianza* del 95 para el estimador de $\hat{\hat{}}$ serán:

$$\hat{\hat{Y}}_{RL_{c_{s_i}}} = \bar{y}_{RL_c} \quad t\sqrt{\hat{V}_{y_{RL_c}}}$$

$$\hat{\hat{Y}}_{RL_{c_{s_i}}} = 629{,}41 \quad 2{,}035\sqrt{290{,}42} \begin{cases} 664{,}08 \\ 594{,}73 \end{cases}$$ Consumo promedio diario de carne en gr.

El resultado anterior significa un consumo diario promedio de carne por familia de 629,41 gramos; pero el verdadero consumo promedio, estará entre 594,73 y 664,08 gramos, con una confianza o seguridad del 95%.

ESTIMACIÓN DE UNA PROPORCIÓN Y EL TOTAL EN CONGLOMERADOS

En el muestreo *aleatorio estratificado* con ***asignación igual***, al estimador de la proporción cuando la unidad se constituye en *conglomerados* es igual a:

$$p_{st} = \frac{\Sigma M_h p_h}{M} \qquad \text{siendo: } p_h = \frac{\Sigma a_{hi}}{\Sigma m_{hi}} \qquad M_h = N\bar{m}_h \quad \text{y} \quad \Sigma M_h = M$$

El *estimador de la varianza* o *error de estimación al cuadrado* será igual a:

$$V_{p_{st}} = \Sigma \frac{(1 - f_h)}{n_h \bar{m}_h^2} \frac{\Sigma a_{hi}^2 - 2p_h(\Sigma a_{hi} m_{hi}) + p_h^2 \Sigma m_{hi}^2}{n_h - 1}$$

Los *límites de confianza* estarán dados por la órmula: $\hat{p}_{st_{s_i}} = p_{st} + \sqrt{\hat{V}_{p_{st}}}$

Para la aplicación de estas fórmulas tengamos en cuenta la información correspondiente a la muestra de 36 familias y como características, el número de mujeres y el total de personas por familia, con el fin de determinar la *proporción* y el *total de mujeres.*

m_i = *Número de personas por Familia*

a_i = *Número de mujeres por Familia*

Calculamos el valor de *p* para cada uno de los estratos:

$\Sigma a_{1i}=18 \quad \Sigma a_{2i}=22 \quad \Sigma a_{3i}=25 \quad \Sigma a_{1i}m_{1i}=62 \quad \Sigma m_{1i}=33 \quad \Sigma m_{2i}=47 \quad \Sigma m_{3i}=42$

$\Sigma a_{2i}m_{2i}=91 \quad \Sigma a_{3i}m_{3i}=99 \quad p_1=\frac{18}{33}=0{,}545 \quad p_2=\frac{22}{47}=0{,}468 \quad p_3=\frac{25}{42}=0{,}595$

Para el cálculo del *estimador de la varianza* se requiere de información complementaria:

$\Sigma a_{1i}^2=38 \quad \Sigma a_{2i}^2=46 \quad \Sigma a_{3i}^2=61 \quad \Sigma m_{1i}^2=111 \quad \Sigma m_{2i}^2=193$

$\Sigma m_{3i}^2=164 \quad \bar{m}_1=2{,}75 \quad \bar{m}_2=3{,}92 \quad \bar{m}_3=3{,}50$

$f_1=\frac{n_1}{N_1}=\frac{12}{162}=0{,}074 \quad f_2=\frac{n_2}{N_2}=\frac{12}{132}=0{,}091 \quad f_3=\frac{n_3}{N_3}=\frac{12}{61}=0{,}197$

El *estimador de la varianza* será igual a:

$$\hat{V}_{p_{st}}=\frac{1\ \ 0{,}74}{12(2{,}75)^2}\ \frac{38\ \ 2(0{,}545)(62)+(0{,}545)^2(111)}{12\ \ 1}+\frac{1\ \ 0{,}091}{12(3{,}92)^2}\ \frac{46\ \ 2(0{,}468)(91)+(0{,}468)^2(193)}{12\ \ 1}+$$

$$\frac{1\ \ 0{,}197}{12(3{,}5)^2}\ \frac{61\ \ 2(0{,}595)(99)+(0{,}595)^2(164)}{12\ \ 1}=0{,}00508$$

Los *límites de confianza* del 95% para la proporción serán:

$$\hat{P}_{st_s}=p_{st}\pm t\sqrt{\hat{V}_{p_{st}}} \qquad p_{st}=\frac{\Sigma M_h p_h}{M}$$

El Total de elementos en los Conglomerados de cada estrato (M_h) se puede calcular de dos maneras: la primera, mediante la aplicación de la fórmula $\Sigma M_h = N m_h$

$M_1=355(2{,}75)=976{,}25 \qquad M_2=355(3{,}92)=1.391{,}6 \qquad M_h=N\bar{m}_h$

$M_3=355(3{,}50)=1.242{,}5 \qquad \Sigma M_h = M = 3.611$

El anterior procedimiento de cálculo de $\mathbf{M}_h$ es muy común, sin embargo se aconseja que en la práctica se haga de una manera mucho más fácil y práctica, siendo así:

$$M=\Sigma N_h\bar{m}_h=\Sigma M_h$$

siendo $M_1=162(2{,}75)=446 \quad M_2=132(3{,}92)=517 \quad M_3=61(3{,}50)=214 \quad M=\Sigma M_h=446+517+214=1.177$

Este último procedimiento lo vamos a utilizar en nuestros cálculos

$$p_{st}=\frac{446(0{,}545)+517(0{,}468)+1.214(0{,}595)}{1.177}=0{,}52=52 \quad \text{son mujeres}$$

$$\hat{p}_{st_{s_i}}=0{,}52 \quad 2{,}035\sqrt{0{,}00508}\ \begin{matrix}\nearrow 66{,}5\\ \searrow 37{,}5\end{matrix} \qquad \text{Porcentaje de mujeres}$$

El anterior resultado nos indica, que un 52% de las personas en el barrio investigado son mujeres y el 48% restante hombres. Sin embargo, la proporción verdadera de mujeres deberá estar entre 37,5% y 66,5%, con una confianza o seguridad en la estimación del 95%.

Si se desea estimar el Total de personas en las 355 familias, además del método anterior, cuyo resultado fue de 1.177 personas, también se puede calcular el valor de M, de varias maneras, así:

$M = N(\bar{x}) = 355(3{,}3104) = 1.175$ personas

También: $M = N(\bar{m}) = 355(3{,}3888) = 1.203$ personas

Los *límites de confianza* del 95% para el total de mujeres en las 355 familias se obtendrán de la siguiente manera:

$$\hat{A}_{st} = Mp_{st} \; t\, M\sqrt{\hat{V}_{p_{st}}} \qquad \hat{A}_{st_i} = 1.177(0{,}52) \quad 2{,}035(1.177)\sqrt{0{,}00508} \begin{matrix} \nearrow 783 \\ \text{Mujeres} \\ \searrow 441 \end{matrix}$$

Lo anterior nos indica que, aproximadamente 612 personas son de sexo femenino y 565 masculino. El verdadero total de mujeres (poblacional) deberá estar entre 441 y 783, con una seguridad del 95%.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

72. Dada la siguiente población estratificada

ESTRATO I		ESTRATO II		ESTRATO III	
No. de orden	y_1	**No. de orden**	y_2	**No. de orden**	y_3
1	5	1	9	1	13
2	7	2	8	2	15
3	3	3	10	3	16
4	2	4	8	4	12
5	1			5	18
6	4				

Se pide extraer una muestra de tamaño 9 (tamaño de la encuesta preliminar) aplicando Afijación igual y calcular el tamaño de la muestra, trabajando con un error del 6% y una confianza del 95%.

73. Se tiene la siguiente población de 20 unidades.

No. DE ORDEN	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
x_i:	10	6	12	11	6	8	14	13	14	10	9	8	7	12	10	16	6	10	15	8

Información adicional: $\Sigma X_i = 205$ $\Sigma X_i^2 = 2.281$

Con base en una muestra aleatoria simple se desea estimar $\hat{\bar{X}}$ y $\hat{X}$

a. Determinar el tamaño de la muestra (aleatoria simple) con un error de 2 y una confianza del 90%.

b. Seleccionar una muestra de tamaño igual al determinado en el punto anterior. Calcular el estimador del promedio y total de x_1, fijando límites del 95%.

74. Dividir la anterior población en dos estratos de tamaños 13 y 7 respectivamente, indicando el criterio empleado al realizar dicha estratificación.
    a. Seleccionar una muestra de tamaño 6 (encuesta preliminar) y mediante la asignación igual calcular el valor n, con un error del 9% y confianza del 95,5%.
    b. Considerando el tamaño de la muestra igual a 6, estimar el promedio y el total de X, fijando límites de confianza de 95%.

75. Con los siguientes datos

ESTRATOS	$_h$	$\mathbf{n}_h$	$_h$	Σ_h	Σ_h^2	Σ_h	Σ_h^2
	270	20	0,35	207,4	2.181,4	169,3	1.456,19
	260	20	0,50	324,3	5.539,01	246,3	3.087,29
	227	20	0,80	543,3	14.860,71	356,1	8.411,97
	354	20	0,20	846,1	36.156,83	676,8	23.147,10

Continuación

$\Sigma x_h y_h$	$\bar{x}_h$	$\bar{y}_h$	S_{x_h}	S_{y_h}	$\bar{X}_h$
1.770,66	8,47	10,37	0,83	1,26	8,2
4.045,23	12,32	16,22	1,69	3,74	14,0
9.763,78	17,81	27,17	10,44	2,32	19,6
29.391,08	33,84	42,31	3,54	4,37	31,2

76. Con los datos del ejercicio 75, se pide:
    a. Estimar el promedio y el total $\hat{}$, fijando límites de confianza del 95%
    b. Estimar el promedio y el total de $\hat{}$, fijando límites de confianza del 95%

77. Con los datos del ejercicio 75, se pide:
    a. Estimar la proporción y el total de $\hat{A}$, fijando límites de confianza del 95%
    b. Estimar el promedio total de Y mediante la regresión lineal (combinada), si $\bar{}$ = 22,6, fijando límites de confianza del 95%.

78. Con los datos del ejercicio 75, se pide:
    a. Estimar el promedio y el total de $\hat{}$, mediante la regresión lineal (separada), fijando límites de confianza del 95%.
    b. Estimar el promedio y el total de $\hat{}$, mediante el método de la razón (combinada), si $\bar{}$ = 22,6 fijando límites de confianza del 95%.
    c. Estimar el promedio y el total de $\hat{}$, mediante el método de la razón (separada) si $\bar{}_{st}$ = 22,6, fijando límites de confianza del 95%.

79. Se tiene la siguiente información:

ESTRATOS	N_h	S_h	$\bar{x}_h$	p_h
Peque o	500	3,48	15,3	0,3
Mediano	300	2,80	22,8	0,2
Grande	200	4,00	29,7	0,4

P = 95
E = 2,1 (promedio)
E = 0,08 (proporción)

80. Con los datos del ejercicio 79, se pide:
    a. Determine el tamaño de la muestra, aplicando el método de afijación igual, suponiendo que los anteriores resultados corresponden a una encuesta preliminar.

b. Estime el promedio y el total de $\bar{}$, fijando límites de confianza del 95%

c. Estime la proporción y el total de $\hat{P}$, fijando límites confianza del 95%

81. Con los siguientes datos:

Estrato I

No. DE ORDEN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X_i :	17	9	9	18	23	32	21	23	15	22	16	15	19	28	25	19	14	18	23	12
Y_i :	9	6	7	10	16	18	14	14	10	17	11	19	12	5	13	11	10	12	13	8
A_i :	si	si	no	si	si	no	no	si	si	si	si	no	si	no	no	si	no	no	si	si

Estrato II

No. DE ORDEN	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X_i :	40	53	49	38	42	54	50	43	47	59	63	58	45	46	54
Y_i :	23	36	30	22	26	32	27	24	21	32	32	35	28	23	29
A_i :	no	no	si	si	no	si	si	no	si	si	no	si	si	no	si

Estrato III

No. DE ORDEN	1	2	3	4	5	6	7	8	9	10	11
X_i :	76	82	74	80	86	91	74	81	78	83	76
Y_i :	51	47	45	54	54	52	45	47	46	52	51
A_i :	si	si	no	si	si	no	si	no	si	si	no

Considerar los datos anteriores como la información de una población y seleccionar una encuesta preliminar de 12 elementos con el fin de determinar el tamaño de la muestra, mediante el método de asignación igual. Error del 4% para la variable y del 10% para la proporción y una confianza del 95%

a. Utilizar la variable X, para realizar los cálculos.

b. Utilizar la característica A para las respuestas <<si>>.

82. Con los tamaños muestrales (asignación igual) para la variable Y , proceda a:

a. Estimar el valor de $\hat{\bar{Y}}_{RL_S}$ y $\hat{\bar{Y}}_{RL_C}$ mediante la regresión lineal, sabiendo que $\bar{X}_1 = 24$ $\bar{X}_2 = 56$ $\bar{X}_3 = 82$ y una confianza del 95%

b. Estimar el valor de $\hat{\bar{X}}$ y $\hat{X}$ (total) utilizando una confianza del 90%

c. Estimar el valor de $\hat{P}$ y $\hat{A}$, utilizando una confianza del 99%

83. Con la siguiente información:

Estrato I

No.	y_i	x_i
1	12,6	8,4
2	13,0	7,2
3	12,8	8,3
4	12,6	9,5
5	13,0	9,4
6	12,9	10,1
7	13,5	8,2
8	14,0	9,6
9	13,2	8,4
10	12,4	8,6

Seleccione una muestra de tamaño 18, aplicando el método de Asignación igual y, estime $\hat{R}_c$, $\hat{R}_s$, fijando límites de confianza del 95%.

Se sabe además que $\bar{\ } = 19{,}18$ $\bar{\ }_1 = 9{,}1$ $\bar{\ }_2 = 17{,}3$ $\bar{\ }_3 = 27{,}1$.

84. Con la misma muestra obtenida en el punto anterior estime $\hat{\bar{Y}}_{RLS}$ y $\hat{\bar{Y}}_{RLC}$ fijando límites de confianza del 95%. (Considere los valores dados en el punto anterior en el caso que los necesite).

84. Con los siguientes datos:

Estrato I

No. ORDEN	No. ALEATORIO	PERSONAS		SERVICIOS		INGRESOS MENSUALES (Miles $)
		H	M	Luz	Agua	
1	70	2	1	no	no	4.500
2	08	1	1	no	si	6.000
3	36	0	2	no	si	5.000
4	42	3	1	si	si	6.200
5	28	2	2	si	no	7.000
6	16	1	3	no	no	5.000
7	22	1	3	si	si	4.500
8	55	1	2	si	si	4.800

Estrato II

No. ORDEN	No. ALEATORIO	PERSONAS		SERVICIOS		INGRESOS MENSUALES (Miles $)
		H	M	Luz	Agua	
1	28	4	3	si	si	6.000
2	33	2	6	no	si	5.800
3	42	3	5	si	no	7.200
4	06	4	4	no	no	6.000
5	01	4	2	si	si	4.800
6	19	3	4	no	si	6.000
7	22	2	5	no	si	5.700
8	38	4	2	si	si	6.400

Estrato III

No. ORDEN	No. ALEATORIO	PERSONAS		SERVICIOS		INGRESOS MENSUALES (Miles $)
		H	M	Luz	Agua	
1	15	3	7	si	si	4.600
2	20	5	4	si	no	5.200
3	16	6	3	no	no	6.300
4	09	5	7	no	no	7.000
5	12	2	8	si	si	6.800
6	18	6	5	si	no	5.900
7	25	4	8	no	si	7.000
8	14	5	5	no	si	6.200

Información complementaria:

M = 150 viviendas $N_1 = 75$ $N_2 = 50$ $N_3 = 25$ Z = 2 (P = 95,5%) E = 5%

Estrato I : Viviendas de menos de 5 personas
Estrato II : Viviendas de 5 a 8 Personas
Estrato III : Viviendas de 9 y más personas

85. Con los datos del ejercicio anterior, se pide:
    a. Considerar que los anteriores resultados fueron obtenidos mediante una Encuesta preliminar (Asignación igual) y determinar el tamaño de la muestra.
    b. Estimar la proporción de mujeres, fijando límites de confianza del 95% (Conglomerados).

86. Con los datos del ejercicio anterior, se pide:
    a. Estimar la Razón combinada: Mujeres/hombres y fijar limites de confianza del 95% para el estimador.
    b. Estimar la Proporción de viviendas con servicio de luz y agua y fijar límites del 95% de confianza para el estimador.

ASIGNACIÓN PROPORCIONAL

Este método nos permite determinar el tamaño óptimo de la muestra, así como los estimados puntuales y límites de confianza para el *promedio*, *proporción, razón*, y *proporciones en conglomerados*, en forma similar al de la Asignación igual, con la diferencia de que los tamaños maestrales de cada estrato, en vez de ser iguales en cuanto al número de unidades, éstos se distribuyen en la misma proporción que las unidades en la población; en otras palabras, el peso relativo dado por el número de unidades en cada estrato en relación al total de elementos de la población, debe ser igual al obtenido en la muestra.

$$W_h = \frac{N_h}{N} \qquad w_h = \frac{n_h}{n} \qquad \text{Siendo:} \quad W_h = w_h$$

Mediante este procedimiento, al igual que en el de *Asignación igual* y *óptima*, se pueden obtener mejores estimaciones que a través del *Muestreo aleatorio simple*, ya que es posible contener en la muestra, unidades de todos los tipos y características, haciéndola de esta manera mucho más representativa de la población.

Tamaño de la muestra

Como en todos los casos tratados anteriormente, cuando no se conoce la *Varianza poblacional* de la característica principal de la investigación, para el cálculo del tamaño muestral, una de las alternativas consiste en realizar una *Encuesta preliminar* o *piloto*.

Recordemos que en la *Asignación igual*, el número de unidades es igual para cada uno de los estratos muestrales. En la encuesta preliminar de 14 familias, ésta fue la razón por la cual se distribuyó un total de (5) cinco familias para cada estrato; en cambio en la *Asignación proporcional*, las 14 unidades se distribuyen en la muestra en la misma proporción como se distribuyen en la población.

El peso relativo sigue siendo:

$$W_1 = \frac{N_1}{N} = \frac{162}{355} = 0{,}46 \qquad W_2 = \frac{N_2}{N} = \frac{132}{355} = 0{,}37 \qquad W_3 = \frac{N_3}{N} = \frac{61}{355} = 0{,}17$$

$$W_1 + W_2 + W_3 = 0{,}46 + 0{,}37 + 0{,}17 = 1 \qquad \Sigma W_h = 1 \quad \text{ó} \quad 100$$

Los anteriores resultados, serán utilizados para calcular los tamaños de la muestra en la encuesta preliminar, para cada uno de los estratos:

$$w_1 = \frac{n_1}{n} = \frac{7}{15} = 0{,}47 \Rightarrow w_1 = W_1 \qquad w_2 = \frac{n_2}{n} = \frac{5}{15} = 0{,}33 \Rightarrow w_2 = W_2$$

$$w_3 = \frac{n_3}{n} = \frac{3}{15} = 0{,}20 \Rightarrow w_3 = W \qquad \Sigma w_h = w_1 + w_2 + w_3 = 1 \text{ ó } 100$$

En esta forma se tendrá que el peso relativo en la muestra deberá ser igual al de la población:

$$n_1 = nW_1 = 14(0{,}46) \cong 7 \qquad n_2 = nW_2 = 14(0{,}37) \cong 5 \qquad n_3 = nW_3 = 14(0{,}17) \cong 3$$

Las diferencias que resultan de las operaciones anteriores, se deben a las aproximaciones realizadas para calcular los tamaños muestrales en cada estrato, pero de todas maneras deberán considerarse como iguales. Las diferencias presentadas carecen de importancia.

Una vez determinados los tamaños muestrales para la Encuesta preliminar, se procede a la selección de las unidades requeridas, que permitirán obtener información y así calcular las varianzas y el error:

Encuesta preliminar – Asignación proporcional

Estrato I

No. de ORDEN	No. ALEATORIO	INGRESOS MENSUALES (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DE CARNE DIARIO Grs.
				Total	Masc.	Fem.	Trabajando	
1	086	1.630	si	3	2	1	1	605
2	151	1.230	no	3	2	1	1	605
3	008	790	si	2	0	2	1	680
4	122	1.260	no	4	2	2	1	520
5	076	1.260	no	3	1	2	1	516
6	024	840	si	4	2	2	1	610
7	055	1.020	no	1	1	0	1	530

Estrato II

No. de ORDEN	No. ALEATORIO	INGRESOS MENSUALES (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DE CARNE DIARIO Grs.
				Total	Masc.	Fem.	Trabajando	
1	129	2.060	si	1	1	4	2	780
2	063	1.860	no	1	2	1	2	586
3	015	2.250	si	1	3	2	2	852
4	039	1.720	si	2	4	2	2	580
5	111	2.050	si	2	5	2	2	734

Estrato III

No. de ORDEN	No. ALEATORIO	INGRESOS MENSUALES (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO DE CARNE DIARIO Grs.
				Total	Masc.	Fem.	Trabajando	
1	03	3.060	si	5	4	1	3	804
2	28	2.960	no	5	2	3	3	548
3	11	2.890	no	4	2	2	2	770

Con la información anterior respecto a la variable Ingresos (miles de $), calculamos el Tamaño óptimo. También se hará con la característica Propietarios de vivienda, únicamente con el fin de aplicar las fórmulas respectivas.

Los *promedios* y *varianzas* para los ingresos en cada estrato son:

$\bar{x}_1 = 1.147{,}14 \quad \bar{x}_2 = 1.988 \quad \bar{x}_3 = 2.970 \quad s_1^2 = 83.923{,}81 \quad s_2^2 = 41.470 \quad s_3^2 = 7.300$

La *media ponderada* será:

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h = 0{,}46(1.147{,}14) + 0{,}37(1.988) + 0{,}17(2.970) = 1.768{,}17$$

El *error de muestreo* con un 5% será igual a:

$E = 0{,}05(\bar{x}_{st}) = 0{,}05(1.768{,}17) = 88{,}41$

La primera aproximación para ***n*** será

$$n_o = \frac{\Sigma W_h S_h^2}{(E/Z)^2} \qquad n_o = \frac{0{,}46(83.923{,}81) + 0{,}37(41.470) + 0{,}17(7.300)}{(88{,}41 \;\; 1{,}96)^2} = 27{,}13$$

$$n = \frac{27{,}13}{1+\frac{27{,}13}{355}} = 26 \text{ amilias}$$

Tambi n mediante la órmula: $n = \frac{NZ^2\Sigma W_h S_h^2}{NE^2+Z^2\Sigma W_h S_h^2}$ se tendrá que:

$$n = \frac{355(1{,}96)^2\ 0{,}46(83.923{,}81) + 0{,}37(41.470) + 0{,}17(7.300)}{355(88{,}41)^2 + 1{,}96^2\ 0{,}46(83.923{,}81) + 0{,}37(41.470) + 0{,}17(7.300)} = 26 \text{ amilias}$$

Veamos ahora, cómo sería el cálculo n en una Proporción. Para ello consideramos como característica cualitativa las familias propietarias de vivienda; para este caso, se establece una confianza del 95% y un error del 10%.

$$p_1 = \frac{3}{7} = 0{,}43 \qquad p_2 = \frac{4}{5} = 0{,}8 \qquad p_3 = \frac{1}{3} = 0{,}33$$

$$n_o = \frac{\Sigma W_h p_h q_h}{(E/Z)^2} \qquad n_o = \frac{0{,}46(0{,}43)(057)+0{,}37(0{,}8)(0{,}2)+0{,}17(0{,}33)(0{,}67)}{(0{,}10\ 1{,}96)^2} = 80{,}49$$

$$n_o = \frac{80{,}49}{1+\frac{80{,}49}{355}} = 66 \text{ amilias}$$

Tambi n mediante la aplicación de $n = \frac{NZ^2\Sigma W_h p_h q_h}{NE^2+Z^2\Sigma W_h p_h q_h}$

$$n = \frac{355(1{,}96)^2\ 0{,}46(0{,}43)(0{,}57) + 0{,}37(0{,}8)(0{,}2) + 0{,}17(0{,}33)(0{,}67)}{355(0{,}10)^2 + 1{,}96^2\ 0{,}46(0{,}43)(0{,}57) + 0{,}37(0{,}8)(0{,}2) + 0{,}17(0{,}33)(0{,}67)} = 66 \text{ amilias}$$

En el tema anterior, para la estimación de promedios y totales por el método de la *Asignación igual*, se había supuesto que el tamaño de la muestra era de 36 familias. En el caso de la *Asignación proporcional*, también consideraremos que n sigue siendo igual a 36. Los tamaños muestrales para cada estrato serán:

$n_1 = nW_1 = 36(0{,}46) = 17$ familias $\qquad n_2 = nW_2 = 36(0{,}37) = 13$ familias

$n_3 = nW_3 = 36(0{,}17) = 6$ familias $\qquad n = 36$ familias

Determinados los tamaños muestrales para cada estrato, seleccionamos aleatoriamente las unidades para la muestra, con lo cual se obtiene la información necesaria, para explicar los diferentes métodos de *estimación puntual* y de *intervalos*.

Muestra de asignación proporcional

Estrato I

No. de ORDEN	No. ALEATORIO	INGRESOS MENSUALES (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO CARNE DIARIO Grs.
				Total	Masc.	Fem.	Trabajando	
1	063	1.350	no	2	2	0	1	638
2	160	1.240	no	2	2	0	1	580
3	124	1.010	no	4	1	3	1	520
4	008	790	si	3	1	2	1	620
5	096	1.130	si	6	2	4	1	700
6	110	850	no	4	1	3	1	520
7	073	890	si	3	2	1	1	520
8	037	1.260	no	3	1	2	1	580
9	021	1.060	no	3	1	2	1	730
10	056	950	si	2	1	1	1	578
11	116	1.080	si	4	1	3	2	716
12	141	950	si	2	1	1	1	520
13	003	1.080	no	3	2	1	1	735
14	018	860	no	1	0	1	1	510
15	076	1.260	no	3	1	2	1	516
16	020	950	no	3	1	2	1	620
17	133	990	si	3	1	2	1	512

Estrato II

No. de ORDEN	No. ALEATORIO	INGRESOS MENSUALES (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO CARNE DIARIO Grs.
				Total	Masc.	Fem.	Trabajando	
1	006	2.350	si	3	2	1	3	732
2	091	1.830	si	3	0	3	2	636
3	121	2.180	no	5	2	3	3	687
4	034	2.020	si	3	1	2	2	680
5	130	1.650	no	2	0	2	1	530
6	115	1.920	si	5	3	2	2	732
7	081	2.150	no	3	1	2	2	613
8	075	1.960	si	2	1	1	1	642
9	008	2.000	si	3	1	2	1	804
10	100	2.260	si	2	1	1	2	712
11	019	1.760	si	3	1	2	2	880
12	023	2.140	si	5	2	3	3	540
13	035	1.930	si	4	2	2	2	520

Estrato III

No. de ORDEN	No. ALEATORIO	INGRESOS MENSUALES (Miles $)	PROPIEDAD VIVIENDA	NÚMERO DE PERSONAS				CONSUMO CARNE DIARIO Grs.
				Total	Masc.	Fem.	Trabajando	
1	31	3.210	no	4	2	2	2	806
2	42	2.960	no	3	1	2	2	684
3	53	3.460	si	4	1	3	3	718
4	20	3.870	no	4	1	3	2	735
5	06	2.600	si	4	2	2	2	752
6	38	2.990	no	4	2	2	2	755

Estimación de promedios y totales

Para este método, únicamente se calculará el *estimador del promedio* y del *total de ingresos por familia,* dejando al lector la aplicación de fórmulas para hallar promedio de personas por familia y consumo diario de carne (en gramos), toda vez que el proceso es similar al utilizado en la *Asignación igual*.

Las *Medias aritméticas* y *Varianzas del Ingreso* (en miles de $) por familia son:

$\bar{x}_1 = 1.041,18$ $\bar{x}_2 = 2.011,54$ $\bar{x}_3 = 3.181,67$ $s_1^2 = 26.611,03$

$s_2^2 = 40.430,77$ $s_3^2 = 195.256,67$ $s_1 = 163,13$ $s_2 = 201,07$

$s_3 = 441,88$ $n_1 = 17$ $n_2 = 13$ $n_3 = 6$

El *es ima i o pun ual* se obtiene así:

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h = 0,46(1.041,18) + 0,37(2.011,54) + 0,17(3.181,67) = 1.764,10$$

El *estimador de la varianza* para el cálculo del error de estimación será:

$$\hat{V}_{\bar{x}_{st}} = \frac{1}{N^2} \Sigma N_h (N_h - n_h) \frac{s_h^2}{n_h}$$

$$\hat{V}_{xst} = \frac{1}{355^2} 162(162 \;\; 17)\frac{26.611,03}{17} + 132(132 \;\; 13)\;\frac{40.430,77}{13} + 61(61 \;\; 6)\;\frac{195.256,67}{6} = 1.545,76$$

$$\hat{S}_{\bar{x}_{st}} = \sqrt{1.545,76} = 39,32$$

y los *límites de confianza* del 95 para el *promedio*:

$$\hat{\bar{X}}_{st_{s_i}} = \bar{x}_{st} \quad t\hat{s}_{\bar{x}_{st}} \qquad \hat{\bar{X}}_{st_{s_i}} = 1.764,10 \pm 2,035(39,32) \begin{cases} 1.844,02 = & 1.844.016,2 \\ 1.684,08 = & 1.684.083,8 \end{cases}$$

$v = n_1 + n_2 + n_3 \quad 3 = 17 + 13 + 6 \quad 3 = 33$

Siendo: $\alpha = 0,05$, se tendrá que el valor de $t = 2,035$

Para el *total* estimado se tiene:

$$\hat{X}_{st_{s_i}} = N\bar{x}_{st} \quad tN\hat{s}_{\bar{x}_{st}} \qquad \hat{X}_{st_{s_i}} = 355(1.764,10) \quad 2,035(355)(39,32) \begin{cases} 654.661,25 & \text{(miles)} \\ 597.849,75 & \text{(miles)} \end{cases}$$

Se estima que el Ingreso total de las 355 familias será aproximadamente de $ 626.255.500 y que el verdadero total estará entre $ 597.849.750 y $ 654.661.250 con una seguridad del 95%.

Estimación de proporciones y totales

Consideramos como ejemplo la característica cualitativa: Propiedad de la vivienda, para estimar la Proporción y el Total de propietarios, fijando límites de confianza del 95%. Se tendrá:

$$_1 = \frac{7}{17} = 0{,}41 \qquad _2 = \frac{10}{13} = 0{,}77 \qquad _3 = \frac{2}{6} = 0{,}33$$

$$_s = \Sigma \qquad _s = 0{,}46(0{,}41) + 0{,}37(0{,}77) + 0{,}17(0{,}33) = 0{,}5296 \cong 52{,}96$$

Se estima que el 52,96 de las amilias son propietarias de vivienda.

$$\hat{}_{s_i} = \quad _s \quad \sqrt{\hat{}\quad _s} \qquad \hat{}_s = \frac{1}{^2} \quad \Sigma \quad (\quad n) \frac{}{n}$$

$$\hat{}_s = \frac{1}{355^2} \; 162(162 \quad 17) \frac{0{,}41(0{,}59)}{17} + 132(132 \quad 13) \frac{0{,}77(0{,}23)}{13} + 61(61 \quad 6) \frac{0{,}33(0{,}67)}{6} = 0{,}006$$

$$\hat{}_{s\,s_i} = 0{,}5296 \quad 2{,}035 \sqrt{0{,}006} \begin{cases} 68{,}72 \\ 37{,}20 \end{cases}$$

Entre 37,20% y 68,72% deberá estar el verdadero porcentaje de familias Propietarias de vivienda. Para calcular el Total de familias propietarias de vivienda, se tendrá.

$$\hat{}_{s_i} = Np_{st} \quad tN\sqrt{\hat{V}_{p_{st}}}$$

$$\hat{}_{s_i} = 355(0{,}53) \quad 2{,}035(355)\sqrt{0{,}006} \begin{cases} 244{,}11 = 244 \\ 132{,}19 = 132 \end{cases} \quad \text{Familias}$$

Se estima que en total, 188 familias son Propietarias de vivienda y el verdadero valor estará entre 132 y 244 familias, con una confianza del 95%.

Estimación de proporciones y totales en conglomerados

Como en los métodos anteriores, consideraremos el estimador de la Proporción en mujeres.

$$p_h = \frac{\Sigma a_{h_i}}{\Sigma m_{h_i}}$$

a_i = Total mujeres en la muestra

m_i = Total de personas (hombres y mujeres) en la muestra

$\Sigma a_1 = 30$ $\Sigma a_2 = 26$ $\Sigma a_3 = 14$ $\Sigma m_1 = 51$ $\Sigma m_2 = 43$ $\Sigma m_3 = 23$

$\Sigma a_1^2 = 72$ $\Sigma a_2^2 = 58$ $\Sigma a_3^2 = 34$ $\Sigma m_1^2 = 173$ $\Sigma m_2^2 = 157$ $\Sigma m_3^2 = 89$

$\Sigma a_1 m_1 = 107$ $\Sigma a_2 m_2 = 92$ $\Sigma a_3 m_3 = 54$ $\bar{m}_1 = 3$ $\bar{m}_2 = 3{,}31$ $\bar{m}_3 = 3{,}83$

$n_1 = 17$ $n_2 = 13$ $n_3 = 6$ $p_1 = 0{,}59$ $p_2 = 0{,}60$ $p_3 = 0{,}61$

El estimador de la varianza o error de estimación estará dado por:

$$\hat{}_s = \Sigma \frac{(1\quad)}{n\quad^{-}{}^2} \frac{\Sigma a^2 \quad 2\quad \Sigma a_{i\quad i} + {}^2\Sigma\ {}^2_i}{n\quad 1}$$

$$\hat{}_s = \frac{1\quad \frac{17}{162}}{17(9)} \frac{72\quad 2(0,59)(107) + 0,59^2(173)}{17\quad 1} + \frac{1\quad \frac{13}{132}}{13(10,96)} \frac{58\quad 2(0,60)(92) + 0,6^2(157)}{13\quad 1} +$$

$$\frac{1\quad \frac{6}{61}}{6(14,67)} \frac{34\quad 2(0,61)(54)+0,61^2(89)}{6\quad 1} =$$

$$\hat{}_s = (0,00585)(0,3726) + (0,0063)(0,3433) + 0,0102(0,2474) = 0,0069$$

$$_s = \Sigma \qquad = 0,46(0,59) + 0,37(0,6) + 0,17(0,61) = 0,597 \cong 0,60$$

$$\hat{}_{s\ s_i} = 0,60 \quad 2,035\sqrt{0,0069} \begin{cases} \nearrow 0,77 = 77 \\ \searrow 0,43 = 43 \end{cases}$$

Se estima que el 59.7% (aproximadamente el 60%) de las personas son mujeres y el porcentaje verdadero deberá estar entre 43% y el 77%.

Como no conocemos el total de personas en las 355 familias, se podrá realizar el estimador de totales, si establecemos una cifra de 1.093 personas, obtenido de la siguiente manera:

M=355(3,08)=1.093 3,08 = Promedio de personas por amilia 355 = Total de amilias

$$\hat{}_{st_{S\ i}} = 1.093(0,6) \quad 2,035(1.093)\sqrt{0,0069} \begin{cases} \nearrow 840,56 = 841 \\ \searrow 471,04 = 471 \end{cases} \text{ Personas}$$

Aproximadamente 655 son mujeres, pero el verdadero valor estará entre 471 y 840 con una seguridad del 95%.

Estimación indirecta del promedio y el total a través del método de la razón

En este caso, al igual que en la *Afijación igual*, la estimación de la Razón se puede hacer por cualquiera de los métodos, ya sea *separada* o *combinada*.

a. En el método Razón Separada, se procede de la siguiente manera:

$$\hat{}_h = \frac{\Sigma\ _{hi}}{\Sigma x_{hi}} \qquad \text{donde:}$$

Si trabajamos con la información obtenida a través de la muestra correspondiente a las 36 familias, se tendrá;

$^{-}_1 = 1,76 \qquad ^{-}_2 = 2,0 \qquad ^{-}_3 = 2,33 \qquad \bar{x}_1 = 1,24 \qquad \bar{x}_2 = 1,31 \qquad \bar{x}_3 = 1,5$

La variable Y la utilizamos para simbolizar al atributo mujeres y la X a la característica hombres, por lo tanto la *razón* a calcular será mujeres/hombres.

$$\hat{}_1 = \frac{1,76}{1,24} = 1,42 \qquad \hat{}_2 = \frac{2,0}{1,31} = 1,53 \qquad \hat{}_3 = \frac{2,33}{1,50} = 1,55$$

$$\hat{}_s = \Sigma W_h \hat{}_h = 0,46(1,42) + 0,37(1,53) + 0,17(1,55) = 1,48$$ por cada 2 hombres hay 3 mujeres.

Para el estimador de la varianza para calcular el error de estimación, se tiene:

$$\hat{V}\hat{}_s = \Sigma \frac{(1 \quad _h)}{n_h \bar{x}_h^2} \quad s^2_{\ h} \quad 2\hat{}_h \rho_h s_h s_{x_h} + \hat{}^2_h s^2_{x_h}$$

Los valores para ρ (rho) son:

$$\rho_1 = \frac{d_{x1}}{2d_1} = \frac{0,45}{2(0,62)} = 0,36 \qquad \rho_2 = \frac{d_{x_2}}{2d_2} = \frac{0,65}{2(0,36)} = 0,90$$

$$\rho_3 = \frac{d_{x_3}}{2d_3} = \frac{0,37}{2(0,22)} = 0,84$$

Siendo:

$$d_{x_1} = \frac{s_{x_1}}{\bar{x}_1} = \frac{0,56}{1,24} = 0,45 \qquad d_{x_2} = \frac{0,85}{1,31} = 0,65 \qquad d_{x_3} = \frac{0,55}{1,5} = 0,37$$

$$d_1 = \frac{s_1}{\bar{\ }_1} = \frac{1,09}{1,76} = 0,62 \qquad d_2 = \frac{0,71}{2,0} = 0,36 \qquad d_3 = \frac{0,52}{2,33} = 0,22$$

Las racciones de muestreo: $_1 = \frac{17}{162} = 0,10 \qquad _2 = 0,08 \qquad _3 = 0,10$

Con la anterior información se podrá calcular el estimador de la varianza para el cálculo del error:

$$\hat{}_{\hat{}_s} = \frac{1 \quad 0,10}{17(1,54)} \quad 1,19 \quad 2(1,42)(0,36)(0,56)(1,09) + 1,42^2(0,32) + \frac{1 \quad 0,08}{13(1,72)}$$

$$0,50 \quad 2(1,53)(0,9)(0,71)(0,85) + 1,53^2(0,73) + \frac{1 \quad 0,10}{6(2,25)} \quad 0,27 \quad 2(1,55)$$

$$(0,84)(0,55)(0,52) + 1,55^2(0,30) = 0,0805$$

Límites del 95 para el estimador:

$$\hat{}_{s_{s_i}} = \hat{}_s \quad \sqrt{\hat{}\,\hat{}_s} \qquad \hat{}_{s_{s_i}} = 1,49 \quad 2,035\sqrt{0,0805} \begin{cases} \nearrow 2,07 \cong 2 \text{ hombres} \\ \searrow 0,91 \cong 1 \text{ hombre} \end{cases}$$

Se estima por cada 3 mujeres hay dos hombres.

b. En el caso de la *razón combinada* ( ), se tendrá que:

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h = 0{,}46(1{,}24) + 0{,}37(1{,}31) + 0{,}17(1{,}5) = 1{,}31$$

$$\bar{}_{st} = \Sigma W_h \bar{}_h = 0{,}46(1{,}76) + 0{,}37(2{,}0) + 0{,}17(2{,}33) = 1{,}95$$

Siendo: $\hat{} = \frac{\bar{}_{st}}{\bar{x}_{st}} = \frac{1{,}95}{1{,}31} = 1{,}49$

El estimador de la varianza para el cálculo del error

$$\hat{V} = \Sigma \frac{(1 \ _h)}{n_h \bar{x}_h^2} \ s^2_h \ 2\hat{} \ \rho \, s_{x_h} s_h + {}^2 s^2_{x_h}$$

$$\hat{V}\hat{} = \frac{1 \ 0{,}10}{17(1{,}54)} \ 1{,}19 \ 2(1{,}49)(0{,}36)(0{,}56)(1{,}09) + 1{,}49^2(0{,}32) + \frac{1 \ 0{,}08}{13(1{,}72)}$$

$$0{,}5 \ 2(1{,}49)(0{,}9)(0{,}71)(0{,}85) + 1{,}49^2(0{,}73) + \frac{1 \ 0{,}10}{6(2{,}25)} \ 0{,}27 \ 2(1{,}49)$$

$$(0{,}84)(0{,}55)(0{,}52) + 1{,}49^2(0{,}30) = 0{,}0781$$

Estimación indirecta del promedio y total mediante la regresión lineal

La estimación del promedio cuando fue realizada mediante la *Asignación igual*, consideramos dos características para la aplicación del método de la *Regresión lineal*, consumo de carne diario en gramos (Y) e ingreso familiar (X). Recordemos que se usaron dos procedimientos de cálculo, el *separado* y el *combinado*.

Ahora serán aplicados, en la *Asignación proporcional* en la misma forma, como se utilizaron para la *Asignación igual.*

Veamos los resultados que se obtendrían en la *Asignación proporcional*.

$$\bar{} = \bar{} + \ (\bar{} \ \bar{x}) \qquad = \frac{\Sigma(\ _i \bar{})(x_i \ \bar{x})}{\Sigma(x_i \ \bar{x})^2} = \frac{n \Sigma \ _i x_i \ (\Sigma x_i)(\Sigma \ _i)}{n \Sigma x^2_i \ (\Sigma x_i)^2}$$

Para el reemplazo de estas fórmulas, tomaremos los resultados obtenidos mediante la muestra de las 36 familias, donde:

Se podrá calcular de la siguiente manera:

$$_h = \frac{\Sigma \ _{h_i} x_{h_i} \ n_h \bar{x}_h \bar{}_h}{\Sigma x^2_{h_i} \ n_h \bar{x}^2_h}$$

El promedio estimado, en la *re resi n lineal separada* se obtiene así:

$\bar{}_s = \Sigma \ _h \bar{}_h$ y sus límites de con ianza $\hat{}_i = \bar{}_s \ \sqrt{\hat{V}\bar{}_s}$

El estimador de la varianza para el cálculo del error $\hat{\ }_{s} = \Sigma \frac{^{2}(1\ \ )}{n}\ \ s^{2}\ \ 2\ \ _{x}\ \ +\ \ ^{2}s_{x}^{2}$

$m_{x_h y_h}$ = Se denomina *covarianza* y se puede calcular de varias maneras, como por ejemplo

$m_{x_{h\ h}} = r_h s_x s$, siendo r_h el coe iciente de *orrela i n* en cada estrato.

$$_{h} = r_h = \frac{n_h \Sigma x_{h\ \ h} \quad (\Sigma x_h)(\Sigma\ _{h})}{\sqrt{n_h \Sigma x_h^2\ \ (\Sigma x_h)^2\ \ n_h \Sigma\ ^{2}_{h}\ \ (\Sigma\ _{h})^2}}$$

Directamente la *Covarianza* se calcula: $m_{x_{h\ h}} = \frac{\Sigma x_{h_i\ \ h_i}}{n_h} \quad \bar{x}_h\ \bar{}_h$

$$m_{x_{h\ h}} = \frac{n_h \Sigma x_{h_i\ \ h_i} \quad (\Sigma_{x h_i})(\Sigma\ _{h_i})}{n_h^2}$$

La aplicación de estas fórmulas, trabajando con las variables ingresos (x) y consumo de carne diario (y), es como sigue:

Se obtiene en primer término, la información necesaria de los tres estratos, en la muestra de 36 familias: n_1= 17 n_2= 13 n_3= 6

Estrato I

s_{x_1}=163,13 $s^2_{x_1}$= 26.611,03 $\Sigma x_{1\ \ 1}$= 10.590.720

$\bar{x}_1$= 1.041,18 Σx_1= 17.700 $\bar{}_1$= 595,0 Σx_1^2 = 18.854.600

$s_{\ 1}$= 82,98 $s^2_{\ 1}$= 6.886,5 $\Sigma_{\ 1}$ = 10.115 Σy_1^2 = 6.128.609

$_1$ = 0,1390 r_1 = 0,27 n_1 = 17

Estrato I I

$\bar{x}_2$= 2.011,54 s_{x_2} = 201,07 $s^2_{x_2}$ = 40.430,77 Σx_2= 26.150 Σx_2^2= 53.086.900

$\Sigma x_{2\ \ 2}$= 17.536.670 $\bar{}_2$= 669,85 $s_{\ 2}$= 106,49 $s^2_{\ 2}$= 11.340,47

$\Sigma\ _2$= 8.708 $\Sigma\ ^{2}_{2}$= 530.086.900 $_2$= 0,0416 r_2= 0,0785 n_2= 13

Estrato III

$\bar{x}_3$= 3.181,67 s_{x_3} = 441,88 $s^2_{x_3}$ = 195.256,67 Σx_3= 19.090 Σx_3^2= 61.714.300

$\Sigma x_{3\ \ 3}$= 14.153.280 $\bar{}_3$= 741,67 $s_{\ 3}$ = 40,87 $s^2_{\ 3}$ = 1.670,67

$\Sigma\ _3$= 4.450 $\Sigma\ ^{2}_{3}$= 3.308.770 $_3$= 0,005261 r_3= 0,057 n_3= 6

Cálculo de los *Coeficientes angulares:*

$$_1 = \frac{17(10.590.720) \quad (17.700)(10.115)}{17(18.854.600) \quad (17.700)^2} = 0{,}1390$$

$$_2 = \frac{13(17.536.670) \quad (26.150)(8.708)}{13(53.086.900) \quad (26.150)^2} = 0{,}0416$$

$$_3 = \frac{6(14.153.280) \quad (19.090)(4.450)}{6(61.714.300) \quad (19.090)^2} = \quad 0{,}0053$$

Ahora se obtendrán los valores de $\bar{y}$

$$\bar{\quad}_1 = 595{,}0 + 0{,}1390(1.080 \quad 1.041{,}18) = 600{,}40$$

$$\bar{\quad}_2 = 669{,}85 + 0{,}0416(1.850 \quad 2.011{,}84) = 663{,}13$$

$$\bar{\quad}_3 = 741{,}67 \quad 0{,}0053(3.220 \quad 3.181{,}67) = 741{,}47 \text{ gramos de consumo de carne diario.}$$

El promedio estimado mediante la *regresión lineal separada* será:

$$\bar{y}_{RL_s} = \Sigma W_h \bar{y}_{RL_h} = 0{,}46(600{,}40) + 0{,}37(663{,}13) + 0{,}17(741{,}47) = 647{,}59 \text{ gramos de consumo.}$$

Para hallar el estimador de la varianza, para los límites de confianza, del 95%, se debe calcular en primer lugar, la *covarianza* para cada estrato:

$$m_{x_1y_1} = \frac{10.590.720}{17} \quad 1.041{,}18(595{,}0) = 3.481{,}43$$

$$m_{x_2y_2} = \frac{17.536.670}{13} \quad (2.011{,}54)(669{,}85) = 1.544{,}55$$

$$m_{x_3y_3} = \frac{14.153.280}{6} \quad (3.181{,}67)(741{,}67) = \quad 869{,}19$$

Ahora, el estimador de la varianza para el cálculo del error de estimación se obtiene mediante la fórmula:

$$\hat{V}_{y_{RL_s}} = \Sigma \frac{W_h^2(1 \quad f_h)}{n_h} \quad s_{y_h}^2 \quad 2b_h m_{x_h y_h} + b_h^2 s_{x_h}^2$$

$$\hat{V}_{\bar{y}_{RL_s}} = \frac{0{,}46^2(1 \quad 0{,}1)}{17} \quad 6.886{,}5 \quad 2(0{,}1390)(3.481{,}43) + 0{,}1390^2(26.611{,}03) \quad + \frac{0{,}37^2 \quad (1 \quad 0{,}8)}{13}$$

$$11.340{,}47 \quad 2(0{,}0416)(1.544{,}55) + 0{,}0416^2(40.430{,}77) \quad + \frac{0{,}17^2 \quad (1 \quad 0{,}10)}{6}$$

$$1.670{,}67 \quad 2(\ 0{,}0053)(\ 869{,}19) + (\ 0{,}0053)^2(195.256{,}67) \quad = 188{,}59$$

Los límites de confianza del 95% para el promedio de consumo de carne diario (grs)

$$\hat{\bar{Y}}_{RL_s} = 649{,}59 \quad 2{,}035\sqrt{188{,}59} \begin{cases} \nearrow 677{,}54 \\ \searrow 621{,}64 \end{cases}$$

Consumo promedio de carne diario en gramos.

Mediante el empleo del método de la *regresión lineal combinada*, el promedio estimado de consumo de carne diario (en gramos) por familia es:

$$\bar{y}_{st} = \Sigma W_h \bar{y}_h \qquad \bar{y}_{st} = 0{,}46(595{,}0) + 0{,}37(669{,}85) + 0{,}17(741{,}67) = 647{,}62$$

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h \qquad \bar{x}_{st} = 0{,}46(1.041{,}18) + 0{,}37(2.011{,}84) + 0{,}17(3.181{,}67) = 1.764{,}21$$

El *coeficiente angular combinado* se obtiene:

$$b_c = \Sigma W_h b_h \qquad b_c = 0{,}46(0{,}1390) + 0{,}37(0{,}0416) + 0{,}17(\ 0{,}0053) = 0{,}0784$$

Siendo:

$$\hat{\bar{Y}}_{RL_c} = \bar{y}_{st} + b_c\,(\bar{X}_{st} - \bar{x}_{st}) \qquad \hat{\bar{Y}}_{RL_c} = 647{,}62 + 0{,}0784(1.729 \quad 1.764{,}21) = 644{,}85$$

El estimador de la varianza para el error de estimación será igual a:

$$\hat{V}_{\bar{y}_{RL_c}} = \Sigma \frac{W_h^2(1\ f_h)}{n_h} \quad s_{y_h}^2 \quad 2b_c m_{x_h y_h} + b_c^2 s_{x_h}^2$$

$$\hat{V}_{\bar{y}_{RL_c}} = \frac{0{,}46^2\ (1\ 0{,}1)}{17} \quad 6.886{,}5 \quad 2(0{,}0784)(3.481{,}43) + 0{,}0784^2(26.611{,}03) + \frac{0{,}37^2\ (1\ 0{,}8)}{13}$$

$$11.340{,}47 \quad 2(0{,}0784)(1.544{,}55) + 0{,}0784^2 + (40.430{,}77 + \frac{0{,}17^2\ (1\ 0{,}10)}{6}$$

$$1.670{,}67 \quad 2(0{,}0784)(\ 869{,}19) + (0{,}0784)^2(195.256{,}67) = 192{,}92$$

$$\hat{\bar{Y}}_{RL_c} = 644{,}85 \quad 2{,}035\sqrt{192{,}92} \begin{cases} \nearrow 673{,}12 \\ \searrow 616{,}58 \end{cases}$$

Consumo promedio de carne diario en gramos.

Como en el caso de la *asignación igual*, se pueden realizar estimativos indirectos para el total; en este caso, consideraremos el *total* consumo de carne diario (en gramos) para las 355 familias del barrio investigado.

$$\hat{Y}_{\substack{RLs_s \\ i}} = N\bar{y}_{RL_s} \quad tN\sqrt{\hat{V}_{\bar{y}_{RLs}}} \qquad \hat{Y}_{\substack{RLs_s \\ i}} = 355(649{,}59) \quad 2{,}035(355)\sqrt{188{,}59} \begin{cases} \nearrow 240.525{,}32 \\ \searrow 220.683{,}53 \end{cases}$$

$$\hat{Y}_{\substack{RLc_s \\ i}} = N\bar{y}_{RL_c} \quad tN\sqrt{\hat{V}_{\bar{y}_{RLc}}} \qquad \hat{Y}_{\substack{RLc_s \\ i}} = 355(644{,}85) \quad 2{,}035(355)\sqrt{192{,}92} \begin{cases} \nearrow 238.955{,}92 \\ \searrow 218.887{,}58 \end{cases}$$

ASIGNACIÓN ÓPTIMA

El cálculo de la muestra dependerá en muchos casos del *costo total* de la investigación (C), del *costo fijo* (C_o) y del *costo unitario* (C_h); este último podrá ser *fijo* o *variable*, para cada uno de los estratos, además, el *grado de variabilidad* (S^2). En conjunto, los anteriores factores determinarán con mayor precisión el tamaño óptimo de la muestra; sin embargo, además de los factores anteriores, podrán considerarse como elementos complementarios, el *margen de error y el grado de confianza.*

Las fórmulas, dependerán de los factores que se disponga o sean considerados para determinar el cálculo del tamaño muestral, con el fin de lograr una mejor asignación de unidades dentro de la muestra, en busca de mayor precisión que en la práctica, algunas veces, es sacrificada, mediante la aplicación de otros métodos más simples, pero que arrojan resultados comparativamente tan seguros como los obtenidos mediante la *Asignación óptima*.

En síntesis, el método de Asignación o Afijación óptima se reduce:

a. A una óptima distribución del tamaño de la muestra entre los diferentes estratos, de tal manera que el error de estimación sea mínimo para un costo total dado.
b. Además, un costo total mínimo para un error preestablecido, en algunos casos, se aplica el grado de variabilidad en cada estrato, combinado con los costos unitarios (fijos o variables) y los tamaños poblacionales de cada estrato.

Tamaño de la Muestra

Se usan varias fórmulas, de acuerdo a los factores que se tomen en cuenta para su obtención. Se verán algunos de tales procedimientos:

a. En el caso de que tan solo se consideren como factores para el cálculo de n, el costo total de la investigación (C), el costo fijo (Co) y el costo unitario por estrato (Ch), se debe aplicar la siguiente fórmula:

$$n = \frac{C - C_o}{C_h}$$

Por ejemplo, si se tiene: C = $ 350.000 C_o= 250.000 C_h= $ 4.000

El tamaño de la muestra será:

$$n = \frac{350.000 \quad 250.000}{4.000} = 25 \text{ amilias de una población de 355.}$$

En esta fórmula no se tienen en cuenta componentes o factores tan importantes como el grado de variabilidad (varianza) de la característica en cada estrato, tamaños de los estratos, nivel de confianza y nivel de error.

Como en este ejemplo, el costo unitario (C_h) es constante para cada uno de los estratos, los respectivos tamaños muestrales para el promedio se obtendrán así:

$$n_h = n\left(\frac{N_h S_h}{\Sigma N_h S_h}\right) \quad \text{ó} \quad n_h = n\left(\frac{W_h S_h}{\Sigma W_h S_h}\right)$$

En el caso de *proporciones*, las fórmulas a aplicar serán:

$$n_h = n\left(\frac{N_h \sqrt{P_h Q_h}}{\Sigma N_h \sqrt{P_h Q_h}}\right) \quad \text{ó} \quad n_h = n\left(\frac{W_h \sqrt{P_h Q_h}}{\Sigma W_h \sqrt{P_h Q_h}}\right)$$

Si consideramos los resultados obtenidos a través de la *encuesta preliminar* de las 15 familias, mediante el método de *Asignación proporcional* a fin de no repetir el proceso de selección y cálculo, la varianza de los ingresos y las proporciones de familias con vivienda en cada estrato, serán:

$s_1^2 = 83.923,81 \qquad s_2^2 = 41.470 \qquad s_3^2 = 7.300$

Las proporciones de familias con vivienda: $p_1 = 0,43 \qquad p_2 = 0,8 \qquad p_3 = 0,33$

Siendo los tamaños en cada estrato: $N_1 = 162 \quad N_2 = 132 \quad N_3 = 61$ se tendrá que:

$$n_1 = 25\ \frac{162\ (\sqrt{83.923,81})}{162\sqrt{83.923,81} + 132\sqrt{41.470} + 61\sqrt{7.300}} = 25\ \frac{46.930,76}{46.930,76 + 26.880,72 + 5.211,98} = 15 \text{ familias}$$

$$n_2 = 25\ \frac{26.880,72}{79.023,46} = 8,5 \cong 9 \text{ familias} \qquad n_3 = 25\ \frac{5.211,98}{79.023,46} = 1,65 \cong 2 \text{ familias}$$

$n = n_1 + n_2 + n_3 = 15 + 9 + 2 = 26$ familias

En las *proporciones*, utilizando como ponderaciones a W_h se tendrá que:

$$n_1 = 25\ \frac{0,46\ (\sqrt{0,43(0,57)})}{0,46\ \sqrt{0,43(0,57)} + 0,37\ \sqrt{0,8(0,2)} + 0,17\ \sqrt{0,33(0,67)}} = 25\ \frac{0,2277}{0,2277 + 0,148 + 0,0799} = 15 \text{ familias}$$

$$n_2 = 25\ \frac{0,148}{0,4556} \cong 8 \text{ familias} \qquad n_3 = 25\ \frac{0,0799}{0,4556} \cong 5 \text{ familias}$$

$n = 13 + 8 + 5 = 26$ familias

b. Otra fórmula para determinar el tamaño óptimo de la muestra, disponiendo de información sobre el *costo total de la investigación* (C), *costo fijo* (C_o), *grado de variabilidad de cada estrato* (S_h), *costos unitarios* (C_h) que pueden ser constantes o variables para cada estrato y el tamaño poblacional de cada estrato (N_h).

Los anteriores factores permiten determinar el tamaño n con mayor representatividad que la obtenida con la fórmula anterior, ya que además de los costos, se tienen en cuenta los tamaños poblacionales de los estratos, y el grado de variabilidad para cada uno de ellos.

$$n = \frac{(C - C_o)\left(\Sigma \frac{N_h S_h}{\sqrt{C_h}}\right)}{\Sigma N_h S_h \sqrt{C_h}}$$

Si consideramos la siguiente información tomada de la *encuesta preliminar; Asignación proporcional*

s_1= 289,70 s_2= 203,64 s_3= 85,44 N_1= 162 N_2= 132 N_3= 61

C_1= 4.000 C_2= 4.000 C_3= 4.000 C= 350.000 C_o= 250.000

Se tendrá que n será igual a:

$$n = \frac{350.000 \quad 250.000 \left(\frac{162(289,70)}{\sqrt{4.000}} + \frac{132(203,64)}{\sqrt{4.000}} + \frac{61(85,44)}{\sqrt{4.000}} \right)}{162(289,7)\sqrt{4.000} + 132(203,64)\sqrt{4.000} + 61(85,44)\sqrt{4.000}} = 24 \text{ amilias}$$

Puede observarse que hemos utilizado costos unitarios (C_h) constantes, en la fórmula anterior.

Coincidencialmente se obtuvo un resultado igual a la anterior fórmula; sin embargo el valor de n puede ser mayor o menor, dependiendo del grado de variabilidad y tamaño de cada estrato. Supongamos ahora que los costos unitarios (C_h) para cada estrato son variables:

C_1= 2.000 C_2= 3.000 C_3= 4.000

Por lo tanto el valor de n será:

$$n = \frac{350.000 \quad 250.000 \left(\frac{162(289,70)}{\sqrt{2.000}} + \frac{132(203,64)}{\sqrt{3.000}} + \frac{61(85,44)}{\sqrt{4.000}} \right)}{162(289,7)\sqrt{2.000} + 132(203,64)\sqrt{3.000} + 61(85,44)\sqrt{4.000}} = 42 \text{ amilias}$$

(Nota: se trabajó en la calculadora con el programa de estadística).

En el caso de que C_h sea constante, los n_h se obtienen mediante las órmulas:

$$n_h = n\left(\frac{N_h S_h}{\Sigma N_h S_h}\right) \quad \text{ó} \quad n_h = n\left(\frac{W_h S_h}{\Sigma W_h S_h}\right)$$

Para cuando C_h es variable para cada uno de los estratos, los n_h se calcularán así:

$$n_h = n\left[\frac{\frac{N_h S_h}{\sqrt{c_h}}}{\Sigma\frac{N_h S_h}{\sqrt{c_h}}}\right] \quad \text{ó} \quad n_h = n\left[\frac{\frac{W_h S_h}{\sqrt{c_h}}}{\Sigma\frac{W_h S_h}{\sqrt{c_h}}}\right]$$

Siendo: n_1 n_2 n_3 los tama os muestrales en cada estrato, estos serán iguales a:

$$n_1 = 42\left[\frac{\frac{162(289,7)}{\sqrt{2.000}}}{\frac{162(289,7)}{\sqrt{2.000}} + \frac{132(203,64)}{\sqrt{3.000}} + \frac{61(85,49)}{\sqrt{4.000}}}\right] = 27,16 \cong 28 \text{ amilias}$$

$$n_2 = 42\left[\frac{490{,}77}{1.622{,}64}\right] = 12{,}71 \cong 13 \text{ amilias} \qquad n_3 = 42\left[\frac{82{,}45}{1.622{,}64}\right] = 2{,}13 \cong 3 \text{ amilia}$$

$$n = 27 + 13 + 3 = 43 \text{ amilias}$$

Si la característica es cualitativa, utilizamos proporciones como en el caso de estimar las *proporciones de familias propietarias de vivienda*. El procedimiento es similar al anterior, tan solo se reemplaza a s_x^2 por $s_p^2 = pq$ y se tendrá que:

$$n = \frac{(C - C_o)\left(\Sigma N_h \sqrt{\frac{P_h Q_h}{C_h}}\right)}{\Sigma N_h \sqrt{P_h Q_h}\sqrt{C_h}}$$

Siendo:

$$n = \frac{350.000 \quad 250.000\left(162\sqrt{\frac{0{,}43(0{,}57)}{4.000}} + 132\sqrt{\frac{0{,}8(0{,}20)}{4.000}} + 61\sqrt{\frac{0{,}33(0{,}67)}{4.000}}\right)}{162\sqrt{0{,}43(0{,}57)}\sqrt{4.000} + 132\sqrt{0{,}8(0{,}2)}\sqrt{4.000} + 61\sqrt{0{,}33(0{,}67}\sqrt{4.000}} = 25 \text{ amilias}$$

Si los C_h varían para cada estrato, se tendrá:

$$n = \frac{350.000 \quad 250.000\left(162\sqrt{\frac{0{,}43(0{,}57)}{2.000}} + 132\sqrt{\frac{0{,}8(0{,}20)}{3.000}} + 61\sqrt{\frac{0{,}33(0{,}67)}{4.000}}\right)}{162\sqrt{0{,}43(0{,}57)}\sqrt{2.000} + 132\sqrt{0{,}8(0{,}2)}\sqrt{3.000} + 61\sqrt{0{,}33(0{,}67}\sqrt{4.000}} = 39 \text{ amilias}$$

Para calcular los *tamaños muestrales* en cada estrato, cuando C_h es variable se procederá así:

$$n_h = n\left[\frac{N_h\sqrt{\frac{P_h Q_h}{C_h}}}{\Sigma N_h\sqrt{\frac{P_h Q_h}{C_h}}}\right] \quad \text{ó} \quad n_h = n\left[\frac{W_h\sqrt{\frac{P_h Q_h}{C_h}}}{\Sigma W_h\sqrt{\frac{P_h Q_h}{C_h}}}\right]$$

Reemplazando se tendrá:

$$n_1 = 39\left[\frac{0{,}46\sqrt{\frac{0{,}43(0{,}57)}{2.000}}}{0{,}46\sqrt{\frac{0{,}43(0{,}57)}{2.000}} + 0{,}37\sqrt{\frac{0{,}8(0{,}20)}{2.000}} + 0{,}17\sqrt{\frac{0{,}33(0{,}67)}{4.000}}}\right] = 22 \text{ amilias}$$

$$n_2 = 39\left[\frac{0{,}00270}{0{,}00905}\right] \cong 12 \text{ amilias} \qquad n_3 = 39\left[\frac{0{,}00126}{0{,}00905}\right] \cong 6 \text{ amilias}$$

El incremento que presenta el valor de n se debe a las aproximaciones, siendo preferible en numerosas ocasiones tomar una unidad de más, que una de menos.

c. El tamaño óptimo de la muestra, en el *muestreo aleatorio estratificado*, por el método de *asignación óptima*, también se puede obtener utilizando el *nivel de confianza, el error de muestreo y el costo unitario* (C_h) para cada estrato, mediante la fórmula:

$$n = \frac{\left[\Sigma W_h S_h \sqrt{C_h}\right]\left[\sum \frac{W_h S_h}{\sqrt{C_h}}\right]}{+\frac{1}{N}\Sigma W_h S_h^2} \qquad = \left(\frac{E}{}\right)^2$$

Dado que los C_h pueden ser constantes o variables, utilizaremos la fórmula y los datos anteriores, con una confianza del 95,5% (Z = 2) y un error del 5% para el promedio y del 10% para el atributo, para calcular n. Veámos cuando C_h es constante.

$$n = \frac{0{,}46(289{,}7)\sqrt{4.000}+0{,}37(203{,}64)\sqrt{4.000}+0{,}17(85{,}44)\sqrt{4.000}\left[\frac{0{,}46(289{,}7)}{\sqrt{4.000}}+\frac{0{,}37(203{,}64)}{\sqrt{4.000}}+\frac{0{,}17(85{,}44)}{\sqrt{4.000}}\right]}{\left(\frac{88{,}41}{2}\right)^2+\frac{1}{355}\ 0{,}46(83.926{,}09)+0{,}37(41.469{,}25)+0{,}17(7.300)} = 24 \text{ amilias}$$

(Nota: se trabajó en la calculadora con el programa de estadística).

El anterior resultado de 24 familias corresponde al tamaño óptimo para calcular el estimador del promedio de ingresos por familia. Veamos ahora que sucede, cuando C_h es variable:

$$n = \frac{0{,}46(289{,}7)\sqrt{2.000}+0{,}37(203{,}64)\sqrt{3.000}+0{,}17(85{,}44)\sqrt{4.000}\left[\frac{0{,}46(289{,}7)}{\sqrt{2.000}}+\frac{0{,}37(203{,}64)}{\sqrt{3.000}}+\frac{0{,}17(85{,}44)}{\sqrt{4.000}}\right]}{\left(\frac{88{,}41}{2}\right)^2+\frac{1}{355}\ 0{,}46(83.926{,}09)+0{,}37(41.469{,}25)+0{,}17(7.300)} = 25 \text{ amilias}$$

En el caso de *estimativos para proporciones*, la fórmula aplicada para determinar el tamaño óptimo de la muestra corresponde a:

$$n = \frac{\left[\Sigma W_h \sqrt{P_h Q_h}\sqrt{C_h}\right]\left[\Sigma W_h \sqrt{\frac{P_h Q_h}{C_h}}\right]}{+\frac{1}{N}\ \Sigma W_h P_h Q_h} \qquad = \left(\frac{E}{}\right)^2$$

En la anterior fórmula como en el caso anterior C_h puede ser constante o variable. Con los datos que hemos obtenido a través de la encuesta preliminar, reemplazamos en la fórmula anterior, cuando C_h es constante.

$$n = \frac{0{,}46\sqrt{0{,}43(0{,}57)}\sqrt{4.000}+0{,}37\sqrt{0{,}8(0{,}2)}\sqrt{4.000}+0{,}17\sqrt{0{,}33(0{,}67)}\sqrt{4.000}}{\left(\frac{0{,}10}{2}\right)^2+\frac{1}{355}\ 0{,}46(0{,}43)(0{,}57)+0{,}37(0{,}80)(0{,}20)+}$$

(contin a)

$$\frac{\left[0{,}46\sqrt{\frac{0{,}43(0{,}57)}{4.000}}+0{,}37\sqrt{\frac{0{,}8(0{,}20)}{4.000}}+0{,}17\sqrt{\frac{0{,}33(0{,}67)}{4.000}}\right]}{+\ 0{,}17(0{,}33)(0{,}67)} = 67 \text{ amilias}$$

En el caso de que C_h sea *variable* se tendrá:

$$n = \frac{0{,}46\sqrt{0{,}43(0{,}57)}\sqrt{2.000} + 0{,}37\sqrt{0{,}8(0{,}2)}\sqrt{3.000} + 0{,}17\sqrt{0{,}33(0{,}67)}\sqrt{4.000}}{\left(\frac{0{,}10}{2}\right)^2 + \frac{1}{355}\ 0{,}46(0{,}43)(0{,}57)+0{,}37(0{,}80)(0{,}20) + \left[0{,}46\sqrt{\frac{0{,}43(0{,}57)}{2.000}} + 0{,}37\sqrt{\frac{0{,}8(0{,}20)}{3.000}} + 0{,}17\sqrt{\frac{0{,}33(0{,}67)}{4.000}}\right]} \quad \text{(contin a)}$$

$$+ 0{,}17(0{,}33)(0{,}67)$$

$$= 68 \text{ amilias}$$

Es necesario anotar que en esta fórmula se puede prescindir de los costos unitarios (C_h) cuando éstos son constantes para cada estrato y, por tal razón, se obtiene una fórmula más simplificada, que nos permite calcular el tamaño de la muestra en una forma más sencilla, veámosla:

a. $$n = \frac{(\Sigma W_h S_h)^2}{+\frac{\Sigma W_h S_h}{N}}$$

b. $$n = \frac{\left(\Sigma W_h \sqrt{P_h Q_h}\right)^2}{+\frac{\Sigma W_h P_h Q_h}{N}}$$

$$V = \left(\frac{E}{Z}\right)^2$$

Reemplazando los datos que ya tenemos, en las fórmulas anteriores se obtiene:

a. $$n = \frac{0{,}46(289{,}7) + 0{,}37(203{,}64) + 0{,}17(85{,}44)^{\ 2}}{\left(\frac{88{,}41}{2}\right)^2 + \left[\frac{0{,}46(83.926{,}09) + 0{,}37(41.469{,}25) + 0{,}17(7.300)}{355}\right]} = 24 \text{ amilias}$$
(variable ingresos)

b. $$n = \frac{\left[0{,}46\sqrt{0{,}43(0{,}57)} + 0{,}37\sqrt{0{,}8(0{,}2)} + 0{,}17\sqrt{0{,}33(0{,}67)}\right]^2}{\left(\frac{0{,}10}{2}\right)^2 + \left[\frac{0{,}46(0{,}43)(0{,}57) + 0{,}37(0{,}8)(0{,}2) + 0{,}17(0{,}33)(0{,}67)}{355}\right]} = 68 \text{ amilias}$$
(proporción propietarios)

Las diferencias obtenidas con la aplicación de estas fórmulas, en relación con aquéllas que tienen en cuenta los *costos unitarios* (C_h), constantes, son pequeñas.

En estos dos casos, los *tamaños muestrales* para cada uno de los estratos, se pueden calcular mediante las fórmulas:

$$n_h = \left[\frac{N_h S_h}{\Sigma N_h S_h}\right] \quad \text{ó} \quad n_h = \left[\frac{W_h S_h}{\Sigma W_h S_h}\right]$$

Las anteriores fórmulas aplican cuando se trata de la *estimación de una variable.* En el caso de trabajar con *atributos*, las fórmulas a aplicar son:

$$n_h = \left[\frac{N_h \sqrt{P_h Q_h}}{\Sigma N_h \sqrt{P_h Q_h}}\right] \quad \text{ó} \quad n_h = \left[\frac{W_h \sqrt{P_h Q_h}}{\Sigma W_h \sqrt{P_h Q_h}}\right]$$

MÉTODO DE NEYMAN

Dentro de las técnicas utilizadas en el *Muestreo Aleatorio Estratificado* para hallar n, puede efectuarse el cálculo en forma independiente para cada uno de los estratos muestrales; siendo conocido como el *Método de Neyman.*

$$n_{o_h} = \frac{W_h S_h^2}{V_h} = \frac{W_h S_h^2}{\left(\frac{E_h}{z_h}\right)^2} \qquad \text{Siendo:} \qquad n_h = \frac{n_{oh}}{1 + \frac{n_{oh}}{N_h}}$$

Es posible en muchos casos, establecer el error para cada uno de los estratos. El procedimiento a seguir es: con los datos obtenidos en la *encuesta preliminar* de 15 observaciones, siendo

$n_1 = 7$ $\quad n_2 = 5$ $\quad n_3 = 3$ (*asignación proporcional*).

$$n_{o1} = \frac{W_1 S_1^2}{(E_1/z_1)^2} = \frac{0{,}46(83.926{,}09)}{(57{,}36/1.96)^2} = 45{,}08 \cong 46 \text{ familias}$$

$E_1 = 0{,}05(1.147{,}14) = 57{,}36$
$z_1 = 1{,}96$

$$n_1 = \frac{n_{01}}{1 + \frac{n_{01}}{N_1}} = \frac{45{,}08}{1 + \frac{45{,}08}{162}} = 36 \text{ familias}$$

$$n_{o2} = \frac{W_2 S_2^2}{(E_2/z_2)^2} = \frac{0{,}37(41.469{,}25)}{(99{,}4/1{,}96)^2} = 5{,}97$$

$E_2 = 0{,}05(1988) = 99{,}4$
$z_2 = 1{,}96$

$$n_2 = \frac{5{,}97}{1 + \frac{5{,}97}{132}} = 5{,}71 \cong 6 \text{ familias}$$

$$n_{o3} = \frac{W_3 S_3^2}{(E_3/z_3)^2} = \frac{0{,}17(7.300)}{(148{,}5/1{,}96)^2} = 0{,}22$$

$E_3 = 0{,}05(2.970) = 148{,}5$
$z_3 = 1{,}96$

$$n_3 = \frac{0{,}22}{1 + \frac{0{,}22}{61}} \cong 1 \text{ familia} \qquad n = n_1 + n_2 + n_3 = 36 + 6 + 1 = 43 \text{ familias}$$

Algo similar ocurre para el cálculo de n cuando se desea hacer la *estimación de una proporción*.

$$n_{o_h} = \frac{W_h P_h Q_h}{(E_h/z_h)^2}$$

$E_1 = 0{,}10$ $\quad E_2 = 0{,}10$ $\quad E_3 = 0{,}10$

$$n_{o_1} = \frac{0{,}46(0{,}43)(0{,}57)}{(0{,}10/1{,}96)^2} = 43{,}31 \qquad n_{o_2} = \frac{0{,}37(0{,}8)(0{,}2)}{(0{,}10/1{,}96)^2} = 22{,}74$$

$$n_{o_3} = \frac{0{,}17(0{,}33)(0{,}67)}{(0{,}10/1{,}96)^2} = 22{,}74$$

$$n_1 = \frac{43{,}31}{1+\frac{43{,}31}{162}} \cong 35 \qquad n_2 = \frac{22{,}74}{1+\frac{22{,}74}{132}} \cong 20 \qquad n_3 = \frac{14{,}43}{1+\frac{14{,}43}{61}} \cong 13 \text{ familias}$$

$$n = n_1 + n_2 + n_3 = 35 + 20 + 13 = 68 \text{ familias}$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

87. Con los datos de una población:

ESTRATO	N_h	S_h
I	285	28
II	415	37
Total	700	–

Calcular el tamaño de la muestra en el estrato I, para un nivel de confianza del 95,5% y un desvío máximo tolerado de $E_1 = 11$. Análogamente para el estrato II, con $Z = 1 \Rightarrow (63{,}8\%)$ y $E_2 = 12{,}5$.

88. Calcular los tamaños de muestra para los estratos, aplicando la afijación proporcional, igual y óptima, tomando como base el tamaño total de la muestra, obtenido en el punto anterior.

89. Con la siguiente información:

ESTRATO	W_h	S_h	C_h	$\bar{y}_h$	N_h
I	0,5	4	1	10	200
II	0,3	5	4	20	120
III	0,2	6	9	32	80

a) Determinar el tamaño de la muestra, por el método de asignación proporcional y óptima. Error del 5% y nivel de confianza del 95%.

b) Si además de la información dada, se tiene que C = \$10.000 y Co = \$9.500, determinar el tamaño de la muestra, por el método de asignación óptima, siendo $n = 148$.

90. Con los datos del ejercicio anterior se pide fijar límites de confianza del 95%, para el estimador del promedio y del total, por los métodos de asignación proporcional y óptima, siendo $n = 177$.

91. Con la siguiente muestra obtenida de una población estratificada (asignación proporcional).

ESTRATO I			ESTRATO II			ESTRATO III		
No. de orden	y_1	x_1	No. de orden	y_2	x_2	No. de orden	y_3	x_3
1	3	2	1	10	5	1	12	8
2	1	1	2	9	4	2	18	10
3	7	3	3	8	6	3	14	8
4	2	1				4	15	10
5	5	2				5	13	7
6	5	3						
7	6	3						

$N_1 = 190$ $N_2 = 80$ $N_3 = 130$ $N = 400$

Se pide:

Considerando que los datos fueron obtenidos mediante una encuesta preliminar, determinar el tamaño de la muestra (proporcional), con nivel de confianza del 95,5% y error del 5%, para las dos variables.

92. Con los datos del ejercicio anterior, suponiendo que el tamaño de la muestra es 15 y los resultados son los del cuadro anterior, se pide estimar:
a) El promedio y el total para la variable X.
b) Mediante razón combinada y separada, estimar el promedio para la variable Y.
c) El promedio para la variable Y, mediante el método de la regresión (combinada y separada).
Información adicional $\bar{X}_1 = 1,6$ $\bar{X}_2 = 5,2$ $\bar{X}_3 = 9,6$ $\bar{X}_{st} = 7,3$

93. En una población cuyas características son conocidas:

ESTRATO	W_h	S_h	C_h	$\bar{y}_h$	P_h
I	0,45	4	2	16	0,62
II	0,35	5	4	21	0,54
III	0,20	6	9	30	0,36

Además: $E = 2,5$ $Z = 1,96$ $N = 200$ $E = 12\%$

Considere la información anterior, para determinar el tamaño de la muestra, mediante: Asignación igual, proporcional y óptima. Obtener los tamaños muestrales para cada estrato con cada uno de los métodos anteriormente dados, tanto para la variable (Y), como para la proporción.

94. Con los datos del ejercicio anterior, suponga que $E_1 = 0,8$ $E_2 = 1,02$ y $E_3 = 2$ para determinar los tamaños de la muestra, de acuerdo al método de Asignación de Neyman.

95. Con los datos del ejercicio 93, fije los límites de confianza del 95% para el estimador del promedio y de la proporción, mediante los métodos de asignación igual y óptima.

96. Insatisfacción de los empleados en su trabajo, puede causar un aumento en los costos de una empresa, debido a la baja calidad de la mano de obra y al ausentismo. El gerente está interesado en determinar el número de días laborales perdidos por ausencias, incluyendo los de enfermedad y accidentes de trabajo. Por razones de tipo administrativo, se usó el método de muestreo aleatorio estratificado con afijación proporcional para seleccionar una muestra de un total de 700 trabajadores que tiene la empresa. Los estratos estaban representados por grupos de obreros, técnicos y personal administrativo.

Supongamos que la encuesta preliminar arroja 15 obreros, 10 técnicos y 2 administradores, como aparecen a continuación:

OBREROS				TÉCNICOS			ADMINISTRATIVOS
8	15	0		4	5		1
0	16	21		0	10		8
6	0	12		8	12		
7	4	3		3	2		
9	5	2		1	8		

Determine el tamaño de la muestra, fijando límites con una confianza del 95% y error del 8%.

97. Suponga que la información anterior es el resultado de una muestra, con la cual se pide estimar el total de días perdidos por los empleados de la empresa.

98. Con los siguientes datos:

Estrato I

NÚMERO DE ORDEN	NÚMERO ALEATORIO	PERSONAS		SERVICIOS		INGRESOS MENSUALES (Miles $)
		Hombres	Mujeres	Luz	Agua	
1	70	2	1	no	no	4.500
2	08	1	1	no	si	6.000
3	36	0	2	no	si	5.000
4	42	3	1	si	si	6.000
5	28	2	2	si	no	7.000
6	16	1	3	no	no	5.000
7	22	1	3	si	si	4.500
8	55	1	2	si	si	4.800

Estrato II

NÚMERO DE ORDEN	NÚMERO ALEATORIO	PERSONAS		SERVICIOS		INGRESOS MENSUALES (Miles $)
		Hombres	Mujeres	Luz	Agua	
1	28	4	3	si	si	6.000
2	33	2	6	si	si	5.800
3	42	3	5	si	no	7.200
4	06	4	4	no	no	6.000
5	01	4	2	si	si	4.800
6	19	3	4	no	si	6.000
7	22	2	5	no	si	5.700
8	38	4	2	si	si	6.400

Estrato III

NÚMERO DE ORDEN	NÚMERO ALEATORIO	PERSONAS		SERVICIOS		INGRESOS MENSUALES (Miles $)
		Hombres	Mujeres	Luz	Agua	
1	15	3	7	sí	sí	4.600
2	20	5	4	no	no	5.200
3	16	6	3	no	no	6.000
4	09	5	7	no	no	7.000
5	12	2	8	sí	sí	6.800
6	18	6	5	sí	no	5.900
7	25	4	8	no	sí	7.000
8	14	5	5	no	sí	6.200

Información complementaria

N = 150 viviendas $N_1 = 75$ $N_2 = 50$ $N_3 = 25$ Z = 2 (P = 95%)
E = 5% (promedio) E = 10% (proporción)

Estrato I : Vivienda de menos de 5 personas

Estrato II : Vivienda de 5 a 8 personas

Estrato III : Viviendas de 9 y más personas

Se pide: suponga una n (encuesta piloto) de 24 viviendas para determinar el tamaño de la muestra mediante la afijación igual para:

a) El ingreso por familiar b) Proporción de vivienda con servicio de luz.

99. Considere que el tamaño de la muestra obtenida en el punto (a) es de 24 y con la información obtenida calcule: los límites de confianza del 95% para el estimativo total de los ingresos de las familias, aplicando la Afijación igual.

100. Con los datos del ejercicio 98 estime el total de viviendas con servicio de agua, trabajando con una confianza del 95%, aplicando Afijación igual.

101. Con los datos del ejercicio 98 considere a n = 24 viviendas como si fuera una sola muestra sin estratificar (M.A.S.) y estime mediante el cálculo la razón de hombres y mujeres con una confianza del 95%.

102. Se tiene el listado de consumo de carne (gramos) de las familias residentes en un barrio de Bogotá.

Se quiere determinar el número de familias que se deben entrevistar, cuyo objetivo es el de estimar la venta de carne en dicho barrio.

Calcular el tamaño de la muestra con un error de 20 gramos y una seguridad del 90%. No se conoce la varianza poblacional.

CONSUMO DIARIO DE CARNE

No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.
1	789	34	782	67	712	100	748	133	746	166	744	199	859	232	742	265	847	298	779
2	807	35	713	68	792	101	863	134	761	167	721	200	763	233	721	266	792	299	746
3	802	36	787	69	782	102	740	135	729	168	790	201	707	234	744	267	785	300	741
4	765	37	736	70	793	103	734	136	775	169	758	202	728	235	735	268	862	301	765
5	735	38	751	71	792	104	780	137	812	170	711	203	811	236	777	269	781	302	796
6	895	39	765	72	754	105	755	138	816	171	792	204	921	237	854	270	774	303	910
7	799	40	815	73	692	106	851	139	767	172	761	205	744	238	788	271	749	304	850
8	749	41	762	74	815	107	763	140	745	173	825	206	808	239	706	272	748	305	736
9	742	42	737	75	832	108	749	141	757	174	734	207	752	240	707	273	795	306	778
10	892	43	733	76	810	109	902	142	730	175	827	208	760	241	722	274	804	307	772
11	864	44	738	77	714	110	747	143	799	176	803	209	733	242	749	275	782	308	798

No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.	No.	Grms.
12	772	45	812	78	804	111	782	144	778	177	776	210	751	243	735	276	748	309	798
13	808	46	739	79	698	112	823	145	834	178	793	211	819	244	844	277	813	310	726
14	732	47	786	80	762	113	797	146	873	179	842	212	755	245	783	278	791	311	753
15	743	48	843	81	778	114	819	147	812	180	833	213	897	246	858	279	742	312	792
16	824	49	808	82	711	115	750	148	837	181	754	214	786	247	768	280	760	313	741
17	804	50	734	83	715	116	774	149	857	182	760	215	790	248	752	281	776	314	765
18	767	51	760	84	762	117	821	150	829	183	790	216	744	249	755	282	773	315	821
19	760	52	745	85	771	118	797	151	730	184	813	217	744	250	776	283	794	316	811
20	842	53	843	86	724	119	776	152	821	185	721	218	791	251	746	284	717	317	865
21	813	54	803	87	778	120	772	153	788	186	748	219	744	252	764	285	729	318	720
22	753	55	737	88	827	121	820	154	868	187	782	220	776	253	718	286	712	319	738
23	723	56	775	89	713	122	865	155	806	188	761	221	742	254	756	287	766	320	740
24	793	57	719	90	819	123	847	156	739	189	791	222	717	255	786	288	854	321	891
25	749	58	775	91	737	124	723	157	780	190	782	223	772	256	804	289	863	322	744
26	763	59	775	92	839	125	797	158	762	191	810	224	749	257	845	290	835	323	735
27	746	60	781	93	791	126	785	159	762	192	810	225	749	258	845	291	835	324	735
28	861	61	744	94	751	127	719	160	815	193	819	226	738	259	811	292	862	325	779
29	787	62	689	95	782	128	789	161	778	194	785	227	810	260	804	293	787	326	745
30	764	63	801	96	729	129	754	162	729	195	870	228	801	261	775	294	792	327	833
31	757	64	764	97	745	130	828	163	759	196	764	229	773	262	744	295	834	$\Sigma x_i = 254.860$	
32	714	65	796	98	744	131	726	164	798	197	772	230	732	263	850	296	763		
33	852	66	766	99	786	132	751	165	854	198	764	231	867	264	757	297	839	$\Sigma x_i = 199.627.173$	

103. Con los datos del ejercicio 102, se desea hacer una estimación del promedio de consumo de carne, con una confianza del 95% (utilizar el tamaño muestral obtenido en el punto anterior).

104. Con los datos del ejercicio 100 y el tamaño obtenido, se desea hacer una estimación del promedio de consumo de carne, con una confianza del 99%.

105. Resuelva los puntos 103 y 104 considerando que el tamaño de la muestra se ha duplicado (no es necesario tomar una nueva muestra sólo debe duplicar la información obtenida), haga un análisis de los resultados obtenidos.

106. Haga una estimación para el total del consumo de carne en libras (resultados del ejercicio 103).
    a. Diaria.
    b. Semanal, considere que el día domingo se consume una libra más por familia.

107. Las familias consumen en hueso la tercera parte de su consumo en carne. De acuerdo al ejercicio 106b realice una estimación para la venta total en arrobas, incluyendo el consumo de hueso.

108. Con los datos del ejercicio 102, se pide probar la hipótesis, con un nivel de significación del 5%, de que el consumo promedio es de 800 gramos ($\mu = 800$) frente a la alternativa de que:

    a. $\mu = 800$ gramos b. $\mu < 800$ gramos c. $\mu > 800$ gramos

    Si el consumo promedio verdadero es de 779 gramos, ¿qué tipo de error se comete en cada una de las alternativas del ejercicio 106?

109. Con los datos del ejercicio 102 determinar el número de familias que deben ser entrevistadas (n) para hacer una estimación del porcentaje de familias que consumen más de 800 gramos; se desea cometer un error del 10% y se establece una confianza del 90%.

    a. Se considera que P = 0,50

    b. No se conoce el valor de P (recurrir a una encuesta preliminar).

    c. Obtener una estimación y fijar los límites de confianza del 95% para la proporción de la población, que consume más de 800 gramos.

110. Con los resultados del ejercicio 109, considere $\alpha = 0,08$ para probar la hipótesis de que P = 0,4, frente a la alternativa de que:

    a) P = 0, 4 b) P < 0,4 c) P > 0,4

111. Con los resultados obtenidos en el ejercicio 110, determinar qué tipo de error se comete, si se considera que la proporción verdadera es de 0,45.

112. Con los datos del ejercicio 102 se pide:

    Dividir la población en tres estratos

    Estrato I : Menos de 780 gramos

    Estrato II : Desde 780 hasta 830 gramos

    Estrato III : Más de 830 gramos.

    a) Determinar el tamaño de la muestra, utilizando los métodos de Afijación igual y proporcional, con una confianza del 95% y un error del 1%

    b) Fijar los límites del 95% para el promedio y el total de consumo de carne (en libras) en la semana.

113. Determinar el tamaño de la muestra, por el método de Afijación óptima, utilizando los diferentes métodos establecidos. Para ello se da la siguiente información:

Costo total	$200.000	
Costo fijo	$150.000	
Costo fijo unitario	$ 2.000	
Costo variable unitario	$ 1.200	(para el primer estrato)
Costo variable unitario	$ 1.500	(para el segundo estrato)
Costo variable unitario	$ 2.300	(para el tercer estrato)

Confianza del 95%

Error del 6% (20) = 1,20	$Z = 1,96$
$N_1 = 120$	$S^2_{x_1} = 0,69$
$N_2 = 180$	$S_{x_1} = 0,83$
$N_3 = 300$	$S^2_{x_2} = 2,35$
$N = 600$	$S_{x_2} = 1,69$
$W_1 = 0,45$	$S^2_{x_3} = 109,03$
$W_2 = 0,35$	$S_{x_3} = 10,4$
$W_3 = 0,20$	

114. Con la siguiente información poblacional:

Estrato I

No. DE ORDEN:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
X_1:	12,6	13,0	12,8	12,5	13,0	12,9	13,5	14,0

Estrato II

No. DE ORDEN:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
X_2	20,3	25,6	25,2	22,0	26,0	24,8	25,7	25,9	23,1	25,0	24,4	26,5	24,4	25,0

Estrato III

No. DE ORDEN:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
X_3	39,2	38,4	40,0	39,6	38,6	37,5	37,9	37,0	39,8	41,3	37,0	42,0	42,3

No. DE ORDEN:	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
X_3:	39,8	37,6	35,3	34,3	36,2	35,0	32,1	34,4	36,5	38,6	39,2	37,6

Se pide:

a) Extraer una muestra de tamaño nueve ($n = 9$) utilizando el método de muestreo Sistemático.
b) Con el anterior resultado estime la media poblacional, fijando límites del 90%.
c) Estime el total de la población, con una confianza del 90%, aplicando el muestreo Estratificado proporcional.

115. Se tiene la siguiente información:

ESTRATOS	N_h	S_h	$\bar{x}_h$	C_h
I Pequeño	500	3,48	15,3	$20
II Mediano	300	2,80	29,7	30
III Grande	200	4,00	22,8	40

z = 1,96 E = 2,1

Con los datos anteriores, se pide determinar el tamaño de la muestra (n) aplicando el método de Afijación óptima, obteniendo los tamaños muestrales en cada estrato.

116. Considerando la información del cuadro como si fuera de la muestra y los tamaños muestrales del punto anterior, fije los límites de confianza del 95% para la media de la población.

117. Con el punto anterior del tema Asignación igual y óptima del punto 104; considere los resultados del numeral b) y seleccione una muestra de tamaño 8 mediante la Asignación proporcional y estime el valor promedio y el total para X, fijando límites de confianza del 95%.

118. Considerando los datos del punto 104, tema Asignación igual y óptima, determine el tamaño de la muestra mediante el método de Asignación proporcional.

119. Considere que los datos del punto 104, tema Asignación Igual y óptima se obtuvo de una muestra de tamaño 80. Determine el estimado para el promedio y total de la variable X, mediante el método de Asignación óptima, con una confianza del 95%.

120. Con los datos del ejercicio 105, tema Asignación igual y óptima, siendo $C_1 = 1.000$ $C_2 = 1.500$ y $C_3 = \$2.000$, además, la diferencia entre el costo total y el costo fijo es de $3.000. Con los anteriores datos determinar el tamaño de la muestra tanto para la variable X, como para el atributo A.

121. Con los datos del ejercicio 105, tema Asignación igual y óptima, seleccionar una muestra preliminar de nueve (9) elementos, distribuidos proporcionalmente para cada estrato. Además, sabiendo que la confianza es del 95%, el error del 3%, costo total $400.000, costo fijo de $250.000 y costos unitarios de $C_1 = \$3.000$ $C_2 = \$4.000$ y $C_3 = \$5.000$, calcular los tamaños muestrales para cada estrato mediante la asignación óptima (aplicar dos de las fórmulas para calcular n, utilizando la variable Y_i).

122. Una firma industrial tiene 4.000 empleados, de los cuales el 72% son operarios de producción; 21% personal administrativo y servicios generales y un 7% directivos. Se quiere determinar una muestra (n) de esos tres grupos de personas, para estimar cuántas utilizan el restaurante de la empresa para almorzar, aplicando el método de afijación proporcional. Algunos funcionarios antiguos aseguran que el restaurante es utilizado por un 48% de operarios; 21% del personal administrativo y el 4% del personal directivo. Utilizar el error del 5% y una confianza del 95,5% para calcular n.

123. De acuerdo al tamaño de la muestra obtenido en el punto anterior, se realizó la investigación con los siguientes resultados: 72% de operarios; 55% personal administrativo y 15% personal directivo. Con dicha información estime la proporción y el total de empleados de la firma industrial, que toman almuerzo en el restaurante de la empresa, fijando límites del 95% de confianza.

124. Con el resultado anterior, se desea probar la hipótesis de que el 70% del total de empleados toma almuerzo en el restaurante, frente a la alternativa de que sea diferente.

125. La casa principal de una cadena de almacenes está interesada en saber el volumen de ventas, cuentas atrasadas y su valor. Para reducir los costos de muestreo, se utilizó la afijación óptima.

ALMACÉN / ESTRATOS	I	II	III	IV
No. clientes: (población)	1.700	2.300	4.600	3.400
Volumen ventas: (millones) (muestra)	800	1.300	2.200	1.800
No. cuentas atrasadas (muestra)	8	10	18	12

Considere el tamaño de la muestra, mediante la afijación óptima de 120 clientes y las varianzas de las ventas para los diferentes estratos son: 160 120 180 140 millones de pesos. Se pide estimar el total de las ventas y el total de cuentas atrasadas con un nivel de confianza del 95%

126. En vista de la recesión económica, una empresa pretende reducir el número de horas laborales por semana a cuatro. Otra alternativa consiste en cerrar una de sus cuatro plantas y despedir a los trabajadores. Para tener una idea de la opinión de los trabajadores, el jefe de personal desea realizar una muestra de empleados en las cuatro plantas para estimar la proporción que prefieren la primera alternativa.
La empresa emplea 90 personas en cada una de las plantas 1 y 2: 75 en la planta 3; y 35 en la planta 4. Se estima que el 80% de los de la planta 4 están a favor de la reducción de la semana de trabajo, mientras que las otras plantas el 70% se inclinan por la segunda alternativa. a) Encuentre el tamaño de la muestra, si se fija una confianza del 95%, un error del 10% y un costo unitario de $1.800. b) Cuál es el tamaño de la muestra para cada una de las plantas. c) Elimine el costo unitario ($1.800) y calcule los tamaños muestrales para cada estrato en forma independiente y directa.

127. Se desea estimar el total de horas-hombre dejadas de trabajar y el total de trabajadores que fueron incapacitados debido a enfermedades o accidentes de trabajo en un mes determinado, en un sector de la economía. Para ello se estratificó en cuatro grupos, teniendo en cuenta el número de trabajadores por establecimiento.

ESTRATO	No. TOTAL DE EMPLEADOS (población)	TOTAL HORAS-HOMBRE NO LABORADAS (Muestra)	NO. EMPLEADOS INCAPACITADOS (Muestra)	DESVIACIÓN ESTÁNDAR Horas-hombre
1	3.400	6.400	100	85
2	5.600	8.000	140	130
3	12.700	10.200	180	160
4	20.000	18.300	200	220

Mediante la Asignación proporcional, se calculó el tamaño muestral siendo de 2.620 empleados y se desea estimar: a) El total de horas-hombre no trabajadas y b) El total de personas que permanecieron incapacitados en el mes. En ambos casos fijar límites de confianza del 95%

128. Se tiene la siguiente población de 18 elementos 16 12 18 17 12 14 20 19 20 16 15 14 13 18 16 22 25 y 16.

a) Divida la población en 2 estratos de tamaño 11 y 7 respectivamente, indicando el criterio empleado.

b) Seleccione al azar una muestra de tamaño 8, aplicando el método de Asignación proporcional, y fije los límites de confianza del 95% para la estimación del promedio.

129. Una empresa desea estimar el número de horas-hombre dejadas de trabajar por diferentes motivos, además, el porcentaje de empleados que llegaron al trabajo con retardos superiores a los 25 minutos, en el mes en que se realizó la investigación, en 6 fábricas localizadas en diferentes zonas del país. Los empleados se estratificaron de acuerdo a los niveles de ocupación.

NIVEL OCUPACIONAL	TOTAL EMPLEADOS	MUESTRA		VARIANZA horas–hombre
		Total Horas–hombre	Número de empleados con retardos superiores a 25 minutos	
Directivos	50	120	4	210
Ejecutivos	120	310	10	196
Profesionales	510	2.100	10	250
Técnico-Adtivos.	830	4.300	18	360
Operarios	1.790	8.800	38	430

a) Se tiene establecido que el costo total de la investigación es de $410.000, el costo fijo de $130.000 y los costos unitarios de $700 para cada estrato. Determinar el valor de n y los n_h.

b) Con la información anterior, estime el total de horas-hombre y el total de empleados con retardos superiores a los 25 minutos. En ambos casos fijar límites de confianza del 95%

MUESTREO SISTEMÁTICO

Más que un método de muestreo, podría ser considerado como un proceso de selección, que algunos denominan *selección a intervalos regulares*. Generalmente se le define como la obtención de los elementos o unidades de una muestra mediante la aplicación de algún *método sistemático,* en el sentido opuesto a los aleatorios. Por tal motivo algunos no lo consideran como un método de muestreo al azar, a menos que la lista o marco se haya conformado al azar.

El *muestreo sistemático* o *en serie*, se usa frecuentemente por ser un método simple, sencillo, directo y económico. Arroja buenos estimativos, cuando la variable está ordenada, por años, valor, cantidad de menor a mayor, etc. Se puede observar en las unidades seleccionadas, que la variable se distribuye en la muestra casi igual que en el *Muestreo Aleatorio Estratificado*, *Asignación proporcional*. La muestra se selecciona, tomando las unidades a intervalos regulares. Veamos cuál es el proceso de selección.

a) Supongamos que la población estudiada está compuesta por 360 familias (en vez de las 355 familias con las cuales se ha venido trabajando), además el tamaño de la muestra (Muestreo Aleatorio Simple o el obtenido por el método de Afijación) es de 30 familias; con esta información se podrá determinar el *intervalo de selección*, simbolizada por I.

$$I = \frac{1}{f} \qquad I = \frac{1}{f} = \frac{N}{n}$$

Reemplazando, se tiene: $I = \frac{360}{30} = 12$ (Cuando es un número entero)

Determinado el intervalo de selección, se debe obtener un *número aleatorio* dentro de ese intervalo; supongamos que entre 001 y 012 se obtuvo el número 004, el cual se denomina punto de arranque. Se ha hecho la primera selección dentro del intervalo, en razón a ser el intervalo (I) un número entero.

Una vez establecido el punto de arranque, mediante la selección aleatoria, se inicia el proceso de selección sistemática, correspondiendo el número 016 a la segunda unidad seleccionada; a ésta se le suma nuevamente el valor del intervalo, para obtener la tercera unidad y así sucesivamente.

004	016	028	040	052	064	076	088	100	112	124	136	148
160	172	184	196	208	220	232	244	256	268	280	292	304
316	328	340	352									

En el caso de que el valor obtenido para el *intervalo de selección* no sea un número entero, se procederá de manera diferente. Siendo la población de 355 familias y el tamaño 30, se tendrá un intervalo igual a:

$$I = \frac{N}{n} = \frac{355}{30} = 11{,}83 \cong 12$$

Una vez calculado el valor del intervalo, deberá redondearse al número inmediatamente superior, por pequeña que sea la fracción. Mediante cualquier método de selección, determinamos un número entre 001 y 355; supongamos que fue 208, a partir del cual (punto de arranque) se va acumulando el valor del intervalo así:

208	220	232	244	256	268	280	292	304	316
328	340	352	364	011	023	035	047	059	071
083	095	107	119	131	143	155	167	179	191

El proceso más utilizado es el de calcular el intervalo de selección sin importar si el número es entero o si presenta decimales, en este último caso, se aproxima al número inmediatamente superior, luego se selecciona aleatoriamente (al azar) un número dentro del intervalo, con el fin de obtener el punto de arranque, al cual se le va acumulando el valor del intervalo, es decir, se procede exactamente igual al primer procedimiento.

Observemos los niveles de ingresos para cada uno de los números aleatorios obtenidos.

No. aleatorio	004	016	028	040	052	064	076	088	100	112
Ingresos (miles $)	1.060	2.060	1.140	950	3.860	1.860	1.840	2.850	2.020	2.620
Propiedad-vivienda	si	si	si	si	no	no	no	no	si	si
No. aleatorio	124	136	148	160	172	184	196	208	220	232
Ingresos (miles $)	840	1.020	1.860	880	1.950	2.870	950	3.890	2.620	1.020
Propiedad vivienda	no	si	si	si	si	si	no	no	no	si
No. aleatorio	244	256	268	280	292	304	316	328	340	352
Ingresos (miles $)	2.260	2.140	1.050	1.280	790	2.790	1.110	2.860	1.320	1.240
Propiedad vivienda	si	no	no	no	no	si	no	si	no	no

Con la anterior información se puede efectuar la estimación puntual, tanto para el promedio como para la proporción, así como también determinar el estimador por intervalo, con una confianza del 95%.

$$s^2 = \frac{123.868.200 - 30\,(1.833,33)^2}{30 - 1} = 794.305,75$$

$$\bar{x} = \frac{55.000}{30} = 1.833,33 \qquad \hat{\bar{X}}_S = \bar{x} \pm t \frac{s}{\sqrt{n}} \sqrt{1 - f}$$

(Nota: se trabajó en la calculadora con el programa de estadística, por lo tanto para s^2, se tomarán todos los decimales en el cálculo de la media)

$$\hat{\bar{X}}_{s_i} = 1.833,33 \pm 2,045\left(\frac{891,24}{\sqrt{30}}\right)\sqrt{1 - \frac{30}{355}} = 1.833,33 \pm 318,39 \begin{cases} 2.151,72 = 2.151.716,06 \\ 1.514,94 = 1.514.943,94 \end{cases}$$

En la *proporción* se tendrá: $p = \frac{15}{30} = 0,50$

$$\hat{P}_{s_i} = p \pm t \sqrt{\frac{pq}{n\text{-}1}} \sqrt{1 - f} \qquad \hat{P}_{s_i} = 0,5 \pm 2,045 \sqrt{\frac{0,5(0,5)}{30\text{-}1}} \sqrt{1 - \frac{30}{355}} \begin{cases} 0,68 = 68\% \\ 0,32 = 32\% \end{cases}$$

Ahora veamos qué hubiera sucedido si la población se encuentra estratificada. Podríamos considerar dos procedimientos, en primer lugar, en el caso de que la población estratificada, esté enumerada en forma continua, desde 001 hasta N.

ESTRATO	NUMERACIÓN POR ESTRATO (A)	NUMERACIÓN CONTINUA (B)
I	Desde 001 hasta 162	Desde 001 hasta 162
II	Desde 001 hasta 132	Desde 163 hasta 294
III	Desde 001 hasta 61	Desde 295 hasta 355
Total	355	

Los números aleatorios seleccionados en el ejercicio anterior, corresponden en la enumeración continua, a los siguientes:

No. ALEATORIO ANTERIOR (B)	EQUIVALENCIA (A)	ESTRATO
004	004	I
016	016	I
018	018	I
040	040	I
052	052	I
064	064	I
076	076	I
088	088	I
100	100	I
112	112	I

No. ALEATORIO ANTERIOR (B)	EQUIVALENCIA (A)	ESTRATO
124	124	I
136	136	I
148	148	I
160	160	I
172	010	II
184	022	II
196	034	II
208	046	II
222	058	II
232	070	II

No. ALEATORIO ANTERIOR (B)	EQUIVALENCIA (A)	ESTRATO
244	082	II
256	094	II
268	106	II
280	118	II
292	130	II
304	10	III
316	22	III
328	34	III
340	46	III
352	58	III

Si consideramos la información de la tabla dada para la Asignación proporcional respecto al Ingreso de las familias (x) en miles de ($) y Propiedad de la vivienda (a), y suponiendo que los tamaños muestrales por estrato sean:

$$n_1 = 14 \qquad n_2 = 11 \qquad n_3 = 5 \qquad n = 30$$

ESTRATO I

No. DE ORDEN	No. ALEATORIO	x_i	a_i
1	004	800	no
2	016	830	si
3	028	950	si
4	040	1.350	si
5	052	1.560	no
6	064	1.300	no
7	076	1.260	no
8	088	880	no
9	100	1.620	no
10	112	1.630	no
11	124	1.010	no
12	136	1.460	si
13	148	980	si
14	160	1.240	no

ESTRATO II

No. DE ORDEN	No. ALEATORIO	x_i	a_i
1	172	2.060	no
2	184	1.870	no
3	196	2.020	si
4	208	2.320	no
5	220	2.060	si
6	232	2.360	no
7	244	1.860	si
8	256	1.990	no
9	268	2.080	no
10	280	1.660	si
11	292	1.650	no

ESTRATO III

No. DE ORDEN	No. ALEATORIO	x_i	a_i
1	304	2.860	si
2	316	3.320	si
3	328	3.620	si
4	340	2.930	si
5	352	2.860	si

$N_1 = 162$ $\quad W_1 = 0{,}46$

$N_2 = 132$ $\quad W_2 = 0{,}37$

$N_3 = 61$ $\quad W_3 = 0{,}17$

$\bar{x}_1 = 1.205$ $\bar{x}_2 = 1.993,64$ $\bar{x}_3 = 3.118$ $p_1 = \frac{6}{14} = 0,43$

$s_1^2 = 87.857,69$ $s_2^2 = 52.025,45$ $s_3^2 = 115.320$ $p_2 = 0,36$

$s_1 = 269,41$ $s_2 = 228,04$ $s_3 = 339,59$ $p_3 = 1,0$

Lo primero que observamos es la forma de distribucion de los elementos en la muestra, igual a la obtenida mediante la *Asignación proporcional*. El tratamiento que se da, para el cálculo de los estimadores puntual e intervalos de confianza, es exactamente igual al utilizado para este método.

$$\bar{x}_{st} = \Sigma W_h \bar{x}_h \qquad \bar{x}_{st} = 0,46\ (1.205) + 0,37\ (1.993,64 + 0,17\ (3.118) = 1.822,01 \cong (\$1.822.006,8)$$

$$p_{st} = \Sigma W_h p_h \qquad p_{st} = 0,46(0,43) + 0,37(0,36) + 0,17(1) = 0,50 \cong 50\%$$

$$\hat{\bar{X}}_{st} = \bar{x}_{st} \pm t \sqrt{\frac{1}{N^2} \Sigma N_h (N_h - n_h) \frac{s_h^2}{n_h}}$$

$$\hat{\bar{X}}_{st} = 1.822,01 \pm 2,052 \sqrt{\frac{1}{355^2}\left[162(162-14)\frac{87.857,69}{14} + 132(132-11)\frac{52.025,45}{11} + 61(61-5)\frac{115.320}{5}\right]} =$$

$$\hat{\bar{X}}_{sts_1} = 1.822,01 \pm 2,052\sqrt{241,85} = 1.822,01 \pm 100,91 \begin{cases} 1.922,91 = \$1.922.910 \\ 1.721,09 = \$1.721.090 \end{cases}$$

$$\hat{P}_{st} = 0,50 \pm 2,052 \sqrt{\frac{1}{355^2}\left[162(162-14)\frac{0,43(0,57)}{14} + 132(132-11)\frac{0,36(0,64)}{14} + 61(61-5)\frac{1}{5}\right]} =$$

$$\hat{P}_{st} = 0,50 \pm 2,052\sqrt{0,005985} = 0,50 \pm 0,16 \begin{cases} 0,66 = 66\% \\ 0,34 = 34\% \end{cases}$$

Específicamente, en el *muestreo aleatorio estratificado*, la aplicación del *método sistemático*, en primer lugar, se debe determinar el tamaño de la muestra y luego los tamaños muestrales para cada uno de los estratos, por cualquiera de los métodos de Asignación que hubiésemos decidido. En cada uno de los estratos muestrales, en forma independiente, se calcula el valor del intervalo de selección con el fin de aplicar la *selección sistemática*.

Supongamos que mediante el método de *Asignación igual* se obtuvo la siguiente información:

$n = 30$ $n_1 = 10$ $n_2 = 10$ $n_3 = 10$

Los *intervalos* son:

$I_1 = \frac{N_l}{n_l} = \frac{162}{10} = 17$ $I_2 = \frac{132}{10} = 14$ $I_3 = \frac{61}{10} = 7$ (Intervalos de selección en cada estrato)

Seleccionamos un número (aleatoriamente) entre 1 y 17, para el primer intervalo por ejemplo 9, se tendrá que los números de selección para este *estrato* serán 002 019 036 053 070 087 104 121 138 155, ya que N_1 es igual a 162. Para el *segundo estrato*, también se seleccionan números aleatorios; supongamos que los resultados fueron 005 019 033 047 061 075 089 103 117 131, en este ejemplo, el tamaño proporcional N_2 es igual a 132; para el *tercer estrato*, los números de selección fueron 01 08 15 22 29 36 43 50 57. Podemos observar que $N_3 = 9$, el cual se explica por la aproximación efectuada en el valor del intervalo, al número inmediatamente superior, tal como generalmente lo hacemos. La experiencia, sin embargo, nos enseña que en algunos casos es preferible trabajar con un número inmediatamente inferior. Otro procedimiento a seguir para este caso, sería continuar la serie, es decir: al ir agregando de 7 en 7, supongamos que el último es 57, al agregarle 7, nos pasamos de 61, pues la suma sería 64, por lo tanto el número que sigue será 03 y con ello se completa el total de elementos para la muestra en ese estrato, si no se hubiere completado, seguiría 10 17 24 31, etc.

En conclusión, en el *muestreo sistemático*, la primera unidad se selecciona mediante utilización de los números aleatorios y el resto se selecciona automáticamente de acuerdo a un proceso predeterminado. Cuando la lista está ordenada por valor, tiempo, de menor a mayor, el procedimiento es equivalente al *Muestreo aleatorio estratificado*, *Asignación proporcional*, en cuanto a las unidades seleccionadas. Es muy aplicado en estimaciones de deudores en un almacén que vende a crédito; para revisar sumas, supongamos que de cada 10 cuentas se revisa minuciosamente una; en el control del horario de entrada y salida de los obreros en una fábrica, por ejemplo: se toman 12 tarjetas diarias y se las revisa; se puede utilizar también para la revisión de obreros con el objeto de evitar el robo de elementos o herramientas, por cada 10 obreros que salen se revisa uno; otros muchos ejemplos se podrían citar, tales como vehículos que se revisan en un retén, barcos pesqueros que atracan en la costa, nombres en un directorio telefónico, etc. Es indispensable preocuparse en la aplicación de este método por las periodicidades escondidas, por ejemplo, las ventas de un almacén, si sólo se tiene en cuenta un día a la semana (siendo el intervalo de 7) por ejemplo los jueves; esta muestra no sería representativa de todas las ventas diarias. Por otra parte en la aplicación de este método se hace necesaria la enumeración y ordenamiento de los elementos en la población, lo cual en algunos casos es casi imposible.

MUESTREO POR CONGLOMERADOS

UNA ETAPA

Si una población se divide en grupo y se realiza una muestra, se dice que se ha realizado una *muestra por conglomerados en una etapa*, en la cual cada grupo sirve como unidad de muestreo. Recordemos que los *estratos* se definen como grupos más o menos homogéneos en cuanto a su composición interna; en cambio, en los *conglomerados*, se espera que la composición interna sea lo más heterogénea posible, de tal forma que cada conglomerado represente en lo posible a la población.

Es muy común que estos *conglomerados* hagan referencia a superficies o áreas en que se ha dividido el terreno, por ejemplo, un barrio o una manzana que pueden considerarse como conglomerados; el primero como un grupo de manzanas, el segundo como un grupo de viviendas. Este método permite reemplazar a las unidades más pequeñas (unidades de selección) de las poblaciones cuando ellas no puedan ser enumeradas, por unidades más grandes que las contienen, haciéndolas fáciles de listar y de manejar y desde luego resulta menos costoso que los métodos anteriormente presentados. Sin embargo, vale la pena señalar que entre más pequeño sea el conglomerado, más exacto será el estimador.

Con frecuencia, a los conglomerados se les denomina *unidades de muestreo primario*. Para este método se presentan varias formas de trabajo para lograr que el estimador del parámetro sea lo más representativo, dependiendo de la característica en la población

a) En primer lugar, vimos en el *muestreo aleatorio simple* como una unidad elemental de selección, familias, en un momento dado se constituía en *conglomerado*, al estar conformado por un grupo de personas; en ese caso, la proporción se estimaba mediante la fórmula:

$$p = \frac{\Sigma a_i}{\Sigma m_i} \qquad \hat{V}_{[p]} = \left[\frac{1-f}{n\bar{m}^2}\right]\left[\frac{\Sigma d_i^2 - 2p\Sigma a_i m_i + p^2\Sigma m_i^2}{n\text{-}^1}\right]$$

b) En segundo lugar, el *muestro sistemático* en algunos casos se parece en su aplicación al muestreo por conglomerados, siendo equivalente la muestra Sistemática de un conglomerado, si se efectúa una selección aleatoria de los (k) conglomerados en que se ha dividido la población. Supongamos una selección sistemática de 30 familias de una población estratificada, para la cual hemos calculado el *intervalo de selección*.

$$I = \frac{355}{30} \cong 12$$

Se podrá decir, que aproximadamente se tienen 12 *grupos* o *conglomerados* de 30 unidades cada uno. Mostraremos algunos de esos conglomerados, para que el estudiante se forme una idea de su conformación.

Conglomerado No. 01

No. ORDEN	No. ALEATORIO	x_i	a_i
1	001	920	no
2	013	1.260	no
3	025	1.350	no
4	037	1.260	no
5	049	1.010	no
6	061	1.130	si
7	073	890	si
8	085	1.010	no
9	097	1.140	si
10	109	1.020	si
-	-	-	-
-	-	-	-
-	-	-	—
28	325	3.210	no
29	337	3.280	no
30	349	2.960	no

Conglomerado No. 02

No. ORDEN	No. ALEATORIO	x_i	a_i
1	002	1.060	sí
2	014	1.040	no
3	026	1.250	si
4	038	950	si
5	050	1.000	no
6	062	840	no
7	074	1.060	no
8	086	1.630	si
9	098	960	si
10	110	850	no
-	-	-	-
-	-	-	-
-	-	-	-
28	326	3.240	si
29	338	3.860	si
30	350	3.290	si

Conglomerado No. 03

No. ORDEN	No. ALEATORIO	x_i	a_i
1	003	1.080	no
2	015	1.450	no
3	027	1.050	si
4	039	980	no
5	051	1.120	si
6	63	1.350	no
7	075	860	no
8	086	1.030	sí
9	099	930	no
10	111	1.450	sí
-	-	-	-
-	-	-	-
-	-	-	-
28	327	3.860	sí
29	339	2.610	sí
30	351	2.860	sí

Conglomerado No. 04

No. ORDEN	No. ALEATORIO	x_i	a_i
1	012	790	no
2	024	840	si
3	036	790	no
4	048	890	si
5	060	980	si
6	072	1.060	no
7	084	880	no
8	087	1.130	sí
9	108	960	no
10	120	-	-
-	-	-	-
-	-	-	-
-	—	-	-
27	324	2.840	no
28	324	2.960	no
29	348	2.790	sí

Observemos al elaborar todos los conglomerados posibles, que en los 7 primeros cada uno tiene 30 unidades y los restantes 5, tan sólo tienen 29 unidades, ya que la población es de 355 familias y no de 360, para que tuviesen igual número de elementos.

Cada *conglomerado* será una pequeña réplica de la población. Una vez establecidos los grupos o conglomerados, utilizamos la Tabla de números aleatorios, con el fin de seleccionar un número entre 01 y 12. Supongamos que el resultado fue 03, correspondiente al conglomerado seleccionado, que se constituye en una muestra de 30 elementos.

c) En general, cuando una población se divide en grupos o conglomerados y se quiere realizar estimativos, se debe usar la siguiente simbología:

N = Número de conglomerados en la población

n = Número de conglomerados en la muestra

M_i = Tamaño de cada conglomerado (total de elementos en el grupo)

$M = \Sigma M_i$ = Total de elementos en la población de N conglomerados

$m = \Sigma m_i$ = Total de elementos en la muestra de n conglomerados

$\overline{M} = \dfrac{M}{N}$ = (tamaño medio de los conglomerados)

y_{ij} = Valor de la variable

y_i = Total del conglomerado en la muestra

$y_i = \overset{m_i}{\Sigma} y_{ij}$

$\bar{y}_i = \dfrac{\overset{m_i}{\Sigma} y_{ij}}{m_i} = \dfrac{y_i}{m_i} =$ (Media del conglomerado) $y = \overset{n}{\Sigma} y_i = \overset{n}{\Sigma}\overset{m_i}{\Sigma} y_{ij}$ (Total general de la muestra)

$\bar{\bar{y}}$ = *Media general de la muestra* Siendo $\bar{\bar{y}} = \frac{y}{\Sigma m_i} = \frac{y}{m}$

$\bar{\bar{y}}$ = *Media de totales o total medio por conglomerado*, en la muestra: $\bar{\bar{y}} = \frac{y}{m}$

Es de anotar, que en el *muestreo estratificado* el estimador de una proporción, por conglomerados se debe calcular con la siguiente fórmula:

$$p_{st} = \frac{\Sigma M_h p_h}{\Sigma m_h} \qquad \text{Siendo} \qquad p_h = \frac{\Sigma a_{h_i}}{\Sigma m_{h_i}}$$

y el error de estimación será igual a:

$$\hat{V}_{[Pst]} = \Sigma\left[\frac{1-f_h}{n_h \bar{m}_h^2}\right]\left[\frac{\Sigma a_{hi}^2 - 2p_h \Sigma m_{hi} a_{hi} + p_h^2 \Sigma m_{hi}^2}{n_h - 1}\right]$$

Veamos dos ejercicios que nos permitan entender el significado y la aplicación con *conglomerados iguales y desiguales,* realizados en una etapa; y al mismo tiempo, observar el proceso que se sigue para el cálculo del estimador, cuando se trata de una variable, ya que algunos de los ejemplos anteriores hacemos referencia al estimador de una proporción, pero es una forma de trabajar con *conglomerados en una etapa* (ver: MAS y MAE).

En primer lugar, para el *muestreo por conglomerados iguales*, suponemos que la población de 355 familias se ha dividido en 71 conglomerados y cada conglomerado se encuentra conformado por 5 familias, así que todos contienen el mismo número de familias, por tal razón se les denomina *conglomerados iguales*.

Consideremos ahora que la muestra está constituida por 12 conglomerados: (n = 12), que han sido seleccionados aleatoriamente, como se observa en la siguiente información:

CONGLOMERADOS	No. SELECCIÓN	No. PERSONAS POR FAMILIA	TOTAL
1	18	3 – 5 – 2 – 1- 4	15
2	03	3 –3 – 1 – 2 – 3	12
3	64	4 – 6 – 3 – 5 – 1	19
4	36	4 – 4 – 3 – 2 – 2	15
5	41	5 – 2 – 6 – 1 – 3	17
6	38	3 – 3 – 4 – 2 – 2	14
7	30	3 – 2 – 3 – 5 – 2	15
8	22	2 – 2 – 4 – 5 – 3	16
9	14	1 – 3 – 6 – 1 – 2	13
10	26	2 – 1 – 5 – 2 – 1	11
11	19	3 – 2 – 3 – 3 – 1	12
12	46	6 – 1 – 4 – 4 – 2	17

N = 71 n =12

Con los datos de la anterior muestra, calcularemos el *promedio* para cada conglomerado.

I	$\bar{x}_i$	$\bar{x}_i$	$\bar{x}_i^2$
1	15	3,0	9,00
2	12	2,4	5,76
3	19	3,8	14,44
4	15	3,0	9,00
5	17	3,4	11,56
6	14	2,8	7,84
7	15	3,0	9,00
8	16	3,2	10,24
9	13	2,6	6,76
10	11	2,2	4,84
11	12	2,4	5,76
12	17	3,4	11,56
Σ	176	35,2	105,76

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}_i}{n} = \frac{35,2}{12} = 2,93 \cong 3$$

$$s^2_{\bar{\bar{x}}} = \left[\frac{N-n}{Nn}\right]\left[\frac{\Sigma \bar{x}_i^2 - n\bar{\bar{x}}^2}{n-1}\right]$$

$$s^2_{\bar{\bar{x}}} = \left[\frac{71-12}{71(12)}\right]\left[\frac{105,76 - 12(2,93)^2}{12-1}\right]$$

$$s^2_{\bar{\bar{x}}} = (0,0692)\,(0,2492) = 0,017$$

El anterior resultado indica que en promedio ($\bar{x}$) hay 3 personas por familia.

En el caso de muestreo por *conglomerados desiguales* con probabilidades iguales, supongamos que el barrio de 355 familias está compuesto por 25 manzanas y cada una de ellas se constituye en un conglomerado desigual con respecto al número de viviendas.

De la población N = 25 (conglomerados), se extrajo una muestra de 10 manzanas (n = 10), con la siguiente información:

i CONGLOMERADO	m_i VIVIENDA	y_i PERSONAS	y_i^2
1	5	18	324
2	12	48	2.304
3	8	40	1.600
4	10	36	1.296
5	14	30	900
6	11	45	2.025
7	13	34	1.156
8	9	28	784
9	8	39	1.521
10	15	61	3.721
Σ	105	379	15.631

$\Sigma m_i^2 = 1.189$

$\Sigma y_i^2 = 15.631$

$\Sigma m_i = 105$

$\Sigma y_i = 379$

$\Sigma m_i y_i = 4.182$

En primer lugar identifiquemos los símbolos que se deben utilizar:

N = Número total de conglomerados en la población (N = 25)

n = Número total de conglomerados en la muestra (n = 10)

M = Número total de elementos en la población (M = 330 viviendas)

M_i = Número de elementos en cada conglomerado poblacional.
$M = \Sigma M_i = 330$ viviendas

$\bar{M}$ = Tamaño medio de los conglomerados

$M = \Sigma M_i = \frac{30}{25} = 13{,}2$ (Promedio de viviendas por conglomerado)

$\bar{y} = \frac{\Sigma y_i}{n}$ = Promedio de personas por conglomerado de la muestra

$\bar{y} = \frac{379}{10} = 3{,}79$ $\qquad \bar{\bar{y}} = \frac{14}{\bar{M}n}\Sigma y_i = \frac{N\bar{y}}{M}$

$\bar{\bar{y}}$ = Estimador de la media por elemento: $\qquad \bar{\bar{y}} = \frac{1}{13{,}2(10)}(379) = 2{,}87$

El anterior estimador también se puede calcular de la siguiente forma:

$\bar{\bar{y}} = \frac{25(37{,}9)}{330} = \frac{947{,}5}{330} = 2{,}87$ Promedio de personas por vivienda

$\hat{V}_{[\bar{y}^*]} = \left[\frac{1-f}{n\bar{M}^2}\right]\left[\frac{\Sigma y_i^2 - n\bar{y}^2}{n-1}\right]$ $\qquad \hat{V}_{[\bar{\bar{y}}]} = \left[\frac{1-0{,}4}{10(3{,}2)^2}\right]\left[\frac{15.631 - 10(37{,}9)^2}{10-1}\right]$

$f = \frac{n}{N} = \frac{10}{25} = 0{,}4$ $\qquad V_{[\bar{\bar{y}}]} = (0{,}00034)\,(140{,}7667) = 0{,}048$

También se puede calcular:

$\hat{V}_{[\bar{\bar{y}}]} = \left[\frac{N-n}{Nn(n-1)}\right]\left[\frac{1}{\bar{M}^2}\Sigma y_i^2 - n\bar{y}^2\right]$ $\quad \hat{V}_{[\bar{\bar{y}}]} = \left[\frac{25\text{-}10}{25(10)(9)}\right]\left[\frac{1}{(13{,}2)^2}(15.631\text{-}10(37{,}9)^2\right] = 0{,}048$

De otra manera se podrá hacer:

$\hat{V}_{[\bar{\bar{y}}]} = (1-f)\left[\left(\frac{N}{M}\right)\left(\frac{1}{n(n-1)}\right)\right]\left[\Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}\right]$ Reemplazando se tendrá que:

$\hat{V}_{[\bar{\bar{y}}]} = \left(1-\frac{10}{25}\right)\left[\left(\frac{25}{330}\right)^2\left(\frac{1}{10(9)}\right)\right]\left[15.631 - \frac{(379)^2}{10}\right] \cong 0{,}048$

$\hat{s}_{[\bar{\bar{y}}]} = \sqrt{0{,}048} = 0{,}22$ (Error estándar de 0,22 personas)

El *error de estimación*, también se puede calcular usando los estimadores de la razón, siendo:

$\hat{V}_{[\bar{\bar{y}}]} = \left[\frac{1-f}{n\bar{M}^2}\right]\left[\frac{\Sigma y_i^2 - 2\bar{\bar{y}}\Sigma M_i y_i + \bar{\bar{y}}^2\Sigma M_i^2}{n-1}\right]$

$$\hat{V}_{[\bar{\bar{y}}]} = \left[\frac{1 - \frac{10}{25}}{10(13{,}2)^2}\right]\left[\frac{15.631 - 2(2{,}87)(4{,}182) + 2{,}87^2(1.189)}{10 - 1}\right] = 0{,}054; \qquad s^2_{\bar{\bar{y}}} = 0{,}23$$

MUESTREO BIETÁPICO

Se le denomina *muestreo por etapas*. Cuando son dos, recibe el nombre de *bietápico* y si son más de dos etapas se le considera *polietápico*. Puede considerarse como un *sub-muestreo* y se trata de una modificación del método de muestreo por conglomerados, ya que la primera muestra está conformada por unidades que son consideradas como conglomerados, una vez seleccionados éstos, se efectúa una nueva selección o submuestreo dentro de cada conglomerado para la segunda etapa y así sucesivamente.

Los conglomerados que resultan de la primera división se les denomina unidades *primarias*, los de la segunda, *secundarias* y así sucesivamente. Los conglomerados pueden ser de igual o de diferentes tamaños, siendo este último caso el más generalizado.

De lo anterior se desprende, que lo característico de este tipo de muestreo es el *proceso de selección que se hace por etapas sucesivas;* en cada una de la unidad es diferente, y las unidades finales (la última muestra) son las que nos permiten la realización de la encuesta, y la recolección de información; las anteriores son consideradas *unidades de selección.*

Son muchos los ejemplos que se podrían citar, en el caso del *Muestreo por etapas*. En primer lugar, para la realización de la encuesta de hogares, podríamos dividir la ciudad en barrios (clasificados por niveles de ingresos, que serían los estratos); en esta *etapa* se eligen en forma aleatoria o sistemática (de acuerdo al procedimiento adoptado), las unidades que conforman el tamaño de la muestra, cuya unidad de selección es el barrio; la *segunda etapa* consistiría en elegir las manzanas en cada uno de los barrios seleccionados en la primera etapa. En la *tercera etapa* podrían seleccionarse las viviendas, de las manzanas obtenidas en la etapa anterior. De continuarse con la *etapa cuarta,* tendría como unidad de selección la familia y podría ser considerado como el elemento, el individuo.

Otro ejemplo, si una empresa tiene sistemas de venta a crédito, para cada deudor se tendrá una tarjeta, lo que implica tener varias gavetas, cada una de ellas con un número de tarjetas. El proceso a seguir aplicando el método *bietápico*, consiste en enumerarlas y seleccionar una muestra cuya unidad sea gaveta, luego se vuelve a muestrear, enumerando las tarjetas en cada una de esas gavetas a fin de efectuar una nueva selección, teniendo como elemento final las tarjetas.

De los anteriores ejemplos se desprende, que si la unidad de muestreo es un conglomerado constituído por elementos o unidades de observación, no es necesario considerar todas las unidades de observación que pertenecen a la unidad seleccionada de muestreo. Esta misma unidad de muestreo, se constituye en una pequeña población de la cual se podrá obtener una muestra y así sucesivamente. Esto requiere la construcción de marcos, para cada una de las etapas.

En el *muestreo bietápico estratificado*, se aplican fórmulas especiales que pueden ser consultadas en los diferentes textos de muestreo, lo mismo sucede en el *Muestreo trietápico* o *polietápico* y algo similar, cuando se trata de atributos. Para estos métodos se recomienda al lector consultar el libro del profesor Francisco Azorín Poch (ver bibliografía) que presenta en forma bastante comprensible esta teoría. No se exponen estos métodos en este texto, por ser considerados extensos para los objetivos propuestos.

MUESTREO POR CONGLOMERADOS DE DOS ETAPAS

En varias oportunidades se ha definido el concepto conglomerado; pero vale la pena recordar que: el *muestreo aleatorio por conglomerados* es aquel en el cual cada unidad de selección es una colección, conjunto o conglomerado de unidades o elementos. Se dice que es de dos etapas o bietápico, ya que se obtiene primero una muestra aleatoria de conglomerados y, luego en una segunda etapa una muestra aleatoria de los elementos de cada conglomerado que han sido seleccionados.

Este método facilita la elaboración del marco en la primera etapa y luego el de la segunda, siendo al final más económico que el obtenido mediante el *Muestreo aleatorio estratificado*.

Estimación de promedios y totales

Supongamos que una firma tiene 110 fábricas distribuidas por todo el territorio nacional y se desea estimar el tiempo y número de máquinas que duraron paralizadas, ya sea por mantenimiento o por daño, durante un trimestre. Por tiempo y distancia se diseña una investigación aplicando el *Muestreo por conglomerados en dos etapas*, para ello se procede a realizar una muestra de doce fábricas y en la segunda etapa, se selecciona una muestra del 20% de las máquinas de cada una de ellas.

El problema se puede presentar de dos maneras: a) Conociendo el total de máquinas para las 110 fábricas, por ejemplo 5.500 máquinas; b) Se desconoce el valor de M, es decir, el número de máquinas.

a) Cuando se conoce el valor de M

Algunos símbolos de entrada que debemos conocer

N = Número de conglomerados en la población $N = 110$
n = Número de conglomerados en la muestra $n = 12$
M = Número de elementos en la población $M = 5.500$ $M = \Sigma M_i$
$\bar{M}$ = Tamaño promedio del conglomerado $\bar{M} = \frac{M}{N} = \frac{5.500}{110} = 55$
m = Número de elementos en cada conglomerado.

Veamos la presentación de la información obtenida en la siguiente tabla:

Información de una encuesta por muestreo aleatorio por conglomerados

FÁBRICA	M_i	m_i	TIEMPO SIN FUNCIONAR (HORAS)	$\bar{x}_i$	S_i^2
1	46	9	3 5 0 0 0 8 6 0 0	2,44	10,03
2	38	8	4 0 0 6 10 0 5 0	3,13	14,13
3	52	10	0 0 0 6 0 4 8 6 1 0	2,50	10,06
4	60	12	3 3 0 1 1 0 0 1 5 0 2 0	1,33	2,61
5	55	11	0 0 3 0 0 6 0 7 0 1 0	1,55	6,87
6	40	8	2 1 5 0 0 0 0 8	2,00	8,86
7	40	8	4 0 0 6 10 2 6 0	3,50	13,43

8	36	7	0 0 3 7 0 1 2	1,86	6,48
9	30	6	5 0 6 0 0 4	2,50	7,90
10	64	13	0 0 0 6 5 0 10 0 6 0 0 0 1	2,15	11,47
11	70	14	2 0 0 0 4 6 2 0 3 6 2 0 0 0	1,79	4,95
12	50	10	2 3 5 0 0 0 0 0 1 2	1,30	2,90

Cálculo del *estimador de la media*:

$$\bar{X}_c = \frac{N\Sigma M_i \bar{x}_i}{Mn} = \frac{1.216{,}09(110)}{5.500(12)} = 2{,}03$$

$$M_i\bar{x}_i = 46(2{,}44) + 38(3{,}13) + 52(2{,}50) + ... = 1.216{,}09$$

$$\Sigma M_i = 46 + 38 + 53 + 60 + ... = 581$$

Los *límites de confianza* del 95% para el estimador de la media es:

$$\hat{\bar{X}}_c = \bar{x}_c \pm t \sqrt{\left[\frac{N-n}{N}\right]\left[\frac{1}{n\bar{M}^2}\right]s_b^2 + \left[\frac{1}{nN\bar{M}^2}\right]\left[\Sigma M_i\,(M_i - m_i)\,\frac{s_i^2}{m_i}\right]}$$

$$s_b^2 = \left[\frac{1}{n\text{-}1}\right]\left[\Sigma(M_i\bar{x}_i)^2 - 2M\bar{x}_c\,\Sigma M_i\bar{x}_i + n\,(\bar{M}\bar{x}_c)^2\right]$$

Reemplazando se tendrá que:

$$\Sigma(M_i\bar{x}_i)^2 = (46 \text{ x } 2{,}44)^2 + (38\text{x}3{,}13)^2 + (52 \text{ x } 2{,}5)^2 + ... = 132.247{,}64$$

$$s_b^2 = \left[\frac{1}{12-1}\right]\left[132.247{,}64 - 2(55)(2{,}03)(1.216{,}09) + 12[55(2{,}03)]^2\right] = 934{,}8$$

$$\Sigma M_i(M_i - m_i)\frac{s_i^2}{m_i} = 46(46-9)\frac{10{,}03}{9} + 38(38-8)\ \frac{14{,}13}{6} + ... = 18.571{,}91$$

Ahora se tendrá que:

$$\hat{\bar{X}}_c = 2{,}03 \pm 2{,}201\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(55)^2}\right]934{,}86 + \left[\frac{1}{12(110)(55)^2}\right](18.571{,}91)}$$

$$\hat{\bar{X}}_c = 2{,}03 \pm 0{,}37 \begin{array}{l}\nearrow 2{,}40 \\ \searrow 1{,}66\end{array} \text{ Horas}$$

Para el Total, los límites del 95% de confianza sabiendo que M = 5.500, se tendrá:

$$\hat{X}_c = 5.500(2,03) \pm 5.500(2,201)\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(55)^2}\right]934,86 + \left[\frac{1}{12(110)(55)^2}\right](18.571,91)}$$

$$\hat{X}_c = 11.165 \pm 2.035 \quad \begin{matrix}\nearrow 13.200 \\ \searrow 9.130\end{matrix} \quad \text{Total de horas}$$

b) *Cuando no se conoce el valor de M*, las fórmulas anteriores sufren una ligera modificación:

$$\bar{x}_c = \frac{\Sigma M_i \bar{x}_i}{\Sigma M_i} = \frac{1.216,09}{581} = 2,09$$

$$s_r^2 = \left[\frac{1}{n-1}\right]\left[\Sigma(M_i\bar{x}_i)^2 - 2\bar{x}_c\Sigma M_i^2\bar{x}_i + \bar{x}_c^2\Sigma M_i^2\right]$$

$$\Sigma M_i^2\bar{x}_i = 46^2(2,44) + 38^2(3,13) + 52^2(2,50) + ... = 60.207,47$$

$$\Sigma M_i^2 = 46^2 + 38^2 + 52^2 + 60^2 + 55^2 + ... = 29.781$$

$$s_r^2 = \left[\frac{1}{12-1}\right]\left[132.247,64 - 2(2,09)(60.207,47) + 2,09^2(29.781)\right] = 969,71$$

Los *límites de confianza* del 95% para el *estimador del promedio* serán:

$$\hat{\bar{X}}_c = 2,09 \pm 2,201\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(49)^2}\right]969,71 + \left[\frac{1}{12(110)(49)^2}\right](18.571,91)}$$

$$M = \frac{\Sigma M_i}{n} = \frac{581}{12} = 49 \qquad \hat{\bar{X}}_c = 2,09 \pm 0,42 \quad \begin{matrix}\nearrow 2,51 \\ \searrow 1,67\end{matrix} \quad \text{Horas}$$

Y el total de horas, con un 95% de confianza, será:

$$\hat{X}_c = 5.390(2,09) \pm 5.390(2,201)\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(49)^2}\right]969,71 + \left[\frac{1}{12(110)(49)^2}\right](18.571,91)}$$

$$\hat{X}_c = 11.265,10 \pm 2.263,8 = \begin{matrix}\nearrow 13.528,9 \\ \searrow 9.001,3\end{matrix} \quad \text{Total de horas}$$

Siendo $M = N\bar{M} = 110\ (49) = 5.390$ máquinas (estimación)

Nota: En la tabla, se puede dar únicamente el tiempo que permanece sin funcionar, para que el estudiante calcule cada uno de $\bar{x}_i$ y s_i; también pueden suministrarse solamente las dos últimas columnas, evitando de esta manera el cálculo de estas medidas.

Estimación de una proporción y total

a) Se conoce el valor de M

Cuando se conoce M, se va a estimar la proporción de máquinas que se paralizan en el trimestre, ya sea por mantenimiento y/o reparación.

Información sobre una muestra por conglomerados doble etapa

FÁBRICA	M_i	m_i	NÚMERO PARALIZADOS	P_i	q_i
1	46	9	4	0,44	0,56
2	38	8	4	0,50	0,50
3	52	10	5	0,50	0,50
4	60	12	7	0,58	0,42
5	55	11	4	0,36	0,64
6	40	8	4	0,50	0,50
7	40	8	5	0,63	0,37
8	36	7	4	0,57	0,43
9	30	6	3	0,50	0,50
10	64	13	5	0,38	0,62
11	70	14	7	0,50	0,50
12	50	10	5	0,50	0,50

Se cuenta el número de máquinas que fueron reparadas o se les hizo mantenimiento. Información adicional.

$\Sigma M_i = 581$

$N = 110$

$M = 5.500$

$\bar{M} = \frac{M}{N} = \frac{5.500}{110} = 55$

$\Sigma M_i^2 = 29.781$

$\bar{M} = \frac{\Sigma M_i}{n} = \frac{581}{12} = 49$

(cuando se estima)

$$p_c = \frac{N}{nM}\Sigma M_i p_i \qquad p_c = \left[\frac{110}{12(5.500)}\right][284,44] = 0,4748 = 47,48\% \cong 47\%$$

Los *límites de confianza* para el estimador se obtienen mediante la siguiente fórmula:

$$\hat{P}_c = p_c \pm t\sqrt{\left[\frac{N-n}{N}\right]\left[\frac{1}{n\bar{M}^2}\right]s_b^2 + \left[\frac{1}{nN\bar{M}^2}\right]\left[\Sigma M_i(M_i - m_i)\frac{p_i q_i}{m_i}\right]}$$

Siendo:

$$s_b^2 = \sqrt{\left[\frac{1}{n-1}\right]\left[\Sigma(M_i p_i)^2 - 2p_c M\Sigma \bar{M}_i p_i + n(\bar{M}p_c)^2\right]}$$

Reemplazando en la fórmula anterior se tiene que:

$$s_b = \left[\frac{1}{12-1}\right]\left[7.172,31 - 2(0,47)(49)(284,88) + 12\,[55(0,47)]^2\right] = 188,12$$

$\Sigma(M_i p_i)^2 = [46(0,44)]^2 + [38(0,5)]^2 + [52(0,5)]^2 + ... = 7.172,31$

$\Sigma M_i p_i = 46(0,44) + 38(0,5) + 52(0,5) + ... = 284,88$

$\Sigma M_i(M_i - m_i)\frac{p_i q_i}{m_i} = 46(46-9)\frac{0,44 \text{ x } 0,56}{9} + 38(38-8)\frac{0,5\text{x}0,5}{8} + ... = 580,24$

Con una confianza del 95%, los límites serán:

$$\hat{P}_c = 0{,}47 \pm 2{,}201\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(55)^2}\right]188{,}12 + \left[\frac{1}{12(110)(55)^2}\right](580{,}24)}$$

$$\hat{P}_c = 0{,}47 \pm 0{,}15 \begin{cases} \nearrow 0{,}62 = 62\% \\ \searrow 0{,}32 = 32\% \end{cases}$$ Horas

Porcentaje de máquinas paralizadas.

Nota: El error obtenido es muy grande, es posible que esto se deba al trabajar con una muestra muy pequeña.

Para establecer los límites de confianza para el Total, será:

$$\hat{A}_c = 0{,}47(5.500) \pm 5.500(2{,}201)\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(55)^2}\right]188{,}12 + \left[\frac{1}{12(110)(55)^2}\right](580{,}24)}$$

$$\hat{A}_c = 2.285 \pm 834 \begin{cases} \nearrow 3.119 \\ \searrow 1.451 \end{cases}$$ Total de máquinas

b) Cuando no se conoce M

Cuando se desconoce M se procede de la siguiente manera:

$$s_r^2 = \left[\frac{1}{n-1}\right]\left[\Sigma(M_i p_i)^2 - 2p_c(\Sigma M_i^2 p_i) + p_c^2 \Sigma M_i^2\right]$$

$$s_r^2 = \left[\frac{1}{12-1}\right]\left[7.172{,}31 - 2(0{,}49)(14.435{,}24) + 0{,}49^2(29.781)\right] = 16{,}02$$

$$\Sigma M_i^2 p_i = 46^2(0{,}44) + 0{,}38^2(0{,}5) + 52^2(0{,}5) + \ldots = 14.435{,}24$$

$$p_c = \frac{\Sigma M_i p_i}{\Sigma M_i} = \frac{284{,}88}{581} = 0{,}49 = 49\%$$

Los *límites de confianza* del 95% para el estimador de la Proporción, será:

$$\hat{P}_c = 0{,}49 \pm 2{,}201\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(49)^2}\right]16{,}02 + \left[\frac{1}{12(110)(49)^2}\right](580{,}24)}$$

$$\hat{P}_c = 0{,}49 \pm 0{,}20 \begin{cases} \nearrow 0{,}69 = 69\% \\ \searrow 0{,}29 = 29\% \end{cases}$$

Para el *estimador del total,* con una confianza del 95% serán:

$$\hat{A}_c = 0{,}49(5.390) \pm 2{,}201(5.390)\sqrt{\left[\frac{110-12}{110}\right]\left[\frac{1}{12(49)^2}\right]16{,}02 + \left[\frac{1}{12(110)(49)^2}\right](580{,}24)}$$

$$\hat{A}_c = 2.641 \pm 1.068 \begin{cases} \nearrow 3.709 \\ \searrow 1.573 \end{cases}$$ Total de máquinas

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

132. Una empresa de investigación de mercados ideó un plan de muestreo para estimar las ventas semanales de un producto X en un área geográfica. La empresa decidió muestrear ciudades dentro del área y luego supermercados dentro de las ciudades. La medición de interés es el número de artículos de dicho producto vendidos en una semana específica: Ocho ciudades son muestreadas entre las 25 del área. Usando los datos de la tabla adjunta, estime el total de artículos vendidos en todos los supermercados del área, para la semana específica, con una confianza del 95%

CIUDAD	No. SUPERM.	No. SUPERM. MUESTREADOS	No. DE ARTÍCULOS VENDIDOS POR SUPERMERCADOS
1	45	9	80 62 26 38 65 74 38 42 54
2	36	7	52 24 32 60 76 42 51
3	20	4	32 36 48 46
4	18	4	30 61 38
5	12	3	36 40 26
6	16	4	34 38 46 51
7	13	4	42 61 24 26
8	22	5	30 54 28 61 26

133. Una cadena de supermercados se encuentra establecida en 70 ciudades. Un director de la compañía quiere estimar la proporción de establecimientos la cadena que no satisfacen un criterio de limpieza específica, ellas se encuentran dentro de cada ciudad al parecer poseen características similares, por lo tanto el Director decide seleccionar una muestra por conglomerados en dos etapas conteniendo la mitad de establecimientos en cada una de las ocho ciudades. El muestreo por conglomerado es conveniente en esta situación debido al costo de traslado. Los datos recopilados se presentan a continuación:

CIUDAD	NÚMERO DE ESTABLECIMIENTOS EN LA CIUDAD	NÚMERO DE ESTABLECIMIENTOS MUESTREADAS	NÚMERO DE ESTABLECIMIENTOS QUE NO SATISFACEN EL CRITERIO DE LIMPIEZA
1	18	9	3
2	20	10	6
3	16	8	5
4	26	13	9
5	14	7	3
6	18	9	5
7	14	7	2
8	12	6	4

Estime el Total de establecimientos que no satisfacen el criterio de limpieza, fijando unos límites de confianza del 95%.

134. El dueño de un vivero quiere estimar la altura de los arbolitos en un extenso terreno, el cual está dividido en 50 lotes que varían levemente de tamaño. El dueño cree que la altura es bastante homogénea en cada lote, pero pueden variar considerablemente de un lote a otro. Por lo tanto decide muestrear un 10% de los árboles dentro de cada uno de los 10 lotes, usando una muestra por conglomerado en dos etapas, los datos se dan en la tabla adjunta:

LOTE	NÚMERO DE ARBOLITOS	NÚMERO DE ARBOLITOS ARBOLITOS MUESTREADOS	ALTURA DE LOS ARBOLITOS (en pulgadas)
1	52	5	12 12 10 13
2	56	6	10 9 7 8 10 12
3	60	6	8 6 6 10 12 8
4	46	5	7 8 7 7 6
5	49	5	12 5 8 7 7
6	51	5	6 5 7 6 6
7	50	5	8 12 10 5 5
8	61	6	5 8 10 10 6 5
9	60	6	7 6 12 6 8 6
10	45	5	10 7 7 10 9

En el vivero hay aproximadamente 2.600 arbolitos. Estime la altura promedio de los arbolitos fijando límites de confianza del 95%.

135. Suponga que un sociólogo quiere estimar el número total de personas jubiladas que viven en una ciudad. El sociólogo decide muestrear manzanas y después casas dentro de las manzanas. Seis manzanas son seleccionadas al azar de entre las 350 de la ciudad. Estime el número total de residentes jubilados en la ciudad y establezca los límites de confianza del 95%

MANZANAS	NÚMERO DE CASAS	NÚMERO DE CASAS MUESTREADAS	NÚMERO DE RESIDENTES JUBILADOS POR CASA			
1	14	4	1	3	0	0
2	14	4	2	0	1	0
3	9	4	3	0	1	1
4	12	4	0	0	1	2
5	16	4	3	2	2	0
6	13	4	1	0	3	2

136. Supongamos que en una ciudad se tienen registrados 620 establecimientos, correspondiente a la pequeña industria y se desea investigar algunas características de su fuerza laboral tales como: edad, nivel educativo y salarial, personas a cargo, tiempo de antigüedad, propiedad de la vivienda,

etc. Se seleccionó una muestra de 10 establecimientos, investigando el 20% de sus trabajadores. Consideremos únicamente la antigüedad y propiedad de la vivienda, con los siguientes resultados:

ESTABLECIMIENTOS	No. DE EMPLEADOS	m_i	ANTIGÜEDAD		PROPIETARIOS DE LA VIVIENDA
			$\bar{y}_c$	s_i^2	
1	25	5	8,3	12,35	2
2	30	6	5,2	4,28	4
3	28	8	6,4	6,54	3
4	20	4	10,6	11,20	1
5	24	5	12,3	8,36	2
6	28	6	8,6	4,16	3
7	36	7	7,8	6,12	5
8	40	8	10,4	5,26	2
9	46	9	14,3	8,35	4
10	28	6	16,2	7,28	4

Fijar límites de confianza del 95%, para la estimación del promedio de antigüedad, así como el total de empleados propietarios de su vivienda.

MUESTREO POR FASES MÚLTIPLES

Es fácil confundirse, en cuanto a la terminología se refiere, en este caso qué es un *Muestreo por etapas* y por *Fases múltiples*. En el *primero* como se vio, el proceso consiste en dividir la población en grupos o conglomerados y tomar de ellos la muestra necesaria. Hecho esto, se subdivide dicha muestra en unidades de segundo grado, de la cual se selecciona una nueva muestra y así sucesivamente. Dependiendo del número de etapas, se habla de muestreo bietápico, trietápico o multietápico.

El *segundo caso*, se presenta, cuando la información requerida, nos permite recogerla en momentos diferentes. Consiste en seleccionar una muestra grande en forma rápida, sencilla y poco costosa, acerca de las características generales de la población, objeto de estudio; generalmente este método nos permite analizar una variable de fácil acceso que está correlacionada con otra de difícil accesibilidad, de ahí que sea empleada con frecuencia en los métodos de estimaciones por *regresión*, para lo cual es necesario el conocimiento del total de la población de la variable independiente (predictor), pero, por desconocerla, es preciso estimarla mediante una muestra.

Una vez desarrollado el anterior proceso, se prosigue con la segunda fase, utilizando unidades ya consideradas como base para la realización de una nueva selección de una muestra mucho más pequeña, destinada al estudio de una característica más específica. Supongamos que se emplea un cuestionario, elaborado en forma sencilla, aplicado a un grupo de 20.000 estudiantes universitarios, tomados como muestra. Recogida la información y de acuerdo a objetivos específicos, nos permitirá elaborar otro formulario detallado para una muestra de 2.000 estudiantes, seleccionados de la anterior. Observemos que en este método de dos fases, llamado por tal motivo *bifásico*, la unidad es la misma. Si fuera de varias

fases o polifásico, la unidad seguirá siendo la misma, en cambio en el *Muestreo polietápico* la unidad de muestreo cambia en cada etapa.

El método *polifásico* permite obtener información, cuando se desea obtener o conocer la proporción de elementos para cada estrato (en el muestreo estratificado) o cuando no se tiene información acerca del tamaño de población en cada estrato.

MÉTODOS MIXTOS

Hemos explicado los métodos más utilizados en el muestreo, vale la pena, hacer referencia a un sistema que en la práctica es muy común y que consiste en la combinación de los diferentes métodos que se han venido exponiendo. No es raro que, en una investigación por muestreo, la población se divida en estratos, en uno de ellos se realice una muestra aleatoria, en otro una muestra sistemática, en otro la muestra sea por etapas, éste último requerirá la conformación de conglomerados. Todo este proceso de combinar dos o más métodos de muestreo, recibe el nombre de *Métodos mixtos*.

Apéndice

CONTENIDO

- **GLOSARIO**
- **TABLAS**
  - Áreas bajo la curva normal de probabilidad
  - Distribución "t" de student
  - Distribución binomial
  - Distribución de j_i cuadrado áreas bajo la curva normal de probabilidad
  - Transformación de r a z
  - Números al azar
  - Distribución de poisson
  - Intervalo de confianza del 95% (2 sigma)
  - Seguridad del 95% (2 sigma)
  - Seguridad del 99,7% (3 sigma)
  - Tabla para la determinación de una muestra (n) sacada de una población finita para márgenes de error de 1%, 2%, 3%, 4%, 5%, 10%, en la hipótesis de p = 50%
  - Probabilidades asociadas con valores tan pequeños como los valores observados de U en la prueba de Mann-Whitney
  - Tabla de los valores críticos de U en la prueba de Mann-Whitney. Valores críticos de U para una prueba de una cola en a = 0,001o para una prueba de dos colas en a = 0,002.
  - Percentillas de las distribuciones (x^2/g.L.)
  - Distribución F
  - Coeficientes de Spearman para la correlación de rango
  - Prueba de Wilcoxon
  - Distribución binomial
- **FÓRMULAS**
  - Medidas de posición, dispersión, asimetría y apuntamiento
  - Fórmulas aplicadas a la regresión y correlación
  - Fórmulas para series de tiempo
  - Fórmulas aplicadas a los números índices
  - Fórmulas para inferencia estadística
- **BIBLIOGRAFÍA**

GLOSARIO

ABSCISA: Es una de las coordenadas en un plano cartesiano, correspondiente al eje horizontal (la otra se denomina ordenada).

AFIJACIÓN: Corresponde al número de elementos asignados en una muestra, cuando ellos provienen de una población estratificada.

AFIJACIÓN IGUAL: Cuando se asignan igual número de elementos en cada una de las muestras estratificadas.

AFIJACIÓN PROPORCIONAL: Cuando los elementos se asignan en la muestra estratificada en la misma proporción en que se distribuyen en la población.

AFIJACIÓN ÓPTIMA: Los elementos se asignan en la muestra estratificada teniendo en cuenta el grado de variabilidad del elemento en cada estrato y en el costo que representa en la investigación. (es conocido como Afijación de Newman).

AJUSTE: Proceso general para representar a uno de los componentes de una serie cronológica denominada *tendencia*. Se habla de ajuste rectilíneo, parabólico, exponencial, logarítmico.

ALEATORIO O AL AZAR: Cuando todos los elementos o unidades que constituyen o conforman la población investigada, tienen la misma posibilidad de ser seleccionados. Generalmente la palabra "azar" implica que el proceso considerado es, en algún sentido probabilístico.

AMPLITUD: Diferencia entre el mayor y menor de un conjunto de valores. Se considera que la amplitud del intervalo, es la diferencia entre el valor mayor y menor en cada uno de ellos, en una distribución de frecuencias. También se considera como una medida de dispersión, denominada "oscilación".

ANÁLISIS DE REGRESIÓN: Consiste en determinar la relación o dependencia que pueda existir entre dos variables, se habla también de análisis de covarianza o variabilidad conjunta de dos variables.

ANÁLISIS DE VARIANZA: Se utiliza la distribución F para el análisis de varianza. Conocida en forma abreviada ANOVA, se basa en probar la hipótesis si las medias de varias muestras provienen de poblaciones que presentan la misma media. Se puede decir que, es la comparación que se realiza suponiendo que tienen la misma varianza.

ASIMETRÍA: Se refiere a las distribuciones que no son simétricas con respecto a la ordenada central. Si una distribución unimodal tiene una "cola" más alargada hacia valores menores de la variable aleatoria, se dice que es asimétrica negativa (hacia la izquierda); en caso contrario, será de asimetría positiva (a la derecha).

ASOCIACIÓN: Es el grado de dependencia o de independencia entre dos o más variables ya sean tomadas en forma cualitativa o cuantitativa.

ATRIBUTO: Característica cualitativa, haciendo una distinción con el término "variable", pues en el primero los datos se expresan mediante palabras (sexo, ocupación, profesión, etc.) o categorías y en el segundo se expresan en forma numérica indicando el valor que toma la característica.

B

BASE: Punto de referencia, siendo el denominador al calcular un número índice. También en una serie de tiempo o cronológica se le toma como punto de partida en la contabilización de los períodos transcurridos en la serie. Hay indices con base fija y con base variable.

BASE FIJA: Ver índice con base fija.

BASE VARIABLE: Ver índice con base variable.

BATERÍA DE PRUEBAS: Término utilizado en psicología, aplicada para designar un grupo de pruebas a las cuales son sometidos algunos sujetos.

BATERÍA PONDERADA: Corresponde a grupos de pruebas psicológicas o educacionales, cuyo peso o importancia denominada ponderación, se le aplica al puntaje obtenido de los sujetos (personas).

BIAS: Es un error sistemático que distorsiona el resultado estadístico, diferente al error aleatorio. A veces es originado por el empleo de ponderaciones incorrectas, también, se refiere a los errores que cometen los empadronadores. (ver errores ajenos al muestreo).

BIMODAL: Distribuciones de frecuencias que presenta dos modas o valores modales (ver moda).

C

CADENA DE MARKOFF: Expresión utilizada en dos sentidos diferentes (a) un proceso (Xt) es llamado como cadena, si el parámetro tiempo es discontinuo: (b) toma esa denominación si los valores de X son discontinuos. El primero mencionado parece preferible esta expresión.

CADENA DE RELATIVOS: Ver encadenamientos de índices con base variable.

CARACTERÍSTICA: Es aquello que se estudia de un elemento, que puede ser cualitativa (expresada mediante palabras) denominada "Atributo" o cuantitativa (expresada en forma numérica) denominada "variable".

CARTOGRAMA: Representación gráfica por medio de mapas, mediante el cual se presentan los datos conforme a distribuciones geográficas.

CATEGORÍA: Es la forma en que se pueden agrupar o clasificar las observaciones en un grupo de frecuencias.

CENSO: Cuando se investiga a la totalidad de los elementos o unidades que constituyen una población objetivo. Ver investigación total.

CHI - CUADRADO: Método de estimación basado en la prueba de la bondad de un ajustamiento, mediante el cual se determinan valores de los parámetros que hacen mínimo el calculado, por medio de las frecuencias observadas (reales) y las frecuencias esperadas (teóricas) expresadas en término de los parámetros.

CLASE: Utilizada en las distribuciones de frecuencias, el agrupar los datos en intervalos, por lo tanto cada uno de ellos tendrá limites inferior y superior en cada clase, que algunos denominan como fronteras de clase. Se habla de intervalos de clase y de frecuencias de clase. (Ver intervalos de clase).

CLASES EXTREMAS ABIERTAS: Cuando no están definidos el primero y el último intervalo de clase, lo cual representa inconveniente en la aplicación de la mayoría de las medidas, permitiendo con frecuencia la utilización de la mediana y en algunos casos el Modo.

CLASIFICACIÓN: Algunos lo denominan como tablas o cuadros. De acuerdo a la forma como se

clasifican las observaciones. Se especifica de una sola entrada o simple; y de múltiples entradas cuando son más de dos características las que se están organizando.

COEFICIENTE DE CONCENTRACIÓN: Corresponde al coeficiente de GINI (1912) utilizada como medida de dispersión.

COEFICIENTE DE CONFIANZA: Medida de la probabilidad asociada a un intervalo de confianza, dentro del cual deberá estar el parámetro con cierto grado de probabilidad.

COEFICIENTE DE CORRELACIÓN: Denominado también como coeficiente de correlación de PEARSON, simbolizado por r o R y nos determina el grado de relación o correlación entre dos variables. Cuando el valor se acerca a 1 o –1, nos estará indicando una fuerte relación positiva o negativa. Por el contrario, al acercarse a cero, nos indicará falta de relación.

COEFICIENTE DE CORRELACIÓN MÚLTIPLE: Aplicada en la regresión múltiple, simbolizado por R. Nos determina el grado de asociación que puede existir entre más de dos variables.

COEFICIENTE DE PARCIAL: Es una forma de determinar la relación en la variable dependiente, que se explica por cada una de las variables independientes, a la vez que se mantienen constante estas variables.

COEFICIENTE DE CORRELACIÓN DE LOS RANGOS: Fue propuesto por Spearman (1906) y corresponde su aplicación a distribuciones no paramétricas. Es una medida de correlación o grado de asociación entre dos conjuntos de rangos, siendo la única manera de calcular la correspondencia entre ellos, basados en datos ordenados de dos variables.

COEFICIENTE DE DETERMINACIÓN: Término equivalente a coeficiente de correlación al cuadrado. El símbolo utilizado es R^2. Se obtiene al dividir la varianza explicada de la variable Y por la varianza de la misma variable. Se considera como la parte de la variación dependiente Y, explicada por la ecuación de la regresión.

COEFICIENTE DE REGRESIÓN: Denominado como pendiente de la línea, nos determina el ángulo de inclinación, además se podrá decir que nos da el crecimiento o decrecimiento en la variable dependiente, por cada unidad que toma la variable independiente.

COEFICIENTE DE VARIACIÓN: Medida de dispersión relativa, resulta al dividir la desviación típica o estándar por la media aritmética de esa misma distribución, para comparar la variabilidad relativa entre dos o más distribuciones que están dadas en unidades de medición diferentes; suele aparecer multiplicado por 100, expresándose en términos porcentuales. Fue propuesto por K. Pearson (1895).

COEFICIENTE ANGULAR: Ver coeficiente de regresión.

COEFICIENTE DE POSICIÓN: En la ecuación de la recta Y=bx+c, en la cual **b** es el coeficiente angular (β_1) y **c** el coeficiente de posición u origen en la ordenada (β_0), es decir, su valor se ubica en la ordenada, cuando utilizamos un plano cartesiano.

COMIENZO AL AZAR: En la selección de una muestra sistemática, cuando los datos están ordenados de menor a mayor (en cantidad, valor, tiempo), se hace a intervalos regulares, suele ser conveniente elegir el primer elemento o unidad de la muestra, como punto de arranque, mediante una selección al azar entre las primeras unidades que conforman el primer intervalo de selección de aquella población. Ver: muestreo sistemático.

COMPONENTES DE UNA SERIE DE TIEMPO: Son cuatro (4) los componentes: Tendencia, Variación estacional, Variación cíclica y Variación aleatoria. Para algunos es la suma Y=T+ VE+VC+VA; para otros es el producto: Y=T x VE x VC x VA y para otros la combinación: Y= T(VE+VC+VA).

CONCORDANCIA: Puede darse en la relación entre variables cuantitativas o cualitativas que tiene modalidades comparables, estableciendo la intensidad de dicha relación.

CONGLOMERADO: Grupo de elementos de gran variabilidad en su interior, pero de gran homogeneidad si son comparados entre sí.

CONSISTENTE: Es otro requisito para que se tenga un buen estimador, su valor será cercano al parámetro, cuando el tamaño de la muestra es grande (n>30).

CONTINGENCIA: Es la diferencia entre las frecuencias reales u observadas y las teóricas o esperadas, consignadas en una tabla de contingencia.

CONTROL DE CALIDAD: Método para controlar la calidad de un producto manufacturado que se produce en grandes cantidades. Su objetivo es detectar y eliminar alteraciones sistemáticas de la calidad o reducirlas a un nivel aceptable.

CORRECCIÓN AL MUESTREO DE POBLACIONES FINITAS: Es aplicado en el cálculo de los límites de confianza, cuando el tamaño de la muestra es bastante pequeño, es decir, si la fracción de muestreo (n/N) es mayor o igual al 5%.

CORRECCIÓN DE YATES: Denominado también como corrección de continuidad, ajuste propuesto por Yates (1934) en el cálculo de Chi-cuadrado en tablas de contingencias de "2x2".

CORRELACIÓN: En el sentido más general, significa la interdependencia entre datos cuantitativos o cualitativos, pudiendo ser extendido a más de dos variables aleatorias. Frecuentemente nos referimos a la relación entre variables aleatorias o entre rangos.

CORRELACIÓN CURVILÍNEA: Expresión para denotar a la correlación en datos bivariantes (dos variables) cuando las regresiones no son lineales. Algunos no están de acuerdo, pues la correlación está dada por un número y no por una función que admite curvatura.

CORRELACIÓN DE LOS RANGOS: Mide la intensidad de la correlación entre dos conjuntos de ordenamientos y el grado de correspondencia entre ellos. Los más conocidos son los de Kendall (1938) y el de Spearman (1904).

CORRELACIÓN DIRECTA: Ver Coeficiente de correlación.

CORRELACIÓN LINEAL: No es aconsejable esa expresión, utilizada muy frecuente en funciones lineales de las observaciones.

COVARIACIÓN: Variación conjunta de dos o más variables.

COVARIANZA: Determina la variación conjunta entre dos variables. Definido como la media aritmética del producto de las desviaciones entre los valores que toman las variables y sus medias aritméticas. Puede tomar signo positivo o negativo indicándonos una relación positiva o negativa.

CUADRO SIMPLE: Se refiere a una tabla o cuadro que presenta la variable y sus frecuencias. Una distribución unidimensional (una variable) de frecuencias es un ejemplo. Es preferible evitar esta expresión en un sentido técnico.

CUARTILES: Se pueden calcular tres cuartiles: (Q) cuartil uno; (Q_2) cuartil dos, siendo igual a la mediana y el (Q_3) cuartil tres. El total de la frecuencia (n) se divide en 4 partes. El primer cuartil nos indica un valor de la variable que supera al 25% de las observaciones y es superado por el 75% del mismo.

CUESTIONARIO: Grupo o sucesión de preguntas diseñado para obtener información sobre uno o varios puntos a los que habrá de responder el informante. Ver formulario.

CURTOSIS: Determina el grado de "apuntamiento" o "achatamiento" de una curva unimodal normal. Fue propuesto por K. Pearson en (1906).

En una distribución normal el resultado deber ser igual a 3. En el caso de que sea menor que 3 será achatada o platicúrtica, si es normal o igual a 3 se denominará mesocúrtica; si es mayor que 3 será apuntada o leptocúrtica.

CURVA NORMAL: Es una curva de frecuencias simétrica, tiene una figura acampanada. Algunos la denominan como: curva de error, de probabilidad o campana de Gauss. Su función de probabilidad se encuentra determinada por un valor central (μ) y la desviación estándar (σ).

CURVA DE COMPERTZ: Denominada también como curva de crecimiento.

CURVA DE LORENZ: Es un método gráfico empleado para mostrar la concentración de la propiedad de cantidades económicas, como por ejemplo ingreso. En la ordenada, se representan las cantidades de las variables y en la abscisa la frecuencia acumulada de la población que posee esas cantidades.

D

DATOS AGRUPADOS: Ver tablas de frecuencias o distribución de frecuencias.

DATOS CUALITATIVOS: Ver Atributo.

DATOS CUANTITATIVOS: Denominados también como variables, se refiere a aquellas características expresadas en forma numérica, tales como medidas (peso, altura, precios, etc.) o recuentos (número de hijos, hogares, autos, etc.)

DATOS ESTADÍSTICOS: Datos numéricos referidos a un agregado de elementos individuales, que la estadística compila, analiza e interpreta.

DECILES: Uno de los nueve valores de una variable en que se divide la frecuencia total en diez partes iguales.

DECISIÓN: Ver Función de decisión.

DECISIÓN ESTADÍSTICA. La prueba de una hipótesis estadística implica la decisión entre dos valores: aceptar o rechazar. En un plan de muestreo, en cada una de sus etapas habrá de adoptarse una de las tres decisiones: aceptar, rechazar o continuar.

DEFECTOS ADMISIBLES: En control de calidad, se refiere al número de defectuosos admisibles en una muestra, como el número crítico tal que si es excedido, la totalidad restante del lote deberá ser examinado o bien rechazar el lote. Ver Número de aceptación.

DEPENDENCIA CAUSAL: Cuando dos variables tienen como consecuencia una relación de causa y efecto.

DEPENDENCIA UNILATERAL: Es una regresión unilateral, donde se puede estimar la variable Y en función de X, pero no viceversa, tal como sucede en la regresión bilateral.

DESCOMPOSICIÓN DE UNA SERIE DE TIEMPO: Por medio de métodos estadísticos se suele separar una serie (cronológica) en sus partes: la tendencia, las variaciones cíclicas, estacionales y aleatorias. Cualquier serie en particular no necesariamente está constituida por todos estos componentes.

DESIGUALDAD DE TCHEBICHEFF – BIENAYME: Obtenida por Bienayme en la mitad del siglo XIX. Generalizada por Tchebicheff. Ver Teorema de Tchebicheff.

DESPLAZAMIENTO CRONOLÓGICO: Diferencia de tiempo por la cual una observación precede (es anterior) a otra.

DESVIACIÓN: Son las diferencias que hay entre los valores que toma la variable y un punto fijo que puede ser la media u otro valor, denominado media teórica.

DESVIACIÓN ABSOLUTA: Valor absoluto, sin considerar el signo de la diferencia entre los valores de la variable y su media aritmética.

DESVIACIÓN CUARTIL: Ver Desviación intercuartílica.

DESVIACIÓN ESTÁNDAR: Es una medida de dispersión, denominada también como desviación típica, siendo obtenida, mediante la raíz cuadrada de la varianza.

DESVIACIÓN ESTÁNDAR O TÍPICA: Es la medida de dispersión más extensamente empleada en una distribución de frecuencias. Es igual a la raíz cuadrada positiva de la varianza.(σ)

DESVIACIÓN ESTÁNDAR PORCENTUAL: Ver coeficiente de variación.

DESVIACIÓN INTERCUARTÍLICA: Es otra medida de dispersión, tomando la diferencia que hay entre el cuartil tres (Q_3) y el cuartil uno (Q_1), siendo dividida entre dos. Se le conoce también como amplitud semi-cuartílica.

DESVIACIÓN MEDIA: Medida de dispersión, derivada de la desviación promedio de las observaciones con respecto a la media, tomados en valor absoluto. Es la media aritmética, de las diferencias entre los valores de la variable y su media aritmética, tomadas en valor absoluto.

DESVIACIÓN MEDIANA: Es la media aritmética de las diferencias entre los valores que toma la variable y su mediana, considerando las mismas en valor absoluto.

DIAGRAMA DE BLOQUES. Ver: Histogramas de frecuencias.

DIAGRAMA DE DISPERSIÓN: Denominado nube de puntos; es una gráfica donde se representan los datos de dos variables y cada par de datos se describen mediante puntos.

DIAGRAMA DE ÁRBOL: Es una forma gráfica, que nos permite elaborar un espacio muestral, donde aparecen los puntos muestrales de dos o más etapas.

DIAGRAMA PASTEL: Nombre que se le da a la gráfica circular.

DEVALUACIÓN: Es la pérdida de poder adquisitivo del peso o unidad monetaria de un país con relación a precios de mercado externo, como por ejemplo frente al dólar o al euro.

DESVALORIZACIÓN: Es la pérdida de poder adquisitivo, que sufre el peso con relación a precios internos de mercado.

DEFLACTAR: Consiste en transformar valores corrientes (a precios de mercado) a precios constantes, con respecto a un período determinado. Denominado tambien como INDEXAR.

DIAGRAMAS DE FRECUENCIAS: Son gráficas, generalmente utilizadas en la teoría estadística, para representar la variable discreta.

DIAGRAMA DE DISPERSIÓN: Denominado también como "nube de puntos", corresponde a una gráfica de un conjunto de puntos que representa a dos variables relacionadas y cada punto representa a un par de observaciones. También es el grado de desparramiento que muestran las observaciones, las cuales se miden como promedio de las desviaciones respecto a algún valor central (desviación media, desviación típica) o por medio de una estadística de orden (desviación cuartílica o amplitud).

DISTRIBUCIÓN ASIMÉTRICA:
Ver asimetría.

DISTRIBUCIÓN BIDIMENSIONAL: Llamado también como distribución bivariante, ya que muestra dos variables relacionadas.

DISTRIBUCIÓN BIMODAL: Distribución de frecuencias con dos modos.

DISTRIBUCIÓN BINOMIAL: También se le denomina distribución Bernoulli por la participación que tuvo en su desarrollo Santiago Bernoulli en 1713. Utilizado para describir los resultados de un experimento conocido también como proceso de Bernoulli, donde cada ensayo tiene dos resultados posibles.

DISTRIBUCIÓN CONTINUA: Se dice que una variable es continua cuando puede tomar todos los valores en un cierto intervalo.

DISTRIBUCIÓN DE BERNOULLI: Sinónimo de distribución binomial.

DISTRIBUCIÓN DE CHI-CUADRADO: Llamado también Ji-cuadrado, distribución dada aparentemente por primera vez, por Helmet en 1875 y redescubierto por Karl Pearson en 1900. Ver: Chi-Cuadrado.

DISTRIBUCIÓN DE FRECUENCIAS: Denominadas también como tablas o cuadros de frecuencias, es una forma de concentrar grandes cantidades en forma resumida donde se muestra el número de veces (frecuencia) que se repite cada valor que toma la variable.

DISTRIBUCIÓN DE LAPLACE: Se la conoce bajo el nombre de Primera Ley de Laplace, distinguiéndola de la Segunda Ley, que es la misma que da la distribución normal.

DISTRIBUCIÓN F: Tiene varias aplicaciones, una de ellas es la realización de una prueba que nos permite comparar dos varianzas para determinar si son iguales; la otra, para determinar si tres o más tratamientos, de información muestral, producen idénticos resultados.

DISTRIBUCIÓN "t": Esta distribución debida originalmente a "Student" (1980) aparece frecuentemente en la forma modificada por R.A. Fisher (1952), utilizado en pruebas de hipótesis y en intervalos de confianza para el estimado de la media poblacional, cuando no se conoce la desviación típica poblacional y además la muestra es pequeña, siendo $n < 30$.

DISTRIBUCIÓN T: Nombre alternativo para la T de Hotelling.

DISTRIBUCIÓN DE MUESTREO: Por lo general, se refiere casi siempre a un esquema de muestreo que implique selección al azar.

DISTRIBUCIÓN DE POISSON: En una distribución binomial, cuando del valor de presentación de un evento es tan pequeño que tiende a cero y además, el número de sucesos es tan grande (generalmente mayor que 50) que se le considera como un suceso de poca ocurrencia, en estos casos, es aconsejable la distribución de Poisson. También en aquellos problemas que involucran espacio, ya sean áreas o tiempo, siendo de gran aplicación.

DISTRIBUCIÓN DE PROBABILIDAD: Es un listado o resultados que se obtienen de un experimento en condiciones normales, donde las probabilidades obtenidas están asociadas con cada resultado.

DISTRIBUCIÓN DE STUDENT: Ver Distribución de la t.

DISTRIBUCIÓN DE UNA VARIABLE: Denominada también como Distribución Unidimensional, es decir, en ella se analiza una sola variable.

DISTRIBUCIÓN HIPERGEOMÉTRICA: Cuando se desarrolla un experimento, se obtiene como resultado las probabilidades para cada uno de los X éxitos en n ensayos, siendo éstos dependientes, para ello utilizamos la función hipergeométrica.

DISTRIBUCIÓN NORMAL: Se le suele denominar Gaussiana, Laplaciana, distribución de Laplace-Gauss o de Gauss-Laplace o bien la segunda Ley de Laplace. Aparentemente fue descubierta por De Moivre como forma límite de la distribución binomial.

DISTRIBUCIÓN BINOMIAL: Cuando desarrollamos un experimento se obtienen las probabilidades para cada uno de los X éxitos en n ensayos independientes, para ello utilizamos la función de probabilidad binomial.

DISTRIBUCIÓN PLURIMODAL: Distribución de frecuencias con más de dos modas.

DISTRIBUCIÓN SIMÉTRICA: Cuando los valores de las frecuencias absolutas y relativas que toma la variable son equidistantes a un valor central. En este caso, la Media, la Mediana y Moda son iguales.

E

ECUACIONES NORMALES: Sistema de ecuaciones al que se llega en una estimación por el método de los mínimos cuadrados.

EMPALME DE NÚMEROS ÍNDICES: En una serie de números índices, muchas veces hay necesidad de suspender una serie, por la poca representatividad que tiene, debido a la aparición de nuevos ítems o desaparición de los que se han venido utilizando, lo cual hace necesario una nueva, con una nueva base. Cuando se hace necesario continuar con el suspendido, se realiza un empalme, (lo mismo para el nuevo), a fin de obtener los índices con la nueva base, realizados mediante una regla de tres simple.

ENCUESTA: Utilizado frecuentemente para denotar a las encuestas por muestreo.

ENCUESTA DE OPINIÓN: Encuesta por muestreo, por medio de la cual se intenta conocer las opiniones de los miembros de una población humana dada.

ENCUESTA EXPLORADORA: Ver Encuesta preliminar o piloto.

ENCUESTA PILOTO: Encuesta generalmente de pequeña escala, llevada con anterioridad a la encuesta principal. Su objetivo principal es aumentar la eficiencia de la investigación, mediante el conocimiento de la reacción de los informantes; problemas en el cuestionario; tiempo promedio por entrevista; familiarización del entrevistador, determinar el tamaño más efectivo de la unidad de muestreo. Se le conoce como encuesta exploradora; encuesta preliminar o pretest.

ENCUESTA POR MUESTREO: Llevada a cabo mediante el empleo de métodos de muestreo, donde sólo se toma una parte de la población.

ENSAYO: Es un intento deliberado de generar un suceso cuya ocurrencia se supone sometida a un esquema probabilístico. Puede ser definido como un experimento controlado, cuyo resultado es de naturaleza cierta.

ENSAYOS A LO BERNOULLI: Sucesión de acontecimientos, ejemplo: lanzamientos de una moneda, siendo sucesos independientes, con probabilidad de "éxito" constante. La distribución de éxitos está dada por la distribución binomial. Ver distribución de Bernoulli y Distribución Binomial.

ENSAYOS INDEPENDIENTES: Cuando la probabilidad del resultado de cualquiera de ellos es independiente de los resultados de los demás, donde la probabilidad es la misma para todos los ensayos.

ERROR: Tiene el mismo sentido que le damos en la vida diaria. En estadística se emplea como la diferencia que puede haber entre el valor esperado y el verdadero.

ERROR AJENO AL MUESTREO: Estos errores pueden surgir por distintos aspectos como: defectos en el marco; la fijación de las unidades de muestreo, como en su selección; errores en el formulario; en la recolección, en la tabulación, negligencia del personal en todo el proceso de la investigación. Son errores sistemáticos que no se pueden cuantificar.

ERROR DE ESTIMACIÓN: En general, es la deficiencia que puede haber entre el valor del estimador y el parámetro (valor verdadero en la población).

ERROR DE INDAGACIÓN: Ver Error ajeno al muestreo.

ERROR DE MUESTREO: Parte de la diferencia entre un valor poblacional (parámetro) y una estimación de la misma, obtenida a través de una muestra aleatoria.

ERROR DE OBSERVACIÓN: Error que surge de imperfecciones en el método para observar, sean éstas debidas a factores humanos o instrumentales.

ERROR DE RESPUESTA: Ver Error ajeno al muestreo.

ERROR DEL TIPO I: Se presenta en una prueba de hipótesis, cuando rechazamos Ho, siendo cierto o verdadero.

ERROR DEL TIPO II: Cuando aceptamos una hipótesis que es falsa.

ERROR ESTÁNDAR: Raíz cuadrada positiva de la varianza de la distribución muestral, es decir, la desviación muestral de un estimador puntual.

ERROR ESTÁNDAR DE UNA ESTIMACIÓN: Hace referencia a la raíz de la varianza residual en una línea de regresión.

ERRORES EN LAS ENCUESTAS: Surgen por efectos del muestreo y de otros no relacionados con el muestreo.

ESPACIO MUESTRAL: Es un conjunto de elementos o sucesos posibles que se obtiene en una prueba. También se puede definir como un conjunto de puntos muestrales posibles, como resultados de un experimento.

ESPERANZA MATEMÁTICA: El valor esperado de una función de variables aleatorias, es su valor medio, no necesariamente lo más frecuente.

ESTADÍSTICA: Valor sumatorio calculado por medio de una muestra de observaciones, aunque no necesariamente, como estimados de algún parámetro.

ESTADÍSTICA CHI-CUADRADO: Ver: Chi-cuadrado y Distribución Chi-cuadrado.

ESTADÍSTICA DESCRIPTIVA: Métodos, procedimientos, que permiten recolectar, organizar, clasificar, cuantificar, los cuales son presentados por medio de cuadros, gráficas, la obtención de valores que son resultados de aplicar algunas medidas que permiten describir el comportamiento de un hecho o conjunto de observaciones.

ESTEREOGRAMAS: Diagramas que intentan representar una figura tridimensional sobre una superficie plana. Diagrama que representa las frecuencias de una distribución de dos variables.

ESTIMACIÓN: Es el valor obtenido al calcular un estimado.

ESTIMACIÓN DE MARKOFF: Estimación de un parámetro derivado de un estimador obtenido mediante el teorema de Markoff.

ESTIMACIÓN POR EL MÉTODO DE REGRESIÓN: Es la estimación del valor de variable Y (dependiente) en función de otra variable (X), supuestamente conocida (independiente).

ESTIMACIÓN POR INTERVALO: Son límites de confianza (inferior o superior) dentro del cual supuestamente debe estar el parámetro (valor verdadero) con cierto grado de confianza.

ESTIMACIÓN PUNTUAL: Es un punto o valor de la variable, considerado como el que representa mejor al parámetro. Es un valor numérico, utilizado como estimador del valor estadístico de la población.

ESTIMADOR: Es una medida aplicada a las características de los elementos o unidades en una muestra. Algunos la denominan como medida estadística o estadístico de la muestra.

ESTIMADOR CONSISTENTE: Al aumentar el tamaño de la muestra, converge en probabilidad al parámetro que estima.

ESTIMADOR EFICIENTE: Se refiere a aquel estimador que tiene la menor varianza entre todos los estimadores posibles.

ESTIMADOR INSESGADO: Es aquella muestra que no tiene error, es decir, que el valor del parámetro es exactamente igual al valor del estimador.

ESTRATIFICACIÓN: Consiste en dividir la población en subpoblaciones, grupos o estratos de tal manera que cada grupo sea lo más homogéneo posible, pero heterogéneo si comparamos los estratos entre sí.

ESTRATO: Ver Estratificación.

EXPERIMENTO: Es un conjunto de pruebas que originan resultados bien definidos.

EXPERIMENTO BINOMIAL: Requiere (a) Que se tenga un número n de ensayos repetidos; (b) solamente se tienen dos resultados (éxito o fracaso); (c) los resultados son independientes; (d) la probabilidad de éxito (p) es fijo.

EXPERIMENTOS DE QUETELET: Consiste en la realización de pruebas, cada vez más grande, observando que a medida que aumenta o disminuye la probabilidad, que al principio fluctúa, se va estabilizando de tal manera que se le considera como la probabilidad de éxito de un suceso en un solo ensayo.

F

FACTOR DE CORRECCIÓN PARA POBLACIONES FINITAS: Se aplica cuando la fracción de muestreo es mayor del 5%, siendo igual a raíz cuadrada de N-n / N-1.

FACTOR DE EXPANSIÓN: Número de veces que se expande el tamaño de la muestra para ser igual al tamaño poblacional.

FRACCIÓN DE MUESTREO: Corresponde al porcentaje de la población, que fue tomado como muestra.

FRECUENCIA ABSOLUTA: El número de veces que se repite cada valor que toma la variable.

FRECUENCIA ABSOLUTA ACUMULADA: En una tabla de frecuencias como su nombre lo indica, son acumuladas las frecuencias absolutas.

FRECUENCIA RELATIVA: Corresponde a cada una de las frecuencias absolutas que toma la variable, dividida por el total de observaciones ya sea en la población o en la muestra.

G

GRADOS DE LIBERTAD: Corresponde al número de valores que pueden ser asignados libremente. Es utilizado en la distribución "t" de Student cuando la muestra es $n \leq 30$ y además no se conoce la varianza poblacional y utilizamos la obtenida en la muestra.

GRÁFICO CIRCULAR: Denominado como pastel "torta", es una representación de datos cualitativos por medio de sectores dentro de un círculo, siendo cada sector el valor proporcional a los componentes del total.

GRÁFICO DE BARRAS: Es una gráfica de frecuencias representadas por medio de barras o rectángulos, trazados con alturas proporcionales. Puede ser de barras simples o compuestas, en forma vertical u horizontal.

GRÁFICO DE CONTROL: Recurso gráfico empleado para mostrar resultados obtenidos de muestras repetidas, provenientes de un proceso de producción.

GRÁFICO LOGARÍTMICO: Gráfico en el cual ambos ejes (ordenada y abscisa) están tomados en valores logarítmicos, correspondientes a las variables X y Y.

GRÁFICO LINEAL: Cuando un gráfico representa las variables por medio de líneas poligonales.

GRÁFICO SEMI-LOGARÍTMICO: Forma de representación gráfica en la que uno de los ejes está dado en escala logarítmica y el otro en forma aritmética.

H

HIPÓTESIS ALTERNATIVA: Correspondiente a la prueba de hipótesis, siendo la alternativa del parámetro sometida a prueba, la cual se considera verdadera cuando rechazamos la hipótesis nula.

HIPÓTESIS ESTADÍSTICA: Es un supuesto que se hace acerca de un valor estadístico de la población, denominado parámetro.

HIPÓTESIS NULA: Es aquella hipótesis que se quiere probar, la cual consideramos como verdadera, distinguiéndola de la hipótesis alternativa.

HISTOGRAMA: Diagrama de frecuencias de una variable (unidimensional) continua, representado por rectángulos de áreas proporcionales a las frecuencias de clase con el eje vertical, sobre los intervalos representados en el eje horizontal.

HOMOGENEIDAD: Cuando el grado de variabilidad tiende a cero.

I

INDEPENDENCIA: Dos sucesos son independientes, si la probabilidad de presentación de uno de ellos es la misma, esté o no dado el otro.

ÍNDICE: Es una cifra relativa expresada en términos porcentuales, que se emplea para indicar la variaciones en los precios, cantidades o valores para un período, con respecto a un período base.

ÍNDICE AGREGATIVO SIMPLE: Es el cuociente de dividir la suma de precios o cantidades de un período, por la suma de los precios o cantidades de un período base y luego multiplicado por 100.

ÍNDICE COMPARATIVO DE MORTALIDAD: Es una tasa de mortalidad promedio. Se puede considerar como un número índice de la forma Marshall - Edgeworth (en el libro aparece de Marshall).

ÍNDICE DE CANTIDADES: Indica variaciones en las cantidades de un período con respecto a un período base.

ÍNDICE DOW JONES: Es un índice que indica las variaciones del precio de las acciones de un grupo industrial que se negocia en la Bolsa de Nueva York.

ÍNDICE DE LASPEYRES: Se tienen índices de precios y de cantidad. En el primero nos determina las variaciones en los precios, manteniendo como ponderaciones las cantidades del período base. En el segundo indica las variaciones en las cantidades manteniendo como ponderaciones los precios del período base.

ÍNDICE DE DROBISCH – SIDEWICK: Este índice es el promedio entre los índices de precios de Laspeyres y el de Paasche. También puede calcularse el promedio de los índices de cantidad, tomando al de Laspeyres y el de Paasche.

ÍNDICE DE KEYNES: Se tienen índices de precios y de cantidad. El primero trata de indicar las variaciones en los precios, comparando los períodos, tomando como ponderaciones las cantidades más pequeñas observadas; en cambio el segundo, indica las variaciones en las cantidades, ponderadas por los precios más pequeños.

ÍNDICE DE FISHER: En realidad es una media geométrica. Se obtiene al calcular la raíz cuadrada, del producto entre el índice de Laspeyres y Paasche. Fue establecido por Irving Fisher en 1927, al principio se consideraba como la fórmula ideal, ya que era la única que cumplía con ciertos criterios que él había establecido.

ÍNDICE DE MARSHAL: Se calculan índices de precios y de cantidad, en los cuales las ponderaciones, corresponden a la suma de los precios (tanto del período base, como del investigado) en el índice de cantidad; y la suma de las cantidades como ponderaciones, en el índice de precios.

ÍNDICE DE PAASCHE: Indica las variaciones en los precios, estando ponderados por las cantidades del período investigado. También, indica las variaciones en las cantidades, cuando están ponderados por los precios del período investigado.

ÍNDICE DE PARETO: Se le denomina así al coeficiente que aparece en la expresión de una curva de Pareto. Proporciona la evidencia de la concentración de las rentas (ingresos) o con más generalidad en la concentración de valores de la variable, en distribuciones de la forma de Pareto.

ÍNDICE DE PRECIOS AL CONSUMIDOR: IPC, es un índice mensual que trata de indicar las variaciones en los precios de los artículos de primera necesidad (bienes y servicios) de un período con respecto a un período base.

ÍNDICE DE PRECIOS AL PRODUCTOR: IPP, es un índice mensual, empleado para indicar las variaciones en los precios de los artículos que directamente vende el productor.

ÍNDICE DE VALORES: Se obtiene al dividir la sumatoria de valores (precio por cantidad) correspondiente al período investigado, por la sumatoria (agregado) de valores correspondiente al período base, multiplicando el resultado por cien. Generalmente se le considera como un relativo de valores.

INFERENCIA ESTADÍSTICA: Consiste en realizar investigaciones pequeñas, denominadas muestras, con las cuales se obtienen valores considerados como estimados de los valores en una población.

INSESGO: Es una de las propiedades que debe reunir un buen estimador, cuando éste es igual al valor del parámetro que es estimado.

INTERVALOS DE CONFIANZA: Cuando se establecen los límites de confianza, es decir, los valores superior e inferior dentro del cual deberá estar el parámetro con cierto grado de confianza, de esta manera se establece el intervalo. Significa que no hay error.

INTERVALO DE MUESTREO: Corresponde al muestreo sistemático.

INTERVALO DE PREDICCIÓN: Intervalo comprendido entre los límites superior e inferior, asociados a un valor estimado o pronosticado.

LEPTOCÚRTICA: Corresponde a las medidas de apuntamiento o curtosis. Se le denomina así a aquella curva más apuntada que la normal.

LÍMITES DE CONFIANZA: Los valores de Ls y Li (límites superior e inferior) de los intervalos de confianza.

LÍMITES DE PROBABILIDAD: Límites superior e inferior asignados a un valor estimado,

con el objeto de indicar el intervalo dentro del cual se supone comprendido el valor poblacional, con cierta probabilidad de que así suceda.

LÍNEA DE REGRESIÓN: Expresión sinónima de curva de regresión, usada generalmente, para denotar a la regresión lineal.

LOTE: Término utilizado en control de calidad en el sentido agregado o colección. Es un conjunto de unidades de un artículo producido bajo condiciones similares.

LOTE DE INSPECCIÓN: Lote que se va a inspeccionar, el cual se puede realizar sobre cada elemento o sobre una muestra de ellos.

MARCAS DE CLASE: Término utilizado para denotar el punto medio del intervalo de clase, es decir, la mitad entre el límite inferior y el superior de clase.

MARCO: Lista o mapa en el cual se identifica con claridad el elemento que pertenece a una población que se va a investigar. Existe un manejo a cada etapa de muestreo, en un plan de muestreo por etapas.

MARGEN DE ERROR: Es aquel valor que se le suma y resta al valor del estimador (estimador puntual), con lo cual fijamos los límites o intervalos de confianza.

MEDIA: Es el término utilizado para identificar al promedio aritmético, geométrico, cuadrático, cúbico, etc.

MEDIA ARITMÉTICA: Es el resultado de dividir a la suma de todos los valores observados por el número total de observaciones. Frecuentemente se omite la calificación "aritmética", de modo que cuando se menciona la palabra "media" se sobreentiende con ella "media aritmética".

MEDIA ARMÓNICA: La media armónica de un conjunto de observaciones, es la recíproca de la media aritmética del recíproco de las observaciones. Su resultado es menor que la geométrica y la aritmética.

MEDIA CUADRÁTICA: Corresponde a la raíz cuadrada de la media aritmética de los cuadrados de los valores de la variable. Su resultado es mayor que el de la media aritmética.

MEDIA CÚBICA: Es la raíz cúbica de la media aritmética del cubo de los valores que toma la variable, su resultado deberá ser mayor que el obtenido con la media cuadrática.

MEDIA GEOMÉTRICA: Es la raíz enésima del producto de los valores de la variable. Cuando ésta existe, deberá estar comprendida entre la media armónica y la media aritmética.

MEDIA MÓVIL: Denominada también como promedio móvil, utilizado en las series de tiempo, generalmente para suavizarla. Se pueden tomar valores pares o impares para promediarlos.

MEDIA PONDERADA: Es la media aritmética para un conjunto de datos agrupados en una tabla de frecuencias, donde a cada valor de la variable se le asigna una frecuencia, que viene a convertirse en un factor de ponderación.

MEDIA TEÓRICA: Denominada como media presunta, cuyo origen es arbitrario denominado como origen de trabajo, utilizada en el cálculo de las desviaciones. No es recomendable esta expresión.

MEDIA VERDADERA: Expresión alternativa, poco empleada, del parámetro o media poblacional, en contraste con la media de la muestra.

MEDIANA: Es el valor de la variable que divide a la frecuencia total en dos partes iguales. Es el

valor central de la variable, que supera la mitad de las observaciones y es superado por la otra mitad.

MEDIDA CUARTÍLICA DE ASIMETRÍA: Es una medida cuartílica de asimetría.

$$As = \frac{(Q_3 - M_e) - (M_e - Q_1)}{Q_3 - Q_1}$$

MEDIDA DE ASIMETRÍA DE PEARSON: Propuesta por Pearson. Esta medida se obtiene dividiendo la diferencia entre la media y el modo, por el valor de la desviación estándar.

MEDIDA DE POSICIÓN: Medida (promedio) que intenta, por medio de un valor central o típico, representar a un conjunto de observaciones, correspondiente a una distribución.

MESOCÚRTICA: En las medidas de apuntamiento, es la forma de denominar a la curva, cuando ella toma la forma correspondiente a la distribución normal.

MÉTODO CLÁSICO DE PROBABILIDAD: Consiste en considerar a la probabilidad como el cuociente de dividir al número de éxitos por el total de casos posibles.

MÉTODO EMPÍRICO DE PROBABILIDAD: Cuando se sabe la probabilidad de un suceso, sin necesidad de realizar el experimento.

MÉTODO SUBJETIVO: Procedimiento o método de asignar probabilidades con base en el juicio personal.

MÉTODO A MANO LIBRE: Denominado método "mano alzada" utilizado para describir una serie de tiempo, en el cual la tendencia se estima trazando a mano libre de manera que pase cerca de los punto originales. Es un método subjetivo y su validez depende del conocimiento que se tenga respecto a esa variable. Es usado en econometría.

MÉTODO DE INFORMACIÓN LIMITADA: No toma en cuenta ciertas restricciones impuestas sobre los parámetros por las ecuaciones estructurales, pero no obstante proporcionan estimadores consistentes.

MÉTODO DEL MÁXIMO DE VEROSIMILITUD: Método utilizado para la estimación de un parámetro o de varios de una población por medio del valor que maximiza (o maximizan) a la verosimilitud de una muestra.

MÉTODO DE LAS MEDIAS MÓVILES: Denominado como promedios móviles.

MÉTODO DE LOS MÍNIMOS CUADRADOS: Se refiere al hacer un ajuste rectilíneo, parabólico o exponencial, seleccionar el mejor de ellos, cuando la suma de los cuadrados de las diferencias, entre los valores de la variable y su estimado es mínimo.

MÉTODO DE LOS SEMI-PROMEDIOS: Método utilizado para obtener una estimación rápida de una recta, en el cual la serie de tiempo se divide en dos partes, de tal manera que cada una de ellas contenga un número impar de períodos, calculando la media de cada grupo y su crecimiento o decrecimiento está dado por el promedio de las diferencias entre los dos períodos.

MÉTODOS NO PARAMÉTRICOS: Cuando no se conoce o no se tiene ningún conocimiento acerca de los parámetros, se habla de métodos no paramétricos o de distribución libre y no es aplicado cuando se dispone de datos nominales y ordinales.

MODELO: Se considera como la expresión formal de una teoría o de la situación causal generadora de los datos observados.

MODELO LINEAL: En el cual las ecuaciones que relacionan a las variables o a las variables aleatorias son de forma lineal.

MODELO MULTIPLICATIVO SERIE DE TIEMPO: Cuando una serie está afectada por los cuatro componentes (Tendencia, Variaciones Cíclicas; Variaciones Aleatorias y Variaciones Estacionales), se multiplican : Y = T x VE x VC x VA.

MODA: Generalizado en 1984 por K. Pearson. Medida de posición que nos determina aquel valor de la variable que más se repite, es decir, aquélla que tiene la mayor frecuencia.

MUESTRA: Parte de una población o subconjunto de un conjunto de elementos, que resulta de la aplicación de algún proceso, generalmente selección deliberada, con el objeto de investigar las características de la población o conjunto del cual proviene.

MUESTRA ALEATORIA: Muestra que ha sido seleccionada por medio de una selección aleatoria. En ella todos los elementos tienen la misma posibilidad de ser seleccionados.

MUESTRAS APAREADAS: Cada uno de los datos de una muestra se compara con el dato obtenido en la otra muestra.

MUESTRA CON BIAS: Es aquélla que incorpora algún elemento sistemático de error, en especial, cuando se confía en el juicio subjetivo por parte de los seres humanos.

MUESTRA CON SELECCIÓN DIRIGIDA: Dentro del muestreo no aleatorio, aparece el muestreo por cuotas, de juicio o intencional, donde las unidades o los elementos son seleccionados en forma deliberada, por tal razón está sujeta a errores.

MUESTRA CON ETAPAS MÚLTIPLES: Conocida como muestreo multietápico, muestreo geográfico, muestreo por áreas y muestreo por conglomerados, cuando se tiene una población de unidades y se hace una primera selección; luego cada unidad seleccionada está compuesta por otra cantidad de unidades y se realiza una nueva selección. Cada unidad seleccionada está compuesta por otras unidades permitiendo una nueva selección, así sucesivamente hasta llegar a la unidad final o al elemento que se requiere ser investigado.

MUESTRA DEFECTUOSA: Especialmente cuando tenemos un marco muestral defectuoso es decir, que contiene elementos que no pertenecen al estudio o está incompleto por no haberse actualizado.

MULTICOLINEALIDAD: Con la aplicación de este término queremos describir la correlación que hay entre las variables independientes.

MUESTRA DE DOS ETAPAS: Denominado también como muestreo bietápico, cuando de una población se extrae una muestra aleatoria y luego se extrae de ella otra muestra, teniendo en cuenta que la unidad cambia en cada etapa de selección.

MUESTRA ESTRATIFICADA: Cuando la población ha sido dividida en grupos, estratos o subpoblaciones y luego se realiza una muestra aleatoria simple en cada uno de los estratos. Se puede hacer mediante Afijación igual, proporcional y óptima.

MUESTRA NO ALEATORIA: Cuando los elementos o unidades son elegidos caprichosamente, pudiéramos decir en forma intencional, conduciendo a errores, por tal razón se le conoce como muestreo errático o circunstancial.

MUESTRA POR CUOTAS: Se instruye a cada recolector de información que recoja un número determinado de elementos de acuerdo a su criterio personal, con base en unas consideraciones que se le han dado.

MUESTRA POR FASES: Cuando de una población se extrae una muestra lo suficientemente grande, que a veces la utilizamos para obtener una información muy general, luego de ella se extrae otra y así sucesivamente, en este caso, el elemento no cambia.

MUESTRA SIMPLE: Se dice, cuando las probabilidades de selección de los elementos son todas iguales y constantes durante el proceso de selección.

MUESTRA SISTEMÁTICA: Cuando se aplica algún método sistemático, opuesto a los aleatorios, como por ejemplo, cuando la selección se hace

a intervalos regulares, teniendo en cuenta que la primera selección se hace en forma aleatoria.

MUESTRAS APAREADAS: Donde cada elemento de una muestra está apareado con el elemento correspondiente de la (o de cada) otra, con respecto a cualidades distintas de aquéllas inmediatamente sometidas a investigación.

MUESTRAS INTERPENETRANTES: Cuando dos o más muestras han sido seleccionadas de la misma población, mediante el mismo proceso de selección.

MUESTRAS INDEPENDIENTES: Son muestras que se extraen de dos o más poblaciones, cada una en forma independiente de la otra

MUESTREO CON REEMPLAZAMIENTO: Cuando cada elemento o unidad de muestreo es extraída de la población finita y después de registrar sus características, se devuelve a la población antes de extraer la siguiente unidad, pudiendo ser seleccionado nuevamente.

MUESTREO MIXTO: Cuando un plan de muestreo supone el uso de dos o más métodos de muestreo.

MUESTREO POR CONGLOMERADO: Son grupos denominados conglomerados, heterogéneos en su composición interna, pero homogéneos al ser comparados entre si, la selección se hace de uno o más conglomerados, observando el total de conglomerados que se han hecho con todos los elementos de la población.

MUESTREO POR CONVENIENCIA: Se le conoce como muestreo intencional, muestreo a juicio, donde los elementos son seleccionados en forma intencional, a juicio del investigador o por conveniencia.

MUESTREO VOLUNTARIO: Cuando el elemento de la muestra, en este caso el informante no es seleccionado, sino por voluntad propia para el suministro de datos.

N

NIVEL DE CONFIANZA: Es una expresión alternativa para el coeficiente de confianza, el cual se asocia con un estimado por intervalo.

NIVEL DE SIGNIFICACIÓN: Es el valor que se le da a la región crítica, cuando realizamos una prueba de hipótesis, cuando el valor "Z" o de "t" obtenido mediante la variante estadística, cae en esa región se estará rechazando la hipótesis nula.

NO PARAMÉTRICO: Denominado como distribución libre, en esos casos no se conoce el valor del parámetro, ni de otro valor estadístico de la población.

NÚMERO ÍNDICE: Es una cifra relativa expresada en forma de porcentaje, empleada para indicar las variaciones de los precios, cantidades y valores de un período con respecto a otro que se denomina base.

NÚMERO ÍNDICE DE BASE FIJA: Cuando el período de referencia o base, se mantiene constante para los demás períodos.

NÚMERO ÍNDICE DE BASE VARIABLE: Se refiere a aquéllos en los cuales el período base de cada índice, corresponde al período inmediatamente anterior.

NÚMERO ÍNDICE DE BASES ENCADENADAS: Cuando se tienen índices de base variable y se desea transformarlo en índices de base fija, con respecto a un período, se encadenan mediante multiplicaciones sucesivas, de los relativos para cada período.

NÚMERO ÍNDICE DE CANTIDADES: Intenta indicar los cambios de las cantidades producidas, vendidas o adquiridas, entre dos o más períodos, independientemente de los cambios en los precios o valores unitarios.

NÚMERO ÍNDICE DE PRECIOS: Intenta indicar las variaciones de los precios o valores unitarios, entre dos o más períodos, independientemente de los cambios en las cantidades.

NÚMERO ÍNDICE PONDERADO: Denominado también como índices de Peso "fijos". Estos índices utilizan ponderaciones que pueden ser precios o cantidades, donde cada ponderación representa el peso de importancia dentro del conjunto.

NÚMEROS ALEATORIOS: Conjunto de números utilizados para la selección de muestras aleatorias.

O

OJIVA: Gráfico utilizado para representar una variable continua, cuya ordenada u eje vertical corresponde a las frecuencias absolutas acumuladas y en el eje horizontal o abscisas se encuentran los intervalos de la variable.

OSCILACIÓN: Algunos la consideran como una medida de dispersión, siendo la diferencia obtenida entre el valor máximo y mínimo que toma la variable.

P

PARÁMETRO: Son valores obtenidos con medidas (promedios, etc.), que han sido aplicadas a las características de los elementos o unidades, que constituyen la población objetivo.

PERCENTILES: Percentilos o Centilos, es un conjunto de valores obtenidos al dividir la frecuencia total en cien partes iguales. Por ejemplo: el P_{30} quiere decir que es un valor de la variable, que supera al 30% de las observaciones y es superado por el restante 70%.

PERÍODO BASE: En los números índices, corresponde al período que se toma como referencia, para indicar las variaciones que se presentan en los precios, cantidades y valores con respecto a otros períodos. En las series de tiempo también hay un período base, en el análisis de la regresión unilateral.

PERÍODO DADO: Corresponde a un período anterior o posterior al período base.

PERÍODO DE REFERENCIA: Es sinónimo de período base.

PICTOGRAMA: Es un método de representación gráfica, por medio de figuras como: casas, carros, barcos, personas, etc., cada uno de ellos representa un valor.

PLAN DE MUESTRA: Corresponde a todo el proceso de la investigación, desde la formulación del problema, el objetivo o finalidad de la investigación, diseño de la muestra, hasta su publicación.

POBLACIÓN: Colectivo finito o infinito de elementos. Conjunto de medidas o recuento de elementos que tienen una característica común. Ver Universo. Población objetivo.

POBLACIÓN FINITA: Población con un número limitado de elementos.

POBLACIÓN HIPOTÉTICA: Población estadística que no tiene existencia real.

POBLACIÓN INFINITA: Colectivo con un número ilimitado de elementos. Población finita con reposición de cada elemento que ha sido extraído o relacionado.

POBLACIÓN NO NORMAL: Población para la cual, la distribución de frecuencias no tiene un comportamiento igual al de la distribución normal.

POLÍGONO DE FRECUENCIAS: Diagrama mediante la cual visualizamos una distribución continua de frecuencias, ubicando las marcas de clase en las abscisas y las frecuencias en la ordenadas, que unidas forman una línea poligonal.

PONDERACIÓN: Peso o importancia que tiene un elemento dentro del conjunto.

POSICIÓN: Se refiere a las medidas de posición, las cuales permiten determinar un valor típico que represente al conjunto de observaciones.

POTENCIA: Es la probabilidad de rechazar la hipótesis alternativa cuando ésta es incorrecta, en una prueba de hipótesis.

PREDICCIÓN: Es el procedimiento que nos permite determinar el comportamiento de una variable en el futuro, en una serie de tiempo. Ver Pronóstico.

PREGUNTAS ABIERTAS: Son preguntas de opinión, que permiten un número ilimitado de respuestas. Presentan dificultad en su tabulación o recuento.

PREGUNTAS CERRADAS: Cuando sólo se tienen dos posibilidades, cuyas respuestas pueden ser clasificadas en las categorías "si o no"

PROBABILIDAD: Creencia que se tiene de que algo puede suceder u ocurrir, es un concepto difícil de definir. Se dice de algo probable o una medida numérica de la ocurrencia de un suceso.

PROBABILIDAD A PRIORI: La que se puede determinar sin necesidad de realizar el experimento, caso contrario a la probabilidad a posteriori.

PROMEDIO: Es un valor que intenta representar la característica de un conjunto de valores, como tal se entiende a algunas medidas de posición y de tendencia central. Por lo general, algunos lo toman como media aritmética.

PROMEDIOS MÓVILES: Es un procedimiento para suavizar una serie de tiempo, también utilizado como método de pronóstico y consiste en promediar cada grupo sucesivo de datos.

PROMEDIO PONDERADO: Cuando el promedio es calculado, en valores ordenados en una tabla de frecuencias, donde cada uno de ellos reflejan su importancia dentro del conjunto.

PRONÓSTICO: También se usa el término "Predicción", para indicar el comportamiento futuro de un hecho.

PROYECCIÓN: Al tener una serie de tiempo, se puede estimar su comportamiento futuro (extrapolar), es decir, nos proyectamos para períodos venideros. Algunos hablan de pronóstico como predicción del comportamiento de una serie de tiempo en el futuro.

PRUEBA BILATERAL: Se refiere a una prueba de hipótesis que tiene dos zonas de rechazo o dos regiones críticas, que permiten rechazar la hipótesis nula (Ho).

PRUEBA UNILATERAL: Puede localizarse a la izquierda, cuando por lo general se quiere probar que sea "menor" "inferior", etc. La región crítica permite rechazar Ho. Hacia la derecha, cuando se emplean los términos "mayor", "superior", "mejor", etc.

PRUEBA DE LA T: De Wilcoxon, aplicado cuando los datos están dados a escala ordinal y las dos muestras se encuentran relacionadas por pares o parejas.

PRUEBA DE MANN-WHITNEY: Es una prueba no paramétrica, aplicada con el fin de determinar las diferencias entre dos poblaciones, teniendo como base el análisis de dos muestras independientes.

PRUEBA DE RANGOS Y SIGNOS DE WILCOXON: Denominada también como prueba de rangos y signos de las diferencias entre dos poblaciones, con base en dos muestras

apareadas, por lo tanto, se realizan pruebas no paramétricas.

PRUEBA DE H DE KRUSKAL-WALLIS: Esta prueba no paramétrica se utiliza para probar si tres o más muestras independientes han sido extraídas de poblaciones que presentan la misma distribución.

PRUEBA DE SIGNO: Pertenece a pruebas estadísticas no paramétricas la cual nos permite identificar diferencias entre dos poblaciones, cuyas observaciones se encuentran apareadas y a la vez se dispone de datos nominales. Sólo se toman los signos positivos y negativos de las diferencias.

PRUEBA DE INDEPENDENCIA: Cuando se tiene una población, de la cual se extrae una sola muestra e interesa probar si un atributo o clasificación es independiente del otro.

PRUEBA DE HOMOGENEIDAD: En este caso, se tienen varias poblaciones de las cuales se extraen muestras, algunas veces de una muestra se obtienen dos o más categorías, e interesa probar si todas ellas presentan una identidad que hagan suponer que provienen de una sola población.

PRUEBA DE LA BONDAD: Se refiere a la realización de una prueba, con el fin de determinar si rechazamos una distribución hipotética para una población.

PRUEBA DE CORRELACIÓN POR RANGOS: Denominada coeficiente de Spearman, aplicada para probar si existe asociación entre dos variables, siempre y cuando se logre un nivel ordinal de medición para cada una de las variables.

PUNTO MUESTRAL: Es un elemento o suceso como resultado de un experimento aleatorio en condiciones normales.

R

RANGO: En algunos casos, se emplea para indicar la diferencia entre el valor máximo y mínimo que toma la variable. Se le denomina también como “recorrido”; algunos utilizan el rango como medida de dispersión y la denominan “oscilación”.

RANGO INTERCUARTÍLICO: Es una medida de dispersión, calculada mediante la diferencia entre el tercer y el primer cuartil.

RAZÓN DE MUESTREO: Corresponde a la denominada fracción de muestreo

$$(f = n/N).$$

REDONDEO: Aproximación a un número cuando se omiten dígitos finales, que pueden ser por exceso o por defecto.

REGIÓN CRÍTICA: También denominada como zona de rechazo, dependiendo de la prueba puede ser unilateral o bilateral. Cuando el valor de la variante estadística (Z, "t" o cualquier otro valor estadístico de prueba) cae o se localiza en esta región se estará rechazando Ho. (Hipótesis nula).

REGIÓN DE ACEPTACIÓN: Cuando la prueba estadística cae en la zona central, se estará aceptando la hipótesis nula.

REGIÓN CRÍTICA: Denominada también como “Región de rechazo”.

REGLA DE STURGES: Regla empírica, utilizada para determinar el número de intervalos o de marcas de clase en una variable continua.

REGRESIÓN: Utilizado por Galton para indicar ciertas relaciones de la herencia biológica; hoy utilizado como método estadístico, para analizar la relación entre dos o más variables. Regresión bidimensional o bivariante y multivariante.

REGRESIÓN EXPONENCIAL: Se aplica cuando la variable, en una serie de tiempo, presenta un crecimiento o decrecimiento geométrico. También se aplica en la regresión bilateral.

REGRESIÓN LINEAL SIMPLE: Es una de las maneras de expresión de la regresión, donde dos variables están relacionadas y se aproximan a una línea recta.

REGRESIÓN MÚLTIPLE: Corresponde a la regresión de una variable dependiente en función de dos o más variables "independientes".

RELATIVOS DE CANTIDADES: Es la división (cuociente) entre cantidades del período investigado y el período base.

RELATIVOS DE PRECIOS: Es la división (cuociente) entre los precios del período investigado y su período base.

REPOSICIÓN: Se emplea el término en el muestreo con reposición o sin reposición, en la selección de los elementos o unidades en una muestra.

RIESGO: Es el valor de equivocación o incumplimiento que se corre en la toma de decisiones.

S

SELECCIÓN AL AZAR: Es un método de selección de elementos o unidades, de tal manera que cada uno de ellos tengan la misma probabilidad de ser seleccionado.

SIMETRÍA: Se da cuando los valores de las frecuencias absolutas y relativas equidistantes de un valor central son iguales. La distribución normal es simétrica, además, la media, mediana y moda son iguales, ubicándose en el centro de la curva.

SERIE CRONOLÓGICA: Es un conjunto de observaciones, que se van registrando a medida que se producen, en un estricto orden cronológico. También se le denomina Serie de Tiempo.

SERIE DE TIEMPO DESESTACIONALIZADO: Corresponde a aquella serie de tiempo, al cual se le ha eliminado uno de los componentes (estacional), dividiendo cada valor original por el correspondiente índice estacional.

SIGNIFICACIÓN: Se dice de una prueba de significación, cuando se intenta verificar la hipótesis de que el efecto no existe. Los puntos o valores críticos en una prueba reciben el nombre de significativos.

SELECCIÓN SISTEMÁTICA: Término o punto aleatorio. En el muestreo sistemático los elementos se seleccionan a intervalos regulares, sin embargo el primer elemento es seleccionado al azar y los demás en forma sistemática.

SESGO: Que hay error.

SUAVIZACIÓN: Proceso de eliminación de las fluctuaciones en una serie de tiempo utilizando para ello totales y promedios móviles.

SUB-MUESTRA: Es una muestra, obtenida de otra muestra. Aplicada en el muestreo por fases múltiples.

SUB-MUESTREO: Es utilizado en el proceso de selección de unidades en el muestreo por etapas y en el muestreo por fases.

SUCESO: Término utilizado en el resultado de una prueba. El conjunto de sucesos constituyen un espacio muestral.

SUCESOS MUTUAMENTE EXCLUYENTES: Denominado como la ley de adición en las probabilidades, cuando dos o más sucesos son tales, que solamente uno de ellos puede ocurrir en un solo ensayo.

SUCESOS INDEPENDIENTES: Cuando la presentación de uno de los sucesos no está influenciado por los demás. Denominado también como Ley multiplicativa.

SUCESOS DEPENDIENTES: Cuando la probabilidad de presentación del segundo suceso, depende de la presentación del primer suceso; el tercero dependerá del resultado anterior y así sucesivamente.

TABLA DE CONTINGENCIA: Es una tabla de doble entrada, que permite clasificar las características, ya sean cualitativas o cuantitativas. En ella se resumen las frecuencias reales y las esperadas.

TABLA DE DOS POR DOS: Es una tabla de doble entrada, que tiene dos filas y dos columnas.

TABLA DE FRECUENCIAS: Es una manera fácil de manejar grandes cantidades de información, sobre una característica, en un cuadro o tabla donde se muestra la frecuencia que presenta cada valor que toma la variable.

TABLA DE VIDA: Forma de representar el número de personas, que de un número dado de nacidos siguen con vida en diferentes edades.

TAMAÑO DE LA MUESTRA: Conocido como tamaño "óptimo", correspondiendo al número de elementos o unidades mínimas que debe tener la muestra para que sea representativa.

TASA DE FERTILIDAD: Número de nacidos vivos en un período unitario, expresado como proporción del número de mujeres potencialmente fértiles, de la población en ese período.

TASA DE INCIDENCIA: Puede corresponder al número de nuevos casos (de una enfermedad) y la población sometida a riesgo.

TASA DE MORTALIDAD: Número de personas fallecidas en un período dado, dividido por la población correspondiente al punto medio del año.

TASA DE NATALIDAD: Es el número de nacimientos vivos registrados en un período (generalmente un año), dividido por la población, para ese mismo año, multiplicado todo por 1.000. Se conoce también como tasa bruta de natalidad.

TASA DE NO RESPUESTA: Proporción de personas que son seleccionadas en una encuesta y se niegan a contestar o responder el cuestionario.

TENDENCIA: Uno de los cuatro componentes de una serie de tiempo. Es un movimiento suave y constante, durante un período relativamente largo. Se representa mediante una función matemática, que puede ser lineal, parabólica, exponencial, etc.

TENDENCIA CENTRAL: Manera de denominar a las medidas de posición, con tendencia a ubicarse en el centro de la distribución.

TENDENCIA LINEAL: Cuando la función de la variable tiempo es lineal.

TENDENCIA RECTILÍNEA: Nombre alternativo para la tendencia lineal.

TEOREMA CENTRAL LÍMITE: Cuando a una población se le considera que no tiene un comportamiento normal y se extraen de ella todas las muestras posibles del mismo tamaño mayor a 30, la distribución de todas esas medias muestrales obtenidas tendrán una distribución igual a la normal, aunque la población no la tenga.

TEOREMA DE CHEBYSHEV: Es el porcentaje de cualquier conjunto de datos, que deben quedar bajo la curva, a uno, dos o tres desviaciones típicas alrededor del promedio, cuando realizamos inferencias.

TEOREMA DE BAYES: Es un método o procedimiento para el cálculo de probabilidades

posteriores, teniendo como base probabilidades a priori.

TEORÍA DE JUEGOS: Es aquella rama de las matemáticas que tiene que ver con la teoría de los resultados, entre dos más jugadores sometidos a normas específicas.

TOTAL MÓVIL: Es la primera parte del proceso para calcular el promedio móvil, es decir, se está dividiendo al total móvil por el número de períodos tomados en él. En cada paso se va agregando un nuevo período a la suma y se le va restando el primero de ese total. Los totales y los promedios pueden ser tomados de 3, de 5, 7, etc. si se desea trabajar con impares, o de 2, 4, 6 si son pares.

U

UNIDAD DE MUESTREO: Se habla de unidad elemental de muestreo, cuando ésta es indivisible y simplemente unidad, cuando se puede descomponer, es decir, divisible. El conjunto se constituye en una población de personas, de hogares, de objetos, etc.

UNIDAD ELEMENTAL: Es la forma más simple, generalmente identificada como elemento, en vez de unidad.

UNIDAD MUESTRAL: Se emplea como sinónimo de unidad de muestreo. Se expresa en algunos casos, como unidad primaria, unidad de segunda etapa, etc.

UNIMODAL: Forma de describir aquella distribución que sólo tiene un valor modal. Si hay varios modos, decimos que es bimodal (dos modos) o plurimodal (más de dos).

UNIVERSO: Término alternativo de población, para algunos cuando contiene varias poblaciones.

V

VALIDACIÓN: En estadística es empleada como una forma de demostrar que los resultados obtenidos a través de una muestra, son representativos de la población y la información obtenida es exacta.

VALOR CRÍTICO: Valor o punto crítico que se obtiene, dado un nivel de significación, en una prueba de hipótesis, el cual nos permite determinar si aceptamos o rechazamos la hipótesis nula.

VALOR ESPERADO: También conocido como esperanza matemática, siendo también la localización central de una variable aleatoria.

VALORES EXTREMOS: Se entiende como el menor y mayor valor que toma la variable, correspondiente a un conjunto de observaciones.

VARIABLE: Cantidad o valor que varía en un conjunto de observaciones.

VARIABLE ALEATORIA: Son los resultados que se obtienen de un experimento en condiciones normales y varía por ser aleatorio.

VARIABLE ALEATORIA DISCRETA: Toma únicamente valores enteros, es decir, que sólo asume una sucesión de valores finitos o infinitos.

VARIABLE ALEATORIA CONTINUA: Variable aleatoria que sólo puede tomar un conjunto de valores discontinuos. La variable puede tomar valores fraccionados, representados en un intervalo o en varios intervalos.

VARIABLE DEPENDIENTE: Variable que se va a predecir. Se denomina también como predictando. Simbolizada, generalmente por *Y*.

VARIACIÓN ESTACIONAL: Otro de los cuatro componentes de una serie de tiempo. Movimiento ligado con las estaciones del año, sin embargo, puede corresponder a períodos más

cortos, determinándose el inicio y finalización del mismo, mostrando un comportamiento periódico de un año o menos.

VARIABLE INDEPENDIENTE: La variable que utilizamos para predecir, se le denomina también Predictor y se simboliza generalmente por X.

VARIACIÓN CÍCLICA: Variaciones ligadas con la economía, que comprende períodos de crisis, recuperación, auge y declinación en su comportamiento. Es otro de los componentes de una serie de tiempo, la longitud del ciclo es indeterminado. Denominado también como componente cíclico.

VARIACIÓN ALEATORIA: Es otro de los componentes de una serie de tiempo, la cual se ve afectada por variaciones irregulares o factores internos y/o externos no predecibles, tales como: inundaciones, paro de trabajadores, falta de energía, etc.

VARIANZA: Medida de dispersión, siendo igual a la media de los cuadrados de las diferencias entre los valores que toma la variable y su media aritmética. Es decir, la media de las desviaciones respecto a su media, elevadas al cuadrado.

VARIANZA EXPLICADA: Es la media de los cuadrados de las diferencias entre los valores estimados y su media aritmética, utilizada en el análisis de regresión.

VARIANZA RESIDUAL: Es la media de los cuadrados, de las diferencias entre los valores de la variable y los estimados en un análisis de regresión.

ZONA DE ACEPTACIÓN: Generalmente denominada como Región de Aceptación, en la realización de una prueba de hipótesis.

ZONA DE RECHAZO: Generalmente denominada como Región Crítica o Región de Rechazo, en una prueba de hipótesis.

ZONA DE RIESGO: Expresión utilizada para señalar la región crítica, en una prueba de hipótesis.

TABLAS

TA LA I. REAS A O LA CURVA NORMAL DE PRO A ILIDAD

A significa el área entre la línea de simetría (es decir, $z = 0$) y el valor dado de *z*.

z		z		z		z	
0.00	0.0000	0.30	0.1179	0.60	0.2258	0.90	0.3159
.01	.0040	.31	.1217	.61	.2291	.91	.3186
.02	.0080	.32	.1255	.62	.2324	.92	.3212
.03	.0120	.33	.1293	.63	.2357	.93	.3238
.04	.0160	.34	.1331	.64	.2389	.94	.3264
.05	.0199	.35	.1368	.65	.2422	.95	.3289
.06	.0239	.36	.1406	.66	.2454	.96	.3315
.07	.0279	.37	.1443	.67	.2486	.97	.3340
.08	.0319	.38	.1480	.68	.2518	.98	.3365
.09	.0359	.39	.1517	.69	.2549	.99	.3389
.10	.0398	.40	.1554	.70	.2580	1.00	.3413
.11	.0438	.41	.1591	.71	.2612	1.01	.3433
.12	.0478	.42	.1628	.72	.2642	1.02	.3461
.13	.0517	.43	.1664	.73	.2673	1.03	.3485
.14	.0557	.44	.1700	.74	.2704	1.04	.3508
.15	.0596	.45	.1736	.75	.2734	1.05	.3531
.16	.0636	.46	.1772	.76	.2764	1.06	.3554
.17	.0675	.47	.1808	.77	.2794	1.07	.3577
.18	.0714	.48	.1844	.78	.2823	1.08	.3599
.19	.0754	.49	.1879	.79	.2852	1.09	.3621
.20	.0793	.50	.1915	.80	.2881	1.10	.3643
.21	.0832	.51	.1950	.81	.2910	1.11	.3665
.22	.0871	.52	.1985	.82	.2939	1.12	.3686
.23	.0910	.53	.2019	.83	.2967	1.13	.3708
.24	.0948	.54	.2054	.84	.2996	1.14	.3729
.25	.0987	.55	.2088	.85	.3023	1.15	.3749
.26	.1026	.56	.2123	.86	.3051	1.16	.3770
.27	.1064	.57	.2157	.87	.3079	1.17	.3790
.28	.1103	.58	.2190	.88	.3106	1.18	.3810
.29	.1141	.59	.2224	.89	.3133	1.19	.3830

(*) Tomado de "*Introduction to probability and statistics*". Henry L. Alder, Edward Roesler. W.H. Freeman and.

TA LA I. REAS A O LA CURVA NORMAL DE PRO A ILIDAD (ontinua i n)

z		z		z		z	
1.20	0.3849	1.55	0.4394	1.90	0.4713	2.25	0.4878
1.21	.3869	1.56	.4406	1.91	.4719	2.26	.4881
1.22	.3888	1.57	.4418	1.92	.4726	2.27	.4884
1.23	.3907	1.58	.4430	1.93	.4732	2.28	.4887
1.24	.3925	1.59	.4441	1.94	.4738	2.29	.4890
1.25	.3944	1.60	.4452	1.95	.4744	2.30	.4893
1.26	.3962	1.61	.4463	1.96	.4750	2.31	.4896
1.27	.3980	1.62	.4474	1.97	.4756	2.32	.4898
1.28	.3997	1.63	.4485	1.98	.4762	2.33	.4901
1.29	.4015	1.64	.4495	1.99	.4767	2.34	.4904
1.30	.4032	1.65	.4505	2.00	.4773	2.35	.4906
1.31	.4049	1.66	.4515	2.01	.4778	2.36	.4909
1.32	.4066	1.67	.4525	2.02	.4783	2.37	.4911
1.33	.4082	1.68	.4535	2.03	.4788	2.38	.4913
1.34	.4099	1.69	.4545	2.04	.4793	2.39	.4916
1.35	.4115	1.70	.4554	2.05	.4798	2.40	.4918
1.36	.4131	1.71	.4564	2.06	.4803	2.41	.4920
1.37	.4147	1.72	.4573	2.07	.4808	2.42	.4922
1.38	.4162	1.73	.4582	2.08	.4812	2.43	.4925
1.39	.4177	1.74	.4591	2.09	.4817	2.44	.4927
1.40	.4192	1.75	.4599	2.10	.4821	2.45	.4929
1.41	.4207	1.76	.4608	2.11	.4826	2.46	.4931
1.42	.4222	1.77	.4616	2.12	.4830	2.47	.4932
1.43	.4236	1.78	.4625	2.13	.4834	2.48	.4934
1.44	.4251	1.79	.4633	2.14	.4838	2.49	.4936
1.45	.4265	1.80	.4641	2.15	.4842	2.50	.4938
1.46	.4279	1.81	.4649	2.16	.4846	2.51	.4940
1.47	.4292	1.82	.4656	2.17	.4850	2.52	.4941
1.48	.4306	1.83	.4664	2.18	.4854	2.53	.4943
1.49	.4319	1.84	.4671	2.19	.4857	2.54	.4945
1.50	.4332	1.85	.4678	2.20	.4861	2.55	.4946
1.51	.4345	1.86	.4686	2.21	.4865	2.56	.4948
1.52	.4357	1.87	.4693	2.22	.4868	2.57	.4949
1.53	.4370	1.88	.4700	2.23	.4871	2.58	.4951
1.54	.4382	1.89	.4706	2.24	.4875	2.59	.4952

TA LA I. REAS A O LA CURVA NORMAL DE PRO A ILIDAD (ontinua i n)

z		z		z		z	
2.60	0.4953	2.95	0.4984	3.30	0.4995	3.65	0.4999
2.61	.4955	2.96	.4985	3.31	.4995	3.66	.4999
2.62	.4956	2.97	.4985	3.32	.4996	3.67	.4999
2.63	.4957	2.98	.4986	3.33	.4996	3.68	.4999
2.64	.4959	2.99	.4986	3.34	.4996	3.69	.4999
2.65	.4960	3.00	.4987	3.35	.4996	3.70	.4999
2.66	.4961	3.01	.4987	3.36	.4996	3.71	.4999
2.67	.4962	3.02	.4987	3.37	.4996	3.72	.4999
2.68	.4963	3.03	.4988	3.38	.4996	3.73	.4999
2.69	.4964	3.04	.4988	3.39	.4997	3.74	.4999
2.70	.4965	3.05	.4989	3.40	.4997	3.75	.4999
2.71	.4966	3.06	.4989	3.41	.4997	3.76	.4999
2.72	.4967	3.07	.4989	3.42	.4997	3.77	.4999
2.73	.4968	3.08	.4990	3.43	.4997	3.78	.4999
2.74	.4969	3.09	.4990	3.44	.4997	3.79	.4999
2.75	.4970	3.10	.4990	3.45	.4997	3.80	.4999
2.76	.4971	3.11	.4991	3.46	.4997	3.81	.4999
2.77	.4972	3.12	.4991	3.47	.4997	3.82	.4999
2.78	.4973	3.13	.4991	3.48	.4998	3.83	.4999
2.79	.4974	3.14	.4992	3.49	.4998	3.84	.4999
2.80	.4974	3.15	.4992	3.50	.4998	3.85	.4999
2.81	.4975	3.16	.4992	3.51	.4998	3.86	.4999
2.82	.4976	3.17	.4992	3.52	.4998	3.87	.5000
2.83	.4977	3.18	.4993	3.53	.4998	3.88	.5000
2.84	.4977	3.19	.4993	3.54	.4998	3.89	.5000
2.85	.4978	3.20	.4993	3.55	.4998		
2.86	.4979	3.21	.4993	3.56	.4998		
2.87	.4980	3.22	.4994	3.57	.4998		
2.88	.4980	3.23	.4994	3.58	.4998		
2.89	.4981	3.24	.4994	3.59	.4998		
2.90	.4981	3.25	.4994	3.60	.4998		
2.91	.4982	3.26	.4994	3.61	.4999		
2.92	.4983	3.27	.4995	3.62	.4999		
2.93	.4983	3.28	.4995	3.63	.4999		
2.94	.4984	3.29	.4995	3.64	.4999		

TA LA II. DISTRI UCI N *t* DE STUDENT

La primera columna señala el número de grados de libertad (*v*). El encabezado de las otras columnas da las probabilidades (P) de que *t* exceda numéricamente al valor de la tabla.

	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
31	0.682	1.309	1.696	2.040	2.453	2.744	3.022	3.375	3.633
32	0.682	1.309	1.694	2.037	2.449	2.738	3.015	3.365	3.622
33	0.682	1.308	1.692	2.035	2.445	2.733	3.008	3.356	3.611
34	0.682	1.307	1.691	2.032	2.441	2.728	3.002	3.348	3.601
35	0.682	1.306	1.690	2.030	2.438	2.724	2.996	3.340	3.591
36	0.681	1.306	1.688	2.028	2.434	2.719	2.990	3.333	3.582
37	0.681	1.305	1.687	2.026	2.431	2.715	2.985	3.326	3.574
38	0.681	1.304	1.686	2.024	2.429	2.712	2.980	3.319	3.566
39	0.681	1.304	1.685	2.023	2.426	2.708	2.976	3.313	3.558
40	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551

TA LA II. DISTRI UCI N *t* DE STUDENT (ontinua i n)

	0.50	0.20	0.10	0.05	0.02	0.01	0.005	0.002	0.001
42	0.680	1.302	1.682	2.018	2.418	2.698	2.963	3.296	3.538
44	0.680	1.301	1.680	2.015	2.414	2.692	2.956	3.286	3.526
46	0.680	1.300	1.679	2.013	2.410	2.687	2.949	3.277	3.515
48	0.680	1.299	1.677	2.011	2.407	2.682	2.943	3.269	3.505
50	0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
52	0.679	1.298	1.675	2.007	2.400	2.674	2.932	3.255	3.488
54	0.679	1.297	1.674	2.005	2.397	2.670	2.927	3.248	3.480
56	0.679	1.297	1.673	2.003	2.395	2.667	2.923	3.242	3.473
58	0.679	1.296	1.672	2.002	2.392	2.663	2.918	3.237	3.466
60	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
62	0.678	1.295	1.670	1.999	2.388	2.657	2.911	3.227	3.454
64	0.678	1.295	1.669	1.998	2.386	2.655	2.908	3.223	3.449
66	0.678	1.295	1.668	1.997	2.384	2.652	2.904	3.218	3.444
68	0.678	1.294	1.668	1.995	2.382	2.650	2.902	3.214	3.439
70	0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
72	0.678	1.293	1.666	1.993	2.379	2.646	2.896	3.207	3.431
74	0.678	1.293	1.666	1.993	2.378	2.644	2.894	3.204	3.427
76	0.678	1.293	1.665	1.992	2.376	2.642	2.891	3.201	3.423
78	0.678	1.292	1.665	1.991	2.375	2.640	2.889	3.198	3.420
80	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	0.677	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
140	0.676	1.288	1.656	1.977	2.353	2.611	2.852	3.149	3.361
160	0.676	1.287	1.654	1.975	2.350	2.607	2.846	3.142	3.352
180	0.676	1.286	1.653	1.973	2.347	2.603	2.842	3.136	3.345
200	0.676	1.286	1.653	1.972	2.345	2.601	2.839	3.131	3.340
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758	2.8070	3.0902	3.2905

Adaptado por J.H. Zar - Análisis bioestadístico - Prentice Hall - Englewood - Clitts - 1974 - pp. 413-414.

TA LA III. DISTRI UCI N INOMIAL

$$P = C_x^n \; p^x \; q^{n-x}$$

n	*x*	.05	.10	.15	.20	.25	. 0	. 5	. 0	. 5	.50
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.4000	.4500	.5000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312
	1	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1562
	2	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078

TA LA III. DISTRI UCI N INOMIAL (ontinua i n)

n	*x*	.05	.10	.15	.20	.25	. 0	. 5	. 0	. 5	.50
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3825	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0312
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2000	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0421	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0018	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0723	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005

TA LA III. DISTRI UCI N INOMIAL (ontinua i n)

n	*x*	.05	.10	.15	.20	.25	. 0	. 5	. 0	. 5	.50
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029
	2	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161
	3	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537
	4	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208
	5	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934
	6	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256
	7	.0000	.0000	.0006	.0033	.0115	.0291	.0591	.1009	.1489	.1934
	8	.0000	.0000	.0001	.0005	.0024	.0078	.0199	.0420	.0762	.1208
	9	.0000	.0000	.0000	.0001	.0004	.0015	.0048	.0125	.0277	.0537
	10	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0025	.0068	.0161
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0029
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.3512	.3672	.2774	.1787	.1029	.0540	.0259	.0113	.0045	.0016
	2	.1109	.2448	.2937	.2680	.2059	.1388	.0836	.0453	.0220	.0095
	3	.0214	.0997	.1900	.2457	.2517	.2181	.1651	.1107	.0660	.0349
	4	.0028	.0277	.0838	.1535	.2097	.2337	.2222	.1845	.1350	.0873
	5	.0003	.0055	.0266	.0691	.1258	.1803	.2154	.2214	.1989	.1571
	6	.0000	.0008	.0063	.0230	.0559	.1030	.1546	.1968	.2169	.2095
	7	.0000	.0001	.0011	.0058	.0186	.0442	.0833	.1312	.1775	.2095
	8	.0000	.0000	.0001	.0011	.0047	.0142	.0336	.0656	.1089	.1571
	9	.0000	.0000	.0000	.0001	.0009	.0034	.0101	.0243	.0495	.0873
	10	.0000	.0000	.0000	.0000	.0001	.0006	.0022	.0065	.0162	.0349
	11	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0012	.0036	.0095
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.3593	.3559	.2539	.1539	.0832	.0407	.0181	.0073	.0027	.0009
	2	.1229	.2570	.2912	.2501	.1802	.1134	.0634	.0317	.0141	.0056
	3	.0259	.1142	.2056	.2501	.2402	.1943	.1366	.0845	.0462	.0222
	4	.0037	.0348	.0998	.1720	.2202	.2290	.2022	.1549	.1040	.0611
	5	.0004	.0078	.0352	.0860	.1468	.1963	.2178	.2066	.1701	.1222
	6	.0000	.0013	.0093	.0322	.0734	.1262	.1759	.2066	.2088	.1833
	7	.0000	.0002	.0019	.0092	.0280	.0618	.1082	.1574	.1952	.2095
	8	.0000	.0000	.0003	.0020	.0082	.0232	.0510	.0918	.1398	.1833
	9	.0000	.0000	.0000	.0003	.0018	.0066	.0183	.0408	.0762	.1222
	10	.0000	.0000	.0000	.0000	.0003	.0014	.0049	.0136	.0312	.0611
	11	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0033	.0093	.0222
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0019	.0056
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0009
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

TA LA III. DISTRI UCI N INOMIAL (ontinua i n)

n	*x*	.05	.10	.15	.20	.25	. 0	. 5	. 0	. 5	.50
15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
	2	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
	3	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
	4	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
	5	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
	6	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
	7	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
	8	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
	9	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
	10	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
	11	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
	12	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0139
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.3706	.3294	.2097	.1126	.0535	.0228	.0087	.0030	.0009	.0002
	2	.1463	.2745	.2775	.2111	.1336	.0732	.0353	.0150	.0056	.0018
	3	.0359	.1423	.2285	.2463	.2079	.1465	.0888	.0468	.0215	.0085
	4	.0061	.0514	.1311	.2001	.2252	.2040	.1553	.1014	.0572	.0278
	5	.0008	.0137	.0555	.1201	.1802	.2099	.2008	.1623	.1123	.0667
	6	.0001	.0028	.0180	.0550	.1101	.1649	.1982	.1983	.1684	.1222
	7	.0000	.0004	.0045	.0197	.0524	.1010	.1524	.1889	.1969	.1746
	8	.0000	.0001	.0009	.0055	.0197	.0487	.0923	.1417	.1812	.1964
	9	.0000	.0000	.0001	.0012	.0058	.0185	.0442	.0840	.1318	.1746
	10	.0000	.0000	.0000	.0002	.0014	.0056	.0167	.0392	.0755	.1222
	11	.0000	.0000	.0000	.0000	.0002	.0013	.0049	.0142	.0337	.0667
	12	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0040	.0115	.0278
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0029	.0085
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.3741	.3150	.1893	.0957	.0426	.0169	.0060	.0019	.0005	.0001
	2	.1575	.2800	.2673	.1914	.1136	.0581	.0260	.0102	.0035	.0010
	3	.0415	.1556	.2359	.2393	.1893	.1245	.0701	.0341	.0144	.0052
	4	.0076	.0605	.1457	.2093	.2209	.1868	.1320	.0796	.0411	.0182
	5	.0010	.0175	.0668	.1361	.1914	.2081	.1849	.1379	.0875	.0472
	6	.0001	.0039	.0236	.0680	.1276	.1784	.1991	.1839	.1432	.0944

TA LA III. DISTRI UCI N INOMIAL (ontinua i n)

n	*x*	.05	.10	.15	.20	.25	. 0	. 5	. 0	. 5	.50
17	7	.0000	.0007	.0065	.0267	.0668	.1201	.1685	.1927	.1841	.1484
	8	.0000	.0001	.0014	.0084	.0279	.0644	.1134	.1606	.1883	.1855
	9	.0000	.0000	.0003	.0021	.0093	.0276	.0611	.1070	.1540	.1855
	10	.0000	.0000	.0000	.0004	.0025	.0095	.0263	.0571	.1008	.1484
	11	.0000	.0000	.0000	.0001	.0005	.0026	.0090	.0242	.0525	.0944
	12	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0081	.0215	.0472
	13	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0021	.0068	.0182
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.3763	.3002	.1704	.0811	.0338	.0126	.0042	.0012	.0003	.0001
	2	.1683	.2835	.2556	.1723	.0958	.0458	.0190	.0069	.0022	.0006
	3	.0473	.1680	.2406	.2297	.1704	.1046	.0547	.0246	.0095	.0031
	4	.0093	.0700	.1592	.2153	.2130	.1681	.1104	.0614	.0291	.0117
	5	.0014	.0218	.0787	.1507	.1988	.2017	.1664	.1146	.0666	.0327
	6	.0002	.0052	.0301	.0816	.1436	.1873	.1941	.1655	.1181	.0708
	7	.0000	.0010	.0091	.0350	.0820	.1376	.1792	.1892	.1657	.1214
	8	.0000	.0002	.0022	.0120	.0376	.0811	.1327	.1734	.1864	.1669
	9	.0000	.0000	.0004	.0033	.0139	.0386	.0794	.1284	.1694	.1855
	10	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214
	12	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0045	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1491	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1462
	9	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762

TA LA III. DISTRI UCI N INOMIAL (ontinua i n)

n	*x*	.05	.10	.15	.20	.25	. 0	. 5	. 0	. 5	.50
19	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1462
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

TA LA IV. DISTRI UCI N DE ¡ CUADRADO
REAS A O LA CURVA NORMAL DE PRO A ILIDAD

A significa el área del extremo derecho para los valores x^2 que aparecen en la tabla. *v* significa el número de grados de libertad.

	= 0	= 0	= 0 5	= 0 0	= 0 0	= 0 0	= 0 50
1	.00016	.00063	.0039	.016	.064	.15	.46
2	.02	.04	.10	.21	.45	.71	1.39
3	.12	.18	.35	.58	1.00	1.42	2.37
4	.30	.43	.71	1.06	1.65	2.20	3.36
5	.55	.75	1.14	1.61	2.34	3.00	4.35
6	.67	1.13	1.64	2.20	3.07	3.83	5.35
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34
11	3.05	3.61	4.53	5.58	6.99	8.15	10.34
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34
15	5.23	5.98	7.26	6.55	10.31	11.72	14.34
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34
22	9.54	10.60	12.34	14.04	16.31	18.10	21.34
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34
26	12.20	13.41	15.38	17.29	19.82	21.78	25.34
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34

TA LA IV. DISTRI UCI N DE ¡ CUADRADO (ontinua i n)

	= 0 0	= 0 20	= 0 10	= 0 05	= 0 02	= 0 01	= 0 001
1	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	28.42	20.46	23.54	26.30	29.63	32.00	39.25
17	19.51	21.62	24.77	27.59	31.00	33.41	40.79
18	20.60	22.76	25.98	28.87	32.35	34.80	42.31
19	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	32.46	35.14	39.09	42.56	46.69	49.59	58.20
30	33.53	36.25	40.26	43.77	47.96	50.69	59.70

TA LA V. TRANS ORMACI N DE *r* A *z* (es e ir $z = 0{,}5\ \text{Log}_e \frac{1+r}{1-r}$)

r	*z*	*r*	*z*	*r*	*z*
.00	.000	.36	.377	.71	.887
.01	.010	.37	.388	.72	.908
.02	.020	.38	.400	.73	.929
.03	.030	.39	.412	.74	.950
.04	.040	.40	.424	.75	.973
.05	.050				
.06	.060	.41	.436	.76	.996
.07	.070	.42	.448	.77	1.020
.08	.080	.43	.460	.78	1.045
.09	.090	.44	.472	.79	1.071
.10	.100	.45	.485	.80	1.099
.11	.110	.46	.497	.81	1.127
.12	.121	.47	.510	.82	1.157
.13	.131	.48	.523	.83	1.188
.14	.141	.49	.536	.84	1.221
.15	.151	.50	.549	.85	1.256
.16	.161	.51	.563	.86	1.293
.17	.172	.52	.576	.87	1.333
.18	.182	.53	.590	.88	1.376
.19	.192	.54	.604	.89	1.422
.20	.203	.55	.618	.90	1.472
.21	.213	.56	.633	.91	1.528
.22	.224	.57	.648	.92	1.589
.23	.234	.58	.662	.93	1.658
.24	.245	.59	.678	.94	1.738
.25	.255	.60	.693	.95	1.832
.26	.266	.61	.709	.96	1.946
.27	.277	.62	.725	.97	2.092
.28	.288	.63	.741	.98	2.298
.29	.299	.64	.758	.99	2.647
.30	.310	.65	.775		
.31	.321	.66	.793		
.32	.332	.67	.811		
.33	.343	.68	.829		
.34	.354	.69	.848		
.35	.365	.70	.867		

TA LA VI. NÚMEROS AL A AR

15 77 01 64 69	69 58 40 81 16	60 20 00 84 22	28 26 46 66 36	86 66 17 34 49
85 40 51 40 10	15 33 94 11 65	57 62 94 04 99	05 57 22 71 77	99 68 12 11 14
47 69 35 90 95	16 17 45 86 29	15 70 48 02 00	59 33 93 28 58	34 32 24 34 07
13 26 87 40 20	40 81 46 08 09	74 99 16 92 99	85 19 01 23 11	74 00 79 41 69
10 55 33 20 47	54 16 86 11 16	59 34 71 55 84	03 48 17 60 13	38 71 23 91 83
05 06 67 26 77	14 85 40 52 68	60 41 94 98 18	62 20 94 03 71	60 26 45 17 92
65 50 89 18 74	42 07 05 15 69	86 97 40 25 88	14 17 73 92 07	93 11 93 45 25
59 68 53 31 55	73 47 16 49 79	69 80 76 16 60	58 53 07 04 53	66 94 94 18 13
31 31 05 36 48	75 16 00 21 11	42 44 84 46 84	83 20 49 17 12	21 93 34 61 16
91 59 46 44 45	49 25 36 12 07	25 90 89 55 25	83 47 17 23 93	99 56 14 39 16
63 59 73 21 67	80 00 25 58 25	72 06 12 86 74	54 79 70 85 88	71 58 21 98 48
89 72 47 46 94	78 56 10 65 97	84 79 42 31 49	94 15 31 13 09	45 43 03 82 81
70 51 21 03 18	50 21 99 49 73	06 99 19 24 96	39 43 10 14 12	94 08 55 54 70
14 15 99 60 44	62 72 38 18 36	63 92 61 55 93	77 66 82 10 91	81 51 67 01 47
92 46 90 39 99	64 08 00 97 27	54 96 63 40 54	34 70 27 48 18	68 59 91 83 32
81 23 17 13 01	37 57 92 16 34	15 80 90 25 64	67 77 29 95 84	80 84 84 87 22
87 54 42 46 56	28 89 02 06 98	59 90 74 13 38	98 66 23 20 23	90 55 31 83 48
74 73 84 98 13	11 48 25 33 39	27 36 08 99 57	60 42 88 68 25	22 89 67 83 16
94 55 14 00 97	32 51 92 47 03	92 33 73 20 21	29 77 37 06 98	64 63 34 31 43
69 21 94 26 20	73 90 70 92 76	49 14 60 34 43	90 51 72 11 07	75 94 19 49 40
82 36 36 89 29	87 70 08 71 98	49 00 89 89 99	29 08 02 72 32	68 16 20 82 19
25 06 22 30 87	87 44 48 90 91	38 53 10 60 29	40 07 58 97 84	09 04 33 56 72
82 37 97 60 92	76 39 17 84 34	67 65 52 89 90	62 97 04 33 81	91 27 56 46 35
83 71 07 22 15	17 55 56 82 62	88 83 86 38 14	63 89 39 81 90	25 62 58 68 87
73 13 79 15 12	18 34 22 24 75	56 47 45 22 81	30 82 38 34 52	57 48 30 34 17
91 28 00 57 30	92 12 38 95 21	15 70 78 50 88	01 07 90 72 77	99 53 04 34 73
33 47 55 62 57	08 21 77 31 05	64 74 04 93 42	20 19 09 71 46	37 32 69 69 89
56 66 25 32 38	64 70 26 27 67	77 40 04 34 63	98 99 89 31 16	12 80 50 28 96
88 40 52 02 29	82 69 34 50 21	74 00 91 27 52	98 72 03 45 65	30 89 71 45 91
87 63 88 23 62	51 07 69 59 02	89 49 14 98 53	41 92 36 07 76	85 37 84 37 47
32 25 21 15 08	82 34 57 57 35	22 03 33 48 84	37 37 29 38 37	89 76 25 09 69
44 61 88 23 13	01 59 47 64 04	99 59 96 20 30	87 31 33 69 45	58 48 00 83 48
94 44 08 67 79	41 61 41 15 60	11 88 83 24 82	24 07 78 61 89	42 58 88 22 16
13 24 40 09 00	65 46 38 61 12	90 62 41 11 59	85 18 42 61 29	88 76 04 21 80
78 27 84 05 99	85 75 67 80 05	57 05 71 70 21	31 99 99 06 96	53 99 25 13 63
42 39 30 02 34	99 46 68 45 15	19 74 15 50 17	44 80 13 86 38	40 45 82 13 44
04 52 43 96 38	13 83 80 72 34	20 84 56 19 49	59 14 85 42 99	71 16 34 33 79
82 85 77 30 16	69 32 46 46 30	84 20 68 72 98	94 62 63 59 44	00 89 06 15 87
38 48 84 88 24	55 46 48 60 06	90 08 83 83 98	40 90 88 25 26	85 74 55 80 85
91 19 05 68 22	58 04 63 21 16	23 38 25 43 32	98 94 65 35 35	16 91 07 12 43
54 81 87 21 31	40 46 17 62 63	99 71 14 12 64	51 68 50 60 78	22 69 51 98 37
65 43 75 12 91	20 36 25 57 92	33 65 95 48 75	00 06 65 25 90	16 29 34 14 43
49 98 71 31 80	59 57 32 43 07	85 06 64 75 27	29 17 06 11 30	68 70 97 87 21
03 98 68 89 39	71 87 32 14 99	42 10 25 37 30	08 27 75 43 97	54 20 69 93 50
56 04 21 34 92	89 81 52 15 12	84 11 12 66 87	47 21 06 86 08	35 39 52 28 09

TA LA VI. NÚMEROS AL A AR (ontinua i n)

48 09 36 95 36	20 82 53 32 89	92 68 50 88 17	37 92 02 23 43	63 24 69 80 91
23 97 10 96 57	74 07 95 26 44	93 08 43 30 41	86 45 74 33 78	84 33 38 76 73
43 97 55 45 98	35 69 45 96 80	46 26 39 96 33	60 20 73 30 79	17 19 03 47 28
40 05 08 50 79	89 58 19 86 48	27 98 99 24 08	94 19 15 81 29	82 14 35 88 03
66 97 10 69 02	25 36 43 71 76	00 67 56 12 69	07 89 55 63 31	50 72 20 33 36
97 58 55 23 12	87 39 84 32 23	26 91 01 11 26	01 24 06 58 20	33 46 38 86 23
84 95 87 34 95	31 23 12 64 75	89 28 38 15 91	81 89 08 86 08	88 20 02 11 67
11 52 38 09 94	32 47 35 42 67	39 33 89 97 16	28 94 86 93 86	96 13 43 85 99
38 69 94 97 10	44 42 85 46 88	56 56 63 58 22	89 19 26 82 25	94 15 54 65 62
23 99 36 33 41	99 76 22 29 19	92 53 92 15 71	47 57 74 69 03	65 57 90 53 17
09 15 95 74 87	09 63 82 63 29	84 57 45 80 07	13 57 40 58 34	21 93 90 39 21
55 75 91 36 57	38 30 89 64 42	01 84 83 12 79	32 09 56 03 81	90 88 00 71 02
84 62 29 92 42	03 92 37 46 19	90 75 68 84 49	53 80 62 19 20	31 14 42 11 17
79 25 70 07 80	85 32 53 87 11	33 79 14 20 04	12 40 31 74 39	80 21 37 65 20
40 10 91 52 27	21 18 64 61 04	85 55 16 90 71	31 95 15 86 74	87 80 75 71 27
93 18 86 63 72	22 53 44 23 89	38 06 46 04 79	67 77 33 21 75	40 51 74 60 53
63 71 69 30 23	12 85 90 05 07	67 33 56 52 60	21 50 72 26 28	48 67 31 87 61
05 29 95 78 06	10 41 62 18 37	42 91 98 43 33	20 58 62 80 65	19 90 07 84 49
30 04 29 90 89	64 25 66 36 41	99 59 15 43 86	34 10 05 99 83	08 02 18 01 22
75 50 83 42 46	80 76 77 34 16	04 05 06 28 86	60 70 04 13 28	98 76 78 43 69
68 82 44 11 33	11 20 42 00 22	40 03 06 12 45	06 32 34 44 18	01 26 36 78 42
51 38 78 69 65	25 98 73 40 31	12 04 99 51 09	49 04 32 68 68	54 64 15 25 68
98 41 81 63 70	58 43 39 93 18	54 46 98 33 01	47 85 39 81 11	48 84 07 64 76
08 44 37 01 53	59 67 11 11 53	16 98 16 52 52	39 32 22 18 22	04 03 06 77 17
17 30 92 82 09	42 37 88 43 35	11 54 89 05 61	10 46 27 43 33	88 92 72 62 01
74 87 89 10 02	19 45 29 65 70	77 81 98 78 67	05 62 57 08 79	30 32 62 91 87
61 81 52 99 80	11 55 21 98 02	08 26 01 20 16	07 42 88 56 51	31 96 14 85 49
55 08 43 08 22	50 28 03 18 00	80 79 60 18 33	92 36 13 50 41	43 59 82 16 65
44 38 47 15 16	96 03 51 42 15	35 96 40 87 91	56 91 13 58 85	40 06 36 04 30
12 45 97 68 57	62 36 61 03 29	46 60 79 85 99	91 13 99 95 58	75 14 74 88 12
19 95 23 05 45	01 87 81 18 92	36 94 07 14 08	90 32 51 29 61	50 60 34 92 25
71 55 86 72 94	77 08 55 65 50	33 53 94 81 52	36 31 53 12 74	88 59 99 35 95
07 32 94 03 20	66 29 98 75 65	70 30 56 59 08	24 51 75 48 73	11 29 77 08 36
10 35 58 59 25	89 62 60 77 71	24 13 38 20 83	02 48 11 67 95	38 97 15 58 18
62 99 34 08 06	81 46 09 16 82	95 17 13 46 36	51 36 87 56 10	80 79 40 48 82
19 44 35 31 20	16 05 25 26 38	98 94 18 38 88	10 90 29 01 12	48 85 52 97 22
77 76 94 64 49	45 39 58 07 88	32 11 43 09 51	32 69 31 63 02	33 47 08 94 85
97 43 81 59 46	59 26 04 63 86	87 31 55 50 66	11 37 04 68 14	57 17 08 82 48
09 77 93 46 95	36 98 08 77 39	71 44 48 10 19	54 80 24 83 47	06 79 01 78 43
71 09 43 23 16	33 93 21 87 89	16 53 05 53 16	98 96 30 89 49	83 32 23 13 32
25 19 47 70 48	16 91 39 59 80	66 77 96 02 08	59 58 48 91 81	04 31 64 65 15
43 23 23 81 42	61 42 37 17 76	75 40 18 81 33	51 68 04 41 00	72 82 28 68 03
50 57 81 53 79	98 04 75 77 30	49 18 17 01 70	06 01 53 04 76	49 93 39 68 00
81 04 78 50 20	33 21 64 10 00	49 43 08 86 53	25 50 24 70 63	01 08 52 66 67
19 62 59 60 23	26 11 30 12 63	26 60 61 15 83	27 41 02 61 80	72 19 91 56 53

TA LA VI. NÚMEROS AL A AR (ontinua i n)

32 52 48 94 61	60 43 08 29 67	86 20 90 03 18	48 22 42 82 59	84 31 00 92 15
79 73 88 64 27	89 92 95 64 78	40 06 16 28 66	54 93 14 19 00	39 11 13 27 55
05 12 93 24 38	18 25 64 65 51	81 15 80 43 36	94 49 89 58 80	80 76 25 65 69
59 72 45 18 64	49 67 78 83 66	72 92 63 42 78	21 14 35 00 16	05 92 74 20 31
22 75 30 52 34	00 43 50 50 91	10 64 18 60 30	48 99 84 23 37	20 03 50 50 05
86 21 48 23 45	01 80 49 33 99	57 92 46 06 55	60 98 81 40 20	72 45 67 83 67
47 02 27 40 96	41 44 06 54 76	83 52 32 56 15	09 45 22 54 07	49 70 54 48 84
36 76 21 72 44	85 55 63 87 29	62 84 18 48 29	23 75 29 90 68	02 56 04 32 34
43 84 04 45 20	18 42 25 25 95	70 15 92 80 82	47 10 21 18 57	83 54 02 09 53
88 82 00 84 16	82 67 66 77 89	78 31 98 11 56	27 07 76 59 71	87 56 99 27 28
83 10 03 87 35	31 89 45 64 40	61 57 87 29 73	69 40 81 83 78	38 48 42 22 29
36 57 72 32 39	32 78 83 08 53	87 69 15 29 29	11 24 78 09 01	35 27 12 80 54
63 83 40 96 33	52 74 75 58 02	66 94 42 87 71	15 33 56 74 49	92 05 98 56 05
85 29 24 77 88	45 67 31 74 66	76 46 33 84 55	50 13 03 00 97	36 18 25 57 31
91 49 28 72 30	03 03 32 57 95	19 01 67 59 15	00 72 44 40 07	93 42 69 48 89
87 60 75 26 58	65 28 70 09 88	76 32 66 08 99	67 99 07 26 67	47 67 42 72 31
93 20 96 93 05	95 71 90 83 58	02 47 28 10 24	80 64 66 52 38	34 94 17 54 28
94 34 73 87 35	53 33 66 29 50	21 38 74 75 20	54 87 63 74 34	80 48 64 70 79
10 42 12 47 60	34 51 84 25 61	15 76 23 67 84	81 43 35 46 20	95 21 48 80 57
42 57 65 70 17	49 44 83 03 32	53 60 38 20 38	27 02 61 96 31	36 84 25 44 05
14 80 92 91 76	39 57 07 38 03	12 39 71 36 50	82 44 43 32 84	37 77 36 93 64
03 00 22 34 59	83 22 53 06 87	63 11 38 68 14	69 63 25 22 53	49 99 06 31 53
35 10 31 74 97	55 56 24 84 81	54 14 11 24 60	97 32 52 97 29	86 25 07 61 67
75 65 43 98 76	14 42 68 43 55	85 35 29 44 75	95 08 42 39 30	34 12 53 00 91
42 44 24 72 86	52 95 61 21 88	53 57 56 76 42	39 29 24 47 87	06 89 29 41 09
88 16 39 38 57	70 64 36 88 96	88 80 42 18 56	79 44 06 33 07	20 64 21 48 29
13 30 95 40 57	41 19 03 38 11	74 30 45 56 81	14 84 18 09 31	20 98 29 04 09
23 36 64 70 84	35 90 33 28 80	31 42 17 42 14	53 97 26 51 79	56 39 88 41 10
41 86 43 21 65	27 78 35 17 70	05 54 68 00 81	57 73 00 65 27	37 88 17 32 78
23 77 82 23 63	32 70 14 93 29	92 35 32 80 34	12 77 83 97 42	98 41 16 08 33
65 67 57 25 37	61 14 90 95 89	63 15 38 22 88	67 54 24 16 70	36 03 43 78 99
67 31 48 80 29	29 75 62 66 20	06 95 93 05 22	42 18 68 82 07	35 85 84 11 10
94 07 25 78 88	71 92 56 13 38	00 21 32 91 42	57 87 41 13 39	46 60 33 81 89
98 40 03 05 13	23 79 44 10 06	62 37 35 74 89	37 32 87 50 25	44 94 20 81 69
99 28 14 84 72	37 60 08 57 93	31 46 40 68 65	88 96 64 46 69	09 77 52 93 91
50 07 12 34 37	43 89 16 85 44	26 59 20 40 26	77 28 35 71 03	00 56 60 15 16
43 43 27 67 69	24 70 00 71 43	41 27 40 43 75	26 53 43 14 09	54 09 37 54 29
23 71 33 20 98	75 46 75 11 95	11 27 73 15 47	01 53 39 55 47	27 77 51 39 16
33 39 03 80 68	63 76 38 21 28	89 65 05 32 04	09 18 15 91 48	77 27 49 07 59
70 60 79 38 93	80 47 18 80 30	23 27 51 39 11	56 13 89 58 19	42 45 52 60 02
42 80 22 71 37	41 78 34 80 50	52 35 37 72 51	91 15 29 08 28	59 22 58 29 34
70 15 04 92 81	58 71 82 88 34	67 53 42 67 05	60 96 28 14 96	88 77 62 92 69
89 87 13 76 96	78 80 29 11 43	96 96 75 43 06	24 08 07 41 67	81 13 28 85 66
34 71 21 83 96	26 78 18 33 22	07 54 10 84 39	64 03 35 82 03	68 81 64 19 12
01 56 07 44 40	86 57 76 42 27	84 48 21 76 80	00 59 65 93 84	58 68 72 17 66

TA LA VI. NÚMEROS AL A AR (ontinua i n)

17 09 69 35 42	26 01 07 58 52	42 61 78 67 23	91 24 18 12 18	09 62 97 22 36
41 91 77 30 25	56 04 03 64 48	31 09 01 73 10	68 86 74 11 01	49 69 39 17 69
36 07 08 01 86	46 94 22 08 18	17 57 91 42 64	35 92 91 01 66	43 36 00 26 20
05 84 14 25 59	55 02 52 38 43	05 06 95 35 82	87 01 42 16 65	82 20 99 51 49
95 65 59 54 30	45 52 24 91 19	53 64 17 80 74	25 76 50 14 95	68 70 26 16 74
02 86 60 83 38	23 99 89 05 41	87 87 85 39 40	95 60 08 08 85	03 56 91 38 35
38 42 83 19 25	46 11 48 40 62	08 44 36 60 81	86 80 71 73 34	55 45 64 10 13
85 91 80 69 35	56 79 85 78 37	59 19 01 02 02	54 38 98 50 14	87 81 68 59 64
12 14 28 93 12	12 71 38 81 60	79 92 19 78 49	56 14 93 03 40	38 33 83 93 39
17 80 74 77 75	62 70 70 90 72	12 96 21 03 70	90 62 53 23 36	67 00 59 11 68
06 50 91 31 95	98 95 81 72 03	37 09 32 11 61	35 40 44 65 14	15 78 17 94 53
26 74 80 36 23	70 43 60 72 48	13 53 12 03 87	46 80 87 39 20	09 69 58 42 84
31 60 92 27 82	42 36 64 13 90	83 89 57 63 85	98 80 30 06 82	85 82 64 86 77
32 68 19 58 65	03 76 48 17 22	61 61 51 26 36	97 54 95 79 57	59 31 78 34 27
90 84 21 49 55	99 37 42 96 82	94 55 34 23 57	13 95 33 65 37	77 37 43 21 41
05 97 03 44 96	49 87 13 01 69	56 19 27 30 31	16 95 62 74 55	62 27 80 55 00
78 07 89 48 22	97 68 14 69 88	21 52 73 88 28	08 27 09 48 56	34 88 92 17 36
24 08 18 32 54	17 10 47 62 05	13 45 48 09 65	88 39 09 21 48	12 46 88 81 22
02 62 56 96 61	58 40 50 70 94	07 38 72 82 85	68 21 50 32 70	50 99 54 26 39
65 85 71 69 31	00 03 29 96 17	55 28 81 32 56	58 65 37 17 49	62 61 84 05 49
96 13 18 60 79	62 09 04 94 05	03 75 77 32 37	13 74 76 59 41	60 49 08 17 32
83 73 96 86 33	43 15 91 12 62	49 52 62 12 27	85 64 63 67 13	72 58 85 85 79
37 10 24 50 84	63 32 25 04 92	56 86 34 21 10	46 28 41 11 27	65 01 80 44 11
83 63 98 19 87	48 93 24 60 58	14 20 24 31 90	06 21 43 33 17	01 06 38 29 32
31 28 23 11 97	92 99 45 16 23	88 32 80 13 98	26 36 83 38 31	40 86 82 06 47
01 55 37 09 08	81 17 93 34 68	48 50 08 41 90	60 26 43 21 86	50 58 48 36 73
88 62 21 77 17	19 83 36 01 24	85 14 14 31 77	37 05 31 81 51	18 19 62 50 12
42 94 20 19 46	97 74 57 42 36	41 15 88 09 02	71 08 54 31 04	44 33 74 61 78
02 94 74 35 47	44 49 19 83 14	64 35 71 96 78	39 82 22 51 72	72 82 37 27 37
98 93 56 53 31	43 87 88 26 07	51 33 57 32 67	94 12 59 72 70	40 60 24 24 88
04 69 96 20 56	34 82 13 40 44	35 73 12 04 47	89 03 15 13 38	57 23 90 10 64
46 17 32 80 88	50 94 76 71 00	66 80 63 10 19	43 71 14 39 67	18 68 83 99 86
03 03 53 25 33	48 56 38 29 97	72 26 86 66 59	45 72 22 50 57	85 61 96 34 51
51 93 74 83 62	36 47 66 48 33	77 88 04 09 60	16 52 91 29 36	82 02 81 08 41
12 57 43 34 86	58 28 03 45 04	17 56 69 15 72	92 11 29 73 60	32 39 72 92 29
53 84 34 35 10	80 58 23 33 53	41 31 36 65 19	36 95 29 39 82	36 44 83 29 54
55 38 24 04 52	34 62 02 75 71	31 79 80 57 90	83 04 81 87 06	39 94 97 18 98
44 18 54 27 28	95 03 69 23 00	22 27 93 81 82	27 62 41 14 88	15 55 08 23 49
70 40 80 38 20	65 19 35 17 50	53 98 92 85 83	17 17 27 36 58	28 81 79 50 86
04 73 99 66 15	80 49 90 75 98	81 37 70 45 42	88 08 71 92 61	36 66 22 05 45

TA LA VI. NÚMEROS AL A AR (ontinua i n)

35 42 90 60 81	67 17 29 99 11	27 47 34 75 24	58 34 16 99 76	47 96 74 27 60
83 09 18 41 03	44 38 21 48 83	47 02 06 64 06	02 79 41 48 98	70 99 85 86 05
70 73 27 11 03	32 39 41 95 31	48 92 07 52 74	04 60 86 62 54	74 82 96 12 31
52 54 07 60 09	11 28 74 68 44	65 54 62 25 13	90 38 80 51 09	63 66 08 69 63
76 73 21 52 45	47 21 92 11 66	09 43 96 59 57	34 35 27 05 22	59 22 31 16 52
82 51 57 91 88	66 47 82 53 13	27 84 57 65 23	27 64 73 11 55	49 12 71 94 95
95 81 86 44 22	67 86 41 16 79	76 11 42 25 48	69 34 95 67 54	37 70 36 35 96
71 15 85 43 23	14 87 31 12 31	46 97 62 75 56	98 94 05 25 97	77 35 25 77 47
68 51 40 99 65	59 34 38 61 88	74 60 22 09 94	69 36 04 99 84	57 92 33 78 45
55 33 10 60 67	39 62 16 35 78	84 93 94 14 10	56 11 50 64 11	13 78 75 52 94
43 56 50 63 55	60 09 73 81 27	36 83 55 12 43	83 05 88 57 87	96 61 34 01 50
19 60 92 82 67	31 06 48 13 20	59 96 66 97 84	72 73 16 83 42	31 61 35 11 32
00 01 71 82 97	15 55 79 08 53	56 52 49 15 44	16 86 30 72 18	54 09 89 88 36
39 18 47 54 33	55 91 02 90 29	81 74 34 98 16	95 03 65 94 61	87 12 02 44 75
60 74 52 41 39	70 85 23 26 19	13 27 95 69 73	54 97 86 65 16	23 66 41 51 77
08 33 98 06 13	75 58 12 37 63	00 43 89 06 97	92 67 83 52 14	40 24 35 63 04
90 78 83 08 33	13 86 55 41 35	26 91 94 70 72	10 46 68 93 19	63 53 12 78 48
55 77 40 85 45	71 63 09 79 12	53 67 21 77 12	01 16 30 72 76	47 79 27 39 59
32 37 88 40 16	08 22 32 48 18	85 65 10 25 54	18 45 24 66 79	89 75 88 70 59
86 14 97 98 50	80 62 48 58 87	03 56 33 46 00	23 79 76 36 62	77 78 78 29 17
82 01 52 21 90	27 95 27 62 34	99 79 54 37 02	69 99 83 70 68	07 39 06 05 24
99 87 92 23 49	06 96 93 21 97	59 73 66 29 74	39 30 89 05 10	02 57 59 89 34
37 67 24 02 62	68 25 66 66 24	48 12 13 94 93	47 54 64 03 40	66 52 45 98 42
44 24 63 02 09	81 52 96 73 00	20 67 70 19 65	80 69 01 80 47	36 75 39 66 96
60 61 42 61 47	50 35 25 28 26	66 25 62 99 76	45 73 32 96 07	64 46 87 73 66
90 78 82 54 47	20 83 80 10 41	35 22 23 02 98	9 74 41 35 05	78 73 95 47 83
78 58 68 87 41	11 08 81 29 89	71 23 10 01 79	5 06 00 45 80	64 70 95 34 29
51 42 21 03 88	20 05 35 93 00	68 12 09 55 09	6 54 95 22 82	48 30 09 56 87
93 15 07 60 86	67 37 94 24 35	82 44 19 92 96	1 84 29 04 29	83 32 05 10 48
27 12 31 66 62	09 54 17 31 23	27 30 37 36 79	5 50 39 57 12	67 23 22 09 33
79 44 83 55 47	96 50 93 56 82	58 16 35 18 87	4 08 22 47 93	86 43 43 30 17
89 73 43 91 03	57 91 35 40 64	13 61 94 37 16	9 93 96 25 87	30 23 42 54 31
29 30 90 00 58	15 99 93 33 67	80 08 59 21 66	3 54 56 85 25	05 32 03 52 52
97 33 17 26 25	04 73 18 10 05	34 40 32 65 07	8 68 29 31 97	89 57 95 55 16
07 15 44 92 47	28 50 93 03 53	37 70 19 68 59	5 39 87 90 46	98 64 46 24 71
82 50 35 50 80	23 67 81 25 02	83 08 12 70 00	5 31 33 80 06	19 86 14 59 27
59 21 86 16 30	27 85 16 26 34	50 15 87 22 69	1 36 95 90 76	90 99 79 63 21
04 19 60 33 05	29 02 33 74 56	38 84 21 07 35	3 54 70 18 47	14 62 75 45 02
96 91 44 09 94	06 89 50 88 83	82 50 11 82 51	0 68 91 06 28	86 65 17 45 20
31 71 03 53 38	94 02 52 72 15	44 49 53 42 43	0 36 97 67 64	12 27 46 00 18

TA LA VI. NÚMEROS AL A AR (ontinua i n)

03 70 22 67 59	98 10 64 68 08	79 06 89 48 41	5 72 10 87 24	96 04 20 68 00
08 45 79 46 89	74 73 67 60 15	70 37 61 44 07	7 89 81 54 26	57 17 63 27 74
37 80 05 75 64	48 51 68 68 27	71 75 45 32 27	6 35 26 58 88	67 74 48 90 94
90 63 56 69 37	19 74 48 63 31	52 36 84 40 66	2 66 03 41 87	65 29 12 36 64
22 69 38 02 88	89 71 43 01 87	41 79 42 99 29	1 08 47 32 19	45 29 59 69 90
05 79 69 67 64	36 14 82 65 26	40 51 63 42 48	5 48 34 12 04	33 26 52 26 52
48 91 53 03 82	64 24 06 31 03	97 44 82 24 89	8 48 66 54 10	41 27 09 11 61
94 64 97 27 25	62 23 94 40 54	56 32 97 78 90	8 86 41 75 19	42 90 85 36 68
15 85 82 52 08	52 96 26 92 88	93 11 03 23 52	8 23 57 85 43	53 90 42 22 22
09 81 37 66 56	99 08 59 19 48	29 69 21 64 95	2 08 15 24 45	59 25 22 76 96
43 83 99 02 76	12 16 45 52 66	35 70 93 09 52	5 40 34 35 62	65 42 27 20 59
31 98 09 80 62	75 26 64 57 26	46 41 47 90 97	9 46 10 51 42	73 28 98 89 91
81 35 42 62 84	37 02 59 78 16	17 96 05 71 39	8 05 34 05 92	22 43 89 66 89
97 95 56 39 75	65 47 61 86 33	14 88 55 33 69	0 87 79 94 46	17 61 72 27 01
37 63 35 93 23	17 30 14 51 51	17 28 21 74 67	2 11 57 19 27	38 70 73 82 92
39 22 96 00 48	52 49 62 09 40	08 30 27 54 70	6 06 52 12 80	36 12 38 68 05
61 29 84 34 51	60 19 77 82 16	64 45 02 27 04	5 55 90 95 04	20 39 29 96 28
38 84 18 10 29	10 09 66 06 78	37 09 60 50 21	2 72 01 52 70	29 65 05 37 16
64 29 48 04 08	55 72 25 25 77	54 26 27 24 39	6 67 06 40 00	99 35 70 69 58
64 02 32 99 63	62 42 89 32 20	81 14 08 40 45	2 15 37 49 38	96 51 19 08 27
13 83 39 51 30	31 49 94 83 66	02 50 95 18 98	8 84 90 58 81	00 40 91 12 46
83 30 90 09 35	41 12 87 93 66	85 96 20 65 34	3 13 05 41 01	91 48 95 59 45
46 63 53 97 63	18 86 37 56 20	35 62 66 11 37	0 91 89 97 51	64 78 06 95 65
54 43 40 02 41	55 70 52 96 87	02 82 61 21 88	0 65 98 42 09	03 61 20 83 01
27 18 65 62 01	97 45 79 51 37	74 47 20 11 48	7 93 73 86 50	46 61 95 01 24
45 42 16 13 20	34 51 08 71 52	39 17 71 39 84	7 27 72 49 42	81 62 32 87 22
35 92 97 02 34	93 32 95 81 13	92 05 40 70 95	1 66 61 24 08	77 32 73 66 79
60 55 35 57 24	52 95 84 90 64	38 39 72 70 17	8 42 85 96 67	41 11 83 17 78
43 17 21 09 60	58 86 12 31 11	66 61 43 96 00	3 97 00 15 20	37 96 73 56 63
07 85 74 58 28	38 74 68 32 61	87 14 71 83 47	0 11 96 70 08	67 04 34 46 08
33 00 29 08 87	42 59 40 24 97	44 99 13 56 87	5 02 47 97 89	23 51 45 37 83
97 14 00 42 23	72 03 19 02 41	11 23 36 98 32	9 91 42 03 58	62 23 74 45 06
68 58 32 80 82	40 49 71 83 37	93 49 99 69 72	8 14 26 88 95	48 69 35 40 63
39 87 38 16 06	82 92 62 32 75	67 64 50 49 39	9 55 53 92 97	04 48 60 53 90
37 73 01 84 87	42 88 30 93 75	01 18 34 73 30	8 44 28 18 01	00 38 26 38 57
15 62 38 72 92	03 76 09 30 75	77 80 04 24 54	67 60 1 79 26	21 60 03 48 14
77 81 15 14 67	55 24 22 20 55	36 93 67 69 37	72 22 4 46 32	56 15 75 25 12
18 87 05 09 96	45 14 72 41 46	12 67 46 72 02	59 06 1 49 12	73 28 23 52 48
08 58 53 63 66	13 07 04 48 71	39 07 46 96 40	20 86 7 11 81	74 11 15 23 17
16 07 79 57 61	42 19 68 15 12	60 21 59 12 07	04 99 8 22 39	75 16 69 13 84

TA LA VII. DISTRI UCI N DE POISSON

($0 < \lambda < 1$)

λ	0	1	2			5	6			
0,0	1,0000	0,9900	0,9802	0,9704	0,9608	0,9512	0,9418	0,9324	0,9231	0,9139
0,1	0,9048	0,8958	0,8869	0,8781	0,8694	0,8607	0,8521	0,8437	0,8353	0,8270
0,2	0,8187	0,8106	0,8025	0,7945	0,7866	0,7788	0,7711	0,7634	0,7558	0,7483
0,3	0,7408	0,7334	0,7261	0,7189	0,7118	0,7047	0,6977	0,6907	0,6839	0,6771
0,4	0,6703	0,6636	0,6570	0,6505	0,6440	0,6376	0,6313	0,6250	0,6188	0,6126
0,5	0,6065	0,6005	0,5945	0,5886	0,5827	0,5770	0,5712	0,5655	0,5599	0,5543
0,6	0,5488	0,5434	0,5379	0,5326	0,5273	0,5220	0,5169	0,5117	0,5066	0,5016
0,7	0,4966	0,4916	0,4868	0,4819	0,4771	0,4724	0,4677	0,4630	0,4584	0,4538
0,8	0,4493	0,4449	0,4404	0,4360	0,4317	0,4274	0,4232	0,4190	0,4148	0,4107
0,9	0,4066	0,4025	0,3985	0,3946	0,3906	0,3867	0,3829	0,3791	0,3753	0,3716

($\lambda = 1\ 2\ \ldots\ldots 10$)

λ	1	2			5	6				10
$e^{-\lambda}$	0,36788	0,13534	0,04979	0,01832	0,006738	0,002479	0,000912	0,000335	0,000123	0,000045

Nota: Para obtener los valores de $e^{-\lambda}$ para otros valores de λ, emplear las leyes de los exponentes.

Ejemplo: $e^{-3,48} = (e^{-3,00})\,(e^{-0,48}) = (0,04979)\,(0,6188) = 0,03081$

TA LA VIII. INTERVALO DE CON IAN A DEL 5 (2 SIGMA)

L mites e error ± 2 o en *d*	VALORES PRESUMI LES DE *P*(+ = 100)									
	0 01	0 02	0 10	0 15	0 20	0 25	0 0	0 5	0 0	0 5
0,1	39.600	78.400	360.000	510.000	640.000	750.000	840.000	910.000	960.000	1.000.000
0,2	9.900	19.600	90.000	127.500	160.000	187.500	210.000	227.500	240.000	250.000
0,3	4.400	8.711	40.000	56.667	71.111	83.333	93.333	101.111	106.667	111.111
0,4	2.475	4.900	22.500	31.875	40.000	46.875	52.500	56.875	60.000	62.500
0,5	1.584	3.136	14.400	20.400	25.600	30.000	33.600	36.400	38.400	40.000
0,6	1.100	2.178	10.000	14.167	17.778	20.833	23.333	25.278	26.667	27.778
0,7	808	1.600	7.347	14.408	13.061	15.306	17.143	18.577	19.592	20.408
0,8	619	1.225	5.625	7.969	10.000	11.719	13.125	14.219	15.000	15.625
0,9	489	968	4.444	6.901	7.901	9.259	10.370	11.235	11.852	12.346
1,0	396	784	3.600	5.100	6.400	7.500	8.400	9.100	9.600	10.000
1,5	176	348	1.600	2.267	2.844	3.333	3.733	4.044	4.267	4.444
2,0	99	196	900	1.275	1.600	1.875	2.100	2.275	2.400	2.500
2,5	63	125	576	816	1.024	1.200	1.344	1.456	1.536	1.600
3,0	44	87	400	567	711	833	933	1.011	1.067	1.111
3,5	32	64	294	416	522	612	686	743	784	816
4,0	25	49	225	319	400	469	525	569	600	625
4,5	20	39	178	252	316	370	415	449	474	494
5,0	16	31	144	204	256	300	336	364	384	400
6,0	11	22	100	142	178	208	233	253	267	278
7,0	8	16	73	104	131	153	171	186	196	204
8,0	6	12	56	80	100	117	131	142	150	156
9,0	5	10	44	63	79	93	104	112	119	123
10,0	4	8	36	51	64	75	83	91	96	100
15,0	2	3	16	23	28	33	37	40	43	45
20,0	1	2	9	13	16	19	21	23	24	25
25,0	0,6	1	6	8	12	12	13	15	15	16

TA LA I . SEGURIDAD DEL 5 (2 SIGMA)

No. e elementos e la muestra *n*	0 01	0 02	0 0	0 0	0 05	0 10	0 15	0 20	0 25	0 0	0 5	0 0	0 5	0 50
	d =													
25	4,0	5,6	6,8	7,8	8,7	12,0	14,3	16,0	17,3	18,3	19,1	19,6	19,8	20,0
50	2,8	4,0	4,9	5,6	6,2	8,5	10,1	11,4	12,3	13,0	13,5	13,9	14,1	14,2
75	2,3	3,2	3,9	4,5	5,0	6,9	8,2	9,2	10,0	10,5	11,0	11,3	11,4	11,5
100	2,0	2,8	3,4	3,9	4,4	6,0	7,1	8,0	8,7	9,2	9,5	9,8	9,9	10,0
150	1,6	2,3	2,8	3,2	3,6	4,9	5,9	6,6	7,1	7,5	7,8	8,0	8,1	8,2
200	1,4	2,0	2,4	2,8	3,1	4,3	5,1	5,7	6,1	6,5	6,8	7,0	7,0	7,1
250	1,2	1,8	2,2	2,5	2,7	3,8	4,5	5,0	5,5	5,8	6,0	6,2	6,2	6,3
300	1,1	1,6	2,0	2,3	2,5	3,5	4,1	4,6	5,0	5,3	5,5	5,7	5,8	5,8
400	0,99	1,4	1,7	2,0	2,2	3,0	3,6	4,0	4,3	4,6	4,8	4,9	5,0	5,0
500	0,89	1,3	1,5	1,8	2,0	2,7	3,2	3,6	3,9	4,1	4,3	4,4	4,5	4,5
600	0,81	1,1	1,4	1,6	1,8	2,5	2,9	3,3	3,6	3,8	3,9	4,0	4,1	4,1
800	0,69	0,98	1,2	1,4	1,5	2,1	2,5	2,8	3,0	3,3	3,3	3,4	3,5	3,5
1.000	0,63	0,90	1,1	1,3	1,4	1,9	2,3	2,6	2,8	2,9	3,1	3,1	3,2	3,2
1.200	0,57	0,81	0,99	1,1	1,3	1,7	2,1	2,3	2,5	2,7	2,8	2,8	2,9	2,9
1.500	0,51	0,73	0,89	1,0	1,1	1,6	1,9	2,1	2,3	2,4	2,5	2,5	2,6	2,6
2.000	0,44	0,61	0,75	0,86	0,90	1,3	1,6	1,8	1,9	2,0	2,1	2,2	2,2	2,2
2.500	0,40	0,56	0,68	0,78	0,87	1,22	1,4	1,6	1,7	1,8	1,9	2,0	2,0	2,0
3.000	0,36	0,51	0,62	0,71	0,79	1,1	1,3	1,5	1,6	1,7	1,7	1,8	1,8	1,8
4.000	0,31	0,44	0,54	0,62	0,69	0,95	1,1	1,3	1,4	1,4	1,5	1,5	1,6	1,6
5.000	0,28	0,40	0,49	0,56	0,62	0,85	1,0	1,1	1,2	1,3	1,4	1,4	1,4	1,4
7.500	0,23	0,32	0,39	0,45	0,50	0,69	0,82	0,92	1,0	1,1	1,1	1,1	1,1	1,2
10.000	0,20	0,28	0,34	0,39	0,44	0,60	0,71	0,80	0,87	0,95	0,95	0,98	0,99	1,0
15.000	0,16	0,23	0,28	0,3	0,36	0,49	0,59	0,66	0,71	0,75	0,78	0,80	0,81	0,82
25.000	0,12	0,18	0,22	0,25	0,27	0,38	0,45	0,50	0,65	0,58	0,60	0,62	0,62	0,63
50.000	0,08	0,11	0,14	0,16	0,17	0,29	0,32	0,32	0,35	0,37	0,48	0,39	0,40	0,40

TA LA . SEGURIDAD DEL (SIGMA)

No. e elementos e la muestra *n*	0 01	0 02	0 0	0 0	0 05	0 10	0 15	0 20	0 25	0 0	0 5	0 0	0 5	0 50
	d =													
25	5,9	8,4	10,3	11,8	13,1	18,0	21,4	24,0	26,0	27,5	28,6	29,4	29,8	30,0
50	4,2	6,0	7,3	8,3	9,3	12,8	15,2	17,0	18,4	19,5	20,3	20,9	21,1	21,3
75	3,4	4,8	5,9	6,8	7,5	10,4	12,3	13,8	14,9	15,8	16,5	16,9	17,1	17,3
100	3,0	4,2	5,1	5,9	6,5	9,0	10,7	12,0	13,0	13,7	14,3	14,7	14,9	15,0
150	2,4	3,4	4,2	4,8	5,4	7,4	8,8	9,8	10,7	11,3	11,7	12,1	12,2	12,3
200	2,1	3,0	3,6	4,2	4,6	6,4	7,6	8,5	9,2	9,8	10,2	10,4	10,6	10,7
250	1,9	2,6	3,2	3,7	4,1	5,7	6,7	7,6	8,2	8,7	9,0	9,3	9,4	9,5
300	1,7	2,4	3,0	3,4	3,8	5,2	6,2	7,0	7,5	8,0	8,3	8,5	8,6	8,7
400	1,5	2,1	2,6	2,9	3,3	4,5	5,4	6,0	6,6	6,9	7,2	7,4	7,4	7,5
500	1,3	1,9	2,3	2,6	2,9	4,1	4,8	5,4	5,8	6,2	6,4	6,6	6,7	6,8
600	1,2	1,7	2,1	2,4	2,7	3,7	4,4	4,9	5,3	5,6	5,9	6,0	6,1	6,2
800	1,0	1,5	1,8	2,1	2,3	3,2	3,7	4,2	4,5	4,8	5,0	5,1	5,2	5,3
1.000	0,95	1,3	1,6	1,9	2,1	2,9	3,4	3,8	4,2	4,4	4,6	4,7	4,8	4,8
1.200	0,86	1,2	1,5	1,7	1,9	2,6	3,1	3,5	3,8	4,0	4,1	4,3	4,3	4,4
1.500	0,77	1,1	1,3	1,5	1,7	2,3	2,8	3,1	3,4	3,6	3,7	3,8	3,9	3,9
2.000	0,65	0,92	1,1	1,3	1,4	2,0	2,4	2,6	2,9	3,0	3,1	3,2	3,3	3,3
2.500	0,59	0,84	1,0	1,2	1,3	1,8	2,1	2,4	2,6	2,7	2,9	2,9	3,0	3,0
3.000	0,54	0,76	0,93	1,1	1,2	1,6	1,9	2,2	2,4	2,5	2,6	2,6	2,7	2,7
4.000	0,47	0,66	0,81	0,92	1,0	1,4	1,7	1,9	2,1	2,2	2,3	2,3	2,4	2,4
5.000	0,42	0,60	0,73	0,73	0,93	1,3	1,5	1,7	1,8	2,0	2,0	2,1	2,1	2,1
7.500	0,34	0,48	0,59	0,68	0,75	1,0	1,2	1,4	1,5	1,6	1,6	1,7	1,7	1,7
10.000	0,30	0,42	0,51	0,59	0,65	0,90	1,1	1,2	1,3	1,4	1,4	1,5	1,5	1,5
15.000	0,24	0,34	0,42	0,48	0,54	0,74	0,89	0,98	1,1	1,1	1,2	1,2	1,2	1,2
25.000	0,19	0,26	0,32	0,37	0,41	0,57	0,67	0,76	0,82	0,87	0,90	0,93	0,94	0,95
50.000	0,12	0,17	0,21	0,24	0,26	0,36	0,43	0,48	0,51	0,56	0,57	0,59	0,60	0,60

TA LA I. En la etermina i n e una muestra (***n***) **sacada de una población finita,** para m rgenes e error e 1 2 5 10 en la ip tesis e = 50

Margen de confianza del 95%

Amplitu e la pobla i n N	Amplitu e la muestra para m rgenes e error abajo in i a os					
	±1	±2	±	±	±5	±10
500 . . .					222	83
1.000 . . .				485	286	91
1.500 . . .			638	441	316	94
2.000 . . .			714	476	333	95
2.500 . . .		1.250	769	500	345	96
3.000 . . .		1.364	811	517	353	97
3.500 . . .		1.458	843	530	359	97
4.000 . . .		1.538	870	541	364	98
4.500 . . .		1.607	891	549	367	98
5.000 . . .		1.667	909	556	370	98
6.000 . . .		1.765	938	566	375	98
7.000 . . .		1.842	949	574	378	99
8.000 . . .		1.905	976	580	381	99
9.000 . . .		1.957	989	584	383	99
10.000 . . .	5.000	2.000	1.000	588	385	99
15.000 . . .	6.000	2.143	1.034	600	390	99
20.000 . . .	6.667	2.222	1.059	606	392	100
25.000 . . .	7.143	2.273	1.064	610	394	100
50.000 . . .	8.333	2.381	1.087	617	397	100
100.000 . . .	10.000	2.500	1.111	625	400	100

TA LA II. En la etermina i n e una muestra (***n***) **sacada de una población finita,** para m rgenes e error e 1 2 5 en la ip tesis e = 50

Margen de confianza del 99,7%

Amplitu e la pobla i n N	Amplitu e la muestra para m rgenes e error abajo in i a os				
	±1	±2	±	±	±5
500					
1.000					474
1.500				726	563
2.000				826	621
2.500				900	662
3.000			1.364	958	692
3.500			1.458	1.003	716
4.000			1.539	1.041	735
4.500			1.607	1.071	750
5.000			1.667	1.098	763
6.000		2.903	1.765	1.139	783
7.000		3.119	1.842	1.171	798
8.000		3.303	1.905	1.196	809
9.000		3.462	1.957	1.216	818
10.000		3.600	2.000	1.233	826
15.000		4.091	2.143	1.286	849
20.000		4.390	2.222	1.314	861
25.000	11.842	4.592	2.273	1.331	869
50.000	15.517	5.056	2.381	1.368	884
100.000	22.500	5.625	2.500	1.406	900

TA LA III. Probabili a es aso ia as on alores tan pe ue os omo los alores obser a os e U en la prueba e Mann itney

$n_2 =$

U \ n_1	1	2	
0	.250	.100	.050
1	.500	.200	.100
2	.750	.400	.200
3		.600	.350
4			.500
5			.650

$n_2 =$

U \ n_1	1	2		
0	.200	.067	.028	.014
1	.400	.133	.057	.029
2	.600	.267	.144	.057
3		.400	.200	.100
4		.600	.314	.171
5			.429	.243
6			.571	.343
7				.443
8				.557

$n_2 = 5$

U \ n_1	1	2			5
0	.167	.047	.018	.008	.004
1	.333	.095	.036	.016	.008
2	.500	.190	.071	.032	.016
3	.667	.286	.125	.056	.028
4		.429	.196	.095	.048
5		.571	.286	.143	.075
6			.393	.206	.111
7			.500	.278	.155
8			.607	. 65	.210
9				.452	.274
10				.548	.345
11					.421
12					.500
13					.579

$n_2 = 6$

U \ n_1	1	2			5	6
0	.143	.036	.012	.005	.002	.001
1	.286	.071	.024	.010	.004	.002
2	.428	.143	.048	.019	.009	.004
3	.571	.214	.083	.033	.015	.008
4		.321	.131	.057	.026	.013
5		.429	.190	.086	.041	.021
6		.571	.274	.129	.063	.032
7			.357	.176	.089	.047
8			.452	.238	.123	.066
9			.548	.305	.165	.090
10				.381	.214	.120
11				.457	.268	.155
12				.545	.331	.197
13					.396	.242
14					.465	.294
15					.535	.350
16						.409
17						.469
18						.531

Reproducida de Mann, H.B. y Whitney, D.R. 1947. En una prueba de si una o dos variables aleatorias es estocásticamente mayor que la otra.

TA LA III. Probabili a es aso ia as on alores tan pe ue os omo los alores obser a os e U en la prueba e Mann itney (ontinua i n)

$n_2 =$

U \ n_1	1	2			5	6	
0	.125	.028	.008	.003	.001	.001	.000
1	.250	.056	.017	.006	.003	.001	.001
2	.375	.111	.033	.012	.005	.002	.001
3	.500	.167	.058	.021	.009	.004	.002
4	.625	.250	.092	.036	.015	.007	.003
5		.333	.133	.055	.024	.011	.006
6		.444	.192	.082	.037	.017	.009
7		.556	.258	.115	.053	.026	.013
8			.333	.158	.074	.037	.019
9			.417	.206	.101	.051	.027
10			.500	.264	.134	.069	.036
11			.583	.324	.172	.090	.049
12				.394	.216	.117	.064
13				.464	.265	.147	.082
14				.538	.319	.183	.104
15					.378	.223	.130
16					.438	.267	.159
17					.500	.314	.191
18					.562	.365	.228
19						.418	.267
20						.473	.310
21						.527	.355
22							.402
23							.451
24							.500
25							.549

Reproducida de Mann, H.B. y Whitney, D.R. 1947. En una prueba de si una o dos variables aleatorias es estocásticamente mayor que la otra.

TA LA III. Probabili a es aso ia as on alores tan pe ue os omo los alores obser a os e U en la prueba e Mann itney (ontinua i n)

$n_2 =$

U \ n_1	1	2			5	6			*t*	Normal
0	.111	.022	.006	.002	.001	.000	.000	.000	3,308	.001
1	.222	.044	.012	.004	.002	.001	.000	.000	3,203	.001
2	.333	.089	.024	.008	.003	.001	.001	.000	3,098	.001
3	.444	.133	.042	.014	.005	.002	.001	.001	2,993	.001
4	.556	.200	.067	.024	.009	.004	.002	.001	2,888	.002
5		.267	.097	.036	.015	.006	.003	.001	2,783	.003
6		.356	.139	.055	.023	.010	.005	.002	2,678	.004
7		.444	.188	.077	.033	.015	.007	.003	2,573	.005
8		.556	.248	.107	.047	.021	.010	.005	2,468	.007
9			.315	.141	.064	.030	.014	.007	2,363	.009
10			.387	.184	.085	.041	.020	.010	2,258	.012
11			.461	.230	.111	.054	.027	.014	2,153	.016
12			.539	.285	.142	.071	.036	.019	2,048	.020
13				.341	.177	.091	.047	.025	1,943	.026
14				.404	.217	.114	.060	.032	1,838	.033
15				.467	.262	.141	.076	.041	1,733	.041
16				.533	.311	.172	.095	.052	1,628	.052
17					.362	.207	.116	.065	1,523	.064
18					.416	.245	.140	.080	1,418	.078
19					.472	.286	.168	.097	1,313	.094
20					.528	.331	.198	.117	1,208	.113
21						.377	.232	.139	1,102	.135
22						.426	.268	.164	.998	.159
23						.475	.306	.191	.893	.185
24						.525	.347	.221	.788	.215
25							.389	.253	.683	.247
26							.433	.287	.578	.282
27							.478	.323	.473	.318
28							.522	.360	.368	.356
29								.399	.263	.396
30								.439	.158	.437
31								.480	.052	.481
32								.520		

Reproducida de Mann, H.B. y Whitney, D.R. 1947. En una prueba de si una o dos variables aleatorias es estocásticamente mayor que la otra.

TA LA IV. Valores r ti os e U en la prueba e Mann itney.
Valores r ti os e U para una prueba e una ola en $\alpha = 0\,001$
o para una prueba e os olas en $\alpha = 0\,002$.

n_1 / n_2		10	11	12	1	1	15	16	1	1	1	20
1												
2												
3									0	0	0	0
4		0	0	0	1	1	1	2	2	3	3	3
5	1	1	2	2	3	3	4	5	5	6	7	7
6	2	3	4	4	5	6	7	8	9	10	11	12
7	3	5	6	7	8	9	10	11	13	14	15	16
8	5	6	8	9	11	12	14	15	17	18	20	21
9	7	8	10	12	14	15	17	19	21	23	25	26
10	8	10	12	14	17	19	21	23	25	27	29	32
11	10	12	15	17	20	22	24	27	29	32	34	37
12	12	14	17	20	23	25	28	31	34	37	40	42
13	14	17	20	23	26	29	32	35	38	42	45	48
14	15	19	22	25	29	32	36	39	43	46	50	54
15	17	21	24	28	32	36	40	43	47	51	55	59
16	19	23	27	31	35	39	43	48	52	56	60	65
17	21	25	29	34	38	43	47	52	57	61	66	70
18	23	27	32	37	42	46	51	56	61	66	71	76
19	25	29	34	40	45	50	55	60	66	71	77	82
20	26	32	37	42	48	54	59	65	70	76	82	88

Tomada abreviadamente de las tablas 1, 3, 5 y 7 de Auble, D. 1953. Tablas extendidas para la estadística de Mann Whitney, Bulletin of the Institute of Educational Research at Indiana University.

TA LA IV. Valores r ti os e U en la prueba e Mann itney.
Valores r ti os e U para una prueba e una ola en $\alpha = 0\ 01$
o para una prueba e os olas en $\alpha = 0\ 02$. (Continua i n)

n_1 / n_2		10	11	12	1	1	15	16	1	1	1	20
1												
2					0	0	0	0	0	0	1	1
3	1	1	1	2	2	2	3	3	4	4	4	5
4	3	3	4	5	5	6	7	7	8	9	9	10
5	5	6	7	8	9	10	11	12	13	14	15	16
6	7	8	9	11	12	13	15	16	18	19	20	22
7	9	11	12	14	16	17	19	21	23	24	26	28
8	11	13	15	17	20	22	24	26	28	30	32	34
9	14	16	18	21	23	26	28	31	33	36	38	40
10	16	19	22	24	27	30	33	36	38	41	44	47
11	18	22	25	28	31	34	37	41	44	47	50	53
12	21	24	28	31	35	38	42	46	49	53	56	60
13	23	27	31	35	39	43	47	51	55	59	63	67
14	26	30	34	38	43	47	51	56	60	65	69	73
15	28	33	37	42	47	51	56	61	66	70	75	80
16	31	36	41	46	51	56	61	66	71	76	82	87
17	33	38	44	49	55	60	66	71	77	82	88	93
18	36	41	47	53	59	65	70	76	82	88	94	100
19	38	44	50	56	63	69	75	82	88	94	101	107
20	40	47	53	60	67	73	80	87	93	100	107	114

Tomada abreviadamente de las tablas 1, 3, 5 y 7 de Auble, D. 1953. Tablas extendidas para la estadística de Mann Whitney, Bulletin of the Institute of Educational Research at Indiana University.

TA LA IV. Valores r ti os e U en la prueba e Mann itney. Valores r ti os e U para una prueba e una ola en $\alpha = 0\,025$ o para una prueba e os olas en $\alpha = 0\,05$. (Continua i n)

n_1 / n_2		10	11	12	1	1	15	16	1	1	1	20
1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

Tomada abreviadamente de las tablas 1, 3, 5 y 7 de Auble, D. 1953. Tablas extendidas para la estadística de Mann Whitney, Bulletin of the Institute of Educational Research at Indiana University.

TA LA IV. Valores r ti os e U en la prueba e Mann itney. Valores r ti os e U para una prueba e una ola en $\alpha = 0\,05$ o para una prueba e os olas en $\alpha = 0\,10$. (Continua i n)

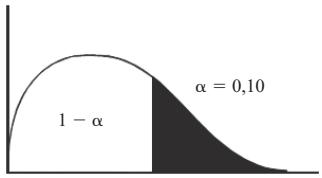
n_1 / n_2		10	11	12	1	1	15	16	1	1	1	20
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

Tomada abreviadamente de las tablas 1, 3, 5 y 7 de Auble, D. 1953. Tablas extendidas para la estadística de Mann Whitney, Bulletin of the Institute of Educational Research at Indiana University.

TA LA V. PERCENTILLAS DE LAS DISTRI UCIONES (x^2 .*l*.)

	.05	.1	.5	1.0	2.5	5.0	10	20	0	0	50	60	0	0	0	5	.5		.5	.	. 5	
1	$.0^{6}39$	$.0^{5}157$	$.0^{4}39$	$.0^{3}16$	$.0^{3}98$	$.0^{2}39$	.016	.064	.148	.275	.455	.708	1.07	1.64	2.71	3.84	5.02	6.64	7.88	10.83	12.12	1
2	.001	.001	.005	.010	.025	.052	.106	.223	.356	.511	.693	.916	1.20	1.61	2.30	3.00	3.69	4.61	5.30	6.91	7.60	2
3	.005	.008	.024	.038	.072	.117	.195	.335	.475	.623	.789	.982	1.22	1.55	2.08	2.60	3.12	3.78	4.28	5.42	5.91	3
4	.016	.023	.052	.074	.121	.178	.266	.412	.549	.688	.839	1.011	1.22	1.50	1.94	2.37	2.79	3.32	3.72	4.62	5.00	4
5	.032	.042	.082	.111	.166	.229	.322	.469	.600	.731	.870	1.03	1.21	1.46	1.85	2.21	2.57	3.02	3.35	4.10	4.42	5
6	.050	.064	.113	.145	.206	.272	.367	.512	.638	.762	.891	1.04	1.21	1.43	1.77	2.10	2.41	2.80	3.09	3.74	4.02	6
7	.069	.085	.141	.177	.241	.310	.405	.546	.667	.785	.907	1.04	1.20	1.40	1.72	2.01	2.29	2.64	2.90	3.47	3.72	7
8	.089	.107	.168	.206	.272	.342	.436	.574	.691	.803	.918	1.04	1.19	1.38	1.67	1.94	2.19	2.51	2.74	3.27	3.48	8
9	.108	.128	.193	.232	.300	.369	.463	.598	.710	.817	.927	1.05	1.18	1.36	1.63	1.88	2.11	2.41	2.62	3.10	3.30	9
10	.126	.148	.216	.256	.325	.394	.487	.618	.727	.830	.934	1.05	1.18	1.34	1.60	1.83	2.05	2.32	2.52	2.96	3.14	10
11	.144	.167	.237	.278	.347	.416	.507	.635	.741	.840	.940	1.05	1.17	1.33	1.57	1.79	1.99	2.25	2.13	2.84	3.01	11
12	.161	.184	.256	.298	.367	.436	.525	.651	.753	.848	.945	1.05	1.17	1.32	1.55	1.75	1.94	2.18	2.36	2.74	2.90	12
13	.177	.201	.274	.316	.385	.453	.542	.664	.764	.856	.949	1.05	1.16	1.31	1.52	1.72	1.90	2.13	2.29	2.66	2.81	13
14	.193	.217	.291	.333	.402	.469	.556	.676	.773	.863	.953	1.05	1.16	1.30	1.50	1.69	1.87	2.08	2.24	2.58	2.72	14
15	.207	.232	.307	.349	.418	.484	.570	.687	.781	.869	.956	1.05	1.15	1.29	1.49	1.67	1.83	2.04	2.19	2.51	2.65	15
16	.221	.246	.321	.363	.432	.498	.582	.697	.789	.874	.959	1.05	1.15	1.28	1.47	1.64	1.80	2.00	2.14	2.45	2.58	16
17	.234	.260	.335	.377	.445	.510	.593	.706	.796	.879	.961	1.05	1.15	1.27	1.46	1.62	1.78	1.97	2.10	2.40	2.52	17
18	.247	.272	.348	.390	.457	.522	.604	.714	.802	.883	.963	1.05	1.14	1.26	1.44	1.60	1.75	1.93	2.06	2.35	2.47	18
19	.258	.285	.360	.402	.469	.532	.613	.722	.808	.887	.965	1.05	1.14	1.26	1.43	1.59	1.73	1.90	2.03	2.31	2.42	19
20	.270	.296	.372	.413	.480	.543	.622	.729	.813	.890	.967	1.05	1.14	1.25	1.42	1.57	1.71	1.88	2.00	2.27	2.37	20
22	.291	.317	.393	.434	.499	.561	.638	.742	.823	.897	.970	1.05	1.13	1.24	1.40	1.54	1.67	1.83	1.95	2.19	2.30	22
24	.310	.337	.412	.452	.517	.577	.652	.753	.831	.902	.972	1.05	1.13	1.23	1.38	1.52	1.64	1.79	1.90	2.13	2.23	24
26	.328	.355	.429	.469	.532	.592	.665	.762	.838	.907	.974	1.05	1.12	1.22	1.37	1.50	1.61	1.76	1.86	2.08	2.17	26
28	.345	.371	.445	.484	.547	.605	.676	.771	.845	.911	.976	1.04	1.12	1.22	1.35	1.48	1.59	1.72	1.82	2.03	2.12	28
30	.360	.386	.460	.498	.560	.616	.687	.779	.850	.915	.978	1.04	1.12	1.21	1.34	1.46	1.57	1.70	1.79	1.99	2.07	30
35	.394	.420	.491	.529	.588	.642	.708	.795	.862	.922	.981	1.04	1.11	1.19	1.32	1.42	1.52	1.64	1.72	1.90	1.98	35
40	.423	.448	.518	.554	.611	.663	.726	.809	.872	.928	.983	1.04	1.10	1.18	1.30	1.39	1.48	1.59	1.67	1.84	1.90	40
45	.448	.472	.540	.576	.630	.680	.741	.820	.880	.933	.985	1.04	1.10	1.17	1.28	1.37	1.45	1.55	1.63	1.78	1.84	45
50	.469	.494	.560	.594	.647	.695	.754	.829	.886	.937	.987	1.04	1.09	1.16	1.26	1.35	1.43	1.52	1.59	1.73	1.79	50
55	.488	.512	.577	.610	.662	.708	.765	.837	.892	.941	.988	1.04	1.09	1.16	1.25	1.33	1.41	1.50	1.56	1.69	1.75	55
60	.506	.529	.592	.625	.675	.720	.774	.844	.897	.944	.989	1.04	1.09	1.15	1.24	1.32	1.39	1.47	1.53	1.66	1.71	60
70	.535	.558	.618	.649	.697	.739	.790	.856	.905	.949	.990	1.03	1.08	1.14	1.22	1.29	1.36	1.43	1.49	1.60	1.65	70
80	.560	.582	.640	.669	.714	.755	.803	.865	.911	.952	.992	1.03	1.08	1.13	1.21	1.27	1.33	1.40	1.45	1.56	1.60	80
90	.581	.602	.658	.686	.729	.768	.814	.873	.917	.955	.993	1.03	1.07	1.12	1.20	1.26	1.31	1.38	1.43	1.52	1.56	90
100	.599	.619	.673	.701	.742	.779	.824	.879	.921	.958	.993	1.03	1.07	1.12	1.18	1.24	1.30	1.36	1.40	1.49	1.53	100
120	.629	.6.48	.699	.724	.763	.798	.839	.890	.929	.962	.994	1.03	1.06	1.11	1.17	1.22	1.27	1.32	1.36	1.45	1.48	120
140	.653	.671	.719	.743	.780	.812	.850	.898	.934	.965	.995	1.03	1.06	1.10	1.16	1.20	1.25	1.30	1.33	1.41	1.44	140
160	.673	.690	.736	.758	.793	.824	.860	.905	.939	.968	.996	1.02	1.06	1.09	1.15	1.19	1.23	1.28	1.31	1.38	1.41	160
180	.689	.706	.749	.771	.804	.833	.868	.910	.942	.970	.996	1.02	1.05	1.09	1.14	1.18	1.22	1.26	1.29	1.36	1.38	180
200	.703	.719	.761	.782	.814	.841	.874	.915	.945	.972	.997	1.02	1.05	1.08	1.13	1.17	1.21	1.25	1.28	1.34	1.36	200
250	.732	.746	.785	.804	.832	.858	.887	.924	.951	.975	.997	1.02	1.04	1.07	1.12	1.15	1.18	1.22	1.25	1.30	1.32	250
300	.753	.767	.802	.820	.846	.870	.897	.931	.956	.977	.998	1.02	1.04	1.07	1.11	1.14	1.17	1.20	1.22	1.27	1.29	300
350	.770	.783	.816	.833	.857	.879	.904	.936	.959	.979	.998	1.02	1.04	1.06	1.10	1.13	1.15	1.18	1.21	1.25	1.27	350
400	.784	.796	.827	.843	.866	.887	.911	.940	.962	.981	.998	1.02	1.04	1.06	1.09	1.12	1.14	1.17	1.19	1.24	1.25	400
450	.795	.807	.837	.852	.874	.893	.916	.944	.964	.982	.999	1.02	1.03	1.06	1.09	1.11	1.13	1.16	1.18	1.22	1.23	450
500	.805	.816	.845	.859	.880	.898	.920	.946	.966	.983	.999	1.01	1.03	1.05	1.08	1.11	1.13	1.15	1.17	1.21	1.22	500
750	. 839	.848	.872	.884	.901	.917	.934	.956	.972	.986	.999	1.01	1.03	1.04	1.07	1.09	1.10	1.12	1.14	1.17	1.18	750
1000	.859	.868	.889	.899	.914	.928	.943	.962	.976	.988	.999	1.01	1.02	1.04	1.06	1.07	1.09	1.11	1.12	1.14	1.15	1000
5000	.936	.939	.949	.954	.961	.967	.974	.983	.989	.995	1.00	1.00	1.01	1.02	1.02	1.03	1.04	1.05	1.05	1.06	1.07	5000
∝	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	∝

TA LA VI. DISTRI UCI N



α = 0 10

$_2$	$_1$ = gra os e liberta el numera or									
	1	2			5	6				10
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
7	3,59	3,26	3,07	2,96	2,88	2,83	2,79	2,75	2,72	2,70
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82
35	2,85	2,46	2,25	2,11	2,02	1,95	1,90	1,85	1,82	1,79
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71
80	2,77	2,37	2,15	2,02	1,92	1,85	1,79	1,75	1,71	1,68
100	2,76	2,36	2,14	2,00	1,91	1,83	1,78	1,73	1,69	1,66
200	2,73	2,33	2,11	1,97	1,88	1,80	1,75	1,70	1,66	1,63
500	2,72	2,31	2,09	1,96	1,86	1,79	1,73	1,68	1,64	1,61
1000	2,71	2,31	2,09	1,95	1,85	1,78	1,72	1,68	1,64	1,61

Nota El valor de α corresponde a probabilidades unilaterales.

TA LA VI. DISTRI UCI N

$\alpha = 0\ 10$

(Continua i n)

$_2$	$_1$ = gra os e liberta el numera or									
	11	12	15	20	25	0	0	50	100	1000
1	60,47	60,71	61,22	61,74	62,06	62,26	62,53	62,69	63,00	63,29
2	9,40	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
3	5,22	5,22	5,20	5,19	5,17	5,17	5,16	5,15	5,14	5,13
4	3,91	3,90	3,87	3,84	3,83	3,82	3,80	3,80	3,78	3,76
5	3,28	3,27	3,24	3,21	3,19	3,17	3,16	3,15	3,13	3,11
6	2,92	2,90	2,87	2,84	2,81	2,80	2,78	2,77	2,75	2,72
7	2,68	2,67	2,63	2,59	2,57	2,56	2,54	2,52	2,50	2,47
8	2,52	2,50	2,46	2,42	2,40	2,38	2,36	2,35	2,32	2,30
9	2,40	2,38	2,34	2,30	2,27	2,25	2,23	2,22	2,19	2,16
10	2,30	2,28	2,24	2,20	2,17	2,16	2,13	2,12	2,09	2,06
11	2,23	2,21	2,17	2,12	2,10	2,08	2,05	2,04	2,00	1,98
12	2,17	2,15	2,10	2,06	2,03	2,01	1,99	1,97	1,94	1,91
13	2,12	2,10	2,05	2,01	1,98	1,96	1,93	1,92	1,88	1,85
14	2,07	2,05	2,01	1,96	1,93	1,91	1,89	1,87	1,83	1,80
15	2,04	2,02	1,97	1,92	1,89	1,87	1,85	1,83	1,79	1,76
16	2,01	1,99	1,94	1,89	1,86	1,84	1,81	1,79	1,76	1,72
17	1,98	1,96	1,91	1,86	1,83	1,81	1,78	1,76	1,73	1,69
18	1,95	1,93	1,89	1,84	1,80	1,78	1,75	1,74	1,70	1,66
19	1,93	1,91	1,86	1,81	1,78	1,76	1,73	1,71	1,67	1,64
20	1,91	1,89	1,84	1,79	1,76	1,74	1,71	1,69	1,65	1,61
21	1,90	1,87	1,83	1,78	1,74	1,72	1,69	1,67	1,63	1,59
22	1,88	1,86	1,81	1,76	1,73	1,70	1,67	1,65	1,61	1,57
23	1,87	1,84	1,80	1,74	1,71	1,69	1,66	1,64	1,59	1,55
24	1,85	1,83	1,78	1,73	1,70	1,67	1,64	1,62	1,58	1,54
25	1,84	1,82	1,77	1,72	1,68	1,66	1,63	1,61	1,56	1,52
26	1,83	1,81	1,76	1,71	1,67	1,65	1,61	1,59	1,55	1,51
27	1,82	1,80	1,75	1,70	1,66	1,64	1,60	1,58	1,54	1,50
28	1,81	1,79	1,74	1,69	1,65	1,63	1,59	1,57	1,53	1,48
29	1,80	1,78	1,73	1,68	1,64	1,62	1,58	1,56	1,52	1,47
30	1,79	1,77	1,72	1,67	1,63	1,61	1,57	1,55	1,51	1,46
35	1,76	1,74	1,69	1,63	1,60	1,57	1,53	1,51	1,47	1,42
40	1,74	1,71	1,66	1,61	1,57	1,54	1,51	1,48	1,43	1,38
50	1,70	1,68	1,63	1,57	1,53	1,50	1,46	1,44	1,39	1,33
60	1,68	1,66	1,60	1,54	1,50	1,48	1,44	1,41	1,36	1,30
80	1,65	1,63	1,57	1,51	1,47	1,44	1,40	1,38	1,32	1,25
100	1,64	1,61	1,56	1,49	1,45	1,42	1,38	1,35	1,29	1,22
200	1,60	1,58	1,52	1,46	1,41	1,38	1,34	1,31	1,24	1,16
500	1,58	1,56	1,50	1,44	1,39	1,36	1,31	1,28	1,21	1,11
1000	1,58	1,55	1,49	1,43	1,38	1,35	1,30	1,27	1,20	1,08

TA LA VI. DISTRI UCI N

$\alpha = 0\,05$

(Continua i n)

^{2}	$_{1}$ = gra os e liberta el numera or									
	1	2			5	6				10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,97
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,73
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85
1000	3,85	3,01	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84

TA LA VI. DISTRI UCI N

$\alpha = 0\,05$

(Continua i n)

$_2$	$_1$ = gra os e liberta el numera or									
	11	12	15	20	25	0	0	50	100	1000
1	242,98	243,91	245,96	248,01	249,26	250,08	251,15	251,77	253,01	254,17
2	19,40	19,41	19,43	19,45	19,46	19,46	19,47	19,48	19,49	19,50
3	8,76	8,74	8,70	8,66	8,63	8,62	8,59	8,58	8,55	8,53
4	5,94	5,91	5,86	5,80	5,77	5,74	5,72	5,70	5,66	5,63
5	4,70	4,68	4,62	4,56	4,52	4,50	4,46	4,44	4,41	4,37
6	4,03	4,00	3,94	3,87	3,84	3,81	3,77	3,75	3,71	3,67
7	3,60	3,57	3,51	3,44	3,40	3,38	3,34	3,32	3,27	3,23
8	3,31	3,28	3,22	3,15	3,11	3,08	3,04	3,02	2,97	2,93
9	3,10	3,07	3,01	2,94	2,89	2,86	2,83	2,80	2,76	2,71
10	2,94	2,91	2,85	2,77	2,73	2,70	2,66	2,64	2,59	2,54
11	2,82	2,79	2,72	2,65	2,60	2,57	2,53	2,51	2,46	2,41
12	2,72	2,69	2,62	2,54	2,50	2,47	2,43	2,40	2,35	2,30
13	2,63	2,60	2,53	2,46	2,41	2,38	2,34	2,31	2,26	2,21
14	2,57	2,53	2,46	2,39	2,34	2,31	2,27	2,24	2,19	2,14
15	2,51	2,48	2,40	2,33	2,28	2,25	2,20	2,18	2,12	2,07
16	2,46	2,42	2,35	2,28	2,23	2,19	2,15	2,12	2,07	2,02
17	2,41	2,38	2,31	2,23	2,18	2,15	2,10	2,08	2,02	1,97
18	2,37	2,34	2,27	2,19	2,14	2,11	2,06	2,04	1,98	1,92
19	2,34	2,31	2,23	2,16	2,11	2,07	2,03	2,00	1,94	1,88
20	2,31	2,28	2,20	2,12	2,07	2,04	1,99	1,97	1,91	1,85
21	2,28	2,25	2,18	2,10	2,05	2,01	1,96	1,94	1,88	1,82
22	2,26	2,23	2,15	2,07	2,02	1,98	1,94	1,91	1,85	1,79
23	2,24	2,20	2,13	2,05	2,00	1,96	1,91	1,88	1,82	1,76
24	2,22	2,18	2,11	2,03	1,97	1,94	1,89	1,86	1,80	1,74
25	2,20	2,16	2,09	2,01	1,96	1,92	1,87	1,84	1,78	1,72
26	2,18	2,15	2,07	1,99	1,94	1,90	1,85	1,82	1,76	1,70
27	2,17	2,13	2,06	1,97	1,92	1,88	1,84	1,81	1,74	1,68
28	2,15	2,12	2,04	1,96	1,91	1,87	1,82	1,79	1,73	1,66
29	2,14	2,10	2,03	1,94	1,89	1,85	1,81	1,77	1,71	1,65
30	2,13	2,09	2,01	1,93	1,88	1,84	1,79	1,76	1,70	1,63
35	2,07	2,04	1,96	1,88	1,82	1,79	1,74	1,70	1,63	1,57
40	2,04	2,00	1,92	1,84	1,78	1,74	1,69	1,66	1,59	1,52
50	1,99	1,95	1,87	1,78	1,73	1,69	1,63	1,60	1,52	1,45
60	1,95	1,92	1,84	1,75	1,69	1,65	1,59	1,56	1,48	1,40
80	1,91	1,88	1,79	1,70	1,64	1,60	1,54	1,51	1,43	1,34
100	1,89	1,85	1,77	1,68	1,62	1,57	1,52	1,48	1,39	1,30
200	1,84	1,80	1,72	1,62	1,56	1,52	1,46	1,41	1,32	1,21
500	1,81	1,77	1,69	1,59	1,53	1,48	1,42	1,38	1,28	1,14
1000	1,80	1,76	1,68	1,58	1,52	1,47	1,41	1,36	1,26	1,11

TA LA VI. DISTRI UCI N

$\alpha = 0\,01$

(Continua i n)

$_2$	$_1$ = gra os e liberta el numera or									
	1	2			5	6				10
1	4.052,00	4.999,50	5.403,00	5.625,00	5.764,00	5.859,00	5.928,0	5.982,0	6.022,0	6.056,0
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,50	27,34	27,22
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
35	7,42	5,27	4,40	3,91	3,59	3,37	3,20	3,07	2,96	2,88
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34

TA LA VI. DISTRI UCI N

$\alpha = 0\,01$

(Continua i n)

$_2$	$_1$ = gra os e liberta el numera or									
	11	12	15	20	25	0	0	50	100	1000
1	6.082,00	6.106,00	6.157,00	6.209,00	6.260,00	6.261,00	6.287,00	6.302,00	6.334,00	6.366,0
2	99,41	99,42	99,43	99,45	99,46	99,46	99,47	99,48	99,49	99,51
3	27,12	27,03	26,85	26,67	26,58	26,50	26,41	26,35	26,24	26,14
4	14,45	14,37	14,19	14,02	13,91	13,84	13,75	13,69	13,58	13,48
5	9,96	9,89	9,72	9,55	9,45	9,38	9,30	9,24	9,13	9,03
6	7,79	7,72	7,56	7,40	7,29	7,23	7,15	7,09	6,99	6,89
7	6,54	6,47	6,31	6,16	6,06	5,99	5,91	5,86	5,75	5,66
8	5,73	5,67	5,52	5,30	5,26	5,20	5,12	5,07	4,96	4,87
9	5,18	5,11	4,96	4,81	4,71	4,65	4,57	4,52	4,41	4,32
10	4,77	4,71	4,56	4,41	4,31	4,25	4,17	4,12	4,01	3,92
11	4,46	4,40	4,25	4,10	4,00	3,94	3,86	3,81	3,71	3,61
12	4,22	4,16	4,01	3,86	3,76	3,70	3,62	3,57	3,47	3,37
13	4,02	3,96	3,82	3,66	3,57	3,51	3,43	3,38	3,27	3,18
14	3,86	3,80	3,66	3,51	3,41	3,35	3,27	3,22	3,11	3,02
15	3,73	3,67	3,52	3,37	3,28	3,21	3,13	3,08	2,98	2,88
16	3,62	3,55	3,41	3,26	3,16	3,10	3,02	2,97	2,86	2,76
17	3,52	3,46	3,31	3,16	3,07	3,00	2,92	2,87	2,76	2,66
18	3,43	3,37	3,23	3,08	2,98	2,92	2,84	2,78	2,68	2,58
19	3,36	3,30	3,15	3,00	2,91	2,84	2,76	2,71	2,60	2,50
20	3,29	3,23	3,09	2,94	2,84	2,78	2,69	2,64	2,54	2,43
21	3,24	3,17	3,03	2,88	2,78	2,72	2,64	2,58	2,48	2,37
22	3,18	3,12	2,98	2,83	2,73	2,67	2,58	2,53	2,42	2,32
23	3,14	3,07	2,93	2,78	2,69	2,62	2,54	2,48	2,37	2,27
24	3,09	3,03	2,89	2,74	2,64	2,58	2,49	2,44	2,33	2,22
25	3,06	2,99	2,85	2,70	2,60	2,54	2,45	2,40	2,29	2,18
26	3,02	2,96	2,81	2,66	2,57	2,50	2,42	2,36	2,25	2,14
27	2,99	2,93	2,78	2,63	2,54	2,47	2,38	2,33	2,22	2,11
28	2,96	2,90	2,75	2,60	2,51	2,44	2,35	2,30	2,19	2,08
29	2,93	2,87	2,73	2,57	2,48	2,41	2,33	2,27	2,16	2,05
30	2,91	2,84	2,70	2,55	2,45	2,39	2,30	2,24	2,13	2,02
35	2,80	2,74	2,60	2,44	2,35	2,28	2,19	2,14	2,02	1,90
40	2,73	2,66	2,52	2,37	2,27	2,20	2,11	2,06	1,94	1,82
50	2,62	2,56	2,42	2,27	2,17	2,10	2,01	1,95	1,82	1,70
60	2,56	2,50	2,35	2,20	2,10	2,03	1,94	1,88	1,75	1,62
80	2,48	2,42	2,27	2,12	2,01	1,94	1,85	1,79	1,65	1,51
100	2,43	2,37	2,22	2,07	1,97	1,89	1,80	1,74	1,60	1,45
200	2,34	2,27	2,13	1,97	1,87	1,79	1,69	1,63	1,48	1,30
500	2,28	2,22	2,07	1,92	1,81	1,74	1,63	1,57	1,41	1,20
1000	2,27	2,20	2,06	1,90	1,79	1,72	1,61	1,54	1,38	1,16

TA LA VI. DISTRI UCI N

$\alpha = 0\,025$

(Continua i n)

$_2$	$_1$ = gra os e liberta el numera or									
	1	2			5	6				10
1	648	800	864	900	922	937	948	957	963	969
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,4
3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5	14,4
4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84
5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62
6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52	5,46
7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82	4,76
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53
12	6,55	5,10	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25
14	6,30	4,86	4,24	3,89	3,66	3,50	3,38	3,28	3,21	3,15
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,80	2,73
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,70
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05

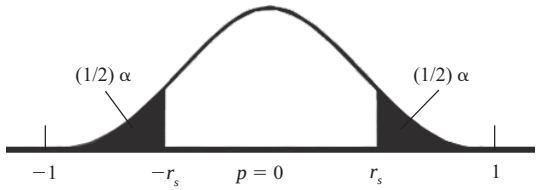
TA LA VI. DISTRI UCI N

$\alpha = 0\,025$

(Continua i n)

$_2$	$_1$ = gra os e liberta el numera or								
	12	15	20	2	0	0	60	120	∞
1	977	985	993	997	1.001	1.006	1.010	1.014	1.018
2	39,4	39,4	39,4	39,5	39,5	39,5	39,5	39,5	39,5
3	14,3	14,3	14,2	14,1	14,1	14,0	14,0	13,9	13,9
4	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26
5	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02
6	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85
7	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14
8	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67
9	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33
10	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08
11	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88
12	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72
13	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60
14	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49
15	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46	2,40
16	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32
17	2,82	2,72	2,62	2,56	2,50	2,44	2,38	2,32	2,25
18	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19
19	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13
20	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09
21	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04
22	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00
23	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97
24	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94
25	2,51	2,41	2,30	2,24	2,18	2,12	2,05	1,98	1,91
30	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79
40	2,29	2,18	2,07	2,01	1,94	1,88	1,80	1,72	1,64
60	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48
120	2,05	1,95	1,82	1,76	1,69	1,61	1,53	1,43	1,31
∞	1,94	1,93	1,71	1,64	1,57	1,48	1,39	1,27	1,00

TA LA VII. COE ICIENTES DE SPEARMAN PARA LA CORRELACI N DE RANGO



Dos e tremos e alores r ti os

Valores r ti os e os olas

n	α = 0 10	α = 0 05	α = 0 02	α = 0 01
5	0,900	-	-	-
6	0,829	0,886	0,943	-
7	0,714	0,786	0,893	-
8	0,643	0,738	0,833	0,881
9	0,600	0,683	0,783	0,833
10	0,564	0,648	0,745	0,794
11	0,523	0,623	0,736	0,818
12	0,497	0,591	0,703	0,780
13	0,475	0,566	0,673	0,745
14	0,457	0,545	0,646	0,716
15	0,441	0,525	0,623	0,689
16	0,425	0,507	0,601	0,666
17	0,412	0,490	0,582	0,645
18	0,399	0,476	0,564	0,625
19	0,388	0,462	0,549	0,608
20	0,377	0,450	0,534	0,591
21	0,368	0,438	0,521	0,576
22	0,359	0,428	0,508	0,562
23	0,351	0,418	0,496	0,549
24	0,343	0,409	0,485	0,537
25	0,336	0,400	0,475	0,526
26	0,329	0,392	0,465	0,515
27	0,323	0,385	0,456	0,505
28	0,317	0,377	0,448	0,496
29	0,311	0,370	0,440	0,487
30	0,305	0,364	0,432	0,478

TA LA VIII. PRUE A DE ILCO ON.

VALORES CR TICOS DE T EN LA PRUE A DE LOS RANGOS
DE PARES IGUALADOS DE ILCO ON

n	Prueba e una ola		
	0 025	0 01	0 005
	Prueba e os olas		
	0 05	0 02	0 01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

Tomada de la tabla I de Wilcoxon, F. 1949, Some rapid approximate statistical procedures. Nueva York: American Cyanamid.

TA LA I . DISTRI UCI N INOMIAL

n	*x*	,05	,10	,25	,20	,25	,30	,35	,40	,45	,50	,55	,60	,65	,70	,75	,80	,85	,90	,95
1	0	,9500	,9000	,8500	,8000	,7500	,7000	,6500	,6000	,5500	,5000	,4500	,4000	,3500	,3000	,2500	,2000	,1500	,1000	,0500
	1	,0500	,1000	,1500	,2000	,2500	,3000	,3500	,4000	,4500	,5000	,5500	,6000	,6500	,7000	,7500	,8000	,8500	,9000	,9500
2	0	,9025	,8100	,7225	,6400	,5625	,4900	,4225	,3600	,3025	,2500	,2025	,1600	,1225	,0900	,0625	,0400	,0225	,0100	,0025
	1	,0950	,1800	,2550	,3200	,3750	,4200	,4550	,4800	,4950	,5000	,4950	,4800	,4550	,4200	,3750	,3200	,2550	,1800	,0950
	2	.0025	,0100	,0225	,0400	,0625	,0900	,1225	,1600	,2025	,2500	,3025	,3600	,4225	,4900	,5625	,6400	,7225	,8100	,9025
3	0	,8574	,7290	,6141	,5120	,4219	,3430	,2746	,2160	,1664	,1250	,0911	,0640	,0429	,0270	,0156	,0900	,0034	,0010	,0001
	1	,1354	,2430	,3251	,3840	,4219	,4410	,4436	,4320	,4084	,3750	,3341	,2880	,2389	,1890	,1406	,0960	,0574	,0270	,0071
	2	,0071	,0270	,0574	,0960	,1406	,1890	,2389	,2880	,3341	,3750	,4084	,4320	,4436	,4410	,4219	,3840	,3251	,2430	,1354
	3	,0001	,0010	,0034	,0080	,0156	,0270	,0429	,0640	,0911	,1250	,1664	,2160	,2746	,3430	,4219	,5120	,6141	,7290	,8574
4	0	,8145	,6561	,5220	,4096	,3164	,2401	,1785	,1296	,0915	,0625	,0410	,0256	,0150	,0081	,0039	,0016	,0005	,0001	,0000
	1	,1715	,2916	,3685	,4096	,4219	,4116	,3845	,3456	,2995	,2500	,2005	,1536	,1115	,0756	,0469	,0256	,0115	,0036	,0005
	2	.0135	,0486	,0975	,1536	,2109	,2646	,3105	,3456	,3675	,3750	,3675	,3456	,3105	,2646	,2109	,1536	,0975	,0486	,0135
	3	,0005	,0036	,0115	,0256	,0469	,0756	,1115	,1536	,2005	,2500	,2995	,3456	,3845	,4116	,4219	,4096	,3685	,2916	,1715
	4	,0000	,0001	,0005	,0016	,0019	,0081	,0150	,0256	,0410	,0625	,0915	,1296	,1785	,2401	,3164	,4096	,5220	,6561	,8145
5	0	,7738	,5905	,4437	,3277	,2373	,1681	,1160	,0778	,0503	,0313	,0185	,0102	,0053	,0024	,0010	,0003	,0001	,0000	,0000
	1	,2036	,3281	,3915	,4096	,3955	,3602	,3124	,2592	,2059	,1563	,1128	,0768	,0488	,0284	,0146	,0064	,0022	,0004	,0000
	2	,0214	,0729	,1382	,2048	,2637	,3087	,3364	,3456	,3369	,3125	,2757	,2304	,1181	,1323	,0879	,0512	,0244	,0081	,0011
	3	,0011	,0081	,0244	,0512	,0879	,1323	,1181	,2304	,2757	,3125	,3369	,3456	,3364	,3087	,2637	,2048	,1382	,0729	,0214
	4	,0000	,0004	,0022	,0064	,0146	,0283	,0488	,0768	,1128	,1652	,2059	,2592	,3124	,3601	,3955	,4096	,3915	,3281	,2036
	5	,0000	,0000	,0001	,0003	,0010	,0024	,0053	,0102	,0185	,0312	,0503	,0778	,1160	,1681	,2373	,3277	,4437	,5905	,7738
6	0	,7351	,5314	,3771	,2621	,1780	,1176	,0754	,0467	,0277	,0156	,0083	,0041	,0018	,0007	,0002	,0001	,0000	,0000	,0000
	1	,2321	,3543	,3993	,3932	,3560	,3025	,2437	,1866	,1359	,0938	,0609	,0369	,0205	,0102	,0044	,0015	,0004	,0001	,0000
	2	,0305	,0984	,1762	,2458	,2966	,3241	,3280	,3110	,2780	,2344	,1861	,1382	,0951	,0595	,0330	,0154	,0055	,0012	,0001
	3	,0021	,0146	,0415	,0819	,1318	,1852	,2355	,2765	,3032	,3125	,3032	,2765	,2355	,1852	,1318	,0819	,0415	,0146	,0021
	4	,0001	,0012	,0055	,0154	,0330	,0595	,0951	,1382	,1861	,2344	,2780	,3110	,3280	,3241	,2966	,2458	,1762	,0984	,0305
	5	,0000	,0001	,0004	,0015	,0044	,0102	,0205	,0369	,0609	,0937	,1359	,1866	,2437	,3025	,3560	,3932	,3993	,3543	,2321
	6	,0000	,0000	,0000	,0001	,0002	,0007	,0018	,0041	,0083	,0156	,0277	,0467	,0754	,1176	,1780	,2621	,3771	,5314	,7351
7	0	,6983	,4783	,3206	,2097	,1355	,0824	,0490	,0280	,0152	,0078	,0037	,0016	,0006	,0002	,0001	,0000	,0000	,0000	,0000
	1	,2573	,3720	,3960	,3670	,3115	,2471	,1848	,1306	,0872	,0547	,0320	,0172	,0084	,0036	,0013	,0004	,0001	,0000	,0000
	2	,0406	,1240	,2097	,2753	,3115	,3177	,2985	,2613	,2140	,1641	,1172	,0774	,0466	,0250	,0115	,0043	,0012	,0002	,0000
	3	,0036	,0230	,0617	,1147	,1730	,2269	,2679	,2903	,2918	,2734	,2388	,1935	,1442	,0972	,0577	,0287	,0109	,0026	,0002
	4	,0002	,0026	,0109	,0287	,0577	,0972	,1442	,1935	,2388	,2734	,2918	,2903	,2679	,2269	,1730	,1147	,0617	,0230	,0036
	5	,0000	,0002	,0012	,0043	,0115	,0250	,0466	,0774	,1172	,1641	,2140	,2613	,2985	,3177	,3115	,2753	,2097	,1240	,0406
	6	,0000	,0000	,0001	,0004	,0013	,0036	,0084	,0172	,0320	,0547	,0872	,1306	,1848	,2471	,3115	,3670	,3960	,3720	,2573
	7	,0000	,0000	,0000	,0000	,0001	,0002	,0006	,0016	,0037	,0078	,0152	,0280	,0490	,0824	,1335	,2097	,3206	,4783	,6983

TA LA I . DISTRI UCI N INOMIAL (Continua i n)

8	0	,6634	,4305	,2725	,1678	,1001	,0576	,0319	,0168	,0084	,0039	,0017	,0007	,0002	,0001	,0000	,0000	,0000	,0000	,0000
	1	,2793	,3826	,3847	,3355	,2670	,1977	,1373	,0896	,0548	,0313	,0164	,0079	,0033	,0012	,0004	,0001	,0000	,0000	,0000
	2	,0515	,1488	,2376	,2936	,3115	,2965	,2587	,2090	,1569	,1094	,0703	,0413	,0217	,0100	,0038	,0011	,0002	,0000	,0000
	3	,0054	,0331	,0839	,1468	,2076	,2541	,2786	,2787	,2568	,2188	,1719	,1239	,0808	,0467	,0231	,0092	,0026	,0004	,0000
	4	,0004	,0046	,0185	,0459	,0865	,1361	,1875	,2322	,2627	,2734	,2627	,2322	,1875	,1361	,0865	,0459	,0185	,0046	,0004
	5	,0000	,0004	,0026	,0092	,0231	,0467	,0808	,1239	,1719	,2188	,2568	,2787	,2786	,2541	,2076	,1468	,0839	,0331	,0054
	6	,0000	,0000	,0002	,0011	,0038	,0100	,0217	,0413	,0703	,1094	,1569	,2090	,2587	,2965	,3115	,2936	,2376	,1488	,0515
	7	,0000	,0000	,0000	,0001	,0004	,0012	,0033	,0079	,0164	,0312	,0548	,0896	,1373	,1977	,2670	,3355	,3847	,3826	,2793
	8	,0000	,0000	,0000	,0000	,0000	,0001	,0002	,0007	,0017	,0039	,0084	,0168	,0319	,0576	,1001	,1678	,2725	,4305	,6636
9	0	,6302	,3874	,2316	,1342	,0751	,0404	,0207	,0101	,0046	,0020	,0008	,0003	,0001	,0000	,0000	,0000	,0000	,0000	,0000
	1	,2986	,3874	,3679	,3020	,2253	,1556	,1004	,0605	,0339	,0176	,0083	,0035	,0013	,0004	,0001	,0000	,0000	,0000	,0000
	2	,0629	,1722	,2597	,3020	,3003	,2668	,2162	,1612	,1110	,0703	,0407	,0212	,0098	,0039	,0012	,0003	,0000	,0000	,0000
	3	,0077	,0446	,1069	,1762	,2326	,2668	,2716	,2508	,2119	,1641	,1160	,0743	,0424	,0210	,0087	,0028	,0006	,0001	,0000
	4	,0006	,0074	,0283	,0661	,1168	,1715	,2194	,2508	,2600	,2461	,2128	,1672	,1181	,0735	,0389	,0165	,0050	,0008	,0000
	5	,0000	,0008	,0050	,0165	,0389	,0735	,1181	,1672	,2128	,2461	,2600	,2508	,2194	,1715	,1168	,0,661	,0283	,0074	,0006
	6	,0000	,0001	,0006	,0028	,0087	,0210	,0424	,0743	,1160	,1641	,2119	,2508	,2716	,2668	,2336	,1762	,1069	,0446	,0077
	7	,0000	,0000	,0000	,0003	,0012	,0039	,0098	,0212	,0407	,0703	,1110	,1612	,2162	,2668	,3003	,3020	,2597	,1722	,0629
	8	,0000	,0000	,0000	,0000	,0001	,0004	,0013	,0035	,0083	,0176	,0339	,0605	,1004	,1556	,2253	,3020	,3679	,3874	,2986
	9	,0000	,0000	,0000	,0000	,0000	,0000	,0001	,0003	,0008	,0020	,0046	,0101	,0207	,0404	,0751	,1342	,2316	,3874	,6302
10	0	,5987	,3847	,1969	,1074	,0563	,0282	,0135	,0060	,0025	,0010	,0003	,0001	,0000	,0000	,0000	,0000	,0000	,0000	,0000
	1	,3151	,3874	,3474	,2684	,1877	,1211	,0725	,0403	,0207	,0098	,0042	,0016	,0005	,0001	,0000	,0000	,0000	,0000	,0000
	2	,0746	,1937	,2759	,3020	,2816	,2335	,1757	,1209	,0763	,0439	,0229	,0106	,0043	,0014	,0004	,0001	,0000	,0000	,0000
	3	,0105	,0574	,1298	,2013	,2503	,2668	,2522	,2150	,1665	,1172	,0746	,0425	,0212	,0090	,0031	,0008	,0001	,0000	,0000
	4	,0010	,0112	,0401	,0881	,1460	,2001	,2377	,2508	,2384	,2051	,1596	,1115	,0689	,0368	,0162	,0055	,0012	,0001	,0000
	5	,0001	,0015	,0085	,0264	,0584	,1029	,1536	,2007	,2340	,2461	,2340	,2007	,1536	,1029	,0584	,0264	,0085	,0015	,0001
	6	,0000	,0001	,0012	,0055	,0162	,0368	,0689	,1115	,1596	,2051	,2384	,2508	,2377	,2001	,1460	,0881	,0401	,0112	,0010
	7	,0000	,0000	,0001	,0008	,0031	,0090	,0212	,0425	,0746	,1172	,1665	,2150	,2522	,2668	,2503	,2013	,1298	,0574	,0105
	8	,0000	,0000	,0000	,0001	,0004	,0014	,0043	,0106	,0229	,0439	,0763	,1209	,1757	,2335	,2816	,3020	,2759	,1937	,0746
	9	,0000	,0000	,0000	,0000	,0000	,0001	,0005	,0016	,0042	,0098	,0207	,0403	,0725	,1211	,1877	,2684	,3473	,3874	,3151
	10	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0001	,0003	,1098	,0025	,0060	,0135	,0282	,0563	,1074	,1969	,3487	,5987
11	0	,5688	,3138	,1673	,0859	,0422	,0198	,0088	,0036	,0014	,0005	,0002	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000
	1	,3293	,3835	,3248	,2362	,1549	,0932	,0518	,0266	,0125	,0054	,0021	,0007	,0002	,0000	,0000	,0000	,0000	,0000	,0000
	2	,0867	,2131	,2866	,2953	,2581	,1998	,1395	,0887	,0513	,0269	,0216	,0052	,0018	,0005	,0001	,0000	,0000	,0000	,0000
	3	,0137	,0710	,1517	,2215	,2581	,2568	,2254	,1774	,1259	,0806	,0462	,0234	,0102	,0037	,0011	,0002	,0000	,0000	,0000
	4	,0014	,0158	,0536	,1107	,1721	,2201	,2428	,2365	,2060	,1611	,1128	,0701	,0379	,0173	,0064	,0017	,0003	,0000	,0000
	5	,0001	,0025	,0132	,0388	,0803	,1321	,1830	,2207	,2360	,2256	,1931	,1471	,0985	,0566	,0268	,0097	,0023	,0003	,0000
	6	,0000	,0003	,0023	,0097	,0268	,0566	,0985	,1471	,1931	,2256	,2360	,2207	,1830	,1321	,0803	,0388	,0132	,0025	,0001
	7	,0000	,0000	,0003	,0017	,0064	,0173	,0379	,0701	,1128	,1611	,2060	,2365	,2428	,2201	,1721	,1107	,0536	,0158	,0014
	8	,0000	,0000	,0000	,0002	,0011	,0037	,0102	,0234	,0462	,0806	,1259	,1774	,2254	,2568	,2581	,2215	,1517	,0710	,0137
	9	,0000	,0000	,0000	,0000	,0001	,0005	,0018	,0052	,0126	,0269	,0513	,0887	,1395	,1998	,2581	,2953	,2866	,2131	,0867
	10	,0000	,0000	,0000	,0000	,0000	,0000	,0002	,0007	,0021	,0054	,0125	,0266	,0518	,0932	,1549	,2362	,3248	,3835	,3293
	11	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0002	,0005	,0014	,0036	,0088	,0198	,0422	,0859	,1673	,3138	.5688

TA LA I . DISTRI UCI N INOMIAL (Continua i n)

12	0	,5404	,2824	,1422	,0687	,0317	,0138	,0057	,0022	,0008	,0002	,0001	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000
	1	,3413	,3766	,3012	,2062	,1267	,0712	,0368	,0174	,0075	,0029	,0010	,0003	,0001	,0000	,0000	,0000	,0000	,0000	,0000	,0000
	2	,0988	,2301	,2924	,2835	,2323	,1678	,1088	,0639	,0339	,0161	,0068	,0025	,0008	,0002	,0000	,0000	,0000	,0000	,0000	,0000
	3	,0173	,0852	,1720	,2362	,2581	,2397	,1954	,1419	,0923	,0537	,0277	,0125	,0048	,0015	,0004	,0001	,0000	,0000	,0000	,0000
	4	,0021	,0213	,0683	,1329	,1936	,2311	,2367	,2128	,1700	,1208	,0762	,0420	,0199	,0078	,0024	,0005	,0001	,0000	,0000	,0000
	5	,0002	,0038	,0193	,0532	,1032	,1585	,2039	,2270	,2225	,1934	,1489	,1009	,0591	,0291	,0115	,0033	,0006	,0000	,0000	,0000
	6	,0000	,0005	,0040	,0155	,0401	,0792	,1281	,1766	,2124	,2256	,2124	,1766	,1281	,0792	,0401	,0155	,0040	,0005	,0000	,0000
	7	,0000	,0000	,0006	,0033	,0115	,0291	,0591	,1009	,1489	,1934	,2225	,2270	,2039	,1585	,1032	,0532	,0193	,0038	,0002	,0000
	8	,0000	,0000	,0001	,0005	,0024	,0078	,0199	,0420	,0762	,1208	,1700	,2128	,2367	,2311	,1936	,1329	,0683	,0213	,0021	,0000
	9	,0000	,0000	,0000	,0001	,0004	,0015	,0048	,0125	,0277	,0537	,0923	,1419	,1954	,2397	,2581	,2362	,1720	,0852	,0173	,0000
	10	,0000	,0000	,0000	,0000	,0000	,0002	,0008	,0025	,0068	,0161	,0339	,0639	,1088	,1678	,2323	,2635	,2924	,2301	,0988	,0000
	11	,0000	,0000	,0000	,0000	,0000	,0000	,0001	,0003	,0010	,0029	,0075	,0174	,0368	,0712	,1267	,2062	,3012	,3766	,3413	,0000
	12	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0001	,0002	,0008	,0022	,0057	,0138	,0317	,0687	,1422	,2824	,5404	,0000

FÓRMULAS

TABLA Fórmulas aplicadas en las medidas de posición, dispersión, asimetría y apuntamiento

MEDIDAS DE POSICIÓN O TENDENCIA CENTRAL

1. Media aritmética

a. $\bar{x} = \dfrac{\sum x_i}{n}$ b. $\bar{y} = O_t + \dfrac{\sum z_i' n_i}{n}$

c. $\bar{y} = O_t + C\dfrac{\sum z_i'' n_i}{n}$

d. $\bar{y} = \dfrac{\sum y_i n_i}{n}$ e. $\bar{y} = \sum y_i h_i$

2. Mediana

A. Variable discreta

a. $N_{j-1} < \dfrac{n}{2} \rightarrow M_e = \dfrac{y_{j-1} + y_j}{2}$

b. $N_{j-1} < \dfrac{n}{2} \rightarrow M_e = y_j$

B. Variable continua

a. $N_{j21} = \dfrac{n}{2} \rightarrow M_e = y'_{j21}$

b. $N_{j21} , \dfrac{n}{2} \rightarrow M_e = y'_{j21} + C\left[\dfrac{n/2 - N_{j-1}}{n_j}\right]$

C. Datos no agrupados

$\dfrac{n+1}{2}$ Posición de la observación central

3. Modo

a. $M_d = Y_j$ b. $M_d = Y'_{j21} + C\left[\dfrac{n_{j+1}}{n_{j+1} + n_{j-1}}\right]$

c. $M_d = Y'_{j21} + C\left[\dfrac{n_j - n_{j-1}}{(n_j - n_{j-1}) + (n_j - n_{j+1})}\right]$

d. $M_d = Y'_{j21} + C\left[\dfrac{n_{j+1} - n_{j-1}}{2(2n_j - n_{j-1} - n_{j+1})}\right]$

e. $M_d = Y'_{j21} + C\left[\dfrac{n_{j-1} - n_{j+1}}{n_j + n_{j+1} + n_{j-1}}\right]$

4. Media cuadrática $M_2 = \sqrt{\dfrac{\sum X_i^2}{n}}$ $M_2 = \sqrt{\dfrac{\sum Y_i^2 n_i}{n}}$

5. Media cúbica $M_3 = \sqrt[3]{\dfrac{\sum X_i^3}{n}}$ $M_3 = \sqrt[3]{\dfrac{\sum Y_i^3 n_i}{n}}$

6. Media armónica $M_{-1} = \dfrac{n}{\sum 1/x_i}$ $M_{-1} = \dfrac{n}{\sum n_i / Y_i}$

7. Media geométrica

a. $M_0 = \sqrt[n]{\pi X_i}$

b. $M_0 = \sqrt[n]{\pi Y_i^{ni}}$

c. $\log M_0 = \dfrac{\sum \log X_i}{n}$

d. $\log M_0 = \dfrac{\sum n_i \log Y_i}{n}$

8. Cuartil, decil, percentil

a. $Q_3 = Y'_{j21} + C\left[\dfrac{\frac{3n}{4} - N_{j-1}}{n_j}\right]$ Tercer cuartil

b. $D_5 = Y'_{j21} + C\left[\dfrac{\frac{5n}{10} - N_{j-1}}{n_j}\right]$ Quinto decil

c. $C_{60} = Y'_{j21} + C\left[\dfrac{\frac{60n}{100} - N_{j-1}}{n_j}\right]$ Sesenta percentil

A. Desviaciones

a. $z_i = x_i - \bar{x}$; $z_i = y_i - \bar{y}$; b. $z'_i = y_i - O_t$; c. $z''_i = \dfrac{y_i - O_t}{C} = \dfrac{z'_i}{C}$

B. Propiedades de la media

a. $M_{[K]} = K$

b. $M_{[K+x]} = K + M_{[X]} = K + \bar{x}$

c. $M_{[Kx]} = KM_{[X]} = K\bar{x}$

d. $\bar{X} = \dfrac{\bar{X}_1 n_1 + \bar{X}_2 n_2}{n_1 + n_2}$

e. $M_{[x-K]} = M_{[x]} - k = \bar{x} - K$

f. $\sum z_i = O$ $\sum z_i n_i = O$

g. $\sum z_i^2 n_i < \sum z_i'^2 n_i$

Otras propiedades

a. $\bar{x} = M_o = M_d$ Simétrica

b. $\bar{x} > M_o > M_d$ Asimétrica 1

c. $\bar{x} < M_o < M_d$ Asimétrica 2

C. Propiedad de la Me

$$\sum |Y_i - M_e| n_i \leq \sum |Y_i - O_t$$

D. Propiedad

$$M_{-1} < M_0 < M_1 < M_2 < M_3$$

TA LA rmulas apli a as en las me i as e posi i n ispersi n asimetr a y apuntamiento. Continua i n.

MEDIDAS DE DISPERSIÓN

1. Varianza

a. $S^2 = \dfrac{\sum z_i^2}{n} = \dfrac{\sum \left(x - \bar{x}\right)^2}{n}$

b. $S^2 = \dfrac{\sum x_i^2 - n\bar{x}^2}{n}$ $\qquad S^2 = \dfrac{\sum x_i^2}{n} - \bar{x}^2$

c. $S^2 = \dfrac{\sum z_i^2 n_i}{n}$ $\qquad S^2 = \dfrac{\sum \left(Y_i - \bar{Y}\right)^2 n_i}{n}$

d. $S^2 = \dfrac{\sum y_i^2 n_i}{n} - \bar{y}^2$ $\qquad S^2 = \dfrac{\sum Y_i^2 n_i - n\bar{Y}^2}{n}$

e. $S^2 = \dfrac{\sum z_i'^2 n_i}{n} - \left(\dfrac{\sum z_i' n_i}{n}\right)^2$

f. $S^2 = C^2 \left\{ \dfrac{\sum z_i''^2 n_i}{n} - \left(\dfrac{\sum z_i'' n_i}{n}\right)^2 \right\}$

2. Desviación típica o estándar

$S = +\sqrt{S^2}$

3. Desviación media

a. $D_a = \dfrac{\sum |Z_i|}{n} = \dfrac{\sum |X_i - \bar{X}|}{n}$

b. $D_a = \dfrac{\sum |Z_i| n_i}{n} = \dfrac{\sum |y_i - \bar{y}| n_i}{n}$

4. Desviación mediana

a. $D_e = \dfrac{\sum |X_i - M_e|}{n}$ $\qquad$ b. $D_e = \dfrac{\sum |Y_i - M_e| n_i}{n}$

5. Coeficiente de variación

$CV = d = \dfrac{S}{\bar{X}} = \dfrac{S}{\bar{Y}}$

6. Puntaje típico $\quad t = \dfrac{X - \bar{X}}{S}$

7. Límites $\quad \hat{Y} \pm Z \dfrac{S}{\sqrt{n}} \quad \dfrac{\hat{Y}_s}{Y_i}$

8. Recorrido $\quad$ Rango 5 $X_{Máx}$ 2 $X_{Mín}$

9. Recorrido intercuartílico:

Q_D 5 Q_3 2 Q_1

10. Desviación cuartil:

$Q_D = \dfrac{Q_3 - Q_1}{2}$

11. Coeficiente de desviación cuartil:

$CQ = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

12. Coeficiente de desviación media: $C_D = \dfrac{D_Q}{\bar{X}}$

A. Propiedades de la varianza

a. S^2 . 0 $\qquad$ b. $V_{[K]}$ 5 0

c. $V_{[K1x]}$ 5 $V_{[x]}$ 5 S_X^2 $\qquad$ d. $V_{[Kx]}$ 5 $K^2 V_{[x]}$ 5 $K^2 S_X^2$

e. $V_{[x2K]}$ 5 $V_{[x]}$ 5 S_X^2

f. $S^2 = \dfrac{S_1^2 n_1 + S_2^2 n_2}{n_1 + n_2} + \dfrac{\left(\bar{X}_1 - \bar{X}\right)^2 n_1 + \left(\bar{X}_2 - \bar{X}\right)^2 n_2}{n_1 + n_2}$

B. Propiedades

a. S . D_a $ D_e

b. D_a 5 0,7979S (en la distribución simétrica)

C. Momentos unidimensionales

Momentos bidimensionales

a. Momentos respecto variable

$a_t = \dfrac{\sum Y_i^r n_i}{n}, \qquad a_{r_1 r_2} = \dfrac{\sum\sum Y_{1i_1}^{r_1} Y_{2i_2}^{r_1} n_{i_1 i_2}}{n}$

b. Momentos respecto a $\bar{X}$

$m_r = \dfrac{\sum Z_i^r n_i}{n}; \qquad m_{r_1 r_2} = \dfrac{\sum\sum Z_{1i_1}^{r_1} Z_{2i_2}^{r_2} n_{i_1 i_2}}{n}$

c. Momentos respecto O_t

$m_r' = \dfrac{\sum Z_i'^r n_i}{n}; \qquad m_{r_1 r_2}' = \dfrac{\sum\sum Z_{1i_1}'^{r_1} Z_{2i_2}'^{r_2} n_{i_1 i_2}}{n}$

d. Momentos respecto a O_t en unidades de C

$m_r' = \dfrac{\sum Z_i'^r n_i}{n}, \qquad m_{r_1 r_2}' = \dfrac{\sum\sum Z_{1i_1}'^{r_1} Z_{2i_2}'^{r_2} n_{i_1 i_2}}{n}$

MEDIDAS DE ASIMETRÍA

a. $A = \dfrac{\bar{X} - M_d}{S}$ $\qquad$ b. $A = \dfrac{3\left(\bar{X} - M_e\right)}{S}$

c. $B = \dfrac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1}$ $\qquad$ d. $g_1 = \dfrac{m_3}{S^3}$

A . 0 Asimétrica 1 $\qquad$ A 5 0 Simétrica 1

A , 0 Asimétrica 2

MEDIDAS DE APUNTAMIENTO

a. $g_2 = \dfrac{m_4}{m_2^2}$ $\qquad$ b. $g_2 = \dfrac{m_4}{S^4}$

g_2 . 3 Apuntada o leptocúrtica $\qquad$ g_2 , 3 Achatada o platicúrtica

g_2 , 3 Achatada o platicúrtica $\qquad$ g_2 5 3 Normal o mesocúrtica

TABLA I Fórmulas aplicadas en la Regresión y Correlación

REGRESIÓN Y CORRELACIÓN LINEAL SIMPLE Y PONDERADA

1. Ecuación general (regresión simple)

   a. $\hat{Y} = b_{yx}X + C_{yx}$ $\qquad \hat{Y} = b_{yx}\left(X - \overline{X}\right) + \overline{Y}$

   b. $\hat{X} = b_{xy}\left(Y - \overline{Y}\right) + \overline{X}$ $\qquad \hat{X} = b_{xy}Y + C_{xy}$

   $\hat{Z}_x = b_{xy}Z_y$ $\qquad \hat{Z}_y = b_{yx}Z_x$

2. Coeficiente angular

   a. $b_{yx} = \dfrac{m_{xy}}{S_x^2}$ $\qquad b_{xy} = \dfrac{m_{xy}}{S_y^2}$

   b. $\hat{Z}_x = b_{xy}\left(Y - \overline{Y}\right)$ $\qquad \hat{Z}_y = b_{yx}\left(X - \overline{X}\right)$

   c. $b_{xy} = \dfrac{\sum Z_x Z_y}{\sum Z_y^2}$ $\qquad b_{yx} = \dfrac{\sum Z_x Z_y}{\sum Z_x^2}$

   $b_{yx} = \dfrac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum x^2 - \left(\sum x\right)^2}$; $\qquad b_{xy} = \dfrac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum y^2 - \left(\sum y\right)^2}$

   $b_{yx} = \dfrac{\sum xy - n\overline{x}\overline{y}}{\sum x^2 - n\overline{x}^2}$; $\qquad b_{xy} = \dfrac{\sum xy - n\overline{x}\overline{y}}{\sum y^2 - n\overline{y}^2}$

3. Coeficiente de posición

   $C_{yx} = \overline{Y} - b_{yx}\overline{X}$ $\qquad C_{xy} = \overline{X} - b_{xy}\overline{Y}$

   $C_{yx} = \dfrac{\sum y - b_{yx}\sum x}{n}$ $\qquad C_{xy} = \dfrac{\sum x - b_{xy}\sum y}{n}$

   Covarianza

   a. $m_{xy} = \dfrac{\sum XY}{n} - \overline{X}\overline{Y}$ $\qquad m_{xy} = \dfrac{\sum Z_x Z_y}{n}$

4. Sistema de ecuaciones

   a. $\sum Y = b_{yx}\sum X + nC_{yx}$; $\qquad$ b. $\sum X = b_{xy}\sum Y + nC_{xy}$

   $\sum XY = b_{yx}\sum X^2 + C_{yx}\sum X$; $\qquad \sum XY = b_{xy}\sum Y^2 + C_{xy}\sum Y$

5. Varianza residual

   a. $S_{yx}^2 = \dfrac{\sum\left(Y - \hat{Y}\right)^2}{n}$ $\qquad S_{xy}^2 = \dfrac{\sum\left(X - \hat{X}\right)^2}{n}$

   b. $S_{yx}^2 = S_y^2 - \dfrac{m_{xy}^2}{S_x^2}$ $\qquad S_{xy}^2 = S_x^2 - \dfrac{m_{xy}^2}{S_y^2}$

   c. $S_{yx}^2 = \dfrac{\sum\left(Z_y - \hat{Z}_y\right)^2}{n}$ $\qquad S_{xy}^2 = \dfrac{\sum\left(Z_x - \hat{Z}_x\right)^2}{n}$

   d. $S_{yx}^2 = \dfrac{\sum Y_i^2 - C_{yx}\sum Y_i - b_{yx}\sum X_i Y_i}{n}$

   $S_{xy}^2 = \dfrac{\sum X_i^2 - C_{xy}\sum X_i - b_{xy}\sum X_i Y_i}{n}$

   $S_{yx}^2 = S_y^2\left(1 - R^2\right)$ $\qquad S_{xy}^2 = S_x^2\left(1 - R^2\right)$

6. Error estándar

   a. $S_{yx} = +\sqrt{S_{yx}^2}$ $\qquad S_{xy} = +\sqrt{S_{xy}^2}$

7. Varianza explicada

   a. $S_{ay}^2 = \dfrac{\sum\left(\hat{Y} - \overline{Y}\right)^2}{n}$ $\qquad S_{ay}^2 = S_y^2 - S_{yx}^2$

   b. $S_{ax}^2 = \dfrac{\sum\left(\hat{X} - \overline{X}\right)^2}{n}$ $\qquad S_{ax}^2 = S_x^2 - S_{xy}^2$

8. Covarianza (ponderada)

   a. $m_{xy} = \dfrac{\sum Z_x Z_y n_i}{n}$ $\qquad$ b. $m_{xy} = \dfrac{\sum X_i Y_i n_i}{n} - \overline{X}\overline{Y}$

   c. $m_{xy} = \dfrac{\sum\left[\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)\right]n_i}{n}$

9. Varianza total

   $S_y^2 = S_{yx}^2 + S_{ay}^2$; $\qquad S_x^2 = S_{xy}^2 + S_{ax}^2$; $\qquad VT = VR + VE$

10. Coeficiente de correlación

    a. $R^2 = \dfrac{S_{ay}^2}{S_y^2}$; $\qquad R^2 = \dfrac{S_{ax}^2}{S_x^2}$;

    $r = \dfrac{m_{xy}}{S_x S_y} = \sqrt{b_{yx}b_{xy}}$

    b. $R^2 = 1 - \dfrac{S_{yx}^2}{S_y^2} = 1 - \dfrac{S_{xy}^2}{S_x^2}$; $\qquad R^2 = 1 - \dfrac{VR}{VT}$

    c. $R^2 = \dfrac{m_{xy}^2}{S_x^2 S_y^2}$; $\qquad R^2 = b_{yx}b_{xy}$; $\qquad R^2 = \dfrac{VE}{VT}$

    $0 \le R^2 \le 1$ $\qquad -1 \le r \le 1$

11. Regresión lineal ponderada

    $\hat{Y} = b_{yx}X + C_{yx}$; $\qquad \hat{X} = b_{xy}Y + C_{xy}$

12. Sistema de ecuaciones (ponderada)

    a. $\sum Y_i n_i = b_{yx}\sum X_i n_i + nC_{yx}$

    $\sum Y_i X_i n_i = b_{yx}\sum X_i^2 n_i + C_{yx}\sum X_i n_i$

    b. $\sum X_i n_i = b_{xy}\sum Y_i n_i + nC_{xy}$

    $\sum X_i Y_i n_i = b_{xy}\sum Y_i^2 n_i + C_{xy}\sum Y_i n_i$

13. Varianza residual (ponderada)

    a. $S_{yx}^2 = \dfrac{\sum Y_i^2 n_i - C_{yx}\sum Y_i n_i - b_{yx}\sum Y_i X_i n_i}{n}$

    $S_{xy}^2 = \dfrac{\sum X_i^2 n_i - C_{xy}\sum X_i n_i - b_{xy}\sum X_i Y_i n_i}{n}$

    b. $S_{yx}^2 = \dfrac{\sum\left(Y_i - \hat{Y}_i\right)^2 n_i}{n}$ $\qquad S_{xy}^2 = \dfrac{\sum\left(X_i - \hat{X}_i\right)^2 n_i}{n}$

TABLA III. Formulas aplicadas en la regresión y correlación. Continuación

REGRESIÓN PARABÓLICA SIMPLE Y PONDERADA

1. Ecuación general (regresión simple)

   a. $\hat{Y} = a_{yx}X^2 + b_{yx}X + C_{yx}$

   b. $\hat{X} = a_{xy}Y^2 + b_{xy}Y + C_{xy}$

2. Sistema de ecuaciones

   a. $\sum Y_i = a_{yx}\sum X_i^2 + b_{yx}\sum X_i + nC_{yx}$

   $\sum X_iY_i = a_{yx}\sum X_i^3 + b_{yx}\sum X_i^2 + C_{yx}\sum X_i$

   $\sum X_i^2Y_i = a_{yx}\sum X_i^4 + b_{yx}\sum X_i^3 + C_{yx}\sum X_i^2$

   b. $\sum X_i = a_{xy}\sum Y_i^2 + b_{xy}\sum Y_i + nC_{xy}$

   $\sum X_iY_i = a_{xy}\sum Y_i^3 + b_{xy}\sum Y_i^2 + C_{xy}\sum Y_i$

   $\sum X_iY_i^2 = a_{xy}\sum Y_i^4 + b_{xy}\sum Y_i^3 + C_{xy}\sum Y_i^2$

3. Sistema de ecuaciones (ponderada)

   a. $\sum Y_in_i = a_{yx}\sum X_i^2n_i + b_{yx}\sum X_in_i + nC_{yx}$

   $\sum Y_iX_in_i = a_{yx}\sum X_i^3n_i + b_{yx}\sum X_i^2n_i + C_{yx}\sum X_in_i$

   $\sum Y_iX_i^2n_i = a_{yx}\sum X_i^4n_i + b_{yx}\sum X_i^3n_i + C_{yx}\sum X_i^2n_i$

   b. $\sum X_in_i = a_{xy}\sum Y_i^2n_i + b_{xy}\sum Y_in_i + nC_{xy}$

   $\sum X_iY_in_i = a_{xy}\sum Y_i^3n_i + b_{xy}\sum Y_i^2n_i + C_{xy}\sum Y_in_i$

   $\sum X_iY_i^2n_i = a_{xy}\sum Y_i^4n_i + b_{xy}\sum Y_i^3n_i + C_{xy}\sum Y_i^2n_i$

4. Varianza residual (simple y ponderada)

   a. $S_{yx}^2 = \dfrac{\sum Y_i^2 - C_{yx}\sum Y_i - b_{yx}\sum X_iY_i - a_{yx}\sum X_i^2Y_i}{n}$

   $S_{xy}^2 = \dfrac{\sum X_i^2 - C_{xy}\sum X_i - b_{xy}\sum X_iY_i - a_{xy}\sum Y_i^2X_i}{n}$

   b. $S_{xy}^2 = \dfrac{\sum X_i^2n_i - C_{xy}\sum X_in_i - b_{xy}\sum Y_iX_in_i - a_{xy}\sum Y_i^2X_in_i}{n}$

   $S_{yx}^2 = \dfrac{\sum Y_i^2n_i - C_{yx}\sum Y_in_i - b_{yx}\sum X_iY_in_i - a_{yx}\sum X_i^2Y_in_i}{n}$

5. Coeficiente de correlación

   $r = \sqrt{1 - \dfrac{S_{yx}^2}{S_y^2}}$; $\quad r = \sqrt{1 - \dfrac{S_{xy}^2}{S_x^2}}$; $\quad r = \sqrt{1 - \dfrac{VR}{VT}}$

6. Límites de confianza

   $\hat{Y} \pm Z\dfrac{S_{yx}}{\sqrt{n}} \quad \begin{matrix}\hat{Y}_s \\ \hat{Y}_i\end{matrix} \qquad \hat{X} \pm Z\dfrac{S_{xy}}{\sqrt{n}} \quad \begin{matrix}\hat{X}_s \\ \hat{X}_i\end{matrix}$

   $\hat{Y} \pm Z \quad S_{yx}\sqrt{1 + \dfrac{1}{n} + \dfrac{\left(x - \bar{x}\right)^2}{\sum X_i^2 - \dfrac{\left(\sum X_i\right)^2}{n}}}$

REGRESIÓN Y CORRELACIÓN MÚLTIPLE

1. Ecuación general

   a. $\hat{X}_1 = b_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$

   b. $\hat{X}_2 = b_{2.13} + b_{21.3}X_1 + b_{23.1}X_3$

   c. $\hat{X}_3 = b_{3.12} + b_{31.2}X_1 + b_{32.1}X_2$

   a. $\hat{X}_1 = b_{12.3}\left(X_2 - \bar{X}_2\right) + b_{13.2}\left(X_3 - \bar{X}_3\right) + \bar{X}_1$

   b. $\hat{X}_2 = b_{21.3}\left(X_1 - \bar{X}_1\right) + b_{23.1}\left(X_3 - \bar{X}_3\right) + \bar{X}_2$

   c. $\hat{X}_3 = b_{31.2}\left(X_1 - \bar{X}_1\right) + b_{32.1}\left(X_2 - \bar{X}_2\right) + \bar{X}_3$

2. Sistema de ecuaciones normales

   a. $\sum X_1 = nb_{1.23} + b_{12.3}\sum X_2 + b_{13.2}\sum X_3$

   $\sum X_1X_2 = b_{1.23}\sum X_2 + b_{12.3}\sum X_2^2 + b_{13.2}\sum X_3X_2$

   $\sum X_1X_3 = b_{1.23}\sum X_3 + b_{12.3}\sum X_2X_3 + b_{13.2}\sum X$

   a. $\sum Z_1Z_2 = b_{12.3}\sum Z_2^2 + b_{13.2}\sum Z_3Z_2$

   $\sum Z_1Z_3 = b_{12.3}Z_2Z_3 + b_{13.2}\sum Z_3^2$

   b. $\sum X_2 = nb_{2.13} + b_{21.3}\sum X_1 + b_{23.1}\sum X_3$

   $\sum X_1X_2 = b_{2.13}\sum X_1 + b_{21.3}\sum X_1^2 + b_{23.1}\sum X_3X_1$

   $\sum X_2X_3 = b_{2.13}\sum X_3 + b_{21.3}\sum X_1X_3 + b_{23.1}\sum X_3^2$

   b. $\sum Z_2Z_1 = b_{21.3}\sum Z_1^2 + b_{23.1}\sum Z_3Z_1$

   $\sum Z_2Z_3 = b_{21.3}\sum Z_1Z_3 + b_{23.1}\sum Z_3^2$

   c. $\sum X_3 = nb_{3.12} + b_{31.2}\sum X_1 + b_{32.1}\sum X_2$

   $\sum X_3X_1 = b_{3.12}\sum X_1 + b_{31.2}\sum X_1^2 + b_{32.1}\sum X_2X_1$

   $\sum X_3X_2 = b_{3.12}\sum X_2 + \sum b_{31.2}\sum X_1X_2 + b_{32.1}\sum X_2^2$

   c. $\sum Z_3Z_1 = b_{31.2}\sum Z_1^2 + b_{32.1}\sum Z_2Z_1$

   $\sum Z_3Z_2 = b_{31.2}\sum Z_1Z_2 + b_{32.1}\sum Z_2^2$

3. Varianza residual

   $S_{1.23}^2 = \dfrac{\sum\left(X_1 - \hat{X}_1'\right)^2}{n} \qquad S_{2.13}^2 = \dfrac{\sum\left(X_2 - \hat{X}_2'\right)^2}{n}$

   $S_{3.12}^2 = \dfrac{\sum\left(X_3 - \hat{X}_3'\right)^2}{n}$

4. Desviaciones

   $Z_1 = X_1 - \bar{X}_1$; $\quad Z_2 = X_2 - \bar{X}_2$; $\quad Z_3 = X_3 - \bar{X}_3$

5. Coeficiente de correlación

   $R_{1.23} = \sqrt{1 - \dfrac{S_{1.23}^2}{S_1^2}} \qquad R_{2.13} = \sqrt{1 - \dfrac{S_{2.13}^2}{S_2^2}}$

   $R_{3.12} = \sqrt{1 - \dfrac{S_{3.12}^2}{S_3^2}}$

TABLA II Fórmulas aplicadas en las Series de tiempo

SERIES CRONOLÓGICAS - TENDENCIA

I. Ajuste rectilíneo

1. Ecuación general

$\hat{Y} = bx + c$

2. Sistema de ecuaciones

cuando $\sum x \neq 0$ $\begin{cases} \sum Y = b\sum x + nc \\ \sum XY = b\sum x^2 + c\sum x \end{cases}$

3. Cálculo de b y c

cuando $\sum x = 0$ $\begin{cases} b = \dfrac{\sum XY}{\sum X^2}; & c = \dfrac{\sum Y}{n} \end{cases}$

4. Estimativos mensuales

A. Valores anuales correspondientes a sumas mensuales

a. $\hat{Y} = \dfrac{b}{144}X + \dfrac{c}{12}$ (impar)

b. $\hat{Y} = \dfrac{b}{72}X + \dfrac{c}{12}$ (par)

B. Valores anuales correspondientes a promedios mensuales

a. $\hat{Y} = \dfrac{b}{12}X + c$ (impar)

b. $\hat{Y} = \dfrac{b}{6}X + c$ (par)

5. Otras ecuaciones

$\hat{Y} = \bar{Y} + \left[\dfrac{\sum XY}{\sum X^2}\right]\bar{X}$

6. Coeficiente de correlación

a.

$$r = \frac{n\sum XY - \sum X\sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

b. $r = \sqrt{1 - \dfrac{S_{yx}^2}{S_y^2}}$

$S_{yx}^2 = \dfrac{\sum\left(Y - \hat{Y}\right)^2}{n}$ varianza residual

$S_y^2 = \dfrac{\sum Y^2}{n} - \bar{Y}^2$ varianza total

(ver S_{yx} tabla XVII)

II. Ajuste parabólico

1. Ecuación general: $\hat{Y} = ax^2 + bx + c$

2. Sistema de ecuaciones

$\sum Y = a\sum X^2 + b\sum X + nc$

$\sum XY = a\sum X^3 + b\sum X^2 + c\sum X$

$\sum X^2Y = a\sum X^4 + b\sum X^3 + c\sum X^2$

$\left(\text{cuando } \sum X \neq 0\right)$

3. Cálculo de a, b y c $\left(\text{cuando } \sum X = 0\right)$

$b = \dfrac{\sum XY}{\sum X^2}$ $\qquad \sum Y = a\sum X^2 + nc$

$\sum X^2Y = a\sum X^4 + c\sum X^2$

4. Otras fórmulas

$c = \dfrac{\sum Y - a\sum X^2}{n}$

$a = \dfrac{n\sum X^2Y - \sum X^2\sum Y}{n\sum X^4 - \left(\sum X^2\right)^2}$

$b = \dfrac{\sum XY}{\sum X^2}\left(\text{cuando } \sum X = 0\right)$

5. Coeficiente de correlación

$r = \sqrt{1 - \dfrac{S_{yx}^2}{S_y^2}}$

La S_{yx}^2 para la parábola aplicar fórmula de la tabla XVII

III. Ajuste exponencial

1. Ecuación general

a. $\hat{Y} = cb^x$ b. $\log Y = \log c + X\log b$

2. Sistema de ecuaciones

$\left(\text{cuando } \sum X \neq 0\right)$

$\sum \log Y = n\log C + \left(\sum X\right)\log b$

$\sum X\log Y = \left(\sum X\right)\log C + \left(\sum X^2\right)\log b$

3. Cálculo de c y b $\left(\text{cuando } \sum X = 0\right)$

$\log b = \dfrac{\sum X\log Y}{\sum X^2}$

$\sum \log c = \dfrac{\sum \log Y}{n}$

4. Coeficiente de correlación exponencial

a. $r = \sqrt{1 - \dfrac{S_{x\log y}^2}{S_{\log y}^2}}$

Siendo

$S_{\log yx}^2 = \dfrac{\sum\left(\log Y - \log \hat{Y}\right)^2}{n}$

b. $r = \sqrt{1 - \dfrac{S_{yx}^2}{S_y^2}}$

Para S_{yx}^2 y S_y^2 ver tabla XVII

IV. Exponencial modificada

$\hat{Y} = K + cb^{\hat{x}}$

V. Curva de Gompertz

$\hat{Y} = K \cdot C^{b^x}$

$\log Y = \log K \cdot b^x \log a$

VI. Logística

$\hat{Y} = \dfrac{1}{K + cb^x}$

VII. Estimación de P_f

a. $P_f = P_i(1 + r)^x$

b. $\log(1 + r) = \dfrac{\log P_f - \log P_i}{x}$

VIII. Límites de confianza

$\hat{Y} = y \pm z\dfrac{S_{yx}}{\sqrt{n}}$ $\begin{matrix} \nearrow \hat{Y}_s \\ \searrow \hat{Y}_i \end{matrix}$

Elaborado por:
Ciro Martínez Bencardino
Ecoe ediciones
Apartado aéreo 30969
Tels.: 2889871/21
Bogotá, D.C

TA LA III rmulas apli a as en los N meros n i es.

NÚMEROS ÍNDICES

1. Índices simples

a. $I_0^t = \frac{X_t}{X_0} \times 100$

b. $I_0^t = \frac{P_t}{P_0} \times 100$

X_t 5 investigado

X_0 5 base

2. Índice agregativo simple

a. $I_0^t = \frac{\sum X_t}{\sum X_0} \times 100$

b. $I_0^t = \frac{\sum P_t}{\sum P_0} \times 100$

c. $I_0^t = \frac{\sum q_t}{\sum q_0} \times 100$

d. $I_0^t = \frac{\sum I_0^t}{n}$

3. Índice ponderado de precios

a. Laspeyres

$L^{t_0} = \frac{\sum P_t q_0}{\sum P_0 q_0} \times 100$

b. Paasche

$P^{t_0} = \frac{\sum P_t q_t}{\sum P_0 q_t} \times 100$

c. Fisher

$F^{t_0} = \sqrt{L^{t_0} + P^{t_0}}$

d. Sidgwick

$S^{t_0} = \frac{L^{t_0} + P^{t_0}}{2}$

e. Keynes

$K^{t_0} = \frac{\sum P_t \text{ mínimo } (q_0 \cdot q_t)}{\sum P_0 \text{ mínimo } (q_0 \cdot q_t)} \times 100$

f. Marshall

$M^{t_0} = \frac{\sum P_t (q_0 + q_t)}{\sum P_0 (q_0 + q_t)} \times 100$

g. Walsh

$W^{t_0} = \frac{\sum P_t \sqrt{q_0 + q_t}}{\sum P_0 \sqrt{q_0 + q_t}} \times 100$

4. Índices ponderados de cantidad

a. $L^{t_0} = \frac{\sum P_0 q_t}{\sum P_0 q_0} \times 100$

b. $P^{t_0} = \frac{\sum P_t q_t}{\sum P_t q_0} \times 100$

c. $F^{t_0} = \sqrt{L^{t_0} + P^{t_0}}$

d. $S^{t_0} = \frac{L^{t_0} + P^{t_0}}{2}$

e. $K^{t_0} = \frac{\sum q_t \text{ mínimo } (P_0 \cdot P_t)}{\sum q_0 \text{ mínimo } (P_0 \cdot P_t)} \times 100$

f. $M^{t_0} = \frac{\sum q_t (P_0 + P_t)}{\sum q_0 (P_0 + P_t)} \times 100$

g. $W^{t_0} = \frac{\sum q_t \sqrt{P_0 + P_t}}{\sum q_0 \sqrt{P_0 + P_t}} \times 100$

5. Examen de las fórmulas

a. Reversibilidad temporal

$p^{t_0} \times p^{t_0} \uparrow 1$

$L^{t_0} \times L^{t_0} \uparrow 1$

$F^{t_0} \times F^{t_0} = 1$

b. Reversibilidad factores

$F^{t_0} \times F^{t_0} = V = \frac{\sum P_t q_t}{\sum P_0 q_0}$

$L^{t_0} \times L^{t_0} \uparrow V$ $\quad P^{t_0} \times P^{t_0} \uparrow V$

$K^{t_0} \times K^{t_0} \uparrow V$ $\quad M^{t_0} \times M^{t_0} \uparrow V$

6. Índice de valores

a. $V = \frac{\sum P_t q_t}{\sum P_0 q_0} = \frac{V_t}{V_0}$

b. V 5 P^I 3 L^J

c. V 5 L^I 3 P^I

7. Otras fórmulas

Porcentaje de alza

a. % alza 5 $100\left[\frac{I_t}{I_0} - 1\right]$

Índice poder adquisitivo

b. I.PA = $100\left[\frac{I_0}{I_t}\right]$

c. % desvalorización 5 $100\left[1 - \frac{I_0}{I_t}\right]$

d. % alza 5 $100\left[\frac{T_t}{T_0} - 1\right]$

e. % devaluación 5 $100\left[1 - \frac{T_0}{T_t}\right]$

f. I. de producción

I.P. = $\frac{\text{producción}}{\text{IPC}} \times 100$

g. I. de productividad

I.P. = $\frac{\text{I. producción}}{\text{I. obreros}} \times 100$

h. Salario real

S.R. = $\frac{\text{S. nominal}}{\text{IPC}} \times 100$

S.R. 5 PA (Salario nominal)

i. Poder adquisitivo

PA = $\frac{1}{\text{IPC}} \times 100$

j. Relación neta de cambio

R = $\frac{\text{IVUX}}{\text{IVUM}} \times 100$

k. Capacidad para importar

CM 5 R 3 IQX

IVUX 5 Índice de valor unitario de exportación.

IVUM 5 Índice de valor unitario de importación.

IQX 5 Índice quantum de exportación.

ESTADÍSTICA
Y
MUESTREO
Ciro Martínez Bencardino

TA LA IV rmulas apli a as en la In eren ia esta sti a

PROBABILIDADES

1. Probabilidades

$$P = \frac{x}{n} - \frac{\text{éxitos}}{\text{total casos}}$$

P 1 q 5 1 $\qquad O \# P \# 1$

2. Propiedades

a. Mutuamente excluyentes (aditiva)

$P(A \text{ o } B) = P(A < B) = P(A) + P(B)$

b. Eventos solapados

$P(A < B) = P(A) + P(B) - P(A > B)$

c. Eventos complementarios

$$P(\bar{E}) = 1 - P_{(E)}$$

d. Eventos independientes

$P(A > B) = P(AB) = P(A)P(B)$

e. Probabilidad condicional

$$P(A / B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B / A) = \frac{P(B \cap A)}{P(A)}$$

f. Fórmula de Bayes

$$P(A / E) = \frac{P(A)\ P(E / A)}{P(E)}$$

$$P(B / E) = \frac{P(B)\ P(E / B)}{P(B)}$$

a. Combinaciones

$$C_r^n = \frac{n!}{(n-r)!\ r!}$$

b. Permutaciones: $Pn = n!$

c. Variaciones

$$V_r^n = \frac{n!}{(n-r)!}$$

d. Permutaciones con repetición

$$P_n = (r_1\ r_2\ r_3) = \frac{n!}{r_1!r_2!r_3!}$$

DISTRIBUCIONES

1. Distribución binomial: $P_x = C_x^n p^x q^{n-x}$
2. Distribución de Poisson

$$P_x = \frac{\lambda^x e^{-\lambda}}{x!}$$

e 5 2,71828 $\qquad$ l 5 np

A. Esperanza matemática: $E = np$

b. Juego equitativo

Apuesta 5 (Prob. éxito) (Premio)

Premio 5 Apuesta 5 Ganancia

3. Ecuación de la curva normal

p 5 3,1416 $\qquad$ e 5 2,71828

a. Función de densidad

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-} \frac{(x-u)^2}{2\sigma^2}$$

b. Cuando m 5 0 y s 5 1

$$Y = \frac{1}{\sqrt{2\pi}} e^{-} \frac{Z^2}{2}$$

c. m 5 np (media dist. binomial)

$\sigma = \sqrt{npq}$ (Desv. dist. binomial)

4. Modelo hipergeométrico:

$$P_x = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

5. Variante estadística

a. Distribución normal: $Z = \frac{x - \mu}{\sigma}$

b. Dist. de medias muestrales

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}; \qquad \mu_{\bar{x}} = \mu;$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

$$Z = \frac{\bar{x} - \mu}{S_{\bar{x}}}; \qquad S_{\bar{x}} = \frac{S_x}{\sqrt{n}}$$

c. Dist. de diferencia $\bar{x} - \bar{y}$

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}}}$$

$$\mu_{\bar{x}-\bar{y}} = \mu_x - \mu_y$$

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\sigma\frac{2}{x} + \sigma\frac{2}{y}}$$

6. Distribución muestral de una proporción

$$Z = \frac{p - P}{\sqrt{\frac{pq}{n}}} \qquad \sigma_{\bar{p}} = \sqrt{\frac{pq}{n}}$$

P 5 mp; $\qquad p = \frac{x}{n}$; $\qquad \sigma_p = \sqrt{pq}$

7. Distribución de diferencias

$$Z = \frac{(p_1 - p_2) - (\mu_{p_1} - \mu_{p_2})}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

8. Distribución "t" de Student $n \# 30$

a. $t = \frac{\bar{x} - \mu}{\hat{s} / \sqrt{n-1}}$; $\qquad t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$

$$S = \hat{s}\sqrt{\frac{n}{n-1}}; \qquad S = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}}$$

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{S_{\bar{x}-\bar{y}}}$$

$$S_{\bar{x}-\bar{y}} = \sqrt{\frac{(n_1-1)S_x^2 + (n_2-1)S_y^2}{n_1 + n_2 - 2}}\ \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_{\bar{x}-\bar{y}} = \sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}$$

$$s^2 = \frac{\sum(x - \bar{x})^2 + \sum(y - \bar{y})^2}{n_1 + n_2 - 2}$$

b. Función $Y = C\left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}$

9. Distribución ji-cuadrado

$$x^2 = \sum \frac{(n_i - n_i')^2}{n_1'}$$

10. Observaciones apareadas

a. Variante z o $t = \frac{\bar{d} - a_{\bar{d}}}{s_{\bar{d}}}$

b. $S_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$

c. $S_{\bar{d}} = \frac{S_d}{\sqrt{n}}$

TABLA IV Fórmulas aplicadas en la Ingeniería estadística. Continuación

MUESTREO

1. Muestreo aleatorio

A. Tamaño muestral (promedios)

a. $n = \dfrac{z^2 S^2 N}{E^2 N + z^2 s^2}$

b. $n = \dfrac{s^2}{\dfrac{E^2}{z^2 n_0} + \dfrac{s^2}{N}}$; $\quad n = \dfrac{N^2 S^2}{\dfrac{E^2}{z^2} + \dfrac{N^2 S^2}{N}}$

c. $n = \dfrac{n_0}{1 + \dfrac{n_0}{N}}$

$n_0 = \dfrac{z^2 S^2}{E^2}$ (promedios)

$n_0 = \dfrac{N^2 z^2 s^2}{E^2}$ (totales)

d. $n = \dfrac{z^2 s^2}{E^2}$ (población infinita)

B. Límites de confianza

$\hat{\bar{Y}}_l^s = \bar{y} \pm z \dfrac{s}{\sqrt{n}} \sqrt{\dfrac{N-n}{N}}$

$\bar{Y}_l^s = N\bar{y} \pm z \dfrac{Ns}{\sqrt{n}} \sqrt{\dfrac{N-n}{N}}$

2. Métodos proporciones

A. Tamaño muestral (proporción)

a. $n = \dfrac{z^2 NPQ}{E^2 N + z^2 PQ} = \dfrac{PQ}{\dfrac{E^2}{z^2} + \dfrac{PQ}{N}}$

b. $n = \dfrac{PQz^2}{E^2}$ (proporción)

$n = \dfrac{N^2 PQ}{\dfrac{E^2}{z^2} + \dfrac{N^2 PQ}{N}}$ (totales)

B. Límites de confianza

$\hat{p}_l^s = p \pm z\sqrt{\dfrac{pq}{n}} \sqrt{\dfrac{N-n}{N-1}}$

$\hat{A}_l^s = Np \pm zN\sqrt{\dfrac{pq}{N}} \sqrt{\dfrac{N-n}{N-1}}$

3. Estimación por razón

a. Tamaño de la muestra

$n_0 = \dfrac{z^2 s_{\hat{R}}^2}{E^2}$ $\qquad n = \dfrac{n_0}{1 + \dfrac{n_0}{N}}$

b. Límites de confianza $\hat{R}_l^s = \hat{R} \pm z\sqrt{\hat{V}_{|R|}}$

c. Varianza estimada

$\hat{V}_{|R|} = \dfrac{1-f}{\bar{x}^2 - n}\left[\dfrac{\sum y^2 - 2\hat{R}\sum xy + \hat{R}^2 \sum x^2}{n-1}\right]$

d. Razón estimada

$\hat{R} = \dfrac{\sum y}{\sum x} = \dfrac{\bar{y}}{\bar{x}}$

e. Razón (población)

$R = \dfrac{\sum Y}{\sum X} = \dfrac{Y}{X} = \dfrac{\bar{Y}}{\bar{X}}$

4. Muestreo estratificado

A. Afijación igual

a. Tamaño muestra (totales)

$n = \dfrac{\sum \dfrac{N_h^2 S_h^2}{W_h}}{(E/Z)^2 + \sum N_h S_h^2}$

b. Tamaño muestra (promedios)

$n_0 = \dfrac{1}{(E/z)^2} \sum \dfrac{W_h^2 S_h^2}{W_h}$

$n = \dfrac{n_0}{1 + \dfrac{1}{N(E/Z)^2} \sum W_h S_h^2}$

$n = \dfrac{NZ^2 \sum W_h S_h^2}{NE^2 + Z^2 \sum W_h S_h^2}$

c. Media muestral estratificada

$\bar{y}_{st} = \dfrac{\sum N_h \bar{Y}_h}{N}$; $\qquad \bar{y}_{st} = \sum W_h \bar{Y}_h$

d. Varianza estimada

$V\left[\hat{\bar{y}}_{st}\right] = \dfrac{1}{N^2} \sum \left[N_h (N_h - n_h) \dfrac{S_h^2}{n_h}\right]$

(promedios)

$V\left[\hat{Y}_{st}\right] = \sum \left[N_h (N_h - n_h) \dfrac{S_h^2}{n_h}\right]$

e. Límites de confianza

$\hat{\bar{Y}} = \bar{y}_{st} \pm z\sqrt{V_{\left[\hat{\bar{y}}_{st}\right]}}$; $\quad \hat{Y} = N\bar{y}_{st} \pm z\sqrt{V_{\left[\hat{y}_{st}\right]}}$

B. Afijación proporcional

a. Tamaño muestra (totales)

$n_0 = \dfrac{\sum N_h S_h^2}{(E/Z)^2}$

b. Tamaño muestra (promedio)

$n_0 = \dfrac{\sum W_h S_h^2}{(E/Z)^2}$ $\qquad$ c. $n = \dfrac{n_0}{1 + \dfrac{n_0}{N}}$

d. Tamaño en cada estrato:

n_h 5 $W_h(n)$ $\qquad W_h = \dfrac{N_h}{N}$

d. Media muestral estratificada

$\bar{y}_{st} = \sum W_h \bar{y}_h$

f. Varianza estimada

$V_{\left[\hat{\bar{y}}_{st}\right]} = \dfrac{N-n}{n} \dfrac{\sum W_h S_h}{n}$

$V_{\left[\bar{y}_{st}\right]} = \dfrac{N-n}{n} \sum N_h S_h^2$

g. Límites de confianza

$\hat{\bar{Y}}_{st} = \bar{y}_{st} \pm t\sqrt{V_{\left[\hat{\bar{y}}_{st}\right]}}$

$\hat{Y}_{st} = N\bar{y}_{st} \pm t\sqrt{V_{\left[\hat{y}_{st}\right]}}$

C. Afijación óptima

a. Tamaño muestra

1. $n = \dfrac{(C - C_0)\left(\sum \dfrac{N_h S_h}{\sqrt{C_h}}\right)}{\sum N_h S_h \sqrt{C_h}}$

2. $n = \dfrac{(C - C_0)}{C_h}$

3. $n = \dfrac{\left(\sum W_h S_h \sqrt{C_h}\right) \sum \dfrac{W_h S_h}{\sqrt{C_h}}}{(E/z)^2 + \dfrac{1}{N} \sum W_h S_h^2}$

4. $n = \dfrac{\left(\sum N_h S_h\right)^2}{(E/z)^2 + \sum N_h S_h^2}$ (totales)

5. $n = \dfrac{\left(\sum W_h S_h\right)^2}{(E/z)^2 + \dfrac{1}{N} \sum W_h S_h^2}$ (general)

b. Tamaño en cada estrato

1. $n_h = n\left(\dfrac{N_h S_h}{\sum N_h S_h}\right)$

2. $n_h = \left[\dfrac{\dfrac{W_h S_h}{\sqrt{C_h}}}{\sum \dfrac{W_h S_h}{\sqrt{C_h}}}\right] n$

c. Varianza estimada

$V_{\left[\bar{y}_{st}\right]} = \dfrac{\left(\sum W_h S_h\right)^2}{n} - \dfrac{\sum W_h S_h^2}{N}$

d. Límites de confianza

$\hat{\bar{Y}}_{st} = \bar{y}_{st} \pm t\sqrt{\hat{V}_{\left[\bar{y}_{st}\right]}}$

$\hat{Y}_{st} = N\bar{y}_{st} \pm tN\sqrt{\hat{V}_{\left[\bar{y}_{st}\right]}}$

BIBLIOGRAFÍA

AZORIN POCH, FRANCISCO. Curso de muestreo y aplicaciones. Edit. Aguilar.

BARBANCHO, ALFONSO G. Estadística elemental moderna. Ariel, 1973, España.

BERENSON/ LEVINE. Estadística básica en administración. Edit. Prentice Hall, 1992, México.

CAMPBELL, STHEPHEN. Equívocos y falacias de la estadística. Ed. Limusa, 1981, México.

CANAVOS C. GEORGE C. Probabilidad y Estadística. Me. Graw Hill., 1988, México.

CANSADO ENRIQUE. Estadística general. Primera parte. Cienes, 1971, Chile.

CHAO, LINCOLN. Estadística para las ciencias administrativas. Me. Graw Hill, 1975, México.

CHEVRY, 'GABRIEL R. Práctica de las encuestas estadísticas. Ediciones Ariel, 1962, España.

CHOU, YA-LUN. Análisis estadístico. Interamericana, 1969, México.

COCHRAM, WILLIAM G. Técnicas de muestreo. Edit. Cecsa

D'OTTONE R. HORACIO. Estadística Elemental. Cooperativa de Cultura y Publicaciones. Santiago de Chile.

DOMS, F.P. La estadística que sencilla. Paraninfo, 1965. Madrid.

ESPINA M., LEÓNIDAS. Estadística. Edit. Universitaria, Chile, 1968.

GILBERT. NORMA. Estadística. Edit. Interamericana, 1976, México.

GRASSAU, S. ERICKA. Elementos de estadística. Edit. Universitaria, 1968, Chile.

HARNETT- MURPHY. Introducción al análisis estadístico. Edit. Adison, 1987, México.

HOEL, PAUL. Estadística elemental. Edit. Cecsa.

JOHNSON, ROBERT. Estadística elemental. Edit. Trillas, 1979, México.

KISH, LESLIE. Muestreo de encuestas. Edit. Trillas, 1979, México.

KOOSIS, DONALD J. Introducción a la teoría de las probabilidades. Edit. Limusa.

LEVIN Y RUBIN. Estadística para administradores. Ed. Prentice Hall. 1996, México.

LEVIN. Fundamentos de estadística en la investigación social, Edit, Haría, 1997, México.

MASÓN Y LIND. Estadística para administración. Edit. Alfaomega, 1992, México.

MC RAE, T.W. Muestreo estadístico para auditoría y control. Ed. Limusa, 1978, México.

MENDENHALL - SCHEAFFER. Elementos de muestreo. Grupo Edit. Iberoamericana, 1986, México.

MENDENHALL. Estadística para las ciencias sociales. Edit. Educativa, 1987, México.

MENDENHALL. Introducción a la probabilidad y la estadística. Iberoamericana, 1982, EEUU.

MENDENHALL/REINMUTH. Estadística para administración y economía, 1978, EEUU.

MENTHA, GERALD. Los test estadísticos aplicados a la empresa. Ed. Gestión Deusto.

MILLS, RICHARD L. Estadística para economía y administración. Me. Graw Hill, 1980, Colombia.

MORONEY, M.J. Hechos y estadísticas. Eudeba, Argentina, 1965.

NOVALES, ALFONSO. Estadística y econometría. Ed. Me Graw Hill, 1996, Madrid-España.

NUÑEZ DEL PRADO, ARTURO. Estadística básica para planificación. Edit. Siglo XXI.

PARTEN, MILDRED B. Encuestas sociales y de mercadeo. Edit. Hispana, Chile.

PEPE, PAUL. Presentación de estadísticas. Ediciones Rialp S.A. España.

RAJ, DES. La estructura de las encuestas por muestreo. F.C.E.

REICHMAN, W.J. Uso y abuso de las estadísticas. Edit. Gestión Deusto, 1965, España.

SÁNCHEZ A., JAVIER I. Estadística básica aplicada. Univ. Nacional-Medellín.

SERVIN, LUIS Y ADELA. Introducción al muestreo. Limusa, 1978, México.

SMITH, MILTON. Estadística simplificada para psicólogos y educadores. Edit. Manual Moderno S.A., 1971. México.

SPIEGEL R. MURRAY. Estadística. Colección Schaum, 1991.

STEVENSON, WILLIAM J. Estadística para administración y economía. Edit. Haría, 1981, México.

SUKHATME, PANDURANG. V. Teoría de encuestas por muestreo. F.C.E. Vesselo. Como leer estadística. Uthea, 1965.

WAYNE, DANIEL W. Estadística con aplicaciones a las ciencias sociales y a la educación. Ed. Limusa, 1979, México.

WEBSTER, ALIEN L. Estadística aplicada a la empresa y economía. Edit. Irwin, 1996, España.

WONNACOTT, THOMAS. Fundamentos de estadística para administración y economía. Edit. Limusa, 1981, México.

YAMANE, TARO. Estadística. Edit. Harla, 1979, México.

ZUWAYLIF, FADIL H. Estadística general aplicada. Fondo educativo interamericano, 1977, México.

Otros títulos de interés:

- **Estadística básica aplicada,**
  Ciro Martínez Bencardino

- **Fundamentos de estadística.**
  Para la investigación en educación,
  Mireya Ardila Rodríguez

- **Álgebra lineal y programación lineal**
  Francisco Soler, Fabio Molina y
  Lucio Rojas.

- **Didáctica de las matemáticas**
  Robinson Castro Puche y
  Rubby Castro Puche.

- **Fundamentos de matemática**
  Francisco Soler Fajardo y
  Reinaldo Nuñez.

- **Matemáticas financieras aplicadas**
  Jhonny de Jesús Meza Orozco

- **Matemáticas financieras**
  empresariales
  Jhonny de Jesús Meza Orozco

- **Matemáticas para informática**
  Ismael Gutiérrez García.

Estadística y muestreo

Se llega con esta publicación a la décima tercera edición, en varias ocasiones con reimpresiones, lo cual indica su gran aceptación durante sus 32 años de vigencia en el mercado educativo. Cada edición ha estado sujeta a una rigurosa depuración, revisión y enriquecimiento continuo, a fin que su contenido se mantenga actualizado y acorde a las necesidades tanto de alumnos como de profesores.

Los 13 capítulos se distribuyen así: los cinco (5) primeros, corresponden a la Estadística Descriptiva, que incluye aspectos metodológicos y definiciones sobre términos básicos; elaboración de tablas de frecuencias, cuadros de informes y gráficas que permiten visualizar la información; interpretación de los datos por medio de la aplicación de medidas de posición (promedios), de dispersión, asimetría y apuntamiento. Los cinco siguientes, hacen referencia a la inferencia estadística e incluyen los elementos básicos de probabilidad, distribuciones muestrales, pruebas de hipótesis, límites de confianza, distribuciones no paranéfricas, Chi cuadrado, Distribución F, etc. La tercera y última parte está dedicada a los números índices, series de tiempo, regresión y correlación simple y múltiple (rectilíneo, parabólico y exponencial), incluyendo al final los elementos teóricos y las aplicaciones prácticas de los diferentes métodos de muestreo aleatorio y no aleatorio.

Además, se proponen aproximadamente mil 300 ejercicios, sin contar con un sinnúmero de ejemplos, cuyas soluciones, procesos y respuestas, están disponible en el Sistema de Información en Línea - SIL. Durante el desarrollo de la teoría, se explica el uso de Excel y la aplicación de las diferentes calculadoras, que permiten agilizar los cálculos. Contiene un resumen de fórmulas, glosario y tablas que serán utilizadas con la aplicación de estos contenidos

Área: Ciencias Exactas
Colección: Estadística.

ECOE EDICIONES

ISBN 978-958-648-702-3