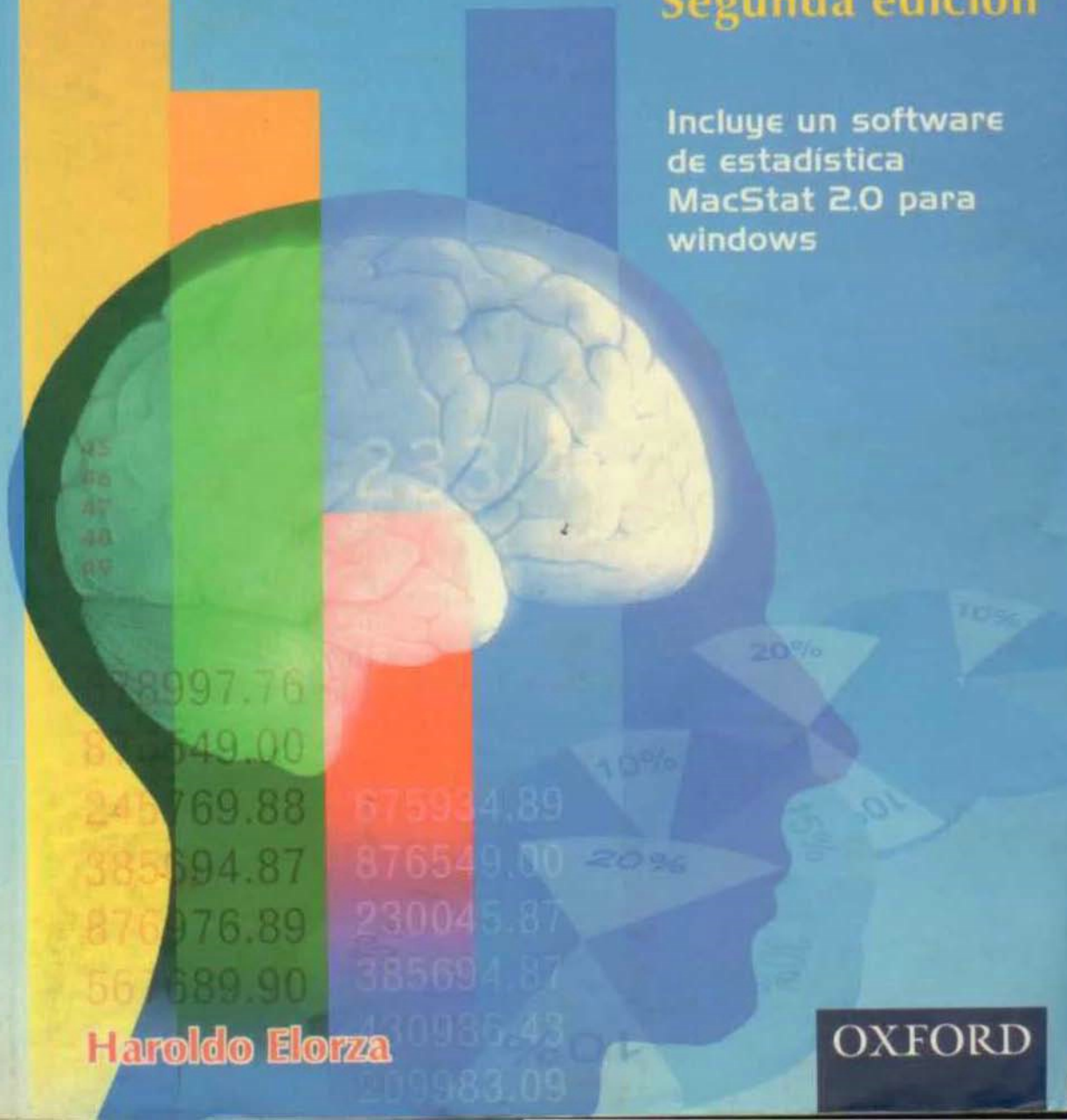
Estadística para las ciencias sociales y del comportamiento

Segunda edición

Incluye un software de estadística MacStat 2.0 para windows



Haroldo Elorza

OXFORD

Haroldo Elorza Pérez-Tejada

Obtuvo la licenciatura en física en la Universidad Nacional Autónoma de México (UNAM) y la maestría en ciencias en el Rensselaer Polytechnic Institute (RPI) de Nueva York (ex becario de CONACYT).

Su actividad docente se ha desarrollado en varias instituciones de nivel medio superior, profesional y de posgrado, como el Russell Sage College, en Nueva York; Universidad Tecnológica de México (UNITEC), Instituto Tecnológico Autónomo de México (ITAM), Universidad Anáhuac (UA), Universidad del Valle de México (UVM, campus Estado de México), Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM, campus Estado de México), Instituto Nacional de Ciencias Penales (INACIPE), Colegio de Bachilleres (SEP) y el Hospital de Especialidades C. M. La Raza (IMSS).

Ha sido conferencista en las universidades de Campeche, Veracruzana, Sonora, del Bajío (León, Gto.) y en el Tecnológico de Villahermosa, Tabasco.

En la UNAM ha sido profesor en el Colegio de Ciencias y Humanidades, la Facultad de Contaduría y Administración y en la ENEP Acatlán. Actualmente colabora en la Facultad de Psicología (en donde fue coordinador del área de matemáticas y estadística, de 1974 a 1981), y es además académico e investigador de tiempo completo en la División de Estudios de Posgrado e Investigación de la Facultad de Odontología.

En el área administrativa se ha desempeñado como jefe del Departamento de Seguimiento y Evaluación de la Dirección General de Asuntos del Personal Académico (DGAPA) de la UNAM.

Miembro de la Asociación Mexicana de Estadística AME, A.C. Asesor de encuestas del Instituto Mexicano de Opinión Pública (IMOP). Miembro del Comité Editorial de la Sociedad de Ex Alumnos de la Facultad de Psicología. Realizó la revisión técnica del libro *Fundamentos de estadística en la investigación social*, de J. Levin (Harla, 1979) y es autor de *Estadística para ciencias del comportamiento* (Harla, 1987).

Estadística para las ciencias sociales y del comportamiento

booksmedicos.org

CONTRAPORTADA

Viaje por el microcosmos y en marcha por la infinidad del espacio de Rafael Ortizgris, 1994.

"De un fondo azul oscuro con tintes violetas, negros y grises, emergen una serie de líneas que forman figuras y manchas de diversos colores. Sólo a cierta distancia se pueden identificar las formas y los temas que de manera entretejida componen el mural. A la izquierda surgen tres grandes figuras que simbolizan la evolución del hombre y su conocimiento. Atrás, dentro de un cubo suspendido que crece secuencialmente, se encuentra un embrión humano en desarrollo impulsado por el movimiento circular que impregna de energía vital toda la obra. Por último, en la parte superior una figura humana se dirige hacia un horizonte infinito. Las figuras no se encuentran aisladas, sino que todos los elementos fluyen para tocarse entre sí. Desde la dimensión personal de Ortizgris, los colores, las formas y las líneas crean una sustancia nueva emergente. El punto central del mural es una gran esfera que marca un inicio / final evolutivo que comienza / termina en otra esfera. Dicha esfera proyecta luces de colores que atraviesan una serie de figuras humanas, para finalmente rematar en un par de manos extendidas. La obra está plena de movimiento. El artista creó una sinfonía, un engranaje en donde cada molécula se sostiene gracias al equilibrio de las demás."

Con la colaboración de:

SERAFÍN MERCADO DOMENECH
Psicólogo, UNAM.
Doctor en psicología, University of Texas (Austin).
Profesor e investigador en la División de Estudios de Posgrado; iniciador de la maestría en psicología ambiental, Facultad de Psicología, UNAM.

ADIP SABAG SABAG
Psicólogo, UNAM.
Maestría en sociología, Universidad de Lovaina, Bélgica.
Doctor en prospectiva, Universidad de París (Sorbona).
Director general del Instituto Superior de Estudios Prospectivos, ISEP.

JUAN CARLOS MEDINA SANDOVAL
Licenciado en matemáticas aplicadas y computación, ENEP Acatlán, UNAM.
Consultor en Dataconsult, S. A. de C. V.
jcms@iname.com.mx

FEDERICO VÁZQUEZ TAPIA
Licenciado en actuaría, Facultad de Ciencias, UNAM.
Profesor y asesor en el sistema de Universidad Abierta en la Facultad de Contaduría y Administración, UNAM.

FRANCISCO GONZÁLEZ ALAMILLA
Licenciado en actuaría, ENEP Acatlán, UNAM.
Gerente de proyectos en el Buró de Investigación de Mercados, S. A. de C. V.
alamilla@bimsa.com.mx

Estadística para las ciencias sociales y del comportamiento

Haroldo Elorza
Universidad Nacional Autónoma de México

Autor del software Macstat 2.0:
Juan Carlos Medina Sandoval

booksmedicos.org

Segunda edición

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Antonio Caso 142, San Rafael,
Delegación Cuauhtémoc, C.P. 06470, México, D.F.
Tel.: 5592 4277, Fax: 5705 3738, e-mail: oxford@oupmex.com.mx

Oxford University Press es un departamento de la Universidad de Oxford.
Promueve el objetivo de la Universidad de la excelencia en la investigación, erudición
y educación mediante publicaciones en todo el mundo en

Oxford México

Atenas Auckland Bangkok Buenos Aires Calcuta
Caracas Chennai Ciudad del Cabo Dar-es-Salaam Delhi Estambul Florencia
Hong Kong Karachi Kuala Lumpur Madrid Melbourne Mumbai
Nairobi Nueva York París Santafé de Bogotá Santiago de Chile São Paulo
Singapur Taipei Tokio Toronto Varsovia

Con compañías afiliadas en Berlín Ibadan

División: College
Área: Estadística

Sponsor editor: Jorge Alberto Ruiz González
Editor: Lilia Guadalupe Aguilar Iriarte
Producción: Antonio Figueredo Hurtado

**ESTADÍSTICA PARA LAS CIENCIAS SOCIALES
Y DEL COMPORTAMIENTO**

Todos los derechos reservados © 2000, respecto a la segunda edición por
Oxford University Press México, S.A. de C.V.
Ninguna parte de esta publicación puede reproducirse, almacenarse en un sistema de recuperación
o transmitirse, en ninguna forma ni por ningún medio, sin la autorización previa y por escrito de
Oxford University Press México, S.A. de C.V.
Las consultas relativas a la reproducción deben enviarse al Departamento de Permisos y
Derechos de Oxford University Press México, S.A. de C.V., al
domicilio que se señala en la parte superior de esta página.
Miembro de la Cámara Nacional de la Industria
Editorial Mexicana, número de registro 723.

ISBN 970-613-181-7

Impreso en México
1 2 3 4 5 6 7 8 9 0

Printed in Mexico
9 8 7 6 5 4 3 2 1 9

Esta obra se terminó de imprimir
en el mes de noviembre de 1999 en
Reprofló, S.A. de C.V.
Chipiona No. 115
Cerro de la Estrella
09880, México, D.F.

El tiraje fue de 3 000 ejemplares.

OXFORD UNIVERSITY PRESS
MEXICO, S.A. DE C.V.
PRUEBA
Nº 83651
PROHIBIDA SU VENTA

Contenido

Presentación **xv**
Introducción **xvii**

PARTE 1
ESTADÍSTICA Y CIENCIA 1

CAPÍTULO 1 ESTADÍSTICA E INVESTIGACIÓN **3**

1.1 Aspectos fundamentales de la ciencia **3**
Nivel filosófico **3**
Introducción **3**
1.1.1 Explicación y teoría **5**
1.1.2 Naturaleza de la investigación **6**
1.1.3 Justificación contra confrontación **9**
1.1.4 ¿De dónde surgen las teorías? **9**
1.2 Relaciones entre estadística e investigación **10**
1.2.1 Error de medida y experimental **10**
1.3 Medición y estadística **11**
1.3.1 Escalas de medición **12**
1.3.2 Limitación de la estadística por nivel de medida **13**
1.4 Inferencia estadística y científica **14**
1.4.1 Diseño experimental **15**
1.4.2 Diseño cuasiexperimental **16**
1.5 Estadística e informe científico **17**
1.5.1 Gráficas **17**
Resumen **17**
Cuestionario **18**

Un modelo de investigación **18**
1.6 Datos iniciales **19**
1.7 Tipo y tamaño de la muestra **21**
1.8 Instrumento de medición **24**
1.9 Procesamiento y análisis de datos **26**
1.10 Informe final **26**

PARTE 2
MODELOS DETERMINÍSTICOS 29

CAPÍTULO 2 DESCRIPCIÓN DE DATOS **31**
Propósitos **31**
Introducción **32**
Caso A. Datos agrupados **32**

2.1 Gráficas y distribuciones de frecuencia **32**
2.2 Sumatorias **47**
2.3 Medidas de tendencia central **49**
2.4 Cuantilas **54**
2.4.1 Diagrama de cajas **64**
2.5 Medidas de dispersión o variabilidad **66**
2.6 Propiedades de la distribución de frecuencias **68**

Caso B. Datos no agrupados **73**

2.7 Medidas de tendencia central **74**
2.8 Medidas de dispersión o variabilidad **76**
2.9 Otros tipos de promedio **83**
2.9.1 Media ponderada $\overline{X}_p$ **83**
2.9.2 Media armónica **84**
2.9.3 Media geométrica **84**
Resumen **85**
Problemas **86**

CAPÍTULO 3 CONJUNTOS, FUNCIONES Y MATRICES **97**

Propósitos **97**
Introducción **98**
3.1 Conjuntos, conceptos y notación **99**
3.1.1 Conjunto **99**
3.1.2 Elemento **99**
3.1.3 Reglas y formas para enunciar los conjuntos **99**
3.1.4 Conjuntos finitos e infinitos **100**
3.2 Relación de pertenencia **100**

3.3 Relaciones entre conjuntos **100**
3.3.1 De inclusión **100**
3.3.1.1 Subconjuntos **101**
3.3.1.2 Número de subconjuntos de un conjunto **101**
3.3.2 Conjuntos ajenos **102**
3.3.3 Conjunto universal **102**
3.3.3.1 Conjunto vacío **102**
3.3.4 Conjuntos iguales **102**
3.3.4.1 Conjuntos similares **103**
3.4 Diagramas de Venn-Euler y de Carroll **103**
3.4.1 Conjunto complemento **106**
3.5 Álgebra de conjuntos **106**
3.5.1 Unión **106**
3.5.2 Intersección **108**
3.5.3 Diferencia o resta aritmética **109**
3.5.4 Diferencia simétrica entre dos conjuntos **111**
3.5.5 Propiedades de los conjuntos **111**
3.6 Conteo de elementos **113**
3.6.1 Cardinalidad **114**
3.6.2 Conteo de elementos para tres conjuntos **115**
3.6.3 Conteo de elementos para más de tres conjuntos **118**
3.7 Producto cartesiano **119**
3.8 Relaciones y funciones **121**
3.8.1 Dominio y contradominio **123**
3.8.2 Variabilidad **126**
3.8.2.1 Variables dependientes e independientes **127**
3.8.3 Intervalos y desigualdades **127**
3.8.4 Clasificación de funciones **129**
3.8.5 Álgebra de funciones **146**
3.9 Matrices **146**
3.9.1 Orden de una matriz **147**
3.9.2 Tipos de matrices **147**
3.9.2.1 Matriz cuadrada **148**
3.9.2.2 Matriz diagonal **148**
3.9.2.3 Matriz escalar **148**
3.9.2.4 Matriz identidad **149**
3.9.2.5 Matriz nula **149**
3.9.2.6 Igualdad de matrices **149**
3.9.3 Determinantes **149**
3.9.4 Álgebra de matrices **151**
3.9.4.1 Transpuesta de una matriz **151**
3.9.4.2 Matriz simétrica **151**
3.9.4.3 Operaciones con matrices **152**
3.9.4.4 Multiplicación de matrices **152**

3.10 Sistemas de ecuaciones lineales **156**
Resumen **166**
Problemas **166**

CAPÍTULO 4 CÁLCULO DE POSIBILIDADES **177**

Propósitos **177**
Introducción **178**

4.1 Experimentos **178**
4.2 Principios fundamentales del conteo **179**
4.2.1 Diagramas de árbol **179**
4.2.2 Principio de la multiplicación **182**
4.2.3 Principio de la adición **182**
4.3 Permutaciones **185**
4.3.1 Cuando $r < n$ **185**
4.3.2 Cuando $r = n$ **186**
4.3.3 Circulares **186**
4.4 Combinaciones **187**
4.5 Coeficiente multinomial o particiones **188**
4.6 Teorema del binomio **189**
Resumen **191**
Problemas **192**

PARTE 3
MODELOS NO DETERMINÍSTICOS **195**

CAPÍTULO 5 CÁLCULO PROBABILÍSTICO **197**

Propósitos **197**
Introducción **198**

5.1 Definiciones, símbolos y terminología **200**
5.2 Axiomas de probabilidad **203**
5.3 Particiones **204**
5.3.1 Eventos mutuamente excluyentes (*me*) **206**
5.4 Probabilidad condicional **212**
5.4.1 Eventos independientes **215**
5.4.2 Probabilidad marginal **219**
5.5 Teorema de la probabilidad total **219**
5.6 Teorema de Bayes **224**
5.7 Procesos estocásticos **233**
5.7.1 Conceptos y definiciones **233**

5.7.2 Cadenas de Markov **234**
5.7.3 Representación gráfica **235**
5.7.4 Representación matricial **236**
Resumen **248**
Problema **249**

CAPÍTULO 6 DISTRIBUCIONES PROBABILÍSTICAS **261**

Propósitos **261**
Introducción **262**

6.1 Variable aleatoria **262**
6.2 Desigualdad de Chebyshev **263**
6.3 Ley de los grandes números **265**
6.4 Distribución Bernoulli **266**
6.5 Distribución binomial **268**
6.6 Distribución de Poisson **274**
6.6.1 Aproximación de la distribución binomial a la de Poisson **276**
6.7 Distribución hipergeométrica **277**
6.8 Distribución normal **278**
6.8.1 Distribución Z **280**
6.8.2 Aproximación de la distribución binomial a la normal **284**
6.8.3 Corrección de continuidad **285**
6.9 Distribución exponencial **287**
Resumen **290**
Problemas **290**

CAPÍTULO 7 TÉCNICAS DE MUESTREO **296**

Propósitos **296**
Introducción **297**

7.1 Fundamentos de muestreo **297**
7.2 Clasificaciones de muestreo **298**
7.2.1 Muestreo determinístico contra probabilístico **300**
7.3 Muestreo determinístico **302**
7.3.1 Muestreo de juicio o criterio **302**
7.3.2 Muestreo por cuotas **303**
7.3.3 Diseño de bola de nieve **303**
7.3.4 Muestreo por conveniencia **303**
7.4 Muestreos probabilísticos **304**
7.4.1 Muestreo aleatorio simple **304**
7.4.2 Muestreo aleatorio estratificado (MAE) **309**
7.4.3 Muestreo sistemático **312**

7.4.4 Muestreo por conglomerados **315**
7.5 Distribuciones muestrales **317**
7.6 Teorema central del límite **317**
Resumen **329**
Problemas **329**

PARTE 4
INFERENCIA ESTADÍSTICA **335**

CAPÍTULO 8 CONCEPTOS BÁSICOS **337**

Propósitos **337**
Introducción **338**

8.1 Estimación puntual **339**
8.1.1 Error estándar de la media muestral **344**
8.2 Estimación intervalar **345**
8.3 Hipótesis estadísticas y científicas **347**
8.4 Clasificación de hipótesis **350**
8.5 Prueba de hipótesis **351**
8.5.1 Error de tipo I y II **351**
8.5.2 Curva característica de operación **354**
8.5.3 Reglas operativas para realizar un contraste de hipótesis **355**
8.6 Relación entre tamaño de muestra y prueba de hipótesis **356**
Resumen **362**
Problemas **363**

CAPÍTULO 9 ESTUDIO DE UNA POBLACIÓN **366**

Propósitos **366**
Introducción **367**

9.1 Uso de las tablas de distribución normal **367**
9.2 Intervalo de confianza y contraste de hipótesis para una media **372**
9.3 Intervalo de confianza y contraste de hipótesis para la media si σ_X^2 no se conoce 377
9.3.1 Pruebas de hipótesis para μ_X si $\sigma_X^2 = ?$ **378**
9.4 Intervalo de confianza y contraste de hipótesis para una proporción, cuando la varianza es desconocida **381**
9.4.1 Resumen de pruebas de hipótesis para la proporción si $\sigma_X^2 = \pi(1 - \pi) = ?$ **384**

9.5 Intervalo de confianza y contraste de hipótesis para una varianza **387**
9.5.1 Pruebas de hipótesis para $\sigma_X^2 = \pi(1 - \pi) = ?$ **390**
Resumen **394**
Problemas **394**

CAPÍTULO 10 ESTUDIO DE DOS POBLACIONES **400**

Propósitos **400**
Introducción **401**

10.1 Intervalo de confianza y contraste de hipótesis para dos medias **401**
10.2 Intervalo de confianza y contraste de hipótesis para dos medias (muestras independientes) **405**
10.2.1 Intervalo de confianza y contraste de hipótesis para dos medias (muestras dependientes, relacionadas o una sola muestra medida dos veces) **413**
10.2.2 Prueba de Sandler (*A*) **420**
10.3 Intervalo de confianza para diferencia entre dos proporciones **421**
10.3.1 Contraste de hipótesis **424**
10.4 Intervalo de confianza y contraste de hipótesis para dos varianzas **426**
10.5 Contraste para más de dos varianzas **431**
10.5.1 La prueba de Cochran consiste en obtener la suma de las varianzas muestrales de las *k* muestras **431**
10.5.2 Prueba de Hartley **433**
10.5.3 Prueba de Bartlett **434**
Resumen **440**
Problemas **440**

CAPÍTULO 11 ANÁLISIS DE VARIANZA DE UN FACTOR **445**

Propósitos **445**
Introducción **446**

11.1 La lógica del análisis de varianza **447**
11.2 Diseño completamente aleatorizado **448**
11.2.1 Para dos tratamientos ($k = 2$) **448**
11.2.2 Caso A) Mismo número de sujetos por tratamiento; ($k > 2$) **456**
11.2.2.1 Prueba de Tukey-Snedecor **462**
11.2.3 Caso B) Diferente número de sujetos por tratamiento: ($k > 2$) **464**
11.2.3.1 Prueba de Tukey-Snedecor (grupos desiguales) **468**

11.2.4 Supuestos del análisis de varianza **469**
11.3 Prueba de Dunnett (d) **470**
Resumen **472**
Problemas **472**

CAPÍTULO 12 ANÁLISIS DE VARIANZA DE DOS FACTORES **479**

Propósitos **479**
Introducción **480**

12.1 Diseño de bloques aleatorizados **480**
12.2 Diseño completamente aleatorizado **486**
12.2.1 El concepto de interacción **491**
Resumen **494**
Problemas **495**

PARTE 5
ASOCIACIÓN 501

CAPÍTULO 13 ANÁLISIS DE REGRESIÓN LINEAL **503**

Propósitos **503**
Introducción **504**

13.1 El modelo de regresión **505**
13.2 Cálculo de la recta de regresión **506**
13.3 Evaluación de la ecuación de regresión **509**
13.3.1 Intervalos de confianza para: α, β, $\sigma^2_{y/x}$, $\mu_{y/x}$ **509**
13.3.2 Pruebas de hipótesis para la regresión **514**
13.4 Análisis de correlación **516**
13.4.1 Intervalo de confianza para el coeficiente de correlación **518**
13.4.2 Valor promedio del coeficiente de correlación **519**
13.4.3 Prueba de hipótesis para el coeficiente de correlación **520**
13.5 Prueba de hipótesis entre dos rectas de regresión **522**
13.5.1 Comparación entre dos coeficientes de correlación **526**
13.6 Análisis de regresión múltiple **528**
13.6.1 Caso particular para dos variables independientes **528**
13.6.2 Relación entre el análisis de varianza y el de regresión múltiple **532**
13.6.3 Evaluación de la ecuación de regresión múltiple **533**
13.6.4 Prueba de hipótesis para los coeficientes de regresión **536**
13.6.5 Coeficiente de correlación múltiple **538**
Resumen **539**
Problemas **539**

CAPÍTULO 14 ANÁLISIS DE FRECUENCIA Y ESTADÍSTICA NO PARAMÉTRICA **545**

Propósitos **545**
Introducción **546**

14.1 Pruebas de bondad de ajuste **547**
14.1.1 Ji cuadrada (χ^2) **547**
14.1.2 Kolmogorov-Smirnov (*K-S*) **557**
14.2 Tablas de contingencia **559**
14.2.1 Prueba de homogeneidad **559**
14.2.2 Prueba de independencia **563**
14.2.3 Procedimiento post hoc **565**
14.2.3.1 Coeficiente fi (ϕ) **568**
14.2.3.2 Coeficiente de contingencia (*C*) **572**
14.2.3.3 Coeficiente *V* de Cramer **575**
14.3 Prueba de la *U* de Mann-Whitney **577**
14.4 Prueba de rangos con signos en pares de Wilcolxon **584**
14.5 Prueba de Mcnemar **588**
14.6 Prueba de la mediana **591**
14.7 Prueba de Kruskal-Wallis (*H*) **594**
14.8 Prueba de Friedman **597**
14.8.1 Prueba de Nemenyi **600**
14.9 Coeficiente de Spearman (r_s) **601**
14.10 Coeficiente tau (τ) de Kendall **607**
14.11 Coeficiente de concordancia, ω, de Kendall **611**
14.12 Coeficiente de correlación (r_{bp}) biserial de punto **613**
Resumen **617**
Problemas **618**

Bibliografía **623**
Anexo 1 Tablas **625**
Anexo 2 Soluciones a los problemas **691**
Índice analítico **769**

Protocolo para la evaluación o desarrollo de un artículo, ponencia, estudio, tesis, etcétera

1. El **título** describe o indica clara y directamente el propósito del estudio.
2. El **resumen** muestra en forma precisa la magnitud de la investigación.
3. Contiene una **introducción** y/o **antecedentes** del tema por tratar.
4. El **problema** se plantea explícita y claramente.
5. El **problema** es relevante, factible y congruente.
6. El **problema** se justifica en forma evidente y exhaustiva.
7. Se relaciona y ubica el **problema** con investigaciones previas y la literatura.
8. Los **objetivos** se establecen en forma clara y organizada.
9. Los **supuestos** están ubicados dentro de un **marco teórico** establecido.
10. Las **limitaciones del estudio** están detectadas y se mencionan.
11. Se identifican y clasifican adecuadamente las **variables** que intervienen en el estudio.
12. Existe interacción de las **variables** con las **hipótesis** y éstas se proponen adecuadamente.
13. Se definen operacionalmente los **términos**, **conceptos** y **constructos** importantes.
14. El **diseño experimental** es el adecuado y el **tipo de estudio** se describe completamente.
15. Se especifica la **población de estudio**, así como los criterios de inclusión, exclusión y eliminación de las unidades experimentales, sujetos, etc., de la **muestra**.
16. El **método de muestreo** y la **muestra** obtenida son confiables y representativos.
17. Se describen los métodos o procedimientos de **recopilación** de la **información**.
18. Las **observaciones** o **mediciones** son objetivas y verificables.
19. La **información** y **datos** obtenidos son relevantes y confiables.
20. El **análisis estadístico** de los datos es el apropiado y se aplica correctamente.
21. Los **resultados** del análisis estadísticos se presentan claramente y sin ambiguedades.
22. Las **conclusiones** son claras, no presentando los resultados negativos como positivos.
23. Las **generalizaciones** están circunscritas a la **población de estudio.**
24. El **informe final** está directamente escrito y lógicamente organizado.
25. El **informe final** no es tendencioso y muestra una actitud científica racional y honesta.
26. La investigación, estudio o experimento no debe implicar conflictos **bioéticos.**

Presentación

Durante la preparación de esta segunda edición, tuve en mente como destinatarios principales a los estudiantes que enfrentan por primera vez a la estadística, ya sea como materia obligatoria en su currículum o porque tienen que realizar un estudio, investigación o experimento y, debido a las características de su carrera profesional, no poseen una sólida preparación matemática. A causa de ello el desarrollo de los temas que conforman este texto constituyó un reto y uno de sus objetivos principales es el razonamiento lógico de la estadística aplicada a distintas áreas de las ciencias sociales y del comportamiento, así como de la salud, sin disminuir la formalidad de los métodos estadísticos.

Esta segunda edición contiene un primer capítulo en el que se considera la relación entre estadística e investigación. En el capítulo 2 se analiza la estadística descriptiva. En el 3, se consideran los conjuntos, las funciones y matrices básicos para entender los siguientes temas: cálculo de posibilidades (capítulo 4), también conocido como *análisis combinatorio*, cálculo probabilístico (capítulo 5) en donde se consideran los conceptos básicos e intuitivos de probabilidad y procesos estocásticos; en el capítulo 6 se desarrollan algunas distribuciones probabilísticas que tienen aplicaciones en casi todas las disciplinas profesionales, y en el 7 técnicas de muestreo, tan necesarias para realizar cualquier investigación.

En la parte de inferencia estadística, el capítulo 8 (conceptos básicos) es la clave para comprender el razonamiento de los modelos y del análisis estadístico; los capítulos 9 y 10 se refieren al estudio de una y dos poblaciones, considerando el contraste de hipótesis y sus respectivos intervalos de confianza. Tres diseños experimentales son considerados en los capítulos 11 y 12 (análisis de varianza de un factor y de dos factores), respectivamente. El modelo de regresión lineal simple y múltiple conforma el capítulo 13. La popular ji cuadrada y varias pruebas no-paramétricas se consideran en el capítulo 14.

Casi todos los temas se ilustran con ejemplos desarrollados paso a paso y se proponen cerca de 300 problemas, todos ellos con su respuesta. Pero el nú-

mero de ejercicios se torna infinito, debido a que este libro está acompañado por el paquete estadístico MacStat para windows, que reduce en forma amigable muchas horas de tediosos cálculos, así como la posibilidad de cometer errores al realizarlos y sobre todo el número ilimitado de ensayos que se pueden realizar tanto con datos reales como imaginarios examinando una gran variedad de estrategias de análisis de datos y modelos estadísticos que se pueden aplicar.

En este reto no estuve solo, por lo que quiero expresar mi agradecimiento a todas las personas que participaron en forma directa o indirecta para llevar a cabo con éxito este proyecto.

En primer lugar, quiero mencionar y al mismo tiempo rendir un tributo a la tolerancia de mis alumnos de la Facultad de Psicología y del Posgrado de la Facultad de Odontología, ambas de la UNAM, quienes fueron los primeros en utilizar este material; a los doctores Serafín Mercado Domenech y Adip Sabag Sabag, por desarrollar con maestría el primer capítulo. Por las largas, agobiantes, pero agradables y fructíferas horas de trabajo que implica el planear y llevar a cabo la realización de este libro, al matemático J. Carlos Medina Sandoval, a los actuarios Federico Vásquez Tapia y Francisco González Alamilla, quienes contribuyeron a la revisión técnica del libro; a la maestra María Elena Junco, por sus sugerencias al capítulo de muestreo; al psicólogo Víctor Gutiérrez, quien diseñó los objetivos educativos. Al doctor Jorge Flores Valdez, director del Centro Universitario de Comunicación de la Ciencia 1997, por las facilidades brindadas para la fotografía del mural *Viaje por el microcosmos y en marcha por la infinidad del espacio* (1994), del maestro Rafael Ortizgris, realizada en forma creativa por el Fis. Arturo Orta Fuentes, jefe del Departamento de fotografía del Museo Universum. A la doctora Josefina Mena, del Grupo Tecnología Alternativa, por su irrestricta ayuda. A la maestra Vivianne Javelly, jefa del Departamento de Personal Académico de la Facultad de Psicología. Al licenciado Carlos Trillas, gerente Editorial de Trillas, S. A., por su generosa autorización para utilizar las tablas estadísticas del libro Métodos estadísticos, un enfoque interdisciplinario de los doctores S. Infante Gil y G. P. Zárate de Lara, así como a la American Statistical Association; a los Trustees de Biometrika y a Oxford University Press de Inglaterra, por su apreciable permiso para reproducir las tablas correspondientes. Y de Oxford University Press México, a la licenciada Claudia Martínez y al ingeniero Antonio Figueredo, por su invaluable apoyo, así como a la licenciada Brenda Reyes de diseño gráfico. En general, a mis amigos, que en los momentos difíciles tuvieron a bien alientarme, mi sincero agradecimiento.

Y por supuesto, me responsabilizo por los errores tanto sistemáticos como aleatorios que siempre están al acecho en este tipo de trabajo, y también por los desagradables momentos que pude haber ocasionado por tratar de realizar este sueño.

Hay que tener presente que cada proyecto que uno idealiza y pretende realizar, pone en juego toda la voluntad posible, sin olvidar una genial frase de Jorge Luis Borges: "qué breves son los años y qué largas son las horas".

Haroldo Elorza Pérez-Tejada

Introducción

La investigación es una característica de los últimos siglos, representa la herencia más clara y de mayor trascendencia hecha por la cultura occidental. La vida se ha transformado vertiginosamente en los últimos siglos por la acumulación de capacidad cognoscitiva para comprender y lidiar con la realidad, sea ésta física, biológica o social.

Con el desarrollo de la tecnología en todos los ámbitos, la vida humana se ha ido permeando por los productos del saber que genera el uso del método científico. Aun campos que parecen muy lejanos a la ciencia, como el arte, se han visto influidos por este proceso. Por ejemplo, las artes visuales han adquirido gran cantidad de medios de expresión, y el conocimiento más profundo de la visión y la cognición visual les ha dado mayores capacidades. La música también se ha beneficiado de la acústica, de la fisiología y de la psicología de la audición.

Por ello, cualquiera que sea nuestro ámbito de trabajo, el entender la naturaleza del quehacer científico y de los datos y las teorías nos hace más competentes. El aprendizaje de la estadística y del diseño experimental se ha vuelto, pues, una necesidad cotidiana para los profesionales e indispensable para los investigadores.

Cuando se realiza una investigación aplicada o básica, se establece un diálogo entre el investigador y el matemático. El investigador tiene una idea clara y precisa de adónde dirigir su estrategia de investigación; el matemático, a su vez, tiene muy claros los modelos estadísticos y sus respectivos diseños experimentales. Una vez iniciado el estudio, investigación o experimento, la información relevante representada por los datos, sean observados, recolectados o extraídos de alguna otra forma, no siempre es obvia y nítida como para aplicar directa y sensiblemente el análisis estadístico adecuado y así concluir acertadamente. También al leer los libros serios y los artículos donde se presentan los resultados

de la investigación, se requiere el conocimiento estadístico para poder comprender los argumentos del investigador y la naturaleza de la evidencia con que apoya sus teorías.

Uno de los objetivos de este libro es minimizar este punto crítico, haciendo compatible la investigación creativa con el modelo estadístico riguroso, desarrollando el concepto de interactividad entre ambos modelos y haciéndolo accesible para quienes sólo serán usuarios de la estadística, ya sean consumidores de información científica o investigadores, y para los que adentrándose en la matemática subyacente trabajarán en el desarrollo de nuevos modelos y técnicas o escribirán los programas de cómputo que nos apoyan como usuarios.

Considerando lo anterior, es necesario:

- Promover una cultura estadística.
- Reducir la distancia entre el alud de información generada y su utilización relevante.
- Promover el interés por consultar las cifras.
- Propiciar la publicación de libros escritos por profesionales nacionales en este campo.

El libro del profesor Elorza conjuga las necesidades del investigador con el rigor de la estadística y de manera sencilla y lógica va superando, ejemplo tras ejemplo, las dificultades de cálculo e interpretación.

Para usar este libro de manera efectiva es necesario no sólo leer la información, sino aprender a usar el paquete estadístico adjunto, examinar cuidadosamente los ejemplos y realizar los ejercicios.

Dr. Serafín Mercado Domenech

Estadística para las ciencias sociales y del comportamiento

Estadística e investigación 1

1.1 ASPECTOS FUNDAMENTALES DE LA CIENCIA

NIVEL FILOSÓFICO

Dr. Serafín Mercado Domenech

INTRODUCCIÓN

La ciencia se basa en la contrastación empírica de las teorías con la evidencia. Las teorías se comprueban tratando de demostrar que son falsas; si no se logra esto, se retiene la teoría. El método de la ciencia es el de las conjeturas audaces e ingeniosas seguidas por intentos rigurosos de refutarlas.

Las teorías tratan de dar sentido a los hechos de la realidad y explicarlos. La investigación es la parte creativa de la ciencia. Procura establecer la relación entre variables. La estadística trata de poner a prueba las hipótesis y usa el concepto de error como ruido. Hay una relación entre la medición, los errores de medida y la estadística.

La ciencia es una de las empresas más productivas que haya desarrollado el ser humano. Si lo que caracteriza al hombre es su excepcional inteligencia, la cual le ha dotado de lenguaje y le ha permitido servirse de él para crear una singular organización social de insólita eficacia para dominar la naturaleza, la ciencia es entonces el logro humano más perfecto y contundente, el cual señala la cúspide de los frutos de su intelecto, único en el Sistema Solar y tal vez en el universo mismo.

La ciencia es el conocimiento que ha permitido una mejor comprensión de la realidad circundante, basada en un proceso analítico y crítico, así como penetrar en los secretos más profundos del mundo, incluido el ser de su propio creador. Ésta también ha facilitado el desarrollo de teorías que exponen la realidad, basadas en un examen de la relación entre los intentos de explicación teórica, evidencia empírica, y con congruencia lógica tanto interna a la explicación como en lo relativo a otras teorías con las que se tienen vínculos. Esto ha implicado que el científico pruebe sus teorías confrontándolas con la evidencia existente que, con objeto de evaluar la teoría de que se trata, se acumula con procedimientos rigurosos. Asimismo, está a la caza de inconsistencias internas en la lógica de las explicaciones, así como de las contradicciones entre las diversas teorías relacionadas.

Aquí se hace relevante la discusión acerca de la naturaleza de las teorías y de su desarrollo de Popper (1959, 1963, 1970, 1974*a*, 1974*b*) y Kuhn (1963, 1970*a*, 1970*b*, 1971, 1974) y elaboradas las posteriores de Kneller (1978), Lakatos (1964, 1968*a*, 1968*b*, 1970, 1971, 1974, 1975); Feyerabend (1962, 1965*a*, 1965*b*, 1970*a*, 1970*b*, 1970*c*), Maxwell (1974) y Weimer (1979), entre otros.

La conceptualización del empirismo y del positivismo acerca de la naturaleza del proceso de acumulación de conocimiento se ha basado siempre en el proceso de inducción. Este principio señala que si observamos una cierta regularidad en los procesos naturales (incluida la naturaleza humana), entonces podemos generalizar, estableciendo una ley. De acuerdo con esta visión, el problema es observar cuidadosamente la naturaleza, evitando caer en errores debidos a la posible confusión de causas. El mejor modo de evitar el error es hacer una cuidadosa observación y medición del fenómeno, y utilizar el método experimental para no confundir la verdadera causa de los fenómenos. De acuerdo con ellos, los hechos observados y establecidos *prueban* una cierta concepción de la realidad. Al ser entonces el proceso científico un proceso lineal y acumulativo, las teorías constituirían simplemente la organización lógica de las leyes empíricas y la explicación de varias de ellas por principios más generales, surgidos de la inducción.

Hume ya había planteado la naturaleza de las limitaciones lógicas del conocimiento inductivo, pues independientemente de cuántas observaciones se hayan hecho de una regularidad, esto no da ninguna "garantía lógica" de que volverá a ocurrir del mismo modo en la siguiente ocasión. La salida planteada por Popper (1972) a este dilema fue en términos de postular que no podemos partir nunca de ninguna certidumbre acerca de nada de lo que creemos. De acuerdo con él, toda teoría, todo modelo o toda ley científica es una conjetura de cómo es la realidad; no importa que su origen sea una *inducción*, un conocimiento tácito, tal vez de carácter personal, o una especulación; la teoría es una conjetura, una suposición, una hipótesis acerca de la realidad.

Las teorías, plantea ese autor, basan su desarrollo en la confrontación crítica con los hechos y con la lógica. De acuerdo con Popper, *ningún conjunto de enunciados contrastadores verdaderos podrá justificar la pretensión de que una teoría explicativa universal es verdadera* (Popper, K.R., 1974, *Conocimiento*

objetivo, Technos, Madrid, p. 20). Sin embargo, afirma: *suponiendo que los enunciados contrastadores sean verdaderos, basándonos en ellos, podemos a veces justificar la pretensión de que una teoría explicativa universal es falsa* (*op. cit.*, p. 20). Esto desplaza el énfasis de la investigación al contrario de como lo plantea el punto de vista tradicional científico, es decir, no probamos que las teorías sean verdaderas: sólo podemos eliminar las falsas. Por ello, Popper señala: *El método de la ciencia es el método de las conjeturas audaces e ingeniosas seguidas por intentos rigurosos de refutarlas* (Popper, *op. cit.*, p. 83).

De acuerdo con su punto de vista, toda ley, todo principio, toda teoría o todo modelo es una conjetura, una suposición. Las teorías no surgen, como supondrían los llamados *justificacionistas*,[1] mediante el proceso de inducción a partir de los datos que, en todo caso, tan sólo proporcionan una *inspiración* inicial para la concepción de una teoría y no son una base empírica para el proceso lógico de la generalización por inducción.

Las teorías se valoran por su poder explicativo y heurístico. Por tanto, son mejores las teorías que explican más hechos conocidos, las que tienen menos hechos que las contradicen y, sobre todo, las que nos internan en lo desconocido haciendo predicciones no triviales y novedosas, sobre cuya base las podemos someter a prueba rigurosa. El carácter rígido del contraste hace que las teorías cuantitativas sean las mejores.

1.1.1 Explicación y teoría

El papel de la teoría es explicar, proporcionar una comprensión de fenómenos, leyes, principios y cualquier otro tipo de *hecho* por medio de postulados generales, mecanismos internos, entes hipotéticos, procesos subyacentes o cualquier otro artificio intelectual; los que se combinan entre sí para proporcionar una estructura que dé cuenta racional de aquello que se pretende explicar. Es decir, las teorías tratan de dar sentido a aquello que explican, ubicándolo en la naturaleza y haciendo explícitas sus propiedades y sus relaciones con otros entes.

El propósito de la explicación es profundizar en la comprensión de los fenómenos. Por ejemplo, en química, Robert Boyle había desarrollado la distinción taxonómica entre elementos y compuestos; Proust elaboró, a partir de esa base, la ley empírica de las proporciones constantes, que propone que los compuestos tienen una combinación de elementos que guardan una proporción de peso y permite relacionarlos por medio de números enteros. Esta ley empírica era suficiente para manejar coherentemente muchos fenómenos de la química. Sin embargo, Dalton, un modesto profesor inglés, introdujo una de las mejores propuestas en la historia de la humanidad: explicó esas regularidades numéricas suponiendo que la materia es discontinua, y retomando la idea de Leucipo y Demócrito postuló la existencia de átomos para explicar esos hechos. No ha

[1] Los justificacionistas son los investigadores que creen que es posible justificar una afirmación teórica mediante pruebas empíricas.

existido una propuesta más fructífera que ésta (Mercado, 1978). Al poco tiempo, no sólo daba cuenta de los fenómenos conocidos de la química, sino que asimiló la ley de Boyle-Mariott de los gases a la explicación atómica, mediante la teoría cinética de los gases, que se basó en una aplicación de la mecánica newtoniana a las partículas atómicas y a las combinaciones de éstas, las moléculas.

Como podemos observar, las teorías son poderosísimos instrumentos intelectuales que nos permiten dar sentido a la apabullante complejidad de la experiencia fenoménica, así como lidiar con la realidad por medio de la creación de un esquema conceptual de ésta, que suponemos que es así en verdad.

1.1.2 Naturaleza de la investigación

La investigación se considera la parte creativa de la ciencia, con la que se busca expandir el conocimiento y la comprensión de la realidad, así como la elaboración del esquema que guíe nuestro objetivo. Los esquemas que proporciona la ciencia no son sólo descriptivos del mundo circundante, sino conceptuales-causales, es decir, son guías en relación con las clases de objetos y eventos y sus conexiones causales recíprocas. Así, en función de esta situación vamos avanzando en el dominio cognoscitivo de la realidad.

La naturaleza de la ciencia y, por ende, la de la investigación han sido explicadas mediante la rama de la filosofía llamada *filosofía de la ciencia.*[2] Esta disciplina es un esfuerzo del razonamiento humano por comprender cuál es el fundamento de esa actividad tan exitosa que denominamos *ciencia.* La filosofía, entendida como la reflexión sobre la naturaleza última de la realidad y de nuestra existencia, lleva a un razonamiento acerca de la relación cognoscitiva existente entre nosotros y la realidad, que es la rama denominada *epistemología.* Dentro de esa reflexión se encuentra ubicado un análisis más específico del proceso de adquirir conocimiento por medio de la ciencia.

La ciencia, como tal, surge en forma sistemática y organizada entre los griegos. Se desarrolla desde el año 600 a.C., en las mentes inquietas e inquisitivas de investigadores de la naturaleza y de filósofos que buscaban la esencia de la realidad, incluida la naturaleza del conocimiento que tenemos de ella; desde la filosofía de la ciencia de Demócrito, Platón, Aristóteles, etc., hasta las contribuciones empíricas y teóricas concretas de Anaxágoras, Aristarco y Arquímedes, entre otros.

Sin embargo, no fue sino hasta que se inicia el Renacimiento cuando surge de nuevo un concepto sistemático del proceder científico para el avance del conocimiento, es decir, una búsqueda activa de la verdad a través de la experiencia y la puesta a prueba empírica de las hipótesis, siendo un hecho que casi todo lo que distingue al mundo moderno de los siglos anteriores es atribuible a la ciencia. Ésta, como práctica, surge al lado de y bajo el cobijo de la filosofía

[2] Se ha llegado al estudio de la naturaleza del conocimiento por una variedad de ramas de la filosofía y de las ciencias particulares, denominadas *epistemología, filosofía de la ciencia* y *metodología.* El carácter va de lo más general en la epistemología, a lo más específico en la metodología.

empirista. Cuatro astrónomos preeminentes en la creación de la ciencia, Copérnico, Kepler, Galileo y Newton, además físicos los dos últimos, dieron el impulso al surgimiento de ésta al ayudar a abrir el camino a la investigación crítica como medio para avanzar en el conocimiento, lo que obtuvo sus logros más espectaculares en el siglo XVII (Russell, 1967).

Junto con los practicantes de la ciencia como método empírico para abordar el conocimiento, surgían los filósofos empiristas, quienes fundamentan el nuevo método de obtener conocimiento. Bacon, Hobbes, Locke, Berkeley y Hume instituyen el *empirismo* como el único camino al conocimiento, al establecer a la experiencia empírica como la única posibilidad para conocer la verdad y a la inducción como el método lógico que hacía posible esto al usar la inferencia para lograr los conocimientos generales a partir de experiencias particulares. Ellos establecieron al conocimiento científico como un camino seguro a la verdad. Intentaban desarrollar un sistema de inferencia racional que hiciese posible la generalización a partir de experiencias particulares y concretas. Suponían también un carácter acumulativo de la ciencia; para ellos los hechos son contactos objetivos con el mundo que, una vez establecidos, quedan perennemente en el acervo de conocimiento verdadero, siendo la ciencia un proceso de acumulación de hechos. En pocas palabras, la concepción de la ciencia se desarrolla como una búsqueda en la experiencia empírica de un camino para una seguridad absoluta que justificase los conocimientos así desarrollados como productos permanentes de un método fehaciente.

Comte dio el siguiente paso en el desarrollo de una concepción de la ciencia. El desarrollo del *positivismo* clásico fue un avance en la concepción de la ciencia empírica y de un sistema metodológico para su ejercicio concreto.[3] El positivismo considera a la experiencia empírica fuente de conocimiento, y los hechos generales o leyes son la única fuente de *certidumbre*. Encontramos a los pensadores Mach, Avenarius, Poincaré y Pearson, entre otros, como estructuradores de una filosofía que establecía a la ciencia sobre una base empírica que se proponía como una guía pragmática para enfrentar la vida. El Universo, incluyéndonos a nosotros, estaría constituido de fenómenos que se conectan causalmente entre sí, conexiones que se podrían descubrir por medio de la inducción controlada, en la medida de lo posible, por el método experimental. Las leyes y las teorías serían símbolos convencionales que reflejarían el orden en las relaciones dentro de la naturaleza.

Tanto el positivismo clásico como el empirismo mantienen una posición radical acerca del conocimiento. El conocimiento putativo no puede considerarse como verdadero a menos de que se le pruebe, y la prueba consiste en poner a este conocimiento bajo la hegemonía de la autoridad epistemológica pertinente, en este caso la experiencia empírica (Weimer, 1979).

En la actualidad encontramos que el trabajo de filósofos con enfoques diferentes, aunque con un núcleo central de acuerdo fundamental, culmina el

[3] Comte fue además el padre de la sociología, que desarrolla dentro del marco filosófico de su *método positivista* de hacer ciencia.

desarrollo de una filosofía de la ciencia empírica. Todos ellos usan la lógica y la lingüística como instrumentos para el desarrollo de una relación entre teoría y realidad, aunque el fundamento de la verdad empírica sigue siendo el criterio epistemológico último. Wittgenstein, Ayer, Carnap, Tarsky y Feigel desde el positivismo lógico; Russell y Whitehead desde una combinación de realismo crítico y filosofía analítica, y Moore, Wittgenstein y Wisdom desde la filosofía analítica abordan la búsqueda de la verdad mediante variantes de un mismo esquema fundamental. Si la inferencia no puede demostrar su validez *absoluta* como método lógico para establecer conocimiento verdadero, es decir, no se le puede probar, el concepto de inducción se sustituye por uno de inducción probabilística. Se fusionan los conceptos de inducción y probabilidad, y es necesario probar el conocimiento en términos de probabilidades.

Este punto de vista de la ciencia prevaleció sin desafío hasta nuestro siglo, pero en la actualidad ha surgido con gran vigor la perspectiva de la ciencia que hemos mencionado, llamada *no justificacionista*, que analiza el proceso de conocimiento científico sin recurrir al de la *justificación empírica* como base para el establecimiento de éste. Como ya vimos, autores como Popper, Kuhn, Lakatos, Feyerabend y Weimer han desempeñado un papel muy importante para dar esa visión alternativa de la ciencia.

La visión de la investigación científica desarrollada por las filosofías empírica y positivista fue relativamente clara. Existen dos tipos de entes: los hechos y las teorías. Los primeros provienen del ingreso sensorial, mientras que las segundas son conjuntos de proposiciones que surgen de los hechos a partir de la inducción. El problema es sencillo: hay que probar las teorías asegurando que sus conceptos tengan una relación unívoca con los hechos establecidos por inducción.

Weimer (1979) llama *justificacionismo*[4] al denominador común de todas estas aproximaciones, porque encuentra a la metateoría como la concepción de que hay una fuente de autoridad que produce una justificación incontrovertible para un método. En esto, afirma que tanto el racionalismo como el empirismo-positivismo parten de una misma posición fundamental, de lo que Dewey llamó *la búsqueda de la certeza*. El racionalismo lo hace apelando a la autoridad del intelecto, mientras que el empirismo-positivismo a la del ingreso sensorial.

Popper (1974) señala que es esa búsqueda de una base firme e incontrovertible la fuente de los problemas. Hace un análisis sobre la reflexión de Hume (1927, 1960) acerca de la inducción, y coincide con él en que no es posible que partiendo de la observación de una serie de casos reiterados de una relación determinada se llegue a una conclusión válida acerca de casos aún no observados; es decir, no se justifica desde el punto de vista lógico la inferencia. La solución que ofrece para no caer en un solipsismo estéril es que, si bien no podemos de modo alguno comprobar teorías, podemos refutarlas. Su solución para el funcionamiento de la ciencia se puede resumir en la idea de que la ciencia opera sobre la base de conjeturas que se someten a una prueba rigurosa ante la evidencia empírica y ante el análisis de la consistencia

[4] El *no justificacionismo* se inicia propiamente a partir del trabajo seminal de Popper y Kuhn, quienes hacen una crítica devastadora del positivismo lógico desde el interior de éste.

lógica. En esta perspectiva no justificacionista, la teoría no surge directamente de los datos a partir de un proceso de inducción, ya que cualquier proposición teórica, desde una simple ley empírica hasta un modelo teórico o una teoría, es una conjetura. El origen puede ser, como se señala anteriormente, cualquier posible fuente: la observación de una o varias regularidades, una especulación teórica, una analogía o algún otro proceso. Lo importante es que las conjeturas científicas se ponen a prueba por medio de la crítica lógica y la empírica (a diferencia de las conjeturas puramente especulativas en otros ámbitos). Sin embargo, si los hechos apoyan la teoría, no podemos pensar que la justifican, sólo que hasta ahora no la han refutado.

1.1.3 Justificación contra confrontación

De acuerdo con Lakatos (1970), un programa de investigación se juzga a partir de su comportamiento comparado con programas rivales.

La conciencia de nuevas variables extrañas generalmente se da en torno a la competencia entre teorías rivales; no nos damos cuenta de qué debemos controlar hasta que otra explicación sugiere los aspectos que hay que controlar cuidadosamente para decidir cuál explicación es la que mejor da cuenta de los hechos. Lakatos (1968*b*) asevera que no es tan importante el choque entre teoría y datos como la competencia entre las teorías rivales. La actitud rigurosa no implica el desecho instantáneo de una teoría, sino la exploración seria y crítica de sus posibilidades frente a otras opciones de explicación. Tal como señala Weimer: *En la mayoría de los casos en la práctica científica actual, el medio más efectivo de crítica disponible a un investigador es permanecer comprometido con una posición para poder articularla plenamente y explorar sus consecuencias* (1979, p. 49).

1.1.4 ¿De dónde surgen las teorías?

Como hemos visto, las teorías son intentos de explicación de la realidad, confrontadas con los hechos de manera rigurosa y éstas compiten entre sí para tratar de encontrar la manera de dar cuenta de los hechos. Las teorías son sistemas de creencias acerca del mundo, más explícitos, claros y precisos que otros conjuntos de creencias y son sometidos a una rigurosa prueba sistemática.

Las teorías pueden tener una génesis muy diversa. Por una parte se encuentra el conocimiento tácito que tenemos de muchos aspectos de la realidad, donde el sentido común y el conocimiento personal son una fuente muy importante de hipótesis científicas (Polanyi, 1958). En la vida cotidiana observamos casualmente muchos hechos que después llevamos al laboratorio y examinamos con más cuidado. Esas mismas observaciones con frecuencia inspiran nuestros primeros intentos de explicación, que al desarrollarse pueden ser la base de una teoría. Otra fuente común son los accidentes en el proceso de investigación, que nos llevan a encontrar lo que buscamos y se le ha llamado

serendipity. En otras ocasiones las teorías surgen de una observación cuidadosa de los hechos, tal vez experimentales, y el desarrollo de una *inferencia* a partir de ellos. Otro origen frecuente de las teorías es la observación de una discrepancia entre algunos hechos y una teoría. Esto puede llevar a una reflexión que dé lugar al desarrollo de una teoría alternativa que resuelva el conflicto.

1.2 RELACIONES ENTRE ESTADÍSTICA E INVESTIGACIÓN

En este capítulo examinaremos el papel que tiene la estadística en la investigación científica. La estadística en realidad es un instrumento muy valioso para organizar la información científica y para tomar decisiones acerca de ella, pues sería imposible concebir la investigación científica moderna sin la estadística.

La investigación, con muy raras excepciones, se refiere a grupos de datos e incluso a grupos de objetos, plantas, animales o personas. Un investigador en astronomía puede tomar varios registros de la distancia a la que se encuentra la Luna o algún objeto lejano con una técnica específica (por ejemplo, usando radar) para controlar el error de medida, y luego usar la estadística para decidir si su nueva medición es igual o diferente que la que obtuvo usando un método más primitivo. Un psicólogo puede medir la ejecución de una tarea por tres grupos de sujetos en un experimento, que difieran en la cantidad de alcohol que han ingerido, para ver el efecto sobre una tarea consistente en colocar palitos en agujeros hechos en una tabla. En este caso es posible usar la estadística para establecer si hay diferencias entre esos grupos de sujetos.

1.2.1 Error de medida y experimental

Existen dos conceptos de gran importancia en los que la estadística tiene un papel preponderante: los errores de medida y los experimentales. Ambos son importantes fuentes de problemas para el investigador y poderosas razones para utilizar la estadística en la investigación.

El error de medida es el que cometemos al medir cualquier cosa, a pesar del cuidado que tengamos. Por una variedad de razones podemos cometer dos tipos de error: el sistemático, que implica una falla regular en una dirección (por ejemplo, un metro un poco más grande de lo debido) o el error aleatorio, que se refiere a inexactitudes de un instrumento al medir con él. El primer tipo de error provoca distorsiones en nuestros datos que implican un error en las conclusiones.

Los errores sistemáticos pueden radicar en fallas en la calibración o estandarización de los instrumentos de medición; éstos deben ser comparados con un estándar, el cual determina que el instrumento arroje los valores adecuados a la escala que se está usando. Por ejemplo, el metro tiene como estándar de calibración una varilla de platino-iridio, colocada sobre un soporte especial en una cámara con temperatura y ambiente controlados que se encuentra en la Oficina de Pesos y Medidas en París, Francia. Los estándares de calibración de

los diversos países se obtienen marcando otra varilla similar en sitios análogos a los de la varilla estándar y conservándolos en condiciones similares. Los instrumentos psicométricos (los *tests*) se estandarizan (una forma de calibración) aplicándolos a una gran parte de la población y luego se establecen calificaciones estándar. Por ejemplo, si se hace un *test* para probar la inteligencia en México y se usan estándares ingleses o argentinos, se estaría produciendo un error sistemático de medida.

Los errores sistemáticos también los provoca la influencia de alguna variable ajena que afecta el proceso de medición, por ejemplo, la presencia de un campo electromagnético cerca de un instrumento de medición, una aguja de bobina o lo que pudiera ser un sonómetro.

Los errores aleatorios (al azar) son los que cometemos accidentalmente, como limitaciones perceptuales, inexactitud en el momento de tomar una medida, cometer un error al leer una escala, leer de manera distraída, y también se deben a la influencia accidental, de carácter temporal, de otras variables, como el estado de ánimo de un sujeto al someterse a un *test*, las variaciones accidentales de la corriente eléctrica al medir con equipo electrónico que use la energía de la red eléctrica, o el efecto de la temperatura en el funcionamiento de un equipo.

La estadística nos permite lidiar con ambos tipos de error. El error sistemático se establece viendo si un grupo de medidas difiere de un estándar bien establecido, por ejemplo, verificar si los metros que se usan en Polonia difieren del metro en la Oficina de Pesos y Medidas en París. Para esto se usan ciertas formas de estadística inferencial.

El error aleatorio se anula usando estadística. Nosotros podemos comparar medidas con error y estimar el valor casi exacto de cierta medida gracias a la estadística.

1.3 MEDICIÓN Y ESTADÍSTICA

La estadística se aplica sobre medidas obtenidas de los diversos objetos de estudio en diferentes condiciones. Por ejemplo, si queremos verificar si un curso de capacitación para soluciones de problemas mejora la inteligencia de los que lo cursaron, podemos tener un grupo al cual le medimos la inteligencia antes y después de llevar el citado curso; es decir, la aplicamos sobre medidas tomadas de los eventos.

Medir, según Torgerson (1958), es asignar números a una propiedad de acuerdo con una regla. Es decir, medir es una forma particular de observación en la cual asignamos números a las propiedades que observamos. Es de notarse que esta asignación no es del todo arbitraria, ya que usamos una regla de asignación de números a los valores de la propiedad.

Nosotros estamos sin duda familiarizados con numerosas formas de medir, por ejemplo, al asignar una ración de alimento para animales de laboratorio. La regla de pesar consiste en comparar el peso del objeto de nuestro interés con el de un estándar. Nosotros asignamos el número (el valor) de acuerdo con la regla de que el peso de nuestro objeto sea igual o un múltiplo del peso del estándar.

Las balanzas nos permiten verificar esto de modo más directo, porque una varilla suspendida horizontalmente por el centro de un postecillo nos indica que se encuentra equilibrada y si colgamos en los extremos unos platillos de igual peso el equilibrio no se altera. En esta balanza ponemos el objeto que queremos pesar y del otro lado el estándar. Por ejemplo, si queremos pesar harina y asegurarnos de que tenemos un kilo, colocamos en uno de los platillos el estándar de un kilo y en el otro la harina. Si el equilibrio se mantiene, entonces tenemos el peso deseado. Si no fuese así, tendríamos que agregar o quitar harina hasta lograr el equilibrio, o podemos cambiar o combinar estándares.

Las básculas modernas tienen un plato de un lado, suspendido sobre el brazo de la báscula, y del otro lado un brazo sobre el cual corre un peso estándar; el efecto del peso varía al correr el estándar sobre el brazo de la palanca.

Otro uso de la estadística en psicología y ciencias afines es el desarrollo de modelos psicométricos. Estos modelos se basan en una teoría que plantea que la respuesta a un problema, una pregunta o algo similar depende de diversas variables. Si seleccionamos una de esas variables para medirla, también podemos escoger varios reactivos que supuestamente la midan, constituyendo una prueba o *test* con ellos. Usando estadísticas como la correlación y el análisis factorial, podemos ver cuán bien funciona cada reactivo (pregunta, etc.) en relación con la prueba y mejorar de modo que obtengamos una medida precisa y que efectivamente mida dicha variable.

Si bien entrar en detalles en cuanto a la teoría psicométrica está fuera del alcance de este libro, esto nos da una idea de la importancia de aprender estadística para poder después aplicarla en la psicometría.

1.3.1 Escalas de medición

Medir es asignar números a propiedades de un objeto de acuerdo con reglas, pero las reglas que podemos usar son de muy diferentes tipos. Al asignar números aprovecharemos las propiedades de los sistemas numéricos. Stevens (1951) definió cuatro tipos de escalas de acuerdo con las propiedades del sistema numérico que se aprovechan por la regla que se usa para la asignación.

El primer tipo, llamado *escala nominal*, emplea los números solamente como nombres de clases de objeto. Éste sería el caso de usar 0 para sexo femenino y 1 para masculino (o viceversa) o usar números diferentes para las personas que escogen distintos tipos de cereal: 1 para los de Corn flakes, 2 para Dulcereal, etcétera.

El segundo tipo, llamado *escala ordinal*, asigna los números de acuerdo con la propiedad ordinal del sistema numérico: los valores están ordenados de menos a más, pero no hay una idea de igualdad en las distancias entre los números. La regla de correspondencia permite entonces asignar los valores numéricos a una propiedad del objeto de estudio de modo que reflejen niveles crecientes de esa propiedad, sin que haya un compromiso de que las distancias en esa propiedad sean iguales. Por ejemplo, en una escala de actitudes podemos asignar números: 1, 2, 3..., a los valores de una actitud. Si decimos: "indique usted el aprecio que tiene por el Presidente de la República: 1. ninguno; 2. poco;

3. regular, y 4. mucho", en esta escala no podemos decir que la distancia en aprecio entre el que responde 1 y el que responde 2 es igual a la que hay entre 3 y 4, pero sí apreciar que el valor 4 es mayor que el 3 en esa dimensión.

El tercer tipo, denominado *escala de intervalo*, no sólo se usa en el ordenamiento, sino que establece que las distancias que hay entre número y número son iguales. Por ejemplo, las temperaturas tomadas con termómetros nos permiten aseverar que la cantidad de incremento de temperatura es igual para distancias iguales en la escala. Por ejemplo, un incremento de 5°C es igual, ya sea cuando se pasa de 0 a 5° o cuando se pasa de 10 a 15°.

En el último nivel de escala, la de *razón*, se usan las propiedades anteriores pero, además, se tiene un cero que refleja la ausencia de la cualidad. Por ejemplo, en el caso de la temperatura que vimos anteriormente, las escalas hacen referencia a un cero que es arbitrario y no refleja la ausencia de la propiedad que se mide (en este caso la temperatura). El cero, en la escala Celsius, es el punto en que el hielo se derrite (o congela). En la escala Fahrenheit, la referencia es el alcohol en vez del agua. Ambos son ceros arbitrarios, por eso las escalas generan números negativos. La escala Kelvin, por el contrario, sí hace referencia a un cero absoluto que implica la ausencia total de movimiento molecular y, por tanto, de temperatura.

Así, como vemos, los diferentes tipos de escalas emplean algunas propiedades de los sistemas numéricos para generar un tipo de medidas que reflejen ciertas propiedades de la dimensión que se pretende reflejar con esa medida. Las escalas nominales, por ejemplo, sirven para medir cosas que tienen que ver con la pertenencia a grupos u otras formas de clasificar las cosas o a las personas. En este caso, los números sólo sirven como nombres y es indistinto el orden que se use. Aquí sólo utilizamos la propiedad de identidad de los números.

Las escalas ordinales usan la propiedad ordinal, esto es, el hecho de que se siga una secuencia. De este modo, sabemos que el 2 es mayor que el 1 o que el 11 es mayor que el 9, sin que eso implique que la distancia entre 9 y 11 tenga que ser mayor que entre 1 y 2; sólo se toma en cuenta el orden.

Las escalas intervalares usan la distancia entre números como algo válido, de manera que la distancia entre 3 y 5 es igual a la distancia entre 7 y 9, pero no hacen referencia a un cero absoluto, de modo que no podemos decir que 8 es el doble de 4.

Las escalas de razón usan todas las propiedades de los números: identidad, orden, igualdad de las distancias y referencia a un cero.

1.3.2 Limitación de la estadística por nivel de medida

El uso de la estadística se ve limitado por el tipo de medidas que usamos. Por ejemplo, las de razón y de intervalo utilizan los modelos más poderosos, llamados *paramétricos*, que emplean la teoría de la curva normal de distribución. Este modelo nos permite llegar a conclusiones más válidas y precisas. Existen otros modelos que se aplican a los casos de las medidas ordinales y nominales y se

les denomina *no paramétricos*. Hay modelos que usan las propiedades de orden como Kolmogorov-Smirnov o la U de Mann-Whitney y otras como la χ (ji) cuadrada, que se utilizan para analizar términos de la probabilidad de clases de eventos. Estos modelos se verán más adelante con todo detalle; lo importante es percatarse que el tipo de medidas que usamos determinan el tipo de estadística.

1.4 INFERENCIA ESTADÍSTICA Y CIENTÍFICA

La estadística funciona para hacer inferencias de las distribuciones de las medidas de los fenómenos. Esta parte de la suposición de que varias muestras pertenecen a la misma población. Cuando la población a la que pertenecen las muestras difiere, tiene un efecto que se refleja en las muestras.

Para entender mejor esto aclaremos qué entendemos por población. La estadística usa la distribución de probabilidad de los estadísticos de las muestras (media, desviación estándar, varianza, etc.). Por ejemplo, la media, que se verá en el capítulo 2, es una medida global que identifica a un grupo de medidas: en el valor, en el punto central o en el equilibrio y, por tanto, representa al grupo. Las medias de muestras varían entre sí, en el tamaño. La frecuencia de estas medias se distribuye de acuerdo con una forma (función) que se denomina *t de Student*. Como vemos, esta función es más alta donde se encuentra la verdadera media o disminuye a medida que se aleja. Cuando tenemos la media de dos muestras podemos calcular la probabilidad de que pertenezcan al mismo grupo usando esta función. Mientras más alejadas se encuentren, es menos probable que pertenezcan al mismo grupo.

La inferencia estadística se basa en llegar a una conclusión a partir de una probabilidad de que las medias de dos o más grupos pertenezcan a la misma población. Si la probabilidad es lo suficientemente baja se concluye que las muestras no pertenecen a la misma población; por tanto, la razón por la cual difiere en los grupos genera diferentes poblaciones en esa medida.

Por ejemplo, si suponemos que el alcohol afecta la comprensión de un texto, podemos usar una medida del grado de comprensión que tiene un sujeto de cierto texto estándar. Esta medida puede obtenerse mediante una serie de preguntas acerca del texto (que deberán ser tratadas psicométricamente). Ahora, supongamos que formamos tres grupos de estudiantes de psicología y a uno no le damos nada de alcohol, al segundo sólo una copa de tequila a cada estudiante y al tercer grupo dos copas. Luego los ponemos a leer el texto y les proporcionamos el cuestionario con el que medimos la comprensión. Si los tres grupos provinieran de la misma población (de comprensión de dicho texto) por probabilidad las medias serían todas parecidas, pero si el alcohol tuvo un efecto en la comprensión de la lectura, estas medias diferirán. El grado de diferencia determina la probabilidad de que éstos pertenezcan a una población homogénea; y cuando la probabilidad es lo suficientemente baja, esto implica que la hipótesis alterna, que los grupos difieren entre sí, prevalece. Este tipo de inferencia, al igual que la no estadística de la que hablamos anteriormente, debemos

tomarla con la reserva debida. De la experiencia personal tal vez surgió la teoría de que el alcohol afecta la comprensión de textos. Esta teoría es puesta a prueba por el experimento y deberá hacerse con diferentes muestras y condiciones.

1.4.1 Diseño experimental

El diseño experimental es simplemente el plan de investigación. Se trata de un plan para hacer que varíe de la manera más amplia posible la variable, o las variables (variables independientes), de la cual nos interesa ver su efecto sobre otra u otras variables (variables dependientes) para establecer relaciones causales o al menos funcionales. Los experimentos están diseñados para poner a prueba rigurosa las hipótesis de investigación, las cuales se derivan de los diferentes planteamientos teóricos. De esta manera, variamos aquello de lo que queremos observar su efecto sobre algo más y controlamos, en la medida de lo posible, todo lo demás.

En las ciencias del comportamiento lo que interesa son las variables estímulo o de la situación, las variables de la conducta y las variables relacionadas con los procesos internos.

El desarrollo actual de la tecnología ha hecho posible medir y controlar aspectos muy refinados de nuestros objetos de estudio. Aunque en la época de Galileo ya se tenían estas nociones acerca del diseño, no era posible observar muchas cosas porque no se contaba con el desarrollo científico y la consecuente tecnología para observar, medir y controlar muchos factores. De esta manera, la ciencia, mediante su propio desarrollo, genera métodos para producir y controlar los diferentes aspectos (variables) que son de su interés, potenciándose a sí misma.

En general, un experimento trata de:

1. Observar y medir lo más exactamente posible las variables dependientes, es decir, aquellas sobre las cuales queremos ver si hay un efecto causal de las variables independientes.
2. Modificar amplia y sistemáticamente las variables independientes o causales, para ver si éstas afectan el fenómeno tal como se plantea en la hipótesis.
3. Controlar las variables extrañas, es decir, las variables que no entran en la hipótesis de investigación, pero que de algún modo podrían influir en los resultados, distorsionándolos. Estas variables son de tres tipos: *a*) La varianza (variación) de error, debida a falla de las medidas, la cual se corrige mejorando las medidas o el proceso de medición; *b*) Variables extrañas controlables, las cuales se controlan llevando a las variables a un estado constante, que no afecte al fenómeno, y *c*) Las variables intrínsecas al sujeto, no susceptibles de control, las cuales se controlan asignando los sujetos al azar a cada situación o usándolos como su propio control, es decir, que el mismo sujeto pase por todas las condiciones experimentales.

Existen diseños más o menos estándar, productos del ingenio y la dedicación de muchas generaciones de investigadores, lo que hace que generalmente no tengamos que inventar nuevos diseños para lograr buenos efectos y control en las investigaciones. Aquí sólo mencionamos el hecho, por lo cual el lector deberá consultar un texto sobre diseño experimental para mayores detalles.

Sin embargo, señalaremos algunos de los diseños experimentales más comunes, que tendrán características diferentes según el nivel de medición que estemos aplicando, tanto a las variables dependientes como a las independientes. El más simple y básico sería el diseño de dos grupos: experimental y control. Este diseño tiene en un grupo, el experimental, una condición que suponemos afecta al proceso y el segundo grupo, el control, carece de esa condición para dar un parámetro de comparación.

Otro diseño más refinado sería el llamado *paramétrico,* o de k grupos. En este caso tendremos un número k de condiciones, tal que $k > 2$ en lugar de tener dos. Generalmente una de las condiciones muestra la ausencia de la variable, sirviendo de grupo control.

Otro diseño muy popular es el *factorial,* donde el sujeto es puesto a condiciones con más de una variable. En ese caso, en lugar de un vector (una hilera de condiciones) con k grupos tenemos una matriz, es decir, un cuadro, un cubo, etc., donde cada dimensión corresponde a una variable y cada cruce corresponde a una cierta combinación de variables. En realidad, el diseño factorial es tan sólo un plan sistemático para producir todas las combinaciones posibles de una serie de vectores.

La estadística nos permitirá obtener resultados en todos los casos, pues ayuda a discernir si las diferencias encontradas se deben al azar, causadas por las variaciones naturales de los grupos, o son debidas al efecto de la variable de interés, la que estamos manipulando.

1.4.2 Diseño cuasiexperimental

Hay ocasiones en que no podemos controlar adecuadamente algunas variables, por ejemplo, en un estudio sobre educación tenemos que tomar los grupos naturales y esto impide la asignación al azar. En estos casos la estadística viene al rescate, pues permite tomar en cuenta el posible efecto de esas variables no controladas. Existen dos métodos experimentales: uno es el análisis de covarianza, que requiere que se midan las variables extrañas potenciales y se introduzcan al modelo estadístico. En unos capítulos más adelante se verá cómo se logra esto. El otro método se refiere al uso de series temporales para extraer la varianza y las relaciones de los fenómenos en el tiempo. Éstos son métodos estadísticos más allá del alcance de este libro, pero los mencionamos para dar una idea general.

Entonces, en los métodos cuasiexperimentales se tienen los mismos elementos que en los experimentales, es decir, maximizar la variación de la variable o variables independientes y controlar las extrañas, pero sólo parte del control es experimental;

el otro es estadístico de las variables extrañas. Estos métodos son más adecuados para estudios en condiciones naturales.

1.5 ESTADÍSTICA E INFORME CIENTÍFICO

El informe es el acto de escribir los resultados de una investigación con el objeto de darlos a conocer, es decir, que se publiquen. El informe tiene estándares técnicos para su organización y existen manuales de redacción, normativos tanto de la estructura como del estilo. Uno muy conocido es el *Manual de la APA* (American Psychological Association), que es ya un estándar internacional. Se trata de que el informe esté ordenado, completo y bien organizado, para que el lector no sólo se dé cuenta de los resultados, sino de sus implicaciones, el modo como se hicieron las cosas y qué se tendría que hacer para reproducir el estudio.

La estadística desempeña un papel, por supuesto, al informar los resultados. Allí, deberán mostrarse cuadros y gráficas, así como describir verbalmente lo que se obtuvo (sin interpretar los resultados, lo que viene más adelante, en la discusión y las conclusiones). Es importante mostrar los datos y señalar qué diferencias fueron significativas estadísticamente.

Si no queremos leer sólo la información repetida de los libros de texto, sino también las investigaciones recientes, tenemos que consultar los artículos de las revistas especializadas. Esto es muy importante si uno quiere mantenerse al día en un campo, pues la información tarda entre tres y 10 años en llegar a los libros. Para poder leer estos informes y comprenderlos, tenemos que entender la estadística que usó el autor y qué significa; sólo así podremos seguir sus argumentos.

1.5.1 Gráficas

Las gráficas son un modo muy eficiente de mostrar resultados. Generalmente los datos se muestran tanto en tablas, donde tenemos los números exactos, como en gráficas, las cuales nos permiten percatarnos mejor de la forma de los datos y del patrón que se da en ellos.

En los capítulos 2 y 6 se analiza el tipo de gráficas y sus usos.

RESUMEN

La ciencia es una actividad muy compleja, por medio de la cual tratamos de entender la realidad. El objetivo de la ciencia es lograr teorías poderosas que le den sentido a esa realidad y la expliquen, lo cual, a veces, lleva a mejorar la capacidad para predecir y controlar los eventos en ella.

La ciencia busca establecer hechos para determinar cuán verdaderas son las teorías, es decir, ponerlas a prueba. Generalmente esto se da en el contexto de la competencia entre teorías para explicar un ámbito de la realidad. La coherencia lógica interna de las teorías, así como su congruencia con otras teorías y con los

hechos son lo que determina su utilidad. También, la capacidad de las teorías para aventurarnos exitosamente en zonas desconocidas de la realidad, es otro factor importante para la evaluación de las teorías.

Los experimentos son una manera rigurosa de establecer los hechos. La estadística ayuda a la experimentación a obtener conclusiones útiles y claras, por tanto, existe una relación muy íntima entre estadística e investigación, pues la estadística permite lograr conclusiones claras y válidas. Aun para entender los informes de investigación es indispensable saber estadística.

En este capítulo se muestra la importancia de la estadística para el desarrollo científico y profesional de los psicólogos, pedagogos, sociólogos, antropólogos y otros en las áreas de las ciencias del comportamiento.

Cuestionario

1. ¿Qué es la ciencia?
2. ¿Qué es una teoría?
3. ¿Qué relación hay entre la teoría y los hechos?
4. ¿En qué difieren el justificacionismo y el no justificacionismo?
5. ¿Qué es la investigación?
6. ¿Qué son el empirismo y el positivismo?
7. ¿Qué metateoría sustenta al justificacionismo?
8. ¿Cómo concibe el no justificacionismo a la teoría?
9. ¿Cómo se confrontan las teorías rivales?
10. ¿De dónde surgen las teorías?
11. ¿Qué es *serendipity*?
12. ¿Qué son los errores de medida y experimental y qué tienen que ver con la estadística?
13. ¿Qué es medir?
14. ¿Cuáles son los niveles de medición y cuál es su papel en la estadística?
15. ¿Qué relación existe entre la inferencia estadística y la inferencia científica?
16. ¿Qué papel desempeñan el diseño experimental y el cuasiexperimental?
17. ¿Qué papel desempeñan la estadística y las gráficas en el informe científico?

UN MODELO DE INVESTIGACIÓN

Dr. Adip Sabag Sabag

Propósitos

El propósito central en esta parte es que el lector sea capaz de realizar un estudio, encuesta o investigación, analizando y procesando los datos obtenidos mediante un instrumento de medición.

El lector:

1. Planteará el problema de investigación en forma clara, sencilla y precisa.
2. Seleccionará de manera adecuada la muestra que participará en dicha investigación.
3. Utilizará un cuestionario confiable, como instrumento de medición.
4. Aplicará el cuestionario en forma ética, supervisando las encuestas y no induciendo las respuestas en ninguna circunstancia.
5. Procesará los datos con la mayor precisión posible utilizando adecuadamente los medios disponibles, sean manuales o automáticos.
6. Presentará los resultados obtenidos, sin ninguna tendencia, sesgo o alteración.
7. Redactará un informe final, señalando los alcances y las limitaciones, las recomendaciones y las conclusiones de dicho estudio.

1.6 DATOS INICIALES

Al emprender una investigación conviene saber a dónde se quiere llegar, para determinar los métodos estadísticos que van a utilizarse y el instrumento de investigación que se va a diseñar. Es un diálogo entre el principio y el final; cuando hablamos de datos iniciales nos referimos a las informaciones que se tienen que especificar. En el caso de las investigaciones de opinión pública, particularmente las electorales, muchas de estas informaciones están contenidas en la ley, por tanto, deben darse a conocer en los resultados. En cualquier investigación, siempre se exige una nota metodológica.

Antes de iniciar una investigación se necesita responder a diversas cuestiones. Una investigación comienza, aunque parezca insólito, por saber *qué se desea investigar.* En muchas ocasiones se tiene una idea imprecisa, difusa, de lo que se quiere investigar. Aún más, en general, las primeras ideas no se sostienen al final. En la práctica un buen número de empresarios solicita investigaciones de mercado y cuando se les pregunta qué desean investigar, la respuesta es: ¡usted es el investigador! Plantear el problema no es tan fácil como parece. También suelen ocurrir situaciones donde el objetivo parece obvio y sin embargo, no es así. En una ocasión una empresa solicitó un estudio sobre tiendas de departamentos. La compañía investigadora realizó un estudio de hábitos de consumo, pero el cliente lo que buscaba era ¡dónde ubicar los establecimientos! Conviene tomarse todo el tiempo requerido hasta formular claramente las metas (fig. 1.1). Otra dificultad que aparece con relativa frecuencia al iniciar una investigación es la formulación de hipótesis. Las hipótesis particularmente se utilizan cuando al procesar los datos se van a usar métodos inferenciales, justamente las pruebas de hipótesis. Existen métodos estadísticos que no requieren hipótesis, como son los multidimensionales o bien, en estadística descriptiva, donde se estudia una población y una variable.

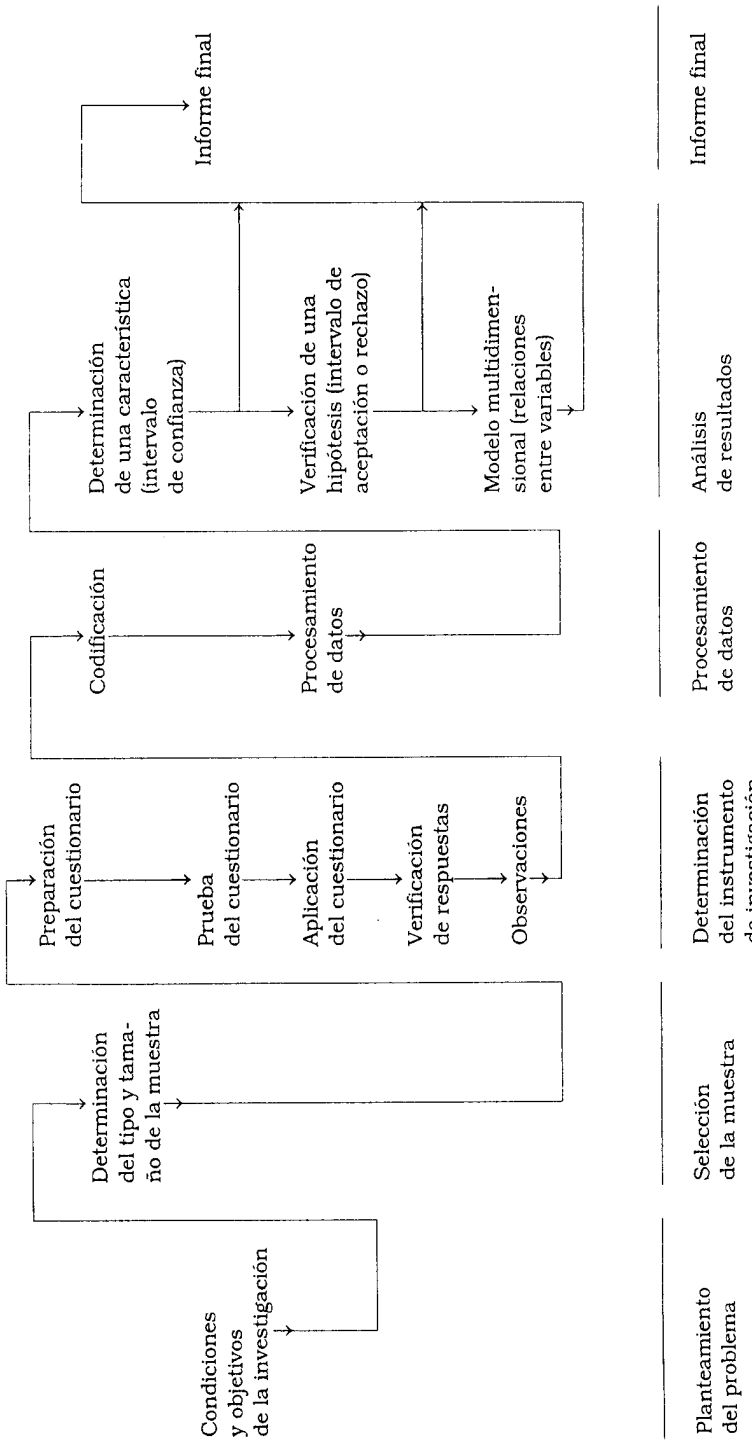


FIGURA 1.1 Etapas de una investigación

¿A qué sectores económicos, políticos o sociales hay que dirigirse? El costo de un estudio aumenta o disminuye de acuerdo con el sector al que se va a dirigir. Cuando se trata de una investigación de mercados, el cuestionario se dirige a una población de consumidores potenciales. Ahora bien, como las clases medias y altas son las que tienen las mayores posibilidades de compra, las encuestas se dirigen principalmente a este universo.

¿Qué regiones abarcará el estudio? Es muy importante especificar la zona donde se va a aplicar el estudio. Una investigación ni cuesta igual ni toma el mismo tiempo en el nivel local que en el nivel nacional. Los resultados se circunscribirán exclusivamente a la región donde se realizó la investigación.

¿De qué tiempo se dispone para el estudio? El tiempo del estudio es una información determinante y se refiere exclusivamente al periodo durante el cual se levanta la encuesta o se realiza la investigación. Es conveniente señalar que una investigación es como una fotografía: cuanto más breve sea la exposición, más alta será la definición de los resultados. Hay investigadores que presumen de haber tardado 10 o 15 años realizando la investigación; desafortunadamente, con el tiempo cambian las circunstancias y, por tanto, los resultados ya no reflejan la realidad estudiada.

¿Con qué recursos humanos, técnicos o económicos se cuenta? Contrario a lo que se piensa, en muchas investigaciones no se requiere una gran cantidad de dinero. Con frecuencia se recurre a infraestructuras ya establecidas y de las que se puede disponer con cierta facilidad. Buena parte de las investigaciones se hacen con estudiantes, a quienes se considera honestos y confiables, con una actitud positiva hacia la investigación. Cuando se trata de encuestas los estudiantes son una mano de obra abundante, confiable, eficiente y con un costo mínimo. Por otra parte, las universidades y otras instituciones disponen de computadoras en las que pueden procesarse los datos de manera confiable y a bajo costo.

¿Qué precisión se desea obtener? La parte más delicada del estudio es justamente la precisión que se desea obtener. Calcular los errores estadísticos es lo que nos hace científicos y nos remite al tipo y tamaño de la muestra. En la práctica, una buena investigación es la que tiene una seguridad entre 95 y 99%, una homogeneidad, cuando no se dispone de estudios previos, de 50/50 y un error aceptable entre 1 y 5%. Antes de iniciar una investigación es necesario decidir la precisión que se desea en el estudio. Para más información sobre muestreo, consúltese el capítulo 7.

1.7 TIPO Y TAMAÑO DE LA MUESTRA

La muestra más sencilla es la de testigos (o testimonios) privilegiados. Consiste en buscar exclusivamente ciertas características en la población (estudiantes, pasajeros de avión, tarjetahabientes, etc.). Se recomienda aplicar un mínimo de 250 encuestas para asegurar la tendencia en las respuestas. Las variables que distinguen a la población, en general, son las preguntas-filtro del cuestionario;

por ejemplo: ¿Tiene tarjeta de crédito? ¿Ha viajado en avión? ¿Escolaridad, edad?...

La muestra más utilizada es la de cuotas, que garantiza la representación de la población, utilizando los datos del censo o de otras fuentes documentales. Algunos investigadores consideran que una muestra de cuotas es equivalente en errores estadísticos a la mitad de una muestra aleatoria. Así, una muestra por cuotas de 800 personas equivale a una muestra aleatoria de 400 personas. Por ejemplo, para una investigación en Naucalpan, con personas mayores de 18 años y en la que se utilizó el método de cuotas, se recurrió a una segmentación a partir de los datos del *X Censo General de Población y Vivienda*, que proporciona los datos por edad y sexo (fig. 2). A partir de los datos el investigador decidió aplicar 1 350 encuestas, que se dividieron en los diferentes grupos de edad y sexo.

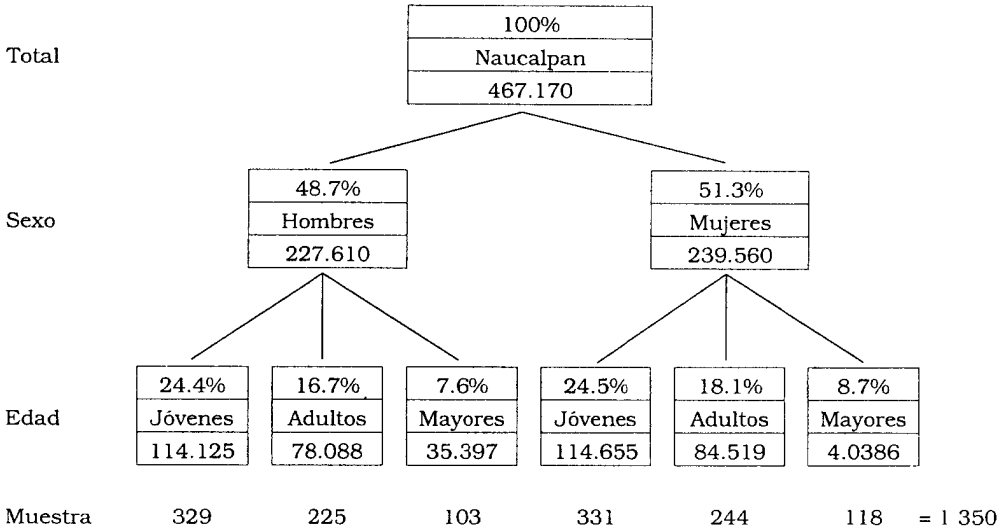


Figura 1.2 Muestra por cuotas (*X Censo General de Población y Vivienda*, INEGI).

La muestra aleatoria permite calcular los errores estadísticos a que está sujeta la investigación. La *seguridad* de una muestra se refiere al error de muestreo. Una seguridad de 95% quiere decir que, si tomamos cien muestras, habrá cinco donde los resultados se alejan de la realidad o, lo que es lo mismo, que hay 95% de probabilidades de tener razón. La *homogeneidad* se refiere al grado de conocimiento de una población. Se indica, por lo general, en términos de una proporción (p/q) y la proporción se expresa en porcentajes 50/50, 70/30 o 60/40. Siempre la suma dará cien. Cuando en una investigación no se tienen antecedentes de ningún tipo, la única proporción que podemos utilizar es 50/50, que es el caso más desfavorable cuando no se conoce nada de la población. Finalmente, el *error aceptable* indica el margen de error en los resultados. Así,

cuando en una encuesta se dice que un candidato obtendrá 53% de votos, éste no es el resultado verdadero. Éste se encontrará sumando y restando el error aceptable. Si es de 4%, quiere decir que el resultado verdadero se encontrará en el intervalo {49-57}.

La fórmula más sencilla para calcular una muestra aleatoria, cuando la seguridad es de 95% y la homogeneidad o proporción es de 50/50 (en estas condiciones se calculan prácticamente todas las muestras), es uno entre el error aceptable al cuadrado ($\frac{1}{e^2}$). Por ejemplo, si $S = 95\%$, $p = 50/50$, $e = 2\%$, el tamaño de la muestra será de ($\frac{1}{e^2} = 1/.02 \times .02$) = 2 500.

Este resultado es válido cuando la muestra es no exhaustiva, es decir, cuando la muestra no agotó al universo. El universo debe ser mayor o igual a siete veces la muestra. Por el contrario, cuando el universo es exhaustivo, es decir, la muestra es mayor que el universo, entonces se recurre a una fórmula de corrección $n' = Nn/(N + n)$, donde N es el universo y n la muestra no exhaustiva. Por ejemplo, si queremos una muestra donde $S = 95\%$, $p = 50/50$, $e = 2\%$, pero un universo de 300 personas, la muestra sería de 2 500, que agota al universo de 300 personas. Por tanto, recurrimos a la fórmula de corrección y nos queda: 2 500 × 300/(2 500 + 300) = 268. En vista del resultado, en lugar de utilizar una muestra es preferible levantar un censo, ya que se evitan los errores estadísticos y se cuenta con la mayor precisión.

Cuando se dispone de una lista de individuos y es necesario seleccionar una muestra aleatoria, en la práctica, las personas dividen el total entre el tamaño de la muestra y a intervalos regulares se van seleccionando los individuos. De esta manera se cree que los individuos fueron seleccionados al azar. Lo que conviene hacer es que el primer número se seleccione por sorteo, o utilizando una tabla de números aleatorios, y a partir del dígito que salió, la serie ya se aplica.

Errores frecuentes. Uno de ellos es pensar que cuanto más grande sea la muestra, mayor será la precisión de los resultados. En unas elecciones presidenciales en Estados Unidos de América, *Literary Digest* lanzó 10 millones de cuestionarios y se recibieron alrededor de dos millones. A pesar del número tan alto de entrevistados el margen de error fue de 18%. Para una muestra aleatoria con ese error habrían bastado 31 entrevistas individuales. Al mismo tiempo, otros institutos, utilizando el cálculo de probabilidades, entrevistaron 4 500 personas y acertaron con 1% de error. La confiabilidad de una muestra no depende de la cantidad de entrevistados, sino de la calidad de sus testigos. Hagámoslo más dramático: en un diagnóstico clínico, el examen de sangre requiere unas cuantas gotas, ¡no litros! Para saber que el agua del océano es salada, basta probar una cucharadita. Parafraseando el proverbio popular, "para conocer a Inés basta una vez".

Otro error frecuente, aun entre profesionales, es creer que una buena muestra es del 10% del universo. O también, según el tamaño, del 6 o 20%. Repetimos: la precisión de los resultados no depende del tamaño del universo, sino de la calidad de los individuos seleccionados.

1.8 INSTRUMENTO DE MEDICIÓN

Hay diferentes instrumentos de investigación: entrevista, observación y experimentación, entre otros. Sin embargo, el método más popular y accesible es el cuestionario. Presenta ventajas importantes: es el mismo para todas las personas; pueden hacerse estudios comparativos tanto longitudinales como transversales y correlativos con las propias variables y con otros estudios, y analizar estadísticamente los datos con relativa facilidad; además, es objeto de derechos de autor y su costo es bajo. Sin embargo, el diseño de un cuestionario es más bien un arte que una ciencia; por tanto, si se desea obtener un instrumento adecuado y sensible, se requiere cierta experiencia.

Un buen cuestionario es el que gana información, no el que la pierde. Por tanto, una pregunta debe ser formulada para este fin y evitar las que ya condensan información, como es el caso de grupos de edad, intervalos de salario, etc. También se considera que es mejor el cuestionario que formula preguntas indirectas. Las personas tienden a evitar y, por consiguiente, a mentir cuando se les interroga directamente.

Antes de iniciar el diseño de un cuestionario conviene considerar las características de las escalas estadísticas, ya que de la escala y el número de sujetos participantes dependerá el método estadístico por utilizar (*véase* la sección 1.3.1).

Según el marco teórico que se esté usando podemos clasificar las variables de un cuestionario en: 1. Control, pasivas, independientes o estructurales y 2. Experimentales, activas o dependientes. Las variables de control son las que estructuran una investigación, dan cuenta de la representatividad de la muestra y permiten construir el perfil de entrevistado, perfil al que se sujetan los resultados. Las variables de control más utilizadas son: edad, sexo, lugar de nacimiento, escolaridad, medio socioeconómico, estado civil, ocupación principal. Otros análisis requieren otras variables, como son: periódico que lee, y barómetros como presión social y preferencias electorales.

El *sexo* es una pregunta, que cuando hay un entrevistador de por medio, no se formula, sino que se observa, aunque pueda haber distintas respuestas, como: masculino, femenino, homosexual, lesbiana, bisexual o transexual. En México, las respuestas a estos últimos reactivos no se dan; por tanto, sólo se clasifican en hombre o mujer.

La *edad* es una pregunta que encierra muchos prejuicios, a tal grado que importantes investigaciones se desechan por haber preguntado la edad, ya que las personas tienden a mentir, a quitarse, aumentarse o redondear su edad y aparecen con frecuencia cifras terminadas en cero y en cinco. Por ejemplo, 30 años, 25 años... Para resolver este problema se recomienda preguntar el año de nacimiento en lugar de la edad, ya que las personas sienten que se trata de un dato oficial impersonal y si intentan mentir muestran un tiempo prolongado de reacción.

Lugar de nacimiento. Como el número de cuestionarios en una investigación es relativamente pequeño, en lugar de formular la pregunta de manera abierta,

conviene agrupar la información en cuatro posibilidades: capital, interior de la República, extranjero y extranjero nacionalizado.

Estado civil. La ley sólo marca dos estados civiles: soltero o casado. Así, divorciado, separado, viudo o unión libre pasan todos a ser solteros. Ahora bien, hay investigaciones que necesitan justamente estos datos, por lo que se recomienda, en ese caso, utilizarlos.

Medio socioeconómico. Hay dos maneras de interrogar sobre el medio socioeconómico: 1. Directamente el ingreso mensual, pero, prácticamente, nadie sabe a ciencia cierta cuánto percibe. Los empleados confunden sus ingresos, ya que se expresan en un salario bruto y otro neto. Por lo general, las personas no hacen esa diferencia. Por tanto, se recurre a los intervalos de ingresos, que tienen mejores resultados. 2. Otra forma es agrupar por estratos, es decir, por las diferentes características de la persona. Además, si se entrevista en el domicilio, se puede observar, sin interrogar directamente, el nivel de vida, la colonia donde vive, escolaridad, estado de la vivienda, etc. Todos esos datos arrojan diferentes estratos: alto, medio alto, medio, medio bajo y bajo. Existen mapas mercadológicos donde se señalan las zonas urbanas por la categoría social.

Estudios realizados. En algunos cuestionarios, la pregunta se reduce a cuántos años de escolaridad completos ha realizado el sujeto. Aunque de esta manera se gana información, se pierde el sentido de los ciclos escolares, por lo que se prefiere interrogar directamente sobre los niveles escolares alcanzados: sabe leer y escribir, primaria, secundaria, estudios técnicos, preparatoria, estudios superiores y posgrado. Todos estos niveles y otros más pueden reagruparse en cuatro: sin instrucción, educación elemental, educación media y educación superior.

Ocupación. El análisis de esta pregunta puede convertirse en un dolor de cabeza. Se habla de más de 70 mil actividades profesionales, que en un gran esfuerzo de síntesis algunos catálogos simplifican en cien. Cien son muchas, por lo que terminamos condensando la información en nueve actividades profesionales: campesino, empleado (público o privado), obrero (personas asalariadas), oficiales (personas que trabajan de forma independiente), profesionistas, amas de casa, estudiantes, comerciantes (donde se incluye a los comerciantes, banqueros). A veces se consideran otras categorías profesionales, como artistas, militares, intelectuales, etcétera.

El *periódico que lee* es una información que permite, además de confirmar el estrato socioeconómico, controlar al entrevistador puesto que en principio él desconoce el perfil de cada periódico o si se trata de investigaciones de otros estados, los nombres y la importancia de los diarios que ahí circulan.

Los *barómetros* permiten traducir el optimismo o el pesimismo de las personas y también su actitud frente a sus gobernantes. En épocas de elecciones la preferencia electoral permite hacer predicciones de voto. Los indicadores de estos barómetros nos dan información relevante al cruzarse con otras variables.

Las variables experimentales, activas o dependientes, son las que constituyen realmente el cuestionario. Una buena pregunta debe satisfacer las siguientes condiciones: despertar el interés, ser concisa, de fácil comprensión y no producir

respuestas tendenciosas. Nunca se deben plantear preguntas personales, indiscretas o que evoquen deseos (sexo, higiene, ingresos, etc.). También hay que evitar cuestiones que requieran hacer un esfuerzo de memoria, las preguntas que nos llevan a respuestas de *sí* o *no*. Demasiados *¿por qué?* pueden predisponer al entrevistado. No es conveniente plantear preguntas ambiguas, difíciles o complejas.

Tanto las preguntas como el cuestionario deben ser breves y ágiles. Resulta fatigante responder a un cuestionario con demasiadas interrogantes. Ahora bien, si al probar un cuestionario en una pregunta se obtiene la misma respuesta (hasta 86%), hay que modificar la cuestión, puesto que más que de una variable se trata de una característica de la población.

1.9 PROCESAMIENTO Y ANÁLISIS DE DATOS

El procesamiento de datos puede ser manual o automático. El procesamiento manual para ciertos cálculos estadísticos resulta muy laborioso y cansado. Sin embargo, es el método más adecuado, aunque parezca paradójico, cuando no se dispone de mucho tiempo. El procesamiento automático (con computadora) es el método más adecuado, sobre todo cuando se desean análisis estadísticos multidimensionales.

Cuando el procesamiento de datos es manual, es muy posible que los resultados sean incorrectos. Para evitar esto, conviene cruzar las respuestas para confirmar los resultados y cotejar los resultados parciales con el total. A pesar de la precisión y rapidez de los medios electrónicos, el problema principal radica en la captura de datos, fuente importante de errores. Algunos programas de cómputo cuentan con algoritmos para controlar la captura de datos, pero siempre existe un margen de error. Hay que tener cuidado con algunos programas de cómputo que arrojan resultados susceptibles de interpretaciones equivocadas.

Ahora bien, los datos pueden analizarse desde la estadística descriptiva, que permite determinar una variable y su intervalo de confianza; en cambio, la estadística inferencial contrasta una hipótesis y su intervalo de aceptación o rechazo y puede manejar simultáneamente un buen número de variables. La estadística multidimensional puede tratar simultáneamente todas las variables necesarias (lo cual depende de la capacidad de la computadora de que se disponga), y aunque existan pocos casos y muchas variables, puede encontrarse con relativa facilidad las relaciones que existen entre ellas.

1.10 INFORME FINAL

Nada mejor que recordar a Santo Tomás al comentar la obra de Aristóteles, *De Anima:* aquel que hace un informe persigue tres objetivos: 1. Ganarse la buena voluntad del lector, al mostrar la utilidad de su investigación; 2. Disponerlo al estudio, al proponerle el orden y el plan, y, finalmente, 3. Mantenerlo atento,

al testimoniarle las dificultades encontradas. Al presentar los resultados hay que ser siempre honesto, sobre todo cuando son negativos, contrarios a lo que se esperaba o irrelevantes. Es indispensable evitar las opiniones personales. Debemos mostrar siempre los hechos y no caer en falacias privilegiando los no hechos o expectativas. También hay que evitar las conclusiones o inferencias de resultados que en realidad no aparecen en el estudio. No se debe presentar como positivos resultados que son negativos, pensando que sólo los resultados positivos son exitosos, cuando en realidad un resultado negativo es tan interesante como uno positivo.

Parte 2

Modelos determinísticos

Descripción de datos 2

PROPÓSITOS

El objetivo central del presente capítulo es que el lector pueda explicar los casos A y B para datos agrupados y no agrupados, respectivamente. Asimismo, el lector podrá explicar y realizar las diferentes representaciones gráficas, distribuciones de frecuencia con sus respectivas curvas, además de aplicar los procedimientos para calcular medidas de tendencia central y dispersión a la solución de problemas, tanto para datos agrupados como no agrupados.

Del mismo modo, el lector podrá:

1. Reconocer la descripción de datos como un instrumento efectivo para interpretarlos.
2. Reproducir los diferentes tipos de gráficos, mediante los pasos correspondientes.
3. Interpretar los datos que se presentan en polígonos de frecuencia, gráficas circulares, histogramas, gráficas de barras, ojivas, etcétera.
4. Aplicar el concepto de asimetría y sus diferentes tipos en el diseño de gráficas.
5. Identificar el uso y cálculo de los estadísticos, intervalo de clase y punto medio.
6. Explicar el uso de los siguientes valores estadísticos: cuartil, decil, porcentil o centil, intervalo centil o porcentil.
7. Diseñar y construir una distribución de frecuencias.
8. Explicar la asimetría.
9. Explicar la importancia de las medidas de tendencia central y dispersión.
10. Reconocer los procedimientos para el cálculo en las medidas de tendencia central y dispersión.
11. Identificar la utilización de los estadísticos más adecuados para la interpretación de la media aritmética, la media geométrica y la media armónica.

12. Identificar las distribuciones de frecuencia para la variable continua, la variable discreta o ambas, con sus respectivos estadísticos y gráficos.
13. Explicar los valores estadísticos para datos agrupados y no agrupados.
14. Calcular mediante datos las siguientes medidas de dispersión: amplitud de variación (rango) o intervalo total e intercuartílico.
15. Identificar el uso de dos medidas de dispersión: la amplitud de variación y el coeficiente de variación (Pearson).
16. Considerar los requisitos para el cálculo de los valores estadísticos de dispersión o variabilidad.
17. Reconocer la utilidad de los contenidos del capítulo en la aplicación a la solución de problemas relacionados con la descripción de datos en otras áreas del conocimiento.
18. Realizar e interpretar un análisis estadístico descriptivo, utilizando el paquete de cómputo MACSTAT.

INTRODUCCIÓN

La estadística se considera un método utilizado para recoger, organizar, concentrar, reducir, presentar, analizar, generalizar y contrastar los resultados numéricos (datos) de las observaciones directas o indirectas de los fenómenos reales, así como de la información obtenida a partir de la experimentación, para estar en condiciones de llevar a cabo tanto evaluaciones como conclusiones adecuadas y tomar decisiones acertadas y confiables. En este capítulo realizaremos un análisis estadístico que consistirá en organizar, concentrar, reducir y presentar (en forma gráfica) la información contenida en una muestra representativa de una población. Este análisis estadístico es de una sola variable, entendiendo por variable aquella característica susceptible de medirse, como la temperatura, el coeficiente de inteligencia, etcétera.

La variable se clasifica como *continua* cuando representa a una característica continua, sin saltos ni rupturas, como la edad cronológica de una persona, la cual varía en años, meses, semanas, días, horas, minutos, segundos, décimas de segundos, etc. En cambio, en la variable *discreta*, aunque puede seguir una razón de cambio, no hay continuidad, o sea, existe una ruptura. Por ejemplo, el número de hijos que tienen en un grupo de personas, que pueden ser cero hijos, uno, dos, etc. La variable continua se puede representar por números fraccionarios y enteros, y la variable discreta únicamente por números enteros.

Caso A Datos agrupados

2.1 GRÁFICAS Y DISTRIBUCIONES DE FRECUENCIA

Para la mayoría de las personas, cuando los datos se presentan en forma numérica tienen poco significado o carecen de él. En cambio, si se representan

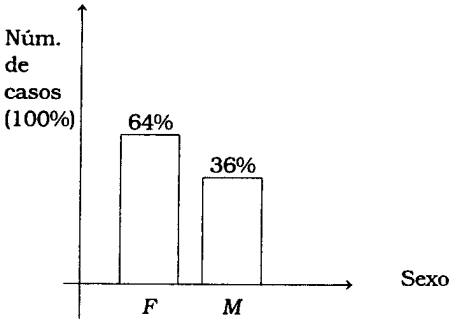


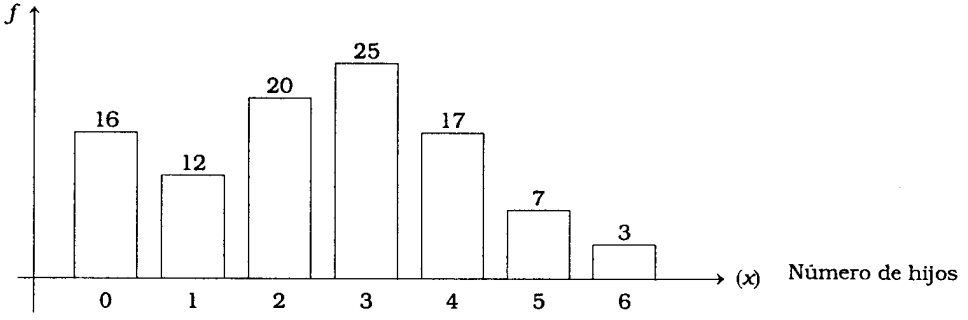
Figura 2.2 Gráfica de barras 1.

Ejemplo 2 En una empresa, se entrevistó a todos los empleados que estaban por cumplir los 20 años de servicio; entre las preguntas de dicha entrevista, consideraremos únicamente el número de hijos que tienen.

A continuación presentamos los datos obtenidos, organizados en una distribución de frecuencias, una forma común de organizar un conjunto de datos agrupándolos en intervalos, categorías o clases, indicando la frecuencia o número de casos de cada uno de ellos. Con esos datos se ha construido una gráfica de barras.

x Núm. de hijos	f Frecuencia[1]
0	16
1	12
2	20
3	25
4	17
5	7
6	3
	100

Cuando la variable de interés es *discreta* como en este ejemplo, la medida de tendencia central más adecuada es la moda, o sea, el dato que tiene la mayor frecuencia de ocurrencia. En este caso la frecuencia mayor es 25 y la moda igual a 3, que representa el 25% de casos.



Nota: El número de padres equivale a la frecuencia de los niveles de la variable (número de hijos).

Figura 2.3 Gráfica de barras 2.

[1] Es el número de veces (casos) que ocurre un evento, que pueden ser calificaciones, observaciones, mediciones, etcétera.

Histograma

Este diagrama es útil cuando se trata de representar distribuciones de frecuencia cuya variable es continua y viene dada en intervalos o clases; dicha gráfica se define y construye como la gráfica de barras, con la diferencia de que las columnas no están separadas sino unidas, lo que le da continuidad.

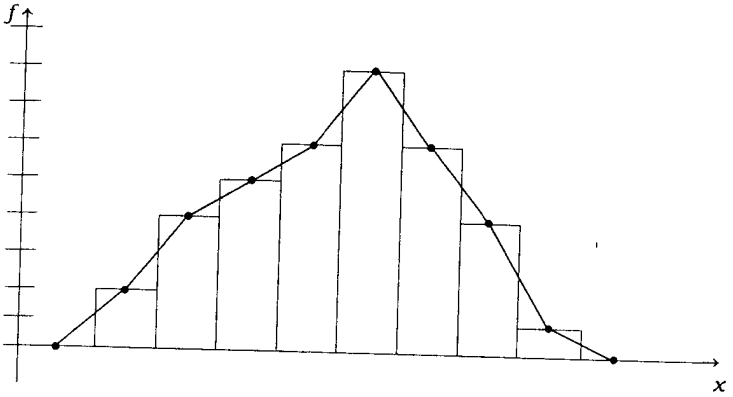


Figura 2.4 Modelo de histograma.

Polígono de frecuencias

Es una gráfica lineal y se construye uniendo por medio de segmentos los puntos medios superiores (marcas de clase) de cada una de las columnas que forman el histograma.[2] El polígono de frecuencias puede contener una amplia variedad de categorías o intervalos, y tiende a destacar la continuidad a lo largo de una escala; por tanto, es útil para representar puntuaciones[3] ordinales y de intervalos.

Polígono de frecuencias acumuladas u ojiva

La representación gráfica de frecuencias acumulada (sumadas progresivamente) se denomina *polígono de frecuencias acumuladas* y también recibe el nombre de *ojiva* o *diagrama de Galton*. Se obtiene uniendo mediante una línea continua los puntos cuyas ordenadas representan las frecuencias acumuladas de los intervalos y su abcisa, el límite real superior (Lrs) de cada uno de ellos. La frecuencia acumulada de cada intervalo representa el número total de casos, dentro y debajo de un intervalo de clase en particular, como se muestra en la figura 2.5.

[2] En este tipo de representación se asume que las frecuencias se ubican en las marcas de clase.
[3] Puntuaciones, datos, calificaciones, observaciones, mediciones, etcétera.

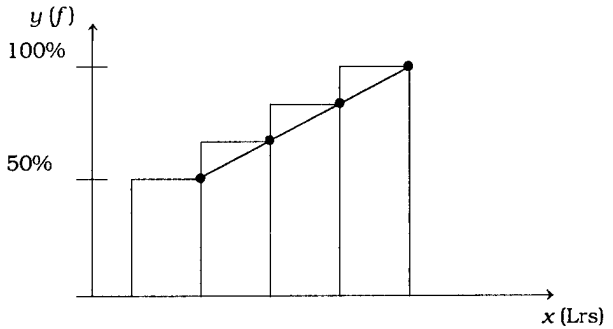


Figura 2.5 Representación de un polígono de frecuencias acumuladas.

TABLAS DE DISTRIBUCIÓN DE FRECUENCIAS

Construcción y representación gráfica

El procedimiento para elaborar una distribución de frecuencias y su tabla respectiva se describirá con base en un ejemplo. En un programa de autocontrol personal del peso, aplicado a 90 personas, los kilogramos que perdieron éstas al terminar dicho programa se muestran a continuación:

Pérdida de peso (kg)

15	8	12	18	44	30	15	18	23	6
23	16	20	17	21	12	12	23	25	13
19	17	17	28	13	17	17	28	18	16
20	7	14	8	15	27	10	19	13	15
18	10	8	11	16	40	18	21	14	27
15	32	28	22	10	9	18	12	25	25
18	20	21	18	18	16	9	8	21	17
29	23	14	14	25	15	12	10	20	16
24	19	15	11	21	12	15	8	17	19

Paso 1. *Agrupamiento u ordenación.* Se forma una tabla ordenando progresivamente los datos, aunque se repitan:

6	10	12	14	16	17	18	20	23	27
7	10	12	15	16	17	18	20	23	28
8	10	12	15	16	17	18	21	23	28
8	10	13	15	16	18	19	21	24	28
8	11	13	15	16	18	19	21	25	29
8	11	13	15	17	18	19	21	25	30
8	12	14	15	17	18	19	21	25	32
9	12	14	15	17	18	20	22	25	40
9	12	14	15	17	18	20	23	27	44

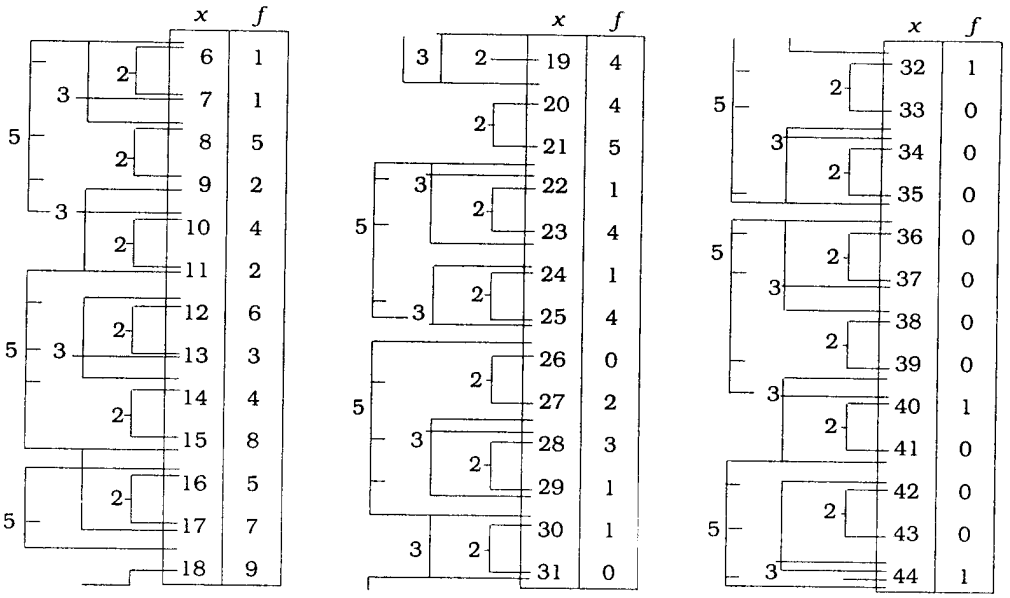
Paso 2. *Marcas de repetición.* Después de ordenar los datos, se forma otra tabla en la que se indican con rayas (/) las veces que se repite cada dato:

6	/	19	/ / / /	32	/
7	/	20	/ / / /	33	
8	/ / / / /	21	/ / / / /	34	
9	/ /	22	/	35	
10	/ / / /	23	/ / / /	36	
11	/ /	24	/	37	
12	/ / / / / /	25	/ / / /	38	
13	/ / /	26		39	
14	/ / / /	27	/ /	40	/
15	/ / / / / / / /	28	/ / /	41	
16	/ / / / /	29	/	42	
17	/ / / / / / /	30	/	43	
18	/ / / / / / / / /	31		44	/

Paso 3. *Frecuencias.* La siguiente tabla que se forma será similar a la anterior, pero en lugar de rayas o marcas se pondrá el número de ellas. Los números obtenidos se llaman *frecuencias*, puesto que indican las veces que se repite un dato. En esta tabla, x representa los datos (kilogramos perdidos) y f la frecuencia.

x	f	x	f	x	f
6	1	19	4	32	1
7	1	20	4	33	0
8	5	21	5	34	0
9	2	22	1	35	0
10	4	23	4	36	0
11	2	24	1	37	0
12	6	25	4	38	0
13	3	26	0	39	0
14	4	27	2	40	1
15	8	28	3	41	0
16	5	29	1	42	0
17	7	30	1	43	0
18	9	31	0	44	1

Paso 4. *Clasificación.* Los datos se agrupan en intervalos o clases. Una clase puede ser de tamaño 2, 3, 5, etc. En la página siguiente se ejemplifica cómo se agrupan los datos en intervalos de tamaños 2, 3 y 5.



Paso 5. *Clasificación en forma práctica.* Los pasos para realizar este procedimiento son los siguientes (para tamaños de clase 2, 3 y 5):

Clases de tamaño 2

x	f	x	f	x	f
6 – 7	2	20 – 21	9	34 – 35	0
8 – 9	7	22 – 23	5	36 – 37	0
10 – 11	6	24 – 25	5	38 – 39	0
12 – 13	9	26 – 27	2	40 – 41	1
14 – 15	12	28 – 29	4	42 – 43	0
16 – 17	12	30 – 31	1	44 – 45	1
18 – 19	13	32 – 33	1		

Clases de tamaño 3

x	f	x	f
6 – 8	7	27 – 29	6
9 – 11	8	30 – 32	2
12 – 14	13	33 – 35	0
15 – 17	20	36 – 38	0
18 – 20	17	39 – 41	1
21 – 23	10	42 – 44	1
24 – 26	5		

Clases de tamaño 5

x	f	x	f
6 – 10	13	26 – 30	7
11 – 15	23	31 – 35	1
16 – 20	29	36 – 40	1
21 – 25	15	41 – 44	1

Es recomendable que los datos se distribuyan entre 5 y 20 clases (es decir, que haya 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 clases o intervalos).

En el ejemplo, el tamaño de clase será 5 y la serie de datos irá en forma creciente, es decir, de 6 a 44 (aunque se puede agrupar también en forma decreciente, de 44 a 6). Ahora determinaremos la amplitud de variación de la siguiente manera:

$$\text{Amplitud } (A) = \text{Puntuación máxima} - \text{Puntuación mínima}$$

De modo que

$$A = 44 - 6 = 38$$

Por tanto

Número de intervalos o clases $= \dfrac{38}{5} = 7.6$, en este texto se denotará por k.

En este caso, para tratar de aproximar lo más posible y obtener el intervalo más exacto, se lleva a cabo el siguiente procedimiento: se aumenta un punto en el extremo superior y se disminuye un punto en el extremo inferior de los datos de la tabla (la frecuencia para estos datos agregados es de 0).

Por tanto, la amplitud queda como

$$A = 45 - 5 = 40$$

Por consiguiente:

Tamaño del intervalo $= \dfrac{40}{8} = 5$, donde $k = 8$ y el tamaño del intervalo se denotará por a,

$$a = 5$$

De este modo, la tabla quedará finalmente así:

x	f
5 – 9	9
10 – 14	19
15 – 19	33
20 – 24	15
25 – 29	10
30 – 34	2
35 – 39	0
40 – 44	2

Y su presentación gráfica será la de la figura 2.6.

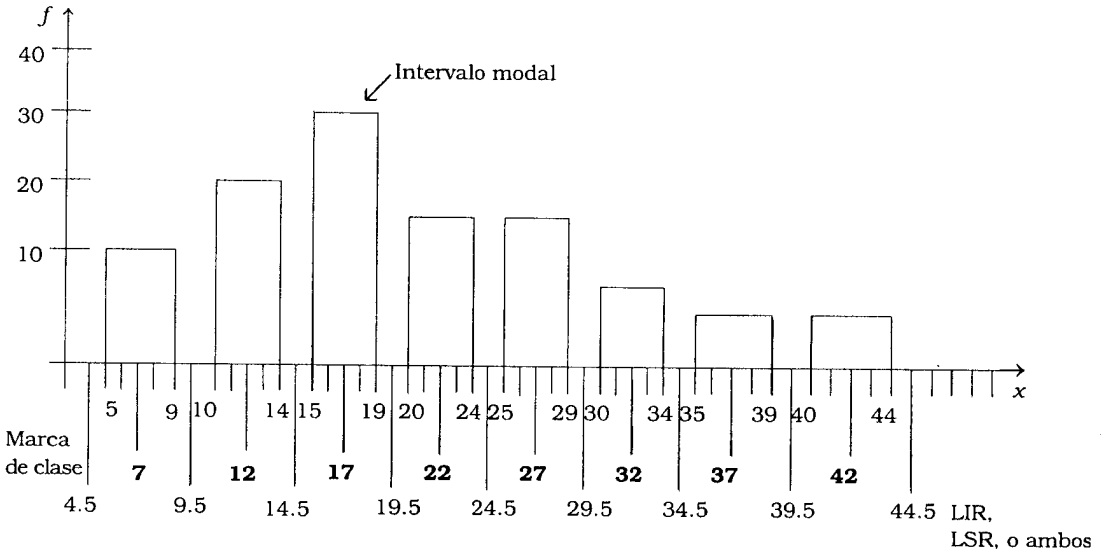


Figura 2.6 Representación gráfica del tamaño de clase.

Paso 6. *Marcas de clase* (*MC*) e *intervalo modal* (*IM*). La marca de clase o punto medio de clase (promedio del intervalo) se obtiene mediante la siguiente fórmula:

$$x = \frac{LI + LS}{2}$$

donde:

x = Marca de clase.
LI = Límite inferior de clase (o intervalo).
LS = Límite superior de clase.

Ejemplos

En el intervalo 5-9:

$$LI = 5,\ LS = 9$$

$$\therefore\ x = \frac{5+9}{2} = 7$$

La marca de clase suele llamarse *punto medio* y en los datos agrupados se considerará como x.

En el intervalo 10 – 14: $LI = 10$, $LS = 14$

o sea, $x = 12$.

Para la totalidad de las clases resulta, entonces:

LI – LS	*f*	*x*
5 – 9	9	7
10 – 14	19	12
15 – 19	33	17
20 – 24	15	22
25 – 29	10	27
30 – 34	2	32
35 – 39	0	37
40 – 44	2	42

Ahora se señalan en negritas las marcas de clase en la gráfica anterior (fig. 2.6).

Intervalo modal (*IM*) es el que contiene la mayor parte de los casos. En consecuencia:

	LI – LS	*f*
	5 – 9	9
	10 – 14	19
IM →	15 – 19	33
	20 – 24	15
	25 – 29	10
	30 – 34	2
	35 – 39	0
	40 – 44	2

En la gráfica se señala también el intervalo modal (*IM*).

Paso 7. *Límites inferior y superior reales.* Como se observa, hay puntuaciones que no están comprendidas. En los intervalos determinados en el paso 4, por

ejemplo, si un dato cayera en 9, 4, o 14.3, no habría clase dónde ubicarlo, por lo que es necesario tomar límites más amplios para cada intervalo. Tales valores son el límite inferior real (*LIR*) y el límite superior real (*LSR*), y se obtienen de la siguiente manera:

a) Como la marca de clase es el punto medio de cada intervalo y la mitad de 5 es 2.5, entonces se tomará la *MC* de cada intervalo y se le restará 2.5, lo cual dará el *LIR*; si se le suma 2.5 se obtiene el *LSR*.
Por consiguiente:

LI – LS	*f*	*x*	*LIR* – *LSR*
5 – 9	9	7	4.5 – 9.5
10 – 14	19	12	9.5 – 14.5
15 – 19	33	17	14.5 – 19.5
20 – 24	15	22	19.5 – 24.5
25 – 29	10	27	24.5 – 29.5
30 – 34	2	32	29.5 – 34.5
35 – 39	0	37	34.5 – 39.5
40 – 44	2	42	39.5 – 44.5

b) Una vez obtenidos el *LIR* y el *LSR*, se marcarán en la gráfica. Al señalar los *LIR* y los *LSR*, el diagrama se convierte en una gráfica de barras y se observa que las mismas ya están unidas, lo que da un histograma. Tomando la marca de clase de cada una de las barras y uniéndolas por líneas se obtiene un *polígono de frecuencias* (fig. 2.7).

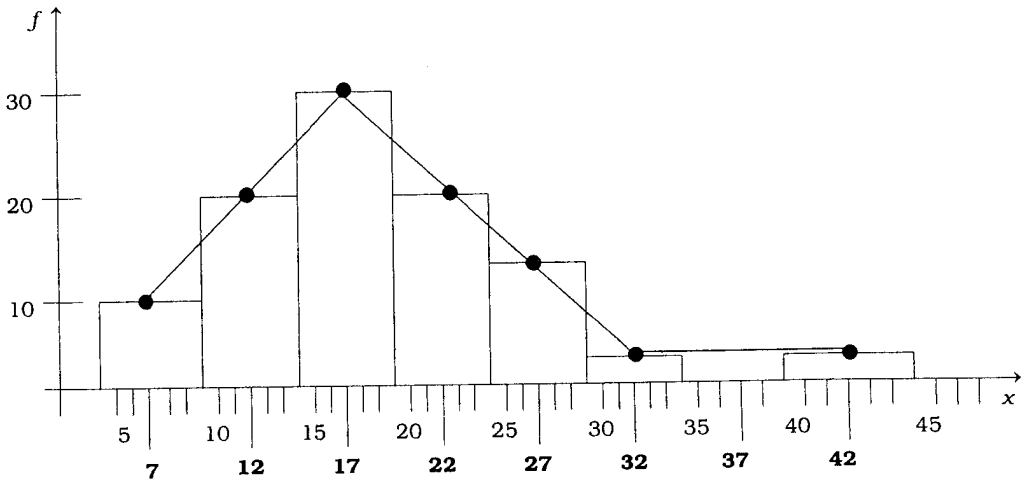


Figura 2.7 Histograma y polígono de frecuencias.

Paso 8. *Frecuencia relativa* y *frecuencia relativa acumulada.* Se ordenan todas las columnas obtenidas anteriormente, y se agregan dos columnas más, que son la de *frecuencia relativa* (f_r) y la *frecuencia relativa acumulada* (f_{ra}).

La columna de f_r se obtiene tomando el porcentaje de la frecuencia de cada intervalo y asignando 100% al número total de frecuencias (en este caso, 90 es el 100%). La f_{ra} se obtiene sumando la frecuencia relativa de un intervalo a la del anterior (en el último intervalo siempre será 100%):

LIR - LSR	*LI - LS*	f	x	f_r	f_{ra}
4.5 – 9.5	5 – 9	9	7	10	10
9.5 – 14.5	10 – 14	19	12	21.11	31.11
14.5 – 19.5	15 – 19	33	17	36.66	67.77
19.5 – 24.5	20 – 24	15	22	16.66	84.43
24.5 – 29.5	25 – 29	10	27	11.11	95.54
29.5 – 34.5	30 – 34	2	32	2.22	97.76
34.5 – 39.5	35 – 39	0	37	0	97.76
39.5 – 44.5	40 – 44	2	42	2.22	100%
		90		100%	

Paso 9. *Ojiva o diagrama de Galton.* Obtenidas las columnas de f_r y de f_{ra} se utilizarán para trazar la curva llamada *ojiva*, lo que se logra situando cada f_{ra} y uniendo con líneas todos los *LSR* de la gráfica. Resulta así el diagrama que se muestra en la figura 2.8.

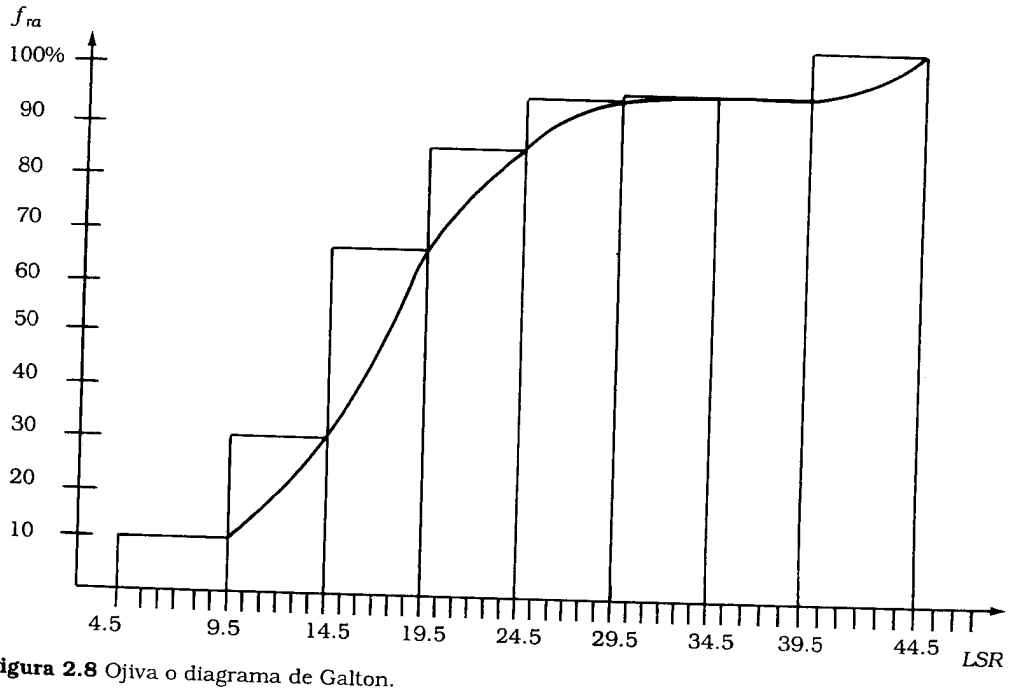


Figura 2.8 Ojiva o diagrama de Galton.

La ojiva obtenida es creciente (o positiva) debido a que los datos fueron dispuestos en forma ascendente. Si se hubieran dispuesto en forma descendente, la ojiva resultante sería decreciente (o negativa).

Agrupamiento de los datos

La información aportada por los datos (puntuaciones, calificaciones, etc.) se puede agrupar o concentrar en pequeños grupos (intervalos o clases); como ya se explicó en el ejemplo anterior, es recomendable que dichas clases sean del mismo tamaño, aunque en algunas no se tenga ningún dato (frecuencia cero). También es recomendable seleccionar una cantidad adecuada de intervalos (entre 5 y 20).

El tamaño y el número de clases se determina en forma empírica dividiendo la amplitud (A = Valor o puntuación más alto – Valor o puntuación más bajo) entre el número de intervalos que deseemos tener; el cociente será el tamaño de los grupos. Suele ocurrir que a la amplitud se le tenga que aumentar una cantidad determinada con fines de redondeo, lo cual de ninguna manera afecta los datos originales y, por consiguiente, no se altera la información original.

El procedimiento para elaborar una distribución de frecuencias y concentrarla en una tabla o cuadro y representarla gráficamente se describirá con base en un ejemplo: un psicólogo realiza un estudio para conocer el tiempo de reacción ante un estímulo visual (una luz roja) que se encienda en un momento determinado, como si fuera un semáforo. A cada uno de los participantes se le pide que accione un pedal, como si frenara un automóvil; al estar en un modelo de comportamiento de conductor de un vehículo, el tiempo de reacción es el que transcurre entre el encendido de la luz y el *frenado*, que se mide en segundos. En este estudio participan 270 personas de un total de 300 que figuraban en la lista, o sea el 90%.

0.20	0.74	1.10	1.34	1.48	1.64	1.90	1.94	2.13	2.22	2.26	2.44	2.62	2.81	2.95	3.23	3.60
0.22	0.75	1.11	1.34	1.50	1.65	1.90	1.95	2.13	2.22	2.26	2.44	2.63	2.82	2.96	3.24	3.61
0.30	0.77	1.13	1.34	1.50	1.65	1.90	1.95	2.15	2.23	2.28	2.45	2.64	2.82	2.97	3.25	3.62
0.31	0.80	1.20	1.35	1.50	1.65	1.90	1.95	2.16	2.23	2.28	2.46	2.65	2.82	3.00	3.28	3.70
0.40	0.81	1.22	1.36	1.53	1.66	1.90	1.96	2.16	2.23	2.29	2.48	2.70	2.83	3.02	3.29	3.74
0.44	0.82	1.23	1.38	1.54	1.70	1.90	1.98	2.17	2.23	2.31	2.49	2.71	2.83	3.05	3.31	3.75
0.50	0.87	1.25	1.38	1.56	1.71	1.91	1.99	2.18	2.23	2.32	2.50	2.71	2.83	3.05	3.33	3.76
0.53	0.88	1.28	1.39	1.57	1.72	1.91	2.01	2.18	2.24	2.32	2.50	2.72	2.84	3.06	3.39	3.78
0.54	0.93	1.30	1.39	1.57	1.74	1.92	2.02	2.18	2.24	2.32	2.51	2.78	2.84	3.08	3.40	3.81
0.60	0.95	1.30	1.41	1.57	1.80	1.93	2.05	2.19	2.24	2.36	2.52	2.78	2.84	3.10	3.40	3.90
0.63	0.98	1.32	1.42	1.58	1.80	1.93	2.06	2.19	2.25	2.36	2.53	2.79	2.85	3.12	3.41	3.95
0.68	1.00	1.32	1.44	1.59	1.81	1.93	2.07	2.19	2.25	2.37	2.54	2.80	2.85	3.13	3.43	4.00
0.70	1.01	1.33	1.45	1.60	1.83	1.93	2.07	2.20	2.25	2.38	2.56	2.80	2.85	3.14	3.45	2.24
0.70	1.03	1.33	1.46	1.62	1.85	1.94	2.07	2.20	2.25	2.39	2.56	2.80	2.85	3.16	3.50	1.64
0.72	1.03	1.33	1.46	1.63	1.86	1.94	2.10	2.21	2.25	2.40	2.60	2.81	2.90	3.20	3.51	
0.73	1.06	1.34	1.46	1.63	1.87	1.94	2.11	2.22	2.26	2.42	2.62	2.81	2.91	3.21	3.55	

Paso 1. Se busca el valor más grande y el valor más pequeño.

Se observa que el valor más grande es igual a 4.00; éste suele denominarse *valor máximo* o *límite superior;* y el más pequeño es 0.20, que suele llamarse *valor mínimo* o *límite inferior.*

Paso 2. Se obtiene la amplitud A que usualmente se llama *rango* (R).

$$A = Rango = 4.0 - 0.2 = 3.8$$

$$\boxed{R = 3.8}$$

Paso 3. Se obtiene el rango incluyente R' *(es el rango anterior* R aumentado un cierto valor de tal forma que sea divisible, sin decimales, entre el número de clases 5, 6, ..., 20). En este caso se puede redondear a 4 o 3.9, ya que 4 se puede dividir entre 5, 8, 10 y 3.9 puede dividirse entre 13. Aquí se dividirá entre 13 (lo deseable es tener el mínimo de decimales posible).

Paso 4. Se calcula el tamaño del intervalo a:

$$a = \frac{R'}{k} \text{ y como } R' = 3.9 \quad \text{y} \quad \boxed{k = 13}$$

$$\boxed{a = 0.3}$$

Paso 5. Se calcula el límite inferior real (*LIR*) y el límite superior real (*LSR*).

Se obtiene la diferencia entre R' y R

$$R' - R = 3.9 - 3.8 = 0.1$$

A esto se le suele llamar *incremento* y se denota por $\Delta = 0.1$. Este incremento se divide entre dos, ya que en cualquier conjunto de datos hay dos extremos: el superior e inferior.

$$\frac{\Delta}{2} = \frac{0.1}{2} = 0.05$$

Calculando el *LIR*, en la siguiente forma, debido a que *LI* es 0.20:

$$LIR = 0.20 - 0.05 = 0.15 \quad \boxed{LIR = 0.15}$$

y como $LS = 4.00$, entonces:

$$LSR = 4.00 + 0.05 = 4.05 \quad \boxed{LSR = 4.05}$$

Paso 6. Se forma la tabla de distribución de frecuencia con los datos obtenidos y se asigna la frecuencia correspondiente para cada intervalo o clase.

Una vez establecido el $LIR = 0.15$ se le suma la constante $a = 0.3$, tanto en forma horizontal como vertical, por lo que el *LSR* de esa primera clase es 0.45 y el límite inferior real (*LIR*) de la segunda clase es también 0.45, así sucesivamente hasta cumplir el número de intervalos obtenidos $k = 13$ y obtener el último *LSR*.

LIR - *LSR*	*f*
0.15 - 0.45	6
0.45 - 0.75	11
0.75 - 1.05	14
1.05 - 1.35	20
1.35 - 1.65	31
1.65 - 1.95	32
1.95 - 2.25	42
2.25 - 2.55	34
2.55 - 2.85	30
2.85 - 3.15	19
3.15 - 3.45	15
3.45 - 3.75	9
3.75 - 4.05	7

Otra forma de obtener el número de intervalos (k) es aplicando la ley de Sturges, que es una regla empírica con base en la siguiente expresión:

$$k = 1 + 3.3\ Log_{10} n$$

para el ejemplo $n = 270$, $Log_{10}\ 270 = 2.4314$

$$k = 1 + 3.3\ (2.4314) = 9.02362$$

y como k debe ser un número entero, se redondea

$$\boxed{k = 9}$$

2.2 SUMATORIAS

Con objeto de describir las propiedades y características de un conjunto de datos, el investigador utiliza ciertos modelos estadísticos y debido a que muchas fórmulas contienen en forma exhaustiva las sumatorias, es necesario conocer las propiedades del operador suma (Σ).

Definición 1

Dado n, un entero positivo, y, sean x, y dos variables

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

Ejemplo

a)
$$\sum_{j=1}^{3} (x_j - y_j) = (x_1 - y_1) + (x_2 - y_2) + (x_3 - y_3)$$

b)
$$\sum_{i=1}^{2} 2x_i^3 = 2x_1^3 + 2x_2^3 + 2x_3^3$$

c)
$$\sum_{k=3}^{4} (x_k - 5) = (x_3 - 5) + (x_4 - 5)$$

Propiedades más importantes del operador (Σ)

1)
$$\boxed{\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}$$

Esto es debido a:

$$\sum_{i=1}^{n} (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + \ldots + (x_n + y_n)$$

agrupando las variables x, y por separado:

$$(x_1 + x_2 + \ldots + x_n) + (y_1 + y_2 + \ldots + y_n)$$

y por definición lo anterior es:

$$\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} (x_i + y_i)$$

Ejemplo

$$\sum_{i=1}^{4} (x_i^2 - x_i)$$

$$\text{Si} \sum_{i=1}^{4} x_i^2 = 60 \qquad \text{y} \qquad \sum_{i=1}^{4} x_i = 12$$

entonces:

$$\sum_{i=1}^{4} (x_i^2 - x_i) = \sum_{i=1}^{4} x_i^2 - \sum_{i=1}^{4} x_i, \text{ sustituyendo lo anterior}$$

se tiene:

$$\sum_{i=1}^{4} (x_i^2 - x_i) = 60 - 12 = 48$$

2) Si k es una constante:

$$\boxed{\sum_{i=1}^{n} kx_i = k \sum_{i=1}^{n} x_i}$$

Si desarrollamos el primer miembro de la igualdad anterior:

$$\sum_{i=1}^{n} kx_i = kx_1 + kx_2 + \ldots + kx_n$$

$$= k(x_1 + x_2 + \ldots + x_n)$$

$$= k \sum_{i=1}^{n} x_i$$

Ejemplo

$$\boxed{\sum_{i=1}^{3} 13 = 3\,(13) = 39}$$

3) Si k es una constante:

$$\boxed{\sum_{i=1}^{n} k = nk}$$

$$\sum_{i=1}^{n} k = \underbrace{k + k \ldots + k}_{\text{“}n\text{” veces}} = nk$$

Ejemplo Si $k = 11$ y $n = 4$:

$$\sum_{i=1}^{n} k = \sum_{i=1}^{4} 11 = 4\,(11) = 44$$

2.3 MEDIDAS DE TENDENCIA CENTRAL

En ocasiones es necesario considerar o disponer de valores que sean representativos de la distribución que forman las puntuaciones, a fin de poder contestar preguntas como las siguientes:

- ¿En qué forma es posible dividir en dos intervalos un conjunto de mediciones de valores, de manera que un grupo contenga la mitad superior del intervalo (50%) y el otro la mitad inferior (el restante 50%)?
- ¿Cuál sería la puntuación promedio de dichos datos (los dos grupos o el total de la muestra)?
- ¿Cuál es la puntuación más común (la que ocurre con más frecuencia en nuestros datos)?

Las preguntas anteriores pueden ser contestadas utilizando las medidas de tendencia central. Estos valores estadísticos por lo general se encuentran en el centro de la distribución, donde se observa una concentración de casos. Las tres medidas de tendencia central utilizadas principalmente, y que responden a las tres preguntas anteriores, son la mediana (*Me*), la moda (*Mo*) y la media aritmética ($\bar{x}$).

Mediana (*Me*)

Se define, como ya vimos, como el punto medio geométrico de la distribución de datos agrupados, o sea, el punto que divide a dicha distribución en dos mitades respecto a las frecuencias. Esta medida de tendencia central es la más adecuada cuando se tienen valores ordinales y se desconocen las puntuaciones extremas de una distribución de frecuencias. Su fórmula de cálculo es la siguiente:

$$Me = \text{LIR} + \left[\left(\frac{\frac{n}{2} - f_{aa}}{f}\right) a\right]$$

donde:

LIR = Límite inferior real del intervalo, donde f_{ra} está cerca del 50%.

n = Número de casos.

f_{aa} = Frecuencia acumulada anterior al intervalo, donde cae la mediana (50%).

f = Frecuencia absoluta donde cae la mediana.

a = Tamaño del intervalo.

Moda (*Mo*)

Es el dato que ocurre más veces (que tiene la mayor frecuencia de ocurrencia); para datos agrupados se obtiene utilizando la fórmula siguiente:

$$Mo = \text{LIR} + \left[\left(\frac{f_p}{f_a + f_p}\right) a\right]$$

donde:

LIR = Límite inferior real del intervalo modal.

f_p = Frecuencia absoluta posterior a la frecuencia del intervalo modal.

f_a = Frecuencia absoluta anterior a la frecuencia del intervalo modal.

a = Tamaño del intervalo.

Media aritmética ($\bar{x}$)

Esta medida suele recibir también el nombre de *media* o *promedio*, y es el valor estadístico de tendencia central más utilizado, ya que es la más confiable de las tres medidas de tendencia central y, por consiguiente, la más representativa de una distribución de datos debido a que es el punto de equilibrio (o "centro de gravedad") de la secuencia de datos. En el caso de datos no agrupados, se define como la suma de todas las puntuaciones o datos, dividida entre el número de ellos.

La fórmula para datos no agrupados es: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

En el caso de los datos agrupados la definición varía, ya que intervienen conceptos nuevos como marca de clase y frecuencia, pero el concepto de media prevalece. La marca de clase la obtenemos sumando el límite inferior y el límite superior de cada intervalo, y dividiendo dicha suma entre 2 (se obtiene así un promedio por cada clase). A tal promedio lo multiplicamos luego por su frecuencia correspondiente. Una vez que se obtiene esto para cada intervalo o clase, se suman todos los resultados y se divide la suma entre el número total de casos, que equivale a la suma de las frecuencias. Es decir:

$$\text{Marca de clase} = \frac{LI + LS}{2} = x = \frac{LIR + LSR}{2}$$

Sucesivamente tenemos:

$$xf, \quad \Sigma\,(fx), \quad \bar{x} = \frac{\Sigma\,(fx)}{\Sigma f}$$

donde:

$$\Sigma f = n$$

Ejemplo Considerando los datos del ejercicio del programa de autocontrol para pérdida de peso aplicado a 90 personas, calculamos a continuación la media, la mediana y la moda, cuando $a = 5$.

a) Cálculo de la media aritmética.

Paso 1. Se calcula la marca de clase para cada uno de los intervalos:

$$\frac{5+9}{2} = 7$$

.

.

.

$$\frac{40+44}{2} = 42$$

Paso 2. Se hacen las operaciones para obtener la columna fx:

f	x	fx
9	7	63
19	12	228
33	17	561
15	22	330
10	27	270
2	32	64
0	37	0
2	42	84

Paso 3. Se determina Σf y $\Sigma (fx)$ con los resultados anteriores:

$$\Sigma f = 9 + 19 + \ldots + 2 = 90; \quad n = \Sigma f = 90$$

$$\Sigma (fx) = 63 + 228 + \ldots + 84 = 1\ 600$$

Paso 4. Se sustituye en la fórmula de la media aritmética y se obtiene así:

$$\bar{x} = \frac{1\ 600}{90} = 17.7778$$

b) Para calcular las otras medidas de tendencia central, la mediana y la moda aplicaremos, respectivamente, las siguientes fórmulas a la tabla que se presenta a continuación:

$$Me = LIR + \left(\frac{\frac{n}{2} - f_{aa}}{f}\right) a \text{ ; asimismo,}$$

$$Mo = LIR + \left(\frac{f_p}{f_a + f_p}\right) a$$

	$LIR - LSR$	f	f_a	x	
	4.5 – 9.5	9	9	7	
	9.5 – 14.5	19*	28++	12	
$LIR\ (IM) \rightarrow$	14.5 – 19.5	33+	61	17	← Valor de la moda en el análisis exploratorio
	19.5 – 24.5	15**	76	22	
	24.5 – 29.5	10	86	27	
	29.5 – 34.5	2	88	32	
	34.5 – 39.5	0	88	37	
	39.5 – 44.5	2	90	42	
		$n = 90$			

* Frecuencia anterior a f_a.
** Frecuencia posterior a f_p.
\+ Frecuencia absoluta donde cae la mediana.
++ Frecuencia acumulada anterior a f.

LIR – LSR	*x*	*f*	f_a	f_r (%)	f_{ra}
4.5 – 9.5	7	9	9	10	10
9.5 – 14.5	12	19	28	21.11	31.11
14.5 – 19.5	17	33	61	36.66	67.77
19.5 – 24.5	22	15	76	16.66	84.43
24.5 – 29.5	27	10	86	11.11	95.54
29.5 – 34.5	32	2	88	2.22	97.76
34.5 – 39.5	37	0	88	0	97.76
39.5 – 44.5	42	2	90	2.22	100
				100%	

En la tabla anterior se localizan todos los elementos que se requieren para calcular tanto la mediana (*Me*) como la moda (*Mo*).

Para obtener la mediana:

LIR = 14.5, ya que el 50% está contenido en la f_{ra} = 67.77%

$$\text{donde:} \quad n = 90, \quad \frac{n}{2} = 45, \quad a = 5$$

$$f_{aa} = 28$$

$$f = 33$$

Sustituimos en la fórmula:

$$Me = 14.5 + \left(\frac{45 - 28}{33}\right) 5 = 17.076$$

Para obtener la moda:

LIR = 14.5 ya que 33 es la frecuencia más alta.

$$f_p = 15$$

$$f_a = 19$$

$$a = 5$$

Sustituimos en la fórmula:

$$Mo = 14.5 + \left(\frac{15}{19 + 15}\right) 5 = 16.706$$

2.4 CUANTILAS

Un conjunto de puntuaciones o mediciones puede dividirse en un cierto número de partes iguales mediante la selección de valores que correspondan a una posición determinada en dicho conjunto. Por ejemplo, la *mediana* divide a un conjunto de valores dados en dos partes iguales, y su posición es, en consecuencia, a la mitad del mismo, de manera que 50% de las puntuaciones quedan a uno u otro lado de dicho valor estadístico.

En general, se llaman *cuantilas* (o *cuantiles*) a estos valores con posición divisora determinada. Pueden considerarse los siguientes cuantiles, además de la mediana:

a) cuartil (o *cuartila*)
b) decil (o "decila")
c) centil (o "centila"), porcentil (o "percentil")

que son, respectivamente, los cuantiles que corresponden a la división entre 4, 10 y 100 partes iguales del conjunto dado.

A continuación se describen estos valores característicos.

A. *En forma gráfica*

Centiles o *porcentiles*.[4] Por lo general, las puntuaciones no elaboradas o crudas (las que se obtienen en forma directa al aplicar una *prueba psicológica* o, en general, cualquier medición) no indican nada en lo relativo al desempeño de las personas, y del lugar que ocupan respecto al resultado de dicha prueba (o bien, de la medición de que se trate) respecto al grupo al cual pertenecen. Por consiguiente, es necesario utilizar varios procedimientos estadísticos, los cuales serán útiles para describir la puntuación o calificación de un individuo particular en relación con otros valores.

Un procedimiento muy adecuado para la comprensión global de los datos obtenidos en un *test* psicológico o en una prueba de conocimientos, comprensión, etc., es determinar los denominados *centiles*. El *centil* de una distribución es el valor dado abajo del cual queda el porcentaje indicado de los valores del conjunto. Un centil indica, entonces, la posición de una puntuación en una distribución porcentual (o en términos de porcentajes). Por ejemplo, si un estudiante obtuvo una calificación que fue más alta que 70% de las puntuaciones en la distribución de estas calificaciones, pero no superior al 71%, el centil correspondiente será el de 70. En otras palabras a tal estudiante le corresponde el 70o. (septuagésimo) centil.

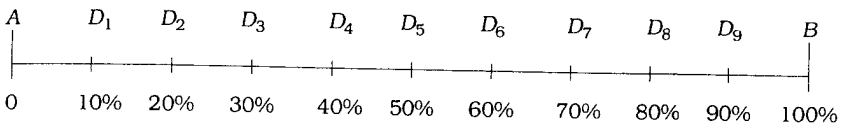
Cuando se emplea el término *centil*, siempre se refiere a un punto en una distribución de puntuaciones o valores, por abajo del cual queda un porcentaje dado de los casos; así, el centil 45 de un conjunto total de 100 puntos es un

[4] Aunque es de uso común el término *porcentil*, es preferible emplear el de *centil*, que concuerda con el nombre de los otros cuantiles.

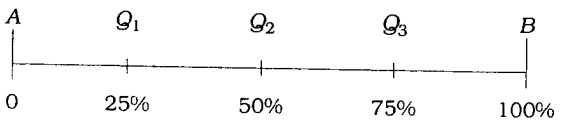
valor o punto por debajo del cual quedan 45 calificaciones. Un centil se representa por C_n (o por P_n, si se utiliza el término *porcentil*).

Deciles. Son los cuantiles que dividen una distribución en 10 tantos a intervalos, por lo que se tienen 9 puntos de división, los *deciles*, que originan los 10 intervalos.

Los deciles, que se representan por D_n, pueden marcarse en una gráfica como la siguiente:



Cuartiles. Son los puntos que dividen a una distribución de valores en cuatro porciones iguales o intervalos. Se representan por Q_1, Q_2, Q_3 y se ilustran en el esquema siguiente:



Debe considerarse la relación que existe entre los centiles, los deciles, los cuartiles y la mediana; su relación en forma sinóptica se representa de la manera siguiente:

Cuartil 1: $Q_1 = P_{25} = C_{25}$ (porcentil 25 o 25o. centil).

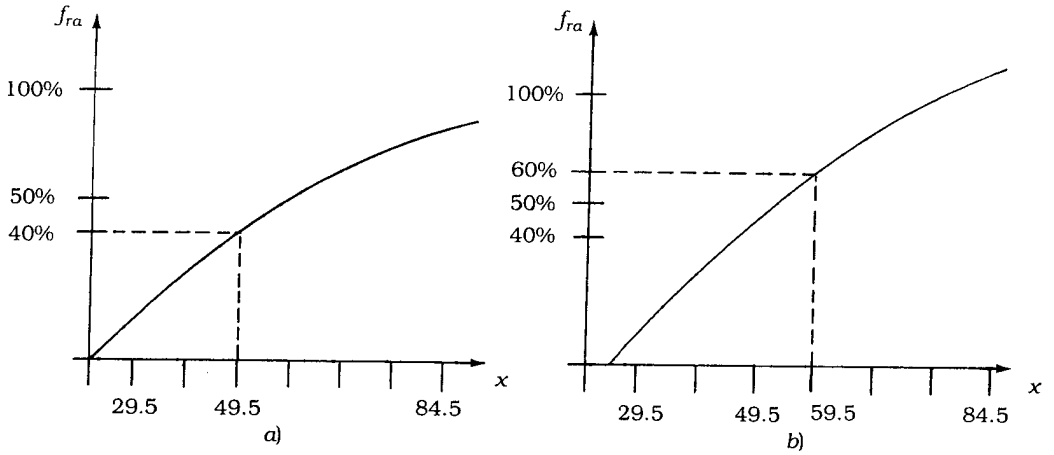
Cuartil 2: $Q_2 = P_{50} = C_{50} = D_5 = Me$ (porcentil 50 o 50o. centil, decil 5, o mediana).

Cuartil 3: $Q_3 = P_{75} = C_{75}$ (porcentil 75 o 75o. centil).

La *ojiva de Galton* (el polígono de frecuencias relativas acumuladas) se puede utilizar en un análisis exploratorio, para ubicar cualquier *cuantil* (cuartiles, deciles, centiles, o la mediana) en una distribución dada.

Ejemplo En el caso de la ojiva de la página siguiente, encontrar:

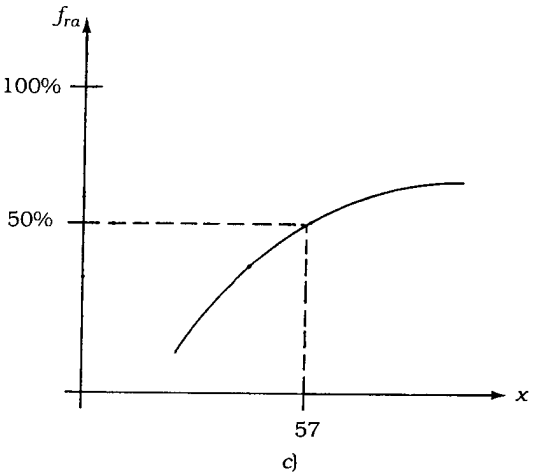
a) El centil de la puntuación 49.5.
b) La puntuación correspondiente a D_6 o C_{60}.
c) La mediana.



a) Se localiza la calificación en el eje horizontal x (puntuaciones) y luego se levanta una línea vertical desde dicho eje hasta encontrar la ojiva; a partir del punto de intersección se traza una línea horizontal hasta llegar al eje vertical f_{ra} (frecuencia relativa acumulada), donde se lee el centil (que es un valor porcentual) correspondiente a la puntuación buscada. Así, se tiene el centil 40, o sea, C_{40}.

b) En este caso se invierte el proceso anterior, o sea, en el eje vertical (fra) se ubica el cuantil deseado (centil, decil o cuartil), en este ejemplo, C_{60} o D_6; se traza luego una horizontal hasta la ojiva, y a partir del punto determinado se lleva una vertical hasta el eje horizontal; se lee a continuación la puntuación o calificación correspondiente al cuantil anterior. En este caso, $C_{60} = 59.5$ (aproximadamente 60 puntos).

c) Se procede como en el inciso anterior, pero buscando en el eje vertical el valor 50% que, como se ha visto con anterioridad, corresponde al C_{50} o D_5 o Q_2, que equivalen a la mediana.



Utilizando este método, la mediana vale aproximadamente 57 puntos.

B. En forma analítica

Se utiliza el modelo para calcular la mediana en datos agrupados, donde $r = 1$

$$Me = LIR + \left(\frac{\frac{r\,n}{2} - f_{aa}}{f} \right) a$$

Cuartiles (Q)

$$Q_r = LIR + \left(\frac{\frac{r\,n}{4} - f_{aa}}{f} \right) a$$

donde r = cuartil que se desea calcular, siendo 1, 2, o 3

Ejemplo Calcular el cuartil 1, $r = 1$ (o sea, el primer 25%)

$LIR = 9.5$. Debido a que el 25% está en la f_{ra}
$n = 90$. Es igual a 31.11%
$f_{aa} = 9$, $f = 19$
$a = 5$

Se sustituyen los valores y se obtiene:

$$Q_1 = 9.5 + \left(\frac{\frac{90}{4} - 9}{19} \right) 5 = 9.5 + \left(\frac{13.5}{19} \right) 5$$

$$Q_1 = 9.5 + 3.5526$$

$$Q_1 = 13.05$$

Deciles (D)

$$Dr = LIR + \left(\frac{\frac{r\,n}{10} - f_{aa}}{f} \right) a$$

donde $r = 1, 2, \ldots 9$

Ejemplo Calcular el decil 8, es decir, el 80%

$LIR = 19.5$
$r = 8$
$n = 90$

$f_{aa} = 61$
$f = 15$
$a = 5$

Sustituimos los valores y obtenemos:

$$D_8 = 19.5 + \left(\frac{\frac{8\,(90)}{10} - 61}{15} \right) 5 = 19.5 + \left(\frac{11}{15} \right) 5$$

$$D_8 = 19.5 + 3.666$$

$$D_8 = 23.17$$

Centiles (C)

$$C_r = LIR + \left(\frac{\frac{r\,n}{100} - f_{aa}}{f} \right) a \quad \text{donde } r = 1, 2, \ldots, 99$$

Ejemplo Calcular el centil 90, es decir, el 90%

$LIR = 24.5$
$r = 90$
$n = 90$
$f_{aa} = 76$
$f = 10$
$a = 5$

Sustituimos los valores y obtener:

$$C_{90} = 24.5 + \left(\frac{\frac{90\,(90)}{100} - 76}{10} \right) 5 = 24.5 + 2.5$$

$$C_{90} = 27$$

Relación de la curva de porcentajes acumulados (ojiva) y las cuantilas

Cuando queremos elaborar una tabla de normas centiles (o porcentiles), la manera más sencilla es utilizar la ojiva. Para ejemplificar esto, utilizaremos los datos de la tabla, de la página siguiente, donde se muestran los intervalos en forma decreciente que en ocasiones se presentan de esta manera. Las mediciones son calificaciones de una prueba que mide capacidad de raciocinio.

(1)	(2)	(3)	(4)	(5)
LI – LS	f	f_a	f_{ra}	f_{ra}%
60 – 64	2	376	1.000	100
55 – 59	12	374	0.995	99.5
50 – 54	20	362	0.963	96.3
45 – 49	32	342	0.907	90.7
40 – 44	46	310	0.824	82.4
35 – 39	58	264	0.702	70.2
30 – 34	64	206	0.548	54.8
25 – 29	58	142	0.377	37.7
20 – 24	42	84	0.223	22.3
15 – 19	23	42	0.112	11.2
10 – 14	15	19	0.050	5.0
5 – 9	4	4	0.011	1.1
	$N = 376$			

Se desarrollará de la siguiente manera:

Paso 1. Se forma la columna (3) y se encabeza con la letra f_a. Esto se hace escribiendo el número total de casos que están por debajo del extremo superior de cada intervalo. Bajo la parte superior del intervalo 5-9 hay 4 casos. Por debajo del extremo superior del intervalo 10-14 hay 19 casos, 4 + 15. Siguiendo este proceso, se obtienen así todos los valores anotados en la columna (3).

Paso 2. Cada uno de los valores de f_a se convierte luego en fracción acumulada f_{ra}. Esto se hace multiplicando cada valor de la columna (3) por el recíproco de 376, o sea 0.00266. Cada producto se redondea al milésimo y se anota en la columna (4).

Paso 3. Cada uno de los valores de la citada columna (4) se multiplica luego por 100 para obtener los porcentajes acumulados que aparecen en la columna (5).

Paso 4. Los datos de la columna (4) se grafican en la figura. En primer lugar se trazan los ejes coordenados de la manera usual: los porcentajes acumulados se sitúan en el eje y, y las puntuaciones, sobre el eje x. En este caso, en el eje y, los valores variarán entre 0 y 100. Al trazar la gráfica de ojiva, el porcentaje acumulado para cada intervalo se marca por encima del *límite superior* de cada intervalo, a diferencia del punto medio que se utilizó en la construcción del polígono de frecuencias. El límite superior del intervalo más bajo es de 9.5.

Por tanto, por arriba de 9.5 sobre el eje x, se hace una marca en 1.1, el f_{ra} para este intervalo. Por encima de 14.5, que es el límite superior del siguiente intervalo, se coloca una marca 5 unidades más adelante en el eje y. Este proceso se continúa hasta que se grafican todos los valores f_{ra}.

Paso 5. A continuación, se dibuja una curva alisada. Muchos de los puntos transportados no aparecerán en tal curva; algunos aparecerán a un lado y algunos en otro. Tales curvas toman la forma que se muestra en la gráfica *c* de la página 56, y con frecuencia se conocen también como *curvas en S*. La curva se lleva a la línea base extendiéndose al siguiente intervalo inferior, dándole a dicho intervalo un f_{ra} de 0.

Paso 6. Los centiles siguientes se leen a partir de la gráfica. Esto se ilustrará obteniendo el quincuagésimo centil (C_{50}) o mediana (Me); ya que por definición, la mediana es el punto en el cual 50% de los casos quedan por debajo de él, se traza una línea desde el f_{ra} de 50 ubicado sobre el eje y, ojiva. Desde este punto de intersección se traza otra recta en ángulo recto hasta el eje x. El punto en el que esta recta corta a dicho eje x es el centil deseado. En la figura 2.9 se han trazado rectas que muestran el valor para C_{80}, C_{50} y C_{25}. Estos valores son aproximadamente 43, 33 y 25, respectivamente. Se observa que estos valores están muy cercanos a los calculados para los mismos valores estadísticos, los cuales son 43.5, 33.1 y 25.4.

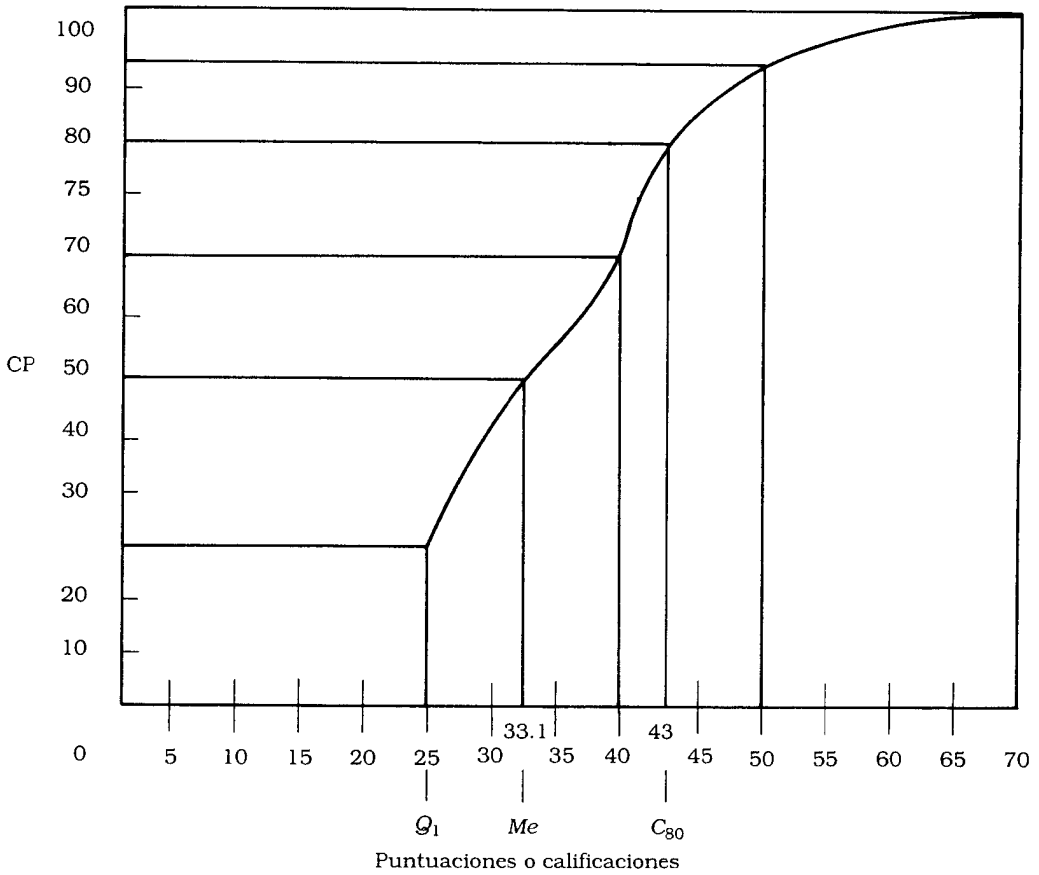


Figura 2.9 Ojiva o curva de porcentajes de acumulados.

Para calcular los porcentiles o centiles, en este caso particular, cuando se tienen los intervalos en forma decreciente, por ejemplo, para calcular el centil 50 que, por definición, tendrá 50% de los casos por encima y 50% por debajo de él. Es decir, es el punto medio de la distribución, y se conoce también como la *mediana*. Dividiendo entre 2 el total N de 376 o tomando 50% del mismo se obtienen 188 casos. Por consiguiente, interesa determinar ese punto de la distribución respecto al cual hay 188 casos por encima y por debajo de él.

Se comienza por contar, en sentido creciente desde abajo, hasta acercarse lo más posible a los 188 casos sin sobrepasar este punto. Esto lleva al punto situado en la parte superior del intervalo 25-29, o al extremo inferior del intervalo 30-34, siendo dicho punto el de 29.5. Hay 142 casos por debajo de este punto. Se requieren 46 casos más de los 64 en el siguiente intervalo. En otras palabras, es necesario avanzar 46/64avos de la distancia a través del tamaño del intervalo, que en este caso es 5. Esto se puede expresar como sigue:

$$C_{50} = 29.5 + \frac{46}{64}(5)$$

$$= 29.5 + \frac{230}{64}$$

$$= 29.5 + 3.59$$

$$= 33.1$$

Lo anterior se comprueba procediendo hacia abajo:

$$C_{50} = 34.5 - \frac{18}{64}(5)$$

$$= 34.5 - \frac{90}{64}$$

$$= 34.5 - 1.4$$

$$= 33.1$$

Si deseamos calcular C_{12} para los mismos datos, en primer lugar tomamos 12% de 376, lo que es igual a 45.12 casos. Es posible contar 42 casos a partir de abajo, lo cual nos lleva a la parte superior del intervalo 19.5. En este punto se debe interpolar:

$$C_{12} = 19.5 + \frac{45.12 - 42}{42}(5)$$

$$= 19.5 + \frac{3.12}{42}(5)$$

$$= 19.5 + \frac{15.60}{42}$$

$$= 19.5 + 0.37$$

$$= 19.9$$

Posteriormente, debemos calcular C_{88}. Podríamos empezar tomando 88% de N y contando en sentido ascendente, como hicimos para el caso de C_{12}. Sin embargo, la tarea se facilita tomando 12% de N y procediendo de arriba abajo.

Por ejemplo:

$$12\% \text{ de } N = 45.12$$

$$C_{88} = 49.5 - \frac{45.12 - 34}{32}(5)$$

$$= 49.5 - \frac{11.12}{32}(5)$$

$$= 49.5 - \frac{55.60}{32}$$

$$= 49.5 - 1.74$$

$$= 47.8$$

Rangos centílicos

Los rangos centílicos son semejantes a los centiles y a menudo se confunden con ellos. Un *rango centílico* designa el porcentaje de *puntuaciones* que caen debajo de una puntuación específica en una distribución. Así, en la figura 2.9 los rangos centílicos se obtienen leyendo la gráfica de manera inversa. Por ejemplo, para obtener el rango centílico de una calificación de 50, a partir de 50 sobre el eje x, se levanta una perpendicular a la ojiva. Desde este punto de intersección se traza una perpendicular al eje y. Este punto de intersección sobre el eje y es el rango centílico. En la figura vemos que una puntuación de 50 tiene un rango centílico de 93. De manera similar, una puntuación de 40 tiene un rango centílico de 69.

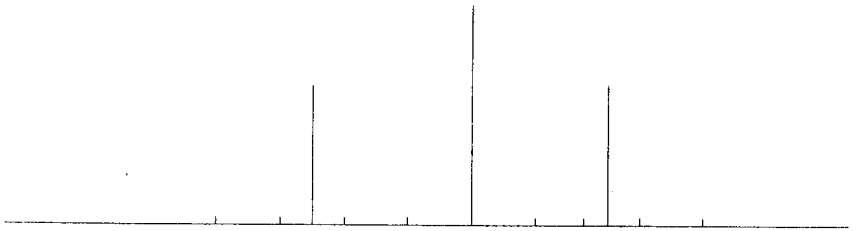
Empleo de centiles

Los centiles se utilizan mucho en las escuelas para informar los resultados de las pruebas estandarizadas. En su favor puede decirse que son muy fáciles de entender. Aun si se desconoce que una persona que tiene una calificación centil

de 77 está en un punto por encima del 77% respecto de aquellos para los que el examen fue estandarizado, por lo menos el 77avo centil parece corresponder a 77%; incluso, el profesor menos informado puede comprender qué quiere decir esto. Puesto que los centiles se parecen a los porcentajes, no hay gran dificultad para comprender su significado. Asimismo, proporcionan una indicación adecuada del rango o la jerarquía de un individuo en un grupo.

Sin embargo, los centiles tienen severas limitaciones, y muchos de los que elaboran y hacen uso de pruebas ya no se molestan por ellas. Si examinamos un conjunto de normas centílicas, o un perfil basado en éstas, observaremos que las normas centílicas se acumulan a la mitad de la distribución. La puntuación (no elaborada) de 33 es equivalente a C_{50}; una puntuación de 36, a un centil 60, y una puntuación de 30, a un centil 40. Un cambio en seis unidades de puntuación es equivalente a un cambio en 20 unidades de centiles. Hay, entonces, una acumulación de centiles en el centro de distribución, y las diferencias entre ellos en esta parte de la curva tienen poco significado.

Establecer que una persona que está en C_{47}, según una prueba, difiere del individuo que está en C_{51}, según la misma evaluación, es darle demasiada importancia a algo que no vale la pena. En el centro de la distribución, el uso de puntuaciones centiles tiende a exagerar diferencias realmente inexistentes. Los centiles son unidades desiguales de medición y no pueden ser tratados aritméticamente. Esto es, no hay justificación para promediarlos, combinarlos o tratarlos de modo matemático. En lo que se refiere a la estadística, los centiles apenas tienen utilidad. No se puede hacer nada más con ellos. Si se desea manejar datos que se han reducido a centiles, deberán convertirse de nuevo en puntuaciones no elaboradas, y operar luego con éstas. Puesto que los centiles son unidades de medida desiguales, algunos estadísticos consideran que es mejor descartarlos. En el caso de pruebas estandarizadas, los centiles se utilizan cada vez menos como método de información. Sin embargo, en ciertos círculos se seguirán usando durante mucho tiempo.



Centiles	5	10	20	40	50	60	80	90	95
Puntuaciones equivalentes	15	20	23	30	33	36	43	49	53

Figura 2.10 Centiles y puntuaciones no elaboradas equivalentes para una serie de datos.

Veamos por qué los centiles se acumulan en el centro de la distribución. Si recordamos la definición de centil como un punto en una distribución con cierto porcentaje de los casos por debajo de él, deducimos que C_{10} es el punto de la

distribución que deja 10% de los casos abajo de él. Decir que 10% del área bajo de una gráfica cae por debajo de C_{10} sería otra forma de expresar esto (fig. 2.10).

Si consideramos ahora C_{20}, por definición, es el punto con 20% de los casos situados por debajo de él. En la figura agregamos otro 10% al área situada por debajo (o antes) de C_{20}. Esto puede continuarse hasta que toda el área, o todos los casos de la distribución, sean incluidos. La distribución resultante queda ilustrada en la figura 2.11, la que en lugar de ser una gráfica común es un rectángulo. La distribución de centiles se describe como rectangular. Las mediciones realizadas en ciencias sociales y biológicas tienden a tomar la forma de la curva normal. Las distorsiones surgen cuando dichos datos se convierten a centiles, que tienen una distribución diferente.

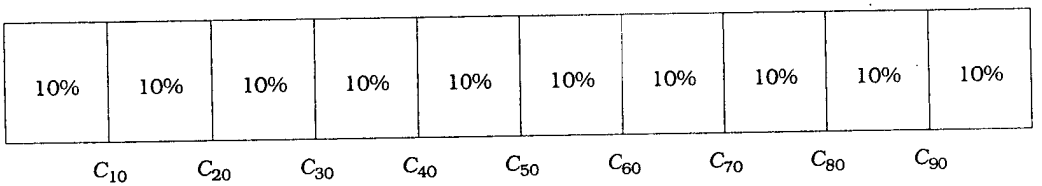


Figura 2.11 Distribución rectangular de centiles.

2.4.1 Diagrama de cajas

Una forma de representar gráficamente a las cuantilas así como a los valores extremos (el mínimo y el máximo) de un conjunto de datos, es una caja rectangular ubicada en un eje vertical u horizontal. En este texto representaremos únicamente el valor mínimo, el máximo y los tres cuartiles Q_1, Q_2, Q_3 (el Q_2 = mediana).

Consideremos el ejemplo de autocontrol de peso donde participan 90 personas ($n = 90$) cuyas cuantilas son:

$$Q_1 = 13.05$$
$$Q_2 = Me = 17.076$$
$$Q_3 = 21.667$$

Valor mínimo = 6
Valor máximo = 44

Una vez que obtenemos los valores anteriores, elaboramos el diagrama que se muestra en la figura 2.12.

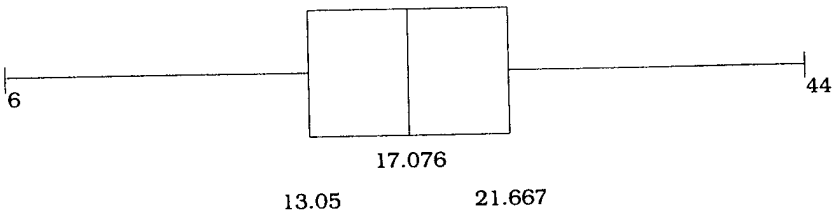


Figura 2.12 Distribución rectangular de centiles.

Este diagrama (fig. 2.13) nos muestra que la distribución de kilogramos perdidos en el programa de autocontrol de peso es simétrico alrededor del valor central, debido a que la longitud izquierda y derecha de la caja respecto a la mediana Q_2 es casi igual.

El diagrama de caja es útil para la comparación de dos o más muestras.

Ejemplo Supongamos que tenemos tres tratamientos cuyos resultados son los siguientes:

Tratamiento 1. 4.79, 3.62, 3.42, 2.38, 2.15, 4.65, 3.33
Tratamiento 2. 0.05, 0.57, 0.05, 0.10, 0.63
Tratamiento 3. 6.61, 3.45, 2.17, 2.27

donde:

$$Me_1 = 3.42$$
$$Me_2 = 0.10$$
$$Me_3 = 2.86$$

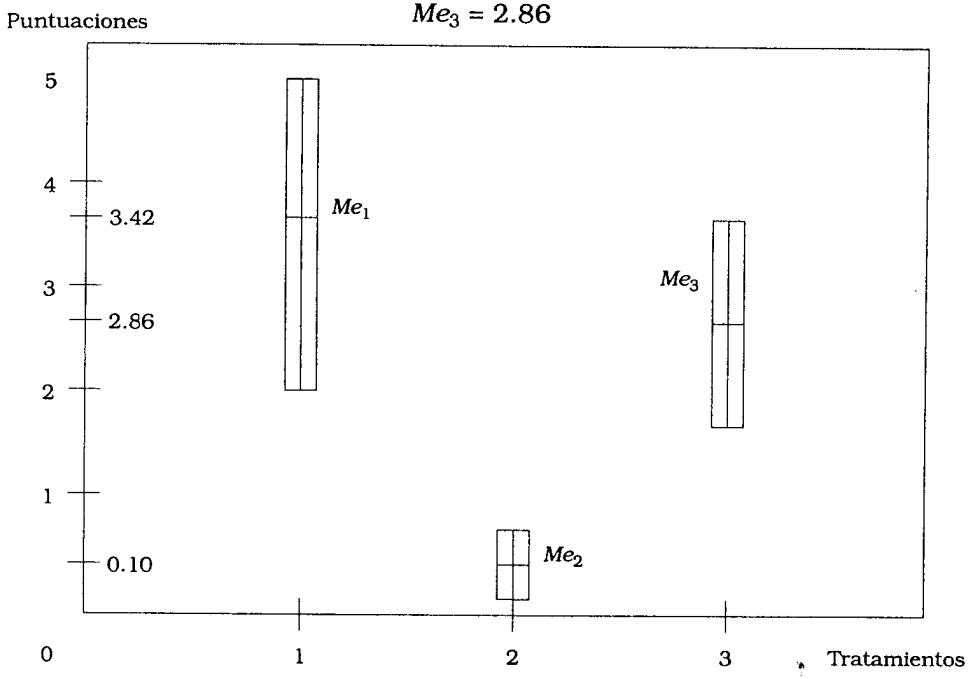


Figura 2.13 Diagrama de cajas.

El diagrama de cajas ilustra en forma muy sencilla la comparación de estos tres tratamientos: el primero tiene una puntuación más alta que el tratamiento 2 y 3, pero en el 3 la puntuación es más alta que en el tratamiento 2.

2.5 MEDIDAS DE DISPERSIÓN O VARIABILIDAD

Ejemplo de cálculo de D_m, s, s^2 y CV

Con los datos del ejemplo de autocontrol de peso, calcularemos ahora las medidas de dispersión o variabilidad: desviación media (D_m), desviación estándar (s), varianza (s^2) y el coeficiente de variación (CV). Para llevar a cabo el cálculo procederemos de la siguiente manera:

Paso 1. Obtenemos la media aritmética por el procedimiento del paso 3 anterior:

$$\bar{x} = 17.7$$

Paso 2. Calculamos ahora la desviación media D_m.

a) Es necesario evaluar la columna $(x_i - \bar{x})$, que corresponde a la marca de clase de cada uno de los intervalos, menos la media aritmética:

$x_i - \bar{x}$	o sea,	$x_i - \bar{x}$
7 – 17.7		- 10.7
12 – 17.7		– 5.7
17 – 17.7		– 0.7
22 – 17.7		– 4.3
27 – 17.7		+ 9.3
32 – 17.7		+ 14.3
37 – 17.7		+ 19.3
42 – 17.7		+ 24.3

b) Una vez obtenida esta columna, se forma otra con los valores absolutos de las diferencias anteriores:

$\lvert x_i - \bar{x} \rvert$
10.7
5.7
0.7
4.3
9.3
14.3
19.3
24.3

c) Cada uno de los valores absolutos se multiplica por su respectiva frecuencia y luego obtenemos la suma total de los intervalos:

$f\,\lvert x_i - \bar{x}\rvert$	o sea,	$f\,\lvert x_i - \bar{x}\rvert$
9×10.7		96.3
19×5.7		108.3
33×0.7		23.1
15×4.3		64.5
10×9.3		93.0
2×14.3		28.6
0×19.3		0.0
2×24.3		48.6
		$\Sigma f\,\lvert x_i - \bar{x}\rvert = 462.4$

d) Al dividir el resultado anterior entre ***n*** obtenemos la desviación media buscada:

$$D_m = \frac{462.4}{90} = 5.138 \quad \text{o bien,} \quad D_m = 5.14$$

Paso 3. Para evaluar la desviación estándar ***s:***

a) Elevamos al cuadrado cada uno de los renglones de la columna obtenida en el paso 2 anterior:

$(x_i - \bar{x})^2$	o sea,	$(x_i - \bar{x})^2$	o sea,	$(x_i - \bar{x})^2$
$(7 - 17.7)^2$		$(-10.7)^2$		114.49
$(12 - 17.7)^2$		$(-5.7)^2$		32.49
$(17 - 17.7)^2$		$(-0.7)^2$		0.49
$(22 - 17.7)^2$		$(4.3)^2$		18.49
$(27 - 17.7)^2$		$(9.3)^2$		86.49
$(32 - 17.7)^2$		$(14.3)^2$		204.49
$(37 - 17.7)^2$		$(19.3)^2$		372.49
$(42 - 17.7)^2$		$(24.3)^2$		590.49

b) Ahora multiplicamos cada uno de estos resultados por su respectiva frecuencia y se determina la suma total, que aparece en la página siguiente:

$f(x_i - \bar{x})^2$	o sea,	$f(x_i - \bar{x})^2$
9×114.49		1 030.41
19×32.49		617.31
33×0.49		16.17
15×18.49		277.35
10×86.49		864.90
2×204.49		408.98
0×372.49		0.0
2×590.49		1 180.98
		$\Sigma f(x_i - \bar{x})^2 = 4\ 396.10$

c) A este resultado lo dividimos entre $n - 1$, y al cociente obtenido le extraemos la raíz cuadrada; de este modo obtenemos la desviación estándar buscada:

$$s = \sqrt{\frac{\Sigma f(x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{4\ 396.10}{89}} = \sqrt{49.39} = 7.028$$

Si en el paso anterior no se extrae la raíz cuadrada al cociente 49.39, entonces lo que se tiene es la varianza (s^2), o sea,

$$\text{Varianza} = (\text{desviación estándar})^2$$

$$\therefore\ s^2 = 49.39$$

Como último paso, se divide la desviación estándar entre la media aritmética y se multiplica por 100 para obtener la medida relativa (en porcentaje), que se conoce como *coeficiente de variación* (de Pearson). Esta medida expresa la proporción en la que la media aritmética no es representativa del conjunto de datos de donde proviene. Eso sólo es posible cuando conocemos la desviación estándar y la comparamos con la media aritmética utilizando dicho coeficiente:

$$CV = \frac{s}{\bar{x}} = \frac{7.02}{17.7} \times 100 = 0.39 \times 100 = 39\%$$

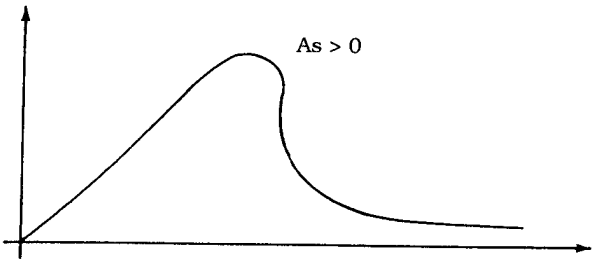
si a 100% – *CV* = confiabilidad de la media, o sea 100% – 39% = 61%.

2.6 PROPIEDADES DE LA DISTRIBUCIÓN DE FRECUENCIAS

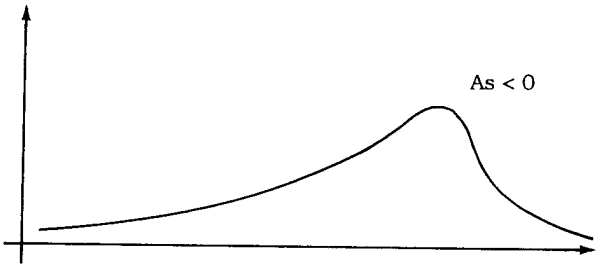
Una de las formas de captar mejor una distribución de frecuencias es por medio de su gráfica. Una vez obtenida ésta, es conveniente observar su asimetría, su curtosis y su modalidad.

Asimetría (As)

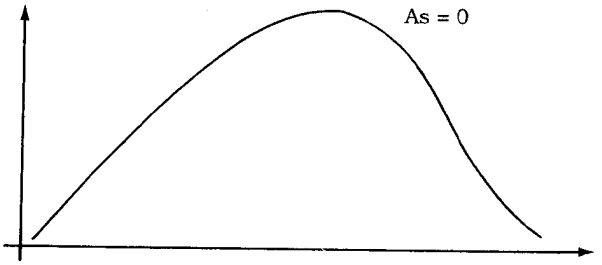
La asimetría se presenta cuando una curva de distribución, en una de sus colas, se extiende más lejos que la otra en una dirección. Así, cuando está extendida hacia la derecha, se dice que la curva tiene *asimetría positiva.*

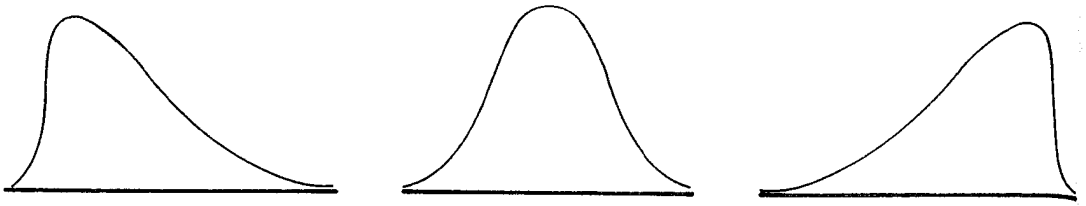


Cuando se extiende hacia la izquierda se tiene una curva con *asimetría negativa:*

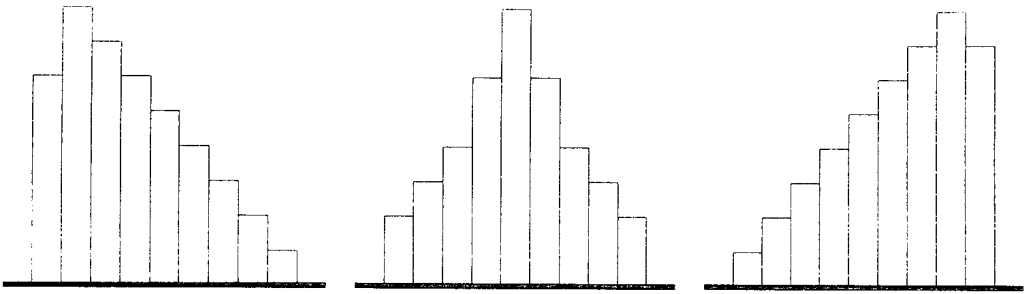


Si la asimetría es igual a cero, entonces se tiene una *curva normal:*





a) Distribuciones continuas



Asimetría positiva Simétrica Asimetría negativa

b) Distribuciones discretas

Figura 2.14 Distribuciones continuas y distribuciones discretas.

El llamado *coeficiente de asimetría* o *sesgo* se calcula con la siguiente fórmula:

$$As = \frac{\bar{x} - Mo}{s}$$

donde:

As = coeficiente de asimetría de Pearson
$\bar{x}$ = media muestral
Mo = moda
s = desviación estándar muestral

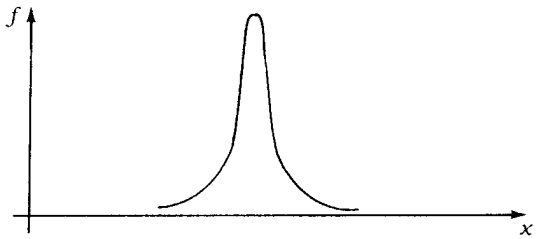
Ejemplo Con los resultados del autocontrol de peso, que consideramos con anterioridad, calculamos el coeficiente de asimetría:

$$As = \frac{17.7 - 16.71}{7.02} = 0.0698$$

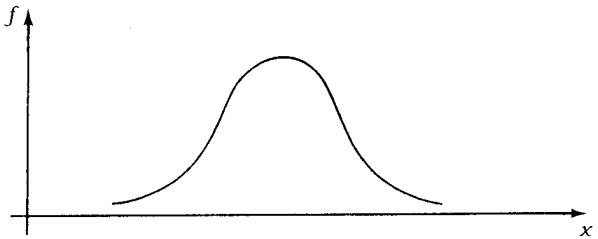
Curtosis

La curtosis es la agudeza que presenta el perfil de una curva unimodal. Por ejemplo, cuando las puntuaciones obtenidas al aplicar un *test* psicológico tienden a agruparse en el centro de la distribución en un intervalo reducido de

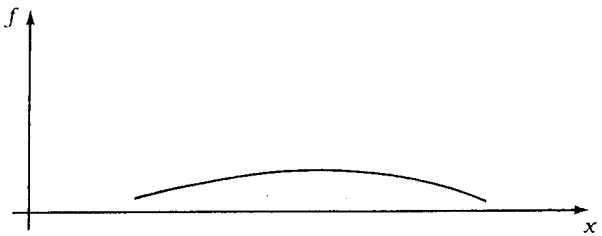
valores, se tiene una curva *aguda* o *leptocúrtica*. Cuando esto ocurre, se dice que el grupo es *homogéneo*[5] respecto a lo que se mide:



Si el intervalo en el que tienden a agruparse las puntuaciones de un grupo no es tan reducido, la curva se denomina *semiaguda* o *mesocúrtica:*



Un caso opuesto a la curva leptocúrtica es el que corresponde a una distribución donde existe un intervalo amplio de puntuaciones con una agrupación mínima en el centro. A esta curva se le llama *aplanada* o *platocúrtica:*

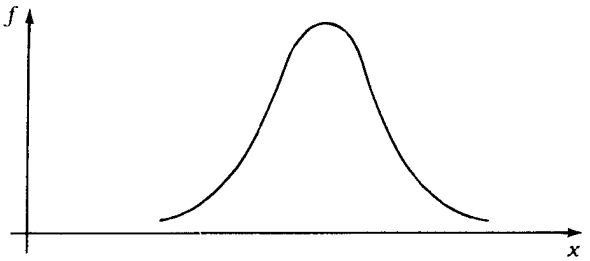


Modalidad

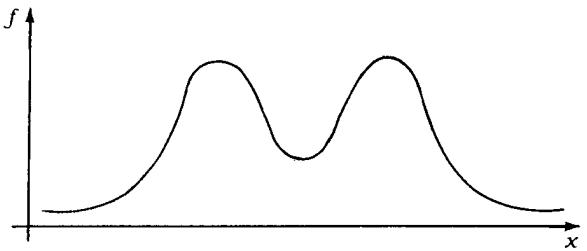
La tercera característica de una curva de distribución es la *modalidad*, que consiste básicamente en el número de picos que presente dicha curva.

[5] Las medidas de dispersión son pequeñas.

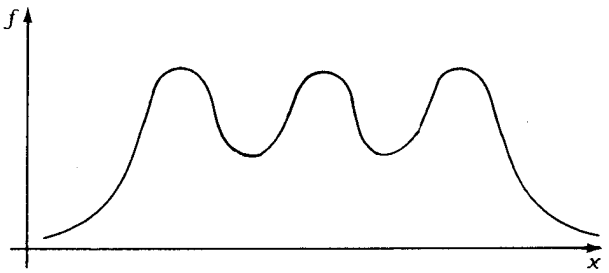
Las curvas que tienen un solo pico se denominan *unimodales*, es decir, de una sola moda:



Otras pueden tener dos picos, por lo que se les conoce como *bimodales*, o con dos modas:



Asimismo, algunas otras pueden tener tres o más picos, y se les llama entonces *multimodales:*



La relación existente entre las tres medidas de tendencia central se muestra en forma gráfica, en la página siguiente.

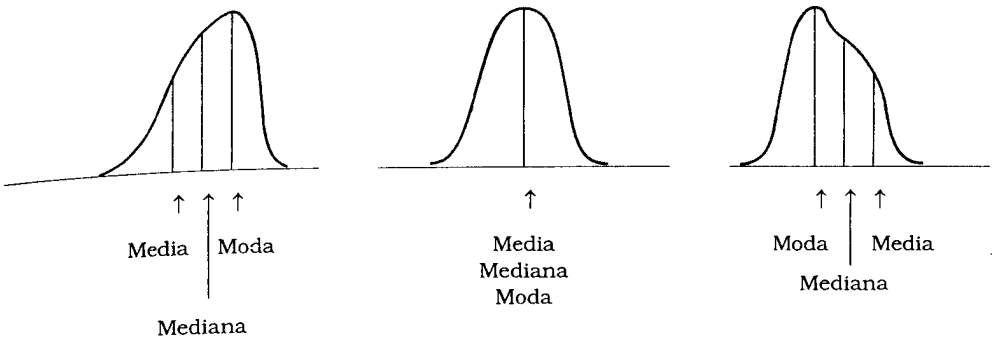


FIGURA 2.15 Cuando una distribución es simétrica, su media, mediana y moda son iguales. De otra manera, la media y la mediana se cargan hacia los valores extremos, la media más que la mediana.

En el caso de que la distribución sea moderadamente sesgada $As \approx 0$, existe una relación aproximada entre las tres medidas de tendencia central:

$$\text{media} - \text{moda} = 3\ (\text{media} - \text{mediana})$$

Es interesante destacar que, en el caso de distribuciones asimétricas con pico muy agudo (curtosis alta, o sea, una curva leptocúrtica), la mediana constituye la medida de tendencia central más útil y representativa.

Caso B Datos no agrupados

Es común que un investigador desee conocer ciertas características importantes de un conjunto grande de objetos o personas, conjunto al que en estadística se le conoce como *población*. En general, es imposible y, además, no es conveniente medir todas las variables de interés (características susceptibles de medición), al igual que todos los elementos de la población en estudio, por lo que seleccionamos un subconjunto de dicha población y medimos únicamente las variables y los elementos que lo componen; a este conjunto de mediciones lo denominaremos *muestra*. Por ejemplo, si deseáramos conocer el coeficiente de inteligencia (CI) de los candidatos a ingresar en una carrera profesional en todas las instituciones de educación superior en México, en vez de dedicarnos a la gigantesca e irrelevante tarea de investigar a todos los candidatos, lo cual implicaría una gran inversión de tiempo y dinero —además del peligro de que, al terminar el estudio, el número de solicitantes ya no fuera el mismo del inicio—, lo más conveniente sería seleccionar una muestra representativa de la población analizada. En los siguientes capítulos estudiaremos las condiciones necesarias y suficientes para obtener la muestra adecuada y representativa de una población dada.

Los resultados así obtenidos serán válidos y precisos en la mayor parte de los casos, debido a que la inversión de tiempo fue mínima y el tamaño de la muestra es constante, cosas que no ocurren al estudiar la población total. La información y los resultados serán una estimación lo bastante confiable como para generalizarlos a la población.

2.7 MEDIDAS DE TENDENCIA CENTRAL

Una de las formas más sencillas y confiables de describir un conjunto de datos, o mediciones que pertenecen a una muestra o población, es conocer su *promedio*. Para esto se hablará de tres *promedios* o *medidas de tendencia central*: la media aritmética, la mediana y la moda. A continuación definiremos cada una de ellas.

Media aritmética ($\bar{x}$)

Se escogieron al azar 25 frascos que contenían un reactivo y se midió la cantidad en exceso del contenido neto que tenían; las cantidades obtenidas fueron 2, 4, 6, 0, 6, 2, 3, 0, 0, 4, 5, 3, 3, 4, 3, 5, 2, 1, 2, 0, 3, 4, 5, 1, 1 mg.

En seguida se designará cada uno de estos números en la siguiente forma:

$$x_1 = 2,\ x_2 = 4,\ x_3 = 6,\ \ldots,\ x_{25} = 1$$

Una vez hecho esto, definimos la *media aritmética* (muestral), $\bar{x}$, como la suma aritmética de las medidas, dividida entre el número total de mediciones. Esto es:

$$\bar{x} = \frac{2 + 4 + 6 + \ldots + 1}{25} = 2.76$$

$$\boxed{\bar{x} = 2.76}$$

Generalizando este ejemplo:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_{25}}{25}$$

y para cualquier número (n) de datos:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

El numerador de esta fórmula se puede escribir en forma abreviada utilizando la notación de sumatoria (denotada por la letra griega sigma mayúscula, Σ[6]):

[6] También llamada *notación sigma*.

$$\Sigma x = x_1 + x_2 + x_3 + \ldots$$

Para n casos:

$$\Sigma x = x_1 + x_2 + x_3 + \ldots + x_n$$

por lo que la fórmula general resultará en la siguiente forma:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Moda (*Mo*)

La llamada *moda* (*Mo*) es la medida que ocurre el mayor número de veces (la que tiene la mayor frecuencia, f) en una muestra dada. Por tanto, en nuestro ejemplo la moda sería:

Cantidad en exceso mg (x)	Frecuencia (f)
0	4
1	3
2	4
Moda = 3	5 = Nivel modal
4	4
5	3
6	2
	$\Sigma f = 25$

La moda es $Mo = 3$ (el valor que tiene la mayor frecuencia, 5); o sea, la mayor parte de los frascos (5) contienen 3 mg. Se dice que este conjunto de datos es unimodal, pues solamente tiene una moda:

$$Mo = 3$$

Mediana (*Me*)

Cuando cada uno de los datos o mediciones de una muestra se han ordenado en forma creciente o decreciente, la *mediana* (*Me*) es el valor cuyo nivel modal es $\frac{n+1}{2}$, o bien, $\frac{n}{2}$, cuando el número de casos es impar o par, respectivamente.

En nuestro ejemplo, al ordenar los datos en forma creciente se tiene:

Datos: 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5, 5, 6, 6.
Orden: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13.

Por consiguiente,

$$Me = 3$$

2.8 MEDIDAS DE DISPERSIÓN O VARIABILIDAD

Ya hablamos del uso de las medidas de tendencia central, las cuales proporcionan información concerniente a un conjunto de datos; no obstante, pueden describir únicamente un aspecto de la población o muestra. Sería incorrecto concluir que dos conjuntos de datos son iguales sólo porque tienen las mismas medidas de tendencia central, es decir, que el valor de la media aritmética sea el mismo para ambos conjuntos, y la mediana y la moda también podrían ser iguales, pero la distribución de dichos datos forma una curva completamente diferente. Esto ocurre porque las "distancias" de los datos tienen diferentes concentraciones respecto al punto de equilibrio, que está representado por la media aritmética. A tales distancias se les suele denominar *medidas de variación* o *dispersión*, a diferencia de la tendencia central, y también son conocidas como *medidas de variabilidad*.

El siguiente ejemplo sirve para ilustrar el porqué de la necesidad de estas medidas de dispersión, que darán a la media aritmética un valor más confiable y realmente representativo de la distribución de los datos o las puntuaciones obtenidos en una muestra y, por consiguiente, representan a la población.

Supongamos que tenemos dos grupos de ocho personas en cada uno (grupo A y grupo B) de una escuela secundaria, y deseamos comparar dichos grupos en el número de errores obtenidos por cada uno de sus ocho integrantes; al aplicarles una prueba que consta de 20 reactivos:

Grupo A	Grupo B
$Mo_A = 5$	$Mo_B = 5$
$Me_A = 5$	$Me_B = 5$
$\bar{x}_A = 5$	$\bar{x}_B = 5$

Aparentemente no hay diferencia entre el grupo A y el grupo B, pero si se observan detenidamente los datos iniciales se verá que ambos grupos no son iguales. En primera instancia se tiene que los errores del grupo A se dispersan más que los del grupo B, que parecen concentrarse alrededor del valor promedio (5).

	Grupo *A*	Grupo *B*	
Puntuación más baja →	1	3	← Puntuación más baja
	1	4	
	2	5	
	5	5	
	5	5	
	5	5	
	9	6	
Puntuación más alta →	12	7	← Puntuación más alta
	$\Sigma x = 40$	$\Sigma x = 40$	

¿Qué ocurre? ¿Por qué la medida de tendencia central, en este caso la media aritmética o promedio, no nos da suficiente información acerca de estos resultados? Porque es necesario contar con algo que señale la dispersión o desviación respecto a la media aritmética; en otras palabras, conocer la *densidad* de los datos, es decir, cuán concentrados se encuentran, cuán homogéneos son, o qué variados están dichos datos.

Amplitud de variación (*A*)

Para responder a lo anterior, se obtiene una de las medidas de dispersión más sencillas, la *amplitud de variación* (o *rango*),[7] que se tiene al restar de la puntuación más grande la puntuación más pequeña.

Mediante el cálculo de la amplitud de cada uno de los grupos se tiene:

Grupo *A*: Puntuación más alta, 12; puntuación más baja, 1. Amplitud $A = 11$.
Grupo *B*: Puntuación más alta, 7; puntuación más baja, 3. Amplitud $A = 4$.

En un primer análisis exploratorio, la amplitud puede ser un valor estadístico útil y es el que nos indica cuánto varían las puntuaciones del primer grupo respecto a los del segundo; no obstante que dicho valor estadístico es una medida simple de calcular y fácil de entender, su simplicidad limita sus aplicaciones.

Desviación media (*Dm*)

Otra medida de dispersión que tiene un significado intuitivo es la *desviación media* (D_m), la cual puede visualizarse evaluando la *distancia* entre cada observación (puntuación) y la media aritmética. El promedio de estas distancias da una medida racional de la dispersión de los datos.

[7] Aunque se emplea mucho la palabra *rango* (traducción poco afortunada del término inglés *range*), por ser incorrecto su significado, en esta obra se preferirá la expresión *amplitud de variación*; en ocasiones se puede decir simplemente *amplitud* o *intervalo total*.

Todo esto se obtiene mediante la fórmula siguiente:

$$D_m = \frac{\Sigma\ |x_i - \bar{x}|}{n}$$

donde las rayas verticales indican valor absoluto, y hay que tratar a todos los números negativos como positivos; cada diferencia $x_i - \bar{x}$ es la distancia o *desviación* de la puntuación respecto a la media aritmética; al sumar estas desviaciones y dividirlas entre el número de casos (sujetos), obtenemos la desviación media.

No obstante que la desviación media es una medida intuitiva de la dispersión, también resulta ser inconveniente, ya que a partir de ella no se pueden obtener otros valores estadísticos. La medida de dispersión más común es la *desviación estándar*, la cual es más compleja y menos intuitiva que la desviación media, pero tiene algunas propiedades matemáticas que la hacen muy valiosa y útil para problemas estadísticos más complicados.

Desviación estándar (s o σ)

Una medida de variabilidad más adecuada es la *desviación estándar*, que se simboliza con s para una muestra, y con σ para una población. Su fórmula general de cálculo es la siguiente:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

Para sustituir los datos en la fórmula, primero se eleva al cuadrado la diferencia de cada una de las puntuaciones respecto al valor de la media aritmética, o sea $(x - \bar{x})^2$; sumando todas las desviaciones cuadradas en cada uno de los grupos, se obtiene $\Sigma\ (x_i - \bar{x})^2$, como se ve a continuación:

Grupo *A*

$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
- 4	16
- 4	16
3	9
0	0
0	0
0	0
4	16
7	49
	$\Sigma\ (x_i - \bar{x})^2 = 106$

Grupo *B*

$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
- 2	4
- 1	1
0	0
0	0
0	0
0	0
1	1
2	4
	$\Sigma\ (x_i - \bar{x})^2 = 10$

Aplicando la fórmula general de s a cada grupo A y B:

$$s_A = \sqrt{\frac{\Sigma\,(x_i - \bar{x})^2}{n-1}} \qquad\qquad s_B = \sqrt{\frac{\Sigma\,(x_i - \bar{x})^2}{n-1}}$$

$$s_A = \sqrt{\frac{106}{7}} \qquad\qquad s_B = \sqrt{\frac{10}{7}}$$

$$s_A = \sqrt{15.14} \qquad\qquad s_B = \sqrt{1.42}$$

$$\therefore\ s_A = 3.89 \qquad \text{y} \qquad \therefore\ s_B = 1.19$$

Si se calculan ahora las desviaciones simples de cada una de las puntuaciones respecto a su media aritmética $(x_i - \bar{x})$ para cada grupo, se tendrá lo siguiente:

Para el grupo A:

x_i	$\bar{x}$	$(x_i - \bar{x})$	$\lvert(x_i - \bar{x})\rvert$
1	5	− 4	4
1	5	− 4	4
2	5	− 3	3
5	5	0	0
5	5	0	0
5	5	0	0
9	5	4	4
12	5	7	7
		$\Sigma = 0$	$\Sigma = 22$

o sea,

$$\Sigma\,(x_i - \bar{x}) = 0 \qquad \text{y} \qquad \Sigma\,\lvert(x_i - \bar{x})\rvert = 22$$

Para el grupo B:

x_i	$\bar{x}$	$(x_i - \bar{x})$	$\lvert(x_i - \bar{x})\rvert$
3	5	− 2	2
4	5	− 1	1
5	5	0	0
5	5	0	0
5	5	0	0
5	5	0	0
6	5	1	1
7	5	2	2
		$\Sigma = 0$	$\Sigma = 6$

o sea,

$$\Sigma\,(x_i - \bar{x}) = 0 \qquad \text{y} \qquad \Sigma\,|(x_i - \bar{x})| = 6$$

En lo anterior se tiene la columna del valor absoluto, que es simplemente el valor numérico de la diferencia (aritmética) de los dos valores, sin tomar en cuenta el signo. Al sumar los valores de esta columna y dividir entre el número de casos, se obtiene la desviación media para los grupos *A* y *B:*

$$(D_m)_A = \frac{\Sigma\,|(x_i - \bar{x})|}{n} = \frac{22}{8} = 2.75$$

$$(D_m)_B = \frac{\Sigma\,|(x_i - \bar{x})|}{n} = \frac{6}{8} = 0.75$$

La desviación estándar en cada grupo es, como se calculó antes, $s_A = 3.89$ y $s_B = 1.19$, y puede apreciarse así la diferencia entre estos valores estadísticos.

Varianza (s^2 o σ^2)

Si elevamos al cuadrado cada una de las desviaciones respecto de la media aritmética, sumamos y dividimos entre $(n - 1)$, obtenemos la llamada *varianza*:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

De manera que:

$$\text{Var.} = (\text{Desv. est.})^2$$

lo que justifica la notación utilizada (s^2).

Quizá el lector se pregunte por qué en el divisor utilizamos $n - 1$ en lugar de n al definir la varianza muestral s^2. Obsérvese que si la suma de las desviaciones cuadradas fuese divida entre n, el cociente resultante sería el promedio de los cuadrados de las desviaciones (es decir, una *desviación cuadrada media*). El divisor $n - 1$ se utiliza también en lugar de n debido a que produce una estimación más precisa de la correspondiente varianza poblacional (σ^2); o sea, un estimador insesgado de dicha varianza. No existe razón alguna para tener dos definiciones de la varianza muestral (s^2), una cuando el numerador se divide entre n, y otra cuando se divide entre $n - 1$, y debido a que esta última forma proporciona una medida satisfactoria de la varianza muestral, la usaremos en todo el texto.

Si aplicamos lo anterior a nuestros datos, tendremos:

$$s_A^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1} \qquad s_B^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

$$s_A^2 = \frac{106}{7} \qquad s_B^2 = \frac{10}{7}$$

$$\therefore\ s_A^2 = 15.14 \qquad \therefore\ s_B^2 = 1.42$$

Coeficiente de variación (CV)

Una de las ventajas de la desviación estándar, como medida de la variación, es que puede ser comparada con las desviaciones estándares de otros grupos, y permite concluir entonces cuál grupo de datos está más disperso que otro. No es posible interpretar la desviación estándar directamente como una medida descriptiva, ya que este valor estadístico es una medida relativa respecto a los datos que se analizan; por ejemplo, el valor que se calculó anteriormente, $s = 3.89$, por sí mismo no significa nada, pero si sabemos que se refiere a 8 medidas con un solo dígito, éstos van a tener más dispersión que si el mismo valor estuviera referido, por ejemplo, a 100 con dos dígitos. La medida que puede resolver este problema es el *coeficiente de variación* (CV), llamado de *Pearson*, que es la relación entre la desviación estándar y la media aritmética. Se multiplica además por 100, para considerar el resultado en forma porcentual (en %):

$$\text{CV} = \frac{s}{\bar{x}} \times 100$$

Por tanto:

$$\frac{s_A}{\bar{x}_A} = \frac{3.89}{5} = 0.778\ ; \qquad \frac{s_B}{\bar{x}_B} = \frac{1.19}{5} = 0.238$$

Al multiplicar estos resultados por 100 resulta finalmente:

$$(\text{CV})_A = 77.8\%$$

y

$$(\text{CV})_B = 23.8\%$$

Cuando se tiene un conjunto de datos muy grande y muchas de las puntuaciones se repiten, o sea, con cierta frecuencia (f), las fórmulas descritas

anteriormente de la desviación media (D_m), la desviación estándar (s^2) y la varianza (s^2), incluirán la frecuencia (f). Por tanto:

$$D_m = \frac{\sum_{i=1}^{n} f_i \left|(x_i - \bar{x})\right|}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{n-1}}$$

$$s^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{n-1}$$

Ejemplos para datos no agrupados

En el departamento de investigación de una distribuidora de frutas se compararon cuatro métodos de congelación y dos clases de conservadores. Se realizó un experimento con fresas, 8 bolsas con el mismo número de producto y de la misma cosecha. Después de 3 meses de congelamiento a 0°C, se midió la pérdida de color en una escala de 1 a 10 considerando la puntuación baja con menos decoloración y la puntuación alta con más pérdida de color. Se obtuvieron los datos siguientes:

Método 1		Método 2	
A	*B*	*A*	*B*
10	10	0.5	6.0
4.0	8.0	9.5	7.5
8.5	7.5	7.5	8.0
9.5	8.0	10	7.0
10	9.5	6.5	6.5
9.0	9.0	5.0	6.0
7.5	7.5	8.5	5.0
8.0	7.0	4.0	5.5
$\bar{x} = 8.31250$	$\bar{x} = 8.31250$	$\bar{x} = 6.43750$	$\bar{x} = 6.43750$
$s = 1.96282$	$s = 1.06695$	$s = 3.17847$	$s = 1.01550$
$s^2 = 3.85268$	$s^2 = 1.13839$	$s^2 = 10.10268$	$s^2 = 1.03125$
$CV = 23.613\%$	$CV = 12.836\%$	$CV = 49.374\%$	$CV = 15.775\%$

Método 3

A	B
1.0	3.0
7.5	5.5
4.5	4.0
5.0	4.5
2.0	3.0
2.5	3.5
3.5	4.0
6.0	4.5
$\bar{x} = 4.0000$	$\bar{x} = 4.0000$
$s = 2.17124$	$s = 0.84515$
$s^2 = 4.71429$	$s^2 = 0.71429$
$CV = 54.281\%$	$CV = 21.129\%$

Método 4

A	B
1.5	2.0
0.5	1.0
1.0	2.5
2.0	3.0
6.5	4.0
4.5	3.5
1.5	2.0
2.5	2.0
$\bar{x} = 2.5000$	$\bar{x} = 2.5000$
$s = 2.01778$	$s = 0.96362$
$s^2 = 4.07143$	$s^2 = 0.92857$
$CV = 80.711\%$	$CV = 38.545\%$

2.9 OTROS TIPOS DE PROMEDIO

2.9.1 Media ponderada $\bar{x}_P$

Se utiliza cuando las puntuaciones tienen pesos o ponderaciones (P) diferentes de la unidad.

Si $P = 1$, se tiene la media aritmética para datos no agrupados $\bar{x} = \frac{\Sigma x}{n}$; la fórmula para calcular la media ponderada es:

$$\bar{x}_P = \frac{\Sigma P x}{\Sigma P}$$

Ejemplo En un curso semestral de una determinada materia, se aplican tres exámenes parciales únicamente, no hay examen final, pero se le asignan los siguientes "pesos": $P_1 = 1$, $P_2 = 2$ y $P_3 = 3$.

Si un estudiante obtiene 10 en el primer examen, 7 en el segundo y 6 en el tercero, calcular su promedio.

como: $x_1 = 10$ $\quad P_1 = 1$

$x_2 = 7$ y $P_2 = 2$

$x_3 = 6$ $\quad P_3 = 3$

$$\bar{x}_P = \frac{(1)(10) + 2(7) + 3(6)}{1 + 2 + 3} = \frac{42}{6}$$

$$\boxed{\bar{x}_p = 7}$$

2.9.2 Media armónica (H)

Se utiliza cuando se desea promediar velocidades, cuando el tiempo se mantiene constante y las distancias varían o se mantienen (constante) y la velocidad varía.

Su fórmula es:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots \frac{1}{x_n}} = \frac{n}{\sum_{i=1} n \frac{1}{x_i}}$$

Ejemplo Si la distancia de A a B es de 1 km y se recorre a 60 $\frac{\text{km}}{\text{h}}$ y de B a A a 30 $\frac{\text{km}}{\text{h}}$, ¿cuál es su velocidad promedio?

$$\text{H} = \frac{2}{\frac{1}{60} + \frac{1}{30}} = \frac{2}{\frac{1+2}{60}} = \frac{2}{\frac{3}{60}} = \frac{120}{3} = 40\,\frac{\text{km}}{\text{h}}\ ; \quad \boxed{H = 40\,\frac{\text{km}}{\text{h}}}\ ;$$

La media aritmética sería $\bar{x} = \dfrac{60\,\frac{\text{km}}{\text{h}} + 30\,\frac{\text{km}}{\text{h}}}{2} = \dfrac{90}{2} = 45;$

$\boxed{\bar{x} = 45\,\frac{\text{km}}{\text{h}}}$ ¡lo cual es erróneo!

2.9.3 Media geométrica ($\bar{x}_g$)

Existe otro tipo de promedio que resulta de interés. Se trata de la *media geométrica*, definida como la raíz n-ésima del producto de n observaciones. Así, la media geométrica x_g, de n observaciones $x_1, x_2, \dots, x_n$ es

$$\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n}$$

Se emplea este promedio cuando se trabaja con observaciones en las que cada una guarda una razón aproximadamente constante respecto a la anterior, por ejemplo, al promediar tasas de crecimiento (aumento o disminución) de una población estadística, según se ilustra en el ejemplo siguiente.

Ejemplo El número de títulos *cum laude* otorgados por una universidad durante seis años consecutivos se presenta en la tabla que aparece en la página 85. ¿Cuál es el porcentaje medio de incremento en la cantidad de títulos otorgados por año?

Para la respuesta se calcula la media geométrica de las razones presentadas en la última columna. Esto es,

$$\sqrt[5]{1.2 \times 1.5 \times 1.67 \times 2.0 \times 1.67} = 1.585$$

es decir, un aumento medio por año de 58.5%.

Cabría preguntar, por qué no puede emplearse la media aritmética, que vale $\frac{1}{5}(1.2 + 1.5 + 1.67 + 2.0 + 1.67) = 1.61$ o sea, un incremento del 61%, que es mayor que el proporcionado por la media geométrica; ¿la media aritmética siempre es mayor que la geométrica? El sesgo en la respuesta dada por la media aritmética es resultado de la influencia de la magnitud absoluta de las razones. Por ejemplo, duplicar un valor representa una razón de 2, en tanto que dividirlo a la mitad origina una razón de $\frac{1}{2}$. De este modo, si consideramos un valor de 100 que desciende a 50 y poco después se eleva a 100, las razones serán $\frac{1}{2}$ y 2, respectivamente. La media geométrica es $\sqrt{\frac{1}{2} \times 2} = \sqrt{1} = 1$, que es la tasa media de incremento. Esta respuesta es correcta en términos intuitivos, dado que el cambio total registrado es nulo. Sin embargo, la media aritmética de las razones es $\frac{1}{2}\left(\frac{1}{2} + 2\right) = 1.25$. Si las razones fueran 3 y $\frac{1}{3}$, la media geométrica seguiría siendo 1, en tanto que la media aritmética sería $1 \times \frac{2}{3}$.

Año	*Número de títulos*	*Razón al valor del año anterior*
1959	5	–
1960	6	1.20
1961	9	1.50
1962	15	1.67
1963	30	2.00
1964	50	1.67

RESUMEN

Por lo general, la persona que realiza un estudio, una evaluación o investigación, desea contestar ciertas preguntas relacionadas con las características o los problemas de una población (conjunto de mediciones); esto casi siempre se reporta por medio de información numérica, o sea, por datos. El trabajo entonces es analizar dicha información, relacionarla con el problema que se desea resolver, y formular luego conclusiones.

Cuando es imposible o resulta caro, en términos de costo y tiempo, efectuar un análisis exhaustivo de todos los elementos que constituyen dicha población, se selecciona en forma aleatoria un subconjunto de tales mediciones (una muestra). Una vez efectuado lo anterior, se procede a calcular la medida de tendencia central más adecuada: la media aritmética ($\bar{x}$), la mediana (*Me*) o la moda (*Mo*). A continuación se complementan con las medidas de dispersión o variabilidad: la amplitud de variación (*A*), la desviación media (D_m), la desviación estándar (*s*), la varianza (s^2) y el coeficiente de variación (de Pearson) (*CV*).

Una vez obtenidas estas medidas para la muestra, se podrán estimar los parámetros correspondientes a la población de donde fue extraída la misma y, por consiguiente, generalizar estos resultados a la población.

En esta etapa de análisis estadístico ya estamos en condiciones de agrupar los datos obtenidos, en función de la frecuencia con que ocurren algunos de ellos, y empezar a construir un cuadro donde se concentren estos datos, que suelen llamarse *tablas de frecuencias*. Tal proceso puede llevarse a cabo sin pérdida de información, aun cuando en la mayoría de los casos la construcción de estas útiles distribuciones de frecuencias implica el agrupamiento en intervalos o clases, y en apariencia suelen perderse los valores exactos de los datos originales. Pero más tarde no sólo se recuperan dichos valores exactos, sino que incluso pueden generar información valiosa, que aparentemente no existe.

Por otro lado, las medidas de dispersión desempeñan un papel importante en el análisis estadístico, ya que nos proveen información del grado de agrupamiento que tienen los datos respecto a las medidas de tendencia central, básicamente de la media aritmética, lo que significa la validación de este estimador, ya que está en función del grado de dispersión de los datos.

Se consideran otros tipos de promedios, como la media ponderada, la armónica y la geométrica, que tienen diversas aplicaciones.

PROBLEMAS

2.1 Desarrolle cada uno de los incisos siguientes:

a) $\sum_{i=1}^{5} Yi$

b) $\sum_{k=3}^{6} X_k^2$

c) $\sum_{j=1}^{7} X_j Y_j$

d) $\sum_{i=1}^{3} (2Xi + 5)$

e) $\sum_{j=2}^{4} X_j^2 \; l_j$

2.2 Utilizando la notación de sumatoria, exprese cada uno de los incisos que se detallan en la página siguiente.

a) $Y_1^2 + Y_2^2 + Y_3^2 + \ldots + Y_{17}^2$

b) $2Z_1 + 2Z_2 + 2Z_3 + \ldots + 2Z_{100}$

c) $X_2\, Y_2^2 + X_3\, Y_3^2 + X_4\, Y_4^2 + X_5\, Y_5^2$

d) $(X_1^2 + 5)^3 + (X_2^2 + 5)^3 + (X_3^2 + 5)^3$

2.3 Si $X_1 = 0$, $X_2 = 2$, $X_3 = -1$, $X_4 = 3$ y $X_5 = 1$

Calcule:

a) $\sum_{i=1}^{5} X_i^2$ *b)* $\sum_{i=3}^{5} X_i$ *c)* $\sum_{i=1}^{5} X_i^3$

2.4 Dado: $X_1 = 2$, $X_2 = 5$, $X_3 = -2$, $Y_2 = -1$ y $Y_3 = 1$

Calcule:

a) $\sum_{k=1}^{3} X_k^2\, Y_k$

b) $\sum_{k=1}^{3} (4X_k + Y_k)$

c) $\sum_{k=1}^{3} (X_k + Y_k)^2$

d) $\left(\sum_{k=2}^{3} X_k\right)\left(\sum_{k=2}^{3} Y_k^2\right)$

2.5 Dado:

$$\sum_{j=1}^{10} X_j^3 = 91, \qquad \sum_{j=1}^{10} X_j^2 = 25 \qquad \text{y} \qquad \sum_{j=1}^{10} X_j = 7$$

Calcule:

$$\sum_{j=1}^{10} (5X_j^3 - 3X_j^2 + X_j + 22)$$

2.6 Con los datos siguientes
53, 52, 49, 56, 58,

calcule:
a) Media aritmética, mediana y moda, ($\bar{x}$, *Me* y *Mo*).
b) Desviación estándar, varianza, (s, s^2).
c) Coeficiente de variación (*CV*).

2.7 El departamento de investigación agropecuario confronta cuatro métodos de comparación y dos clases de conservadores naturales, por lo cual realiza un experimento con fresas; considera ocho bolsas con el mismo número de

la misma cosecha. Después de tres meses de congelamiento a 0°C, se mide la pérdida de color en una escala de 1 a 10, considerando la puntuación baja con menos decoloración y la alta con más pérdida de color. De los cuatro métodos de congelación (1, 2, 3, 4), y las dos clases de conservadores *A* y *B*, se obtienen los datos siguientes:

Método 1		Método 2	
A	*B*	*A*	*B*
10	10	0.5	6.0
4.0	8.0	9.5	7.5
8.5	7.5	7.5	8.0
9.5	8.0	10	7.0
10	9.5	6.5	6.5
9.0	9.0	5.0	6.0
7.5	7.5	8.5	5.0
8.0	7.0	4.0	5.5

Método 3		Método 4	
A	*B*	*A*	*B*
1.0	3.0	1.5	2.0
7.5	5.5	0.5	1.0
4.5	4.0	1.0	2.5
5.0	4.5	2.0	3.0
2.0	3.0	6.5	4.0
2.5	3.5	4.5	3.5
3.5	4.0	1.5	2.0
6.0	4.5	2.5	2.0

Realice un análisis estadístico para cada método y obtenga conclusiones de cada uno de ellos y una general, comparándolos.

2.8 En un experimento sobre aprendizaje y memoria espacial con nueve ratas de la cepa Wistar, aparentemente sanas y sin aprendizaje previo (de diferente peso), se les aplicó una misma dosis de escopolamina y se registró el tiempo que tardaron en atravesar un laberinto abierto. Se obtuvieron los siguientes datos:
1 min, 2.5 min, 3 min, 1.5 min, 2 min, 1.5 min, 1 min, 0.9 min, 30 min.

Cuando se realiza un análisis estadístico de datos en el que intervienen medidas temporales (minutos, horas), hay que homogenizarlos en un solo tipo de unidad, tomando en cuenta las fracciones antes de realizar el análisis y después regresar a las unidades originales: minutos, horas, días, etcétera.

a) ¿Cuál de las tres medidas de tendencia central es la más apropiada en este caso?

2.9 Al cambiarles la dieta de líquida a sólida, el peso de 20 bebés, en un mes, se incrementó de la siguiente forma:

0.4	1.9	1.5	0.9	0.3	1.6	0.4	1.4	1.2	0.8
0.9	0.7	0.9	0.7	0.9	1.5	0.5	1.5	1.7	1.8

Realice un análisis estadístico.

2.10 En una encuesta a 300 personas se les preguntó el número de hijos que tienen y se obtuvieron los datos siguientes:

x	f
0	78
1	90
2	32
3	25
4	50
5	15
6	10

Considere que x es una variable discreta. Haga su gráfica.

2.11 El departamento de ecología de la universidad adquiere 400 convertidores catalíticos cuya vida promedio se distribuye así:

Intervalo	f
300 — 399	14
400 — 499	46
500 — 599	58
600 — 699	78
700 — 799	68
800 — 899	62
900 — 999	48
1000 — 1088	22
1150 — 1199	4
	400

Realice un análisis estadístico y grafique utilizando el paquete MACSTAT.

2.12 Realice un análisis estadístico y grafique.

LIR — LSE	*f*
0 — 1.5	7
1.5 — 3.0	5
3.0 — 4.5	6
4.5 — 6.0	15
6.0 — 7.5	7
7.5 — 9.0	10

2.13 Dada la siguiente tabla de frecuencias, realice un análisis estadístico:

LRI — LRS	*f*
0.523 — 0.526	3
0.526 — 0.529	6
0.529 — 0.532	7
0.532 — 0.535	12
0.535 — 0.538	10
0.538 — 0.541	6
0.541 — 0.544	4
0.544 — 0.547	2

2.14 Realice un análisis estadístico y grafique:

	f
de 120 a < 140	2
de 140 a < 160	23
de 160 a < 180	75
de 180 a < 200	73
de 200 a < 220	22
de 220 a < 240	18
de 240 a < 260	4
de 260 a < 280	8
de 280 a < 330	8

a) Calcule la media, la mediana y la moda.
b) Trace el polígono de frecuencias de la distribución.

2.15

Ingresos anuales	f
Más de 18 000	5
15 000 — 17 999	15
14 000 — 14 999	80
13 000 — 13 999	40
12 000 — 12 999	25
11 000 — 11 999	30
10 000 — 10 999	18
9 000 — 9 999	10
8 000 — 8 999	6
7 000 — 7 999	6
Menos de 7 000	4

a) Realice un análisis estadístico y sus gráficas.
b) Halle los cuantiles Q_1 y Q_3.

2.16 A los operadores de una línea de autobuses foráneos, como parte de un programa de seguridad para ellos y los usuarios, se les aplican algunos exámenes para conocer su estado de salud física y mental. Una de dichas pruebas consiste en evaluar sus reflejos, o sea, determinar el tiempo de reacción ante la presencia sorpresiva de algún obstáculo.

A una muestra de 34 conductores, mayores de 40 años, se les aplicó dicha prueba y se obtuvieron los datos siguientes.
Sea x = Tiempo de reacción en segundos.

x	0.40	0.42	0.43	0.46	0.47	0.48	0.49	0.50	0.51	0.52	0.53
Frecuencia	1	2	3	2	2	5	6	4	2	3	3
x	0.55										
f	1										

Realice un análisis estadístico y obtenga conclusiones.

2.17 El gobierno de un estado realiza un programa de concientización sobre el uso racional del agua, en una colonia de clase media.

Para conocer la efectividad de dicho programa, se realizó un muestreo a familias de cuatro integrantes y se observó la reducción de consumo de

agua. El número de familias encuestadas fue de 40 y se obtuvieron los datos en m^3. Las observaciones se muestran en la tabla siguiente.

2.2	3.5	3.2	3.0
3.4	3.1	3.8	4.7
2.5	3.4	2.9	3.9
3.3	3.7	3.2	1.9
4.7	3.2	3.9	4.2
4.1	4.5	3.7	2.6
1.6	3.3	3.1	3.7
4.3	3.6	3.3	3.1
3.1	4.4	4.1	3.4
3.8	2.6	3.0	3.5

Realice un análisis estadístico y sus gráficas, utilizando el paquete MACSTAT.

2.18 Dentro del mismo programa, sobre el uso racional del agua, en dicha entidad se capacitó a los alumnos de una escuela preparatoria que cursaban la materia de ecología, sin considerar el número de integrantes por familia. Se obtuvieron los datos siguientes (con una $n = 60$).

3.6	1.9	2.1	0.3	0.8	0.2
1.0	1.4	1.8	1.6	1.1	1.8
0.3	1.1	0.5	1.2	0.6	1.1
0.8	1.7	1.4	0.2	1.3	3.1
0.4	2.3	1.8	4.5	0.9	0.7
0.6	2.8	2.5	1.1	0.4	1.2
0.4	1.3	0.8	1.3	1.1	1.2
0.8	1.0	0.9	0.7	3.1	1.7
1.1	2.2	1.6	1.9	5.2	0.5
1.8	0.3	1.1	0.6	0.7	0.6

Realice un análisis estadístico y compare con los resultados del problema anterior, utilizando MACSTAT.

2.19 Con los siguientes datos, realice un análisis estadístico utilizando el paquete MACSTAT:

79	97	86	76
93	87	98	68
84	88	81	91
86	87	70	94
77	92	66	95
63	68	98	88
46	72	59	79

2.20 En una sucursal bancaria donde trabajan 25 personas (desde gerente, ejecutivos de cuenta, administrativos y cajeros), la hora de salida a comer,

en día normal de actividades, es a las 3:00 pm, y se registran las siguientes horas de salida:

2.50	3.22	4.15
2.55	3.22	3.25
3.10	4.05	2.58
3.07	3.05	3.18
3.50	3.38	3.19
3.00	3.30	3.24
3.00	4.00	3.25
3.20	4.30	3.32
3.22		

Calcule el promedio de la hora de salida, desviación estándar, mediana y moda.

2.21 Dados los 70 datos siguientes, realice un análisis estadístico con diferente número de intervalos ($k = 8$, $k = 10$, $k = 12$); utilice el paquete estadístico MACSTAT:

0.65	0.99	1.25	1.63	1.74	1.86	2.08
0.65	1.00	1.28	1.64	1.75	1.88	2.09
0.72	1.05	1.37	1.64	1.75	1.90	2.10
0.73	1.09	1.40	1.67	1.75	1.92	2.11
0.75	1.11	1.47	1.68	1.79	1.92	2.17
0.85	1.11	1.47	1.68	1.79	1.93	2.28
0.85	1.15	1.51	1.69	1.82	1.97	2.31
0.85	1.20	1.58	1.69	1.83	2.00	2.37
0.93	1.21	1.60	1.70	1.85	2.03	2.46
0.99	1.24	1.60	1.72	1.85	2.05	2.55

2.22 En un estudio, 20 pacientes presentaban lesiones de tipo vesicular, la cual generalmente tiene repercusión en la cavidad bucal, tomando en cuenta la edad, sexo, raza, localización de la lesión en la cavidad bucal (lengua, paladar, piso de boca, mucosa yugal, labios, tejido gingival), las repercusiones sistemáticas (zonas cutáneas y mucosas en general), la frecuencia de recidiva y la duración de la lesión dada en días (los datos se muestran en el siguiente problema). Realice un análisis estadístico.

Núm. de caso	Edad (años)	Sexo	Raza	Sitio de la lesión	Lesión extraoral	Frecuencia de recidiva	Duración de la lesión (días)
1	20	0	Cauc.	0	1	1	4 (365 días)
2	25	1	Cauc.	0	1	1	2
3	40	0	Cauc.	3	1	1	2
4	38	0	Cauc.	1	1	1	0
5	27	1	Cauc.	2	0	1	1
6	22	1	Negro	1	0	0	3
7	20	0	Cauc.	1	1	1	4
8	26	0	Cauc.	3	1	1	0
9	34	0	Cauc.	3	0	0	0
10	35	0	Cauc.	4	1	1	1
11	39	1	Cauc.	5	1	1	2
12	40	1	Cauc.	5	0	1	1
13	38	1	Cauc.	2	0	0	0
14	36	1	Cauc.	1	0	1	4
15	28	0	Cauc.	0	1	0	3
16	25	1	Cauc.	2	1	1	2
17	22	0	Cauc.	2	0	1	0
18	20	0	Cauc.	3	0	1	1
19	24	1	Cauc.	1	1	0	4
20	35	0	Cauc.	4	1	0	3

Las claves en que se basó la tabla anterior son las siguientes:

Sexo:
masculino = 0
femenino = 1

Recidiva:
no = 0
sí = 1

Sitio de la lesión:
lengua = 0
paladar = 1
piso de boca = 2
mucosa yugal = 3
labios = 4
encías = 5

Lesión extraoral:
zonas cutáneas = 0
mucosas = 1

2.23 Los datos de la tabla que se presenta en la página siguiente se obtuvieron de una clínica odontológica. Realice un análisis estadístico de cada variable de la tabla, ya sea por datos agrupados o no agrupados.

En una investigación se tomó en cuenta a 20 pacientes atendidos en la Clínica de Endodoncia de Posgrado de la Facultad de Odontología de la UNAM.

A cada paciente se le tomaron datos personales, como edad, sexo, dientes, número de conductos de la pieza tratada, estado pulpar del diente, tipo de irritante que motivó la lesión de la pulpa, técnica de obturación que se utilizó en el tratamiento y otras modificaciones que se les hicieron, según lo indicaba el caso.

Se debe hacer, en cada caso obtenido, su propia estadística ya sea con datos agrupados o no, según sea conveniente.

Sexo	Edad	Diente por tratar	Número de conductos	Vitalidad	Antecedentes del diente por tratar	Tratamiento
1	58	1	1	0	4	1
1	56	1	1	1	0	0
0	22	0	1	0	1	2
1	37	0	1	1	1	3
0	27	2	3	1	3	1
1	55	2	3	0	2	1
1	19	0	1	1	1	2
1	20	1	2	0	2	3
0	48	2	3	0	3	1
1	28	2	3	1	0	1
1	23	0	1	1	1	2
0	28	2	3	1	3	1
0	48	1	2	0	2	2
1	21	2	2	1	3	1
1	48	1	1	0	1	0
0	18	0	1	1	1	3
0	33	0	1	1	1	0
0	12	2	3	1	3	1
1	21	1	2	0	2	1
1	59	0	1	0	0	3

Nota:

Sexo: masculino = 0
femenino = 1

Diente por tratar:
anterior = 0
premolar = 1
molar = 2

Vitalidad:
vital = 0
no vital = 1

Antecedentes del diente por tratar:
caries = 0
resina = 1
amalgama = 2
incrustación = 3
otros = 4

Tratamiento:
recubrimiento = 0
tratamiento de conductos = 1
cirugía periapical = 2
otros = 3

Realice el análisis estadístico adecuado.

2.24 Dados los siguientes datos muestrales:

11	1
6	15
−1	−4
0	
10	
−2	

determine media, mediana, moda, desviación estándar, varianza y coeficiente de variación.

2.25 En un país de Latinoamérica, los censos señalan el número de habitantes (millones) y su tasa de crecimiento (porcentaje) de 1940 a 1990. Calcule el promedio de la tasa de crecimiento.

Año	Habitantes (millones)	Tasa de crecimiento (porcentaje)
1940	20.3	—
1950	28.7	41.4
1960	40.5	41.2
1970	60.0	48.2
1980	73.8	23.3
1990	95.4	29.3

2.26 Un entrenador de perros compra alimento especial, pero en los últimos cuatro meses, los precios por kilogramo han aumentado, pero la cantidad que él compra permanece constante. Se obtuvo la siguiente información:

	1er. mes	2o. mes	3o. mes	4o. mes
Precio en pesos por kg	1.00	1.25	1.60	2.00

¿Cuál es el costo promedio del alimento, si el entrenador gastó 1 000 pesos por mes?
Realice un análisis estadístico.

2.27 En un semestre lectivo, el maestro aplica 3 exámenes parciales y uno final; el valor que le asigna el profesor a éstos es que el final equivale 2 veces más que los parciales. Si un estudiante obtiene las siguientes calificaciones: 6.5, 7.0, 8.5, y 9 en el examen final, calcule el promedio en el semestre.

Conjuntos, funciones y matrices 3

PROPÓSITOS

El objetivo central del presente capítulo es que el lector pueda comprender, manejar y solucionar situaciones relacionadas tanto con las abstracciones y los significados, como con las propiedades y los axiomas correspondientes a la teoría de conjuntos, las funciones y las matrices.

De igual modo, el lector podrá:

1. Reconocer a la teoría de conjuntos como posibilidad de sistematización de información relevante.
2. Desarrollar el razonamiento lógico deductivo en el análisis de situaciones complejas.
3. Manejar lenguaje matemático con suficiente precisión y conciencia de los significados presentes en los conjuntos, funciones y matrices.
4. Explicar los diagramas de Venn-Euler y Lewis Carroll.
5. Relacionar los siguientes elementos que corresponden a los diagramas de los autores citados: conjunto, elemento y complemento; adicionalmente, los símbolos $\subset$, $\supset$, $\varnothing$, $x|x$, $\{\ \}$, entre otros.
6. Desarrollar propiedades y operaciones entre conjuntos.
7. Explicar los procesos matemáticos del razonamiento de las propiedades y operaciones entre conjuntos.
8. Desarrollar la técnica de conteo de elementos.
9. Explicar el producto cartesiano.
10. Graficar la localización de puntos en un plano cartesiano.

11. Detallar las nociones de función y relación, así como los términos que las caracterizan.
12. Explicar los conceptos de dominio y contradominio en relaciones y funciones.
13. Reconocer y desarrollar los diferentes tipos de funciones: algebraicas (constante, lineal, cuadrática, cúbica y valor absoluto), trascendentes (exponencial, logarítmicas).
14. Desarrollar el álgebra de funciones.
15. Enunciar el concepto de matriz y sus propiedades.
16. Identificar los diferentes tipos de matrices.
17. Explicar qué es un determinante.
18. Identificar las operaciones que caracterizan el álgebra de matrices.
19. Calcular las diferentes operaciones del álgebra de matrices; por ejemplo, transpuestas de matriz, matriz simétrica, operaciones con matrices, suma algebraica, multiplicación por un escalar, inversa aditiva y producto.
20. Comprender el sistema de ecuaciones lineales.
21. Desarrollar el proceso de solución correspondiente a cada sistema de ecuaciones lineales: solución única, número infinito de soluciones y el insoluble.
22. Solucionar ejercicios propios de los temas abordados en este capítulo: conjuntos, relaciones y funciones, conteo de elementos, producto cartesiano, matrices, así como sistema de ecuaciones lineales.
23. Aplicar los conocimientos adquiridos en la solución de problemas cotidianos y describirlos de manera adecuada.
24. Establecer la continuidad con temas que se abordarán en capítulos siguientes: probabilidad y procesos estocásticos, entre otros.

INTRODUCCIÓN

No obstante que el estudio de conjuntos, funciones y matrices no siempre forma parte de un curso de estadística, tales temas proveen los conceptos fundamentales que se utilizan en esta disciplina.

La teoría de conjuntos es un sistema con un lenguaje específico para el manejo de ciertos problemas, un instrumento adecuado para la sistematización de la información relevante que nos permite enfocar un problema en su totalidad, deslindando en él lo que es fundamental; asimismo, dicho sistema activa nuestra capacidad de análisis, facilitando la visualización de las probables relaciones que existan entre un problema y una solución propuesta.

Dada la metodología propia de los conjuntos y su razonamiento deductivo-inductivo, nos es posible hacer un análisis lógico, de situaciones complejas, así como sistematizar nuestra capacidad analítica en el manejo de una información determinística (no probabilística).

En general, la teoría de conjuntos facilitará en gran medida la comprensión de la teoría de la probabilidad, los procesos estocásticos y el muestreo, que son básicos, tanto en su razonamiento como en su mecánica de cálculo, para una mejor comprensión y aplicación de la estadística.

3.1 CONJUNTOS, CONCEPTOS Y NOTACIÓN

3.1.1 Conjunto

Es tan común en nuestra vida diaria el uso de los conjuntos que resulta fácil captar intuitivamente su noción, diciendo que *conjunto* es una colección definida de objetos, personas, animales o ideas y que por lo general se representa con letras mayúsculas: A, B, C, etc., encerradas entre paréntesis o llaves.

Ejemplo $A = \{a, e, i, o, u\}$ se lee: el conjunto A con los elementos a, e, i, o, u.
$F = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ el conjunto F de dígitos.

3.1.2 Elemento

Se llama *elemento* a cada uno de los miembros de un conjunto determinado; así, en $A = \{a, e, i, o, u\}$ a, e, i, o, u son elementos de A.

3.1.3 Reglas y formas para enunciar los conjuntos

Para enunciar correctamente los conjuntos hay que tomar en cuenta las reglas siguientes:

a) La colección ha de estar bien definida; no se debe dar cabida a confusiones, que un elemento esté y no esté en el conjunto. Debe ser preciso, por ejemplo: sea B el conjunto de los días de la semana, B = {domingo, lunes, martes, miércoles, jueves, viernes, sábado}.

b) Los elementos no se repiten. En un conjunto no deben aparecer dos elementos iguales, por ejemplo: sea D el conjunto de las letras de la palabra *Guadalajara:* $D = \{G, u, a, d, l, j, r\}$.

c) El orden de los elementos es irrelevante. Los elementos pueden aparecer en cualquier orden, por ejemplo: sea M el conjunto de las vocales, $M = \{a, i, o, u, e\}$; $M = \{u, i, o, a, e\}$, etcétera.

Para representar un conjunto podemos utilizar distintas formas:

- Enumerando sus elementos, se enuncian uno a uno los elementos del conjunto

$$H = \{1, 2, 3, 4, 5\}; \quad E = \{a, e, i, o, u\}.$$

- Describir una característica de manera verbal que englobe a todos los elementos: A = {los meses del año}; S = {los números naturales pares}. Sabemos que los elementos de A son: enero, febrero, marzo, abril, mayo, junio, julio, agosto, septiembre, octubre, noviembre, diciembre.
- Mediante notación matemática o al utilizar alguna propiedad definitoria:

$G = \{x | x$ sea un número natural donde N es el conjunto de los números naturales$\}$.[1]

[1] $\{x | x \in N\}$ Se lee: x tal que x pertenece al conjunto de los números naturales.

$J = \{y \mid y > 5,\ y < 10\}$, se lee: el conjunto J es el conjunto de las y tal que y sea mayor que 5 y menor que 10.

Existen conjuntos que admiten expresarse en más de una forma; sin embargo, otros sólo se pueden expresar de una forma:

$J = \{y \mid y > 5,\ y < 10\}$

3.1.4 Conjuntos finitos e infinitos

Para nuestros fines, un conjunto finito será el que conozcamos con exactitud los elementos que contiene, es decir, posee cardinalidad.[2]

Ejemplo $E = \{2, 4, 8, 10, 12\}$, E es un conjunto finito, tiene cardinalidad = 5
$J = \{y \mid y$ sea una letra de *papá*$\}$ J es un conjunto finito, tiene cardinalidad = 4.

Los conjuntos infinitos son aquellos de los cuales no se conoce su cardinalidad por ser demasiado extensos.

Ejemplo P = {las estrellas}, P es un conjunto infinito.
O = {la arena del mar}, O es un conjunto infinito.

3.2 RELACIÓN DE PERTENENCIA

Esta relación se da entre los elementos y el conjunto; se representa con el signo $\in$, que significa "*pertenece a*", o "*es elemento de.*" Sea el conjunto $A = \{a, e, i, o, u\}$ se tiene que A se lee: *el elemento a pertenece a A*, así también $e \in A$, $i \in A$, o $\in$ A, $u \in A$, en el caso contrario, cuando un elemento no pertenece, se cruza $\in$ ($\notin$), ejemplo $b \notin A$, significa "b no pertenece a A".

3.3 RELACIONES ENTRE CONJUNTOS

3.3.1 De inclusión

Esta relación ocurre cuando un conjunto contiene a otro conjunto y se denota con $\subset$, que significa *está incluido en* o *está contenido en.* Por ejemplo, sean los siguientes conjuntos A, B, C, que forman parte del mismo conjunto universal:

$$A = \{a, e, i, o, u\},\ B = \{a, e, o\} \text{ y } D = \{i, u\}$$

se observa que $B \subset A$: El conjunto B está incluido en el conjunto A.
se observa que $D \subset A$: El conjunto D está incluido en el conjunto A.

---

[2] Cardinalidad es el número de elementos que contiene un conjunto.

Este símbolo de inclusión se puede invertir y se tiene lo siguiente:

$A \supset B$ el conjunto A incluye al conjunto B.
$A \supset D$ el conjunto A incluye al conjunto D.

Cuando un conjunto no está incluido o no incluye a otro conjunto, simplemente se utiliza $\not\subset$ o $\not\supset$, por ejemplo:

$D \not\subset B$ el conjunto D no está incluido en el conjunto B.
$B \not\supset D$ el conjunto B no incluye al conjunto D.

Debemos tener cuidado de no confundir la relación de pertenencia con la de inclusión, o sea:

$a \subset A$ y $B \in A$ son errores
lo correcto es $a \in A$ y $B \subset A$

3.3.1.1 Subconjuntos

La relación de inclusión sólo es posible cuando los elementos del conjunto B están contenidos en el conjunto A. Luego entonces, B es subconjunto de A. $B \subset A$: el conjunto B es subconjunto del conjunto A, o simplemente, B es subconjunto de A. Esto significa que todos los elementos del conjunto B son elementos del conjunto A. Sea $B \subset A$, donde:

$$A = \{a, e, i, o, u\} \text{ y } B = \{a, e, o\}.$$

Existe un caso particular de la relación de inclusión, cuando todos los elementos del subconjunto también pertenecen al conjunto que lo contiene.

Ejemplo $A = \{1, 2, 4, 5, 6\}$ y $B = \{2, 4, 1, 5, 6\}$ cuando esto ocurre, se dice que A es subconjunto propio de B, o también que B es subconjunto propio de A. Puesto que $A \subset B$, y $B \subset A$, esto sólo es válido si $A = B$.

3.3.1.2 Número de subconjuntos de un conjunto

Si tomamos al conjunto $N = \{1, 2, 3, 4\}$ y se desean obtener todos los subconjuntos posibles, se obtendrá:

$$\{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\},$$
$$\{1, 2, 3\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}, \{\ \}\}.$$

El número de subconjuntos obtenidos es de 16. En general, el número de subconjuntos de un conjunto dado se obtiene aplicando: 2^n, donde n es el número de elementos del conjunto dado. Sean conjuntos con:

1 elemento, se tendrán 2 subconjuntos = 2^1
2 elementos, se tendrán 4 subconjuntos = $4 = 2^2$

3 elementos, se tendrán 8 subconjuntos = 8 = 2^3
4 elementos, se tendrán 16 subconjuntos = 16 = 2^4
y así sucesivamente, considerando que el número de subconjuntos del conjunto vacío es uno, esto es porque $2^0 = 1$.

3.3.2 Conjuntos ajenos

Si el conjunto A no tiene ningún elemento común con el conjunto B, se dice que A y B son conjuntos ajenos (disjuntos). A) (B.[3]
Sea:

$$A = \{a, e, i, o, u\} \text{ y } B = \{1, 2, 3, 4, 5, 6\} \text{ son ajenos.}$$

3.3.3 Conjunto universal

Es aquel conjunto que agota a todos los elementos con una característica común y se denota por U; también se utiliza el símbolo Ω. Además, el conjunto universal suele ser tomado arbitrariamente. Así, al hablar del conjunto de las vocales, se antoja pertinente llamar *conjunto universal* a las letras del abecedario, o si hablamos de los dígitos, entonces el conjunto universal es el de los enteros positivos.

Ejemplo Sea $U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $E = \{1, 3, 5, 7, 9\}$, $H = \{0, 2, 4, 6, 8\}$
se tendrá que: $E \subset U$
$H \subset U$ o bien $U \supset E$, y
$U \supset H$, respectivamente.

3.3.3.1 Conjunto vacío

Este conjunto es el que no contiene elementos y se denota por

$\varnothing$ o por { }, se define a continuación:
$A = \{x \mid x \neq x\}$ entonces, podemos decir que el conjunto A es vacío.

3.3.4 Conjuntos iguales

Se dice que dos conjuntos A, B son iguales ($A = B$), si y sólo si constan de exactamente los mismos elementos. Por ejemplo $M = N$, si:

$$M = \{1, 2, 3, 4, 5\} \text{ y } N = \{z \mid 1 \leq z < 6\}$$

Obsérvese que los elementos del conjunto N son aquellos números mayores que cero y menores que seis. Entonces, los elementos únicos que cumplen las dos condiciones serán: 1, 2, 3, 4, 5 y por tanto $M = N$.

[3] Algunos autores utilizan esta notación para conjuntos ajenos.

En general, dos conjuntos, A y B, son iguales si cumplen la siguiente condición: $A = B$ si y sólo si, el número de elementos (cardinalidad) es igual y sus elementos son los mismos sin importar el orden.

3.3.4.1 Conjuntos similares

Son aquellos que tienen la misma cardinalidad pero con elementos diferentes.

Ejemplo $A = \{a, b, c\}$, $B = \{2, 4, 6\}$ $A \approx B$

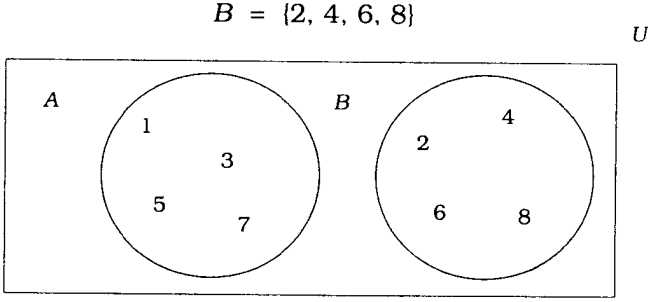
3.4 DIAGRAMAS DE VENN-EULER Y DE CARROLL

En el estudio de los conjuntos, el uso de los diagramas ha permitido explicar de manera gráfica esta importante herramienta matemática, gracias a los trabajos realizados por el matemático inglés John Venn y por el matemático suizo Euler. Estos diagramas representan a los conjuntos y a los subconjuntos, así como a las operaciones que con ellos se realicen.

Estos diagramas se representan por medio de un rectángulo, que indica el conjunto universal U, por ser este conjunto el que agota a todos los elementos con una misma cualidad, y dentro del rectángulo los conjuntos que sea necesario expresar, así como las operaciones por realizar.

Ejemplo

$$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$
$$A = \{1, 3, 5, 7\}$$
$$B = \{2, 4, 6, 8\}$$



Ejemplo B = Conjunto formado por todos los limítrofes (considerando el coeficiente de inteligencia).

CI entre 70 y 79 (según el *test* Wais)

A = Conjunto formado por todas las personas inferior a lo normal.

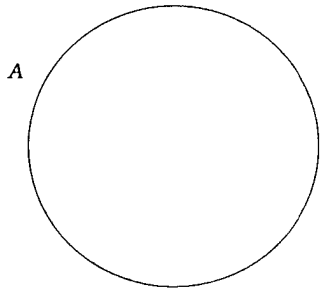
CI entre 0 y 89 (según el *test* Wais)

Por tanto, el conjunto B es subconjunto del conjunto A.

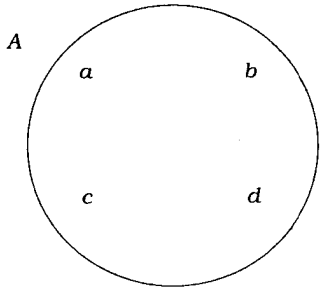
$$B \subset A, \text{ o bien, } A \supset B$$

Esto significa que B es un subconjunto de A; o sea, todos los elementos de B están en A, y sólo algunos de A están en B. En caso de que A fuera igual que B se cumple que $A \subset B$ y $B \subset A$ (siempre que $A = B$, esto es, que los dos contengan los mismos elementos); en consecuencia, la igualdad de conjuntos (=) significa que dos conjuntos, A y B, son iguales ($A = B$) si $A \subset B$ y $B \subset A$.

Los diagramas de Venn-Euler son figuras geométricas cerradas que representan un conjunto dado; por ejemplo, para representar a cualquier conjunto A se tendría:



Si $A = (a, b, c, d)$

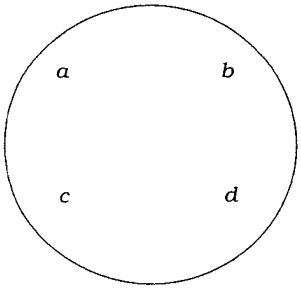


Como vemos, a es elemento de A, así como b, c, d; por tanto:

$$a \in A \qquad c \in A$$
$$b \in A \qquad d \in A$$

Como observamos, todas las demás letras del alfabeto no son elementos de A, pero en este caso forman el complemento de nuestro conjunto universal (Ω):

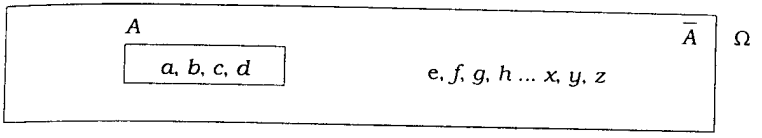
Conjunto A



Conjunto universal (Ω) o U

$a, b, c, d, \ldots, x, y, z$

Por tanto, la representación gráfica será:



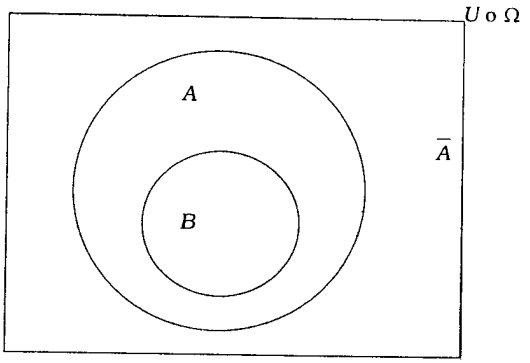
donde el conjunto $A = \{a, b, c, d\}$ y el complemento $\overline{A} = \{e, f, g, h \ldots, x, y, z\}$.

Este conjunto se denota usualmente con el nombre del conjunto y el símbolo: $(\bar{v})$, $(^c)$, $(')$, (c).

En nuestro caso: $(\overline{A})$, (A^c), (A'), (CA).

Por lo general, el universo se representa por un rectángulo.

La representación gráfica del ejemplo de subconjunto será:



donde:

Ω = Todas las personas a las que se aplicó el examen de inteligencia.

A = Las personas que tienen un *CI* entre 0-89.

B = Las personas que tienen un *CI* entre 70-79.

Los diagramas de Carroll (L. Carroll, autor de *Alicia en el país de las maravillas*), también llamadas *tablas de contingencia* o *doble entrada*, tienen una ventaja sobre los diagramas de Venn-Euler ya que pueden contener más de tres conjuntos y permiten visualizar sus relaciones entre ellos muy claramente.

Ejemplo:

		Escolaridad				
		2o. semestre	*4o. semestre*	*6o. semestre*	*8o. semestre*	*Total*
Sexo	Masculino	10	5	12	14	41
	Femenino	12	15	18	14	59
Total		22	20	30	28	100

10 estudiantes del sexo masculino y en el segundo semestre, 12 del sexo femenino y en el mismo semestre, etcétera.

3.4.1 Conjunto complemento

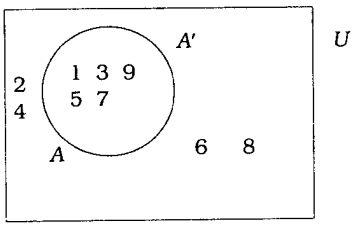
Se llama *conjunto complemento de A* a aquel conjunto $\bar{A}$ con elementos que pertenecen al conjunto universal y no pertenecen al conjunto A.

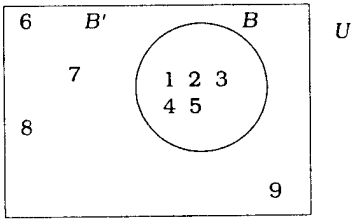
Ejemplo Sea $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{1, 3, 5, 7, 9\}$

$$A' = \{x \mid x \in U, x \notin A\} = \{2, 4, 6, 8\}$$

$$B = \{1, 2, 3, 4, 5\} \quad B' = \{6, 7, 8, 9\}$$

Expresado en diagramas de Venn-Euler tendríamos:





Algunos autores consideran al conjunto complemento como una operación entre conjuntos: la complementación.

3.5 ÁLGEBRA DE CONJUNTOS

Quizás una de las aplicaciones más importantes de los conjuntos es el álgebra de los mismos, es decir, las operaciones que se realizan con éstos aplicando los principios que el álgebra utiliza en el sistema de los números reales. Veamos las operaciones siguientes.

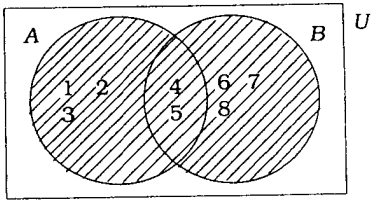
3.5.1 Unión ($\cup$)

La unión de dos conjuntos A y B (denotada $A \cup B$) es aquel conjunto formado por los elementos que pertenecen a A o pertenecen a B, o a ambos. También suele llamársele *suma lógica* y se representa de la manera siguiente:

$$A \cup B = \{x \mid x \in A \text{ o } x \in B \text{ o ambos}\}$$

se lee: el conjunto A unión B es aquel conjunto de las x tal que x pertenece a A o x pertenece a B, o a ambos.

Ejemplo $A = \{1, 2, 3, 4, 5\}$, $B = \{4, 5, 6, 7, 8\}$, $A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\}$
Mediante el diagrama de Venn-Euler quedaría:



$A \cup B$

Se recomienda observar en el diagrama que cuando se realiza una operación, el resultado debe achurarse (cruzarlo con rayas diagonales tenues).

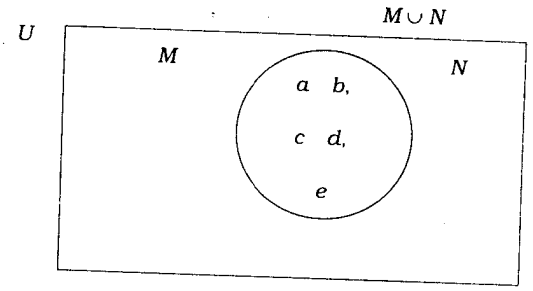
CASOS

Unión de conjuntos iguales

$$\text{Sea } M = \{a, b, c, d, e\},\ N = \{b, c, d, e, a\},\ M \cup N = \{a, b, c, d, e\}$$

Obsérvese que la unión de conjuntos iguales es un conjunto igual a cualesquiera de ellos, pues hay que recordar que los elementos de un conjunto no deben repetirse.

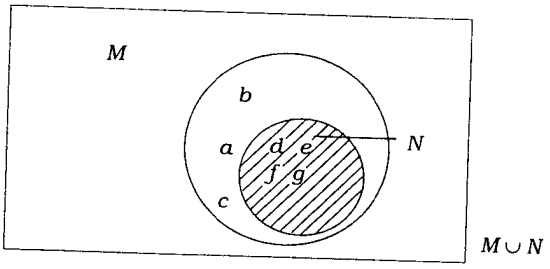
El diagrama de Venn-Euler será:



Unión de un subconjunto y un conjunto

$$\text{Sea } M = \{a, b, c, d, e, f, g\} \quad \text{y} \quad N = \{d, e, f, g\}$$

entonces la unión será: $M \cup N = \{a, b, c, d, e, f, g\}$, por lo que la unión de un conjunto con un subconjunto es igual al mismo conjunto:



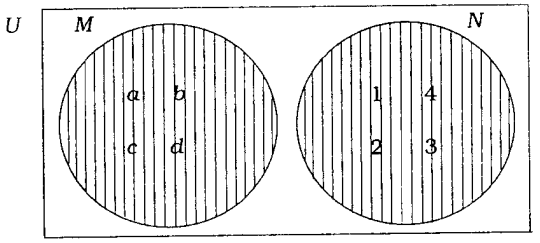
Unión de conjuntos ajenos (disjuntos)

$$\text{Sea } M = \{a, b, c, d\} \text{ y } N = \{1, 2, 3, 4\}$$

la unión es $M \cup N = \{a, b, c, d, 1, 2, 3, 4\}$

Obsérvese que la unión está formada por todos los elementos de los dos conjuntos; esto significa que la cardinalidad del conjunto unión es la suma aritmética

de la cardinalidad de cada uno de los conjuntos que la conforman: $n(M) = 4$ y $n(N) = 4$, por lo que $n(M \cup N) = 8$.



$M \cup N$

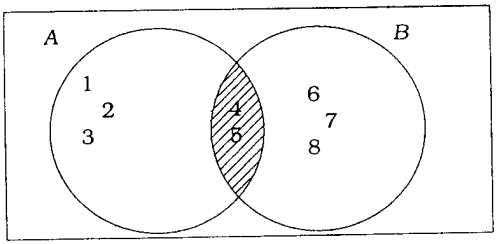
3.5.2 Intersección ($\cap$)

La intersección de dos conjuntos A y B, $(A \cap B)$ es el conjunto formado por los elementos que pertenecen a A y a B simultáneamente. También suele llamarse *producto lógico de dos conjuntos* y se representa de la manera siguiente:

$$A \cap B = \{x \mid x \in A \text{ y } x \in B\}$$

Este nuevo conjunto es la intersección de A y B, formado por todos los elementos que pertenecen a A y a B simultáneamente.

Ejemplo Sea $A = \{1, 2, 3, 4, 5\}$ y $B = \{4, 5, 6, 7, 8\}$, entonces la intersección será: $A \cap B = \{4, 5\}$, representada con el diagrama siguiente:



$A \cap B$

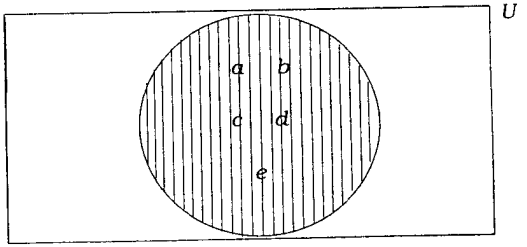
CASOS

Intersección de conjuntos iguales

$$\text{Sea } M = \{a, b, c, d, e\} \text{ y } N = \{b, c, d, e, a\}$$

La intersección será $M \cap N = \{a, b, c, d, e\}$.

La intersección de dos conjuntos iguales es igual a un solo conjunto. El diagrama correspondiente será:



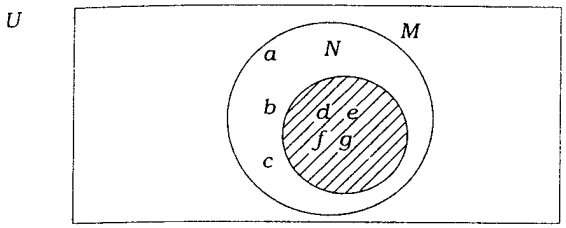
$M \cap N$

Intersección de un conjunto y un subconjunto

$$\text{Sea } M = \{a, b, c, d, e, f, g\} \text{ y } N = \{d, e, f, g\}$$

La intersección es $M \cap N = \{d, e, f, g\}$.

La intersección de un conjunto y uno de sus subconjuntos es igual al subconjunto dado.



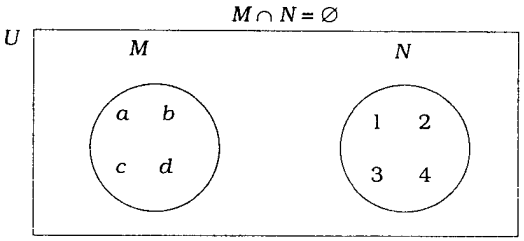
$M \cap N = N$

Intersección de conjuntos ajenos (disjuntos)

$$\text{Sea } M = \{a, b, c, d\} \text{ y } N = \{1, 2, 3, 4\}$$

La intersección es $M \cap N = \varnothing$.

La intersección de conjuntos disjuntos o ajenos es el conjunto vacío, porque no existen elementos comunes.



3.5.3 Diferencia o resta aritmética ($A - B$)

La diferencia de dos conjuntos A y B es el conjunto formado por los elementos que pertenecen a A y no pertenecen a B. También se le suele llamar *complemento relativo*. Su definición es:

$A - B = \{ x \mid x \in A \text{ y } x \notin B\}$. "El conjunto A menos B es el conjunto de las x tal que x pertenece a A y no pertenece a B".

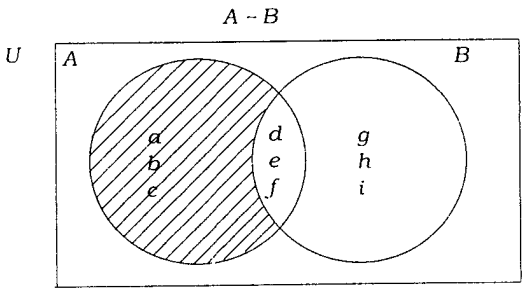
"El conjunto B menos el conjunto A es el conjunto formado por todos los elementos que están en B, pero que no estén en A", $A - B \neq B - A$, o sea:

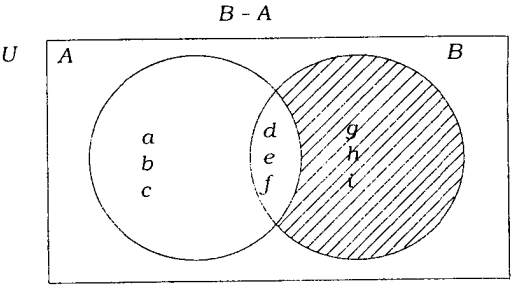
$B - A = \{ x \mid x \in B \text{ y } x \notin A\}$, por lo que se concluye que $A - B \neq B - A$ no cumple la propiedad conmutativa.

Ejemplo Sean dos conjuntos A y B, tales que:

$$A = \{a, b, c, d, e, f\} \text{ y } B = \{d, e, f, g, h, i\}$$
$$A - B = \{a, b, c\}$$
$$B - A = \{g, h, i\}$$

Los diagramas correspondientes serán:





CASOS

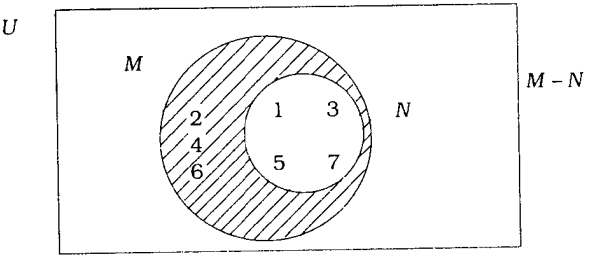
Diferencia de conjuntos iguales

Sea $M = \{1, 2, 3, 4, 5\}$ y $N = \{1, 2, 3, 4, 5\}$

$M - N = \{\ \} = \varnothing$. En este caso se cumple $M - N = N - M$

Diferencia de un conjunto y uno de sus subconjuntos

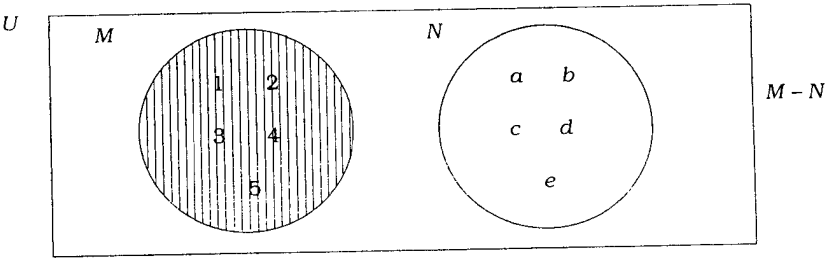
Sea $M = \{1, 2, 3, 4, 5, 6, 7\}$ y $N = \{1, 3, 5, 7\}$
$M - N = \{2, 4, 6\}$



Diferencia de conjuntos ajenos (disjuntos)

Sea $M = \{1, 2, 3, 4, 5\}$ y $N = \{a, b, c, d, e\}$
$M - N = \{1, 2, 3, 4, 5\} = M$

La diferencia de dos conjuntos ajenos es igual al conjunto minuendo.



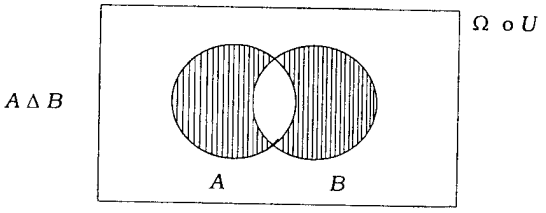
3.5.4 Diferencia simétrica entre dos conjuntos (A Δ B)

La diferencia simétrica de conjuntos es el formado por todos aquellos elementos que están en la unión pero no en la intersección.

Sean los conjuntos A y B; la diferencia simétrica (Δ) queda establecida de la manera siguiente:

$$A \Delta B = \{x \mid x \in (A \cup B) \text{ y } x \notin (A \cap B)\}$$

Lo anterior queda representado en el siguiente diagrama de Venn-Euler:



donde $A \Delta B$ es el área sombreada

$$A \Delta B = \{(A \cup B) - (A \cap B)\}$$

3.5.5 Propiedades de los conjuntos

Si intercambiamos los operadores (+) por $(U) \cup$ y $\cap(X)$ por $(X)_0$ (·) el conjunto universal (U) por 1 y el conjunto vacío (Φ) por 0, así como los numerales por conjuntos, respectivamente, entonces por razones nemotécnicas sólo se podrán utilizar las reglas de la aritmética en forma análoga con las propiedades de los conjuntos.

Esto se presenta en la tabla siguiente; por supuesto, en el caso de conjuntos, dichas propiedades se pueden verificar o justificar por medio de diagramas de Venn-Euler o de Lewis Carroll, según el caso, y en general se podrán comprobar utilizando axiomas o teoremas; esto último queda fuera del nivel de esta obra.

Ley aritmética	*Propiedad*	*Ley de conjuntos*
1) $a + b = b + a$ $a \cdot b = b \cdot a$	Conmutativa	$A \cup B = B \cup A$ $A \cap B = B \cap A$
2) $(a + b) + c = a + (b + c)$ $(ab)\,c = a\,(bc)$	Asociativa	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
3) $a\,(b + c) = ab + ac$	Distributiva	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
4) $a + 0 = a$ $a\,0 = 0$	Existencia de "0"	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cup \varnothing = A \quad A \cup \Omega = \Omega$
5) $a\,1 = a$	Existencia de "1"	$A \cap \varnothing = \varnothing$ $A \cap \Omega = A$

Por último, existe una propiedad exclusiva para conjuntos que no tiene equivalente en aritmética, esto es, lo relacionado con el complemento de conjuntos:

6. Para cualquier conjunto A hay solamente un conjunto complemento $\overline{A}$, tal que:

$$A \cup \overline{A} = \Omega$$
$$A \cap \overline{A} = \varnothing$$
$$(\overline{A})^c = A$$
$$\overline{\Omega} = \varnothing$$
$$\overline{\varnothing} = \Omega$$

7. Leyes de De Morgan:

$$a)\ \overline{A \cup B} = \overline{A} \cap \overline{B}$$
$$b)\ \overline{A} \cup \overline{B} = \overline{A \cap B}$$

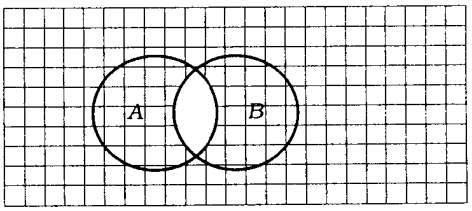
8. Leyes idempotentes:

$$A \cup A = A$$
$$A \cap A = A$$

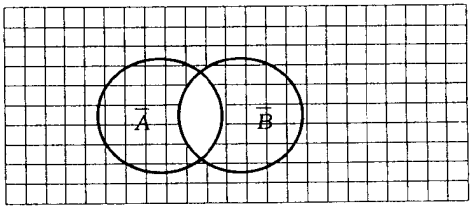
a) Si $A = B$ y $C = D$ implica que $A \cup C = B \cup D$
b) Si $A = B$ y $C = D$ implica que $A \cap C = B \cap D$
c) Si $A = B$ implica que $\overline{A} = \overline{B}$

Justificación por medio de los diagramas de Venn-Euler de una de las leyes de De Morgan:

$$\overline{A} \cup \overline{B} = \overline{A \cap B}$$



$\overline{A} \cup \overline{B}$



$\overline{A \cap B}$

De esta forma el lector podrá verificar todas las propiedades de los conjuntos; es recomendable que lo efectúe para tener una buena idea gráfica de los conjuntos.

Relación de igualdad

1. $A = A$	propiedad de reflexión
2. $A = B \Leftrightarrow B = A$	propiedad de simetría
3. $A = B$ y $B = C \Leftrightarrow A = C$	propiedad de transitividad

Relación de inclusión

1. $A \subset B \Leftrightarrow B \supset A$	propiedad de reflexión
2. $A \subset B \Leftrightarrow B \subseteq A$	propiedad de no simetría
3. $A \subset B$ y $B \subset C \Leftrightarrow A \subset C$	propiedad de transitividad

Operaciones algebraicas

1. $A \cup B = B \cup A$	propiedad conmutativa
2. $A \cap B = B \cap A$	propiedad conmutativa
3. $A - B \neq B - A$	no se cumple la propiedad conmutativa para la diferencia entre dos conjuntos
4. $(A \cup B) \cup C = A \cup (B \cup C)$	propiedad asociativa
5. $(A \cup B) \cap C = A \cap (B \cap C)$	propiedad asociativa
6. $(A - B) - C \neq A - (B - C)$	no se cumple la propiedad asociativa para la resta
7. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	propiedad distributiva de la intersección
8. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	propiedad distributiva de la unión

Otras propiedades

1. $A \cup \varnothing = A$
2. $A \cap U = A$
3. $A \cup U = U$
4. $A \cap \varnothing = \varnothing$

} propiedades de identidad

1. $A \cup A = A$
2. $A \cap A = A$

} propiedades de idempotencia

1. $\varnothing' = U$
2. $U' = \varnothing$
3. $(\overline{A'}) = A$

} propiedades de complementación

1. $(A \cup B)' = A' \cap B'$
2. $(A \cap B)' = A' \cup B'$

} propiedad de De Morgan.

3.6 CONTEO DE ELEMENTOS

Uno de los problemas con los que se enfrenta un investigador es el de evaluar la ocurrencia de ciertos eventos cuando se lleva a cabo una investigación, estudio o experimento. Aunque este tipo de problemas pertenece al campo de la pro-

babilidad, en muchos casos se podrán resolver enumerando (conteo de elementos) los puntos (eventos) en un espacio de eventos.

El principio fundamental del conteo se establece por medio del teorema siguiente.

Teorema Si una operación puede llevarse a cabo en n_1 formas, y si por cada una de estas operaciones se puede efectuar una segunda operación en n_2 formas, entonces las dos operaciones pueden efectuarse simultáneamente en $n_1 \cdot n_2$ formas.

3.6.1 Cardinalidad

La cardinalidad de un conjunto A es el número de elementos que contiene el conjunto, y se denota como $n(A)$.

Supongamos que A y B son dos conjuntos cualesquiera; cabe considerar dos posibilidades:

1. Si $A \cap B = \varnothing$, entonces, $n(A \cup B) = n(A) + n(B)$
2. Si $A \cap B \neq \varnothing$, entonces, $n(A \cup B) = n(A) + n(B) - n(A \cap B)$

En tal caso, se observa que $A \cup B$ incluye todos los elementos que pertenecen tanto a A como a B, pero cada uno de ellos debe contarse sólo una vez. Se observa que $n(A) + n(B)$ excede a $n(A \cup B)$, por $n(A \cap B)$

$$\text{o sea: } n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

Ejemplo Al entrevistar a 100 familias, se observó que 75 de ellas tenían suscripción al periódico *Tiempo*, 55 a *La Prensa*, y 10 a ninguno de ellos. ¿Cuántas familias están suscritas en ambos?

Sea A el conjunto de familias que están suscritas al periódico *Tiempo*.
Sea B el conjunto de familias que están suscritas al periódico *La Prensa*.

Conocemos los siguientes datos:

$$n(U) = \text{conjunto universal} = 100$$

$$n(A \cup B) = 90$$
$$n(A) = 75$$
$$n(B) = 55$$

Nos encontramos con la posibilidad:

$$A \cap B \neq \varnothing$$

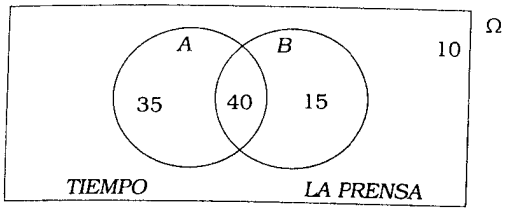
Es por eso que usaremos la fórmula siguiente:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B).$$

Sustituimos los datos que nos da el problema:

$$90 = 75 + 55 - n(A \cap B)$$
$$n(A \cap B) = 75 + 55 - 90$$
$$n(A \cap B) = 40$$

En un diagrama de Venn-Euler podemos visualizarlo:

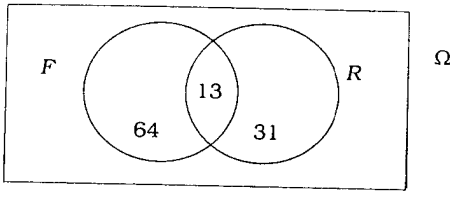


40 familias están suscritas en ambos periódicos.

Ejemplo En una encuesta efectuada en cierto centro de idiomas se encontró que 77 personas estudiaban francés, 44 ruso y 13 tanto francés como ruso. ¿Cuántas personas estaban estudiando por lo menos un idioma extranjero?

Observamos que un alumno que estudia tanto francés como ruso se encuentra en los dos grupos, de manera que la suma 77 + 44 cuenta a los estudiantes dos veces; para corregir este error se debe restar 13 (los que estudian ambos idiomas simultáneamente). Así, 77 + 44 − 13 = 108 es el número de estudiantes que aprenden por lo menos una lengua extranjera.

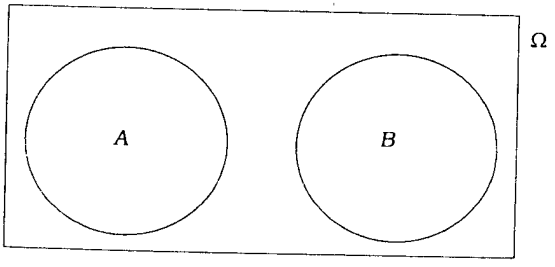
$$n(F \cup R) = n(F) + n(R) - n(F \cap R) = 77 + 44 - 13 \Rightarrow n(F \cup R) = 108$$



F = Los que estudian francés.
R = Los que estudian ruso.

Si $A \cap B = \varnothing$, la ecuación anterior se reduce a:

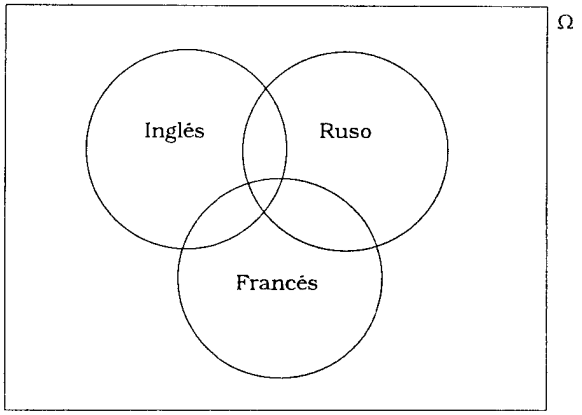
$$n(A \cup B) = n(A) + n(B)$$



3.6.2 Conteo de elementos para tres conjuntos

Como en el caso de dos conjuntos, cuando la información a clasificar requiere tres conjuntos, se pueden seguir utilizando los diagramas de Venn-Euler para tal efecto.

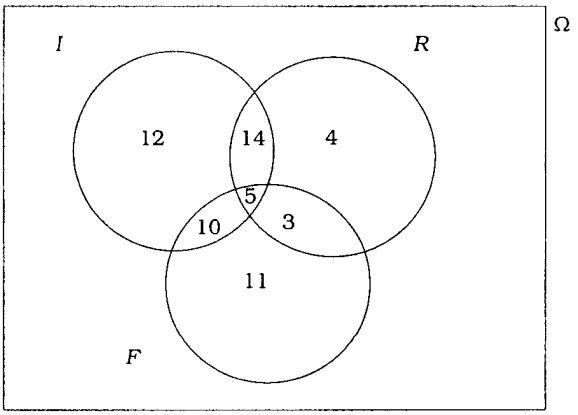
Ejemplo En una encuesta de 100 estudiantes se obtuvo la siguiente información: 41 alumnos estudian inglés, 29 francés, 26 ruso, 15 estudian inglés y francés, 8 francés y ruso, 19 inglés y ruso y 5 estudian los tres idiomas. Esto es:



donde:

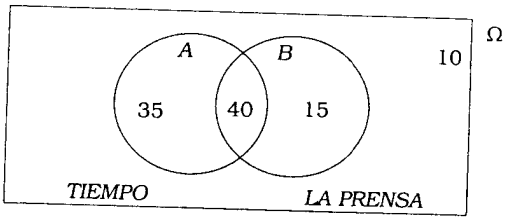
I = Los que estudian inglés
F = Los que estudian francés
R = Los que estudian ruso

Una vez ubicada la información en un diagrama de Venn-Euler, tenemos lo siguiente:



Al contar todos estos elementos, el resultado es 59 estudiantes, pero la muestra es de 100 estudiantes. ¿Dónde se encuentran los demás estudiantes y qué significa esto?

En un diagrama de Venn-Euler podemos visualizarlo:

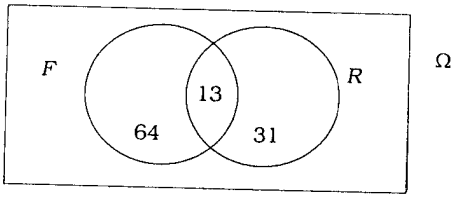


40 familias están suscritas en ambos periódicos.

Ejemplo En una encuesta efectuada en cierto centro de idiomas se encontró que 77 personas estudiaban francés, 44 ruso y 13 tanto francés como ruso. ¿Cuántas personas estaban estudiando por lo menos un idioma extranjero?

Observamos que un alumno que estudia tanto francés como ruso se encuentra en los dos grupos, de manera que la suma 77 + 44 cuenta a los estudiantes dos veces; para corregir este error se debe restar 13 (los que estudian ambos idiomas simultáneamente). Así, 77 + 44 − 13 = 108 es el número de estudiantes que aprenden por lo menos una lengua extranjera.

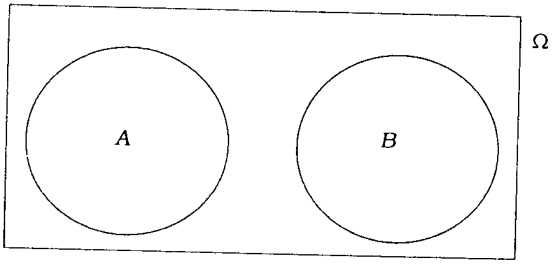
$$n(F \cup R) = n(F) + n(R) - n(F \cap R) = 77 + 44 - 13 \Rightarrow n(F \cup R) = 108$$



F = Los que estudian francés.
R = Los que estudian ruso.

Si $A \cap B = \varnothing$, la ecuación anterior se reduce a:

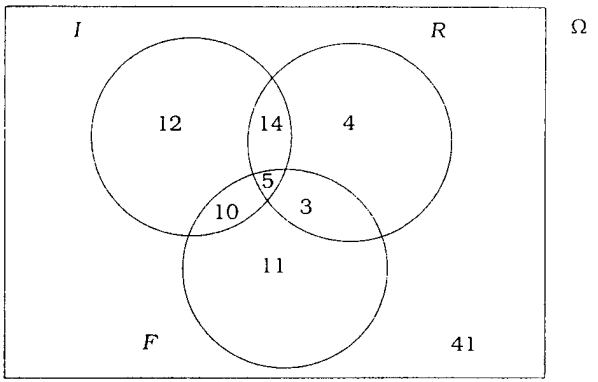
$$n(A \cup B) = n(A) + n(B)$$



3.6.2 Conteo de elementos para tres conjuntos

Como en el caso de dos conjuntos, cuando la información a clasificar requiere tres conjuntos, se pueden seguir utilizando los diagramas de Venn-Euler para tal efecto.

Como $n(\overline{I \cup R \cup F})$ significa todos aquellos alumnos que no estudian inglés, ruso o francés que, ubicados en el diagrama de Venn-Euler, da:



Además, este diagrama nos proporciona la siguiente información: 12 + 11 + 4 = 27 son los alumnos que estudian exactamente un idioma (inglés, francés o ruso, respectivamente).

3.6.3 Conteo de elementos para más de tres conjuntos

Cuando la información relevante se tiene que ubicar en más de tres conjuntos, la representación gráfica por medio de los diagramas de Venn-Euler se complica y en lugar de aclarar y facilitar la solución del problema, lo torna confuso. Pero, afortunadamente, existe otro tipo de representación gráfica que acepta un número grande de conjuntos y combinaciones entre ellos.

Estos diagramas —bastante lógicos— fueron innovados por Lewis Carroll. Esas tablas se conocen como *tablas de contingencia* o *diagramas de Carroll.* A continuación los utilizaremos en un problema donde intervienen siete conjuntos.

Ejemplo En una estancia infantil (con una población de 9 educadoras, 22 niños y 20 niñas) se hizo un estudio en el que, entre otras cosas, se les pidió que señalaran su color favorito. De este modo se obtuvo lo siguiente:

22 {
- 10 niños mencionaron el color rojo
- 5 niños señalaron el color azul
- 3 niños indicaron el color amarillo
- 4 niños mencionaron el color verde

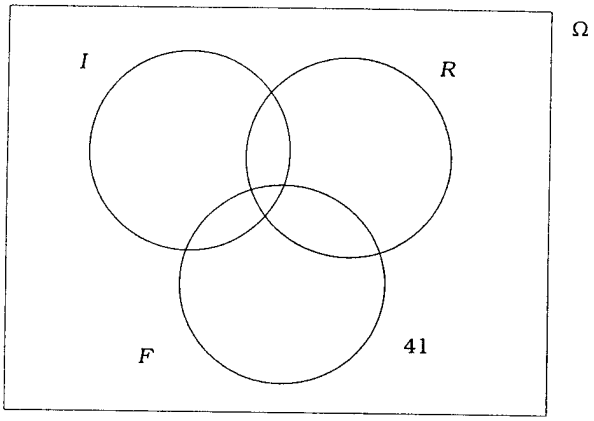
+

20 {
- 6 niñas mencionaron el color rojo
- 11 niñas indicaron el color azul
- 2 niñas señalaron el color amarillo
- 1 niña mencionó el color verde

+

9 {
- 3 educadoras señalaron el color rojo
- 4 educadoras indicaron el color azul
- 1 educadora mencionó el color amarillo
- 1 educadora señaló el color verde

51



41 alumnos no estudian ningún idioma.

Para tres conjuntos donde la intersección es diferente del vacío ($A \cap B \cap C \neq \varnothing$) se cumple lo siguiente:

$$\begin{aligned} n(A \cup B \cup C) = {} & n(A) + n(B) + n(C) \\ & - n(A \cap B) - n(B \cap C) - n(C \cap A) \\ & + n(A \cap B \cap C) \end{aligned}$$

Si se considera el complemento de $(\overline{A \cup B \cup C})$ que significa: todos aquellos elementos que están en el conjunto universal (Ω), excepto los elementos que están en $(A \cup B \cup C)$, o, en otras palabras:

$$n(\Omega) - n(A \cup B \cup C)$$

o sea,

$$\begin{aligned} n(A \cup B \cup C) = {} & n(\Omega) - n(A) - n(B) - n(C) \\ & + n(A \cap B) + n(B \cap C) + n(C \cap A) \\ & - n(A \cap B \cap C) \end{aligned}$$

Que aplicado al ejemplo anterior da como resultado lo siguiente:

$$\begin{aligned} n[(\overline{I \cup R \cup F})] = {} & n(\Omega) - n(I) - n(R) - n(F)) \\ & + n(I \cap R) + n(R \cap F) + n(F \cap I) \\ & - n(I \cap R \cap F) \end{aligned}$$

$$n(\overline{I \cup R \cup F}) = 100 - 41 - 26 - 29 + 19 + 8 + 15 - 5 = 41$$

Estos datos se ordenan en la tabla siguiente de contingencia:

	Rojo	*Azul*	*Amarillo*	*Verde*	*Total*
Niños	10	5	3	4	22
Niñas	6	11	2	1	20
Educadoras	3	4	1	1	9
Total	19	20	6	6	51

Ejemplo En una encuesta a 600 estudiantes de primer ingreso a la Facultad de Psicología (400 mujeres, 200 hombres) se les pidió que señalaran dos posibles áreas de especialidad de su interés.

	Clínica y social	*Clínica e industrial*	*Social e industrial*	*Educativa e industrial*	*Clínica y educativa*	*Educativa y social*	*Total*
Hombres	20	25	110	15	5	25	200
Mujeres	80	120	20	50	100	30	400
Total	100	145	130	65	105	55	600

Compruebe lo siguiente:

1. $n\,[(H \cap C \cap I) \cup (M \cap C \cap I)] =$
2. $n\,[(C \cup S)] =$
3. $n\,(C \cup S \cap E) =$
4. $n\,[(C \cap S) \cup (C \cap E)] =$
5. $n\,[(M \cup C \cap I)] =$
6. $n\,[(\overline{E} \cap I) \cup H] =$
7. $n\,[(\overline{H \cup C})]^c =$
8. $n\,[(H \cup C)] =$
9. $n\,[(\overline{M} \cup H \cap I)] =$
10. $n\,[(I \cup (E \cap I)] =$
11. $n\,[(\overline{C \cup E}) \cup (M \cap H)] =$
12. $n\,[H - (C \cap S) \cup (E \cap I) \cup (C \cap I)] =$
13. $n\,(C \Delta I) =$
14. $n\,(\overline{E} \cup H \cap C \cap I) =$

3.7 PRODUCTO CARTESIANO

Antes de definir y explicar qué es un producto cartesiano daremos a conocer qué es un par ordenado.

Sean dos conjuntos A y B no vacíos; tomaremos un elemento de cada conjunto. Queremos formar un conjunto que dependa de dichos elementos y del orden en el que se presenta.

Sea $x \in A$ ^[4] $y \in B$ y hacemos (x, y); la expresión anterior se llama *par ordenado*. "Tenemos un conjunto que dependen de x ^ y" que tienen un orden, primera componente y segunda componente, respectivamente.

Veamos ahora una definición de plano cartesiano.

Definición El producto cartesiano de dos conjuntos A y B es el conjunto cuyos elementos son todos los pares ordenados cuya primera componente pertenece a A y la segunda a B. Se denota de la siguiente manera.

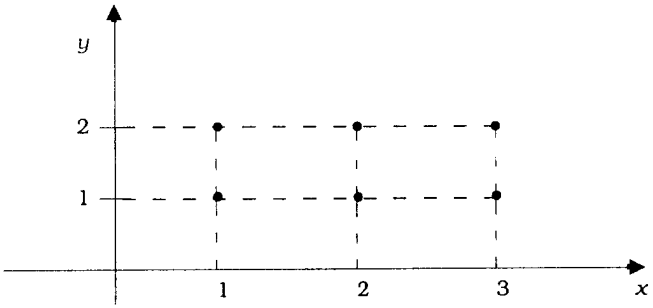
$$A \times B = \{(x, y) \mid x \in A \wedge y \in B\}$$

Ejemplo Sea $A = \{1, 2, 3\}$ ^ $B = \{1, 2\}$

$$A \times B = \{(1, 1)\ (1, 2)\ (2, 1)\ (2, 2)\ (3, 1)\ (3, 2)\}$$

La expresión anterior puede ser representada en un plano: son dos rectas perpendiculares, una llamada *abscisa* y la otra *ordenada*.

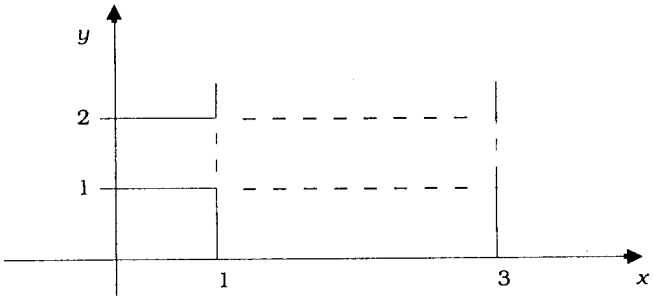
$$(x, y) \Rightarrow (x \rightarrow \text{abscisa},\ y \rightarrow \text{ordenada})$$



También podemos dibujar áreas en el plano cartesiano, incluyendo desigualdades, de la siguiente manera:

$(a, b) = \{x \in \mathbb{R} \mid J < x < 3\}$
$(c, d) = \{y \in \mathbb{R} \mid J < y < 2\}$
$\mathbb{R}$ = Conjunto de números reales.

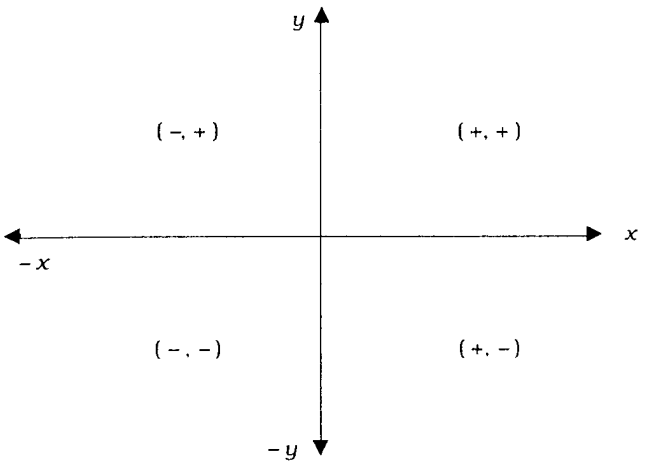
Si realizamos el producto cartesiano obtenemos
$[a, b]\ x\ [c, d] = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq x \leq 3 \wedge 1 \leq y \leq 2\}$
Geométricamente tenemos el rectángulo siguiente.



[4] ^ significa y.

El tema de desigualdades se retomará más adelante, con su uso y propiedades.

Como se observa, por medio del producto cartesiano se dibuja un plano, que se denota como $\mathbb{R}^2$. A continuación se muestra con sus signos respectivos:



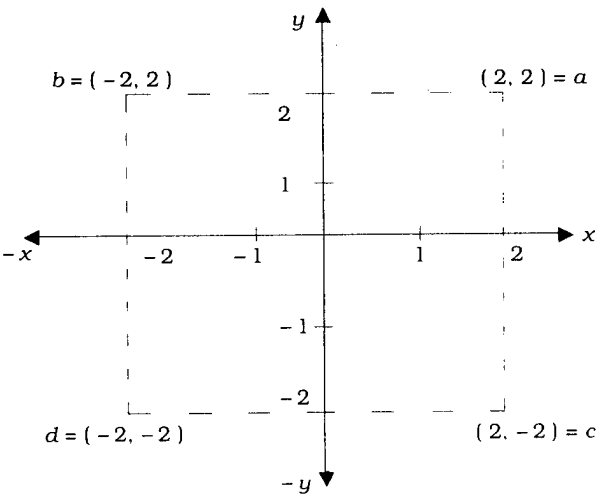
Localice los siguientes puntos en el plano cartesiano:

$a = (2, 2)$

$b = (-2, 2)$

$c = (2, -2)$

$d = (-2, -2)$



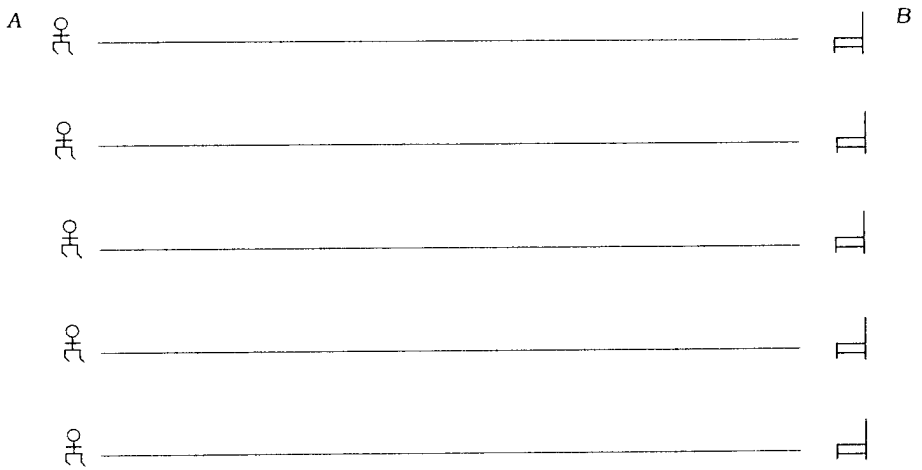
3.8 RELACIONES Y FUNCIONES

Una vez entendido el concepto de producto cartesiano, será fácil para el lector comprender que una relación es un subconjunto del plano cartesiano.

Definición Si A y B son dos conjuntos no vacíos definimos una relación de A a B (R) como sigue:

$$R = \{(x, y) \mid x \in A \wedge y \in B\} \in A \times B$$

Ejemplo En un salón de clase hay cinco alumnos y cinco sillas; la relación será la siguiente:



El dibujo anterior nos indica la relación siguiente: a cada niño le corresponde una silla, o de otra forma, a cada silla le corresponde un niño.[5]

El número de alumnos o de sillas puede ser diferente, se pueden hacer comparaciones entre conjuntos, a éstas se les llama *relaciones binarias*.

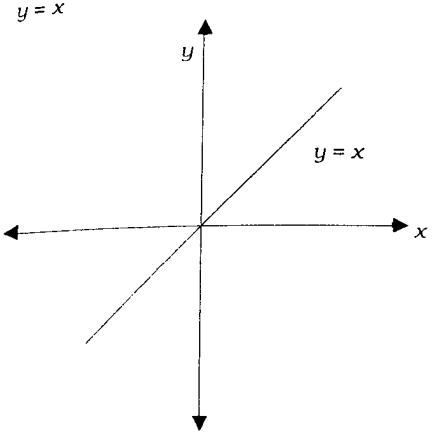
Las relaciones binarias de mayor uso en las matemáticas son las siguientes:

$=$	igual a	$\geq$	mayor o igual
$\neq$	diferente de	$<$	menor
$\Leftrightarrow$	equivale a	$\leq$	menor o igual
$\in$	pertenece a	$\perp$	es perpendicular a
$\subset$	está contenido en	$\mid\mid$	es paralelo a
$>$	mayor		

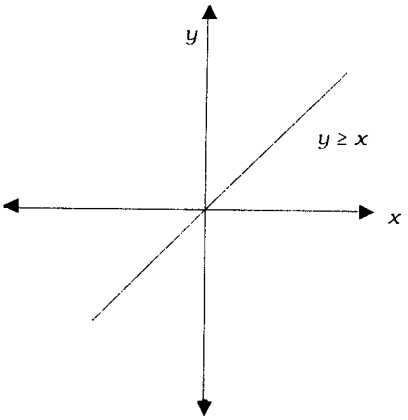
[5] Podemos determinar que a los elementos del conjunto A se les llame x, y y. A los del conjunto B es decir, a cada alumno le corresponde una x y a cada silla le corresponde una y, y así formamos la pareja ordenada (x, y).

Algunas de estas relaciones pueden ser representadas en el plano cartesiano, como son las siguientes:

$y = x$



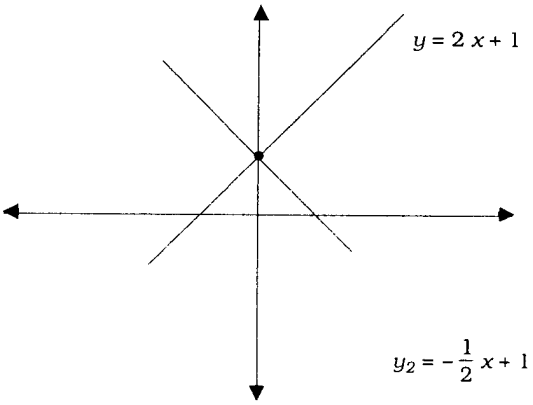
$y \geq x$



$y = 2x + 1$

$y_2 = -\frac{1}{2}x + 1$

$\therefore\ y_1 \perp y_2$



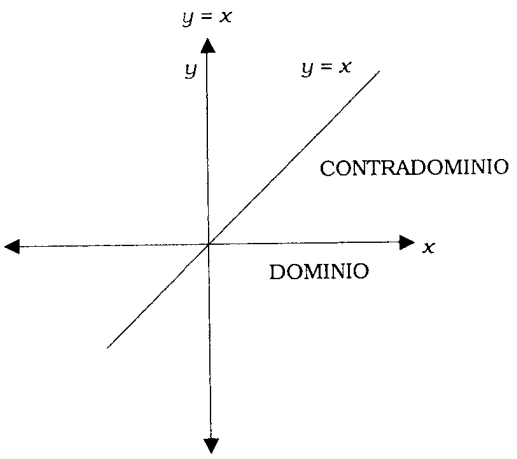
y_1 y y_2 son perpendiculares ya que la intersección de ambas rectas forma un ángulo de 90°.

3.8.1 Dominio y contradominio

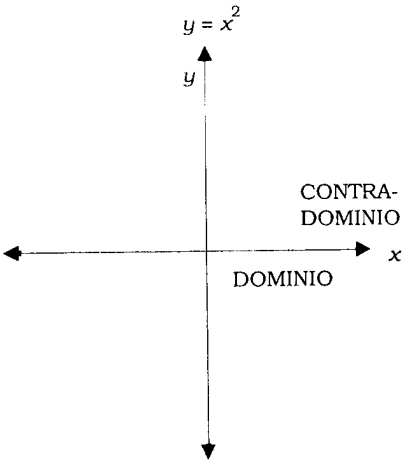
Para una mejor comprensión de dominio y contradominio, se presentarán algunos ejemplos para después dar las definiciones.

En el ejemplo de las sillas, se llama *dominio* al conjunto de los alumnos y *contradominio* al conjunto de las sillas.

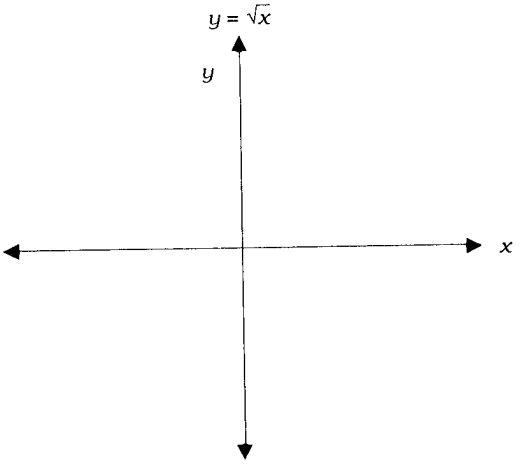
Para:



DOMINIO = $x \in \mathbb{R}$
CONTRADOMINIO = $y \in \mathbb{R}$



DOMINIO = $x \in \mathbb{R}$
CONTRADOMINIO = $y \geq 0$



DOMINIO = $x \geq 0$
CONTRADOMINIO = $y \geq 0$

Algunos autores usan los términos *imagen, codominio, rango* o *ámbito* cuando se refieren al contradominio.

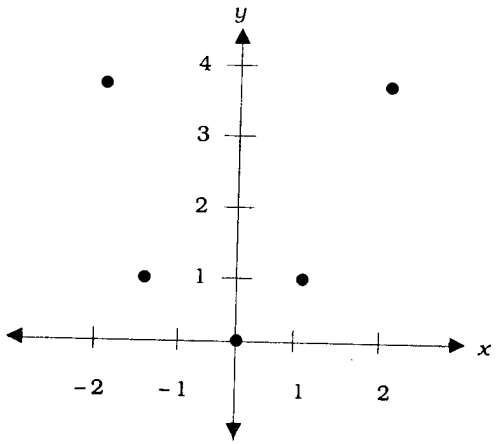
En textos avanzados de matemáticas, la imagen tiene un concepto diferente del de contradominio. En esta obra se considerará como lo mismo.

Se observa que a cada primer elemento le corresponde un segundo elemento. Se dibujarán sólo algunos puntos de la siguiente relación, formándose los pares ordenados:

$y = x^2$

x	y
−2	4
−1	1
0	0
1	1
2	4

(−2, 4)
(−1, 1)
(0, 0)
(1, 1)
(2, 4)



A todos los posibles valores que pueda aceptar el eje x se les llamará *dominio.*

A todos los posibles valores que pueda aceptar el eje y se les conoce como *contradominio.*

A continuación se establecen las definiciones formales de estos dos puntos. Sean A y B dos conjuntos no vacíos.

El dominio de R es la totalidad de los elementos de A, que admite contradominio en B.

$$D_R = \{x \in A \mid (x, y) \in R\}$$

El contradominio de R es el conjunto de los elementos de B, que admite un antecedente en A.

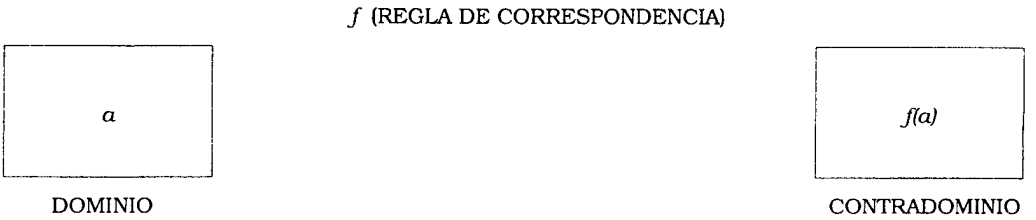
$$C_D = \{y \in B \mid (x, y) \in R\}$$

donde R es la relación.

El lector estará en condiciones de aprender las funciones.

Definición Una función es una terna formada por:

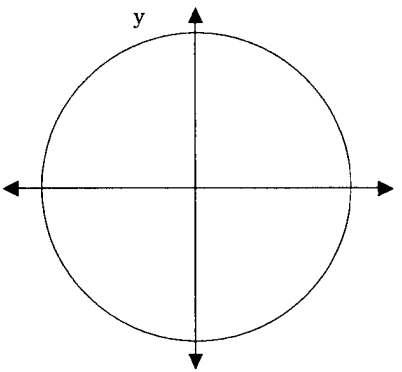
1. Un primer conjunto no vacío llamado *dominio de la función.*
2. Un segundo conjunto no vacío llamado *contradominio de la función.*
3. Una regla de correspondencia que tenga las siguientes propiedades:
   a) Por medio de ella, a cualquier elemento del dominio de la relación se le puede asociar un elemento del contradominio.
   b) Ningún elemento del dominio se quedará sin su asociado con el contradominio.
   c) Ningún elemento puede tener más de un asociado en el contradominio.



De la definición anterior se concluye que toda función es una relación, pero no toda relación es una función.

El siguiente es un ejemplo para ilustrar la conclusión anterior:

$$f(x) = \pm\sqrt{25 - x^2} \qquad \ldots \text{(I)}$$



x	$f(x)$
-3	± 4
-2	$\pm\sqrt{21}$
-1	$\pm\sqrt{24}$
0	0
1	$\pm\sqrt{24}$
2	$\pm\sqrt{21}$
3	± 4

Al realizar la regla de correspondencia de algunos puntos, se tiene que:

$$f(4) = \pm\sqrt{25 - (4)^2} = \pm\sqrt{25 - 16} = \pm\sqrt{9} = 3$$

Con el valor obtenido se forman las siguientes parejas ordenadas:

$$(4, 3)$$
$$(4, -3)$$

Como se observa, para la primera componente se tienen dos valores diferentes en la segunda componente.

El resultado anterior no cumple con la condición 3*C* de la definición; por tanto, la expresión (*I*) no es una función, sino una relación.

3.8.2 Variabilidad

Al referirse a una relación matemática, se ha observado que algo varía en función a otra u otras; por lo general, a este tipo de valores se les conoce con el nombre de *variables*. Así, si una variable está dada en función de otra, se le llama *variable dependiente* y a la otra *variable independiente*, cuya característica es de ser susceptible de medirse y se expresa de la forma siguiente:

$$y = f(x)$$

3.8.2.1 Variables dependientes e independientes

y es una función de x, donde y es la variable dependiente y la variable independiente es x.

Ejemplo 1 A todo estímulo (E) le corresponde una respuesta (R); mediante la notación funcional, se tiene:

$$R = f(E)$$

variable dependiente

variable independiente

Ejemplo 2

$$y = 3x^3 + 1$$

variable independiente

variable dependiente

Variable independiente porque puede tomar cualquier valor "x", es decir, se le puede asignar $x = 1$, $x = \pi$, $x = 1.12$, $x = \sqrt{2}$, etcétera.

Por otro lado, y depende de los valores de x para poder tener algún resultado, por eso se le nombra *variable dependiente.*

Antes de continuar con el desarrollo de las funciones es conveniente explicar en qué consisten los intervalos, ya que el uso de éstos se verá en el transcurso de los siguientes capítulos.

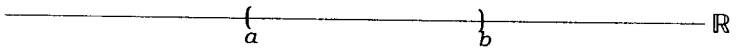
3.8.3 Intervalos y desigualdades

Sea $\mathbb{R}$ el conjunto de números reales, pero al fijar nuestra atención en un solo segmento de recta, entonces necesitamos definir conceptos y notación. Esta recta la podemos limitar por medio de desigualdades o igualdades, incluso, cuando hablemos de un punto.

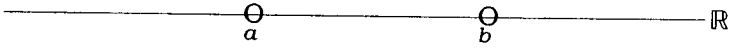
Los intervalos pueden ser *cerrados*, *abiertos* o *híbridos*. Se presentan los siguientes casos:

$$(a, b) = \{x \mid a < x < b\} \text{ Intervalo abierto}$$

La expresión anterior nos indica que x nunca tomará el valor de a ni de b; en la recta numérica queda de la manera siguiente:



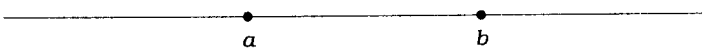
o también



Sea el intervalo:

$$[a, b] = \{x \mid a \leq x \leq b\} \quad \textit{Intervalo cerrado}$$

En este caso, x sí tomará el valor a y b. Como podemos observar, las desigualdades nos indican si incluye al extremo que limita al intervalo. En la recta numérica queda expresado de la siguiente forma:



o también

Los *híbridos*, también llamados *semiabiertos*, son los siguientes:

$[a, b) = \{x \mid a \leq x < b\}$

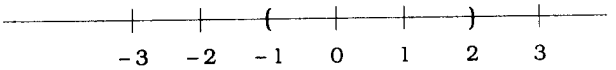
$(a, b] = \{x \mid a < x \leq b\}$

Antes de explicar algún ejemplo numérico daremos uno cotidiano.

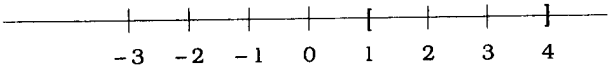
Cuando se señala un periodo de vacaciones como del 28 de febrero al 10 de marzo, no se menciona si los días 28 y 10 se tomarán también como vacaciones; pero cuando se dice *inclusive*, nos indica que los días no hábiles serán tomados incluyendo 28 y 10, respectivamente. Así también con los intervalos que contienen las desigualdades con mayor o igual o menor o igual.

Ejemplo Localice en la recta numérica los intervalos siguientes:

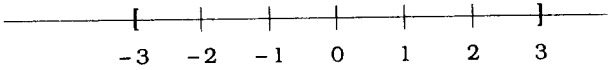
$(-1, 2)$



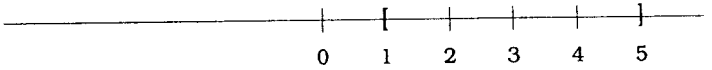
$[1, 4]$



$[-3, 3]$

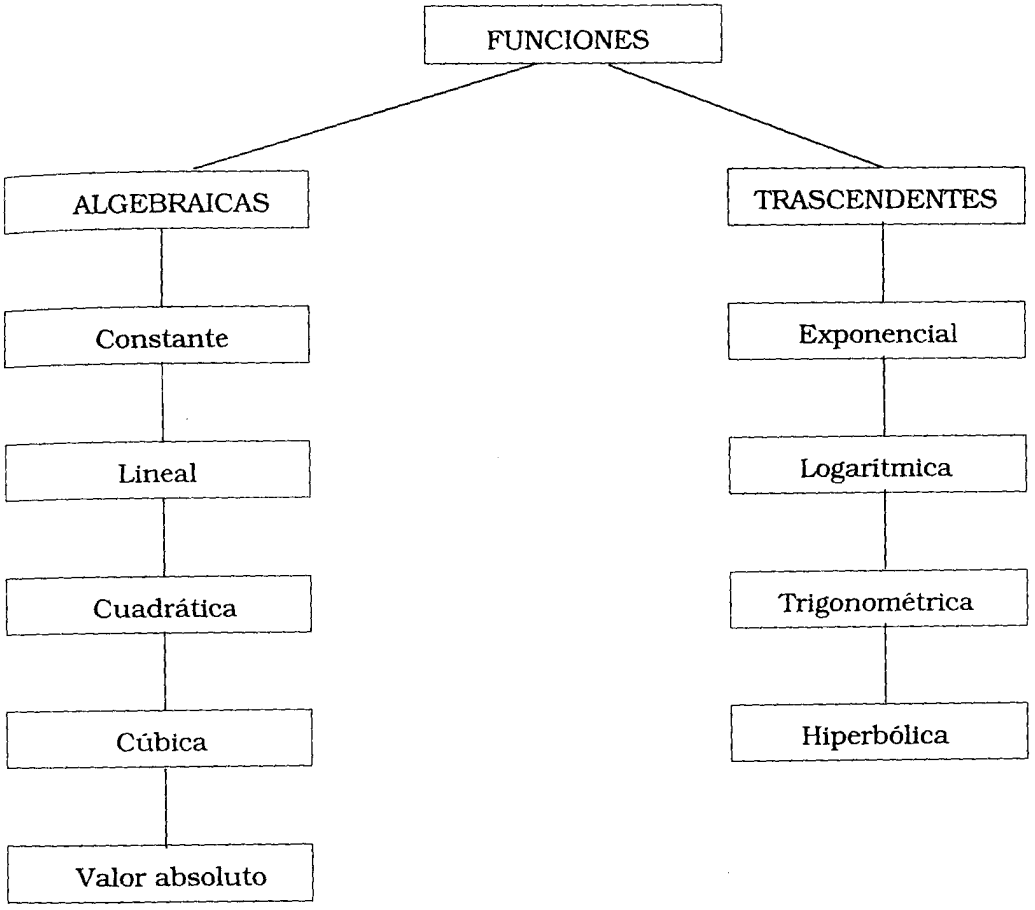


$[1, 5]$



3.8.4 Clasificación de funciones

En cursos avanzados de álgebra superior se estudian varias clasificaciones de funciones, pero nosotros analizaremos sólo una, con la finalidad de ubicar al lector dentro del universo de las matemáticas.

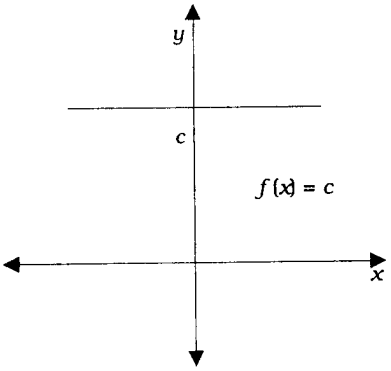


En el cuadro anterior se trata la función *valor absoluto* como un caso especial de las funciones algebraicas. En cuanto a la parte de trascendentes, estudiaremos las funciones logarítmicas y exponenciales, ya que son la base para el análisis de algunos temas estadísticos, como distribuciones de probabilidad y análisis de regresión. No estudiaremos las funciones trigonométricas ni las hiperbólicas, ya que están fuera del alcance de este libro; pero sí se considerarán las funciones algebraicas, las cuales se catalogan en la forma siguiente: constante, lineal, cuadrática, cúbica, valor absoluto.

Constante Es una función con $\mathbb{R}$ como su dominio y contradominio es sólo un número real.

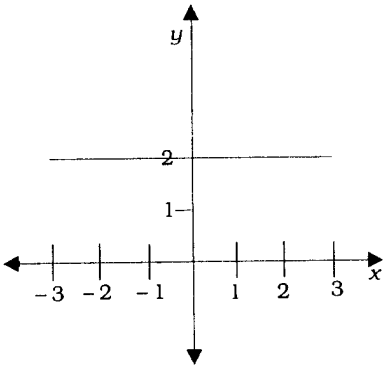
Sea $C \in \mathbb{R}$, los pares ordenados serán los siguientes:

$$\{(x, c) \mid x \in \mathbb{R}\}, \quad f(x) = C$$



x	y
−3	2
−2	2
−1	2
0	2
1	2
2	2
3	2

Ejemplo $f(x) = 2$

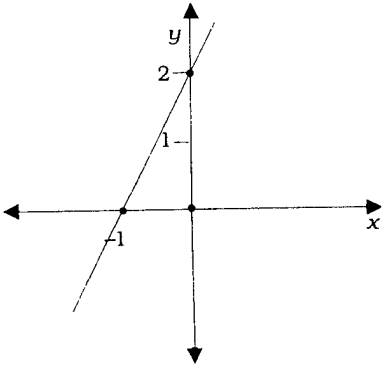


Lineal Sea $f(x) = ax + b$ en donde a y b son números reales y $a \neq 0$. Su dominio como su contradominio es $\mathbb{R}$.

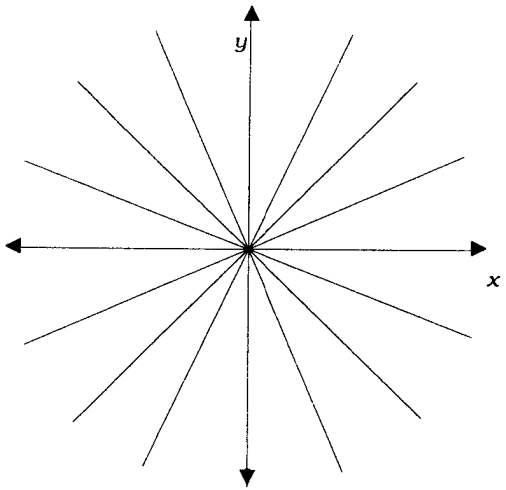
Se utiliza el término *lineal* por lo que la gráfica de $f(x) = ax + b$ es una línea recta.

Ejemplo $f(x) = 2x + 2$

x	y
−3	−4
−2	−2
−1	0
0	2
1	4
2	6
3	8

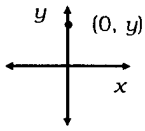
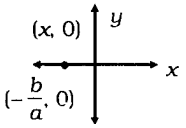
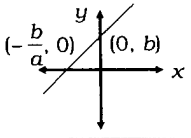


Ejemplo $f(x) = ax$ donde a es cualquier número real.



A todas las posibles rectas que pasan por el punto (0, 0), llamado *origen*, se le conoce como *familia de rectas que pasan por el origen.*

A continuación se presenta un procedimiento alternativo para obtener una gráfica de cualquier función lineal:

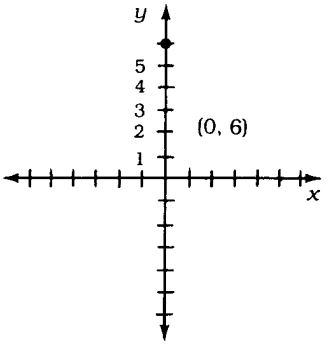
Paso	1	Sea $x = 0$	Para $y = ax + b$
Paso	2	Obtenga y	$y = a(0) + b,\ y = b$
Paso	3	Grafique el punto $(0, y)$	
Paso	4	Sea $y = 0$	$0 = ax + b$
Paso	5	Obtenga x	$x = -\frac{b}{a}$
Paso	6	Grafique el punto $(x, 0)$	
Paso	7	Una los dos puntos con una línea recta	

Ejemplo $y = 2x + 6$

Paso 1. Sea $x = 0 \quad \Rightarrow \quad y = 2(0) + 6$

Paso 2. $y = 6$

Paso 3. Grafique el punto $(0, y) = (0, 6)$



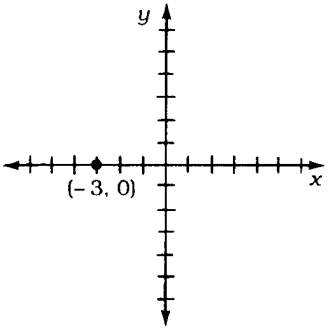
Paso 4. Sea $y = 0 \quad \Rightarrow \quad 0 = 2x + 6$

Paso 5. Obtenga $x \quad \Rightarrow \quad -6 = 2x$

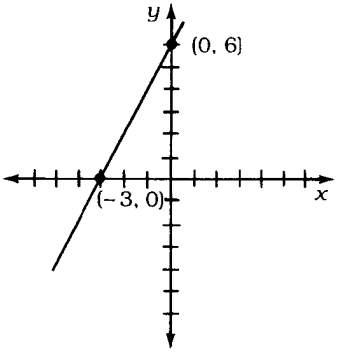
$$-\frac{6}{2} = x$$

$$\Rightarrow x = -3$$

Paso 6. Grafique el punto $(-3, 0)$



Paso 7. Una los dos puntos con una línea recta:



Toda línea recta tiene una inclinación; a ésta se le llama pendiente, se denota por m y se define como:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

Como podemos observar, para el cálculo de m es necesario conocer dos puntos:

$$P_1 = (x_1, y_1),\ P_2 = (x_2, y_2)$$

Existe otra forma de la ecuación lineal, que es la de punto y pendiente, dada por la siguiente fórmula:

$$y - y_1 = m(x - x_1)$$

La pendiente puede presentar cuatro casos, los cuales se explicarán con su respectiva ecuación de punto y pendiente.

CASO 1 $m > 0$

Sea $P_1 = (-4, 6)$, $P_2 = (-1, 18)$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{18 - 6}{-1 - (-4)} = \frac{12}{3} = 4$$

Con el punto P_1 formamos la ecuación de la recta:

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y - 6 &= 4(x - (-4)) \\ y - 6 &= 4x + 16 \\ y &= 4x + 16 + 6 \end{aligned}$$

$$\boxed{y = 4x + 22}$$

CASO 2 $m < 0$

Sea $P_1 = (5, -7)$, $P_2 = (3, -4)$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{-4 - (-7)}{3 - 5} = \frac{-4 + 7}{-2} = \frac{3}{-2}$$

Con el punto P_1 formamos la ecuación de la recta:

$$y - (-7) = \frac{3}{-2}(x - 5)$$

$$y + 7 = \frac{-3}{2}x + \frac{15}{2}$$

$$y = -\frac{3}{2}x + \frac{15}{2} - \frac{14}{2}$$

$$\boxed{y = -\frac{3}{2}x + \frac{1}{2}}$$

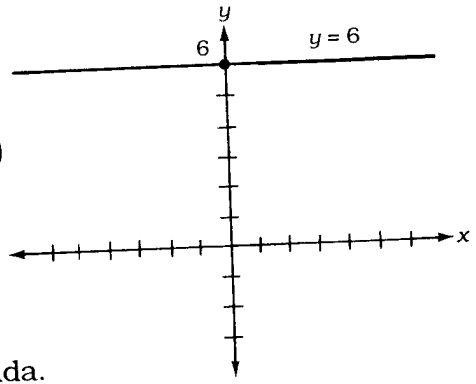
CASO 3 $m = 0$

Sea $P_1 = (3, 6)$, $P_2 = (-2, 6)$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{6-6}{-2-3} = \frac{0}{-5} = 0$$

Con el punto P_1 formamos la ecuación de la recta:

$$y - 6 = 0\,(x - x_1)$$
$$y = 6$$

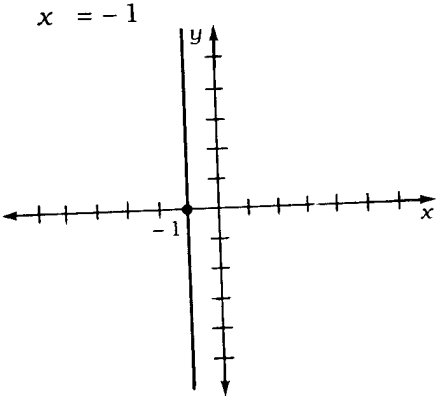


CASO 4 $m = \infty$ no está definida.

Sea $P_1 = (-1, -3)$, $P_2 = (-1, 2)$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{2-(-3)}{-1-(-1)} = \frac{5}{0} = \infty$$

la ecuación es $x = -1$



En suma, la ecuación de una recta es:

$$y = mx + b$$

donde: m = es la pendiente de la recta.

b = es conocida como la *ordenada al origen*.

En este libro, a la pendiente m la llamaremos a.

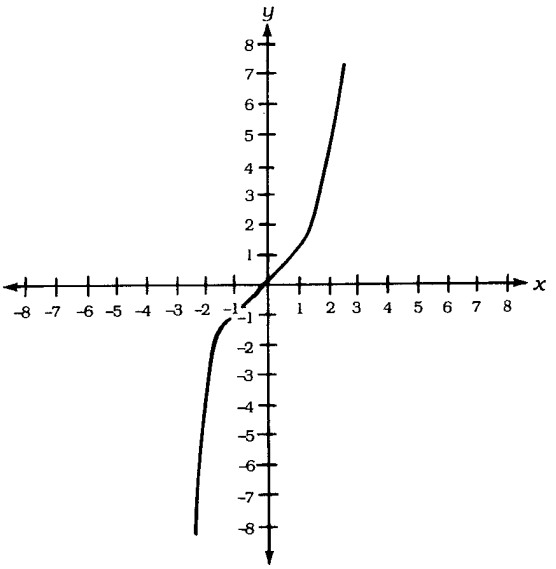
Cúbica Una función $f(x)$ es una función cúbica si

$$f(x) = ax^3 + bx^2 + cx + d$$

donde a, b, c y d son números reales y $a \neq 0$. Su dominio y contradominio siempre será $\mathbb{R}$.

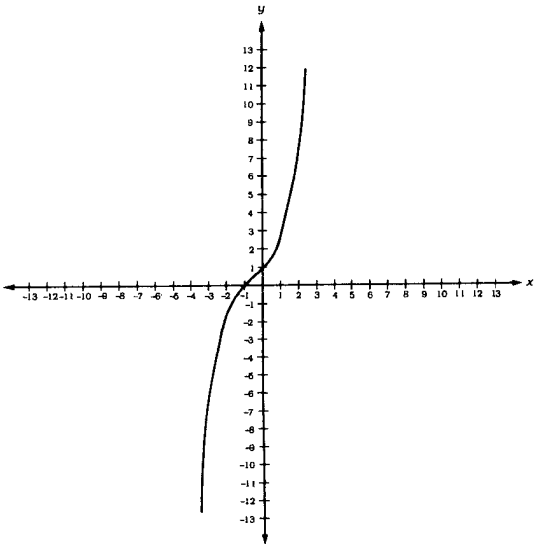
Ejemplo Trazar la gráfica de $f(x) = x^3$

x	y
−2	−8
−1	−1
0	0
1	1
2	8



Ejemplo Trazar la gráfica de $f(x) = x^3 + x^2 + 1$

x	y
−2	−3
−1	1
0	1
1	3
2	13



Todos estos conceptos serán retomados al abordar los sistemas de ecuaciones, tema que se explica en (3.10) y regresión lineal, en el capítulo 13.

Cuadrática Una función $f(x)$ es una función cuadrática si

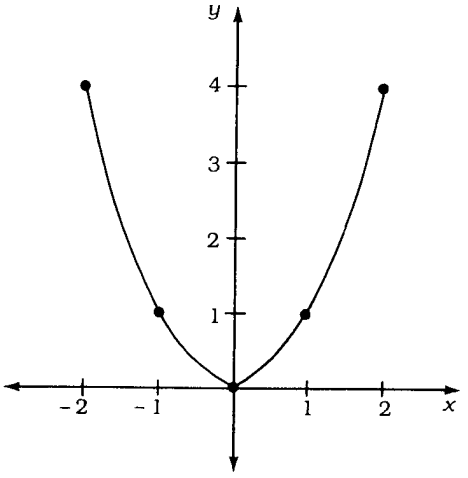
$$f(x) = ax^2 + bx + c$$

donde a, b y c son números reales y $a \neq 0$, a y b pueden tomar el valor de cero.

Ejemplo Trazar la gráfica de $f(x) = x^2$

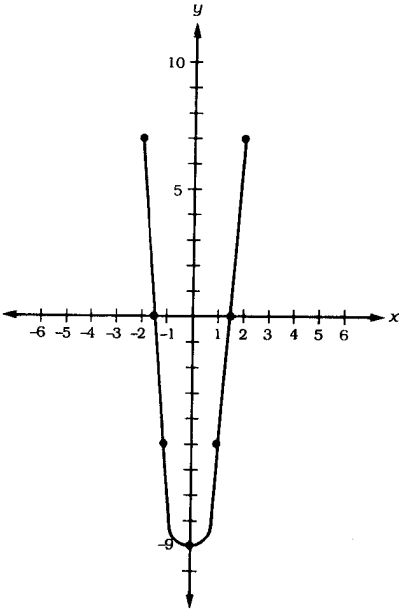
x	y
-2	4
-1	1
0	0
1	1
2	4

Dominio $\Rightarrow x \in \mathbb{R}$ y contradominio $\{y \mid y \geq 0\}$ para toda $y \in \mathbb{R}$



Ejemplo Trazar la gráfica de $f(x) = 4x^2 - 9$

x	y
-2	7
-1	-5
0	-9
1	-5
2	7
3/2	0
-3/2	0



El dominio de la función será todos los reales ($\mathbb{R}$) y contradominio es:

$$\{y \mid y \geq -9\} \Rightarrow \text{Eje } y$$

Al comprobar los resultados tenemos:

$$|-2-2| = 4 \qquad\qquad |6-2| = 4$$
$$4 = 4 \qquad\qquad 4 = 4$$

Es importante aclarar que no siempre encontraremos dos soluciones; en algunos casos obtendremos sólo una, es decir, no siempre las dos cumplirán con la igualdad.

Ejemplo Resuelva:

$$|x+1| = 2x-4$$

$$\begin{aligned} -(x+1) &= 2x-4 \\ -x-1 &= 2x-4 \\ -x-2x &= -4+1 \\ -3x &= -3 \\ x &= \frac{-3}{-3} \\ x_1 &= 1 \end{aligned}$$

o bien

$$\begin{aligned} x+1 &= 2x-4 \\ x-2x &= -4-1 \\ -x &= -5 \\ x_2 &= 5 \end{aligned}$$

Las soluciones posibles son $x_1 = 1$, $x_2 = 5$.
Realizamos la comprobación:

Si $x_1 = 1$

$$\begin{aligned} |1+1| &= 2(1)-4 \\ |2| &= 2-4 \\ 2 &= -2 \end{aligned}$$

El resultado anterior es falso; por tanto $x_1 = 1$ no es solución a la ecuación propuesta.

Si $x_2 = 5$

$$\begin{aligned} |5+1| &= 2(5)-4 \\ |6| &= 10-4 \\ 6 &= 6 \end{aligned}$$

La única solución a la ecuación es $x_2 = 5$, ya que así se cumple la igualdad.

CASO 2 $\qquad |a| < b$

La solución al caso 2 es la siguiente:

$$a > -b \quad \text{y} \quad a < b$$

Las funciones hasta aquí estudiadas se denominan polinomiales y se expresan en forma general de la manera siguiente:

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_{n-1}x^{n-1} + a_nx^n$$

La suma anterior se puede expresar como:

$$f(x) = \sum_{i=0}^{n} a_i x^i$$

para todo $a_i \in \mathbb{R}$ y $a_n \neq 0$.

Puesto que las operaciones utilizadas para definir el polinomio son simplemente la suma y la multiplicación, resulta evidente que para cada $x \in \mathbb{R}$, y es un número real único. Por tanto, esta ecuación define una función cuyo dominio es $\mathbb{R}$ y cuyo contradominio es algún subconjunto de $\mathbb{R}$. Se dice, entonces, que la función es una *polinomial de grado n*.

Valor absoluto La función valor absoluto, denotada por $|\ |$, es la función con dominio $\mathbb{R}$ y regla de correspondencia:

$$|x| = \begin{cases} x, & \text{si } x \geq 0 \\ -x, & \text{si } x < 0 \end{cases}$$

A continuación se explicarán las soluciones a ecuaciones y desigualdades en las que está implicado el valor absoluto.

CASO 1 $\qquad |a| = b \qquad \ldots$ (I)

La solución al caso 1 es la siguiente:

$$-a = b \qquad \text{o} \qquad a = b \qquad \ldots \text{(II)}$$

Ejemplo Resolver

$$|x - 2| = 4$$

Para resolver la ecuación hacemos $a = x - 2$ y $b = 4$, de (I) y pasamos a las dos expresiones que se indican en (II) para obtener:

$-(x-2) = 4$	o bien	$x - 2 = 4$
$-x + 2 = 4$		$x = 4 + 2$
$-x = 4 - 2$		$x = 6$
$-x = 2$		
$x = -2$		

Así pues, la ecuación tiene las dos soluciones:

$$x_1 = -2 \quad \text{y} \quad x_2 = 6$$

Ejemplo Resuelva:

$$|3 + 2x| \leq 2$$

$3 + 2x \geq -2$	y	$3 + 2x \leq 2$
$2x \geq -2 - 3$		$2x \leq 2 - 3$
$2x \geq -5$		$2x \leq -1$
$x \geq \dfrac{-5}{2}$		$x \leq \dfrac{-1}{2}$

El conjunto solución es el siguiente:

$$x \in \left[\frac{-5}{2}, \frac{-1}{2}\right]$$

En el caso 2 tenemos una desigualdad; por eso, la solución es un intervalo, es decir, encontramos un conjunto de posibles soluciones que satisfacen la desigualdad propuesta.

CASO 3 $|a| > b$:

La solución al caso 3 es la siguiente:

$$-a > b \quad \text{o} \quad a > b \qquad \text{... (III)}$$

Una de las propiedades de las desigualdades nos indica que si multiplicamos por (–1) a ambos lados de la desigualdad, ésta se invierte.

$$-a > b$$
$$(-1)(-a) < (-1)b$$
$$a < -b \qquad \text{... (IV)}$$

Sustituimos el resultado (IV) en la expresión (III) y así obtenemos:

$$a < -b \quad \text{o bien} \quad a > b \qquad \text{... (V)}$$

Ejemplo Resolver la desigualdad siguiente:

$$|4x - 3| > x + 2$$

Hacemos $a = (4x - 3)$ y $b = (x + 2)$; ahora sustituimos en (V) y obtenemos:

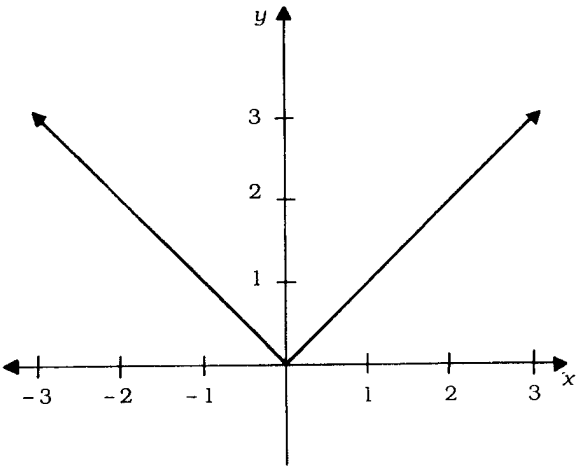
$(4x - 3) < -(x + 2)$	o bien	$(4x - 3) > x + 2$
$4x - 3 < -x - 2$		$4x - x > 2 + 3$
$4x + x < -2 + 3$		$3x > 5$
$5x < 1$		$x > \dfrac{5}{3}$
$x < \dfrac{1}{5}$		

El conjunto solución a la desigualdad es $x \in (-\infty, \frac{1}{5}) \cup (\frac{5}{3}, \infty)$

La unión nos indica el conectivo lógico que se da en la expresión (V).

Ejemplo Trazar la gráfica de $f(x) = |x|$. Aplicamos la definición anterior.

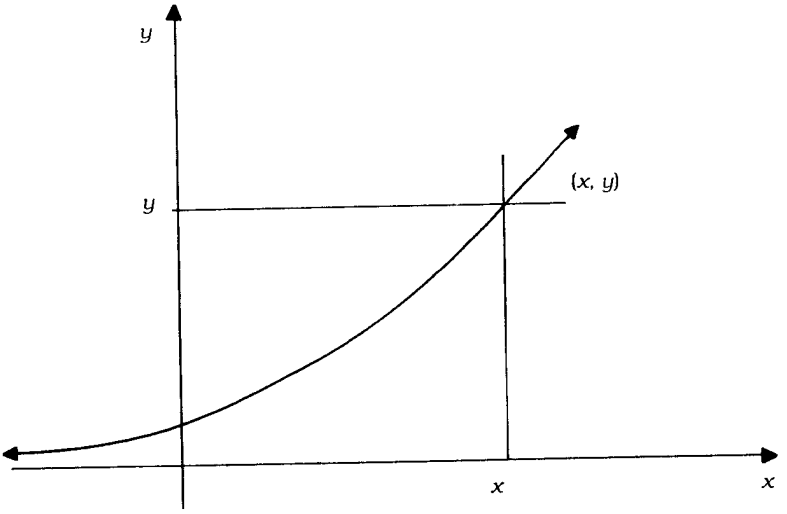
x	y
-3	3
-2	2
-1	1
0	0
1	1
2	2
3	3



Todas las funciones antes mencionadas se consideran algebraicas, pero existen otras funciones usuales que se conocen como trascendentes: exponencial y logarítmica.

Exponencial Sea b cualquier número real positivo distinto de 1. Entonces, una función $f(x)$ se llama *función exponencial de base b* si y sólo si

$$f = \{(x, y) \mid y = b^x, x \in \mathbb{R}\}$$



Como podemos observar, en la gráfica de una función exponencial típica de inmediato salta a la vista la siguiente propiedad: toda recta paralela al eje x y por encima de él, corta a la curva en un punto exactamente. Esto significa que si b es un número real dado, tal que $b > 0$ y $b \neq 1$, entonces, para cada número real positivo y existe un solo número real x, tal que $y = b^x$.

La función exponencial es muy usada para observar el crecimiento de cierta población; así, es común escuchar que China tuvo un crecimiento exponencial en los últimos cinco años, dato de fácil comprensión si se conoce el comportamiento de tales funciones. En el mundo de los negocios es usual el término *interés compuesto*, concepto que implica una función exponencial, como es la siguiente:

$$f(x) = P(1 + i)^x$$

donde P = monto inicial o principal
i = interés al que se invierte el principal

Así como los dos ejemplos anteriores, existen otros en los que utilizan funciones, herramienta necesaria en el estudio de varias disciplinas.

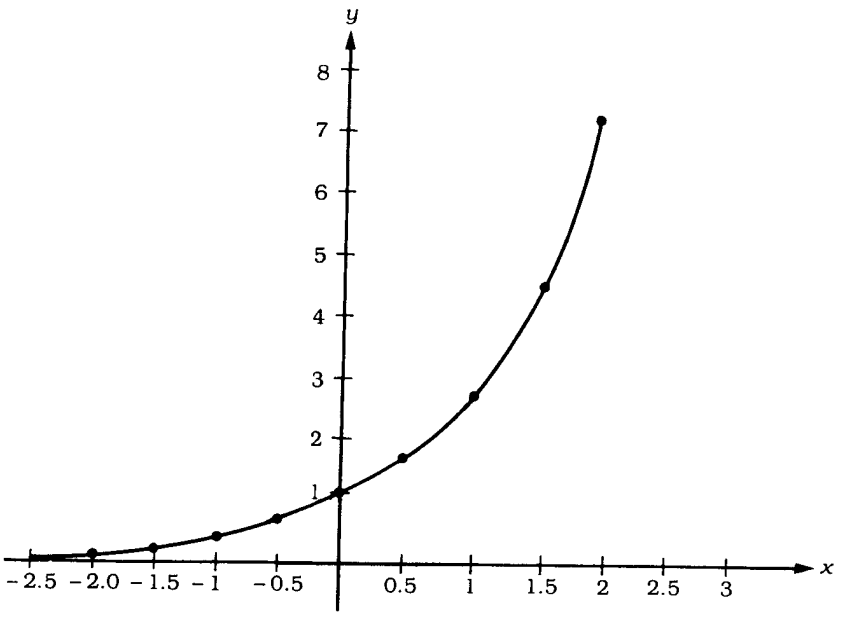
La función exponencial natural está definida de la forma siguiente:

$$f(x) = e^x$$

donde $e = 2.7182818$

A continuación presentamos su gráfica con algunos de sus respectivos valores:

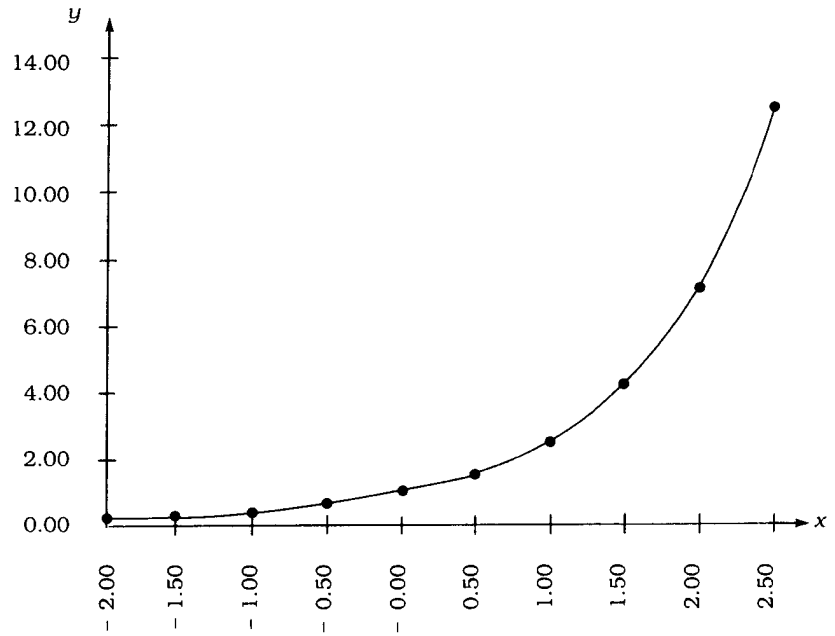
x	$f(x) = e^x$
-2	0.14
-1.5	0.22
-1	0.37
-0.5	0.61
0.0	1.00
0.5	1.65
1	2.72
1.5	4.48
2	7.39
2.5	12.18



Otra forma de representar gráficamente las funciones es trasladando el eje y a la izquierda para visualizar en forma más clara la curva que se dibuja, que queda de la manera siguiente:

Función exponencial

x	$y = e^x$
-2.00	0.14
-1.50	0.22
-1.00	0.37
-0.50	0.61
0.00	1.00
0.50	1.65
1.00	2.72
1.50	4.48
2.00	7.39
2.50	12.18



Realizaremos una transformación a la función exponencial, tratándola como un caso especial.

En el capítulo 2 hablamos de una distribución normal con una forma muy singular: los datos observados tenían cierta característica para comportarse de esa manera.

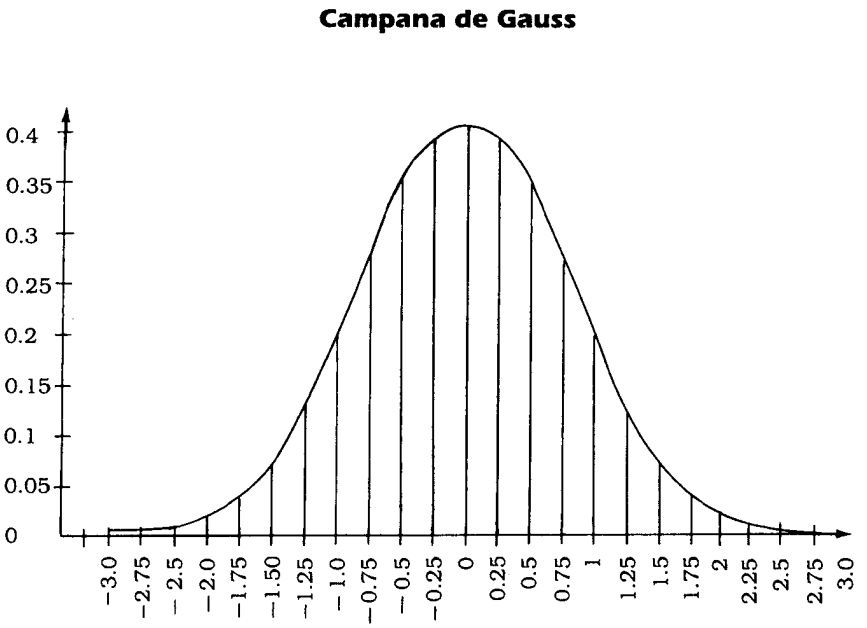
La cuestión que nos ocupa en el presente capítulo es la función que dibuja la distribución normal estándar y está dada por la siguiente expresión:

$$f(x) = \frac{1}{\sqrt{2\pi}} \quad e^{\left(-\frac{1}{2}x^2\right)}$$

El dominio de la función son todos los reales $x \in (-\infty, \infty)$.

La gráfica de la función anterior es la siguiente:

x	y
-3	0.00443185
-2.75	0.00909356
-2.5	0.0175283
-2.25	0.03173965
-2	0.05399096
-1.75	0.0862773
-1.5	0.12951757
-1.25	0.18264905
-1	0.24197068
-0.75	0.30113737
-0.5	0.35206526
-0.25	0.38666804
0	0.3989422
0.25	0.38666804
0.5	0.35206526
0.75	0.30113737
1	0.24197068
1.25	0.18264905
1.5	0.12951757
1.75	0.0862773
2	0.05399096
2.25	0.03173965
2.5	0.0175283
2.75	0.00909356
3	0.00443185



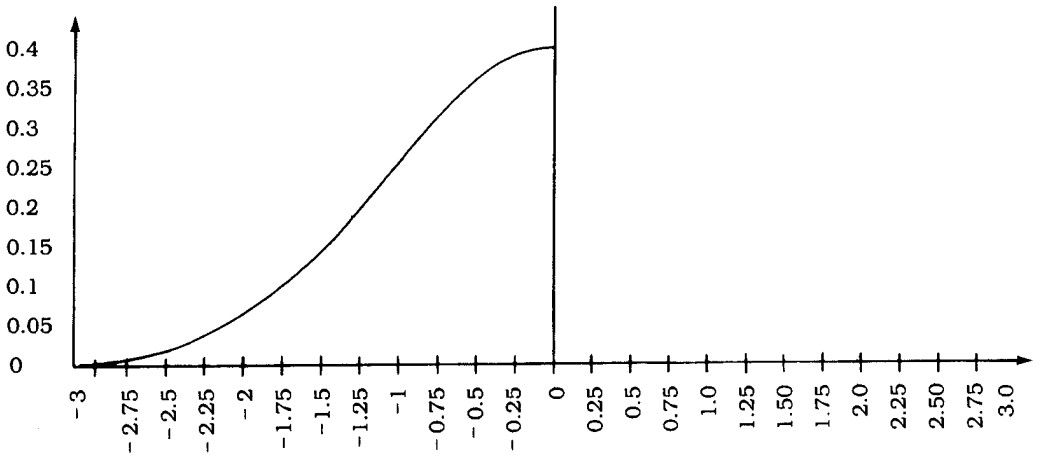
De Moivre (1733) fue el primero en obtener la ecuación de la curva normal. A principios del siglo XIX, Gauss y Laplace ampliaron el desarrollo del concepto de tal curva y el de la probabilidad. Hoy la curva normal se conoce también como *curva de Gauss*, *curva de De Moivre*, *curva de la campana*, etcétera.

Este concepto se retomará en el capítulo 6, *Distribuciones probabilísticas*, por eso explicaremos algunos ejercicios con dicha función.

Ejemplo Marque en la gráfica el área que se indica en el siguiente intervalo:

$$\{x \mid x \leq 0\}$$

Campana de Gauss



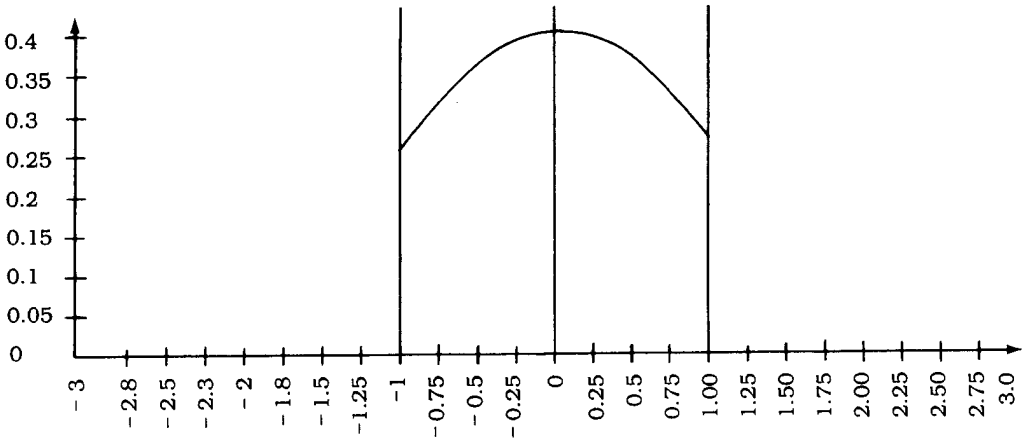
Ejemplo Marque en la gráfica el área que se indica en el siguiente intervalo:

$$\{x \mid x \leq -1 \quad \text{o bien} \quad x \geq 1\}$$

También se escribe de la siguiente forma:

$$\{x \mid -1 \geq x \geq +1\}$$

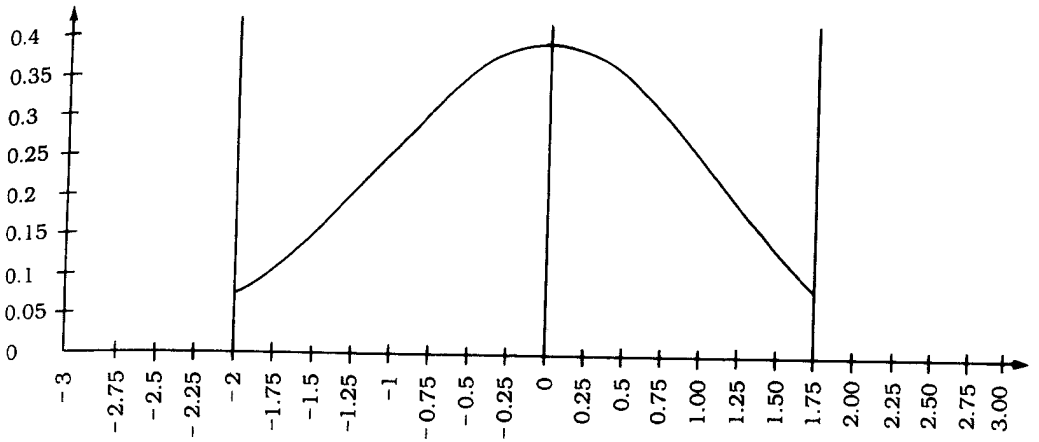
Campana de Gauss



Ejemplo Marque en la gráfica el área que se indica en el intervalo siguiente:

$$\{x \mid -2 < x < 2\}$$

Campana de Gauss



Logaritmos En la sección anterior se explicó la definición siguiente:

$$f = \{(x, y) \mid y = b^x, x \in \mathbb{R}\}$$

donde: $b \neq 1$

y = es cualquier número positivo dado, entonces existe un número real único x.

Se dice que el número x es el logaritmo de y de base b y lo escribimos:

$$x = \text{Log}_b\, y \quad \Rightarrow \quad f(x) = \text{Log}_b\, x$$

Esta definición se puede enunciar en forma un poco más concisa como sigue:

Definición Si y y b son números positivos y si $b \neq 1$, entonces

$$\text{Log}_b\, x \quad = \quad f(x) \qquad \text{si y sólo si,} \qquad y = b^x$$

Según esto, notamos que el concepto de un exponente y el de un logaritmo son simplemente dos formas diferentes de definir el mismo concepto. Las dos ecuaciones $y = b^x$ y $f(x) = \text{Log}_b\, x$ son equivalentes y podemos cambiar de una a otra para obtener la forma que resulte más conveniente a nuestro propósito.

Si se utiliza la base $e = 2.718218$, entonces se hablará de un *logaritmo natural* (*Ln*) y se denota como sigue:

$$f(x) = \text{Log}_e\, x \quad \Rightarrow \quad f(x) = Lnx$$

Aplicando la definición de logaritmo, queda de la forma siguiente.

$$y \;=\; Lnx \quad \text{si y sólo si} \quad x = e^y$$

3.8.5 Álgebra de funciones

Las operaciones elementales para funciones son las siguientes:

Sea f_1 y f_2 dos funciones; entonces:

$(f_1 + f_2)(x) = f_1(x) + f_2(x)$, suma de dos funciones.
$(f_1 - f_2)(x) = f_1(x) - f_2(x)$, resta de dos funciones.
$(f_1 \ f_2)(x) = f_1(x) f_2(x)$, producto de dos funciones.
$(f_1/f_2)(x) = f_1(x)/f_2(x)$, división de dos funciones.
$(f_1 \circ f_2)(x) = f_1(f_2(x))$, función compuesta.

El dominio de cualquier operación con funciones es la intersección de sus dominios.

Ejemplos Realizar las cinco operaciones algebraicas de las funciones siguientes:

$$f_1 = 3x + 6, \qquad f_2 = 2x + 5$$

i) $(f_1 + f_2)(x) = (3x + 6) + (2x + 5) = 5x + 11$

ii) $(f_1 - f_2)(x) = (3x + 6) - (2x + 5) = 3x + 6 - 2x - 5 = x + 1$

iii) $(f_1 f_2)(x) = (3x + 6)(2x + 5) = 6x^2 + 15x + 12x + 30 = 6x^2 + 27x + 30$

$= 6x^2 + 27x + 30$

iv) $(f_1/f_2)(x) = \dfrac{3x + 6}{2x + 5}$

v) $(f_1 \circ f_2)(x) = f_1(2x + 5) = 3(2x + 5) + 6$
$= 6x + 15 + 6 = 6x + 21$

3.9 MATRICES

El manejo del álgebra matricial es importante para la resolución de diversos problemas de estadística y para la comprensión de algunos capítulos posteriores. Para eso es el desarrollo de la teoría elemental de las matrices y su relación con algunos modelos estadísticos lineales.

Concepto Una matriz es un conjunto (arreglo) de elementos ordenados tanto en renglones como en columnas. Los elementos de una matriz se denotan por a_{ij}, los renglones mediante la letra m y las columnas por la letra n. Así:

Sea una matriz que tiene los renglones – 1, 2 y 0, 1; entonces, $m = 2$ y $n = 2$ es una matriz 2×2, y se representa en la forma siguiente:

$$A_{2 \times 2} = \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix}$$

3.9.1 Orden de una matriz

El orden de una matriz estará expresado mediante un par ordenado, donde el primer componente nos indica los renglones que tiene la matriz y el segundo componente las columnas.

Ejemplo

$$B_{2\times 3} = \begin{bmatrix} 3 & 1 & 0 \\ -2 & -4 & -1 \end{bmatrix} \qquad \text{La matriz es de orden } 2 \times 3$$

$$C_{1\times 3} = [0,\ 1,\ -1] \qquad \text{La matriz es de orden } 1 \times 3$$

$$E_{3\times 1} = \begin{bmatrix} 6 \\ 7 \\ -9 \end{bmatrix} \qquad \text{La matriz es de orden } 3 \times 1$$

Una matriz de orden $m \times n$, será:

$$A_{m\times n} = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1j} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2j} \ldots a_{2n} \\ . & \\ . & \\ . & \\ a_{i1} & a_{i2} \ldots a_{ij} \ldots a_{in} \\ . & \\ . & \\ . & \\ a_{m1} & a_{m2} \ldots a_{mj} \ldots a_{mn} \end{bmatrix}$$

Las matrices son básicas en la teoría económica contemporánea y también en el análisis factorial, cuya aplicación es muy amplia tanto en psicología como en la administración moderna y, por consiguiente, en la investigación, proceso industrial y control de la calidad total. En las ciencias de la salud, los médicos y biólogos utilizan las matrices como un método muy útil en el estudio de interrelaciones genéticas de la herencia.

3.9.2 Tipos de matrices

Las matrices toman diferentes denominaciones, atendiendo a su orden.

Ejemplo *Matriz renglón o vector renglón* $A_{1\times n}$

$$A_{1\times 4} = (2,\ -1,\ 0,\ 1),$$

matriz columna $A_{m \times 1}$

$$A_{4 \times 1} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 2 \end{bmatrix} \qquad A_{3 \times 1} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix} \qquad A_{2 \times 1} = \begin{bmatrix} 5 \\ -4 \end{bmatrix}$$

También se conoce como *vector columna.*

3.9.2.1 Matriz cuadrada

Se denomina así cuando tiene el mismo número de renglones y de columnas, o sea $m = n$.

Ejemplo

$$A_2 = \begin{bmatrix} -2 & 1 \\ 0 & 2 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 0 & 1 & 3 \\ 2 & -1 & 0 \\ 1 & -2 & 0 \end{bmatrix}$$

3.9.2.2 Matriz diagonal

Es una matriz cuadrada, cuyos elementos $a_{ij} = 0$ para toda $i \neq j$.

Ejemplo

$$\begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

En este tipo de matrices, los únicos valores diferentes de cero se encuentran sobre la diagonal principal (de izquierda a derecha).

3.9.2.3 Matriz escalar

Es una matriz diagonal, en la que los a_{ij} son iguales.

Ejemplo

$$\begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} \qquad \begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -4 \end{bmatrix}$$

Por lo común, la matriz escalar se representa de la manera siguiente:

$$\begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix} = \lambda \qquad \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = 1$$

3.9.2.4 Matriz identidad

Llamada también *matriz unitaria*, es una matriz escalar donde $\lambda = 1$ y se representa por I, de tal manera que para conocer el orden se le coloca el subíndice respectivo.

Ejemplo

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3.9.2.5 Matriz nula

Es aquella matriz en la cual todos sus elementos son iguales a cero.

3.9.2.6 Igualdad de matrices

Decimos que dos matrices, A y B, son iguales cuando son del mismo orden y todos y cada uno de sus elementos tienen el mismo valor algebraico.

Ejemplo

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 \\ -1 & x \end{bmatrix}$$

$A = B$ si y sólo si $x = 2$

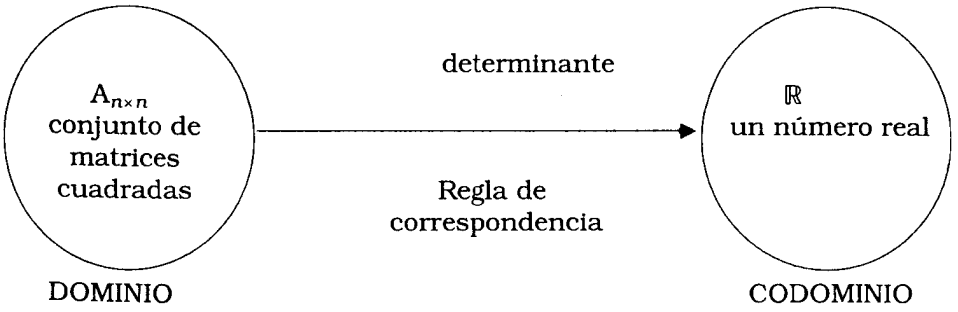
$$C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \qquad D = \begin{bmatrix} w & x \\ y & z \end{bmatrix}$$

$C = D$ si y sólo si $w = 1$, $x = 2$, $y = 3$, $z = 4$

3.9.3 Determinantes

Antes de abordar el desarrollo del álgebra de matrices se explicará el concepto y algunas propiedades de los determinantes. Éstos se estudiarán más adelante.

Un determinante es un número real asociado con una matriz cuadrada mediante una función llamada *determinante*, denotada como det A, $|A|$, Δ siendo A una matriz cuadrada. El dominio es el conjunto de las matrices cuadradas y el codominio es un número real, también llamado *escalar*.



Ahora explicaremos cómo se calcula el determinante de una matriz 2 × 2:
Es el producto de los elementos de la diagonal principal, *menos* el producto de los elementos de la otra diagonal.

Ejemplo

$$\begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix} = A \qquad \det A = \begin{vmatrix} 2 & 1 \\ 3 & 0 \end{vmatrix} = 2 \times 0 - 3 \times 1 = -3$$

$$\det\ A = -3$$

Se obtuvo como resultado un número real, y es importante mencionar que este número real es único para cada matriz.

A continuación explicaremos cómo se calcula el determinante para matrices de 3 × 3. Para ello es necesario dar a conocer dos conceptos.

Sea A una matriz de 3 × 3

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Como podemos observar, cada elemento tiene su posición ij, donde i = número de renglón o también llamada *fila*, y j = número de columna.

Definición Si A es una matriz cuadrada de orden 3, entonces el menor M_{ij} de un elemento a_{ij} es el determinante de la matriz de orden 2 que se obtiene al omitir el renglón i y la columna j de A.

Definición El cofactor A_{ij} del elemento a_{ij} se define como

$$A_{ij} = (-1)^{i+j} M_{ij}$$

Explicaremos las definiciones anteriores con el ejemplo siguiente.
Sea

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -8 & -1 & 7 \\ -6 & 4 & 0 \end{bmatrix}$$

Como primer paso, se calculan los menores, es decir, M_{ij}:

$$M_{11} = \begin{vmatrix} -1 & 7 \\ 4 & 0 \end{vmatrix} = (-1)(0) - (4)(7) = -28$$

$$M_{12} = \begin{vmatrix} -8 & 7 \\ -6 & 0 \end{vmatrix} = (-8)(0) - (-6)(7) = +42$$

$$M_{13} = \begin{vmatrix} -8 & -1 \\ -6 & 4 \end{vmatrix} = (-8)(4) - (-1)(-6) = -32 - 6 = -38$$

El segundo paso consiste en calcular los cofactores A_{ij}:

$$A_{11} = (-1)^{1+1} \qquad M_{11} = (-1)^2\,(28) = -28$$
$$A_{12} = (-1)^{1+2} \qquad M_{12} = (-1)^3\,(42) = -42$$
$$A_{13} = (-1)^{1+3} \qquad M_{13} = (-1)^4\,(-38) = -38$$

El tercer y último paso es obtener el determinante por medio de la siguiente definición:

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\,A_{11} + a_{12}\,A_{12} + a_{13}\,A_{13}$$

En notación de suma podemos expresarlo de la siguiente forma:

$$\det A = \sum_{k=1}^{3} a_{1k}\,A_{1k}$$

Sustituimos los valores y obtenemos:

$$\det A = 1\,(-28) + 2\,(-42) + 3(-38) = -28 - 84 - 114 = -226$$

$$\boxed{\det A = -226}$$

3.9.4 Álgebra de matrices

3.9.4.1 Transpuesta de una matriz

Es la matriz que resulta de intercambiar el primer renglón con la primera columna, el segundo renglón con la segunda columna y así sucesivamente. El orden de la matriz queda intercambiado: si es de orden (2 × 3), quedará (3 × 2). La transpuesta de una matriz A se representa por A^T.

Ejemplo

$$A_{2\times 2} = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix} \qquad A^T{}_{2\times 2} = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$$

$$E_{2\times 3} = \begin{bmatrix} 1 & 3 & -2 \\ 0 & 1 & -1 \end{bmatrix} \qquad E^T{}_{3\times 2} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \\ -2 & -1 \end{bmatrix}$$

3.9.4.2 Matriz simétrica

Se denomina así cuando una matriz es igual a su transpuesta $A = A^T$.

Ejemplo

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 3 \end{bmatrix} \qquad A^T = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 2 & -1 & 3 \end{bmatrix}$$

Obsérvese que esto sólo es posible cuando las matrices son del mismo orden, o sea, una matriz simétrica es cuadrada.

3.9.4.3 Operaciones con matrices

Suma de matrices

Para que la suma de dos matrices se pueda realizar, se requiere que sean del mismo orden.

Ejemplo

$$\begin{bmatrix} -2 & 1 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 3 & -1 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} -2+3 & 1+(-1) \\ 0+0 & 3+(-2) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

La suma algebraica se realiza con los elementos correspondientes. Sean:

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ -1 & 3 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & -1 & 1 \\ 1 & 2 & 2 \\ 1 & 0 & 4 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ -1 & 3 & 1 \end{bmatrix} + \begin{bmatrix} 3 & -1 & 1 \\ 1 & 2 & 2 \\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1+3 & 2+(-1) & -1+1 \\ 0+1 & 1+2 & 0+2 \\ -1+1 & 3+0 & 1+4 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 2 \\ 0 & 3 & 5 \end{bmatrix}$$

Propiedades de la suma de matrices:

1. $A + B = B + A$ Propiedad conmutativa
2. $(A + B) + C = A + (B + C)$ Propiedad asociativa

3.9.4.4 Multiplicación de matrices

Producto de una matriz por un escalar

El resultado de multiplicar una matriz por cualquier número real (escalar) es el producto de cada uno de los componentes por dicho escalar.

Sea multiplicar el escalar (5) por la matriz A, donde:

$$A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & 3 & 0 \\ -1 & 1 & 4 \end{bmatrix} \qquad 5A = \begin{bmatrix} 10 & 0 & -5 \\ 5 & 15 & 0 \\ -5 & 5 & 20 \end{bmatrix}$$

Matriz opuesta

Se obtiene cuando una matriz se multiplica por (–1).

Ejemplo

$$(-1)\begin{bmatrix} 5 & -3 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} -5 & 3 \\ -2 & -6 \end{bmatrix}$$

Resta de matrices

Es la suma algebraica de una matriz dada con la matriz opuesta de la matriz que se quiere restar, o sea:

$$A - B = A + (-1)B$$

Ejemplo Sean las matrices A y B, tales que:

$$\begin{bmatrix} 7 & 3 \\ 4 & -2 \end{bmatrix} - \begin{bmatrix} 5 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 4 & -2 \end{bmatrix} + (-1)\begin{bmatrix} 5 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 4 & -2 \end{bmatrix} + \begin{bmatrix} -5 & 1 \\ -1 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 4 \\ 3 & -4 \end{bmatrix} = A - B$$

Producto de dos matrices

En la multiplicación de dos matrices. $(A \times B)$, el número de columnas de A debe ser igual al número de renglones de B.

Ejemplo Sean A y B tales que:

$$\begin{array}{ccc} A & B & (A_{2\times 2}) \times (B_{2\times 3}) \end{array}$$

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & -1 & 3 \\ 2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 2(1) + 1(2) & 2(-1) + 1(1) & 2(3) + 1(0) \\ 3(1) + 4(2) & 3(-1) + 4(1) & 3(3) + 4(0) \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -1 & 6 \\ 11 & 1 & 9 \end{bmatrix} = C_{2\times 3}$$

Así, si A es de orden $m \times n$ y B es de orden $n \times p$, el producto es una matriz C es de orden $m \times p$. Sean las siguientes matrices; realizar sus productos:

$$(-1,\ 2,\ 1)\begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} = [(-1)3 + 2(1) + 1(4)] = [3]$$

$$\begin{bmatrix} 5 \\ -2 \end{bmatrix}[1, -1, 3] = \begin{bmatrix} 5(1) & 5(-1) & 5(3) \\ -2(1) & -2(-1) & -2(3) \end{bmatrix} = \begin{bmatrix} 5 & -5 & 15 \\ -2 & 2 & -6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1(1) + 0(0) & 1(0) + 0(1) \\ 2(1) + 3(0) & 2(0) + 3(1) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$$

Cualquier matriz A multiplicada por la unitaria, será la misma matriz A.
Propiedades del producto de matrices.

i)	$A \times B \neq B \times A$	No cumple la propiedad conmutativa
ii)	$A \times (B \times C) = (A \times B) \times C$	Propiedad asociativa
iii)	$A \times (B + C) = A \times B + A \times C$	Propiedad distributiva del producto respecto de la suma y la resta
	$A \times (B - C) = A \times B - A \times C$	

Inversa de una matriz

Es importante señalar que la división entre matrices no está definida.

Para resolver este problema es necesario definir la inversa multiplicativa, también llamada *inversa de una matriz* que, multiplicada por la original, da como resultado la unitaria o identidad (I); o sea:

$$\frac{A}{B} = I \quad \text{siempre que} \quad AA^{-1} = I \quad AB = I \quad B = A^{-1}$$

$B = A^{-1}$ (siempre que la matriz inversa de A exista; no todas las matrices tienen inversa). Las condiciones que debe cumplir una matriz para tener inversa, son las siguientes:

1. Que sea una matriz cuadrada ($A_{m \times m}$).
2. El determinante de A (det $A \neq 0$) sea diferente de cero; cuando esto ocurre se dice que la matriz A es *no singular*.
3. Si det $A = 0$, A es una matriz singular y no tiene inversa.

Procedimiento para formar la inversa de A por el método de cofactores:

1. Reemplazar cada elemento de A por su cofactor, tomando en cuenta la siguiente relación de signos:

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}$$

Este caso es para una matriz de orden 3×3.

2. Dividir cada elemento de la nueva matriz, formada por sus cofactores, por el determinante de A, ($|A|$).
3. Obtener su transpuesta, A^T
4. Comprobar que $AA^{-1} = I$

Ejemplo

Obtener la inversa de $A = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & -3 \\ 3 & -2 & 1 \end{bmatrix}$

Paso 1 Calcular $|A|$

Como se tiene un renglón con dos ceros, entonces el determinante de A es simplemente:

$$|A| = -(-3)\begin{vmatrix} 1 & -1 \\ 3 & -2 \end{vmatrix} = 3(-2+3) = 3$$

considerando los signos de referencia

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}$$

Paso 2 Obtener la matriz de cofactores:

$$\begin{bmatrix} \begin{vmatrix} 0 & -3 \\ -2 & 1 \end{vmatrix} & -\begin{vmatrix} 0 & -3 \\ 3 & 1 \end{vmatrix} & \begin{vmatrix} 0 & 0 \\ 3 & -2 \end{vmatrix} \\ -\begin{vmatrix} -1 & 2 \\ -2 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & -1 \\ 3 & -2 \end{vmatrix} \\ \begin{vmatrix} -1 & 2 \\ 0 & -3 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 0 & -3 \end{vmatrix} & \begin{vmatrix} 1 & -1 \\ 0 & -0 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} -6 & -9 & 0 \\ -3 & -5 & -1 \\ 3 & 3 & 0 \end{bmatrix}$$

Paso 3 La matriz de cofactores se divide entre el determinante de A, o sea $|A| = 3$, y se obtiene:

$$\begin{bmatrix} -2 & -3 & 0 \\ -1 & -\frac{5}{3} & -\frac{1}{3} \\ 1 & 1 & 0 \end{bmatrix}$$

Paso 4 Se obtiene la transpuesta:

$$\begin{bmatrix} -2 & -1 & 1 \\ -3 & -\frac{5}{3} & 1 \\ 0 & -\frac{1}{3} & 0 \end{bmatrix}$$

Paso 5 Se establece la inversa:

$$A^{-1} = \begin{bmatrix} -2 & -1 & 1 \\ -3 & -\frac{5}{3} & 1 \\ 0 & -\frac{1}{3} & 0 \end{bmatrix}$$

Paso 6 Se comprueba:

$$AA^{-1} = I$$

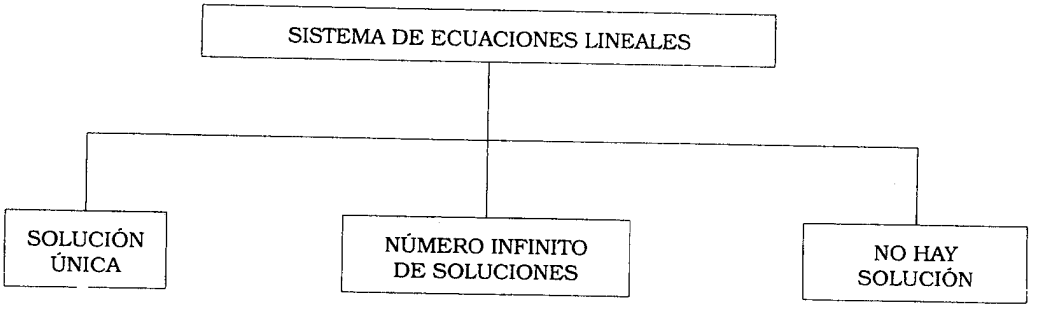
$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & -3 \\ 3 & -2 & 1 \end{bmatrix} \begin{bmatrix} -2 & -1 & 1 \\ -3 & -\frac{5}{3} & 1 \\ 0 & -\frac{1}{3} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3.10 SISTEMA DE ECUACIONES LINEALES

Sea $Ax = B$ un sistema de ecuaciones lineales; en forma matricial tenemos lo siguiente:

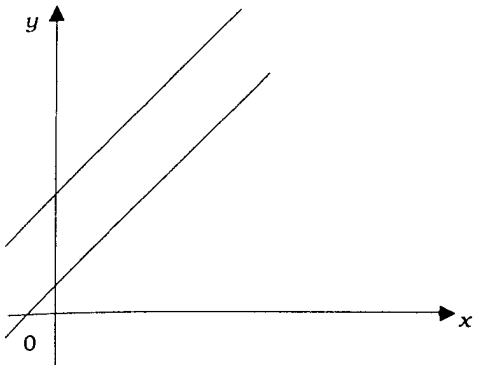
A = es la matriz cuyos elementos son los coeficientes de las variables.
x = es el vector columna, que forman las n variables.
B = es el vector columna, que contiene el valor de las constantes, también llamados *términos independientes.*

El sistema $Ax = B$ puede presentar una de las tres situaciones siguientes:



Los tres casos anteriores se explicarán geométricamente, con sólo dos ecuaciones y dos incógnitas, para su mejor comprensión.

CASO 1



$$a_{11}\,x + a_{12}\,y = b_1$$
$$a_{21}\,x + a_{22}\,y = b_2$$

No hay solución; las rectas tienen pendientes iguales, es decir, $m_1 = m_2$; no existe punto de intersección.

Ejemplo Explicar si el sistema de ecuaciones tiene solución:

$$2x - 8y = 5 \qquad \ldots \text{(I)}$$
$$-3x + 12y = 8 \qquad \ldots \text{(II)}$$

Se despeja la variable y de ambas expresiones:

Ecuación (I)

$$2x - 8y = 5$$

$$-8y = 5 - 2x$$

$$y = \frac{5 - 2x}{-8}$$

$$y = \frac{2}{8}x - \frac{5}{8}$$

$$\boxed{y = \frac{1}{4}x - \frac{5}{8}}$$

Ecuación (II)

$$-3x + 12y = 8$$
$$12y = 8 + 3x$$

$$y = \frac{8 + 3x}{12}$$

$$y = \frac{3}{12}x + \frac{8}{12}$$

$$\boxed{y = \frac{1}{4}x - \frac{5}{8}} \qquad \ldots \text{(I')}$$

$$\boxed{y = \frac{1}{4}x + \frac{2}{3}} \qquad \ldots \text{(II')}$$

Una vez que se obtuvo la ecuación lineal de la forma

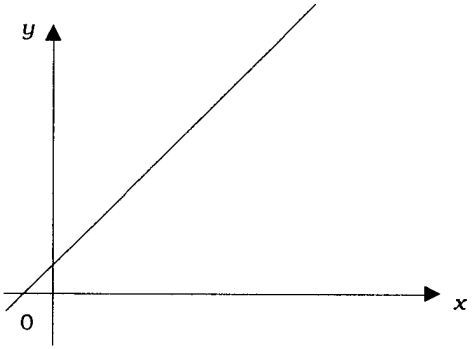
$$y = mx + b,$$

se puede concluir que ambas ecuaciones (I') y (II') tienen la misma pendiente:

$$m_1 = m_2 = \frac{1}{4}$$

No existe punto de intersección; por tanto, el sistema de ecuaciones no tiene solución.

CASO 2



$$a_{11}\,x + a_{12}\,y = b_1$$
$$a_{21}\,x + a_{22}\,y = b_2$$

El número de soluciones es infinito; las rectas son coincidentes.

Ejemplo Explicar si el sistema de ecuaciones tiene solución.

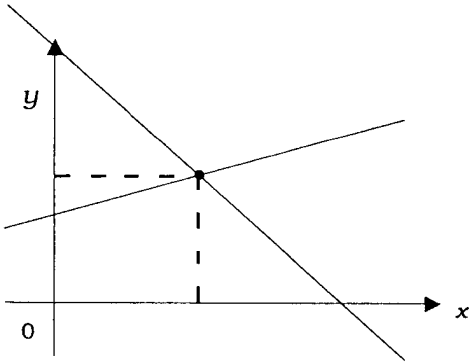
$$4x - 6y = 0 \qquad \ldots \text{(I)}$$
$$-2x + 3y = 0 \qquad \ldots \text{(II)}$$

Multiplicamos la ecuación (II) por (– 2), obtenemos la ecuación (I):

$$(-2)\,(-2x) + (-2)\,(3y) = (-2)\,(0)$$
$$4x - 6y = 0$$

Luego entonces, es la misma recta solamente multiplicada por un escalar, es decir, el sistema de ecuaciones tiene un número infinito de soluciones.

CASO 3



$$a_{11}x + a_{12}y = b_1$$
$$a_{21}x + a_{22}y = b_2$$

El sistema tiene solución única; las pendientes de las rectas son diferentes.

$$m_1 \neq m_2$$

Ejemplo Resolver el sistema de ecuaciones:

$$3x + 4y = 3 \quad \ldots \text{(I)}$$
$$x - 2y = -4 \quad \ldots \text{(II)}$$

Se resolverá por el método de igualación.

Paso 1. Se despeja y de la ecuación (I) y se obtiene:

$$y = \frac{3}{4} - \frac{3}{4}x$$

Paso 2. Se despeja y de la ecuación (II) y se obtiene:

$$y = 2 + \frac{x}{2}$$

Paso 3. Se igualan las variables despejadas; por eso el método se llama *de igualación,*

Como $y = y$ entonces

$$\frac{3}{4} - \frac{3}{4}x = 2 + \frac{x}{2}$$

Paso 4. Se despeja la variable x de la expresión anterior:

$$\frac{3}{4} - \frac{3}{4}x = 2 + \frac{x}{2}$$

$$\frac{3}{4} - \frac{8}{4} = \frac{3}{4}x + \frac{1}{2}x$$

$$-\frac{5}{4} = \frac{5}{4}x$$

$$x = -1$$

Paso 5. Una vez que se encontró el valor de una de las variables, se sustituye ese número en cualquiera de las dos ecuaciones. Se considera la expresión (I) para encontrar el valor de y:

$$-1 \quad -2y = -4$$

$$-2y = -4 + 1$$

$$y = \frac{3}{2}$$

Paso 6. Se forma el punto que se buscaba:

$$(x, y) = (-1, 3/2)$$

Paso 7. Se comprueba que dicho punto satisfaga las dos ecuaciones:

Ecuación (I)

$$3(-1) + 4\left(\frac{3}{2}\right) = 3$$

$$-3 + 6 = 3$$

$$3 = 3$$

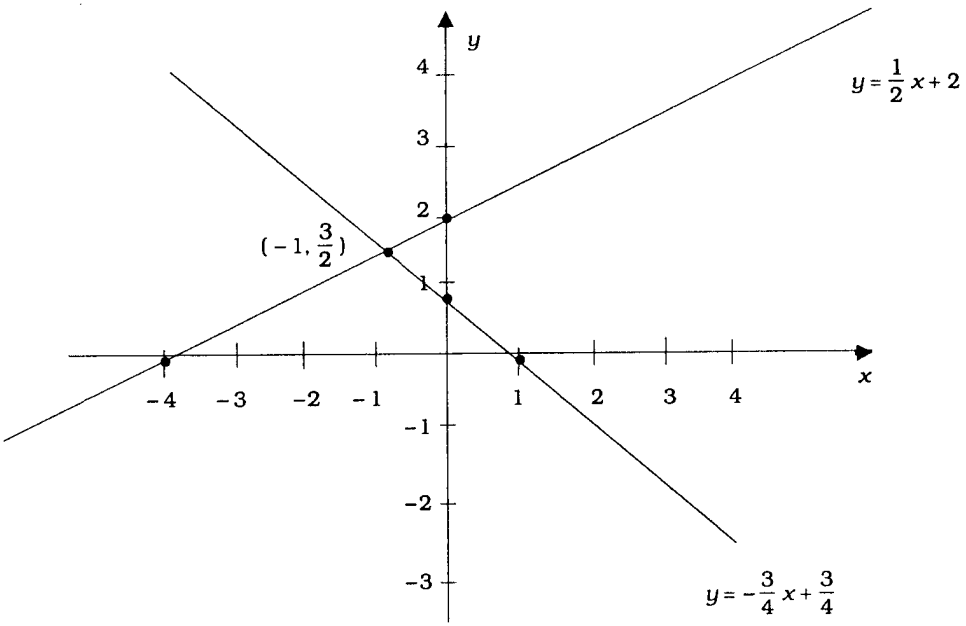
Ecuación (II)

$$(-1) - 2\left(\frac{3}{2}\right) = -4$$

$$-1 - 3 = -4$$

$$-4 = -4$$

Si se grafican las ecuaciones anteriores se obtiene lo siguiente:



Al dibujo anterior se le conoce como el *método gráfico.*

Ejemplo Resolver el siguiente sistema de ecuaciones:

$$2x - 4y = -16 \qquad \text{... (I)}$$
$$x + y = 1 \qquad \text{... (II)}$$

Se resolverá por el método de reducción, también llamado *suma* o *resta.*

Paso 1. Se multiplica por (– 2) a la ecuación (II) y el sistema queda de la manera siguiente:

$$2x - 4y = -16 \qquad \ldots \text{(I')}$$
$$-2x - 2y = -2 \qquad \ldots \text{(II')}$$

Paso 2. Se suman las expresiones (I') y (II'):

$$\begin{array}{r} 2x - 4y = -16 \\ -2x - 2y = -2 \\ \hline 0 - 6y = -18 \end{array}$$

Paso 3. Del resultado de la suma se obtuvo una ecuación con una incógnita; ahora despejamos:

$$-6y = -18$$

$$y = \frac{-18}{-6}$$

$$y = 3$$

Paso 4. Se sustituye el valor de $y = 3$ en cualquiera de las dos ecuaciones. Se considera la ecuación (II), $x + y = 1$

$$x + 3 = 1$$
$$x = 1 - 3$$
$$x = -2$$

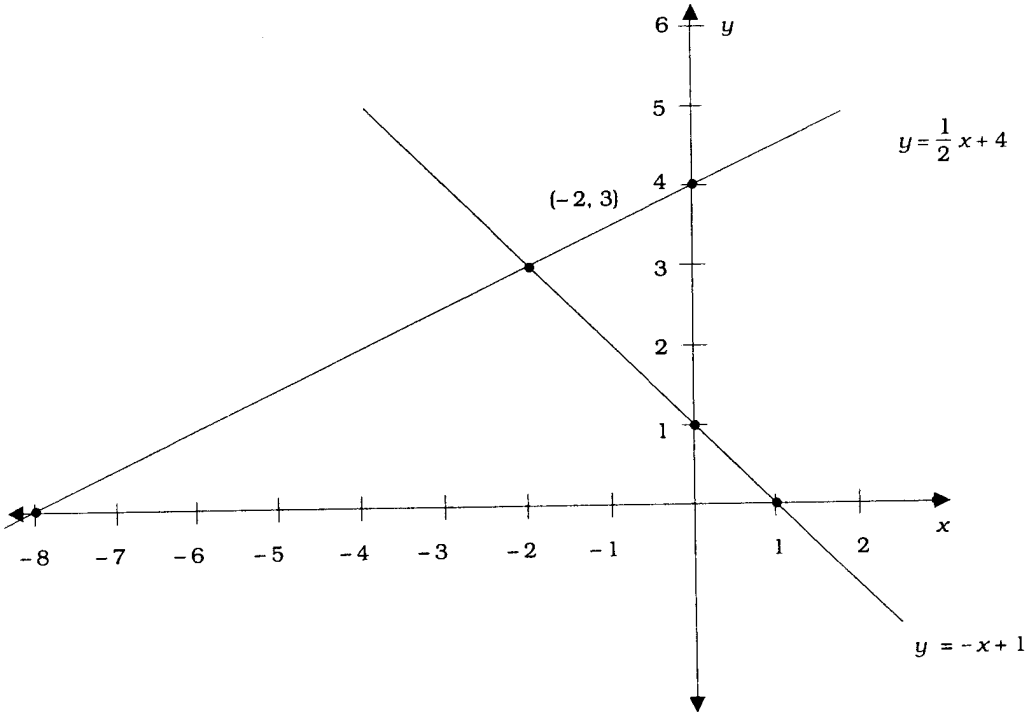
Paso 5. Se obtiene el punto $(x, y) = (-2, 3)$ y se comprueba en (I) y en (II):

(I)

$$2(-2) - 4(3) = -16$$
$$-4 - 12 = -16$$
$$-16 = -16$$

(II)

$$-2 + 3 = 1$$
$$1 = 1$$



Ejemplo Resolver el sistema de ecuaciones por el método de determinantes.

$$2x + 3y = 4$$
$$3x + 4y = 5$$

Paso 1. Se calcula el determinante formado por los coeficientes de las variables:

$$\Delta = \begin{vmatrix} 2 & 3 \\ 3 & 4 \end{vmatrix} = (2)\,(4) - (3)\,(3) = 8 - 9 = -1$$

$$\Delta = -1$$

Se puede presentar el caso donde $\Delta = 0$, lo cual indica que el sistema no tiene una solución única, o no tiene solución.

Paso 2. Para conocer el valor de la variable x se calcula el determinante formado por los términos constantes en la primera columna y los coeficientes de la variable y en la segunda columna; éste se divide entre el determinante calculado en el paso 1:

$$x = \frac{\begin{vmatrix} 4 & 3 \\ 5 & 4 \end{vmatrix}}{-1} = \frac{(4)\,(4) - (5)\,(3)}{-1} = \frac{16 - 15}{-1} = \frac{1}{-1} = -1$$

$$x = -1$$

Paso 3. En forma análoga calculamos el valor de la variable y. El determinante contendrá en la primera columna los coeficientes de las variables x y en la segunda columna los términos independientes, también llamados *constantes.*

$$y = \frac{\begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix}}{-1} = \frac{10-12}{-1} = \frac{-2}{-1} = 2$$

Paso 4. Los valores $(x, y) = (-1, 2)$ cumplen con ambas ecuaciones propuestas; ahora comprobamos:

(I)

$$\begin{aligned} 2(-1) + 3(2) &= 4 \\ -2 + 6 &= 4 \\ 4 &= 4 \end{aligned}$$

(II)

$$\begin{aligned} 3(-1) + 4(2) &= 5 \\ -3 + 8 &= 5 \\ 5 &= 5 \end{aligned}$$

El método de Cramer puede extenderse a sistemas de n ecuaciones con n incógnitas; la mecánica del cálculo sufre cambios.

Para sistemas de tres ecuaciones y tres incógnitas o más no es recomendable su resolución por los métodos de igualación, reducción o gráfico. Existen varios algoritmos para la resolución de sistemas de ecuaciones; en esta obra se explicarán sólo dos: el método de Cramer, también llamado de *determinantes*, y el método de la matriz inversa.[6]

Ejemplo Resolver el sistema de ecuaciones por el método de Cramer:

$$\left| \begin{array}{rrrr} 2x - 3y + & 2z = & -3 \\ -3x + 2y + & z = & 1 \\ 4x + y - & 3z = & 4 \end{array} \right|$$

Paso 1. Se calcula el determinante formado por los coeficientes de las variables:

$$\Delta = \begin{vmatrix} 2 & -3 & 2 \\ -3 & 2 & 1 \\ 4 & 1 & -3 \end{vmatrix} = 2 \begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} + 3 \begin{vmatrix} -3 & 1 \\ 4 & -3 \end{vmatrix} + 2 \begin{vmatrix} -3 & 2 \\ 4 & 1 \end{vmatrix}$$

$$= 2(-6-1) + 3(9-4) + 2(-3-8) = -14 + 15 - 22 = -21$$

[6] Se sugiere utilizar el paquete de cómputo MACSTAT.

Paso 2. Para conocer el valor de la variable x se calcula el determinante formado por los términos constantes en la primera columna y los coeficientes de la variable y en la segunda columna, y la tercera columna está formada por los coeficientes de la variable z. El resultado se divide entre el determinante calculado en el paso 1:

$$x = \frac{\begin{vmatrix} -3 & -3 & 2 \\ 1 & 2 & 1 \\ 4 & 1 & -3 \end{vmatrix}}{\Delta} = \frac{-3\begin{vmatrix} 2 & 1 \\ 1 & -3 \end{vmatrix} + 3\begin{vmatrix} 1 & 1 \\ 4 & -3 \end{vmatrix} + 2\begin{vmatrix} 1 & 2 \\ 4 & 1 \end{vmatrix}}{-21}$$

$$= \frac{-3(-6-1)+3(-3-4)+2(1-8)}{-21} = \frac{+21-21-14}{-21} = \frac{-14}{-21} = \frac{2}{3}$$

Paso 3. Para encontrar el valor de la variable y, el determinante estará formado de la siguiente forma: los elementos de la primera columna serán los coeficientes de la variable x, los elementos de la segunda columna serán las constantes; en la tercera y última los elementos serán los coeficientes de la variable z. El determinante se dividirá entre Δ :

$$y = \frac{\begin{vmatrix} 2 & -3 & 2 \\ -3 & 1 & 1 \\ 4 & 4 & -3 \end{vmatrix}}{\Delta} = \frac{2\begin{vmatrix} 1 & 1 \\ 4 & -3 \end{vmatrix} + 3\begin{vmatrix} -3 & 1 \\ 4 & -3 \end{vmatrix} + 2\begin{vmatrix} -3 & 1 \\ 4 & 4 \end{vmatrix}}{-21}$$

$$= \frac{2(-3-4)+3(9-4)+2(-12-4)}{-21} = \frac{-14-15-32}{-21} = \frac{-31}{-21} = \frac{31}{21}$$

Paso 4. Por último se encontrará el valor de la variable z; en forma análoga, el determinante estará formado de la siguiente manera: la primera y segunda columnas por los coeficientes de las variables x y y, respectivamente, la tercera y última columna se formará con los términos constantes. Como siempre, se divide entre Δ.

$$z = \frac{\begin{vmatrix} 2 & -3 & -3 \\ -3 & 2 & 1 \\ 4 & 1 & 4 \end{vmatrix}}{\Delta} = \frac{2\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} + 3\begin{vmatrix} -3 & 1 \\ 4 & 4 \end{vmatrix} - 3\begin{vmatrix} -3 & 2 \\ 4 & 1 \end{vmatrix}}{-21}$$

$$= \frac{2(8-1)+3(-12-4)-3(-3-8)}{-21} = \frac{14-48+33}{-21} = \frac{-1}{-21} = \frac{1}{21}$$

Paso 5. Concluimos que los valores para las variables son:

$$x = \frac{2}{3} \qquad y = \frac{31}{21} \qquad z = \frac{1}{21}$$

Por último se estudiará la forma de resolver sistemas de ecuaciones lineales por medio de la matriz inversa.

Sabemos que toda matriz multiplicada por su inversa da como resultado la matriz identidad, entonces:

$$Ax = b$$

Si multiplicamos ambos miembros de la igualdad por A^{-1}, en el primer miembro obtendremos el vector x, y en el segundo tendremos $A^{-1}b$, es decir:

$$\begin{aligned} A^{-1}\,Ax &= A^{-1}\,b \\ I\,x &= A^{-1}\,b \\ x &= A^{-1}\,b \end{aligned}$$

Ejemplo Resolver el sistema de ecuaciones por medio de la matriz inversa:

$$\begin{aligned} x - y + 2z &= 1 \\ -3z &= 3 \\ 3x - 2y + z &= 3 \end{aligned}$$

Paso 1. Expresamos el sistema de ecuaciones de la forma

$$Ax = b$$

donde:

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & -3 \\ 3 & -2 & 1 \end{bmatrix} \qquad x = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}$$

Paso 2. Una vez que se conoce A se calcula su inversa:

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & -3 \\ 3 & -2 & 1 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} -2 & -1 & 1 \\ -3 & -5/3 & 1 \\ 0 & -1/3 & 0 \end{bmatrix}$$

Paso 3. Realizamos la operación siguiente:

$$x = A^{-1}\,b$$

$$x = \begin{bmatrix} -2 & -1 & 1 \\ -3 & -\frac{5}{3} & 1 \\ 0 & -\frac{1}{3} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -5 \\ -1 \end{bmatrix}$$

$$x = -2$$

$$y = -5$$

$$z = -1$$

Paso 4. Se comprueba que los resultados obtenidos satisfagan las ecuaciones propuestas.

(I)

$$\begin{aligned} -2-(-5)+2(-1) &= 1 \\ -2+5-2 &= 1 \\ 5-4 &= 1 \\ 1 &= 1 \end{aligned}$$

(II)

$$\begin{aligned} -3(-1) &= 3 \\ 3 &= 3 \end{aligned}$$

(III)

$$\begin{aligned} 3(-2)-2(-5)-1 &= 3 \\ -6+10-1 &= 3 \\ 10-7 &= 3 \\ 3 &= 3 \end{aligned}$$

RESUMEN

Puesto que el investigador emplea la estadística inferencial para generalizar los resultados de una muestra a la población objetivo y ya que tanto la población como la muestra son conjuntos de mediciones, fue necesario desarrollar el tema de conjuntos y funciones. Debido a que la información presente puede interaccionarse utilizando las leyes del álgebra de conjuntos, así como los diagramas de Venn y de Carroll, cuando se considere o no el orden de los elementos que constituyen a los conjuntos, entraremos en el contexto del análisis combinatorio[7] y al álgebra de conjuntos, relaciones y funciones, que son los elementos básicos para desarrollar con éxito la teoría de probabilidad que, entre otras cosas, formaliza a la estadística como modelo matemático donde interviene un elemento aleatorio. Tanto el modelo estadístico como el matemático nos ayudan a describir fenómenos determinísticos e indeterminísticos. Las matrices son el elemento básico para desarrollar los procesos estocásticos y junto con los sistemas de ecuaciones, se podrá construir el modelo de regresión lineal simple y el múltiple.

PROBLEMAS

3.1 ¿Cuál de los siguientes enunciados es un conjunto finito?

a) Los estudiantes que asisten a la Universidad.
b) Los números negativos.

[7] También conocido como *análisis de posibilidades*.

c) Las letras del abecedario.
d) Las mujeres profesionales de un país.
e) Los presidentes de los países latinoamericanos.
f) Todas las chicas hermosas de su grupo.

3.2 Dado el conjunto $\{a, b, c\}$

a) Enuncie todos los subconjuntos y cuántos son.
b) ¿Cuáles de los anteriores son propios?

3.3 Si $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

$$A = \{1, 3, 5, 7, 9\}$$

¿Cuál es A'?

3.4 Dados los siguientes conjuntos, defina cualquier relación que exista entre ellos e ilustre estas relaciones por medio de un diagrama de Venn:

$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
$B = \{1, 3, 5, 7, 9\}$
$C = \{x \mid x \text{ es un número real}\}$

3.5 A 12 personas se les pregunta qué tipo de transporte público utilizan para acudir a su trabajo. Las dos formas de trasladarse son:

(A) = Autobús
(M) = Metro

Construya un diagrama de Venn en el que se resuman los resultados. Suponga que el conjunto universal incluye a las 12 personas, si:
$A = \{1, 2, 4, 7, 9, 10, 12\}$
$M = \{2, 6, 3, 7, 9, 10\}$

3.6 A partir de la información del problema 3.5, enuncie los conjuntos utilizando la definición por construcción.

3.7 Si las personas son las siguientes:

$$U = \{a, b, c, d, e, f, g, h, i, j, k, l\}$$

y las que utilizan metro son:

$$M = \{a, d, e, f, g, h, i\}$$

Las que utilizan autobús son:

$$A = \{b, c, d, i, j, k, l\}$$

De las siguientes expresiones, determine cuál es falsa (F) o verdadera (V):

1. $b \in A$
2. $a \in A$
3. $d \in M$
4. $d \in A$
5. $i \subset M$
6. $C \not\subset M$
7. $i \in M$
8. $i \notin \mu$
9. $A \subset U$
10. $U \supset M$

3.8 De los siguientes conjuntos, diga cuáles son iguales y cuáles similares:

$C = \{1, 3, 5, 7\}$
$D = \{\Delta, \bigcirc, \square, \star\}$
$E = \{1, 3, 5\}$
$F = \{7, 3, 5, 1\}$
$G = \{a, a, a\}$
$H = \{a, a, a, a\}$

3.9 A partir de los datos del problema 3.7, ubíquelos en un diagrama de Venn-Euler.

3.10 En un grupo de 100 estudiantes de secundaria, 60 son varones, de los cuales 45 fuman; de las mujeres, 10 no fuman. Ubíquelos en un diagrama de Carroll.

3.11 Dados los siguientes conjuntos:

$A = \{1, 2, 3, 4, 5\}$
$B = \{1, 3, 5, 7, 9\}$
$C = \{2, 4, 6, 8, 10\}$
$U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

encuentre:

a) $A \cap B$
b) $A \cap C$
c) $B \cap C$
d) $A \cap A'$
e) $A \cap A \cup A \cup B$
f) $B \cap U$
g) $\overline{A \cap B} \cup C$
h) $A \cup B$
i) $A \cap B \cap C$
j) $A \cap (B \cup C)$
k) $C \cap (A \cup B)$
l) $(A \cap C) \cup B$
m) $(A \cap B) \cup C$
n) $\overline{A} \cap \overline{B}$
o) $\overline{A} \cap \overline{B} \cap \overline{C}$

3.12 Del problema anterior, ilustre con un diagrama de Venn-Euler los incisos *a, b, c, d, e, f.*

3.13 Dados los conjuntos:

$A = \{0, 1, 2, 3, 4, 5, -1, -2, -3, -4, -5\}$

(V):

$B = \{-2, -4, 0, 2, 4\}$
$C = \{1, 2, 3, 4, 5\}$

encuentre:

a) $A \cup (B \cup C)$
b) $A \cap (B \cup C)$
c) $(A \cap C) \cup B$
d) $C \cap (A \cup B)$

3.14 Se realizó una encuesta a 1 000 personas en relación con sus adquisiciones en distintas librerías. Los resultados de la investigación revelaron que:

300 personas compraron en la librería A.
200 en la B.
250 en la C.
50 en la A y B.
75 en la A y C.
60 en B y C únicamente.
25 en las 3.

Por medio de un diagrama de Venn-Euler diga:

a) ¿Cuántas personas no compraron en ninguna de las 3?
b) ¿Cuántas compraron únicamente en la A?
c) ¿Cuántas compraron solamente en la B y C?

3.15 Dibuje un diagrama de Venn-Euler para cada conjunto:

a) $(A \cup B)'$
b) $(A' \cap B')^C$
c) $A \cap B - (A \cap B \cap C)$

3.16 En la línea de producción se obtienen 900 bombillas en un día y pueden tener los defectos siguientes:

A = filamento roto
B = rosca dispareja
C = fisura

$n(A) = 39$
$n(B) = 19$
$n(C) = 43$

$n(A \cap B) = 15$
$n(A \cap C) = 21$
$n(B \cap C) = 12$
$n(A \cap B \cap C) = 8$

Exprese, utilizando las leyes entre conjuntos, cuántos artículos hay en cada uno de los siguientes incisos:

a) Los que tienen los defectos tipo B y C, pero no A.
b) Los que tienen defecto tipo B, únicamente.

c) Que no tienen defectos.
d) Aquellos que tienen el defecto tipo A o B o C, pero no tienen B y C.
e) Exactamente 1 defecto.
f) Exactamente 2 defectos.
g) Al menos 1 defecto.
h) Al menos 2 defectos.
i) A lo sumo 1 defecto.
j) A lo sumo 2 defectos.

3.17 En una encuesta a 120 personas, se les preguntó acerca de su preferencia de tres marcas de café (A, B, C); los entrevistados se clasificaron de acuerdo con sus nacionalidades y preferencias.

H = hondureños
M = mexicanos
Cu = cubanos

a) H y $A = 18$
b) H y $B = 11$
c) H y $C = 9$
d) M y $A = 24$
e) M y $B = 8$
f) M y $C = 10$

Cu y $A = 9$
Cu y $B = 19$
Cu y $C = 12$

a) ¿Cuántos entrevistados se encuentran en a, b, c, d, e, f ?
b) ¿Cuántos de cada nacionalidad existen en total?

3.18 A 61 personas (entre profesoras y alumnos) de una escuela se les preguntó sobre su color favorito:

M = niños
F = niñas
P = profesoras

V = violeta
R = rojo
A = azul
a = amarillo

hay 10 niños a quienes les gusta el rojo
a seis niñas les encanta el rojo
a tres profesores les gusta el rojo
a cinco niños les agrada el azul
a tres niños les gusta el color amarillo
a cuatro niños les agrada también el violeta
11 niñas prefieren el color azul
a 12 niñas también les gusta el amarillo

una niña prefiere el violeta
para cuatro profesores, el azul es su color favorito
a un profesor le agrada el amarillo
a un profesor le gusta el color violeta

1. ¿Cuántos elementos hay en cada uno de los siguientes incisos?

a) $n(V) =$
b) $n(R \cup A) =$
c) $n(M \cup a) =$
d) $n[P \cap (A \cup a)] =$
e) $n(F \cap V)' =$
f) $n[M \cup (F \cup A')] =$
g) $n(F \cup V)' =$

2. Indique cuántos elementos hay en:

a) $n(M \cap R) =$
b) $n(F \cap R) =$
c) $n(M \cap R) =$
d) $n(M \cap A) =$
e) $n(M \cap a) =$
f) $n(M \cap V) =$
g) $n(F \cap A) =$
h) $n(F \cap a) =$
i) $n(F \cap V) =$
j) $n(P \cap A) =$
k) $n(P \cap a) =$
l) $n(P \cap V) =$

3.19 Si A y B son dos conjuntos ajenos, calcule:

a) $n(A \cup B) =$
b) $n(A \cap B) =$

Si A y B son conjuntos mutuamente excluyentes, calcule a y b.

3.20 Existen tres tipos de antígenos denotados por A, B y Rh. Todas las personas están clasificadas doblemente; si tienen el antígeno Rh (+) se denotan por R y si tienen el Rh (−) por R′; pueden ser del tipo AB, A, B o pueden ser O si no tienen el antígeno A ni B. Si Ω es el conjunto de todas las personas sanas:

a) ¿Cuántos grupos sanguíneos existen?
b) Haga una lista de dichos subconjuntos.

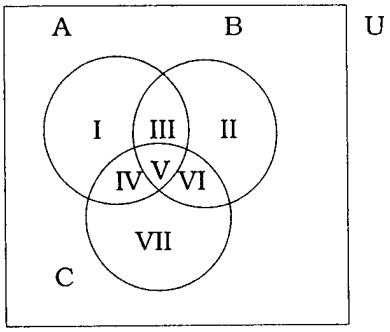
3.21 A 100 personas se les hacen pruebas de sangre y se obtienen los datos siguientes:
48 poseían el antígeno A.
50 el B.
38 el R.
22 los A y B.
10 los A y R.
14 los B y R.
cuatro los tres.

a) ¿Cuántas personas de dicha muestra poseen A y B y R′ ?
b) ¿Cuántas personas del grupo son del tipo O?

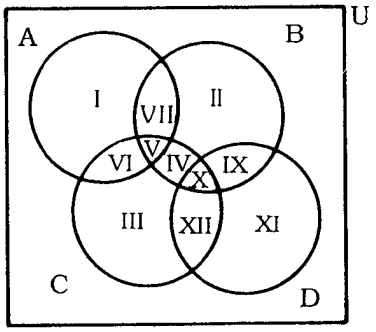
3.22 En un vuelo de cierta línea aérea viajan 18 muchachos, 178 hombres, 10 mexicanos del sexo masculino, dos muchachos mexicanos, 26 personas de nacionalidad mexicana y 14 muchachas extranjeras. ¿Cuántos pasajeros se encuentran a bordo del avión?

3.23 Exprese, utilizando las operaciones y relaciones posibles entre conjuntos, para cada una de las regiones de los siguientes diagramas:

a)



b)



3.24 Las muchachas de un grupo de cierta escuela fueron clasificadas en: A = altas, B = bajas, H = hermosas, L = listas, P = pedantes, T = tontas. Por consiguiente, había:

22 hermosas, listas y altas;
17 hermosas, listas y bajas;
13 altas, pedantes y listas;
cuatro pedantes, altas y tontas;
18 pedantes, listas y bajas;
11 hermosas, tontas y altas;
siete bajas, hermosas y tontas;
cinco tontas, bajas y pedantes

¿Cuántas muchachas hay en:

I) *a)* $A \cup B$? *b)* $H \cup P$? *c)* $A \cap P$? *d)* $A \cup (P \cap L)$? *e)* $A' \cup (P \cap A)$? *f)* $(P \cap L) \cup (H \cap T) \cup B'$? *g)* $(B \cap H \cap L) \cup (B' \cap H' \cap T')$? *h)* En todo el grupo?

II) Escriba el significado de cada uno de los incisos anteriores.

3.25 Dados los siguientes conjuntos:

$A = \{x \mid x = 1, 2, 3, \ldots, 10\}$
$B = \{x \mid 1 \leq x \leq 10\}$
$C = \{x \mid x = 0, 1, 2, 3, 4, 5, 6\}$
$D = \{0, 10, 20, 30\}$

calcule:

a) $A \cup B$.
b) $A \cap B$.
c) $A \cup C$.
d) $A \cap C$.
e) $A \cup D$.
f) $A \cap D$.
g) $B \cup C$.
h) $B \cap C$.
i) $B \cup D$.
j) $B \cap D$.
k) $C \cup D$.
l) $C \cap D$.
m) $A \cup B \cup C$.
n) $A \cap (B \cup C)$.
o) $A \cup (B \cap C)$.
p) $A \cap B \cap C$.
q) $C \cup (A \cap D)$.
r) $(A \cup B) \cap (C \cup B)$.

3.26 Simplifique: *a)* $(A \cup B) \cap (C \cup C')$. *b)* $(A \cup \varnothing) \cap B$. c) $[(A \cup B) \cap A']'$.

3.27 Treinta y tres estudiantes prefieren el sistema trimestral en lugar del sistema semestral. Si la encuesta se realizó a 50 estudiantes de licenciatura y 35 de posgrado:

a) ¿Cuántos estudiantes prefieren el sistema semestral, si 23 estudiantes de licenciatura prefieren el sistema semestral?
b) ¿Cuántos estudiantes de posgrado prefieren el sistema trimestral?

3.28 Una agencia noticiosa vende tres periódicos diferentes: *El imparcial*, *El correo* y *El noticioso*; 70 lectores compran *El imparcial*, 60 *El correo* y 50 *El noticioso*; 17 compran *El imparcial* y *El correo*; 15 *El correo* y *El noticioso*; y 16 *El noticioso* y *El imparcial*, mientras que tres lectores compran los tres periódicos.
¿Cuántos lectores de estos periódicos existen?

3.29 Después que 100 ratones corrieron por un laberinto, un psicólogo reportó los datos siguientes: 50 ratones eran machos, 50 fueron previamente entrenados, 42 tomaron a la izquierda en la primera oportunidad, 21 fueron machos previamente entrenados, 7 machos tomaron a la izquierda, 30 ratones previamente entrenados tomaron a la izquierda y 5 machos previamente entrenados tomaron a la izquierda.

a) Dibuje un diagrama de Venn y determine el número de:
  i) Hembras previamente entrenadas que tomaron a la izquierda.
  ii) Las hembras que tomaron a la derecha.
b) A cada una de las regiones del diagrama anterior, señale las relaciones entre dichos conjuntos.
c) Resuelva el problema utilizando un diagrama de Carroll.

3.30 Dibuje la gráfica del siguiente conjunto de pares ordenados e indique si es función o relación:

$A = \{(4, 1), (-1, -3)\}$

3.31 Encuentre el dominio y contradominio de la siguiente expresión y señale si es función o relación:

$B = \{(3, 5), (2, 5)\}$

3.32 Indique si es función o relación cada uno de los incisos siguientes:

a) $|y| = x$
b) $y = |x - 5|$
c) $y = -|x + 2|$
d) $y = |x| - 5$
e) $y = 5 - |x + 1|$

3.33 Calcule la ecuación de la recta y su pendiente que pasa por los puntos A, B,

donde
$A\ (6, -3)$, $B\ (-7, 3)$

3.34 Construya la ecuación de la recta si la pendiente es $m = 4$ y pasa por el punto $A\ (2, -3)$.

3.35 Sea $y = 4x - 3$; encuentre la pendiente y la ordenada al origen.

3.36 Sea L_1 la recta que pasa por los puntos $A\ (1, 2)$, $B\ (3, 6)$ y sea m_1 su pendiente; sea L_2 la recta que pasa por los puntos $C\ (2, -5)$, $D\ (-1, 7)$ y m_2 su pendiente. Calcule las m_1, m_2, grafique y concluya.

3.37 Dada la recta L_1 que tiene la ecuación $2x - 3y = 12$ y la recta L_2 cuya ecuación es $3x + 3y = 6$, dibuje un esquema de cada recta y obtenga las coordenadas del punto de intersección de ambas rectas:

a) Por el método gráfico.
b) Indique cuáles son sus pendientes y ordenadas al origen.
c) Calcule las coordenadas del punto de intersección de ambas rectas algebraicamente.

3.38 Dados tres puntos: $A\ (-3, -4)$, $B\ (2, -1)$, $C\ (7, 2)$, por medio de sus pendientes determine si los tres se encuentran en la misma línea (colineales).

3.39 Dada la matriz A

$$A = \begin{bmatrix} 4 & 3 & 3 \\ -1 & 0 & -1 \\ -4 & -4 & -3 \end{bmatrix}$$

Calcule $A^2 = A \cdot A$

3.40 Sea la matriz A

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{bmatrix}$$

Calcule A^3.

3.41 Sea A

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -5 \\ 3 & -5 & 6 \end{bmatrix}$$

Calcule la transpuesta de dicha matriz, (A^T).

3.42 Sea A

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -5 \\ 3 & -5 & 6 \end{bmatrix} \text{ y } B = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 4 & -5 \\ 1 & -5 & 0 \end{bmatrix}$$

Compruebe que $AB = (BA)^T$

3.43 Dada la matriz A

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 1 & 1 & -2 \\ 2 & 4 & 6 \end{bmatrix}$$

Calcule A^{-1}

3.44 Sea B

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 1 & 1 & 1 \end{bmatrix}$$

Calcule B^{-1}

3.45 Sea el siguiente sistema de ecuaciones

$2x - 8y = 5$
$-3x + 12y = 8$
Resuélvalo utilizando:

a) Método gráfico.
b) Cualquier método algebraico.

3.46 Dado el siguiente sistema de ecuaciones

$3x + y = 0$
$2x - 3y = 0$
Resuélvalo por

a) Método gráfico.
b) Determinantes.

3.47 Sea el sistema de ecuaciones

$$x - 3y = 4$$
$$-4x + 2y = 6$$

Resuelva por:

a) Método gráfico.
b) Determinantes.

3.48 Resuelva el siguiente sistema de ecuaciones:

$$x - 2y + 3z = 11$$
$$4x + y - z = 4$$
$$2x - y + 3z = 10$$

Por el método de:

a) Determinantes.
b) Usando la matriz inversa.
c) Utilizando el paquete MACSTAT.

3.49 Sea el siguiente sistema

$$3x_1 + 6x_2 - 6x_3 = 9$$
$$2x_1 - 5x_2 + 4x_3 = 6$$
$$-x_1 + 16x_2 - 14x_3 = -3$$

a) Resuelva por medio de la matriz inversa.
b) Utilizando el paquete MACSTAT.

3.50 Sea el siguiente sistema de ecuaciones

$$3x_1 - 6x_2 + 9x_3 = 400$$
$$2x_2 + 5x_3 = -200$$
$$-2x_1 + 4x_2 - 6x_3 = 1$$

a) Resuelva utilizando el paquete MACSTAT.
b) Por matrices.

Cálculo de posibilidades 4

PROPÓSITOS

El propósito central del presente capítulo consiste en que el lector sea capaz de explicar las nociones fundamentales del cálculo de posibilidades; adicionalmente, solucionará los ejercicios que se presentan, para reconocer su aplicación en las ciencias sociales y del comportamiento.

Del mismo modo, el lector podrá:

1. Explicar las siguientes notaciones: experimento, diagramas de árbol, principios de multiplicación y adición, permutaciones, combinaciones, coeficiente multinominal, así como el teorema del binomio.
2. Desarrollar los procedimientos para la aplicación de los principios de conteo, permutaciones, combinaciones y otros más.
3. Generalizar los principios de conteo en la teoría de conjuntos (análisis combinatorio).
4. Desarrollar ejemplos de diagramas de árbol.
5. Aplicar las operaciones en la solución de los ejercicios propuestos en el capítulo.
6. Plantear situaciones reales (cotidianas) donde se aplique el cálculo de posibilidades (análisis combinatorio).
7. Diferenciar entre permutaciones y combinaciones.
8. Reconocer la relación de los contenidos de este capítulo con los que se presentan en los próximos tres capítulos, referentes a modelos no determinísticos o probabilísticos.

INTRODUCCIÓN

En los capítulos anteriores hemos analizado algunos modelos estadísticos y matemáticos, ya que son básicos para la utilización adecuada de la estadística, así como para realizar algunos estudios, investigaciones o experimentos, cuando el resultado esté completamente determinado. Por lo general, a estos *modelos* en los cuales los efectos del cambio (tanto en los parámetros como en las variables de interés) puedan determinarse previamente, aun antes de efectuar dicho experimento, estudio o investigación (también se puede considerar una evaluación), se les conoce como *determinísticos*. Pocas veces se utilizan en situaciones donde participe la conducta del ser humano, debido a que el comportamiento humano no se rige por leyes inexorables análogas a las de la física o la química. Por ello se han desarrollado modelos matemáticos, donde interviene la incertidumbre o aleatoriedad; a estos modelos se les suele conocer como *no determinísticos*, *estocásticos*, *aleatorios* o *probabilísticos*, ya que están basados en la probabilidad de ocurrencia de dichos eventos o fenómenos.

En este capítulo se describirán todos los resultados o eventos posibles que puedan ocurrir en un fenómeno o experimento dado, de allí el nombre de *análisis de posibilidades*, también conocido como *análisis combinatorio*.

En este texto nuestro interés radica en la probabilidad finita, o sea, en los conceptos de probabilidad aplicable a situaciones en las cuales exista un conjunto finito o contable de eventos discretos posibles. Por tanto, nuestro primer paso es desarrollar formas de conteo de dichos eventos posibles.

4.1 EXPERIMENTOS

Cualquier actividad, operación o procedimiento —bien definido— en el cual el resultado se observa y se mide, pero que no puede predecirse con certeza, es un experimento. Según Pavlov, *la observación recoge lo que ofrece la naturaleza; la experimentación, en cambio, toma de la naturaleza lo que desea.*

Se dice que están sujetos a un experimento al azar cuando se efectúan mediciones de varias clases. Es común encontrar que los resultados varían de una observación a otra debido a que no pueden predecirse con exactitud.

Aun el resultado de un único ensayo al azar no puede predecirse; la experiencia con una secuencia de experimentos ha revelado una clase muy útil de regularidad. Un modelo o una estructura matemática que describa tal regularidad es de gran valor, y la teoría de la probabilidad ha desarrollado modelos para estos experimentos aleatorios.

La estadística realiza inferencias del modelo probabilístico verdadero o adecuado para un experimento dado, y utiliza los resultados recientes, pero lleva a cabo el experimento una o varias veces para obtener datos. Por lo general no es posible conocer el modelo verdadero en forma detallada, y esto se debe a que uno obtiene los datos y los utiliza como la base para llegar a conclusiones acerca

de aspectos desconocidos del modelo real. Dichas conclusiones pueden ser muy cercanas a la realidad para los propósitos prácticos.

Cuando aplicamos la matemática a cualquier situación real, es necesario construir un modelo.

Ejemplo La segunda ley de Newton del movimiento establece que un objeto que esté sujeto a una fuerza se mueve con una aceleración proporcional a dicha fuerza ($F = ma$). Ese modelo queda definido por esta ecuación, donde la fuerza es la función de la masa y la aceleración del objeto. Esta ley establece que ese modelo matemático describe con toda precisión el movimiento de un objeto. Dicha ley no es demostrable desde un punto de vista lógico, pero puede comprobarse en forma experimental; más aún, la ecuación básica conduce matemáticamente a varias consecuencias que pueden verificarse; por ejemplo, la distancia de un objeto que cae en un tiempo determinado se deriva de esta ecuación. Esta consecuencia en el modelo entonces se verifica dejando caer objetos y midiendo los tiempos de caída.

Muchos de los modelos de la física clásica y de otras ciencias o disciplinas son determinísticos. Esto significa que, dada una configuración inicial de los elementos que participan, el resultado de un experimento queda completo y precisamente determinado por la ecuación del modelo. Existen muchos fenómenos en los cuales los mecanismos que los originan son tan complejos y las configuraciones iniciales tan difíciles de medir, que cualquier intento para aplicar un modelo determinístico es inútil. Pero se requiere alguna descripción matemática de este fenómeno.

Cuando ocurre el nacimiento de un bebé existen dos posibilidades, niño o niña. La proporción en que ocurren estos eventos se llama *probabilidad*; cuando dicha proporción es constante se suele llamar *regularidad estadística* (fig. 5.1).

En este capítulo se analizarán todos los posibles resultados de un experimento.

4.2 PRINCIPIOS FUNDAMENTALES DEL CONTEO

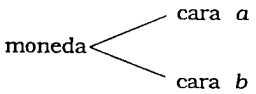
Para conocer todas las posibles ocurrencias de un evento determinado es necesario contar el número de elementos de las combinaciones entre varios conjuntos; una forma es utilizar los diagramas de Venn-Euler, de Carroll o ambos, y otra forma es emplear los diagramas de árbol.

4.2.1 Diagramas de árbol

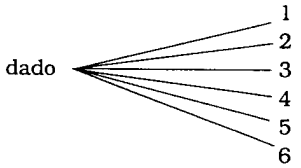
Es un esquema que muestra lógica y explícitamente todos los posibles resultados (eventos) por obtener de un experimento.

Se ejemplificará este método utilizando casos sencillos, pero aumentando el grado de dificultad de los mismos.

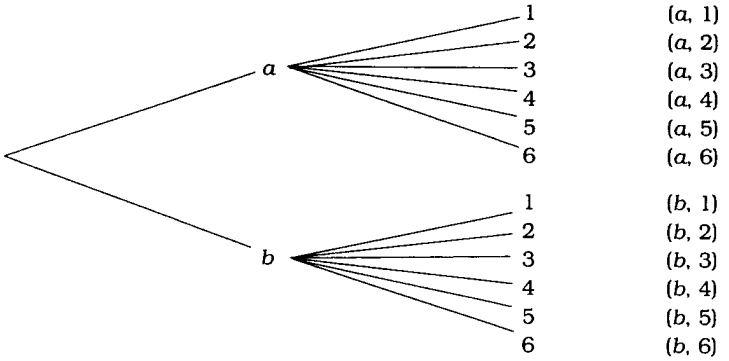
Ejemplo 1 Al lanzar una moneda al aire, cuando cae muestra la cara *a* o la cara *b*:



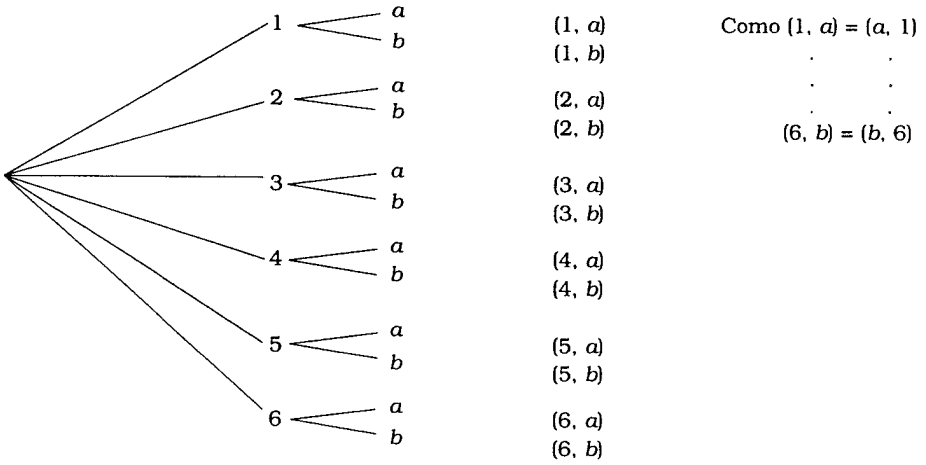
Ejemplo 2 Si se lanza un dado con seis caras numeradas de 1 a 6, los resultados observados son:



Ejemplo 3 Si se hace el experimento de lanzar simultáneamente el dado y la moneda, los resultados que se obtendrán son los siguientes:



o también,



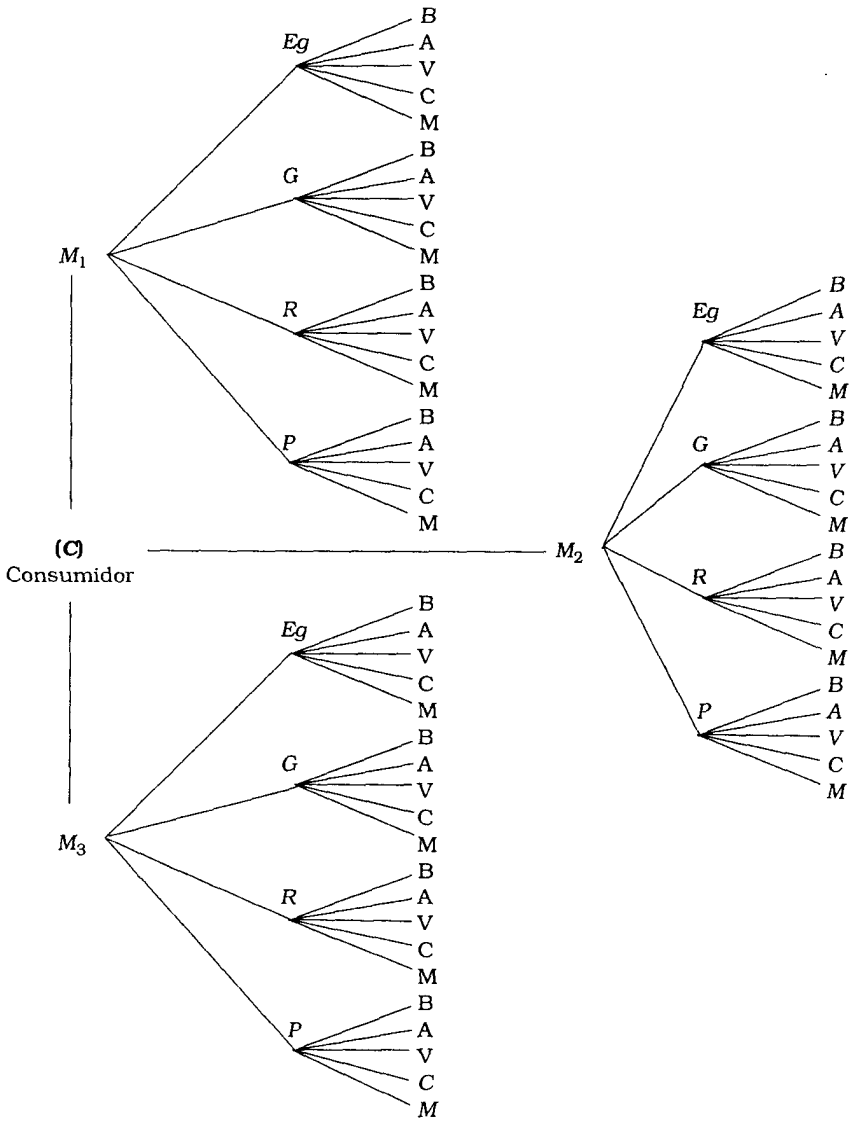
ambos diagramas de árbol son iguales.

Ejemplo En una investigación de mercado vemos que un consumidor se enfrenta a lo siguiente: si desea comprar un refrigerador, visita una tienda que distribuye tres

marcas de refrigeradores, cada marca viene en cuatro tamaños y en cinco colores. ¿Cuántos diferentes refrigeradores existen para escoger según la marca, el tamaño y el color?

Donde:

C	= Consumidor	*B*	= Blanco
M	= 1, 2, 3 (tres marcas diferentes)	*A*	= Amarillo
Eg	= Extra grande	*V*	= Verde
G	= Grande	C	= Cobre
R	= Regular	M	= Madera
P	= Pequeño		



El diagrama de árbol es un medio básico para mostrar un *patrón de conteo.*

En la situación anterior de los refrigeradores, la simetría del patrón de conteo hace más simple utilizar una fórmula en lugar de un diagrama. En otras situaciones, este método es más útil para aclarar que todas las posibilidades han sido contadas una sola vez. El diagrama es particularmente valioso cuando todas las combinaciones, en una o más de las partes de una secuencia, no son compatibles con todas las demás posibilidades en las otras partes. Sin embargo, la dificultad del uso de dicho método aumenta al incrementarse el número de todos los posibles resultados.

Ejemplo En un examen de opción múltiple formado por 10 preguntas, cada una de las cuales tiene cuatro alternativas, habrá $4^{10} = 1\ 048\ 576$ diferentes formas de responder a las preguntas, lo cual sería imposible de representar en un diagrama de árbol.

4.2.2 Principio de la multiplicación

Sea A_1 y A_2 dos conjuntos no vacíos y subconjuntos de U:

$n(A_1)$ = Número de elementos de A_1
$n(A_2)$ = Número de elementos de A_2

entonces:

$$n(A_1 \times A_2) = n(A_1) \times n(A_2)$$

Por lo en general, el siguiente principio se cumple:

$$n(A_1 \times A_2 \times A_3 \times \ldots \times A_{k-1} \times A_k) = n(A_1 \cap A_2 \cap A_3 \cap A_k) = n(A_1) \times n(A_2) \times (A_3) \times \ldots \times n(A_{k-1}) \times n(A_k)$$

Para el caso de los refrigeradores se tiene:

$A_1 = M_1, M_2, M_3$ (las tres marcas de refrigerador).
$A_2 = Eg, G, R, P$ (los cuatro tamaños del refrigerador).
$A_3 = B, A, V, C, M$ (son los cinco colores del refrigerador).

Aplicando el principio se tiene lo siguiente:

$$n(A_1 \times A_2 \times A_3) = 3 \times 4 \times 5 = 60$$

4.2.3 Principio de la adición

CASO 1

Sean A_1 y A_2 dos conjuntos ajenos y no vacíos y subconjuntos de U; entonces:

$$n(A_1 + A_2) = n(A_1) + n(A_2)$$

Por lo general, el siguiente principio se cumple:

$$n(A_1 + A_2 + \ldots + A_{k-1} + A_k) = n(A_1 \cup A_2 \cup A_3 \cup \ldots$$
$$\cup A_{k-1} \cup A_k) = n(A_1) + n(A_2) + n(A_3) + \ldots + n(A_{k-1}) + n(A_k)$$

CASO 2

Pero si los conjuntos no son ajenos, entonces:

$$n(A_1 \cup A_2) = n(A_1) + n(A_2) - n(A_1 \cap A_2)$$

Y para tres conjuntos:

$$n(A_1 \cup A_2 \cup A_3) = n(A_1) + n(A_2) + n(A_3) - n(A_1 \cap A_2)$$
$$- n(A_1 \cup A_3) - n(A_2 \cap A_3) + n(A_1 \cap A_2 \cap A_3)$$

Ejemplo Se lanzan dos dados; todos los resultados posibles se muestran en la matriz siguiente:

(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

A ésta usualmente se le conoce como *espacio muestral* o *espacio de eventos.*

Para facilitar este ejemplo, se considerará que de cada pareja que forma la matriz anterior, el primer elemento corresponde a un dado y el segundo elemento al otro dado.

CASO 1

a) Los resultados que arroja un dado son 6.

$$n(A_1) = 6$$

b) Los resultados que arroja el segundo dado son 6.

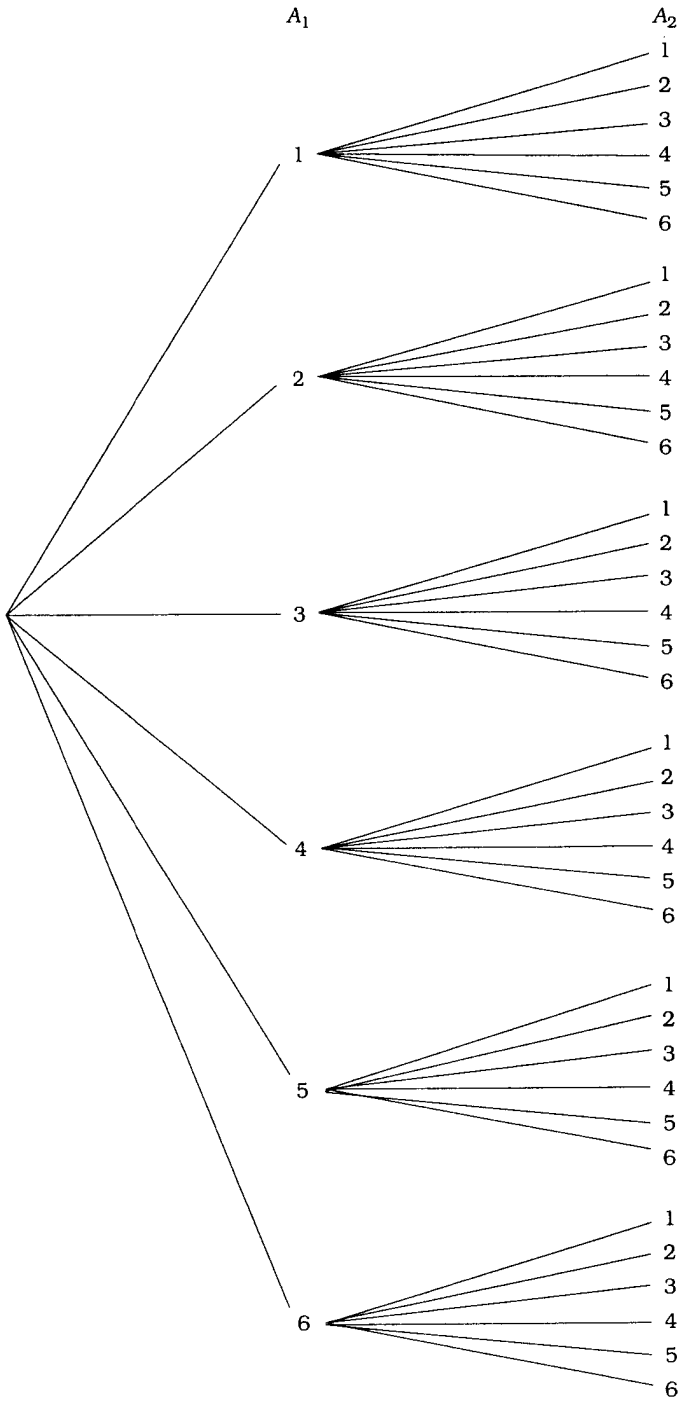
$$n(A_2) = 6$$

CASO 2

a) Si el dado A_1 muestra el número 1 y el número del dado A_2 no importa. Los resultados son:

$$n(A_1) = 6 = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)\}$$

DIAGRAMA DE ÁRBOL AL TIRAR SIMULTÁNEAMENTE DOS DADOS



b) ¿Cuántos resultados arroja la siguiente condición?
A_2 = no importa el número y el dado A_1 que muestre el número uno.

$$n(A_2) = 6 = \{(1,1), (2,1), (3,1), (4,1), (5,1), (6,1)\}$$

c)
$$n(A_1 \cup A_2) = n(A_1) + n(A_2) - n(A_1 \cap A_2)^* = 6 + 6 - 1 = 11$$

sin importar de qué dado proviene.

d) De todos los resultados posibles, excluir los que contiene el número 1, llamando este conjunto de eventos A.

Al observar el espacio muestral y excluyendo los resultados que tienen 1, se obtiene:

$$n(A) = 25$$

4.3 PERMUTACIONES ($P_{n,r}$)

Una permutación es un arreglo de n objetos tomados de tamaño r (el orden de los r elementos es importante).

4.3.1 Cuando $r < n$

$$P_{n,r} = \frac{n!}{(n-r)!}$$

donde:

$n!$ = Se lee como "factorial de un número n, $n > 0$" y n debe ser un entero, $n \in z^{\dagger}$. El factorial de un número son todos los productos sucesivos decrecientes de los factores que lo componen hasta llegar a la unidad:

$$n! = n(n-1)(n-2)(n-3) \ldots [n-(n-1)]$$

Ejemplo: 5!

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \boxed{120}$$

Algunas propiedades de $n!$ son:

i) $0! = 1$

ii) si $a + b = c$, entonces, $(a + b)! = c!$ análogamente para la resta.
si $a - b = d$, entonces, $(a - b)! = d!$

---

* Si $A_1 = 1$ y $A_2 = 1$ entonces $n(A_1, A_2) = n(A_1, A_2)$ y sólo se considera una vez.

† z es el conjunto de los números enteros.

Ejemplo

$$(5 + 3)! = 8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \boxed{40\,320}$$

$$(5 - 2)! = 3! = 3 \cdot 2 \cdot 1 = 6$$

iii) si $\frac{a}{b} = e$, entonces, $\left(\frac{a}{b}\right)! = e!$

Ejemplo

$$\left(\frac{10}{5}\right)! = 2! = 2$$

$$\text{cuando } \frac{a!}{b!},\ a > b$$

Ejemplo

$$\frac{10!}{5!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5\,!}{5\,!} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30\,240$$

4.3.2 Cuando *r* = *n*

$$Pn,r = n!$$

Ejemplo Encontrar la permutación de 5 elementos tomados de tamaño 5.

$$P_{5,5} = 5! = 120$$

Cuando $r < n$ se le suele llamar *ordenación* y se denota por

$${}_nO_r,\quad O_{n,r},\quad O_r^n,\quad {}_nP_r,\quad P_{n,r},\quad P_r^n$$

4.3.3 Circulares

Cuando los objetos se encuentran arreglados en forma circular, el número de permutaciones se obtiene de la manera siguiente:

CASO 1 $n = r$

$$PC_{(n,n)} = (n - 1)!$$

Ejemplo El director de una empresa desea ser democrático y llama a seis de sus colaboradores a una junta. ¿De cuántas formas diferentes se pueden sentar las siete personas *alrededor* de la mesa circular?

$$PC_{(7,7)} = (7-1)! = 6! = 720$$

CASO 2 $n \neq r$

$$PC_{(n,r)} = \frac{n!}{r\,(n-r)\,!}$$

Ejemplo Si se tiene a 10 integrantes en la administración y sólo siete asisten a la junta, ¿de cuántas formas se pueden sentar alrededor de una mesa circular?

$n = 10$
$r = 7$

$$PC(_{10,\,7}) = \frac{10!}{7\,(10-7)\,!} = \frac{10!}{7\,(3\,!)} = \frac{10!}{7 \cdot 6} = \frac{10!}{42}$$

$$PC(_{10,\,7}) = \quad 86\,400$$

4.4 COMBINACIONES

Si dos conjuntos tienen los mismos elementos, sin tomar en cuenta su orden, se consideran iguales. A esta característica se le conoce como una *combinación.*

Cuando el conjunto tiene n elementos, se dice que se pueden formar 2^n subconjuntos, de los cuales $(2^n - 1)$ son no vacíos (capítulo 3). Dichos subconjuntos son todas las combinaciones posibles que se pueden obtener del conjunto de n elementos tomados de un tamaño determinado.

Ejemplo De un grupo de tres personas, ¿cuántos grupos de tamaño tres se pueden formar?, ¿cuántos de tamaño dos?, ¿cuántos de tamaño uno?

Sean las personas (a, b, c).

i) Un grupo de tres personas $(a, b, c) = (a, c, b) = (b, a, c) = (b, c, a) = (c, b, a) = (c, a, b)$

Como no importa el orden de las personas, se obtiene un conjunto de un solo grupo.

ii) Si los subconjuntos por formar son de dos personas, se tienen tres grupos: (a, b), (a, c), (b, c)

como $(a, b) = (b, a)$ se considera sólo una combinación
como $(a, c) = (c, a)$ se considera sólo una combinación
como $(b, c) = (c, b)$ se considera sólo una combinación
Se habla de tres grupos.

iii) Para un subconjunto formado por una sola persona se tienen tres grupos: (*a*), (*b*), (*c*)

Sumando todas las combinaciones anteriores se tienen $2^n - 1 = 2^3 - 1 = 8 - 1 = 7$ subconjuntos o grupos diferentes.

En general, para denotar el número de combinaciones que se pueden obtener de un conjunto de n elementos tomados de r diferentes tamaños, se utiliza la simbología siguiente:

$$C_{(n,r)},\ C_r^n,\ {}_nC_r,\ \binom{n}{r} \text{ donde } r < n$$

Si $n = r$, todas las combinaciones de n elementos de tamaño r es un solo conjunto. Como en el ejemplo del grupo de tres personas tomadas de tamaño tres, se habla de las mismas tres personas que forman el grupo.

Generalizando, tenemos que:

$$C_{(n,r)} = \frac{n!}{r!\,(n-r)!}$$

De la expresión anterior podemos observar que:

$$r!\,[C_{(n,r)}] = \frac{n!}{(n-r)!} = P(_{n,r})$$

Ejemplo En el de las tres personas tenemos:

i) $n = 3,\ r = 3$

$$C_{(3.3)} = \frac{3!}{3!\,(3-3)!} = \frac{3!}{3!\,(0)!} = \frac{3!}{3!} = 1$$

ii) $n = 3,\ r = 2$

$$C_{(3,\,2)} = \frac{3!}{2!\,(3-2)!} = \frac{3!}{2!\,(1)!} = \frac{3 \times 2 \times 1}{2 \times 1} = 3$$

iii) $n = 3,\ r = 1$

$$C_{(3,\,1)} = \frac{3!}{1!\,(3-1)!} = \frac{3!}{2!} = \frac{3 \times 2 \times 1}{2 \times 1} = 3$$

4.5 COEFICIENTE MULTINOMIAL O PARTICIONES

Cuando mencionamos en el inciso anterior los subconjuntos de un conjunto dado implícitamente, estábamos refiriéndonos a *particiones*. Más adelante, en

una población de estudio podremos hacer particiones en términos de la edad, ocupación, ingreso, etc. (capítulo 7).

Ejemplo Indicar cuántas palabras diferentes se pueden formar con el término:

Lisa $4! = 24$ palabras diferentes.

Lola $\frac{4!}{2!} = 12$ palabras diferentes.

A estas 12 palabras diferentes se les conoce como *particiones ordenadas.*

Chihuahua $\frac{9!}{3!\,2!\,2!} = \frac{362\,880}{24} = 15\,120$ palabras diferentes.

En general, tenemos que las particiones ordenadas (PO) se obtienen utilizando

$$(PO) = \frac{n!}{r_{1!}\, r_{2!}\, r_{3!} \cdots r_{k!}}$$

donde las r son las letras que se repiten.

Una de las aplicaciones de los conceptos anteriores, como combinaciones, permutaciones, diagramas de árbol y en general técnicas de conteo, tendrán su aplicación en la probabilidad, herramienta para la construcción de distribuciones probabilísticas básicas en la estadística inferencial.

4.6 TEOREMA DEL BINOMIO

Para desarrollar el teorema del binomio se necesitan dos conceptos ya familiares:

1. Combinaciones, que en este caso serán los coeficientes binomiales:

$$\binom{n}{r}$$

2. El factorial de un número:
   donde $n! = n\,(n-1)!$
   si hacemos $n = 1$ aplicando la fórmula anterior, tenemos que:

$$1! = 1\,(1-1)! = 1 \cdot 0!$$

Despejamos 0! de la expresión anterior:

$$\frac{1!}{1} = 0! \Rightarrow 0! = 1$$

Para el desarrollo binomial se requieren dos elementos (a, b) y un exponente n.

$$(a + b)^0 = 1$$

$$(a + b)^1 = (a + b)$$

$$(a + b)^2 = a^2 + 2ab + b^2$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

$$\vdots \qquad \vdots$$

y así sucesivamente.

Si se consideran los coeficientes numéricos, únicamente se tiene un arreglo conocido como el *triángulo de Pascal:*

1	$(a + b)^0$
1 1	$(a + b)^1$
1 2 1	$(a + b)^2$
1 3 3 1	$(a + b)^3$
1 4 6 4 1	$(a + b)^4$
1 5 10 10 5 1	$(a + b)^5$

En general, las propiedades del desarrollo binomial son:

1. Existen $n + 1$ términos en el desarrollo.
2. La suma de los exponentes de a y b en cualquier término es n; el exponente de a disminuye en una unidad, y el exponente de b aumenta en una unidad de un término al siguiente.
3. El primer término en el desarrollo es:

$$a^n \text{ o lo que es lo mismo } \binom{n}{0} a^n b^0$$

El segundo término es:

$$\frac{n}{1} a^{n-1} b \text{ o en forma semejante: } \binom{n}{1} a^{n-1} b$$

El tercer término es:

$$\frac{n(n-1)}{2 \cdot 1} a^{n-2} b^2 \text{ o de igual modo } \binom{n}{2} a^{n-2} b^2$$

y así sucesivamente.

PROBLEMAS

4.1 Un granjero tiene tres campos, cada uno debe ser cultivado ya sea con maíz o trigo. No siembra el mismo grano en los tres campos.

a) Construya un diagrama de árbol para representar las maneras en que este experimento puede desarrollarse.
b) Enumere los elementos generados por el diagrama de árbol.

4.2 En un partido de tenis de dos jugadores, uno gana tres de cinco sets; el juego se puede extender a tres, cuatro o cinco sets. Describa, por medio de un diagrama de árbol, los posibles eventos del partido y las probables combinaciones de sets en que puede ganar un jugador.

4.3 Un estudiante intenta ordenar su horario de verano; quiere tomar tres clases durante 8:00, 9:00, 10:00 y 11:00 periodos de tiempo. Las clases en que está interesado se ofrecen como sigue:

8:00	9:00	10:00	11:00
Estética	Música	Química	Música
Inglés	Estética	Matemáticas	Inglés
Matemáticas	Biología		

Las condiciones que el estudiante desea satisfacer en su ordenamiento son:

a) No tomar tres clases consecutivas.
b) No tomar matemáticas y ciencias juntas.
c) No tomar dos ciencias.
d) No acomodar música y arte consecutivas.

Organice el horario por medio de un diagrama de árbol.

4.4 En un pequeño restaurante del centro hay un letrero que dice: "Si quiere usted comer con variedad, coma en el restaurante de Pepe. Tenemos más de 200 platillos diferentes todos los días". Al ver el menú, el comensal encuentra la siguiente lista:

A escoger entre:
Sopa o ensalada.
Tres diferentes platillos.
Dos clases de papas.
Tres diferentes tipos de verduras.
Café, té o leche
Postre

¿Cuántos diferentes postres debe ofrecer el restaurante para que el letrero sea verdadero? Entonces, ¿cuál sería el número exacto de diferentes comidas que se tendría ahora?

El último término es:

$$b^n \text{ o lo que es lo mismo } \binom{n}{n} b^n$$

La fórmula general para cualquier n entero positivo es:

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \ldots$$

$$+ \binom{n}{r} a^{n-r} b^r + \binom{n}{n-1} ab^{n-1} + \binom{n}{n} b^n$$

Si se introduce la notación de suma:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$$

Ejemplo Desarrolle:

$$(a+b)^4 = \binom{4}{0} a^4 + \binom{4}{1} a^3 b + \binom{4}{2} a^2 b^2 + \binom{4}{3} a \quad b^3 + \binom{4}{4} b^4$$

RESUMEN

Casi todos los métodos no determinísticos están basados en principios de aritmética elemental y de conocimientos de permutaciones, combinaciones y, en general, de las reglas de conteo.

De acuerdo con este principio, el número total de formas en que dos acciones consecutivas, E_1 y E_2, pueden llevarse a cabo (una acción seguida por la otra), es simplemente el producto del número de formas en que E_1 y E_2 pueden realizarse de manera separada. Formulado como una regla, este principio de conteo se reduce a:

Regla 1. Si la acción E_1 puede llevarse a cabo en n_1 formas, y si la acción E_2 puede realizarse en n_2 formas, entonces su acción conjunta, E_1 seguida por E_2, puede ocurrir en $n_1 \cdot n_2$ formas.

Regla 2. La mayor parte de los cálculos probabilísticos están basados en la suposición de que todas las permutaciones de los resultados son igualmente probables.

4.5 En una clase de apreciación de pintura se muestran 10 colores a cada estudiante, y se les pide que escojan los cinco que más les gusten y hagan una lista de ellos en orden de su preferencia. ¿Cuántas listas diferentes se pueden hacer?

4.6 Un químico desea observar el efecto de la temperatura, de la presión y de la cantidad de catalizador sobre la producción de un compuesto particular por medio de una reacción química. Si en el experimento decide usar dos niveles de temperatura, tres de presión y dos de catalizador, ¿cuántos experimentos deben llevarse a cabo de modo que cada combinación temperatura-presión-catalizador se considere una sola vez?

4.7 Una universidad clasifica a sus estudiantes como hombre (H), mujer (M); de primero, segundo, tercero, cuarto año y por último si aprueban o no un examen final de conocimientos en cada grado:

a) Dibuje un diagrama de árbol.
b) Encuentre el número de clasificaciones posibles.

4.8 Una agencia ofrece tres diferentes tipos de motores en siete modelos y en 14 colores. ¿Cuántos autos distintos debe tener la agencia para satisfacer al cliente?

4.9 En un robo a un banco, un hombre perceptivo dijo que la matrícula del auto en que los ladrones huyeron era un número de seis dígitos, de la cual las dos primeras cifras era 35. Aunque no recuerda el resto de las demás, está seguro de que los dígitos eran todos diferentes. ¿Cuántas matrículas de auto debe revisar la policía?

4.10 Cierta distribuidora de helados ofrece 32 sabores diferentes. ¿Cuántos helados dobles se pueden escoger si

a) el orden de las bolas de helado no es importante?
b) el orden de las bolas de helado es importante?

4.11 Se quiere saber cuántas matrículas de automóvil pueden formarse si cada una consta de tres dígitos y tres letras:

a) Si no se repiten letras ni números.
b) Se pueden repetir letras y números.
c) Considerando la condición en que primero sean dígitos y después letras, y ambos casos.

Se tomarán en cuenta 26 letras y 10 dígitos (de cero a nueve).

4.12 En el popular juego del Melate, en el que se pueden seleccionar seis números de 44, ¿cuántas combinaciones pueden formarse?

4.13 ¿De cuántas formas pueden colocarse 12 cuentas de diferente color para formar un collar?

4.14 Considerando el problema anterior, si en las 12 cuentas existen tres grupos de diferente color con cuatro canicas cada uno, ¿cuántos collares diferentes pueden formarse?

4.15 Si se tienen 6 objetos ($n = 6$), calcule todas las combinaciones posibles, cuando $r = 6$, $r = 5$, $r = 4$, $r = 3$, $r = 2$, $r = 1$, $r = 0$.

4.16 Si en una reunión hay siete mujeres y cinco hombres, ¿cuántas parejas (hombre-mujer) se forman?

4.17 En una asociación de solteros, durante un mes se inscribieron 230 hombres y 400 mujeres. ¿Cuántas parejas pueden formarse?

4.18 Quince niños cumplen las condiciones establecidas por un entrenador de futbol y son aceptados por el grupo; ¿cuántos equipos diferentes de 11 integrantes pueden formarse?

4.19 Un examen consta de 10 preguntas y el estudiante debe escoger seis de ellas. ¿Cuántos posibles exámenes pueden formarse?

4.20 *a)* ¿Cuántos nombres diferentes pueden formarse con Nadia?
b) ¿Cuántos con Alma?
c) ¿Cuántos con Gabriela?
d) ¿Cuántos con Kalimán?
e) ¿Cuántos con la palabra *México*?

4.21 Una pareja tiene r hijos y cada uno de éstos tiene r hijos, etcétera.

a) ¿Cuántos bisnietos tendría?
b) Supongamos que la pareja inicial y todos sus descendientes continúan vivos y casados, excepto los bisnietos. ¿Cuántos lugares se deben considerar para la cena de Navidad, si todos traen a sus propias familias?
c) ¿Cuál es el tamaño de la familia, si $r = 6$?

Parte 3

Modelos no determinísticos

Cálculo probabilístico 5

PROPÓSITOS

Con los contenidos y ejercicios propuestos en este capítulo, el lector será capaz de realizar los aprendizajes específicos para conocer, comprender y generalizar el cálculo probabilístico, y comprenderá los procedimientos para el cálculo de los modelos estocásticos.

Del mismo modo, el lector podrá:

1. Explicar en qué consiste la probabilidad.
2. Identificar las diferentes concepciones que existen en el cálculo probabilístico.
3. Explicar los fundamentos de las escuelas: *a)* objetivista y *b)* subjetivista, con sus características que la definen con el cálculo probabilístico.
4. Explicar la terminología y la simbología empleada.
5. Reconocer las condiciones o axiomas que se trabajan en el cálculo probabilístico.
6. Enunciar las características de las particiones.
7. Identificar las condiciones que hacen a un evento mutuamente excluyente.
8. Explicar las propiedades de la probabilidad condicional.
9. Reconocer las características de los eventos independientes.
10. Identificar las diferencias existentes entre eventos mutuamente excluyentes e independientes.
11. Explicar en qué consiste la probabilidad marginal.
12. Desarrollar los procedimientos propios del teorema de Bayes.
13. Explicar en qué consiste la probabilidad previa o *a priori* y la probabilidad posterior o *a posteriori*.
14. Aplicar los procedimientos desarrollados, propios del cálculo probabilístico, en la solución de problemas.

15. Solucionar diferentes ejercicios de los temas propuestos como continuidad de los capítulos anteriores y posteriores.
16. Manejar los conceptos asociados a los procesos estocásticos, como: estado, estado absorbente, etapa, transición, probabilidad de transición, vector de probabilidad, proceso estocástico, cadena de Markov, cadena de Markov absorbente, etcétera.
17. Desarrollar los estadísticos citados anteriormente a través de los ejercicios correspondientes.
18. Ampliar sus conocimientos en la utilización de diagramas de árbol, de transición o ambos, en el proceso de representación gráfica.
19. Elaborar con habilidad diagramas de transición a partir de las características de un problema o función.
20. Combinar la construcción de diagramas, esto es, de un diagrama de transición pasar a uno de árbol y a partir de él elaborar una matriz de transición.
21. Interpretar los sentidos de una matriz y un esquema de transición y, en general, de los procesos estocásticos.
22. Identificar los grados de complejidad y confusión al graficar diagramas de transición.
23. Generalizar los conocimientos de matrices revisados en el capítulo 3, representando los procesos estocásticos a través de matrices.
24. Aplicar las diferentes características de la matriz de transición estocástica cuadrada.
25. Desarrollar los pasos requeridos para una matriz de transición en cualquier cadena de Markov.
26. Aplicar estos estadísticos en la representación y el abordaje de situaciones reales asociadas al estudio del comportamiento.

INTRODUCCIÓN

En la interpretación del concepto de probabilidad se han seguido dos escuelas. La primera, *clásica* u *objetivista*, interpreta la probabilidad como la frecuencia relativa de la ocurrencia de algún evento o resultado (consecuencia) en el desempeño final o la repetición de un experimento, entendida esta última palabra como una actividad que pudiera conducir a los resultados previstos, sin olvidar que el resultado particular observado en cualquier etapa del experimento también está generado por el azar.

La escuela clásica u objetivista, del siglo XVII, tendría como principal interés determinar la probabilidad de éxito en los juegos de azar. En este caso, los experimentos en cuestión eran tirar un dado, girar la ruleta, manejar las cartas, etc. Los matemáticos de ese siglo, B. Pascal, P. Ferment y P. S. Laplace, entre otros, fueron los que más contribuyeron a formalizar los conceptos de probabilidad.

La segunda escuela es la *subjetivista* o *bayesiana*, la cual utiliza los resultados *a priori* para calcular la probabilidad de ocurrencia de varios estados del

universo (también llamado *espacio de eventos*), en una etapa particular del experimento de que se trata.

El conocimiento *a priori* es la forma en que emergerá la distribución de probabilidades, dado el conocimiento de la naturaleza del experimento que se va a realizar.

El enfoque subjetivista permite al investigador asignar probabilidades o eventos particulares (algunos posibles resultados del experimento que se esté llevando a cabo).

Como mencionamos, la escuela subjetivista también se conoce como *bayesiana* (en este libro, la etiqueta *bayesiana* significa deseo de incorporar probabilidades subjetivas en el análisis estadístico), en contraparte con los seguidores de la escuela objetivista, que no desean incorporar *presentimientos subjetivos* en probabilidad a la estructura formal del análisis estadístico, pues no creen que el concepto de probabilidad pueda aplicarse a la verosimilitud de ocurrencia de varios acontecimientos en un experimento sin repetición.

El uso de las probabilidades *a posteriori* implica la aceptación de la creencia de que ciertas fuerzas que operaron en el pasado causan que un modelo particular de distribución de probabilidad continúe de la misma manera.

Como muchos sistemas matemáticos, la teoría de probabilidad se convierte en un modelo útil cuando sus elementos se relacionan e identifican con cosas particulares; en este caso, los resultados de experimentos reales o conceptuales. De ese modo, la teoría nos permite deducir ciertas proposiciones acerca de la verosimilitud de varios resultados.

Debido a que el observador tiene una ignorancia total o parcial, es decir, una incertidumbre acerca del resultado de una investigación, su actitud será eventualmente inferior a adivinar qué ocurrirá. En un sentido, la probabilidad refleja la incertidumbre acerca del resultado de un experimento, por lo que la probabilidad puede pensarse como el lenguaje matemático de la incertidumbre, y ésta es igual para el investigador y para el jugador de azar.

En los últimos años se ha incrementado el uso de los modelos probabilísticos, debido a que presentan muchas ventajas al aplicarse a situaciones reales. En los modelos probabilísticos interviene un conjunto de variables aleatorias, las cuales se rigen por algún parámetro, como tiempo o espacio. A los procesos aleatorios también se les conoce como *procesos estocásticos*.

En toda investigación o estudio, una vez que se han recopilado datos empíricos se adopta un modelo teórico (alguna distribución teórica), con el objeto de extraer más información de los datos. Si la adaptación es buena, las propiedades del conjunto de datos pueden aproximarse a las del modelo teórico. De manera similar, a un proceso real que posea las características de un proceso aleatorio se le podrá adaptar un proceso estocástico. De este modo, el conocimiento de las propiedades y características del proceso estocástico es altamente deseable para una mejor comprensión del fenómeno real en estudio.

Las incertidumbres asociadas con el comportamiento humano, en diferentes situaciones, se adaptan de una manera adecuada a los modelos probabilísticos y a los procesos estocásticos.

Estos modelos podrían utilizarse para analizar:

- Movilidad social de los individuos.
- Movilidad industrial de trabajadores o empleados.
- Sistemas educativos.
- Procesos de una enfermedad.
- Sistemas de información.

5.1 DEFINICIONES, SÍMBOLOS Y TERMINOLOGÍA

En nuestro lenguaje es común utilizar las expresiones *probablemente, probabilidad, es posible que suceda, es probable que*, etc. Esas frases o palabras se usan para marcar la ocurrencia de un fenómeno, evento o experimento; en esta sección se analizará, en forma sucinta los conceptos matemáticos de probabilidad así como los símbolos y la terminología.

Aún no existe una interpretación única para describir *probabilidad*, es decir, los estadísticos, filósofos y científicos en la materia no han podido homogeneizar el concepto, por lo que, existen tres interpretaciones:

a) *Frecuencia relativa.* Este enfoque es una aproximación empírica del concepto de probabilidad; existe cuando se tiene un número muy grande de observaciones o repeticiones del mismo fenómeno o evento. A continuación, se enunciará una definición un poco más formal.

Si un experimento se ejecuta n *veces en las mismas condiciones y hay* x *resultados,* $x \leq n$ *en que ocurrió un evento (hecho, suceso), entonces una estimación de la probabilidad de ese evento es la razón* x/n, *siempre que* $n \to \infty$.[1]

Algunos autores manejan el concepto de eventos equiprobables para decir que los posibles resultados de un experimento son igualmente probables o en las mismas condiciones.

Ejemplo La probabilidad de obtener cara al tirar una moneda es igual a la razón del número de maneras que el evento pueda ocurrir (uno: una cara) al número total de formas que el resultado del experimento pueda acontecer (dos: cara o cruz), siempre que éste se repita n veces ($n \to \infty$).

Así pues, la probabilidad de que caiga cara es 1/2. Es importante notar que la posibilidad obtenida al usar esta regla se refiere a la frecuencia de ocurrencia. Tal relación se describe en la figura 5.1.

---

[1] Ya-Lun Chou, *Análisis estadístico*, McGraw-Hill, México, 1992, p. 79. En estadística, ∞ es cualquier número mayor que 30.

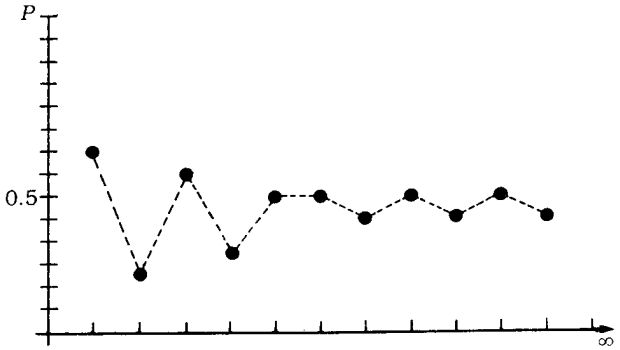


Figura 5.1

b) *Teoría clásica.* La definición clásica de probabilidad fue dada por Laplace y desde entonces se ha repetido en casi todos los libros sobre teoría de probabilidades. En su forma primitiva, dice: *Probabilidad es la razón del número de casos favorables al número total de casos igualmente posibles.*

Algunos autores manejan el concepto de probabilidad *a priori* (antes de); es decir, se determina una vez que se conoce la naturaleza del experimento. A continuación se presentan dos casos de probabilidad clásica:

- La probabilidad de obtener un 3 al tirar un solo dado.

Casos favorables = 1

Todos los casos posibles = 6

$$\text{Teoría clásica} = \frac{\text{Núm. de casos favorables del experimento}}{\text{Núm. de casos posibles del experimento}}$$

Luego entonces, la probabilidad de obtener un 3 al tirar un solo dado es de (1/6).

- La probabilidad de sacar un as de una baraja que consta de 52 cartas (existen 4 ases) es $\left(\dfrac{\text{4 ases}}{\text{total de cartas}}\right) = \dfrac{4}{52}$

c) *Probabilidad subjetiva.* Las probabilidades obtenidas mediante el enfoque clásico o el de frecuencia relativa reciben el nombre de *probabilidades objetivas*, ya que se derivan de hechos. Sin embargo, hay numerosas situaciones en las que no se puede emplear el enfoque objetivo, es decir, situaciones en las que los resultados no son igualmente probables, y no se dispone fácilmente de datos históricos. En este caso, se debe hacer una evaluación *subjetiva* de probabilidad.

Las probabilidades subjetivas son el resultado de un esfuerzo por cuantificar las expectativas o creencias respecto a algo.

En suma, la probabilidad subjetiva es una evaluación personal de la posibilidad de *que* ocurra un evento. (Para una mejor apreciación de las definiciones de probabilidad, véase la figura 5.2.)

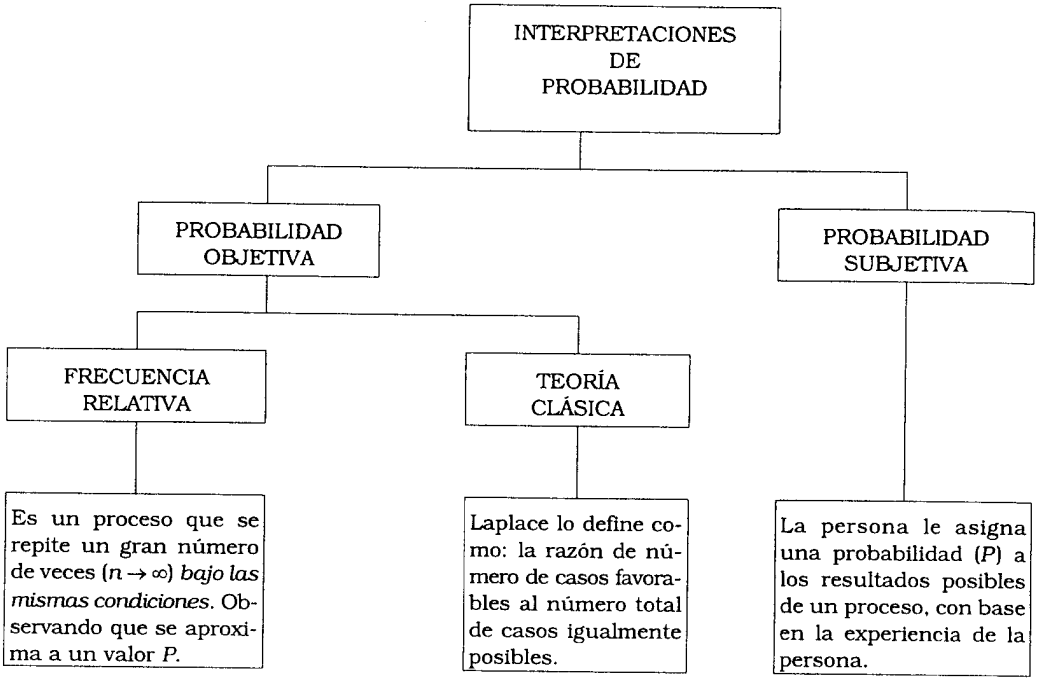


Figura 5.2 Esquema de las definiciones de probabilidad.

Símbolos y terminología

Experimento. Es cualquier actividad bien definida capaz de repetirse en condiciones esencialmente estables; una situación en la que está presente la incertidumbre.

Espacio muestral. Es el conjunto que está formado por todos los posibles resultados de un experimento, denotado por Ω.

Evento o suceso. Es un subconjunto del espacio muestral. Si el subconjunto está formado por un solo resultado del espacio muestral, entonces se denomina *evento simple*. Si el subconjunto está formado por más de un resultado posible, se llama *evento compuesto*; ambos se denotan con una letra mayúscula (*E*, *A*, etcétera).

A continuación se presenta parte de la terminología (respecto a probabilidad) que se usará en los siguientes capítulos.

$P(A)$ = Probabilidad de que ocurra el evento A.

$P(\overline{A}) = P(A') = P(A^c)$ = Probabilidad de que no ocurra el evento A, también llamada *probabilidad del complemento de A*.

$P(A \cap B)$ = Probabilidad de que ocurra el evento A y el evento B, también llamada *probabilidad de la intersección*.

$P(A \cup B)$ = Probabilidad de que ocurra el evento A o el evento B, también llamada *probabilidad de la unión*.

$P(A \mid B)$ = Probabilidad de que ocurra A, puesto que ocurrió, ocurre u ocurrirá B. (*Véase* la secc. 5.4).

5.2 AXIOMAS DE PROBABILIDAD

A todo evento se le asigna una probabilidad, que se da en términos de números reales y debe cumplir con las siguientes condiciones:

i) $P(E) \geq O$, para todo $E \subset \Omega$.
No existen probabilidades negativas.

ii) $P(E \cup F) = P(E) + P(F)$ para todo E y $F \subset \Omega$ y $E \cap F = \phi$
La probabilidad del conjunto universal es igual a uno, o sea la suma de todas y cada una de las probabilidades de que se compone el espacio muestral, $\sum_{i=1}^{n} P(E_i) = 1$.

iii) $P(\Omega) = 1$ para todo Ω.
La probabilidad de ocurrencia de un evento deberá estar siempre en un intervalo [0, 1].

Ejemplo Se lanza una vez un dado honesto.
El espacio muestral es:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Es el conjunto de todos los posibles resultados (Ω).
Sea E_1 = Obtener un 1 en la tirada.
$P(E_1)$ = Probabilidad de obtener un 1 en la tirada.

$$P(E_1) = \frac{1}{6}$$

Como E_1 consta de sólo un elemento del espacio muestral, se le considera un *evento simple*.
Considerando todo el espacio muestral se tiene:

$$\sum_{i=1}^{6} P(E_i) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{6} = 1$$

$P(E_2)$ = Probabilidad de obtener un 2.
Hasta $P(E_6)$ = Probabilidad de obtener un 6.

Luego entonces, el experimento sí cumple con los axiomas de probabilidad.

Sea E_7 = Obtener un 2 o un 5 en la tirada.
$P(E_7)$ = Probabilidad de obtener un 2 o un 5 en la tirada de un dado honesto:

$$P(E_7) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Como E_7 está formado por los elementos del espacio muestral, entonces se le llama *evento compuesto.*

Teoremas: $P(\phi) = 0$ y $P(E^c) = 1 - P(E)$
Ahora se desea calcular la siguiente probabilidad:
$P(\overline{E}_7) = P(E_7^c) = P(E'_7)$ = Probabilidad de no obtener un 2 o un 5 en la tirada de un dado honesto:

$$P(E_7^c) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

Es decir, excluimos los eventos E_2 y E_5

$$P(E_2) = \frac{1}{6}$$

$$P(E_5) = \frac{1}{6}$$

Una forma alterna de resolver es:

$$P(E'_7) = 1 - P(E_7)$$

$$= 1 - \frac{1}{3} = \frac{3}{3} - \frac{1}{3} = \frac{2}{3}$$

$$\Rightarrow P(E'_7) = \frac{2}{3}$$

5.3 PARTICIONES

La cardinalidad de un evento o conjunto $n(E)$ = número de elementos de E; $n(\phi) = 0$

Sea $A_1, A_2, A_3, \ldots, A_n$ subconjuntos del conjunto Ω. La colección de subconjuntos, $A_1, A_2, \ldots, A_n$ forman una partición de Ω si y sólo si

i) $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n = \Omega$
ii) $A_i \cap A_j = \varnothing$ para toda $i \neq j$

Para mayor entendimiento, la definición anterior se analizará con un ejemplo.

Una empresa está constituida por 200 empleados, de los cuales 110 son hombres y 90 mujeres.

Ω = Todos los empleados de la empresa, $n(\Omega) = 200$
A_1 = Todos los hombres que trabajan en la compañía $n(A_1) = 110$
A_2 = Todas las mujeres que trabajan en la compañía $n(A_2) = 90$
Ahora se verificará que en verdad A_1 y A_2 forman una partición de Ω.
i) $A_1 \cup A_2 = \Omega$
{hombres} o {mujeres} = Ω
sí cumple *i)*.
ii) $A_1 \cap A_2 = \varnothing$
$n(A_1 \cup A_2) = n(A_1) + n(A_2) - n(A_1 \cap A_2) = 110 + 90 - 0 = 200$

Como no existe elemento en común, la intersección es igual a vacío.

Luego entonces, los subconjuntos A_1 y A_2 son una partición del conjunto Ω. Es importante hacer notar que si no cumple una de las dos condiciones, deja de ser partición.

Supóngase que en la fiesta de fin de año se desea dar premios a algunos de los empleados. Para ello se tienen dos tómbolas, una para hombres y otra para mujeres.

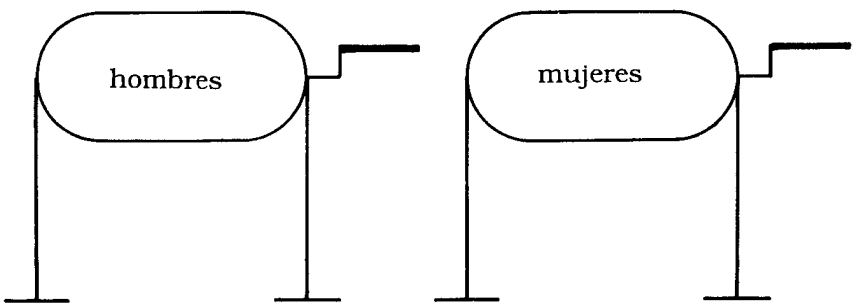


Figura 5.3

Sea
$P(A_1)$ = Probabilidad de que un hombre obtenga el premio

$$P(A_1) = \frac{1}{110}$$

$P(A_2)$ = Probabilidad de que una mujer obtenga el premio

$$P(A_2) = \frac{1}{90}$$

La definición y el ejemplo explicado en la presente sección constituyen la base para entender algunos conceptos que se analizarán más adelante.

Si hacemos una sola rifa, entonces la probabilidad que salga premiado un hombre es $\left(\frac{110}{200}\right)$ y la probabilidad de que salga premiada una mujer es $\left(\frac{90}{200}\right)$

$$\left(\frac{90}{200}\right) + \left(\frac{110}{200}\right) = \frac{200}{200} = 1$$

5.3.1 Eventos mutuamente excluyentes (*me*)

Dos o más eventos se consideran mutuamente excluyentes si la ocurrencia de uno de manera automática elimina la posibilidad de la ocurrencia de los otros.

Ahora se está en condiciones de establecer que todos los subconjuntos de una *partición* son eventos mutuamente excluyentes; en forma gráfica se tiene:

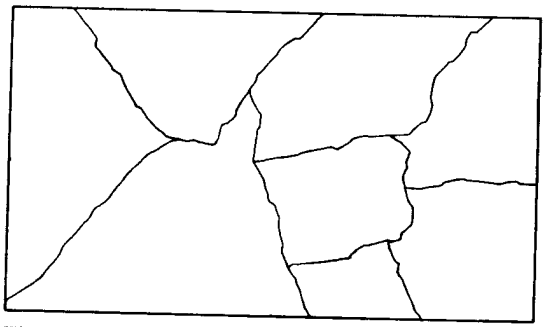


Figura 5.4

En diagramas de Venn-Euler:

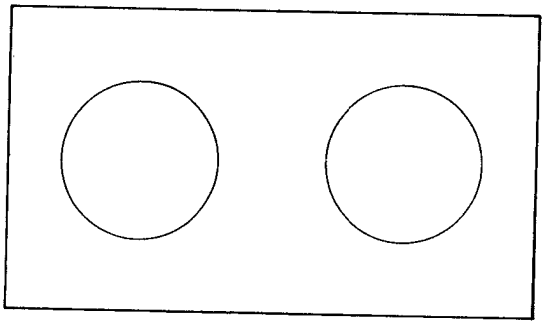


Figura 5.5

Una base útil para entender la probabilidad es la habilidad para usar los conceptos y las relaciones que intervienen en ellas; para la comprensión del concepto de eventos mutuamente excluyentes se utilizará un ejemplo.[2] Se describe el desarrollo académico de los 300 estudiantes que completaron el curso de matemáticas el semestre pasado. Los datos obtenidos se muestran en la tabla 5.1.

[2] Tomado de R. Parsons, *Statistical analysis*, Harper & Row, Nueva York, 1974, p. 112.

Observemos que los 300 estudiantes se clasificaron con base en su calificación obtenida en el curso de matemáticas y también tomando en cuenta si son hombres o mujeres. La tabla contiene 10 celdas o casillas para consignar los datos. Las categorías están definidas en tal forma que cada estudiante es asignado para una y solamente una casilla; por ejemplo, 20 de los estudiantes han sido asignados para la celda asociada con la obtención de MB en matemáticas y formar parte del grupo de hombres. El número total de posibles resultados en el experimento corresponde al número total de elementos en el conjunto universal, el cual se define como el conjunto de todos los posibles resultados (N = 300). Puesto que un conjunto es una colección o grupo de elementos con una característica bien definida, en este caso el espacio muestral consta de aquellos individuos que completaron el curso de matemáticas el semestre pasado.

Tabla 5.1

Calificación en matemáticas	*Hombres (evento 6)*	*Mujeres (evento 7)*	*Total*
(Evento 1) MB	20	5	25
(Evento 2) B	15	50	65
(Evento 3) S	5	90	95
(Evento 4) NA	0	80	80
(Evento 5) NP	0	35	35
Total	40	260	300

El número de subconjuntos puede ser identificado dentro del conjunto universal. Para ilustrar el subconjunto constituido por los individuos que forman el grupo de estudiantes masculinos, son 40 estudiantes en la primera columna. El subconjunto de estudiantes que obtuvieron una S consiste en los 95 que forman parte del tercer renglón; el subconjunto de estudiantes que obtuvieron S y de sexo masculino son los 5 miembros tanto de la primera columna como del tercer renglón, o sea, los miembros de la casilla que es la intersección de la primera columna y el tercer renglón, etcétera.

(E_1) se refiere al evento 1. Éste es definido como *el estudiante seleccionado que obtuvo MB en matemáticas.* $P(E_1)$ se refiere a la probabilidad de ocurrencia del evento 1. También puede ser interpretada como la probabilidad de éxito, donde el éxito se define como el individuo seleccionado o escogido del conjunto universal que también es miembro del subconjunto designado como E_1. Se dice que ocurre el evento uno si cualquiera de los 25 individuos que obtuvieron MB en matemáticas pudieron ser seleccionados. Puesto que cada uno de los 300 individuos en el conjunto universal tienen la misma oportunidad de ser seleccionados, $P(E_1)$ es igual a la razón del número de elementos del subconjunto E_1 al número total de elementos del conjunto universal:

$$P(E_1) = \frac{\text{Núm. de elementos del subconjunto } E_1}{\text{Núm. de elementos del conjunto universal}} = \frac{n(E_i)}{n(\Omega)} = \frac{25}{300}$$

Los eventos mutuamente excluyentes cumplen dos propiedades:

1. $P(A \cup B) = P(A \text{ o } B) = P(A) + P(B)$
   La probabilidad de la unión de dos eventos es igual a la suma de las probabilidades de cada evento.
2. El complemento de un evento A es A^c; entonces $P(A^c) = 1 - P(A)$

$$\Rightarrow P(A) + P(A') = 1.$$

Sea $P(E_1 \cup E_2)$ representado en un diagrama de Venn-Euler.

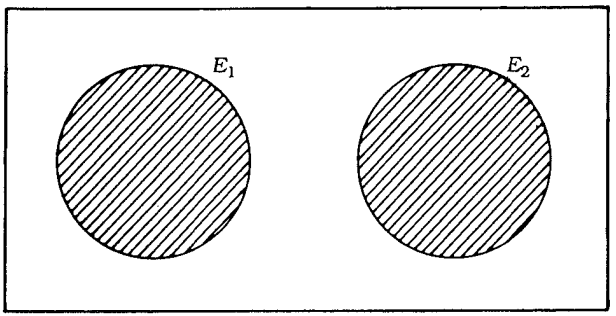


Figura 5.6

$$P(E_1 \cup E_2) = \frac{E_1}{N} + \frac{E_2}{N} = \frac{E_1 + E_2}{N}$$

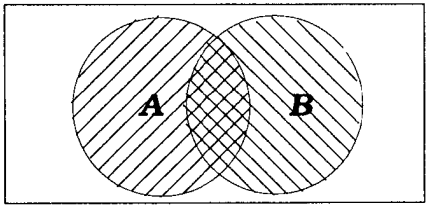
E_1 se refiere al subconjunto de estudiantes que obtuvieron MB en matemáticas. E_2 se refiere al subconjunto de estudiantes que obtuvieron B en matemáticas.

$$P(E_1) = \frac{n(E_1)}{n(N)} = \frac{\text{Núm. de elementos del subconjunto } E_1}{\text{Núm. de elementos del universal}} = \frac{n(E_1)}{n(\Omega)} = \frac{25}{300}$$

$$P(E_2) = \frac{n(E_2)}{n(N)} = \frac{\text{Núm. de elementos del subconjunto } E_2}{\text{Núm. de elementos del universal}} = \frac{n(E_2)}{n(\Omega)} = \frac{65}{300}$$

Reconsiderando, la regla general de adición:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



La regla especial de adición:

$$P(A \cup B) = P(A) + P(B), \text{ si } A \cap B = \phi$$

La función del término $P(A \cap B)$ es para rectificar el doble conteo que se lleva a cabo cuando se suman $P(A)$ y $P(B)$. La regla especial de adición no tiene el término $P(A \cap B)$, debido a que se aplica a eventos mutuamente excluyentes, ya que en este caso no hay intersección y, por tanto, no hay doble conteo.

Si dos eventos son mutuamente excluyentes, la probabilidad de ocurrencia simultánea es igual a cero, es decir:

$$P(A \cap B) = 0 \quad \text{si } A \text{ y } B \text{ son m.e.}$$

Aplicando los conceptos anteriores, así como los axiomas de la probabilidad simple al ejemplo propuesto en la tabla 5.1, donde E_1 se refiere al subconjunto formado por aquellos estudiantes que obtuvieron MB en matemáticas y E_6 a los estudiantes que son hombres. Al presentar esta información en un diagrama de Venn-Euler se tiene (fig. 5.7):

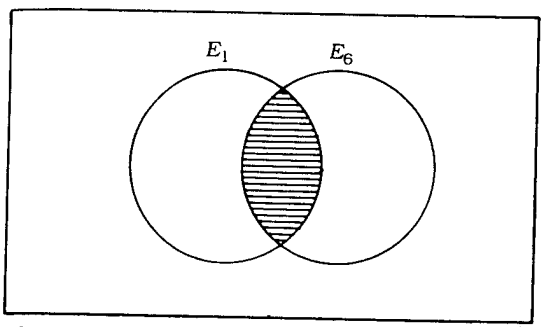


Figura 5.7

El número de MB que pueden ocurrir es igual a la suma de los miembros de los siguientes tres subconjuntos:

subconjunto $(E_6 \cap \overline{E}_1)$ + subconjunto $(\overline{E}_6 \cap E_1)$ + subconjunto $(E_6 \cap E_1)$ = subconjunto $(E_6 \cup E_1)$.

a) $(E_6 \cap \overline{E}_1)$ = Los 20 estudiantes hombres que no obtuvieron MB en matemáticas.
b) $(\overline{E}_6 \cap E_1)$ = Los cinco estudiantes que recibieron MB en matemáticas eran mujeres.
c) $(E_6 \cap E_1)$ = Los 20 estudiantes que eran hombres y también obtuvieron MB en matemáticas.
d) $(E_6 \cup E_1)$ = Los 45 estudiantes que obtuvieron MB o eran hombres.

En el lenguaje de los diagramas de Venn-Euler (fig. 5.8):

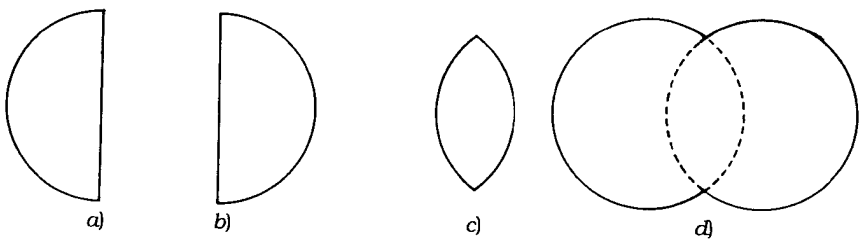


Figura 5.8

$$(E_6 \cap \overline{E}_1) \cup (\overline{E}_6 \cap E_1) \cup (E_6 \cap E_1) = E_6 \cup E_1$$

Una manera (más fácil) para determinar el número de elementos en el subconjunto $E_6 \cup E_1$, es como sigue:

$$n(E_1) + n(E_6) - n(E_6 \cap E_1) = n(E_6 \cup E_1).$$

Esto es:

(Los 25 estudiantes que obtuvieron MB en matemáticas) + (Los 40 estudiantes que eran hombres) − (Los 20 estudiantes que obtuvieron MB en matemáticas y que eran hombres)

= Los 45 que obtuvieron MB, eran hombres. En diagramas de Venn-Euler:

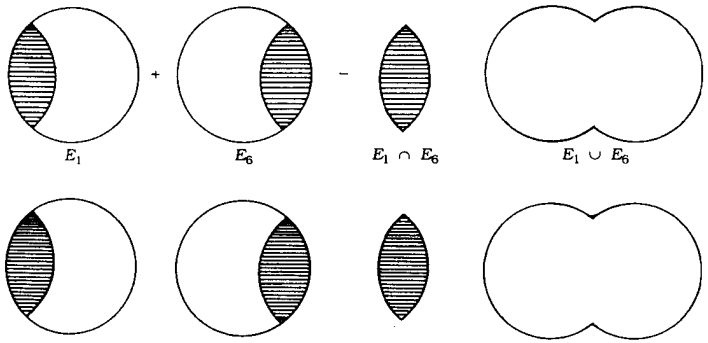


Figura 5.9

$$n(E_1) + (E_6) - n(E_1 \cap E_6) = n(E_6 \cup E_1)$$

Al dividir ambos lados de la expresión entre N, por ejemplo, el número de elementos del conjunto universal Ω, se mantienen las siguientes igualdades:

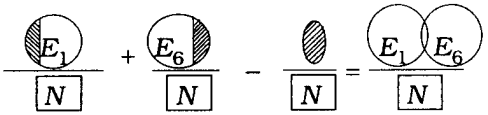


Figura 5.10

donde:

$$P(E_1) = \frac{n\,(E_1)}{n\,\boxed{N}} \qquad P(E_6 \cap E_1) = \frac{n\,(\;)}{n\,\boxed{N}}$$

$$P(E_6) = \frac{n\,(E_6)}{n\,\boxed{N}} \qquad P(E_6 \cup E_1) = \frac{n\,(\;)}{n\,\boxed{N}}$$

La expresión anterior es equivalente a:

$$P(E_1 \cup E_6) = P(E_1) + P(E_6) - P(E_1 \cap E_6) = \frac{25}{300} + \frac{40}{300} - \frac{20}{300} = \frac{45}{300}$$

A partir de la relación anterior se puede derivar una regla general de adición:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

La probabilidad de la unión de dos eventos es igual a la probabilidad marginal de la primera, más la probabilidad marginal de la segunda, menos la probabilidad de su intersección.

La función del término $P(A \cap B)$ rectifica el doble conteo que se lleva a cabo cuando se suman $P(A)$ y $P(B)$. La regla general de adición puede extenderse a la probabilidad de la unión de tres eventos. La regla general de adición para tres eventos es: $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

Ejemplo El conjunto A consta de cuatro regiones etiquetadas 1, 2, 3 y 4 en el diagrama de Venn.

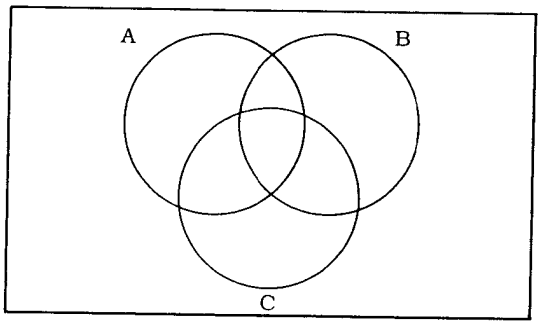


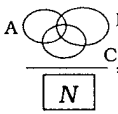
Figura 5.11

El conjunto B está formado por cuatro regiones etiquetadas, 2, 3, 6 y 7, y el C consiste en cuatro regiones etiquetadas 3, 4, 5 y 6.

Observemos los efectos de combinaciones de subconjuntos referidos en la regla general de adición:

Conjunto	A	=	1	+ 2	+ 3	+ 4			
Más conjunto	B	=		2	+ 3			+ 6	+ 7
Más conjunto	C	=			3	+ 4	+ 5	+ 6	
Menos conjunto	$(A \cap B)$	=		- 2	- 3				
Menos conjunto	$(A \cap C)$	=			- 3	- 4			
Menos conjunto	$(B \cap C)$	=			- 3			- 6	
Más conjunto	$(A \cap B \cap C)$	=			+ 3				
	$(A \cup B \cup C)$	=	1	+ 2	+ 3	+ 4	+ 5	+ 6	+ 7

Así $(A \cup B \cup C) = A + B + C - (A \cap B) - (A \cap C) - (B \cap C) + (A \cap B \cap C)$. Por consiguiente:



$$= \frac{\text{Núm. de elementos de los nuevos subconjuntos formados por la unión de los tres subconjuntos}}{\text{Núm. total de elementos del conjunto universal } (\Omega).} = \frac{n(A \cup B \cup C)}{n(\Omega)}$$

$$= P(A \cup B \cup C)$$

5.4 PROBABILIDAD CONDICIONAL

Teorema. La probabilidad de que ocurra un evento A, dado que ocurrió el evento B, se llama *probabilidad condicional* de A dado B y se expresa como sigue:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) \neq 0$$

Análogamente, la probabilidad condicional de B dado A se define como:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \qquad P(A) \neq 0$$

La probabilidad condicional, como su nombre lo indica, está condicionada a un evento que ya sucedió o sucederá. En la vida diaria se dice: "*es posible que llueva ya que está nublado*", y la frase anterior está compuesta de dos eventos:

A = La probabilidad de que llueva.

B = El día está nublado.

Si regresamos al ejemplo de la sección anterior tenemos lo siguiente: $P(E_1 \mid E_6)$ es una probabilidad condicional y se identifica por una raya vertical que separa E_1 y E_6. Éste es el símbolo usado para representar la expresión *dado que*. Es decir, $P(E_1 \mid E_6)$ se refiere a la probabilidad que el evento E_1 ocurra, dado que, condicionado por información, puede ser que el evento E_6 ya haya ocurrido. La diferencia básica entre probabilidad condicional y las que examinamos previamente es que, en el caso de las probabilidades condicionales, las selecciones se realizan dentro del subconjunto indicado a la derecha de la raya, en tanto que para $P(E_1)$, $P(E_1 \cup E_6)$, y $P(E_1 \cap E_6)$, las selecciones se llevan a cabo a partir del conjunto universal.

$$\begin{aligned} P(E_1) &= P(E_1 \mid \Omega) \\ P(E_1 \cup E_6) &= P(E_1 \cup E_6 \mid \Omega) \\ P(E_1 \cap E_6) &= P(E_1 \cap E_6 \mid \Omega) \end{aligned}$$

Ocasionalmente $P(E_1)$ se escribe como $P(E_1 \mid \Omega$ universo), lo cual se lee *dado que la extracción es hecha fuera del espacio muestral*, que es la probabilidad de que el elemento seleccionado sea también un miembro del subconjunto E_1. Por lo regular se escribe $P(E_1)$.

A menos de que se afirme otra cosa, se supone que la selección se hace a partir del *espacio muestral*. El formato condicional por lo común se reserva para esas ocasiones donde es necesario prestar atención al hecho de que las extrac-

ciones se llevan a cabo a partir de un subconjunto del *espacio muestral* en lugar del mismo *espacio muestral.*

$P(E_1 \mid E_6)$ se refiere a la probabilidad que un individuo seleccionado al azar del subconjunto E_6 sea también un miembro del subconjunto E_1. E_6, en el ejemplo, consiste en el grupo de los hombres. El evento se define como que el individuo también es un miembro del subconjunto E_1 (el grupo que obtuvo MB).

$$P(E_1 \mid E_6) = \frac{\text{Núm. de elementos del subconjunto } E_6 \text{ que son también miembros del subconjunto } E_1}{\text{Núm. de elementos del subconjunto } E_6} = \frac{n(E_1 \cap E_6)}{n(E_6)} = \frac{20}{40}$$

De manera semejante:

$$P(E_6 \mid E_1) = \frac{\text{Núm. de elementos del subconjunto } E_1 \text{ que son también miembros del subconjunto } E_6}{\text{Núm. de elementos del subconjunto } E_1} = \frac{n(E_1 \cap E_6)}{n(E_1)} = \frac{20}{25}$$

Reconsiderada la definición de probabilidad condicional, se estableció que $P(E_1 \mid E_6)$ se refiere a la probabilidad de que un individuo sea seleccionado del subconjunto E_6 y sea también miembro del subconjunto E_1.

Otra manera de establecer esto es la siguiente: "de los elementos que son miembros del subconjunto E_6, ¿qué porcentaje son también miembros del subconjunto E_1?"

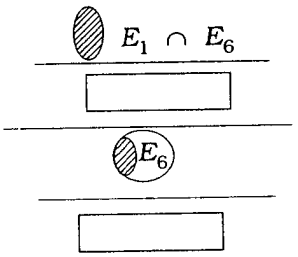


Figura 5.12

$$P(E_1 \mid E_6) = \frac{n(E_1 \cap E_6)}{E_6}^{3} = \frac{\text{Núm. de elementos de } E_1 \cap E_6}{\text{Núm. de elementos de } E_6} = \frac{n(E_1 \cap E_6)}{n(E_6)}$$

[3] El símbolo E_6 se interpreta como la representación de todo el subconjunto E_6.

Un cociente no se afecta cuando numerador y denominador se dividen entre el mismo valor positivo. Dividiendo ambos, numerador y denominador, entre el número de elementos del espacio muestral, se obtiene:

$$P(E_1 \mid E_6) = \frac{\dfrac{n(E)}{n(\Omega)}}{\dfrac{n(E_6)}{n(\Omega)}} = \frac{\dfrac{\text{Núm. de elementos del subconjunto } E_1 \cap E_6}{\text{Núm. de elementos del espacio muestral } (\Omega)}}{\dfrac{\text{Núm. de elementos del subconjunto } E_6}{\text{Núm. de elementos del espacio muestral } (\Omega)}} = P(E_1 \mid E_6 = \frac{P(E_1 \cap E_6}{P(E_6)} = \frac{\dfrac{n(E_1 \cap E_6)}{n(\Omega)}}{\dfrac{n(E_6)}{n(\Omega)}}$$

Por definición:

$$P(E_1 \cap E_6) = \frac{n}{n\boxed{\Omega}} \qquad P(E_6) = \frac{n}{n\boxed{\Omega}}$$

La región sombreada es simplemente para destacar la región del subconjunto E_6 asociada a la intersección. Así pues,

$$P(E_1 \mid E_6) = \frac{P(E_1 (\cap E_6)}{P(E_6)} = \frac{\dfrac{n(E_1 (\cap E_6)}{1}}{\dfrac{n(E_6)}{1}} \times \frac{\dfrac{1}{n(\Omega)}}{\dfrac{1}{n(\Omega)}} = \frac{\dfrac{n(E_1 \cap E_6)}{n(\Omega)}}{\dfrac{n(E_6)}{n(\Omega)}} = \frac{P(E_1 \cap E_6)}{P(E_1)}$$

Multiplicando ambos lados $P(E_1 \mid E_6) = \dfrac{P(E_1 \cap E_6)}{P(E_6)}$ por $P(E_6)$, se generan las siguientes igualdades:

$$P(E_6) \times P(E_1 \mid E_6) = \frac{P(E_1 \cap E_6)}{P(E_6)} \times P(E_6)$$

Cancelando, se obtiene:

$$P(E_6) \times P(E_1 \mid E_6) = P(E_1 \cap E_6)$$

Entonces,

$$P(E_1 \cap E_6) = P(E_6) \times P(E_1 \mid E_6)$$

Un razonamiento análogo se aplica a $P(E_6 \mid E_1)$ y se obtiene la siguiente relación:

$$P(E_6 \mid E_1) = \frac{n}{n(E_1)} = \frac{\dfrac{n}{n(\Omega)}}{\dfrac{n(E_1)}{n(\Omega)}} = \frac{P(E_6 \cap E_1)}{P(E_1)}$$

5.4.1 Eventos independientes

Lema. Dos eventos, A y B, definidos en el mismo espacio muestral, se dice que son independientes si y sólo si la probabilidad de la ocurrencia conjunta ($P(A \cap B)$) de A y B es igual al producto de sus respectivas probabilidades individuales:

$$P(A \cap B) = P(A)\, P(B)$$

Ahora, si se introduce el concepto de probabilidad condicional, queda como sigue:

i) $P(A \mid B) = P(A)$ si y sólo si $P(A \cap B) = P(A)\, P(B)$, siempre que $P(B) \neq 0$.

$$\left(P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\, P(B)}{P(B)}\right)$$

ii) $P(B \mid A) = P(B)$ si y sólo si $P(A \cap B) = P(A)\, P(B)$ siempre que $P(A) \neq 0$.

Como vimos en la sección anterior, $P(A \mid B)$ significa la probabilidad de que ocurra el evento A dado que ocurra el evento B. Aplicamos el teorema:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\, P(B)}{P(B)} = P(A)$$

Esta proporción nos brinda más información que el solo conocimiento de $P(A)$, llamada *probabilidad simple* o *no condicionada.* Cuando la ocurrencia de un evento A no depende de la ocurrencia de otro evento B, entonces los eventos A y B son independientes, es decir, la ocurrencia de un evento no depende de la ocurrencia del otro evento.

El lema puede generalizarse para k eventos independientes, de la forma siguiente:

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots \cap A_k) = P(A_1)\, P(A_2)\, P(A_3) \ldots P(A_k)$$

Ejemplo Supóngase que se tiene una urna que contiene dos bolas, una roja y una blanca. El experimento consiste en tomar una bola, anotar el color y devolverla a la urna. El procedimiento se repite dos veces. Para cada selección la probabilidad de tomar una bola roja es igual a 1/2.

El espacio muestral está formado por el conjunto

$$\Omega = \{RR, RB, BR, BB\}$$

La primera letra representa la primera extracción (Bola *Roja*) y la segunda letra representa la segunda extracción (Bola *Blanca*), es decir $\{RB\}$.

Sea Z_1 = Sacar una bola roja en la primera extracción:

$$Z_1 = \{RR, RB\}$$

la probabilidad de Z_1:

$$P(Z_1) = \frac{2}{4} = \frac{1}{2}$$

Sea Z_2 = Sacar una bola roja en la segunda extracción:

$$Z_2 = \{RR, BR\}$$

la probabilidad de Z_2:

$$P(Z_2) = \frac{2}{4} = \frac{1}{2}$$

Ahora se desea conocer $Z_1 \cap Z_2$:

$$Z_1 \cap Z_2 = \{RR\}$$

la probabilidad es:

$$P(Z_1 \cap Z_2) = \frac{1}{4}$$

Por otra parte, se tiene que:

$$P(Z_1)\,P(Z_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P(Z_1 \cap Z_2)$$

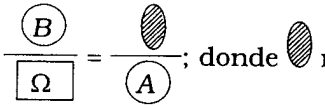
Como los eventos Z_1 y Z_2 cumplen con la igualdad propuesta en el teorema, entonces se dice que los eventos son independientes.

¿Eventos mutuamente excluyentes o independientes?

¿Pueden dos eventos ser mutuamente excluyentes e independientes al mismo tiempo?

Si los eventos A y B son independientes, se cumple $P(B) = P(B|A)$. Esto significa que si el número de elementos del conjunto B es 10% del número de elementos del conjunto universal; entonces, el número de elementos del subconjunto $(A \cap B)$ debe ser también igual a 10% del número de elementos del subconjunto A. Esto quiere decir que $(A \cap B)$ debe contener algunos elementos.

$P(B) = P(B|A)$ implica que:



; donde representa $A \cap B$

Por otro lado, si los eventos A y B son mutuamente excluyentes, no pueden ocurrir asociadamente y el subconjunto $(A \cap B)$ debe ser un conjunto nulo (vacío). Es evidente que ambas situaciones no pueden ocurrir al mismo tiempo; por consiguiente, *dos eventos no pueden ser mutuamente excluyentes e independientes al mismo tiempo.*[4]

Otra incongruencia es que cuando los eventos A y B son independientes $P(B \mid A) = P(B)$, de cualquier manera; entonces, los eventos A y B son mutuamente excluyentes: $P(B \mid A) = 0$, $P(B \cap A) = 0$.

Otro modo de percibir la relación es pensar en términos de una tabla que describa los tipos de relación que pueden existir entre dos eventos (tabla 5.2).

La ocurrencia o no ocurrencia de un evento de ninguna forma afecta la probabilidad condicional de ocurrencia del otro evento, cuando son independientes:

$$P(B) = P(B \mid A) = P(B \mid \overline{A})$$

Considérese un espacio muestral que consta de 100 elementos parecidos y dos subconjuntos: el subconjunto A con 20 elementos y el B con 40 elementos.

1. Si A y B son independientes

$$\left.\begin{array}{r}\text{Aunque sea } E \text{ el evento ser hombre} \\ \text{y sea } F \text{ el evento ser mujer}\end{array}\right\} \Rightarrow \begin{array}{c} E \cap F = \varnothing \\ \text{y} \\ E \text{ y } F \text{ son independientes}\end{array}$$

$$P(A) = P(A \mid B) = P(A \mid \overline{B}) = \frac{20}{100} = \frac{8}{40} = \frac{12}{60}$$

TABLA 5.2 Tipos de relaciones que pueden existir entre dos eventos.

	Dependiente	
Mutuamente excluyente	*No mutuamente excluyente*	*Independiente*
La ocurrencia de un evento en forma automática causa que la probabilidad condicional del segundo evento sea cero: $P(B \mid A) = 0$	La ocurrencia o no ocurrencia de un evento afecta la probabilidad condicional del otro evento que ocurre, aunque no incluye la otra ocurrencia de los otros eventos; por ejemplo: no causa que la probabilidad condicional del otro evento sea cero: $P(B) \neq P(B \mid A) \neq P(B \mid \overline{A}) \neq 0$	La ocurrencia o no ocurrencia de un evento no afecta la probabilidad condicional de ocurrencia del otro. $P(B) = P(B \mid A) = P(B \mid \overline{A})$

[4] Sólo cuando $P(B) = 0$.

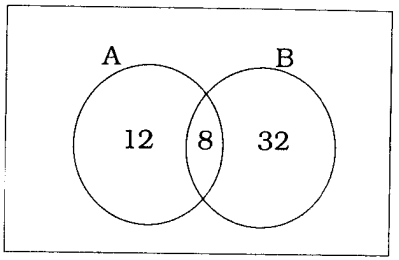


Figura 5.13

2. Si A y B son mutuamente excluyentes

$$P(A) = \frac{20}{100} \neq P(A \mid B) = \frac{0}{40} = 0$$

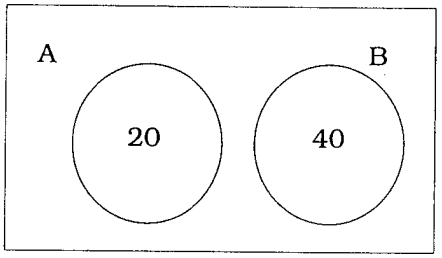


Figura 5.14

3. Si A y B son eventos dependientes pero no mutuamente excluyentes

$$P(A) = \frac{20}{100} \neq P(A \mid B),$$

puede haber una intersección y el número de elementos contenidos en $A \cap B$ es cualquier número entero no negativo hasta 20, excepto 0 (lo cual significaría que no hay intersección), y que los eventos son mutuamente excluyentes y ocho (lo cual significaría que los eventos son independientes). Claramente, el número de elementos en la intersección nunca excede del número de elementos en el más pequeño de los dos subconjuntos, en este caso 20.

En suma, se presentan los siguientes casos:

1. Considérense dos eventos A y B cuyas probabilidades $P(A)$ y $P(B) \neq 0$, A y B son mutuamente excluyentes:
   Por definición $A \cap B = \varnothing$
   $P(A \cap B) = 0$
   Para que A y B sean independientes deben cumplir $P(A) \times P(B) = P(A \cap B)$

pero como $P(A) \neq 0$ y $P(B) \neq 0$

$$P(A) \times P(B) \neq 0 \Rightarrow P(A \cap B) \neq P(A) \times P(B)$$

Luego entonces, A y B no pueden ser independientes, si son mutuamente excluyentes.

2. Sean A y B dos eventos independientes; por definición se tiene:

$$P(A) \times P(B) = P(A \cap B)$$

ya que $P(A) \neq 0$, $P(B) \neq 0$ y $P(A \cap B) \neq 0$
$A \cap B \neq \varnothing$, por tanto, A y B no son mutuamente excluyentes.

5.4.2 Probabilidad marginal

"En muchas circunstancias es conveniente considerar que un suceso ocurra siempre conjuntamente con otros sucesos. Por ejemplo, puede ser útil identificar no sólo los artículos defectuosos en un proceso de producción, sino también especificar qué máquinas (o qué operarios) han originado esos defectos. A las compañías de seguros no sólo les interesa la magnitud del daño asociado con cada accidente automovilístico, sino también, entre otras cosas, la ciudad donde ocurrió el accidente y la edad y el sexo del conductor. De manera análoga, los estudiantes de una universidad son identificados no sólo por su sexo, sino también por el año que cursan y por su especialidad. Una persona que solicita un préstamo puede ser clasificada no sólo por la suma de dinero solicitada, sino también por los años de permanencia en su empleo, por sus ingresos personales, por el monto de sus deudas, etcétera".[5]

5.5 TEOREMA DE LA PROBABILIDAD TOTAL

La fórmula que se enunciará a continuación recibe varios nombres: *principio de expansión*, *teorema de la probabilidad total* y *probabilidad marginal*, entre otros.

Sea $H_1 \cup H_2 \cup H_3 \cup \ldots \cup H_n = \Omega$, donde $H_1, H_2, H_3, \ldots, H_n$ son sucesos exhaustivos y mutuamente excluyentes, y el espacio muestral Ω. Sea E un evento arbitrario originado por Ω. Entonces:

$$E = (E \cap H_1) \cup (E \cap H_2) \cup \ldots \cup (E \cap H_n) \text{ y } (E \cap H_i) \cap (E \cap H_j)$$

---

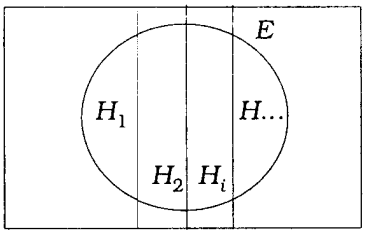
[5] Tomado de D. Harnett, J. Murphy, D. *Introducción al análisis estadístico*, Addison Wesley, México, 1987, p. 86.

$$P(E) = P[(E \cap H_1) \cup (E \cap H_2) \cup \ldots \cup (E \cap H_n)]$$

$$P(E) = P = (E \cap H_1) + P(E \cap H_2) + \ldots + P(E \cap H_n)$$

$$P(E) = \sum_{i=1}^{n} P(E \cap H_i) = \sum_{i=1}^{n} P(E \mid H_i)\, P(H_i)$$

$$\boxed{P(E) = \sum_{i=1}^{n} P(E \mid H_i)\, P(H_i)}$$



Ejemplo Supongamos que tenemos dos máquinas, I y II, que fabrican zapatos. Sea H_1 el suceso de fabricación de zapatos de la máquina I, y H_2 el suceso por el que los zapatos se fabrican en la máquina II. Sea A el suceso que representa un zapato sin defecto; entonces:

$$A = AH_1 \cup AH_2$$

puesto que AH_1 y AH_2 son mutuamente excluyentes, se tiene:

$$P(A) = P(AH_1) + P(AH_2) = P(A \mid H_1) \times P(H_1) + P(A \mid H_2)\, P(H_2)$$

Si la máquina I fabrica el 60% de los zapatos, entonces:

$$P(H_1) = 0.60\% \qquad P(H_2) = 0.40\%$$

Además, si el 10% de los zapatos que hace la máquina I tienen defecto, y el 20% de los zapatos hechos por la máquina II tienen defecto, resulta:

$$P(A \mid H_1) = 0.90^{6} \qquad P(A \mid H_2) = 0.80^{7}$$

---

[6] Probabilidad de no defectuoso ya que se fabricó en la máquina I.
[7] Probabilidad de no defectuoso ya que se fabricó en la máquina II.

Es decir, la probabilidad de fabricar un zapato sin defecto es: $P(A) = (0.90)(0.60) + (0.80)(0.40) = 0.86$:

$$\boxed{P(A) = 0.86}$$

$$P(A) = (0.9)(0.6) + (0.8)(0.4) = 0.54 + 0.32 = 0.86$$

Esquematizando las fórmulas anteriores se tiene:

Regla general de adición

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Regla especial de adición (aplicable sólo a los eventos mutuamente excluyentes)

$$P(A \cup B) = P(A) + P(B)$$

Regla general de multiplicación:

$$P(A \cap B) = P(A) \times P(B \mid A) \text{ o } P(B) \times P(A \mid B)$$

Regla especial de multiplicación (aplicable sólo a eventos independientes)

$$P(A \cap B) = P(A) \times P(B)$$

Antes de considerar casos de aplicación específica es conveniente idear un experimento simple para observar cómo se calculan las *probabilidades a posteriori.*

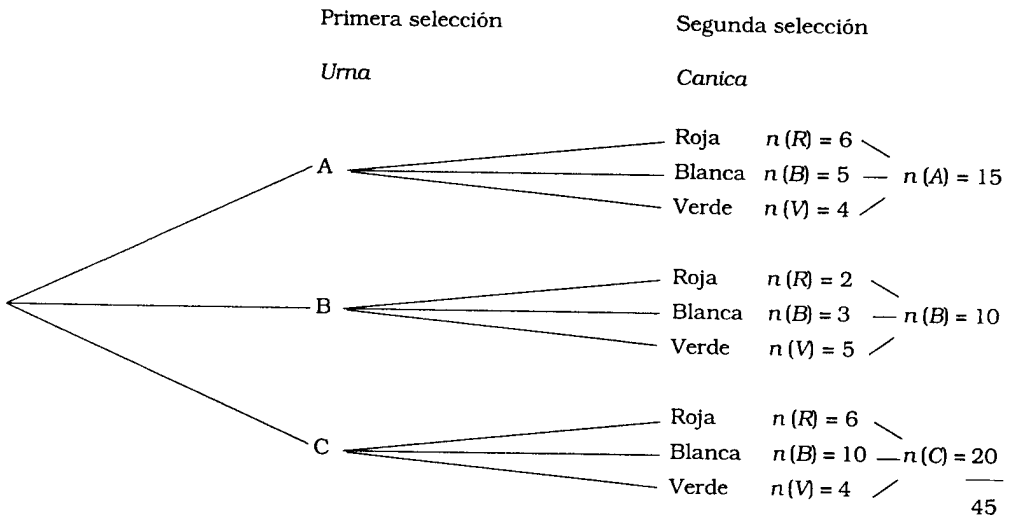
Considérese que tiene tres urnas:[8] *A*, *B*, *C*; la urna *A* contiene seis canicas rojas, cinco blancas y cuatro verdes. La urna *B* contiene dos canicas rojas, tres blancas y cinco verdes, y la *C* contiene seis canicas rojas, 10 blancas y cuatro verdes.

En un experimento común se le vendan los ojos a una persona y se le pide seleccionar una urna (el sujeto no sabe qué urna selecciona); extrae una canica y se considera su color. Dado el color de la canica; ¿cuál es la probabilidad de que la persona haya seleccionado la urna *A*, *B* o *C*? En un experimento de simple selección al azar de una urna, la probabilidad de seleccionar cualquiera de las tres es 1/3. Existe una forma para hacer una selección específica de tres igualmente posibles. ¿De qué manera afectará esta probabilidad la consideración del color de la canica extraída?

---

[8] Tomado R. W. Negus, *Fundamental of finite Mathematics*, John Wiley & Sons, Nueva York, 1974.

Para iniciar el análisis de este proceso se hará un diagrama de árbol:



Los resultados en el experimento constan de pares de eventos: el primer elemento del par es la urna seleccionada y el segundo elemento el color de la canica seleccionada.

Ejemplo (A, R_2) y (C, V_3) representan la selección de la urna A y la segunda canica roja de dicha urna, y la siguiente representa la urna C y la tercera canica verde, etcétera.

Obsérvese que hay un total de 45 canicas en las tres urnas, por lo que existen 45 resultados posibles que no son igualmente probables debido al número diferente de canicas en cada una de las urnas; o sea, no se puede obtener la probabilidad de extraer una canica dada al pensar que hay un total de 14 canicas rojas y formar la proporción de 14/45, sino que se deben aplicar las fórmulas para eventos mutuamente excluyentes, que consisten en la selección de una urna y una canica. De este modo, la probabilidad de seleccionar una canica roja $P(R)$ está dada por:

$$P(R) = P(A \cap R) + P(B \cap R) + P(C \cap R)$$

donde:

$P(A \cap R)$ = Probabilidad de seleccionar la urna A y extraer una canica roja.
$P(B \cap R)$ = Probabilidad de seleccionar la urna B y extraer una canica roja.
$P(C \cap R)$ = Probabilidad de seleccionar la urna C y extraer una canica roja.

Con base en la probabilidad condicional para dos eventos E_1 y E_2 (la urna y la canica):

$$P(E_2 \mid E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)}$$

donde:

$$P(E_2 \cap E_1) = P(E_1) \times P(E_2 \mid E_1)$$

Por tanto, para nuestro caso particular:

$$P(A \cap R) = P(A) \times P(R \mid A)$$

$$P(B \cap R) = P(B) \times P(R \mid B)$$

$$P(C \cap R) = P(C) \times P(R \mid C)$$

Puesto que la urna A contiene 15 canicas de las cuales seis son rojas, la urna B contiene 10 canicas de las cuales dos son rojas y la urna C contiene 20 canicas de las cuales seis son rojas, ocurre lo siguiente:

$$P(R \mid A) = \frac{6}{15} = \frac{2}{5}$$

$$P(R \mid B) = \frac{2}{10} = \frac{1}{5}$$

$$P(R \mid C) = \frac{6}{20} = \frac{3}{10}$$

La probabilidad de escoger cualquiera de las tres urnas es la misma:

$$P(A) = P(B) = P(C) = \frac{1}{3}$$

Por consiguiente:

$$P(A \cap R) = P(A) = P(R \mid A) = \frac{1}{3} \cdot \frac{2}{5} = \frac{2}{15}$$

$$P(B \cap R) = P(B) = P(R \mid B) = \frac{1}{3} \cdot \frac{1}{5} = \frac{1}{15}$$

$$P(C \cap R) = P(C) = P(R \mid C) = \frac{1}{3} \cdot \frac{3}{10} = \frac{3}{30} = \frac{1}{10}$$

Por tanto:

$$P(R) = P(A) = P(R \mid A) = + P(B) = P(R \mid B) + P(C) = P(R \mid C) = \frac{2}{15} + \frac{1}{15} + \frac{1}{10} = \frac{9}{30} = \frac{3}{10}$$

La pregunta que se formuló al principio de este ejemplo era que después de observar que la canica extraída es roja, ¿cuál es la probabilidad de que provenga de la urna A?

$$P(A \mid R) = \frac{P(A \cap R)}{P(R)} = \frac{P(A)\,P(R \mid A)}{P(A)\,P(R \mid A) + P(B)\,P(R \mid B) + P(C)\,P(R \mid C)}$$

$$P(A \mid R) = \frac{\frac{2}{15}}{\frac{3}{10}} = \frac{4}{9}$$

Esto significa que, después de haber visto que la canica seleccionada es roja, la probabilidad de que la urna A fuera la urna seleccionada ya no es de 1/3, sino que se ha incrementado a 4/9 y, entonces, para las demás urnas (B y C) se tiene:

$$P(B \mid R) = \frac{P(B \cap R)}{P(R)} = \frac{\frac{1}{15}}{\frac{3}{10}} = \frac{2}{9}$$

$$P(C \mid R) = \frac{P(C \cap R)}{P(R)} = \frac{\frac{1}{10}}{\frac{3}{10}} = \frac{1}{3}$$

5.6 TEOREMA DE BAYES

El experimento de las urnas es un modelo de proceso aleatorio de dos etapas, ya que en nuestro análisis hemos hecho algunas inferencias acerca de lo que pasó en la primera etapa del proceso. Cuando esto se desconoce, basado en el resultado final del experimento, el cual también es desconocido, la relación crucial se denomina *teorema de Bayes*, que es una generalización de la probabilidad condicional:

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

donde:
E_1 = Cualquiera de los eventos de la primera etapa.
E_2 = El evento observado en la segunda etapa.
$P(E_2)$ será calculada utilizando el teorema de probabilidad total.

$$P(E_2) = \sum_{i=1}^{n} P(E \mid H_i)\, P(H_i)$$

Otra aplicación en las ciencias del comportamiento consiste en el contraste de hipótesis. Los investigadores frecuentemente proponen diferentes explicaciones de un fenómeno dado, cada uno de los cuales predice distintas probabilidades para el resultado de un experimento particular. Se asigna una probabilidad a cada hipótesis, esto es, a la verosimilitud relativa de cada hipótesis

(explicación del fenómeno) se le asigna una medida numérica y después se lleva a cabo el experimento y se observa el resultado. En general, el **teorema de Bayes** está dado por:

$$P(H_j \mid E) = \frac{P(H_j \cap E)}{P(E)} = \frac{P(E \cap H_j)}{P(E)} = \frac{P(E \mid H_j)\, P(H_j)}{\sum_{i=1}^{n} P(E \cap H_j)\, P(H_i)} = \frac{P(H_j)\, P(E \mid H_j)}{\sum_{i=1}^{n} P(H_i)\, P(E \cap H_i)}$$

Retomando el ejemplo anterior, en el que dos máquinas fabrican zapatos, supóngase que una persona está interesada en la siguiente cuestión: ¿Cuál es la probabilidad de que un zapato sin defecto se fabrique en la máquina I? Si A era el suceso de que un zapato no tenga defecto y H_1 era el suceso de que el zapato fuera fabricado por la máquina I, nuestra cuestión en símbolos es

$$P(H_1 \mid A)$$

Es decir, dado un zapato sin defecto, ¿cuál es la probabilidad de que se haya fabricado en la máquina I?

Con la fórmula de probabilidad condicional, la probabilidad $P(H_1 \mid A)$ es:

$$P(H_1 \mid A) = \frac{P(H_1 \cap A)}{P(A)}$$

y $P(H_1 \cap A) = P(A)\, P(H_1 \mid A)$ Regla general de multiplicación; y también

$$P(H_1 \cap A) = P(A \cap H_1)$$

Pero con el teorema de probabilidad total, la $P(A)$ resulta:

$$P(A) = P(A \mid H_1)\, P(H_1) + P(A \mid H_2)\, P(H_2)$$

se tiene:

$$P(A) = \sum_{i=1}^{n} P(A \mid H_i)\, P(H_i)$$

Sustituyendo la expresión anterior en la definición de probabilidad condicional se tiene:

$$P(H_1 \mid A) = \frac{P(H_1 \cap A)}{\sum_{i=1}^{n} P(A \mid H_i)\, P(H_i)} = \frac{P(A \mid H_1)\, P(H_1)}{\sum_{i=1}^{n} P(A \mid H_i)\, P(H_i)} \qquad \text{(I)}$$

Obsérvese que $P(H_1)$ es la probabilidad de que un zapato sea fabricado por la máquina I, mientras que $P(H_1 \mid A)$ es la probabilidad de que un zapato fabri-

cado por la máquina I no tenga defecto. A la probabilidad $P(H_1)$ se le llama *probabilidad previa* (*a priori*) y a $P(H_1 \mid A)$ se le llama *probabilidad posterior* (*a posteriori*).

Ejemplo 1 Considérese la situación de un estudio (retrospectivo), realizado por un psicólogo acerca de la satisfacción en el trabajo. En una empresa, en la cual los empleados se ubican en alguna de las siguientes tres áreas: gerencial, administrativa y ventas se obtuvo la siguiente información: del 100% de entrevistados, 20% pertenecen al área gerencial, 30% a la administrativa y 50% a la de ventas.

El cuestionario que se les aplica se puede también resumir en tres áreas, que son: satisfacción en el trabajo, indiferencia en el trabajo y no satisfacción en el trabajo. Los porcentajes obtenidos en dicha encuesta se concentraron en la tabla 5.3.

TABLA 5.3

	Gerencial	*Administrativa*	*Ventas*	
	20%	30%	50%	100%
Satisfecho en el trabajo	50%	10%	30%	
Indiferente	30%	40%	50%	
No satisfecho en el trabajo	20%	50%	20%	
	100%	100%	100%	

Es necesario destacar la independencia que existe entre el área a la cual pertenece un empleado (gerencia, administrativa, ventas) y su actitud hacia el trabajo (satisfecho, indiferente y no satisfecho). Ambos eventos son independientes, pero no mutuamente excluyentes.

Para simplificar el desarrollo consideraremos la siguiente notación:

G = Gerencia.
A = Administrativo.
V = Ventas.
S = Satisfecho.
I = Indiferente.
$\overline{S}$ = No satisfecho.

Si escogemos un empleado al azar obtenemos:

$$P(G) = 0.20$$

$$P(A) = 0.30$$

$$P(A) = \frac{0.50}{100}$$

Observemos la información contenida en la tabla anterior, por ejemplo:

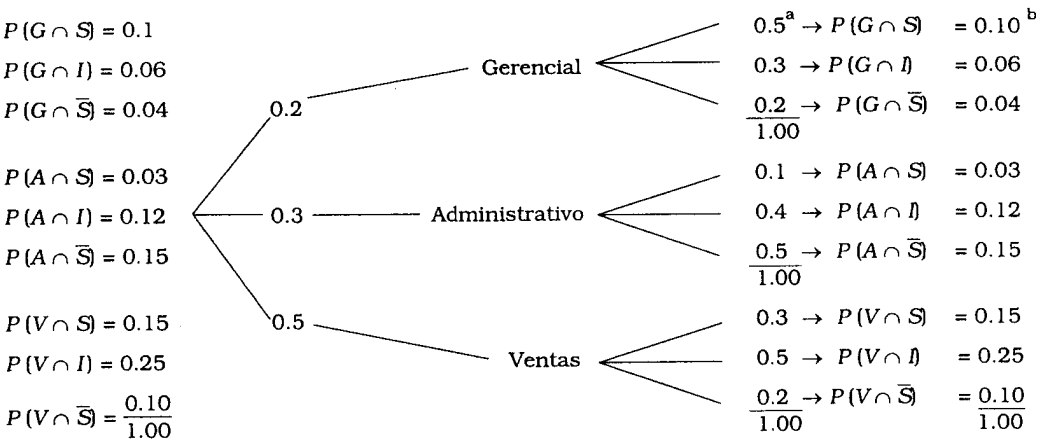
En el primer renglón de satisfecho con el trabajo, el 50% del 20% que corresponde a los gerentes están satisfechos con el trabajo, así como el 10% del 30% que forman los administrativos, y el 30% del 50% del área de ventas; igual para el caso de las personas indiferentes y no satisfechos con el trabajo.

Como dijimos anteriormente, los eventos formar parte de la gerencia y estar satisfecho con el trabajo son independientes; entonces, se cumple lo siguiente:

$$P(G \cap S) = P(G) \times P(S) = (0.20 \times 0.50) = 0.10$$

y así para todos los casos.

DIAGRAMA DE ÁRBOL



Y si se suman $P(G \cap S) + P(A \cap S) + P(V \cap S) = P(S)$, siendo ésta la probabilidad del evento S; así como también para el caso $P(I)$ y $P(\bar{S})$.

El estudio consiste en seleccionar al azar un empleado de la empresa y preguntarle cuál es su actitud respecto a su trabajo; con base en su respuesta, determinar la probabilidad de que se encuentre en alguno de los tres niveles (gerencial, administrativo o ventas). Si se desea conocer si un empleado escogido al azar está satisfecho con su trabajo $P(S)$, estará dado por la suma de las probabilidades *joint* o mancomunadas, descritas anteriormente:

$$P(S) = P(G \cap S) + P(A \cap S) + P(V \cap S) \qquad \ldots \qquad 1$$

---

[a] Estas probabilidades son condicionales $P(S \mid G) = 0.5$.

[b] A estas probabilidades se les suele conocer como *probabilidades joint*.

y consecuentemente:

$$P(I) = P(G \cap I) + P(A \cap I) + P(V \cap I) \quad \ldots \quad 2$$

$$P(\bar{S}) = P(G \cap \bar{S}) + P(A \cap \bar{S}) + P(V \cap \bar{S}) \quad \ldots \quad 3$$

A continuación se desarrollará $P(S)$ que es la probabilidad marginal.

Considerando la regla general de multiplicación, que es consecuencia inmediata del teorema de Bayes, tenemos:

$$P(S \mid G) = \frac{P(G \cap S)}{P(G)} \text{ donde } P(G)\, P(S \mid G) = P(G \cap S)$$

Para cada una de las expresiones que forman parte de $P(S)$ obtenemos lo siguiente, y sustituyendo:

$$P(G \cap S) = P(G) \times P(S \mid G) = (0.2)(0.5) = 0.10$$

$$P(A \cap S) = P(A) \times P(S \mid A) = (0.3)(0.10) = 0.03$$

$$P(V \cap S) = P(V) \times P(S \mid V) = (0.5)(0.3) = 0.15$$

Al sustituir los valores anteriores en la expresión 1 se tiene $\boxed{P(S) = 0.28}$

Si elegimos al azar a un empleado, la probabilidad de que esté satisfecho con su trabajo es del 28%, y para indiferentes y no satisfechos tenemos:

$$P(I) = P(G \cap I) + P(A \cap I) + P(V \cap I)$$

$$P(I) = 0.06 + 0.12 + 0.25$$

$\boxed{P(I) = 0.43}$ Indiferente con el trabajo 43%

$$P(\bar{S}) = P(G \cap \bar{S}) + P(A \cap \bar{S}) + P(V \cap \bar{S})$$

$$P(\bar{S}) = 0.04 + 0.15 + 0.10$$

$\boxed{P(\bar{S}) = 0.29}$ No satisfecho con el trabajo 29%

100% Considerando a todos los empleados.

Es interesante conocer la probabilidad de que un empleado satisfecho con su trabajo provenga de la gerencia $P(G \mid S)$ o que provenga del área administrativa $P(A \mid S)$ o de ventas $P(V \mid S)$, así como $P(S \mid G)$ que significa, un empleado que esté satisfecho, dado que es gerente, y todos los casos posibles. Esto se obtiene utilizando el teorema de Bayes:

$$P(G \mid S) = \frac{P(G \cap S)}{P(S)} = \frac{0.10}{0.28} = 0.35714, \text{ y}$$

$$P(S \mid G) = \frac{P(G \cap S)}{P(G)} = \frac{0.10}{0.20} = 0.5$$

donde:

$$P(G \cap S) = P(S \cap G) \text{ y } P(G \mid S) \neq P(S \mid G)$$

Al calcular todos los casos posibles con los resultados anteriores, se obtiene lo siguiente:

$$P(G \mid S) = \frac{P(G \cap S)}{P(S)} = \frac{0.10}{0.28} = 0.35714$$

$$P(A \mid S) = \frac{P(A \cap S)}{P(S)} = \frac{0.03}{0.28} = 0.10714$$

$$P(V \mid S) = \frac{P(V \cap S)}{P(S)} = \frac{0.15}{0.28} = 0.53571 \qquad 1$$

$$P(G \mid I) = \frac{P(G \cap I)}{P(I)} = \frac{0.06}{0.43} = 0.13953$$

$$P(A \mid I) = \frac{P(A \cap I)}{P(I)} = \frac{0.12}{0.43} = 0.27907$$

$$P(V \mid I) = \frac{P(V \cap I)}{P(I)} = \frac{0.25}{0.43} = 0.58140 \qquad 1$$

$$P(G \mid \bar{S}) = \frac{P(G \cap \bar{S})}{P(\bar{S})} = \frac{0.04}{0.29} = 0.13793$$

$$P(A \mid \bar{S}) = \frac{P(A \cap \bar{S})}{P(\bar{S})} = \frac{0.15}{0.29} = 0.51724$$

$$P(V \mid \bar{S}) = \frac{P(V \cap \bar{S})}{P(\bar{S})} = \frac{0.10}{0.29} = 0.34483 \qquad 1$$

Las siguientes probabilidades $P(S \mid A)$, $P(S \mid V)$, $P(I \mid G)$..., son las probabilidades ([b]) que se mostraron en el diagrama de árbol anterior.

Ejemplo 2 Un investigador estudia el proceso de aprendizaje y la manera que un conjunto de experiencias influye en la proporción en que determinada actividad se domine. Dicho investigador tiene tres ideas diferentes acerca de cómo una persona utiliza la experiencia; éstas son tres hipótesis alternativas. Se diseña un experimento en el cual una persona aprende a dominar una tarea fácil simplemente practicando una parte de dicha actividad. Sus diferentes

hipótesis hacen predicciones acerca de cuántas veces una persona debe practicar antes de que domine la actividad o tarea. Dichas hipótesis son:

La probabilidad de que $P(H_1)$ ocurra es 40%, $P(H_2) = 30\%$ y $P(H_3) = 30\%$. (Estos valores se obtienen por estudios previos.)

$$P(H_1) = 0.4$$

$$P(H_2) = 0.3$$

$$P(H_3) = 0.3$$

Sea A igual a dos o menos prácticas, B igual a tres prácticas y C más de tres:

	H_1	H_2	H_3
Probabilidad	40%	30%	30%
A	30%	50%	20%
B	50%	40%	30%
C	20%	10%	50%

Si H_1 y A ocurren simultáneamente, su probabilidad es (0.4)(0.3) = 0.12 = 12%, o sea (40%)(30%) = 12%

$$P(H_1 \cap A) = 0.12$$

y sucesivamente:

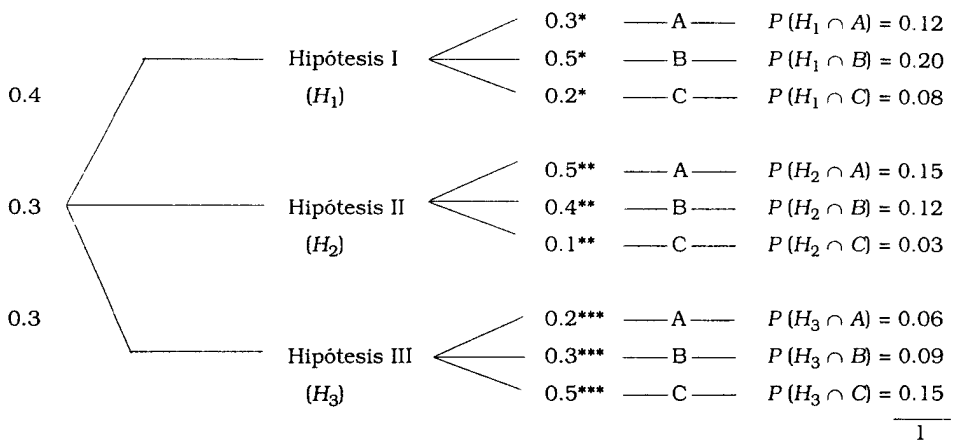
$$P(H_1 \cap B) = 0.20 \qquad P(H_2 \cap C) = 0.03$$

$$P(H_1 \cap C) = 0.08 \qquad P(H_3 \cap A) = 0.06$$

$$P(H_2 \cap A) = 0.15 \qquad P(H_3 \cap B) = 0.09$$

$$P(H_2 \cap B) = 0.12 \qquad P(H_3 \cap C) = 0.15$$

Considerando las probabilidades anteriores en el *diagrama de árbol* se tiene:



Se ubican en una tabla de doble entrada:

	H_1	H_2	H_3	TOTAL
A	0.12	0.15	0.06	0.33
B	0.20	0.12	0.09	0.41
C	0.08	0.03	0.15	0.26
TOTAL	0.40	0.30	0.30	1

Y todos los eventos posibles son:

$$P(A) = 0.12 + 0.15 + 0.06 = 0.33$$

$$P(A) = 0.33$$

$$P(B) = 0.20 + 0.12 + 0.09 = 0.41, \quad P(B) = 0.41$$

$$P(C) = 0.08 + 0.03 + 0.15 = 0.26, \quad P(C) = 0.26$$

$$P(H_1) = 0.12 + 0.20 + 0.08 = 0.40, \quad P(H_1) = 0.40$$

$$P(H_2) = 0.15 + 0.12 + 0.03 = 0.30, \quad P(H_2) = 0.30$$

$$P(H_3) = 0.06 + 0.09 + 0.15 = 0.30, \quad P(H_3) = 0.30$$

$$P(H_1 \mid A) = \frac{P(H_1 \cap A)}{P(A)} = \frac{P(A \cap H_1)}{P(A)} = \frac{0.12}{0.33} = 0.36364$$

$$P(A \mid H_1) = \frac{P(A \cap H_1)}{P(H_1)} = \frac{P(H_1 \cap A)}{P(H_1)} = \frac{0.12}{0.40} = 0.30^*$$

$$P(H_1 \mid B) = \frac{P(H_1 \cap B)}{P(B)} = \frac{0.20}{0.41} = 0.4878$$

$$P(B \mid H_1) = \frac{P(H_1 \cap B)}{P(H_1)} = \frac{0.20}{0.40} = 0.50^*$$

$$P(H_1 \mid C) = \frac{P(H_1 \cap C)}{P(C)} = \frac{0.08}{0.26} = 0.30769$$

$$P(C \mid H_1) = \frac{P(H_1 \cap C)}{P(H_1)} = \frac{0.08}{0.40} = 0.20^*$$

$$P(H_2 \mid A) = \frac{P(H_2 \cap A)}{P(A)} = \frac{0.15}{0.33} = 0.45455$$

$$P(A \mid H_2) = \frac{P(H_2 \cap A)}{P(H_2)} = \frac{0.15}{0.30} = 0.50^{**}$$

$$P(H_2 \mid B) = \frac{P(H_2 \cap B)}{P(B)} = \frac{0.12}{0.41} = 0.29268$$

$$P(B \mid H_2) = \frac{P(B \cap H_2)}{P(H_2)} = \frac{0.12}{0.30} = 0.40^{**}$$

$$P(H_2 \mid C) = \frac{P(H_2 \cap C)}{P(C)} = \frac{0.03}{0.26} = 0.11538$$

$$P(C \mid H_2) = \frac{P(H_2 \cap C)}{P(H_2)} = \frac{0.03}{0.30} = 0.10^{**}$$

$$P(H_3 \mid A) = \frac{P(H_3 \cap A)}{P(A)} = \frac{0.06}{0.33} = 0.18182$$

$$P(A \mid H_3) = \frac{P(H_3 \cap A)}{P(H_3)} = \frac{0.06}{0.30} = 0.20^{***}$$

$$P(H_3 \mid B) = \frac{P(H_3 \cap B)}{P(B)} = \frac{0.09}{0.41} = 0.21951$$

$$P(B \mid H_3) = \frac{P(H_3 \cap B)}{P(H_3)} = \frac{0.09}{0.30} = 0.30^{***}$$

$$P(H_3 \mid C) = \frac{P(H_3 \cap C)}{P(C)} = \frac{0.15}{0.26} = 0.57692$$

$$P(C \mid H_3) = \frac{P(H_3 \cap C)}{P(H_3)} = \frac{0.15}{0.30} = 0.50^{***}$$

Recomendamos al lector interpretar estos resultados.

Algunos autores aplican directamente el teorema general de Bayes (*I*) para resolver los problemas; por ejemplo, para el caso anterior:

$$P(H_1 \mid A) = \frac{P(H_1)\,P(A \mid H_1)}{P(H_1)\,P(A \mid H_1) + P(H_2)\,P(A \mid H_2) + P(H_3)\,P(A \mid H_3)}$$

$$= \frac{P(H_1)\,P(A \mid H_1)}{\sum_{i}^{3} P(H_i)\,P(A \mid H_i)}$$

Sustituyendo los valores obtenidos anteriormente:

$$P(H_1 \mid A) = \frac{0.40 \times 0.30}{0.40 \times 0.30 + 0.30 \times 0.50 + 0.30 \times 0.20}$$

$$P(H_1 \mid A) = \frac{0.12}{0.12 + 0.15 + 0.06} = \frac{0.12}{0.33}$$

$$P(H_1 \mid A) = 0.36364$$

Es el resultado obtenido anteriormente.

5.7 PROCESOS ESTOCÁSTICOS

5.7.1 Conceptos y definiciones

Estado. Cada una de las situaciones o maneras mutuamente excluyentes, en las cuales se puede encontrar un sistema.

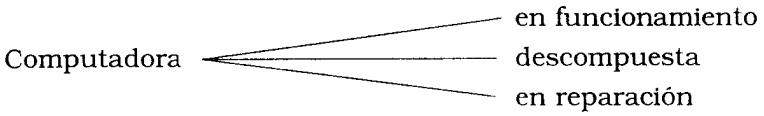
Ejemplos ***1.*** Una computadora puede estar en alguno de los siguientes estados: *en funcionamiento, descompuesta* o *en reparación.*

2. La probabilidad de cada resultado (estado) depende exclusivamente del ensayo (etapa) anterior.
3. La probabilidad de pasar de un resultado a otro permanecerá constante en el transcurso de todos los ensayos (etapas).

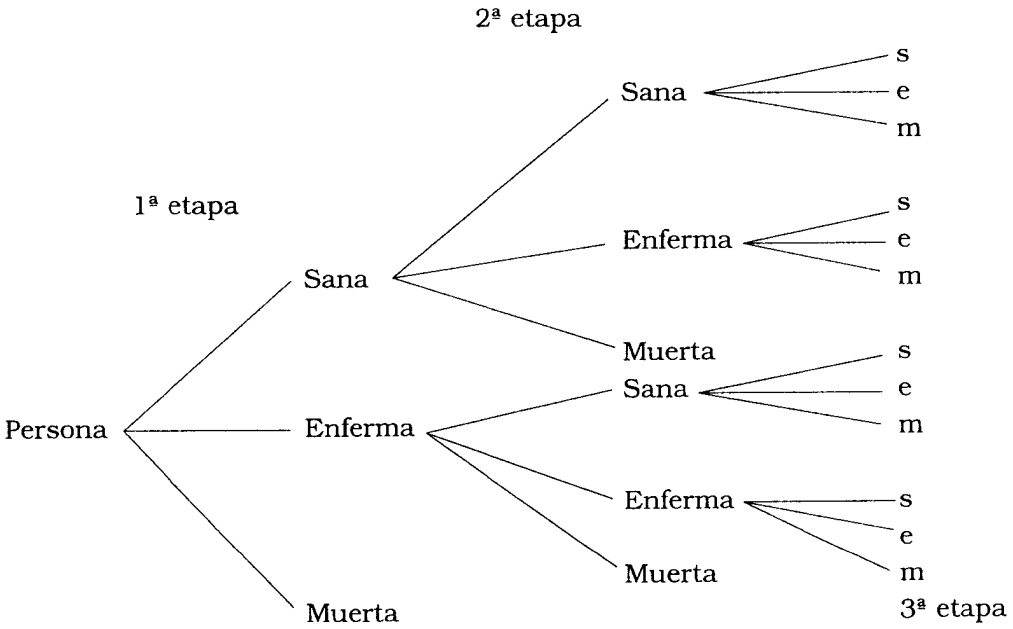
Para una definición más formal de cadena de Markov, *véase* E. Parzen, (1972) y L. Breiman (1969).

5.7.3 Representación gráfica

Para representar gráficamente los posibles cambios de los estados y sus etapas en un proceso estocástico, se pueden utilizar tanto los diagramas de árbol como los de transición. Así, el diagrama de árbol que corresponde al ejemplo de la computadora es el siguiente:



En el ejemplo de la persona el diagrama es el siguiente:



En la medida que se requieran mostrar más etapas, la construcción e interpretación del diagrama de árbol resultan menos prácticas, ya que aumentan el grado de complejidad y de confusión.

2. Una persona puede encontrarse en alguna de estas situaciones: *enferma*, *sana* o *muerta*.

Estado absorbente. Situación a la cual ha llegado un sistema y de la cual ya no puede pasar a otro estado. Además, un estado absorbente sólo puede tener entradas, pero no salidas. En el ejemplo anterior, cuando la persona llega al estado de muerte, ya no puede pasar a ningún otro. Por tanto, la muerte es un estado absorbente.

Etapa. Periodo que dura el desarrollo de una acción o proceso (establecido previamente por el investigador). Al término de cada etapa se observa el sistema, para determinar el estado en el cual se encuentra. El periodo puede abarcar una unidad convencional (segundo, día, mes, año, entre otros) o el tiempo que toma un lance del juego, o la acción que corresponda según el sistema en estudio.

Transición. Cualquier cambio de estado es una transición. En el ejemplo de la computadora, es transición cuando pasa de estar en funcionamiento a descomponerse, o viceversa.

Probabilidad de transición. Probabilidad de que se produzca el cambio de un estado a otro. Lo ideal es que dicha probabilidad, que se puede determinar, permanezca constante en el transcurso de la vida de un sistema.

Vector de probabilidad. Vector que representa la probabilidad de que el sistema se encuentre en un estado particular. En el ejemplo de la computadora, el vector de probabilidad inicial es 1, 0, 0, donde la probabilidad 1 señala la computadora funcional y que hay probabilidad 0 de que esté descompuesta o en reparación. Un vector de probabilidad que podría presentarse en algún momento de la vida del sistema es 0.6, 0.1, 0.3. Esto significa que la computadora tiene una probabilidad de 0.6 (60%) de estar en funcionamiento, de 0.1 (10%) de estar descompuesta y de 0.3 (30%) de estar en reparación.

Proceso estocástico. Sucesión de las diferentes etapas de un sistema, fenómeno o actividad. El estado en el que se encuentra el sistema en cada una de las etapas ocurre en forma aleatoria.

Cadena de Markov. Proceso estocástico mediante el cual el estado de un sistema se puede predecir con cierta probabilidad a partir del conocimiento de su estado anterior. La probabilidad de transición entre los estados deberá permanecer constante en el transcurso de la vida del sistema.

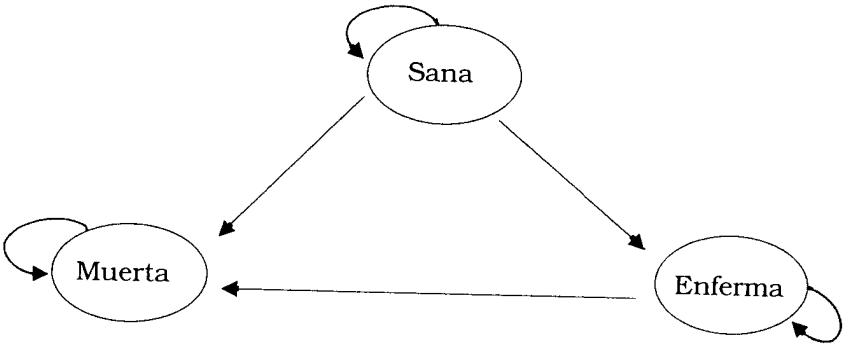
Cadena de Markov absorbente. Cadena de Markov, que tiene uno o varios estados absorbentes. Cuando el sistema llega a uno de éstos, ya no puede salir de él.

5.7.2 Cadenas de Markov

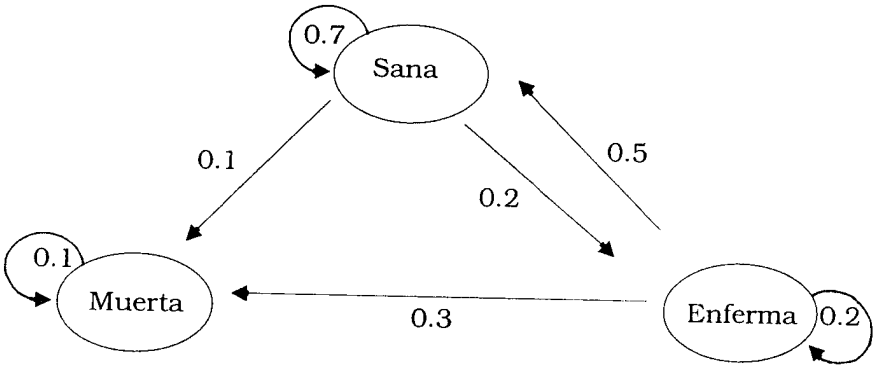
Es un experimento que consiste en una secuencia de ensayos y cumple con las siguientes propiedades:

1. El resultado de cada ensayo (etapa) es uno de un conjunto finito de posibles resultados, llamados *estados*.

Para solucionar estos problemas existe otro tipo de gráfica, denominada *diagrama de transición*, que es muy fácil de construir y lógica de entender:



En el diagrama las flechas indican las posibles transiciones de un estado a otro. Además, para indicar las probabilidades de transición, en el diagrama se pone dicho valor junto a la flecha, por ejemplo:



El diagrama indica que una persona sana tiene en la siguiente etapa una probabilidad de 0.7 de estar sana, de 0.2 de estar enferma y de 0.1 de estar muerta. Por otra parte, una persona enferma se encontrará en la siguiente etapa sana, enferma o muerta, con sus respectivas probabilidades de 0.5, 0.2 y 0.3. Por último, una persona muerta seguirá muerta en la siguiente etapa; por tanto, la probabilidad de permanecer muerta asciende a 1, mientras que la de estar sana o enferma en la siguiente etapa es 0.

5.7.4 Representación matricial

Un proceso estocástico completo puede representarse de modo más conveniente por una matriz, donde los renglones representan los estados iniciales y las columnas los finales. Debido a que los elementos de la matriz son las probabilidades de transición, a aquélla se le suele conocer como *matriz de transición* T, y se puede formar para cualquier cadena de Markov:

Sea:

$$\begin{array}{cc} & \text{Estado final} \\ & \begin{array}{ccc} A & B & C \end{array} \\ \text{Estado inicial} \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{pmatrix} = T \end{array}$$

Las características de la matriz de transición son las siguientes:

a) Es estocástica, donde la suma de los elementos de cada uno de los renglones o columnas constituye la unidad.
b) Es cuadrada, debido a que el número de estados iniciales es igual al de estados finales.
c) $T = T'$ representa la etapa 1 y T^n representa la enésima etapa del proceso estocástico.

Ejemplo *La ruina de un jugador*

Dos oponentes (un jugador y la banca) participan en un juego consistente en arrojar un dado honesto de seis caras. El jugador escoge los números pares (2, 4, 6) o los nones (1, 3, 5), por lo que cada uno de los oponentes tiene una probabilidad de ganar de 50% (0.5).

Sea el caso cuando el jugador cuenta con dos pesos y en cada lance tiene que apostar uno, que puede ganar o perder; por otra parte, la banca cuenta con tres pesos. El jugador suspenderá cuando pierda sus dos pesos o cuando gane los tres de la banca. ¿Qué probabilidad hay de que el jugador participe más de seis veces?

Solución:

Los posibles estados en que puede estar el jugador son:

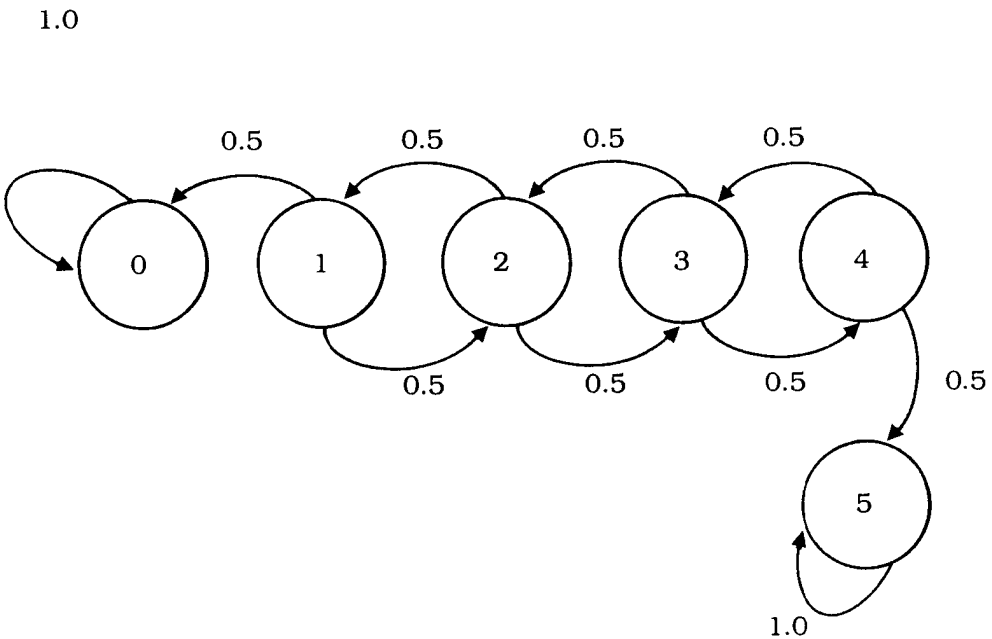
- 0 pesos (cuando pierde todo)
- 1 peso (cuando pierde un peso)
- 2 pesos (cuando inicia el juego)
- 3 pesos (cuando gana un peso a la banca)
- 4 pesos (cuando gana dos pesos a la banca)
- 5 pesos (cuando gana los tres pesos a la banca)

Cada una de estas situaciones posibles representan el número de pesos que tiene el jugador y se denominarán respectivamente, estados 0, 1, 2, 3, 4 y 5.

En cada jugada "etapa" (gane o pierda), el jugador cambiará de estado. Estos cambios de un estado a otro se llaman *transiciones*.

Como el jugador cuenta con dos pesos, se dice que inicia en el estado 2; a la siguiente jugada, se encontrará en el estado 3 si gana, o en el 1 si pierde. Si se halla en el estado 1, todavía le queda un peso por apostar (recuérdese que sólo podrá abandonar el juego cuando gane o pierda todo). Si apuesta ese peso, puede ganar o perder en la siguiente jugada; en el primer caso se encontrará nuevamente en el estado 2, en el segundo, pasará al estado 0, y debe abandonar el juego, pues perdió todo. Del estado 2 puede transitar al 3 si gana, y de ahí al estado 4 y al 5, si se mantiene ganando.

Estas transiciones se notan claramente en el siguiente diagrama de transición del proceso, en donde se aprecia que el jugador puede caer en los estados 0, 1, 2, 3, 4 o 5 (donde tiene 0, 1, 2, 3, 4 o 5 pesos):



En el diagrama se muestran las posibles transiciones y sus probabilidades asociadas entre los estados; por ejemplo, si el jugador se encuentra en el estado 2 hay una probabilidad de 0.5 de pasar al estado 1 (si pierde) y una probabilidad de 0.5 de pasar al 3 (si gana). Los estados 0 y 5 son en este caso los absorbentes, ya que cuando el jugador llega al estado 0 carece de dinero para jugar y, por tanto, acaba el juego, mientras que si llega al 5, significa que ganó los tres pesos de la banca y el juego termina.

Los estados absorbentes se identifican fácilmente en un diagrama de transición por tener sólo entradas pero no salidas hacia otros estados.

Una vez que se tiene el diagrama de transición se puede construir la matriz de transición:
Hacia:

$$
\begin{array}{cc}
 & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\text{De:} \quad \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

Ahora veamos cómo se interpreta la matriz:

- Se puede pasar del estado 0 hacia el 0 con una probabilidad de 1, ya que es un estado absorbente. Si el jugador no tiene dinero, así permanecerá.
- La probabilidad de pasar del estado 0 a cualquier otro estado es 0, ya que una vez sin dinero el jugador no puede apostar y entonces no podrá ganar.
- La probabilidad de pasar del estado 1 al estado 0 es 0.5, ya que si el jugador tiene un peso, la probabilidad de perderlo y quedar con 0 pesos es de 0.5.
- La probabilidad de tener un peso y conservarlo en el siguiente lance, es decir, pasar del estado 1 al 1 de 1 a 1 es 0, ya que al jugar, o gana un peso y pasa a 2, o pierde un peso y pasa a 0.
- La probabilidad de tener un peso y pasar a dos pesos en el siguiente lance (o sea, el estado 2) es de 0.5, ya que hay una probabilidad de 0.5 de ganar.
- La probabilidad de pasar de 1 a 3, 4 o 5 es 0, ya que el jugador sólo puede ganar un peso; entonces, como máximo tendrá dos pesos en la próxima jugada. De manera análoga se interpretan los renglones que parten de 2, 3 y 4.
- Si se tienen cinco pesos significa que el jugador ganó los tres del contrincante; entonces, la probabilidad de conservarlos es de uno, ya que el juego terminó al no haber más apuestas.

En una matriz de transición es tan fácil identificar los estados absorbentes como en los diagramas de transición: los estados que tengan 1 en la diagonal principal son los absorbentes.

En este caso, el estado 0 y el estado 5 tienen 1 en la diagonal principal, es decir, de 0 a 0 hay probabilidad de 1, y de 5 a 5 también 1. Por tanto, los estados 0 y 5 son absorbentes, tal como se había observado en el diagrama de transición.

En el problema, interesa saber qué probabilidad existe de que se jueguen más de seis lances. Con este fin habría que determinar la probabilidad de que el jugador esté en el estado 1, 2, 3 o 4 en el juego 6, o sea, en la transición 6 de la matriz (T^6). Entonces, se eleva la matriz T a la sexta potencia:

$$T = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad T^2 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0.25 & 0 & 0.25 & 0 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 & 0 \\ 0 & 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0 & 0.25 & 0 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$T^3 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.625 & 0 & 0.25 & 0 & 0.125 & 0 \\ 0.25 & 0.25 & 0 & 0.375 & 0 & 0.125 \\ 0.125 & 0 & 0.375 & 0 & 0.25 & 0.25 \\ 0 & 0.125 & 0 & 0.25 & 0 & 0.625 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \qquad T^4 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.625 & 0.125 & 0 & 0.1875 & 0 & 0.0625 \\ 0.375 & 0 & 0.3125 & 0 & 0.1875 & 0.125 \\ 0.125 & 0.1875 & 0 & 0.3125 & 0 & 0.3750 \\ 0.0675 & 0 & 0.1875 & 0 & 0.125 & 0.625 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$T^5 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.6875 & 0 & 0.1563 & 0 & 0.0938 & 0.0625 \\ 0.3750 & 0.1563 & 0 & 0.25 & 0 & 0.2188 \\ 0.2188 & 0 & 0.25 & 0 & 0.1563 & 0.3750 \\ 0.0675 & 0.0938 & 0 & 0.1563 & 0 & 0.6875 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \quad \text{y finalmente} \quad T^6 = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.6875 & 0.0781 & 0 & 0.1250 & 0 & 0.1094 \\ 0.4531 & 0 & 0.2031 & 0 & 0.125 & 0.2188 \\ 0.2188 & 0.1250 & 0 & 0.2031 & 0 & 0.4531 \\ 0.1094 & 0 & 0.1250 & 0 & 0.0781 & 0.6875 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

La matriz T^6 indica la probabilidad que hay de pasar de un estado inicial a cualquier otro después de seis transiciones (en este caso, juegos).

Como se sabe que el jugador comienza con dos pesos, céntrese la atención en el renglón del estado 2:

Hacia:

De:		0	1	2	3	4	5	
	2	[0.4531	0	0.2031	0	0.125	0.2188	]

Es decir:

- La probabilidad de que comience con dos pesos y tenga 0 después de seis juegos es 0.4531
- La probabilidad de que comience con dos pesos y tenga uno después de seis juegos es 0

- La probabilidad de que comience con dos pesos y tenga dos después de seis juegos es 0.2031
- La probabilidad de que comience con dos pesos y tenga tres después de seis juegos es 0
- La probabilidad de que comience con dos pesos y tenga cuatro después de seis juegos es 0.125
- La probabilidad de que comience con dos pesos y tenga cinco después de seis juegos es 0.2188

Como se desea conocer la probabilidad de que se jueguen más de seis lances se deben considerar únicamente los estados que permitirán al jugador proseguir: 1, 2, 3 y 4, ya que si cuenta con 1, 2, 3 o 4 pesos continuará apostando. Los estados 0 y 5 no se toman en cuenta, ya que si se llega a cualquiera de éstos, el juego se detiene al no tener dinero para apostar o al haber ganado todo el dinero del contrincante.

Entonces, si se suman las probabilidades que permiten seguir jugando después de 6 lances (estados 1, 2, 3 y 4), resulta:

$0 + 0.2031 + 0 + 0.125 = 0.3281$

Así, la probabilidad de que el jugador juegue más de seis veces es 0.3281.

Ejemplo En una escuela preparatoria, cuyo ciclo escolar es de tres años, se realiza un estudio estadístico. De éste se obtienen las siguientes proporciones (porcentajes) respecto a la situación escolar del alumnado: 15% de los estudiantes que ingresan en primer grado tuvieron que recursarlo; 60% pasaron a segundo grado y 25% se dieron de baja. La situación de los estudiantes que cursan el segundo grado es la siguiente: 13% lo recursaron, 80% pasaron y 7% se dieron de baja. De los alumnos de tercer grado, 22% tuvieron que recursarlo, 8% se dieron de baja y 70% se graduaron.

La administración de la escuela desea saber la probabilidad de que un alumno se gradúe o se dé de baja, considerando el grado que cursa. Actualmente, el primer grado lo cursan 1 000 alumnos; el segundo año, 820, y el tercero, 640. Si la Secretaría de Educación Pública recomienda a la escuela que se gradúen cuando menos 650 estudiantes, conservando el excelente nivel académico de la institución, calcúlese cuántos alumnos se graduarán de los 1 000 que ingresan. Adicionalmente, si este número es menor que 650, determínese el número de alumnos que tienen que ingresar para cumplir con la recomendación de las autoridades educativas.

Solución

Paso 1. Construir el diagrama de transición que represente la situación en la que se pueda encontrar un alumno, considerando los estados siguientes:

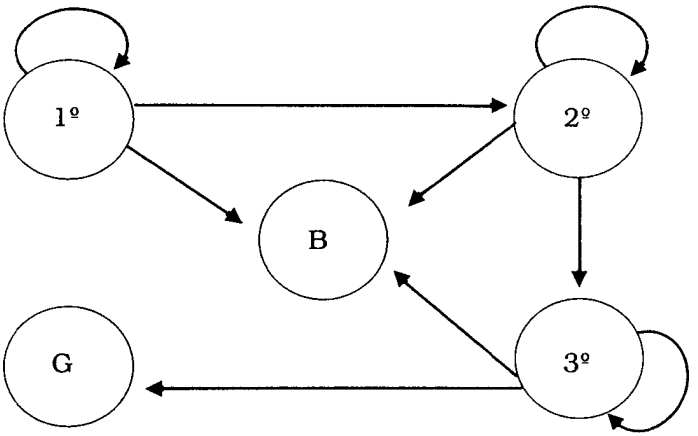
- Primer año.
- Segundo año.
- Tercer año.

- Baja.
- Graduado.

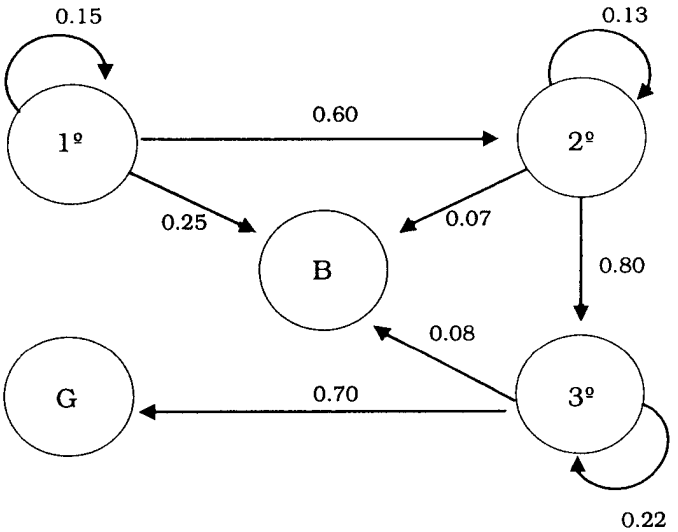
Las posibles transiciones son:

- Pasar al siguiente grado (siempre que esté en primero o segundo año).
- Repetir el grado.
- Darse de baja.
- Graduarse (si está en tercer año).

El diagrama de transición tiene la siguiente forma:



Considerando las probabilidades estimadas anteriormente, se tiene:



Paso 2. Construir la matriz de transición.

$$T = \begin{array}{c} \\ 1^{\circ} \\ 2^{\circ} \\ 3^{\circ} \\ B \\ G \end{array} \begin{array}{c} \begin{array}{ccccc} 1^{\circ} & 2^{\circ} & 3^{\circ} & B & G \end{array} \\ \begin{pmatrix} 0.15 & 0.60 & 0 & 0.25 & 0 \\ 0 & 0.13 & 0.80 & 0.07 & 0 \\ 0 & 0 & 0.22 & 0.08 & 0.70 \\ 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \end{pmatrix} \end{array}$$

Paso 3. Determinar la matriz estacionaria (utilizando el *software* educativo). Se encuentra que a partir de T^{10}, la matriz se estaciona.

$$T^{10} = \begin{array}{c} \\ 1^{\circ} \\ 2^{\circ} \\ 3^{\circ} \\ B \\ G \end{array} \begin{array}{c} \begin{array}{ccccc} 1^{\circ} & 2^{\circ} & 3^{\circ} & B & G \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0.4175 & 0.5825 \\ 0 & 0 & 0 & 0.1748 & 0.8252 \\ 0 & 0 & 0 & 0.1026 & 0.8974 \\ 0 & 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \end{pmatrix} \end{array}$$

Paso 4. De esta matriz se obtiene la probabilidad de que un estudiante se dé de baja o se gradúe, considerando el año que cursa. Para un alumno que cursa el primer año, la probabilidad de darse de baja asciende a 41.75% (0.4175), mientras que la de graduarse es 58.25% (0.5825). Si está en segundo año, las probabilidades del alumno son: 17.48% (0.1748) y 82.52% (0.8252) de darse de baja y graduarse, respectivamente. Por otra parte, un estudiante de tercer grado tiene una probabilidad de 10.26% (0.1026) de darse de baja y 89.74% (0.8974) de graduarse.

Paso 5. Actualmente ingresan 1 000 alumnos en primer grado y la probabilidad de que se gradúen asciende a 58.25%. Así, egresarán 1 000 × 0.5825 = 582.5 (583 alumnos), un número de graduados inferior al de 650, recomendado por la Secretaría de Educación Pública. ¿Qué acción se podría emprender para incrementar el número de graduados?

Paso 6. Una solución es incrementar el número de alumnos de nuevo ingreso. En ese caso, si de 1 000 estudiantes que ingresan 583 se gradúan, ¿cuántos alumnos deben inscribirse para que se egresen al menos 650?

Si 1 000 es a 582.5, ¿cuánto es a 650?

1 000 → 582.5

? → 650

(1 000 × 650) ÷ 582.5 = 1 115.88 (aproximadamente 1 116)

Con 1 116 alumnos que se acepten en el primer grado se espera que se gradúen 650 (1 116 × 58.25% = 650.07)

Ejemplo Los locales de un nuevo centro comercial están a punto de ofrecerse en alquiler. Existe un centro comercial en servicio con características similares. El administrador decide implantar una norma según la cual el arrendatario deberá pagar un depósito por tres meses de alquiler. Así, el local podrá encontrarse en alguna de las siguientes situaciones:

a) Al día en sus pagos.
b) Atrasado con un pago.
c) Atrasado con dos pagos.
d) Atrasado con tres pagos.
e) Disponible para ofrecerse en alquiler.

En el último inciso se pueden presentar dos casos:

1. Se deben tres pagos y ninguno se efectúa posteriormente.
2. El arrendatario decide desocupar el local (en el caso de que adeude uno, dos o tres meses o incluso cuando se halle al día).

El centro comercial dispone de 130 locales, de los cuales 85 fueron alquilados a partir del día de la inauguración.

El administrador desea saber en qué situación se encontrarán estos 85 locales después de un año, es decir, el porcentaje que estará:

a) Al día con sus pagos.
b) Atrasado con uno, dos o tres pagos.
c) Desocupado.

Asimismo, de los 45 locales no alquilados el día de la inauguración, el administrador desea conocer el porcentaje que estará ocupado un año después.

Por otra parte, la experiencia con el centro comercial en servicio durante varios años ha proporcionado las estadísticas siguientes:

- De los locales al día con los pagos, 63% se mantiene así el siguiente mes; 35% deben una mensualidad, y 2% estarán desocupados.
- De los locales que deben una mensualidad, 22% estarán al día el siguiente mes, 28% seguirán atrasados con un alquiler; 45% estarán atrasados con dos y 5% se desocuparán.
- De los locales que están atrasados con dos pagos, 10% se pondrán al día el siguiente mes, 37% estarán atrasados únicamente con un pago, 33% con dos; 11% con 3, y 9% estarán desocupados.
- De los locales donde el pago se atrasó tres meses, sólo 6% se pondrán al día el siguiente mes; 34% deberán una mensualidad; 27% deberán dos pagos; 17% tres, y 16% tendrán que desocupar, pues están atrasados con tres mesualidades y no realizan pago alguno.
- De los locales desocupados, 60% seguirán así al siguiente mes, mientras que 40% se ocuparán.

Además, el administrador desea determinar el momento en que se detendrá el proceso, el porcentaje de los locales que permanecerá al día en sus pagos, el

porcentaje que permanecerá desocupado, así como el porcentaje que no estará al día con sus pagos.

Solución

Paso 1. Construir el diagrama de transición que represente la situación en la que puede encontrarse un local. Los estados respectivos son:

- Al día en sus pagos (0).
- Atrasado con un pago (1).
- Atrasado con dos pagos (2).
- Atrasado con tres pagos (3).
- Desocupado (D).

Las transiciones posibles entre los estados son:

a) Si un local está al corriente, al siguiente mes puede:
- Permanecer al día.
- No pagar y, por tanto, atrasarse con un pago.
- Desocuparse.

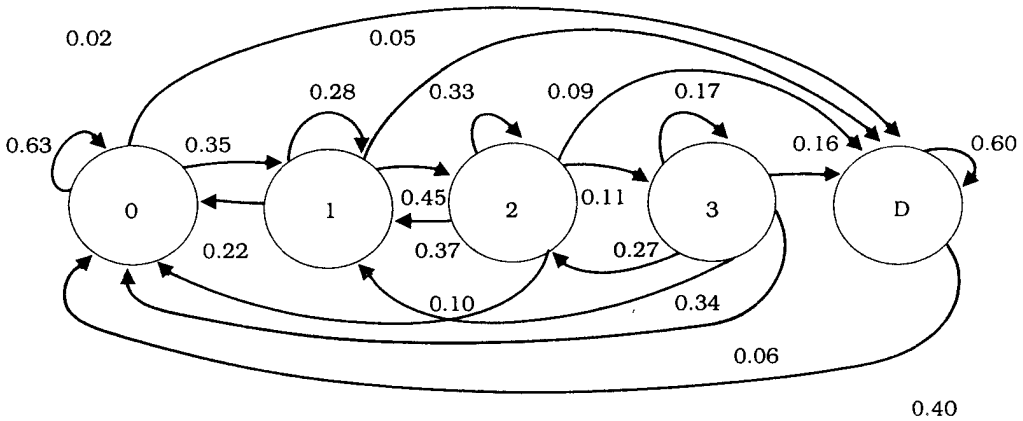
b) Si un local está atrasado con un pago, al siguiente mes puede:
- Estar al día (si paga el mes atrasado y el mes correspondiente).
- Seguir atrasado con un pago (si sólo paga un mes).
- Atrasarse con dos meses (si no efectúa pago alguno).
- Desocuparse.

c) Si un local está atrasado con dos pagos, al siguiente mes puede:
- Ponerse al día (si paga tres meses).
- Estar atrasado con un mes (si paga sólo dos).
- Seguir atrasado con dos meses (si paga uno solamente).
- Estar atrasado con tres meses (si no paga).
- Desocuparse.

d) Si un local está atrasado con tres pagos, al siguiente mes puede:
- Ponerse al día (si paga cuatro meses).
- Estar atrasado con un mes (si paga tres).
- Estar atrasado con dos meses (si paga dos).
- Estar atrasado con tres meses (si paga uno).
- Desocuparse (si no paga).

e) Si un local está desocupado, al siguiente mes puede:
- Seguir desocupado.
- Estar ocupado.

El diagrama de transición respectivo es el siguiente:



Paso 2. Construir la matriz de transición:

$$T = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ D \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & D \end{array} \\ \begin{pmatrix} 0.63 & 0.35 & 0 & 0 & 0.02 \\ 0.22 & 0.28 & 0.45 & 0 & 0.05 \\ 0.10 & 0.37 & 0.33 & 0.11 & 0.09 \\ 0.06 & 0.34 & 0.27 & 0.17 & 0.16 \\ 0.40 & 0 & 0 & 0 & 0.60 \end{pmatrix} \end{array}$$

Paso 3. Obtener T^{12}, ya que esta matriz representará la situación de los locales en el mes 12.

Utilizando el programa MacStat, se obtiene:

$$T^{12} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ D \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & D \end{array} \\ \begin{pmatrix} 0.3569 & 0.2939 & 0.2088 & 0.0277 & 0.1126 \\ 0.3574 & 0.2938 & 0.2084 & 0.0276 & 0.1127 \\ 0.3576 & 0.2938 & 0.2083 & 0.0276 & 0.1127 \\ 0.3577 & 0.2939 & 0.2083 & 0.0276 & 0.1126 \\ 0.3576 & 0.2972 & 0.2086 & 0.0276 & 0.1120 \end{pmatrix} \end{array}$$

El administrador puede obtener de esta matriz el porcentaje de los 85 locales ocupados al momento de la inauguración. Estos locales comenzaron a partir de

la inauguración, por lo que entonces estaban al día (estado 0). En este caso, había que considerar el renglón 0:

$$\begin{array}{cccccc} & 0 & 1 & 2 & 3 & D \\ D & [0.3569 & 0.2939 & 0.2088 & 0.0277 & 0.1126] \end{array}$$

Entonces:

a) 35.69% de los 85 locales estarán al día en un año, esto es, aproximadamente 30 locales (85 × 0.3569 = 30.3365).

b) 29.39% estarán atrasados con un pago en un año, lo que da 25 locales aproximadamente (85 × 0.2939 = 24.9815).

20.88% estarán atrasados dos pagos en un año, aproximadamente 18 locales (85 × 0.2088 = 17.748).

2.77% estarán atrasados con tres pagos dentro de un año, es decir, dos locales aproximadamente (85 × 0.0277 = 2.3545).

c) 11.26% estarán desocupados dentro de un año, o sea, 10 locales aproximadamente (85 × 0.1126 = 9.571).

En resumen, de los 85 locales, después de un año:

- 35.69% estarán al día (alrededor de 30 locales).
- 53.04% estarán atrasados con uno o varios pagos (aproximadamente 45 locales).
- 11.26% estarán desocupados (10 locales aproximadamente).

Ahora, el administrador desea saber, de los 45 locales desocupados al momento de la inauguración, qué porcentaje estará ocupado un año después. Por tanto, habrá que considerar el renglón D de T^{12}:

$$\begin{array}{cccccc} & 0 & 1 & 2 & 3 & D \\ D & [0.3576 & 0.2942 & 0.2086 & 0.0276 & 0.1120] \end{array}$$

Los locales que estarán ocupados son los que se encuentran al día o deben algún pago, esto es, 35.76 + 29.42 + 20.86 + 2.76 = 88.8 por ciento.

Una manera más simple de obtener el porcentaje es restar a 100% de los locales desocupados en la inauguración el porcentaje de locales desocupados después de un año, es decir, 100 − 11.20 = 88.8 por ciento.

De los 45 locales, 88.8% equivalen a 40 aproximadamente (45 × 0.888 = 39.96), esto es alrededor de 40 de los 45 locales desocupados en la inauguración se habrán ocupado después de un año.

Por último, al administrador le interesa establecer el momento en que se estaciona el proceso, a fin de obtener el porcentaje de locales al día, atrasados y desocupados. Para obtener esto, se continúa elevando la matriz hasta que se

Si los eventos son dependientes pueden ser o no mutuamente excluyentes. Algunas leyes y propiedades que rigen la probabilidad clásica permiten calcular probabilidades a partir de otras probabilidades.

El teorema de Bayes, por tanto, es una de las leyes más importantes de la probabilidad; aunque hemos desarrollado nuestro concepto de probabilidad en el supuesto de que en el espacio muestral todos los resultados son igualmente probables, de hecho estas mismas propiedades y leyes no se aplican únicamente al modelo clásico.

Si el principal objetivo de la estadística es hacer inferencias acerca de una población, con base en la información contenida en una muestra, la probabilidad es el mecanismo adecuado para ello.

Muchos experimentos, en las ciencias del comportamiento, concluyen ya sea en resultados numéricos o en valores nominales a los que fácilmente se les asignan valores numéricos. Estos valores pueden ser representados por una variable aleatoria. La diferencia entre los dos tipos de variables aleatorias, discretas y continuas, es que se asignan probabilidades a valores individuales y a intervalos, respectivamente.

Las distribuciones de probabilidad para las variables aleatorias pueden derivarse matemáticamente utilizando la teoría de probabilidad, o aproximarse de manera empírica repitiendo el experimento un gran número de veces y construyendo un histograma de la frecuencia relativa de los resultados. Estas distribuciones son modelos de la población y se caracterizan por la media y la desviación estándar.

No son las probabilidades numéricas el principal objeto de la teoría; su propósito es descubrir leyes generales y elaborar modelos teóricos satisfactorios.

PROBLEMAS

5.1 En una caja hay 10 canicas rojas, cinco blancas y cinco azules. Si R es el conjunto de todas las canicas rojas, B el de las blancas y A el de las azules, calcule:

a) N
b) $P(R)$
c) $P(B)$
d) $P(A)$

5.2 Si $S = \{2, 4, 6, 8, 10\}$ y $A = \{2, 4\}$

Calcule:

a) $P(A)$
b) $P(\overline{A})$

Compruebe $P(A) + P(\overline{A}) = 1$

estacione; así, mediante el programa MacStat resulta que la matriz se estaciona en T^{20}:

$$T^{20} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ D \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & D \end{array} \\ \begin{pmatrix} 0.3573 & 0.2939 & 0.2085 & 0.0276 & 0.1126 \\ 0.3573 & 0.2939 & 0.2085 & 0.0276 & 0.1126 \\ 0.3573 & 0.2939 & 0.2085 & 0.0276 & 0.1126 \\ 0.3573 & 0.2939 & 0.2085 & 0.0276 & 0.1126 \\ 0.3573 & 0.2939 & 0.2085 & 0.0276 & 0.1126 \end{pmatrix} \end{array}$$

De aquí se obtiene que a partir del mes 20:

- 35.73% de los locales estarán al día.
- 29.39% de los locales estarán atrasados con un pago.
- 20.85% de los locales estarán atrasados con dos pagos.
- 2.76% de los locales se atrasarán con tres pagos.
- 11.26% de los locales permanecerán desocupados.

RESUMEN

Tanto en situaciones profesionales como en las cotidianas, diseñamos[9] decisiones, encarándonos con la incertidumbre o ignorancia.

El objetivo de analizar el concepto de probabilidad es poder expresar de manera congruente, significativa y confiable, nuestra forma de concebir la incertidumbre respecto a la ocurrencia de los eventos. La tarea de definir la probabilidad puede considerarse mediante procedimientos diferentes, pero al mismo tiempo relacionados:

a) el *empírico*, en el cual la probabilidad de ocurrencia de un evento se concibe como la frecuencia relativa, y

b) el *clásico*, en el que la probabilidad de que ocurra un evento se define como la proposición de los resultados en un espacio muestral suponiendo que todos los resultados en el espacio de eventos son equiprobables.

Ciertos principios de conteo que incluyen la noción de permutaciones, ordenaciones o combinaciones, ayudan a encontrar los resultados en un espacio muestral o en un evento; en esta forma se evita enumerar todos los posibles resultados.

Un conjunto de dos o más eventos puede ser descrito, en la mayor parte de los casos, como independiente o dependiente.

---

[9] Erróneamente se le suele llamar *toma de decisiones*.

c) Calcule la probabilidad de que sea hembra, esté previamente entrenada y haya tomado a la izquierda.
d) Calcule la probabilidad de que sea macho, no esté entrenado y tomará a la izquierda.
e) Calcule la probabilidad de que sea hembra, esté entrenada y no haya tomado a la izquierda.

5.9 Se realizó una entrevista a 885 amas de casa y se recabó la información siguiente:

600 veían telenovelas
400 series policiacas
620 programas deportivos
195 telenovelas y series policiacas
140 series policiacas y deportivas
400 telenovelas y deportivos
y todas, al menos, ven uno de estos programas

Calcular la probabilidad de:

a) Que vean únicamente telenovelas
b) Que vean telenovelas y series policiacas
c) Escoger una ama de casa que no vea ningún programa
d) Que vean series policiacas y programas deportivos
e) Que vean series policiacas y programas deportivos, pero no telenovelas, o telenovelas y deportes, pero no policiacas

5.10 A dos pacientes del mismo hospital, un hombre y una mujer, se les ha diagnosticado cáncer en el estómago. Los médicos especialistas pronostican que vivirán cuando menos un año más, y la probabilidad es la siguiente: para el hombre 30% y para la mujer 40%.

Si se considera que los eventos son independientes, calcule que:

a) Los dos vivan
b) El hombre no viva
c) La mujer no viva
d) Ni la mujer ni el hombre vivan
e) Al menos uno viva
f) Ambos no vivan
g) Ubique las probabilidades en un diagrama de Venn-Euler

5.11 Un noticiero anuncia el pronóstico del tiempo. Se estima la probabilidad de que llueva hoy $P(E_2) = 0.20$, de que llueva mañana $P(E_1) = 0.22$ y de que llueva hoy y mañana $P(E_1 \cap E_2) = 0.14$.

5.3 En un cesto hay 30 perritos y 20 perritas. Si se va a sacar un cachorro al azar, y M es el evento "sacar una perrita" y H es "sacar un perrito", calcule:

a) N *b)* $n(M)$ *c)* $n(H)$ *d)* $P(M)$ *e)* $P(H)$

5.4 Un experimento consiste en el lanzamiento de una moneda y un dado simultáneamente; si E_1 es el hecho de que salga "cara" en el lanzamiento de una moneda y E_2 es el hecho de obtener 3 o 6 en el lanzamiento del dado, explique qué significado tiene cada uno de los siguientes eventos:

a) $\overline{E}_1$ *b)* $\overline{E}_2$ *c)* $P(E_1 \cap \overline{E}_2)$

5.5 En el mercado existen cuatro marcas diferentes de agua purificada para beber, en presentación de un litro. Si se considera que éstas tienen la misma calidad y una persona desea adquirirla sin considerar alguna en especial (A, B, C, D):

a) Describa el espacio muestral y sus probabilidades correspondientes.
b) Para el caso de que tenga que adquirir dos litros, ¿cuál será el espacio muestral y las probabilidades correspondientes si:
  i) los dos litros deben ser de marcas diferentes?
  ii) las marcas se pueden repetir?

5.6 Dados $S = \{1, 2, 3\}$, $A = \{1\}$, $B = \{3\}$, $C = \{2\}$, encuentre:

a) $P(C)$
b) $P(A \cup B)$
c) $P(\overline{A})$
d) $P(\overline{A} \cap \overline{B})$
e) $P(\overline{A} \cup \overline{B})$
f) $P(B \cup C)$

5.7 Dado un evento en el que $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \cup B) = \frac{2}{3}$, calcule:

a) $P(\overline{A})$
b) $P(\overline{B})$
c) $P(A \cap B)$
d) $P(\overline{A} \cap \overline{B})$
e) $P(\overline{A} \cup \overline{B})$
f) $P(A \cap \overline{B})$
g) $P(\overline{A} \cup \overline{B})$
h) $P(\overline{A} \cup B)$

5.8 Después de que 100 ratones corrieron por un laberinto, se detectaron los siguientes datos: 50 ratones eran machos, 50 fueron previamente entrenados, 42 tomaron a la izquierda en la primera oportunidad, 21 eran machos previamente entrenados, siete machos tomaron a la izquierda, 30 ratones previamente entrenados tomaron a la izquierda y cinco machos previamente entrenados tomaron a la izquierda.

a) Ubique los datos en un diagrama de Venn.
b) Ubique los datos en un diagrama de Carroll.

ras, cuando están fuera de servicio, dependa de la condición actual de las otras dos. El funcionamiento de cada una es independiente entre si. Calcule la probabilidad de que:

a) Las tres fotocopiadoras estén fuera de servicio.
b) La número 1 esté fuera de servicio, pero la 2 y 3 sigan funcionando.
c) Las número 1 y 2 estén fuera de servicio y la 3 siga funcionando.
d) Una de las tres esté fuera de servicio.
e) Dos de las tres estén fuera de servicio.
f) Al menos una esté fuera de servicio.
g) Por lo menos dos estén fuera de servicio, donde: x = fotocopiadora 1 funcionando, y = fotocopiadora 2 funcionando, y z = fotocopiadora 3 funcionando. Obviamente, $\bar{x}$, $\bar{y}$, $\bar{z}$ significa que no funciona la fotocopiadora, respectivamente.

5.14 Considerando el ejercicio anterior, si el tiempo fuera de servicio de las fotocopiadoras aumenta un 20%, calcule la probabilidad de que:

a) Al menos una fotocopiadora esté fuera de servicio.
b) Dos o más fotocopiadoras estén fuera de servicio.

5.15 La probabilidad de que una persona con síndrome de inmunodeficiencia adquirida tenga una reacción positiva es del 89%; la probabilidad de que una persona sin sida tenga una reacción positiva es del 2%; en una comunidad del continente africano, el 20% tiene sida. Calcule todos los eventos posibles (aplique el teorema de Bayes).

5.16 Sean dos eventos A y B con su no ocurrencia $\bar{A}$ y $\bar{B}$ (los eventos A y B no son mutuamente excluyentes).

a) Ubique las probabilidades en un diagrama de Carroll (tabla de dos entradas).
b) Deduzca la probabilidad condicional correspondiente.
c) Deduzca el teorema de Bayes correspondiente.

5.17 En una empresa los empleados tienen las características siguientes:

317 son hombres
316 son casados
25 son mujeres casadas sin profesión
72 son hombres casados sin profesión
83 son hombres profesionales solteros
15 son mujeres profesionales solteras
125 son hombres profesionales casados
49 son mujeres solteras sin profesión

a) Calcule el número total de empleados.
b) ¿Cuál es la probabilidad de que sea profesional?

a) Complete la siguiente tabla:

		Llueva mañana E_1	No llueva mañana E_1	Total
Llueva hoy	E_2	$(E_1 \cap E_2)$ 0.14		0.20
No llueva hoy	E_2			
		0.22		1

b) Si la probabilidad de que llueva mañana, dado que llueva hoy, es $P(E_1 \mid E_2) = 0.7$ y de que no llueva mañana, dado que no llueva hoy, es 0.9, calcule la probabilidad de que llueva mañana dado que no llueva hoy, y de que llueva mañana dado que no llueva hoy. Utilice un diagrama de árbol.

c) Calcule la probabilidad de que llueva mañana.

5.12 En un examen de estadística, aplicado a 300 alumnos, se obtuvieron las calificaciones de la tabla siguiente:

Eventos	Hombres (E_6)	Mujeres (E_7)	Total
E_1 : MB	20	5	25
E_2 : B	15	50	65
E_3 : S	5	90	95
E_4 : NA	0	80	80
E_5 : NP	0	35	35

a) Obtenga $P(E_1)$
b) Obtenga $P(E_1 \cup E_6)$
c) Obtenga $P(E_1 \cap E_6)$
d) Obtenga $P(E_6 \mid \omega)$
e) $P(E_1 \cup E_6 \mid \omega)$
f) $P(E_1 \mid E_6)$
g) $P(E_6 \mid E_1)$

donde Ω = Universo

5.13 En el departamento de fotocopiado de una universidad existen tres copiadoras que fueron adquiridas al mismo tiempo, con las mismas características técnicas para una gran demanda de trabajo. Este tipo de copiadora está fuera de servicio 10% del tiempo de uso (por mantenimiento y reparación). Suponga la posibilidad de que ninguna de las fotocopiado-

5.24 Veinte por ciento de los empleados de una compañía son profesionales. De éstos, 75% están asignados a la dirección general; de los que no son profesionales 20% está también en la dirección general.

a) Haga un diagrama de árbol.
b) ¿Cuál es la probabilidad de que sea profesional y esté asignado a la dirección general?
c) ¿Cuál es la probabilidad de que sea profesional y no esté asignado a la dirección general?
d) ¿Cuál es la probabilidad de que sea profesional y esté asignado a la dirección general?
e) ¿Cuál es la probabilidad de que sea profesional, dado que está asignado a la dirección general?
f) ¿Cuál es la probabilidad de que esté asignado a la dirección general, dado que es profesional?
g) ¿Cuál es la probabilidad de que no esté asignado a la dirección general, dado que es profesional?
h) ¿Cuál es la probabilidad de que no sea profesional, dado que está asignado a la dirección general?
i) ¿Cuál es la probabilidad de que esté asignado a la dirección general, dado que no es profesional?

5.25 Una empresa aplica un examen de aptitud para un trabajo técnico. Su experiencia es que la probabilidad de que un candidato pueda aprobar el examen es de 0.6.

Si un candidato aprueba el examen, la probabilidad de que realice el trabajo satisfactoriamente es de 0.80.

Si no pasa el examen, la probabilidad de que no realice el trabajo satisfactoriamente es de 0.40.

a) Haga un diagrama de árbol.
b) ¿Cuál es la probabilidad de que pase el examen, dado que realiza satisfactoriamente el trabajo?
c) ¿Cuál es la probabilidad de que pase el examen, dado que no realiza satisfactoriamente su trabajo?
d) ¿Cuál es la probabilidad de que realice satisfactoriamente su trabajo, dado que pasó el examen?
e) ¿Cuál es la probabilidad de que no realice su trabajo satisfactoriamente, dado que pasó el examen?
f) ¿Cuál es la probabilidad de que realice su trabajo satisfactoriamente, dado que no pasó el examen?
g) ¿Cuál es la probabilidad de que no realice su trabajo satisfactoriamente, dado que no pasó el examen?

5.26 Las estudiantes de la Facultad de Psicología fueron clasificadas en altas, bajas, hermosas, listas, pedantes y tontas.

5.18 Con base en su experiencia, un médico ha recabado la siguiente información, relativa a las enfermedades de sus pacientes: 5% creen tener cáncer y lo tienen, 45% creen tener cáncer y no lo tienen, 10% no creen tener cáncer y lo tienen y, finalmente, 40% creen no tenerlo y es cierto. De entre los pacientes del médico estos porcentajes implican las siguientes probabilidades para un paciente seleccionado al azar. Calcule:

a) P (lo tenga | crea).
b) P (lo tenga | no crea).
c) P (lo crea | no tenga).
d) P (lo crea | lo tenga).

5.19 Suponga que la ciencia médica ha desarrollado una prueba para el diagnóstico del cáncer que tiene 95% de exactitud, tanto en los que tienen cáncer como entre los que no lo tienen. Se sabe que el 0.005 de la población realmente tiene cáncer.

a) Calcule la probabilidad de que determinado individuo tenga cáncer si la prueba dice que lo tiene.
b) ¿Cuál es la probabilidad de que determinado individuo no tenga cáncer si la prueba dice que lo tiene?
c) ¿Cuál es la probabilidad de que determinado individuo no tenga cáncer si la prueba dice que no lo tiene?
d) ¿Cuál es la probabilidad de que determinado individuo tenga cáncer si la prueba dice que no lo tiene?

5.20 Un apostador de carreras de caballos estima que la probabilidad de que gane hoy es de 30%, la probabilidad de que gane con el mismo caballo la siguiente semana es de 40%, y de que gane hoy y la siguiente semana es de 8%. Calcule todos los eventos posibles.

5.21 Se selecciona al azar un estudiante de una escuela preparatoria, en la cual 3% padece insomnio. Se aplica una prueba psicológica para detectar dicho padecimiento; la probabilidad de que el resultado de dicha prueba sea positivo dado que tiene insomnio es de 95%; la probabilidad de que una persona sin insomnio, pero con resultados positivos es de 2%. Calcule todos los eventos posibles.

5.22 Un apostador de carreras de galgos estima que la probabilidad de que gane hoy es de 0.40, la probabilidad de que gane la siguiente semana es de 0.30, de que gane hoy y la siguiente semana es del 0.08. Calcule todos los eventos posibles.

5.23 Se aplica una prueba de aptitud para candidatos a ingresar en una universidad. Dicho examen lo aprueba 60% de los aspirantes. De los que exentaron, el 80% termina exitosamente el semestre; se selecciona una muestra aleatoria de los aspirantes que no aprobaron, pero cursan el semestre, y 50% de este grupo lo terminan con éxito. Calcule todos los eventos posibles.

a) ¿Cuál es la probabilidad de que una persona tomada al azar tenga un nivel de colesterol por debajo de 160 mg/100 ml?
b) ¿Cuál es la probabilidad de que una persona tomada al azar tenga un nivel de colesterol mayor de 340 mg/100 ml?
c) ¿Cuál es la probabilidad de que una persona tomada al azar tenga un nivel menor de 160 mg/100 ml o mayor de 340 mg/100 ml?

5.29 En un centro de salud se desea hacer un estudio sobre las necesidades médicas y dentales que existen en su población, considerando también el sector donde laboran. Los resultados obtenidos se muestran en la tabla siguiente.

	NECESIDADES		
	Dental	*Médica*	*Total*
Público	470	280	750
Privado	110	140	250
Total	580	420	1000

a) ¿Cuál es la probabilidad de que una persona tomada al azar trabaje en el sector público?
b) ¿Cuál es la probabilidad de que una persona tomada al azar necesite atención médica?
c) Calcule la probabilidad de $P(A \cup B)$ (según los incisos anteriores).
d) ¿Cuál es la probabilidad de seleccionar una persona al azar que necesite servicio dental y trabaje en el sector privado?

5.30 Dadas las matrices siguientes, construya su diagrama de transición:

a) $\begin{bmatrix} 0.667 & 0.333 \\ 0.250 & 0.750 \end{bmatrix}$

b) $\begin{bmatrix} 0.167 & 0.500 & 0.333 \\ 0.500 & 0.250 & 0.250 \\ 0.417 & 0.250 & 0.333 \end{bmatrix}$

c) $\begin{bmatrix} 0.500 & 0.300 & 0.200 & 0.000 \\ 0.00\underline{0} & 0.00\underline{0} & 0.50\underline{0} & 0.50\underline{0} \\ 0.33\overline{3} & 0.33\overline{3} & 0.33\overline{3} & 0.33\overline{3}^{a} \\ 0.250 & 0.750 & 0.000 & 0.000 \end{bmatrix}$

d) $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

[a] El número $0.33\overline{3}$ significa que el número 3 se repite.

22 son hermosas, listas y altas
17 son hermosas, listas y bajas
13 son altas, pedantes y listas
cuatro son pedantes, altas y tontas
18 son pedantes, listas y bajas
11 son hermosas, tontas y altas
siete son bajas, hermosas y tontas
cinco son tontas, bajas y pedantes

a) Haga un diagrama de Carroll.
b) Calcule todos los eventos posibles y sus probabilidades correspondientes.

5.27 En un vuelo de cierta línea aérea viajan 18 muchachos, 178 hombres, 10 mexicanos del sexo masculino, dos muchachos mexicanos, 26 personas de nacionalidad mexicana y 14 muchachas extranjeras, sin contar la tripulación.

a) Haga un diagrama de Carroll.
b) Describa todos los eventos posibles y sus probabilidades correspondientes.

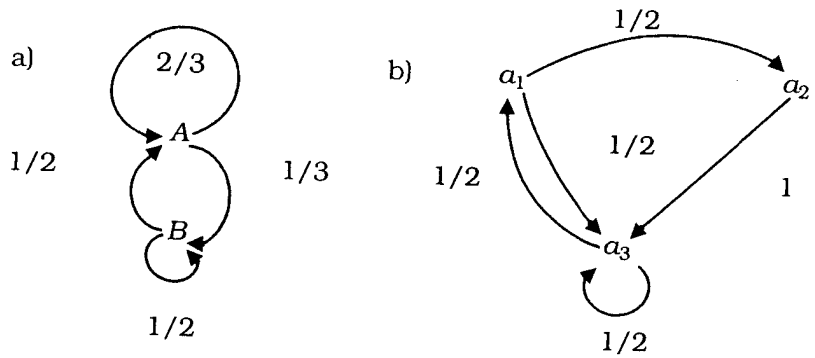
5.28 En un hospital se desea realizar un estudio sobre colesterol en hombres sanos de 40 a 60 años de edad. A continuación se presenta una tabla con los datos obtenidos.

NIVELES DE COLESTEROL EN HOMBRES SANOS
(de 40 a 60 años)

Colesterol (mg/100 ml)					
LRI	*LRS*	*MC*	f	Frecuencia relativa (porcentaje) [a]	Frecuencia acumulada (porcentaje) [a]
119.5	139.5	129.5	10	1	1
139.5	159.5	149.5	21	2	3
159.5	179.5	169.5	37	3.5	6.5
179.5	199.5	189.5	97	9.3	15.8
199.5	219.5	209.5	152	14.5	30.3
219.5	239.5	229.5	206	19.7	50.0
239.5	259.5	249.5	195	18.7	68.7
259.5	279.5	269.5	131	12.5	81.2
279.5	299.5	289.5	96	9.2	90.4
299.5	319.5	309.5	47	4.5	94.9
319.5	339.5	329.5	30	2.9	97.8
339.5	359.5	349.5	13	1.2	99.0
359.5	379.5	369.5	6	0.6	99.6
379.5	399.5	389.5	4	0.4	
				100%	100%

[a] Valores aproximados.

5.31 Dado el diagrama de transición, encuentre su matriz:



5.32 Un empleado del área B tiene 0.5 de probabilidad de ser transferido a A y 0.5 de ser transferido a C. Ningún empleado permanece en B por una segunda semana consecutiva. En el caso de los empleados del área C, la probabilidad es de 0.2 de ser transferidos a A, 0.3 de ser transferidos a B y 0.5 para permanecer en C.

Consideremos cada reubicación del personal como un ensayo en el cual el resultado está controlado por probabilidades. Primero veamos la naturaleza de las condiciones que controlan las transferencias; obsérvese que la probabilidad de ser transferido a cada una de las áreas depende solamente del área a la cual es asignado el candidato en el tiempo de la transferencia. No es necesario considerar la cantidad de tiempo que el candidato ha permanecido en cada área en el pasado; el único efecto de asignación anterior es que la ubicación de la semana previa está en el área actual.

Nótese que esta política describe una secuencia de transferencias que se habrán de efectuar con el mismo conjunto de probabilidades. El patrón de transferencia no varía.

a) Construya la matriz estocástica.
b) Construya el diagrama de transición.
c) Construya un diagrama de árbol.

5.33 Rosalba y Laura son dos jugadoras de tenis del mismo nivel y de características parecidas. Son escogidas al azar para sostener algunos juegos; sin embargo, cada vez que Rosalba gana un juego, adquiere mayor confianza en sí misma y la probabilidad de que gane el siguiente juego es de 0.60; por otro lado cada vez que Rosalba pierde un juego se deprime y la probabilidad de que gane el siguiente disminuye a 0.30:

a) ¿Cuál es la probabilidad de que Rosalba gane en el primer juego?
b) Construya la matriz de transición para Rosalba.
c) ¿Cuál es la probabilidad de que gane en el segundo juego?
d) ¿Cuál es la probabilidad de que gane en el tercer juego?
e) Construya un diagrama de árbol y resuelva el promedio por este método.

5.34 Se recopilaron datos durante 15 temporadas de lluvia en el norte de México. Una buena descripción de la ocurrencia de lluvias está dada por una cadena de Markov de orden dos. Los datos obtenidos fueron los siguientes:

		DÍA POSTERIOR	
		Seco = E_1	Mojado = E_2
DÍA ANTERIOR	Seco = E_1	1049	350
	Mojado = E_2	351	687

a) Calcule la matriz de transición, si E_1 = estado seco y E_2 = mojado.
b) Haga un diagrama de transición.

5.35 Por informaciones recientes se sabe que en la ciudad de México y zona metropolitana 0.1% de sus habitantes transfirieron su residencia a provincia en un año; en cambio 0.8% de los habitantes de provincia se mudaron a la ciudad de México y zona metropolitana. Suponiendo que dichos porcentajes permanecen constantes, determine la distribución de probabilidad de los residentes de provincia y de la ciudad de México con su zona metropolitana.

	Ciudad de México y zona metropolitana	Provincia
A =	(0.25	0.75)

a) Construya la matriz de transición.
b) A los 3 años.
c) A los 10 años.

5.36 Suponga que la probabilidad de que un padre con intolerancia a la leche herede a su hijo dicha deficiencia enzimática es de 0.7; la probabilidad de otro padre sin ninguna deficiencia es de 0.4:

a) ¿Cuál es la probabilidad de que el padre intolerante tenga un bisnieto similar a él?
b) ¿Cuál es la distribución de probabilidad?
c) ¿Depende de que los abuelos sufren la misma intolerancia?

5.37 Una ensambladora de automóviles está promoviendo un modelo (E_1) de auto. El resultado de dicha promoción es que 75% de las personas que prefirieron ese modelo el primer año de su fabricación, adquieran el modelo reciente al siguiente año; de las personas que adquirieron otra marca (E_2) el año anterior, 35% cambia al modelo E_1 al siguiente año.
¿Qué porcentaje de automovilistas adquirirán el modelo E_1 después de dos años, si 50% adquiere dicho modelo ahora?

a) Construya la matriz de transición.
b) Dibuje un diagrama de árbol.
c) Concluya.

5.38 En un estudio realizado en la ciudad de México, con mujeres, se presentó la siguiente tabla de proporción de una actitud de aceptación hacia el aborto:

Escolaridad de las madres	Escolaridad de las hijas		
	Primaria	Bachillerato	Profesional
Primaria	0.486	0.699	0.011
Bachillerato	0.247	0.503	0.054
Profesional	0.068	0.484	0.448

a) ¿Cuál será la actitud de la siguiente generación?
b) ¿Cuál será la actitud de la quinta generación?

5.39 En una familia se observó, durante la primera generación, una enfermedad hereditaria (VIH). Suponga que cada síntoma puede ser considerado una unidad de tiempo, en una cadena de Markov con cuatro síntomas: sarcoma de Kaposi (S), cuadro gripal (G), neumocistitis (N), pérdida de peso (P).

A continuación se muestra la matriz de probabilidad de cada síntoma en la primera generación:

$$\begin{array}{c} \\ S \\ G \\ N \\ P \end{array} \begin{array}{c} \begin{array}{cccc} S & G & N & P \end{array} \\ \begin{bmatrix} 0.75 & 0.10 & 0.10 & 0.05 \\ 0.05 & 0.75 & 0.15 & 0.05 \\ 0.20 & 0.40 & 0.30 & 0.10 \\ 0.10 & 0.30 & 0.20 & 0.40 \end{bmatrix} \end{array}$$

Si comenzamos nuestras observaciones (primera generación) con sarcoma de Kaposi (S), calcule la probabilidad de que:

a) Aparezca cuadro gripal en la próxima generación.
b) El cuadro gripal aparezca en la segunda generación, y sarcoma de Kaposi en la tercera.
c) El cuadro gripal aparezca en la segunda o en la tercera generación, o en ambas.
d) El cuadro gripal aparezca primero en la tercera generación.

Distribuciones probabilísticas 6

PROPÓSITOS

Siguiendo con el análisis y la comprensión de los contenidos, así como con la solución de los ejercicios propuestos, el lector explicará en qué consisten las variables aleatorias (discretas y continuas).

También, explicará en qué consisten las distribuciones de probabilidad y desigualdad de Chebyshev.

Del mismo modo, el lector podrá:

1. Identificar cómo la desigualdad de Chebyshev pone de relieve el papel primordial de la desviación estándar como medida de dispersión.
2. Interpretar el gráfico de la desigualdad de Chebyshev.
3. Explicar la ley de los grandes números.
4. Comprender el proceso de la distribución de probabilidad más simple (Bernoulli).
5. Comprender el proceso de las distribuciones:
   a) binomial, *b*) Poisson, *c*) hipergeométrica y *d*) normal.
6. Identificar las características básicas de las distribuciones citadas.
7. Identificar las aplicaciones para cada distribución.
8. Reconocer los casos en los que puede ocurrir o no un evento o ensayo.
9. Explicar y calcular para cada distribución:
   a) media aritmética, *b*) varianza y *c*) desviación estándar.
10. Explicar en qué consiste la distribución normal estándar.
11. Desarrollar el proceso de cálculo de la distribución normal estándar.

12. Desarrollar verbalmente en qué consiste la aproximación de la binomial a la normal.
13. Desarrollar el proceso de cálculo de la aproximación binomial a la normal.
14. Identificar el uso de la aproximación citada en el inciso anterior.
15. Explicar el proceso de corrección de continuidad.
16. Desarrollar el cálculo del proceso anterior.
17. Identificar problemas y ejercicios aplicando los procedimientos de las distribuciones siguientes:
    a) binomial, *b)* Bernoulli, *c)* Poisson, *d)* hipergeométrica y *e)* normal.

INTRODUCCIÓN

Considerando que el valor asociado con un aspecto particular de cada resultado en el espacio muestral varía y depende del resultado observado, una variable cuyo valor está determinado por el resultado de un experimento u observación es una variable aleatoria. Los valores que toma una variable aleatoria se asocian con eventos aleatorios, en el espacio muestral de un experimento realizado.

Existen dos clases de variables aleatorias:

1. Discretas
2. Continuas

En muchos problemas es necesario determinar las probabilidades de que una variable aleatoria tome valores específicos en su rango de valores posibles. Con ese fin, resulta útil conocer todos los valores posibles de una observación o experimento, así como de las correspondientes frecuencias relativas de una variable aleatoria y sus probabilidades respectivas. Dicho resumen se llama *distribución de probabilidad.* Ésta se considera como la expresión de una función aleatoria (discreta o continua) que tiene como dominio un conjunto de eventos mutuamente excluyentes y completamente exhaustivos.

6.1 VARIABLE ALEATORIA

Para construir las distribuciones probabilísticas, que son punto de interacción entre la probabilidad y la estadística, es necesario entender el concepto de variable aleatoria.

Definición Una variable aleatoria es una regla bien definida para asignar valores numéricos a todos los resultados probables de un experimento.

Los distintos resultados que se obtienen al lanzar un dado forman una variable, que abarcará de 1 a 6, y que no son predecibles. A todo valor de una variable aleatoria corresponderá una probabilidad. Ese conjunto de todos

los valores posibles de la variable aleatoria y de sus respectivas probabilidades recibe el nombre de *función* o *distribución de probabilidad* de la variable aleatoria.

Cualquier función de una variable aleatoria también será de resultados impredecibles y, por tanto, también aleatoria. En general, cualquier operación algebraica entre dichas funciones dará lugar a otra variable aleatoria.

6.2 DESIGUALDAD DE CHEBYSHEV

Esta desigualdad muestra que, hasta cierto punto, la varianza controla cuán lejos se extienden las probabilidades.

Teorema Sea X una variable aleatoria con una función (densidad)[1] de probabilidad $f(X)$ de manera tal que tanto $E(X) = \mu_X$ como $\text{var}(X) = \sigma_X^2$ tienen un valor finito. Entonces:

$$P(\,|X - \mu_X\,| \leq k\sigma_X) \geq 1 - \frac{1}{k^2}$$

o

$$P(\,|\,X - \mu_X\,| > k\sigma_X) \leq \frac{1}{k^2}$$

para cualquier constante $k \geq 1$.

La demostración de Chebyshev es muy importante, ya que permite determinar los límites de las probabilidades de variables aleatorias discretas o continuas sin tener que especificar sus funciones (densidad) de probabilidad. Este teorema calcula que la probabilidad de que una variable aleatoria no se aleje más allá de k desviaciones estándar de la media.

Por ejemplo, Si $k = 2$

$$P(|\,X - \mu_X\,| \leq 2\,\sigma_X) \geq 1 - \frac{1}{4}$$

[1] Algunos autores llaman a la función de probabilidad, *función de densidad de probabilidad* o *masa de probabilidad*.

Para cualquier variable aleatoria X con media μ_X y varianza σ_X^2 finita, la desigualdad de Chebyshev en forma gráfica se puede observar de la siguiente forma:

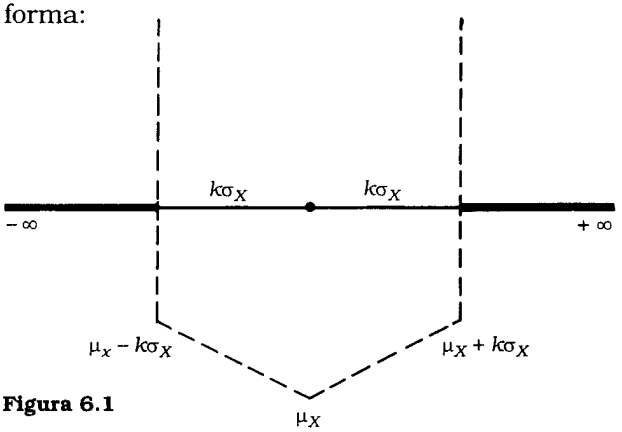


Figura 6.1

Intervalo de la variable en el cual se cumple:

$$|X - \mu_X| > k\sigma_X \qquad \text{... (I)}$$

De la expresión I despejamos X aplicando la definición de valor absoluto:

$$|X - \mu| > k\sigma_X \quad \begin{cases} X - \mu_X > k\sigma_X & \text{para } (X - \mu_X) > 0 \\ X > \mu_X + k\sigma_X & \\ -(X - \mu_X) > k\sigma_X & \text{para } (X - \mu_X) < 0 \\ -X + \mu_X < k\sigma_X & \\ X - \mu_X < -k\sigma_X & \\ X < \mu_X - k\sigma_X & \end{cases}$$

En suma, se tiene:

$$|x - \mu_X| > k\sigma_X \quad \begin{cases} X > \mu_X + k\sigma_X & \text{para } (X - \mu_X) > 0 \\ X < \mu_X - k\sigma_X & \text{para } (X - \mu_X) < 0 \end{cases} \qquad \text{... (II)}$$

En la figura 6.1 se establece el valor de μ_X y dos puntos equidistantes, como $\mu_X - k\sigma_X$ y $\mu_X + k\sigma_X$ (que se obtiene de la expresión II). Quedaron así delimitados tres intervalos de variación de la variable aleatoria:

$$(-\infty, \mu_X - k\sigma_X)$$
$$[\mu_X - k\sigma_X, \mu_X + k\sigma_X]$$
$$(\mu_X + k\sigma_X, +\infty)$$

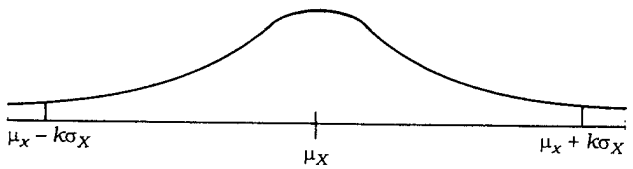
En estas condiciones, calculamos la varianza de la variable X, de la función de densidad $f(X)$.

Teorema de Chebyshev. Si μ_X y σ_X son, respectivamente, la media y la desviación estándar de la variable aleatoria X, entonces, para una constante positiva k cualquiera, la probabilidad es cuando menos $[1 - \frac{1}{k^2}]$ de que X tomará un valor contenido en k desviaciones estándar de la media; en forma simbólica se tiene:

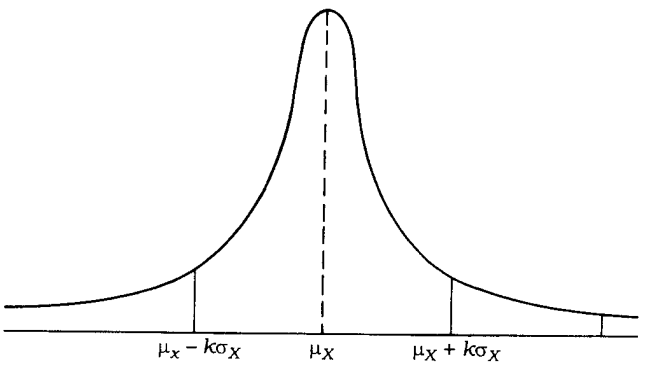
$$P(\mid X - \mu_X \mid < k\sigma_X) \geq 1 - \frac{1}{k^2}$$

La importancia de este teorema reside en que pone en relieve el papel primordial de la desviación estándar como medida de dispersión. Efectivamente:

CASO 1



CASO 2



Para un valor de k determinado, si σ_X es grande, la masa de probabilidad existente entre los extremos $\mu_X - k\sigma_X$ y $\mu_X + k\sigma_X$ se repartirá con gran dispersión alrededor del punto μ_X, tal como lo marca el *caso 1*; si, en cambio, el valor de σ_X es pequeño para el mismo valor de k, la misma masa de probabilidad habrá de repartirse alrededor del punto μ_X con mayor concentración, como se observa en el *caso 2*. Así, se obtiene que a mayor σ_X, mayor dispersión o menor concentración y, por el contrario, a menor σ_X, menor dispersión o mayor concentración.

6.3 LEY DE LOS GRANDES NÚMEROS

La ley de los grandes números, así como el teorema central del límite, se relacionan con la noción de la suma de variables aleatorias. Se sabe que si

$$X_i \text{ para toda } i = 1, \ldots, n$$

son variables aleatorias, la suma

$$S_n = X_1 + X_2 + \ldots + X_n$$

es también una variable aleatoria. La ley de los grandes números afirma que el promedio de un número de variables aleatorias distribuidas en forma idéntica e independientes converge hacia un valor esperado de la distribución cuando aumenta el número de las variables aleatorias.

Teorema Supóngase que X_1, X_2, ... X_K ... es una secuencia de variables aleatorias independientes distribuidas idénticamente, cada una con media μ_X y varianza σ_X^2. Defina la nueva secuencia de los valores X_i mediante

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad n = 1, 2, 3, \ldots$$

Por tanto, $\overline{X}_1 = X_1$, $\overline{X}_2 = \frac{1}{2}(X_1 + X_2)$, $\overline{X}_3 = \frac{1}{3}(X_1 + X_2 + X_3)$ ...

Entonces, $\lim_{n \to \infty} P(|\overline{X}_n - \mu_X| > \epsilon) = 0$ para cualquier ϵ* > 0

Esto significa que al aumentar el tamaño de muestra ($n \to \infty$), $\overline{X}_n - \mu_X$, tiende a cero o sea $\overline{X}_n$ tiende a μ_X.

6.4 DISTRIBUCIÓN BERNOULLI

Es la distribución de probabilidad más simple, con solamente dos resultados posibles: éxito o fracaso.

Definición Si se considera una secuencia o serie de experimentos en la que cada experimento es el mismo y se realiza en las mismas condiciones, entonces la función de probabilidad es la misma para cada uno de ellos, siempre que se realicen en forma independiente. A esta serie o secuencia se le conoce como *ensayos repetidos* (Bernoulli).

Una de las aplicaciones más importantes que presenta en la actualidad esta distribución es el control de calidad. La manufactura de un producto determinado puede considerarse como resultado de ensayos repetidos, al final de los cuales se cuenta con un producto excelente, si pasa ciertas especificaciones, o defectuoso, si no las pasa. Así, en este proceso de producción existen sólo dos resultados posibles: defectuoso o no defectuoso.

En general, un experimento o ensayo en el que sólo ocurra uno de los dos eventos posibles se llama *ensayo de Bernoulli*, y a un experimento consistente en una serie de ensayos independientes, idénticos y dicotómicos se le conoce como *proceso de Bernoulli*. Por lo general, a uno de los eventos o resultado de dicho proceso se le denomina *éxito*, y al otro *fracaso*.

* Donde ϵ es un número positivo suficientemente pequeño.

A la probabilidad de que ocurra éxito se le conoce como p y al no éxito o fracaso como q, con lo que se cumple la siguiente expresión:

$p + q = 1$ o $p = 1 - q$ y $q = 1 - p$ donde $p = P(E)$ y $q = P(F)$.

De ensayo a ensayo, tanto p como q permanecen constantes; a esta característica se le conoce como *estacionario*.

En resumen, todo proceso que se caracteriza como *dicotómico*, *independiente* y *estacionario* es un proceso de Bernoulli.

Si llamamos E = éxito y F = fracaso, entonces $P(E) = P$ y $P(F) = q$, la notación para cada uno de los siguientes casos:

i) $P(E, E) = P \cdot P = P^2$
ii) $P(E, F) = P \cdot q = q \cdot P = P(f, e)$, donde el orden no importa
iii) $P(E, E, F) = P^2 q$
iv) $P(E, E, E, E) = P^4$

En general

v) $\underbrace{P(E, E, E, \ldots E)}_{X \text{ veces}} = P^X$ donde el número de éxitos es x

La probabilidad de una secuencia dada de n ensayos independientes de Bernoulli depende sólo del número de éxitos (X) y de su probabilidad de éxito p.

Como x es el número de éxitos y $n - X$ es el de fracasos, entonces la expresión anterior queda $p^X q^{n-X}$.

En el caso de pruebas repetidas independientes y si cada experimento es idéntico a los demás, la determinación de los correspondientes espacios de resultados y de sus probabilidades se obtiene sin saber qué ocurrió con los demás experimentos. Esto es posible si se efectúa un muestreo al azar y con reemplazo (reposición); entonces se trata de un *modelo de pruebas de Bernoulli*, para lo cual es necesario codificar los dos posibles resultados de la variable Bernoulli en 1 y 0, donde X (Éxito) = 1 y X (Fracaso) = 0. Entonces, su función de probabilidad se define con el modelo siguiente:

X	$f(X)$	$Xf(X)$	$X - \mu$	$(X - \mu)^2$	$(X - \mu)^2 f(X)$
1	p	p	$1 - p$	q^2	$q^2 P$
0	q	0	$0 - p$	P^2	$P^2 q$
		$\mu x = p$			

$$q^2 P + P^2 q = (1 - P)^2 P + P^2 (1 - P)$$
$$= (1^2 - 2p + P^2) P + P^2 (1 - P)$$
$$= P - 2P^2 + P^3 + P^2 - P^3$$
$$= P - P^2 = P(1 - P) = Pq = \sigma_x^2$$

y los parámetros (características) o momentos que describen la distribución Bernoulli son:

$$\mu_X = p \text{ (media aritmética)}$$

$$\sigma_X^2 = \text{var } X = pq \text{ (varianza) y}$$

$$\sigma_X = \sqrt{p\,q} \text{ (desviación estándar)}$$

Ejemplo 1 Por las características de su producción, 10% de los artículos fabricados son defectuosos; al inspeccionar seis de ellos, dos tienen defectos. La probabilidad de esta secuencia particular es la siguiente:

$$p = 0.10$$

$$q = 0.90$$

$$\binom{6}{2} P^2 q^4 = \frac{6\,!}{2\,!\,4\,!} \quad (0.10)^2\,(0.90)^4 = 15\,(0.01)\,(0.656) = 0.0984$$

$$\text{y como} \quad \mu_X = P \quad \text{entonces;} \quad \mu_X = 0.1$$

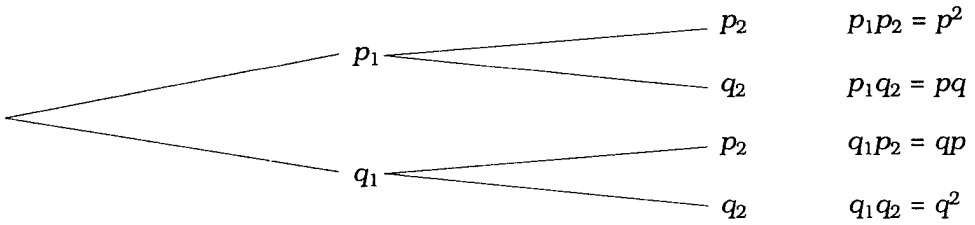
$$\sigma_X^2 = pq = 0.10 \times 0.90 = 0.09$$

$$\sigma_X = \sqrt{pq} = \sqrt{0.09} = 0.30$$

6.5 DISTRIBUCIÓN BINOMIAL

La distribución binomial produce la descripción adecuada a las probabilidades de ocurrencia de los resultados posibles a un experimento, siempre que sean generados por un proceso Bernoulli. Aunque un experimento tenga muchos resultados probables, por ejemplo, el lance de un dado tiene seis resultados posibles, quizá satisfaga esa primera condición del proceso de Bernoulli si se dicotomizan los resultados: E = obtener un 6, F = no obtener un 6, o también E = número par, F = número impar, etcétera.

Ejemplo 2 Si los resultados posibles de un experimento que se realiza dos veces se grafican por medio de un diagrama de árbol, se tiene lo siguiente:



si $p = p_1 = p_2 = 0.5$ entonces $p = 0.5 = \frac{1}{2} = 50\%$

$$q = q_1 = q_2 = (1 - p) = 1 - 0.5 \quad \text{entonces} \quad q = 0.5 = \frac{1}{2} = 50\%$$

$$p + q = 1 = 100\%$$

Entonces:

$$p \cdot p = (0.5)\,(0.5) = 0.25 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 25\%$$

$$pq = (0.5)\,(0.5) = 0.25 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 25\%$$

$$qp = (0.5)\,(0.5) = 0.25 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 25\%$$

$$q \cdot q = (0.5)\,(0.5) = 0.25 = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 25\%$$

$$1.0 \quad = \quad 1 \quad = \quad 100\%$$

Pero como $pq = qp$ (el orden no importa),

entonces $p_1q_2 + q_1p_2 = 25\% + 25\% = 50\% = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

En relación con el diagrama de árbol anterior, como son eventos independientes (basado en el postulado de Bernoulli) y por la definición de eventos independientes (véase el cap. 5):

$$P\,(2 \text{ éxitos}) = P\,(E_1E_2) \;=\; P\,(E_1)\,P\,(E_2 \mid E_1) \;=\; (0.5)\,(0.5) \;=\; 0.25$$

$$P\,(1^{\text{o}} \text{ éxito } 2^{\text{o}} \text{ fracaso}) = P\,(E_1F_2) \;=\; P\,(E_1)\,P\,(F_2 \mid E_1) \;=\; (0.5)\,(0.5) \;=\; 0.25$$

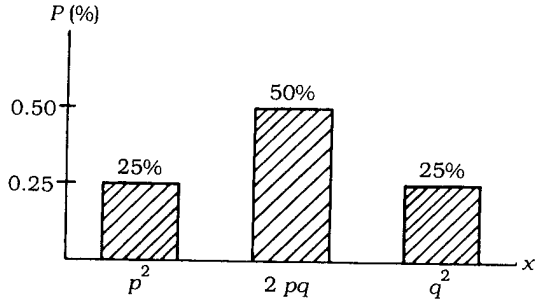
$$P\,(1^{\text{o}} \text{ fracaso } 2^{\text{o}} \text{ éxito}) = P\,(F_1E_2) \;=\; P\,(F_1)\,P\,(E_2 \mid F_1) = (0.5)\,(0.5) \;=\; 0.25$$

$$P\,(0 \text{ éxitos})\,P\,(2 \text{ fracasos}) = P\,(F_1F_2) \;=\; P\,(F_1)\,P\,(F_2 \mid F_1) \;=\; (0.5)\,(0.5) \;=\; 0.25$$

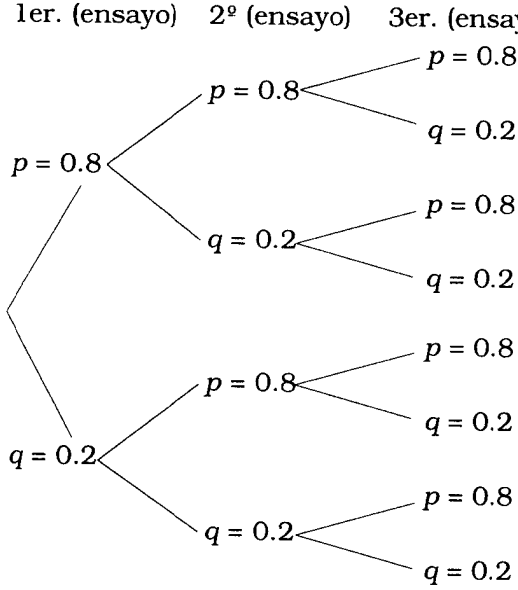
y como P (1 éxito y 1 fracaso) = P (1º éxito y 2º fracaso) + P (1º fracaso y 2º éxito)

para $n = 2$ $\quad p = q = 0.5$
P (2 éxitos) = 0.25
P (1 éxito) = 0.50
P (0 éxitos) = 0.25
1

Estos resultados se pueden mostrar en una gráfica de barras.



Ejemplo 3 Si el experimento se ejecuta tres veces y se supone que $p = 0.8 = 80\%$, entonces $q = 0.2 = 20\%$; el diagrama de árbol queda estructurado como se muestra a continuación:



Resultados	Probabilidad
$ppp = p^3$	$= (0.8)(0.8)(0.8) = 0.512$
$ppq = p^2q$	$= (0.8)(0.8)(0.2) = 0.128$
$pqp = p^2q$	$= (0.8)(0.2)(0.8) = 0.128$
$pqq = pq^2$	$= (0.8)(0.2)(0.2) = 0.032$
$qpp = qp^2$	$= (0.2)(0.8)(0.8) = 0.128$
$qpq = q^2p$	$= (0.2)(0.8)(0.2) = 0.032$
$qqp = q^2p$	$= (0.2)(0.2)(0.8) = 0.032$
$qqq = q^3$	$= (0.2)(0.2)(0.2) = 0.008$
	1.00

En resumen, para el caso $n = 3$:

$P = 0.8$ y $q = 0.2$

P (3 éxitos) = $= 0.512$

P (2 éxitos) = $0.128 + 0.128 + 0.128 = 0.384$

P (1 éxito) = $0.032 + 0.032 + 0.032 = 0.096$

P (0 éxitos) = $= \underline{0.008}$

1.000

La característica común de la distribución binomial es que puede generarse por medio de la expresión matemática conocida como el *teorema del binomio*

(cap. 4) $(p + q)^n$, donde p = probabilidad de éxito, q = probabilidad de fracaso y n = número de ensayos.

Para el ejemplo 2, se tiene:

$$(p + q)^2 = p^2 + 2\,pq + q^2 \text{ como } p = 0.5 \text{ y } q = 0.5 \qquad n = 2$$
$$= 0.25 + 0.50 + 0.25 = 1.0$$

Para el ejemplo 3, se tiene:

$$p = 0.8 \text{ y } q = 0.2,\ n = 3$$

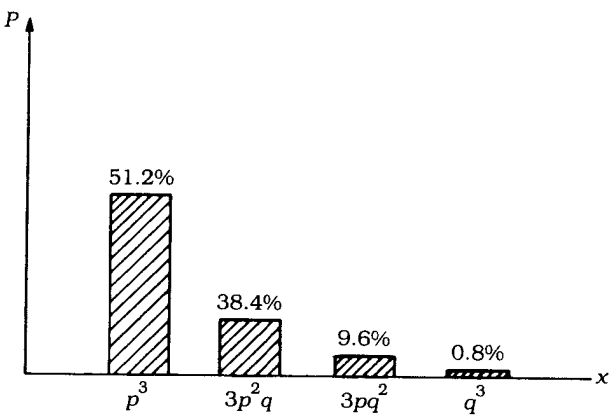
$$(p + q)^3 = p^3 + 3\,p^2q + 3\,pq^2 + q^3$$

$$0.512 + 3\,(0.128) + 3\,(0.032) + 0.008 = 1.0$$

$$51.2\% + 9.6\% + 38.4\% + 0.8\% = 100\%$$

Generalizando $(p + q)^n = \sum_{X=0}^{n} \binom{n}{X} p^X q^{n-X}$

entonces, P (X [éxitos] | n [ensayos Bernoulli], P [= P (éxito)]) = $\binom{n}{X} p^X q^{n-X}$
donde $X = 0, 1, \ldots, n$



El teorema del binomio elimina la necesidad de construir un diagrama de árbol para generar todos los resultados posibles de un experimento.

Algunas veces no se necesita obtener la distribución completa de los resultados de un experimento binomial, pero sí un resultado particular. Para calcularlo, se dispone de la siguiente fórmula:

$$P\,(X \mid n, p) = \binom{n}{X} p^X q^{n-X}$$

donde:

n = número de ensayos del experimento
p = probabilidad de éxito de cualquier ensayo

q = probabilidad de fracaso de cualquier ensayo
X = número de éxitos de los n ensayos
$P(X|n, P)$ = probabilidad de obtener X éxitos de n ensayos con una probabilidad constante de p.

$\binom{n}{X}$ significa dividir a un grupo de n [experimentos] en dos grupos, 1 con X [éxitos] elementos y el otro $(n - X)$ [fracasos] elementos (sección 4.4.).

Entonces, la fórmula queda como:

$$P(X \mid n, p) = \frac{n!}{X!\,(n - X)!}\, p^X q^{n-X}$$

Donde $X = 0, 1, 2, \ldots, n$

Ejemplo 4 Supóngase que al manufacturar un artículo, la probabilidad de que resulte defectuoso es de 20% = 0.2. Si en una línea de producción en un momento determinado se selecciona al azar una muestra de seis artículos, calcular la probabilidad de que cero, uno, dos, tres, cuatro, cinco o seis artículos estén defectuosos.

Solución:

Como el evento "salir defectuoso" se considerará un éxito, entonces:
$P = 20\% = 0.2$ y $q = 80\% = 0.8$

$n = 6$ y $X = 0, 1, 2, 3, 4, 5, 6$

$$P(0 \mid 6, 0.2) = \frac{6!}{0!\,(6-0)!}(0.2)^6\,(0.8)^0 = 1(0.000064)(1) = 0.000064 = 0.0064\,\%$$

$$P(1 \mid 6, 0.2) = \frac{6!}{1!\,(6-1)!}(0.2)^5\,(0.8)^1 = 6\,(0.00032)\,(0.8) = 0.001536 = 0.1536\,\%$$

$$P(2 \mid 6, 0.2) = \frac{6!}{2!\,(6-2)!}(0.2)^4\,(0.8)^2 = 15\,(0.0016)\,(0.64) = 0.015360 = 1.5360\,\%$$

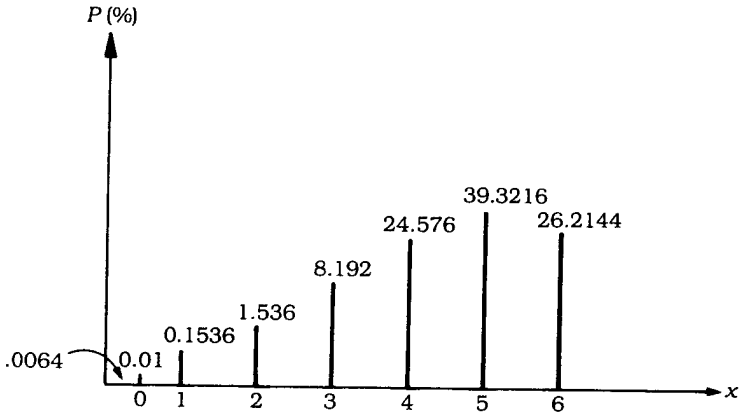
$$P(3 \mid 6, 0.2) = \frac{6!}{3!\,(6-3)!}(0.2)^3\,(0.8)^3 = 20\,(0.008)\,(0.512) = 0.081920 = 8.1920\,\%$$

$$P(4 \mid 6, 0.2) = \frac{6!}{4!\,(6-4)!}(0.2)^2\,(0.8)^4 = 15\,(0.04)\,(0.4096) = 0.245760 = 24.576\,\%$$

$$P(5 \mid 6, 0.2) = \frac{6!}{5!\,(6-5)!}(0.2)^1\,(0.8)^5 = 6\,(0.2)\,(0.32768) = 0.393216 = 39.3216\,\%$$

$$P(6 \mid 6, 0.2) = \frac{6!}{6!\,(6-6)!}(0.2)^0\,(0.8)^6 = 1\,(1)\,(0.262144) = \underline{0.262144} = \underline{26.2144\,\%}$$

$$1.00 \qquad 100\,\%$$



Algunas veces estos cálculos se tornan tediosos, pero se obtienen mediante una tabla de distribución binomial.

Características o parámetros de la distribución binomial

Media aritmética $\mu_X = np$

Varianza $\sigma_X^2 = npq$

Desviación estándar $\sigma_X = \sqrt{npq}$

Calcular μ_X, σ_X^2 y σ_X para el ejemplo anterior:

$$n = 6$$

$$p = 0.2$$

$$q = 0.8$$

$$\mu_X = (6)\ (0.2) = 1.2$$

$$\sigma_X^2 = (6)(0.2)(0.8) = 0.96$$

$$\sigma_X = \sqrt{(6)\ (0.2)\ (0.8)} = 0.9798$$

6.6 DISTRIBUCIÓN DE POISSON

Puede demostrarse que $\binom{n}{X} p^X q^{n-X} \to e^{-\lambda} \frac{\lambda^X}{X!}$

$n\,p \to \lambda$

$n \to \infty$

y se usa cuando n es grande, por ejemplo, $n > 10$

Esto significa que la ocurrencia al azar de muchos procesos reales está en función de una unidad temporal o espacial (longitud, superficie), que sigue un modelo exponencial con las siguientes consideraciones: los intervalos tanto espaciales como temporales deben ser tan pequeños como para que la probabilidad de dos o más eventos, en un intervalo dado, sea tan pequeña que se considere cero. Por consiguiente, el evento *acontece* o *no acontece* en el intervalo dado, sin afectar la probabilidad de que esto suceda en otro intervalo, por lo que se considerarán independientes.

Hasta aquí todo parece indicar que se asiste a un experimento binomial. Aunque en este modelo interesa determinar el número de éxitos de un número dado de ensayos, dichos eventos son observables, mientras que en el modelo de Poisson el número de éxitos es la importancia primordial, pero resulta virtualmente imposible de calcular.

Al determinar el total de ensayos (todas las oportunidades de que dicho evento acontezca) es posible percibir cuando el evento ocurrió, pero la dificultad prevalece cuando se quiere observar que el evento no ocurre.

Algunos ejemplos de aplicación del proceso de Poisson son el número de intercambios cromosómicos en una célula, la demanda de servicios en una institución, el número de arribos o salidas, o ambos, en un aeropuerto, la cantidad de meteoritos que caen en una hectárea en el desierto o en el mar, el número de autos que pasan por una caseta de cobro en un periodo, así como el número de fallas en un equipo o sistema.

El conocimiento de estas variables es de mucha utilidad, ya que mediante éstas se considera, por ejemplo, cuántas casetas de cobro o teléfonos públicos deben instalarse en un lugar establecido o el número de empleados a contratar en un tiempo determinado. Estos temas se estudian con más profundidad en una disciplina llamada *teoría de colas*.

A medida que n se incrementa y la longitud del intervalo t (temporal o espacial) disminuye, la probabilidad P de un acontecimiento en cualquier intervalo o ensayo (instancia) también disminuye.

Al respecto, hay que considerar que la subdivisión del intervalo en t subintervalos no tiene efecto alguno en el número de acontecimientos esperado en dicho periodo t.

Entonces, el término no permanece constante. El modelo de la distribución de Poisson se establece de la siguiente manera:

$$P(X) = \frac{e^{-np}(np)^X}{X!}, \quad X = 0, 1, 2, \ldots$$

donde:

$P(X)$ = probabilidad total
X = número de éxitos, acontecimientos, ocurrencias
e = 2.71828 (constante, base de los logaritmos naturales)
p = probabilidad de ocurrencia

Como np es constante, es común llamarla λ o sea $\lambda = np$. Entonces, la fórmula anterior queda:

$$P(X) = \frac{e^{-\lambda}\lambda^X}{X!}$$

donde $X = 1, 2, 3, \ldots$

Características o parámetros de la distribución de Poisson

Como $X \geq 0$ (no puede asumir valores negativos) y por la distribución binomial $\mu_X = np$,

entonces $\mu_X = \lambda$
$\sigma_X^2 = \lambda$
$\sigma_X = \sqrt{\lambda}$

Ejemplo 5 En una oficina de la Procuraduría Federal de Protección al Consumidor cuentan con un conmutador de gran capacidad. Diariamente se reciben muchas llamadas, 2% de las cuales (probabilidad obtenida por experiencias anteriores) se deben a problemas relacionados con inmobiliarias. Calcular la probabilidad de que en una hora determinada (12-13 h), en la que se captan 300 llamadas: *a)* 10 de ellas se refieran a problemas inmobiliarios; *b)* más de 10 y menos de 19 correspondan a este tipo de situaciones.

Solución:

a) $n = 300$
$p = 2\% = 0.02$
$X = 10$
$\lambda = np = 300\,(0.02) = 6$

$$P(X) = \frac{e^{-\lambda}\lambda^X}{X!}$$

$$P(10) = \frac{e^{-6}\,6^{10}}{10!} = 0.04132$$

$$P(10) = 0.04132 = 4.132\%$$

b) aplicando sucesivamente para $x = 10, 11, \ldots, 18$ se tiene:

$P(10) = 0.04132$	$P(15) = 0.0009$
$P(11) = 0.02254$	$P(16) = 0.0003$
$P(12) = 0.0113$	$P(17) = 0.0001$
$P(13) = 0.0052$	$P(18) = 0.00004$
$P(14) = 0.0022$	0.0838

$P(10 \leq x < 19) = 0.0838$

$\mu_X = \lambda$ significa $\mu_X = 6$

$\lambda = \sigma_X^2$ significa $\sigma_X^2 = 6$

$\sigma_X = 2.44949$

6.6.1 Aproximación de la distribución binomial a la de Poisson

La distribución de Poisson es una buena aproximación de la binomial, si n es grande y p cercana a cero; en forma práctica, si $\frac{n}{p} > 500$ la aproximación es excelente, ya que la distribución de Poisson constituye el caso límite de la binomial.

Ejemplo 6 Si se considera el ejemplo anterior:

$$n = 300$$

$$p = 0.02$$

$$X = 10$$

Al aplicar el modelo binomial resulta:

$$P(10 \mid 300, 0.02) = {}_{300}C_{10}\,(0.02)^{10}\,(0.98)^{290}$$

$$= \frac{300\,!}{10\,!\,290\,!} \times 1.02400 \times 10^{-17} \times 0.00285$$

$$= 0.04132$$

como $$\frac{n}{p} = \frac{300}{0.02} = 15\,000 >> 500$$

se usa Poisson

$$P(10) = \frac{e^{-6}\,6^{10}}{10\,!} = 0.04132$$

6.7 DISTRIBUCIÓN HIPERGEOMÉTRICA

La distribución binomial se basa en los supuestos de que las observaciones son independientes entre sí y de que es un proceso estacionario, por lo que la probabilidad de éxito es constante. Esto resulta válido cuando se realiza un muestreo con reemplazo o cuando la población es infinita; sin embargo, cuando la población es finita o se realiza un muestreo sin reemplazo, los ensayos sucesivos no son idénticos ni independientes y, por supuesto, la probabilidad de éxito varía de ensayo a ensayo.

En la mayor parte de los problemas de control de calidad se realiza siempre muestreo sin reemplazo y se utiliza de manera amplia la distribución hipergeométrica.

Como ${}_nC_r = \binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$, entonces el modelo hipergeométrico será:

$$P(r_1, r_2, r_3, \ldots, rk) = \frac{\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_k}{r_k}}{\binom{n}{r}} \quad \text{donde} \quad \begin{cases} \sum_{i=1}^{n} n_i = n \\ \sum_{i=1}^{k} r_i = r \end{cases}$$

donde $r = 0, 1, 2, \ldots, n$

Ejemplo 7 En un conjunto de 10 canicas, ocho son rojas y dos verdes. De una muestra de cinco canicas, calcúlese la probabilidad de que tres de ellas sean rojas.

$n = 10$, $n_1 = 8$ rojas, $n_2 = 2$ verdes
$r = 5$, $r_1 = 3$ rojas, $r_2 = 2$ verdes

$$P(r_1, r_2) = \frac{\binom{8}{3}\binom{2}{2}}{\binom{10}{5}} = \frac{\dfrac{8!}{3!\,5!} \times \dfrac{2!}{2!\,0!}}{\dfrac{10!}{3!\,5!}} = 0.22222$$

Los parámetros o características de la distribución hipergeométrica son:

$$\mu_X = np$$

$$\sigma_X^2 = \frac{npq\binom{n}{r}}{n-1}$$

$$\sigma_X = \sqrt{\frac{npq\binom{n}{r}}{n-1}}$$

Calcular μ_X, σ_X^2 y σ_X $\quad \mu_X = np = 10 \times 0.22222 = 2.2222$

$$\sigma_X^2 = \frac{10 \times 0.22222 \times 0.77778 \times 252}{9} = \frac{435.55}{9}$$

$$\boxed{\sigma_X^2 = 48.39472}$$

$$\boxed{\sigma_X = 6.95663}$$

6.8 DISTRIBUCIÓN NORMAL

La distribución normal es sin duda la más conocida y usada de todas. Muchos fenómenos naturales tienden a dar como resultado una distribución normal. Entre otras, longitud, altura y grosor de animales o plantas; mediciones de cantidades de azúcar en la sangre; cantidad de glóbulos blancos; incidencia de las enfermedades del oído interno; medidas en el aspecto conductista, emocional o psicológico de las acciones, aptitudes o capacidades humanas. La distribución de los errores de medida (desviaciones en relación con un valor específico en los diámetros de pistones, cilindros o cañones de armas de fuego; pesos de productos empaquetados, e incluso las longitudes de las cintas métricas) tiende a ser normal, al igual que la distribución del grado de perfección de diversos procesos de producción.

Debido a que la distribución normal describe de manera satisfactoria muchos fenómenos naturales, se ha convertido en un patrón de referencia para muchos problemas probabilísticos. Además, en ciertas condiciones, las distribuciones binomial y de Poisson se aproximan mediante la distribución normal.

En el capítulo 2 se agruparon datos y se dibujó una gráfica en forma de campana; en el capítulo 3 se describió la función y se encontraron algunos intervalos. Ahora se estudiará la función como distribución de probabilidad con sus correspondientes propiedades.

La distribución normal es continua, en la que la variable aleatoria X es capaz de adoptar cualquier valor comprendido en el siguiente intervalo (*véase*, el cap. 3).

$$\{-\infty < X < \infty\}$$

Dos parámetros describen a la distribución normal:

$$\mu_X = \text{media}$$

$$\sigma_X^2 = \text{varianza}$$

y se denota como:

$$(\mu_X, \sigma_X^2)$$

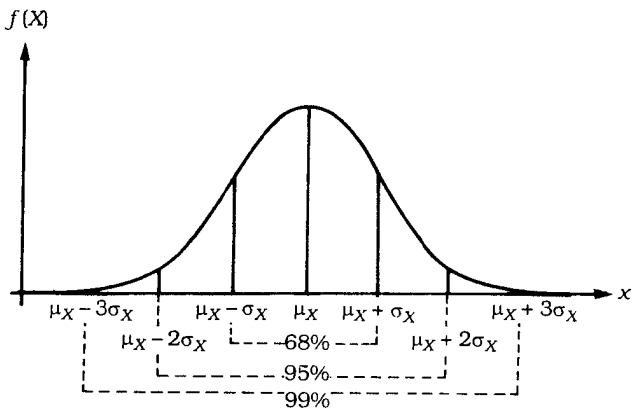
La función de densidad normal (μ_X, σ_X^2) es:

$$f(X) = \frac{1}{\sqrt{2\pi\sigma_X}} e^{-\frac{1}{2}\left[\frac{X-\mu_X}{\sigma_X}\right]^2} \quad \text{donde} \quad x \in \square \qquad \ldots(I)$$

donde

$$\mu_X = E(X) \text{ tal que } -\infty < \mu < \infty$$
$$\text{var}(X) = \sigma_X^2, \quad \sigma_X^2 > 0$$
$$e = 2.71828, \quad \pi = 3.14159$$

La gráfica de la expresión I es la siguiente:

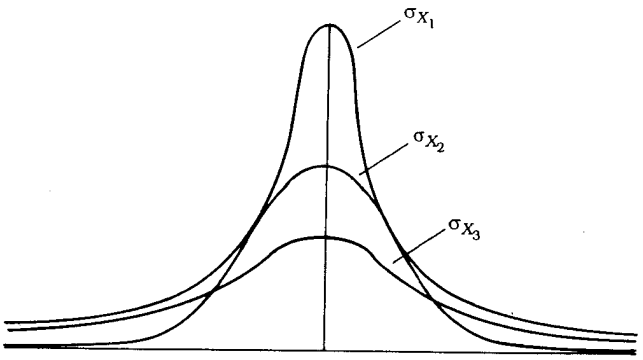


El área bajo la curva se llama *masa de probabilidad* o simplemente *probabilidad*. Ésta se calculará por medio de tablas.

Se retomará el concepto de la desigualdad de Chebyshev.

CASO 1

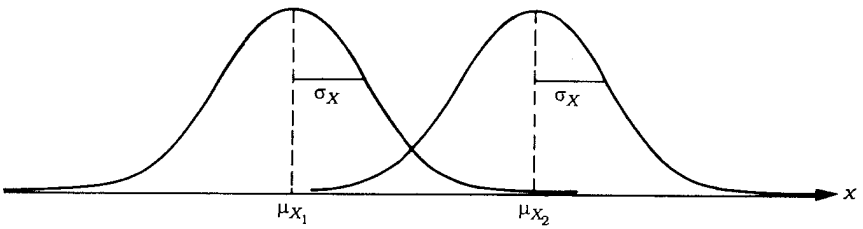
Funciones de densidad de tres variables normales con la misma media y diferentes desviaciones estándar ($\sigma_1 < \sigma_2 < \sigma_3$).



Como se comentó al considerar la desigualdad de Chebyshev, cuanto mayor sea el valor de la desviación estándar, existe mayor dispersión de los datos. Ahora sucede lo mismo, con la única diferencia de que existe una función de densidad en particular.

CASO 2

Funciones de densidad de dos variables normales con diferente media y la misma varianza, por ejemplo $(\mu_1 < \mu_2)$.



Por la forma de la distribución se observan dos propiedades:

a) La distribución normal es simétrica alrededor de μ_X.

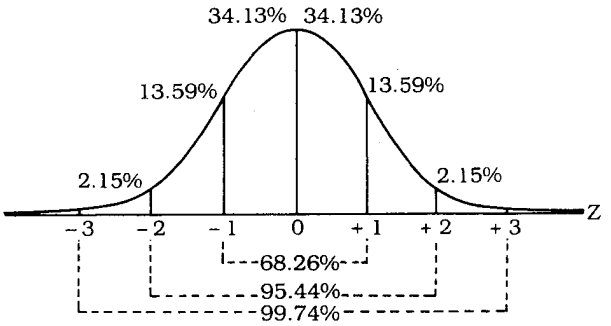
b) La media, moda y mediana coinciden por la forma simétrica.

6.8.1 Distribución *Z*

Si una variable X (puntuaciones, datos, calificaciones, etc.) se halla normalmente distribuida, entonces las estadísticas tipificadas o estandarizadas estarán definidas por:

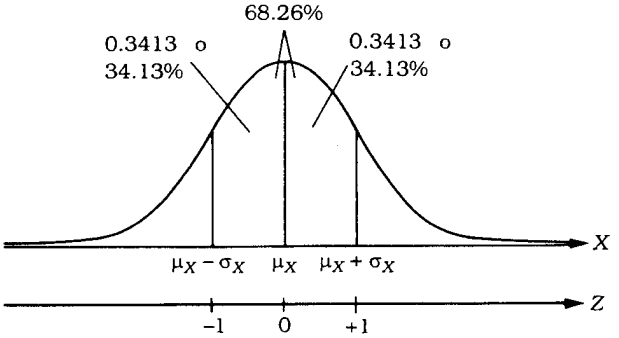
$$Z = \frac{X - \mu_X}{\sigma_X}$$

Esta nueva variable estará normalmente distribuida con media 0 y varianza 1; como se ve, Z es en realidad el número de desviaciones estándar en que se encuentra la puntuación X respecto a la media aritmética, la cual se presenta en la figura siguiente:



Para conocer el área bajo la curva que se encuentra entre la media aritmética (μ_X) y una desviación estándar ($\pm\sigma_X$), se ubica en la figura 6.13 el valor asociado con $Z = 1$ (1 desviación estándar arriba o abajo de la media aritmética), y se obtiene el valor de 34.13%, o 0.3413 (*véase* figura 6.13).

Por lo general el valor de 34.13% se denota por 0.3413 cuando representa la probabilidad de encontrar una puntuación entre la media (μ_X) y la primera desviación estándar ($\pm\sigma_X$); el total (o suma) de probabilidades bajo la curva es igual, mientras que cuando se denota el porcentaje constituye la proporción de área bajo la curva, y su total será 100 por ciento.



Ejemplo 8 A un grupo de niños se les aplica una prueba de inteligencia (WISC); supóngase que las puntuaciones se distribuyen en forma normal y tienen los parámetros siguientes: $\mu_X = 100$ y $\sigma_X = 15$.

En términos de CI[2], utilizando esta información se resolverán los siguientes problemas:

1. Si la proporción normal de inteligencia en términos de CI está entre 90 y 110, ¿qué porcentaje de niños están en este intervalo; $P(90 \leq X \leq 110)$?

Paso 1. Se calculan los valores de Z para 90 y 110, o sea:

$$X_1 = 90 \text{ y } X_2 = 110$$

Puntuación Z para 90:

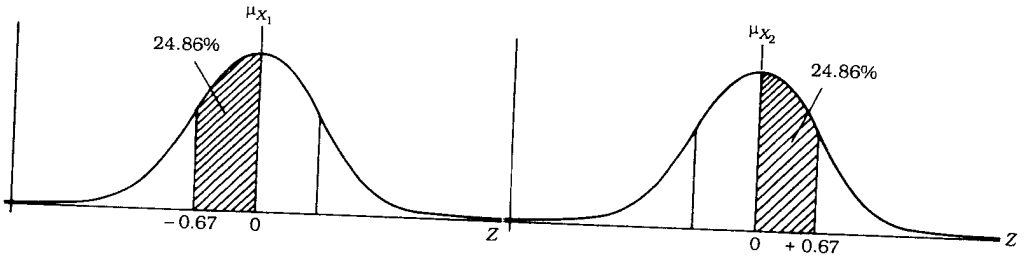
$$Z = \frac{X_1 - \mu_X}{\sigma_X} = \frac{90 - 100}{15} = \frac{-10}{15} = -0.67$$

Puntuación Z para 110:

$$Z = \frac{X_2 - \mu_X}{\sigma_X} = \frac{110 - 100}{15} = \frac{10}{15} = 0.67$$

Donde el signo menos únicamente indica que la puntuación Z se encuentra a la izquierda de la media aritmética.

Paso 2. Se determina el porcentaje entre la media y cada una de las puntuaciones Z obtenidas anteriormente, utilizando la tabla Z se obtiene que el área bajo la curva entre la media y la puntuación $Z(-0.67)$ es 24.86%; y para $Z(0.67)$ es 24.86%, lo que da un total de 49.72 por ciento.



Será 49.72% el porcentaje esperado de niños que presentan un CI normal (entre 90 y 110).

---

[2] $CI = \dfrac{\text{Edad mental}}{\text{Edad cronológica}} \times 100$

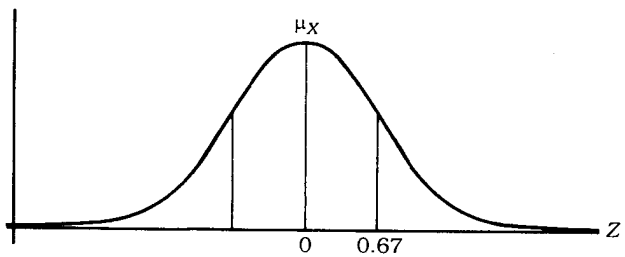
El concepto de normalidad en las ciencias de la salud es un concepto netamente estadístico y se obtiene utilizando la distribución normal.

2. ¿Qué porcentaje de niños obtuvieron un CI mayor que 110?, es decir, $P(X \geq 110)$?

Paso 1. Se calculan los valores Z de 110.
Puntuación Z para 110:

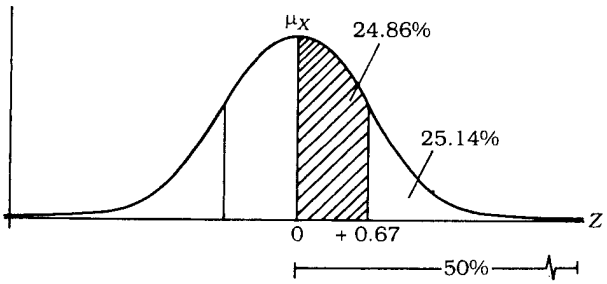
$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{110 - 100}{15} = \frac{10}{15} = 0.67$$

Paso 2. Se ubica en una gráfica el valor de Z obtenido:



Paso 3. Se localiza en la tabla Z el valor del área que se habrá de encontrar.

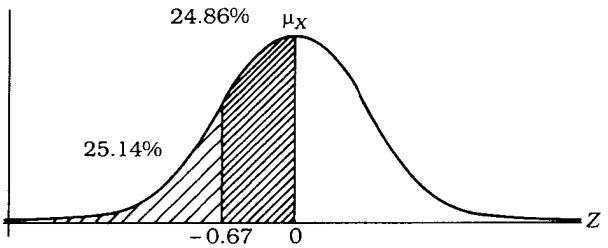
Como el área que se requiere es la del extremo derecho y el valor anterior, 24.86% es el área de la media a 110; y dado que dicha área es de 50%, se resta a 50% el valor de 24.86% y éste será el porcentaje del área buscada: 50 − 24.86 = 25.14 por ciento.



Esta área en la gráfica equivalente a 25.14% significa el porcentaje de niños superior al término medio.

3. ¿Qué porcentaje de niños tienen un CI menor que el término medio o menos de 90; $P(X \leq 90)$?

$$Z = \frac{90 - 100}{15} = \frac{-10}{15} = -0.67$$



$$P(X \leq 90) = 25.14\%$$

6.8.2 Aproximación de la distribución binomial a la normal

Se considera una muestra aleatoria (al azar) de una población finita y el atributo o la característica observada deberá ser dicotómico, es decir, éxito o fracaso, a favor, en contra, etcétera.

Este resultado es una consecuencia del teorema central del límite, que se explicará con detalle en el capítulo 7.

Teorema Sea X una variable aleatoria binomial con media np y desviación estándar $\sqrt{np(1-p)}$. La distribución de la variable aleatoria tiende a la normal estándar cuando el número de ensayos independientes n se repite un número muy grande de veces ($n \to \infty$):

$$Z_n = \frac{X - np}{\sqrt{np(1-p)}}$$

Ejemplo 9 Una empresa periodística llevó a cabo una encuesta entre 400 personas seleccionadas de manera aleatoria en un estado acerca del control de armas. De las 400 personas, 220 se pronunciaron a favor de un control estricto.

¿Qué tan probable resulta el hecho de tener 220 o más personas a favor del control de armas, si la población en este estado se encuentra dividida en opinión de igual manera?

Las características del problema llevarían a una distribución binomial. Existe el atributo dicotómico a favor y en contra en cada ensayo; la encuesta se realizó a cada persona, por tanto, es discreto. Como puede observarse, el número de encuestados es 400; si desea resolverse el problema por medio de la distribución binomial, quedaría de la siguiente forma:

$$P(X \geq 220) = 1 - \sum_{i=0}^{219} \binom{400}{i} p^i (1-p)^{400-i}$$

Debido a que resultaría muy laborioso realizar los cálculos correspondientes, se recomienda en estos casos aplicar la aproximación de la binomial a la normal:

sea $n = 400$

$p = 0.5$ $\qquad np = 400(0.5) = 200$

$q = 0.5$

El valor de $p = 0.5$ corresponde al enunciado que indica "La población en ese estado se encuentra dividida de igual manera". Sustituimos los valores:

$$P(X \geq 220) = 1 - P(X < 220)$$

$$= 1 - P\left(Z < \frac{220 - 200}{\sqrt{400\,(0.5)\,(0.5)}}\right)$$

$$= 1 - P\left(Z < \frac{20}{10}\right)$$

$$= 1 - P(Z < 2)$$

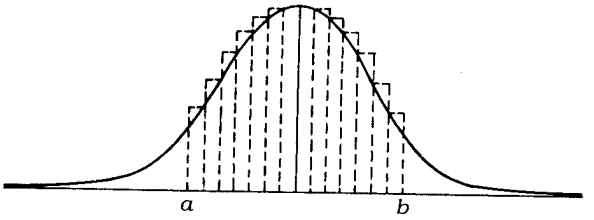
Se busca el valor de $P(Z < 2)$ en la tabla de la distribución normal:

$$P(X \geq 220) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

El valor de 0.0228 es una buena aproximación a la distribución binomial (probabilidad exacta).

6.8.3 Corrección de continuidad

Al realizar la aproximación existe una diferencia mínima, que consiste en que los rectángulos salen de la curva (figura de la página siguiente), por lo que tiene que efectuarse un ajuste, como se muestra en la expresión.



$$P(a \leq X \leq b) = P\left(\frac{a - np - 0.5}{\sqrt{np(1-p)}} \leq Z \leq \frac{b - np + 0.5}{\sqrt{np(1-p)}}\right)$$

El ejemplo anterior se resolverá con la corrección de continuidad:

$$n = 400$$

$$p = 0.5$$

$$P(X \geq 220) = 1 - P(X < 220)$$

Para

$$np = 400\ (0.5)$$

$$b = 220$$

$$P(X \geq 220) = 1 - P\left(Z < \frac{220 - 200 + 0.5}{\sqrt{400\ (0.5)(0.5)}}\right)$$

$$= 1 - P\left(Z < \frac{20.5}{10}\right)$$

$$= 1 - P(Z < 2.05)$$

$$= 1 - 0.9798$$

Entonces

$$\boxed{P(X \geq 220) = 0.0202}$$

Ejemplo 10 Si 23% de los pacientes con presión sanguínea elevada tienen efectos colaterales nocivos por la ingestión de cierto medicamento, utilice la aproximación

normal para obtener la probabilidad de que entre 120 de estos pacientes tratados con este medicamento, más de 32 presentarán efectos colaterales nocivos.

Solución:
Los datos que proporciona el problema son los siguientes:
$p = 0.23$
$n = 120$

Se plantea el problema

$$P(X \geq 32) = 1 - P(X < 32)$$

El mismo ejemplo sugiere utilizar la aproximación normal:

$$P(X \geq b) = 1 - P\left(Z < \frac{b - np + 0.5}{\sqrt{np\,q}}\right)$$

Se sustituyen los valores en la expresión anterior:

$$P(X \geq 32) = 1 - P\left(Z < \frac{32 - (0.23)(120) + 0.5}{\sqrt{120\,(0.23)\,(1 - 0.23)}}\right)$$

$$P(X \geq 32) = 1 - P\left(Z < \frac{4.9}{4.6099}\right)$$

$$= 1 - P(Z < 1.0629092)$$

$$= 1 - 0.8554$$

Entonces

$$\boxed{P(X \geq 32) = 0.1446}$$

La probabilidad de que más de 32 pacientes presenten efectos colaterales nocivos es de 0.1446 de 120 pacientes tratados.

6.9 DISTRIBUCIÓN EXPONENCIAL

El origen de esta función de probabilidad se halla en el proceso Poisson; tal probabilidad se relaciona con la posibilidad de ocurrencia de un número

específico de éxitos, donde el número de éxitos es la variable aleatoria. Ahora, si se invierten los papeles, el número de éxitos no será la variable aleatoria, sino el tiempo; en otras palabras, una variable exponencial X es el intervalo de tiempo, o espacio, requerido para obtener un número específico de éxitos.

Definición Se dice que una variable aleatoria continua X, que toma todos los valores no negativos, tiene una distribución exponencial con parámetro θ positivo si su función de probabilidad está dada por:

$$f(X)\begin{cases}\dfrac{1}{\theta}e^{[-X/\theta]} & \text{en otro caso} \quad X>0\\ 0\end{cases}$$

$$F(X)\begin{cases}1-e^{-\alpha X} & \text{en otro caso} \quad X\geq 0 \quad \text{donde} \quad \alpha=\dfrac{1}{\theta}\\ 0\end{cases}$$

Ejemplo 11 La vida de cierta marca de bombilla está distribuida exponencialmente con $\theta = 1\ 000$ horas.

a) ¿Cuál es la probabilidad de que la bombilla dure menos de 1 000 horas?

$$P(X<1\ 000)=1-e^{-\left(\frac{X}{\theta}\right)}=1-e^{-\left(\frac{1\ 000}{1\ 000}\right)}=1-e^{-1}\Rightarrow$$

$$P(X<1\ 000)=\boxed{0.63212}$$

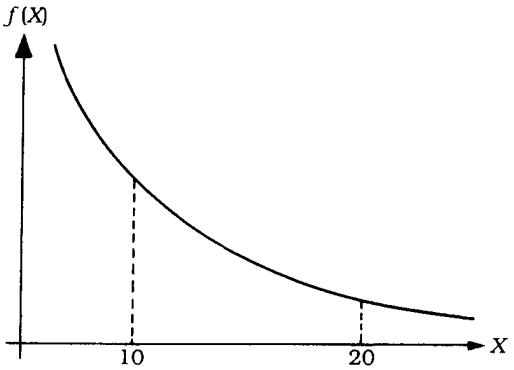
b) ¿Cuál es la probabilidad de que la bombilla dure más de 1 200 horas?

$$P(X>1\ 200)=1-P(X\leq 1\ 200)=1-[1-e^{-\left(\frac{1\ 200}{1\ 000}\right)}]$$

$$=e^{-1.2}\Rightarrow \boxed{P(X>1\ 200)=0.30119}$$

Ejemplo 12 Cierta computadora digital, que funciona 24 horas al día, sufre 10% descomposturas por hora. Si la computadora ha funcionado satisfactoriamente durante 10 horas, ¿cuál es la probabilidad de que continúe así durante las 10 horas siguientes?

$$\alpha = 0.1$$



$$P(10 < X < 20) = \left[1 - e^{-\left(\frac{20}{10}\right)}\right] - \left[1 - e^{-\left(\frac{10}{10}\right)}\right]$$

$$= 0.86466 - 0.63212$$

entonces $\boxed{P(10 < X < 20) = 0.23254}$

Ejemplo 13 El recorrido (en miles de kilómetros) que obtienen los propietarios de automóviles con cierto tipo de neumático radial es una variable aleatoria que tiene una distribución exponencial con $\theta = 40$. Obtenga la probabilidad de que uno de estos neumáticos dure:

a) Cuando menos 20

$$P(X > 20) = 1 - P(X \leq 20)$$

como $\theta = 40 \Rightarrow \alpha = \dfrac{1}{\theta} \Rightarrow \alpha = 0.025$

$$P(X > 20) = 1 - \left[1 - e^{-\left(\frac{20}{40}\right)}\right] = e^{-0.5}$$

$$\boxed{P(X > 20) = 0.6065306}$$

b) Como máximo 30

$$P(X \leq 30) = 1 - e^{-\left(\frac{30}{40}\right)} = 1 - 0.4723665 = 0.5276334$$

RESUMEN

Las distribuciones discretas de probabilidad como la binomial, Poisson e hipergeométrica se relacionan con el proceso de Bernoulli. En el caso de la binomial, las probabilidades individuales son los términos del desarrollo binomial (teorema del binomio). La hipergeométrica considera las muestras de una población finita y es posible aproximarse por la binomial cuando el tamaño de la muestra es pequeña respecto a la población. La distribución de Poisson considera las observaciones efectuadas en un continuo (por ejemplo, el tiempo), en lugar de una secuencia finita de ensayos; Poisson puede utilizarse para aproximar a la binomial cuando el número de ensayos es grande y el parámetro "p" es cercano a cero o a uno.

Una distribución muy importante y necesaria para justificar el desarrollo de casi todas las distribuciones probabilísticas, es su gama.

PROBLEMAS

6.1 Supóngase que un grupo de estudiantes ha realizado un examen en el que los valores posibles son 0, 1, 2, ..., 10. Se elige una muestra de 2 estudiantes, y se calcula la media de la muestra $\overline{X}$.

a) Describa el espacio muestral de la variable aleatoria X.
b) En una recta identifique el intervalo de todos los posibles resultados que puede tomar la variable aleatoria X.

6.2 Clasifique cuál de las siguientes variables aleatorias es discreta o continua.

a) Tiempo de reacción ante el mismo medicamento a un grupo de pacientes.
b) Cantidad de ozono en la atmósfera en la ciudad de México.
c) Número de personas que sufren de infarto en un municipio.
d) Cantidad de colesterol de una persona.
e) Número de cheques no pagados en una sucursal bancaria.

6.3 Dado un intervalo de una variable aleatoria X menor que 20 y mayor a -4, cuya media aritmética es $\mu = 8$ y $\sigma = 3$, encuentre el límite inferior y superior (aplicando la desigualdad de Chebyshev).

6.4 La proporción de fumadores en una ciudad es $p = 0.3$. Se toma una muestra aleatoria de tamaño $n = 16$ personas. ¿Cuál es la probabilidad de que haya $k = 4$ fumadores en la muestra? Supóngase que aunque la muestra se tomó sin reemplazamiento, la probabilidad $p = 0.3$ permanece igual.

6.5 La proporción de familias que usan el jabón A en cierta ciudad es $p = 0.2$. Se toma una muestra aleatoria de tamaño $n = 20$ familias. ¿Cuál es la probabilidad de que $k = 5$ familias usen ese jabón?

6.6 La proporción de familias que usan los productos de la compañía K en cierta ciudad es $p = 0.7$. Se toma una muestra aleatoria de $n = 16$ familias. ¿Qué probabilidad hay que $k = 9$ de esas familias en la muestra usen los productos de la compañía K?

6.7 La proporción de estudiantes que reciben la calificación S es de $p = 0.7$. Se toma una muestra aleatoria de tamaño $n = 10$ estudiantes. ¿Cuál es la probabilidad de que haya $k = 6$ estudiantes con esa calificación en la muestra?

6.8 Los ojos color verde son una característica genética, supongamos que dicho genotipo de los futuros padres de una familia, que planean procrear 6 hijos es del 80%. Calcular todos los eventos posibles.

$n = 6$
$r = 0, 1, 2, 3, 4, 5, 6$
$p = 0.8$
$q = 0.2$

Calcular μ y σ, graficar la distribución.

6.9 Los ojos color café son una característica genética, supongamos que en dicho genotipo de los futuros padres de una familia exista una probabilidad de ocurrencia del 75%. Calcular los eventos posibles si se planea procrear 6 hijos.

$n = 6$
$r = 0, 1, 2, 3, 4, 5, 6$
$p = 0.75$
$q = 0.25$

Calcular μ y σ.

Graficar la distribución.

6.10 Se lanzan dos dados, sea X la suma que muestran los dos dados:

a) Encuentre la función de probabilidad en forma tabular.
b) Encuentre μ, σ^2 y σ.
c) Graficar la distribución.

6.11 El 12% de los estudiantes que se inscriben a un curso de estadística en la Facultad de Psicología, tendrán que repetir el curso; si dicho grupo consta de 15 alumnos, ¿cuál es la probabilidad de que

a) menos de 6 tengan que repetir el curso?
b) exactamente 10 aprueben?
c) más de 12 aprueben?
d) Calcular μ y σ, graficar la distribución.

6.12 En el año de 1995, un huracán causó, entre otros daños, una gran inundación en una región del sureste de México, donde el 44% contrajeron el cólera. Si el sector salud atendió a una familia compuesta por 13 miembros, ¿cuál es la probabilidad de que tengan el cólera

a) 6 o más miembros?
b) 7 o menos?
c) 9 o más?
d) Calcular μ y σ, graficar la distribución.

6.13 La Lotería Nacional tiene una probabilidad de cometer errores en la impresión de sus billetes, dicha probabilidad es 0.5% ($P = 0.005$). Estos billetes no salen a la venta. ¿Cuál es la probabilidad de que en un lote de 1 000 billetes de un sorteo determinado

a) ningún billete tenga errores?
b) 10 billetes tengan errores?
c) 15 exactamente tengan errores?

6.14 Una tienda de autoservicio se queja de que el 3% de los clientes que pagan con tarjetas de crédito, el banco emisor no les autoriza su compra; si un día 200 personas realizan sus compras en dicha tienda, ¿cuál es la probabilidad de que

a) exactamente 10 no se les autorice su compra?
b) exactamente 5 no se les autorice su compra?

6.15 La probabilidad de obtener cierta enfermedad infecciosa es de 0.001; si a 10 000 estudiantes dispuestos a ingresar a una universidad se les realiza un examen médico, ¿encuentre la probabilidad de que:

a) 6, 7 u 8 tengan dicha enfermedad?
b) Calcular μ y σ correspondiente.

6.16 La probabilidad de votos equivocados anulados en una determinada casilla es de 0.002. ¿Cuál es la probabilidad que de las 2 000 personas que emitieron su voto en dicha casilla, menos de 5 les sea anulado su voto?

Calcule μ y σ.

6.17 Supóngase que, en promedio, 1 casa de cada 2 000 en cierto distrito, se incendia durante el año. Si hay 6 000 casas en el distrito, ¿cuál es la probabilidad de que exactamente 5 casas se incendien durante el año?

6.18 Tal vez la probabilidad de marcar un número equivocado es de 0.05. Entonces, ¿cuál es la probabilidad de marcar un número equivocado exactamente 3 veces en 100 llamadas telefónicas?

6.19 La probabilidad de que un cajero de un almacén dé el vuelto equivocado es de 0.02. ¿Cuál es la probabilidad que el cajero dé el cambio equivocado exactamente 5 veces en 200 ventas que hace?

6.20 Estímese que de cada 5 000 carros uno tiene problemas con las llantas en cierta autopista. Si 1 000 carros transitan por la autopista durante cierto día, ¿cuál es la probabilidad de que exactamente dos carros tengan problemas con las llantas?

6.21 En una tienda de autoservicio se colocan 100 bombillas, de las cuales $n_1 = 60$ (son de 75 watts), $n_2 = 40$ (son de 100 watts). Se toman al azar $r = 10$ bombillas, ¿cuál es la probabilidad de que haya $X = 4$ de 75 watts?

6.22 Consideremos a 10 estudiantes universitarios, de los cuales 6 son hombres y 4 son mujeres, si se escoge una muestra al azar de tamaño 5, ¿cuál es la probabilidad de que haya 3 hombres en esta muestra?

6.23 Considérese el caso de 10 electores, de los cuales 5 son liberales, 3 conservadores y 2 independientes. Si se toma una muestra al azar de tamaño 5, ¿cuál es la probabilidad de que haya 2 liberales y 2 conservadores?

6.24 De una distribución normal la media es de 26.1 y la desviación estándar de 10.7.

a) Calcule las calificaciones Z que equivalen a los siguientes datos de dicha distribución: 5, 18, 30, 39, 46, 51.
b) Transforme cada una de estas calificaciones Z en una puntuación X con media de 50 y desviación estándar de 10.

6.25 Dadas las siguientes puntuaciones Z:

a) 0.27	*c)* 3.00	*e)* −0.65	*g)* −0.80
b) 2.47	*d)* 1.00	*f)* −0.05	*h)* −3.20

Calcule el área comprendida entre la media y cada puntuación. Suponga que la distribución es normal.

6.26 Una prueba conocida como C.I. para niños de determinada edad es estandarizada para tener un promedio de 100 y una desviación estándar de 15. Para una muestra de 500 niños, de una cierta edad, calcule el porcentaje del área bajo la curva de:

a) Superior a 130.
b) Superior a 150.
c) Debajo de 80.

d) Entre 90 y 100.
e) Fuera de los límites de 80 y 120.

6.27 En una distribución normal, calcule la proporción de casos que quedan entre:

a) $Z = -1.00$ y 1.00.
b) $Z = -2.00$ y 2.00.
c) $Z = -1.25$ y 0.75.
d) $Z = -0.75$ y 1.25.
e) $Z = -1.30$ y -0.30.

6.28 En una distribución normal, calcule la proporción de casos que están.

a) Entre $Z = -0.25$ y -1.25.
b) Entre $Z = 0.25$ y 1.25.
c) Entre $Z = -0.62$ y -2.15
d) Fuera de los límites de $Z = 0.40$ y -0.20.
e) Fuera de los límites de $Z = -2.00$ y 2.00.

6.29 En una distribución normal, calcule la proporción de casos que quedan:

a) Superior a $Z = 1.00$.
b) Abajo de $Z = -2.00$.
c) Arriba de $Z = 3.00$.
d) Abajo de $Z = 0$.
e) Entre $Z = -1.28$ y 0.
f) Arriba de $Z = -1.28$.
g) Entre $Z = 0$ y 1.64.

6.30 Para la distribución de los resultados de una prueba de introversión-extroversión que se aplicó a un numeroso grupo de estudiantes, $X = 82$ y $s = 12$ (los resultados altos están en dirección de la introversión). Para los siguientes estudiantes convierta los resultados nuevos a un resultado Z.

a) Juan 88 *b*) Jorge 96 *c*) María 72 *d*) Alfredo 64

6.31 Calcule el valor de Z asociado a las siguientes áreas bajo la gráfica de la distribución normal.

a) El área a la izquierda de Z es 0.9732.
b) A la derecha de Z es 0.6406.
c) A la izquierda de Z es 0.0409.
d) A la derecha de Z es 0.2389.
e) Entre Z y μ es 0.3212.
f) A la derecha de Z es 0.025.
g) A la derecha Z es 0.01.
h) Entre Z y $-Z$ es 0.99.

6.32 Un estudio, realizado por un instituto de investigación en la docencia, trata de determinar el porcentaje de maestros que alcanzaron el 70% de control sobre sus grupos, que es lo que el instituto establece como puntuación mínima de control. Si se investigaron 100 maestros y se encontró un porcentaje medio de 67, con una desviación estándar de 12, determine la cantidad de los que alcanzaron a cubrir el requisito.

6.33 En un hospital de traumatología se encontró que las contracciones de 300 injertos se distribuyen normalmente con una media de 68 mm y una desviación estándar de 3 mm. A los 6 meses:

a) ¿Cuántos injertos tiene una contracción mayor de 72 mm?
b) ¿Cuántos menor o igual de 62 mm?
c) ¿Cuántos entre 65 y 71 mm?

Aproximación de la binomial a la normal.

6.34 Una variable aleatoria binomial tiene una media de 6 y una desviación estándar de 1.5 ¿Qué porcentaje del área bajo el histograma de probabilidad estará a la izquierda de 5; a la derecha de 7?

6.35 Se aplica un examen de opción múltiple que consta de 200 preguntas $n = 200$, cada una con 4 posibles respuestas, de las cuales sólo 1 es la respuesta correcta. ¿Cuál es la probabilidad de que un estudiante conteste de 25 a 30 correctamente, para una sección de 80 de las 200, en la cual el estudiante no tiene conocimiento?

6.36 Un especialista en control de calidad determina que la duración efectiva de un componente electrónico, producido por una empresa, sigue una distribución exponencial.

$$f(X) = \frac{1}{1\,000} e^{-X/1000} \qquad \text{si } X \geq 0$$

y

$$f(X) = 0 \qquad \text{si } X < 0$$

Calcule la probabilidad de que al seleccionar en forma aleatoria un componente electrónico tenga una duración de menos de 1 000 horas.

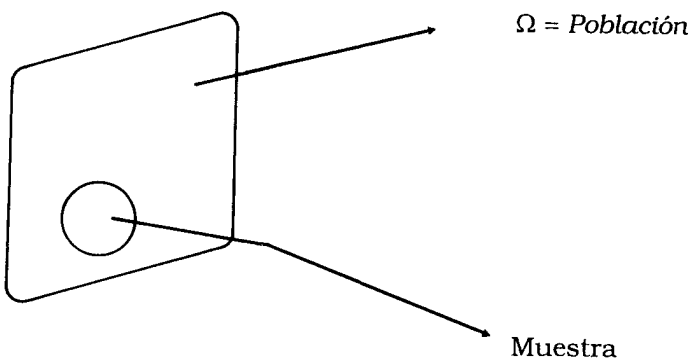
INTRODUCCIÓN

A manera de preámbulo, es importante destacar cuál es el objetivo de la estadística, a fin de entender la trascendencia del muestreo. Según cierto postulado, "el objetivo fundamental de la estadística es hacer inferencias acerca de una población con base en la información contenida en una muestra." Desde esta perspectiva, la estadística es un instrumento que ayuda a conocer determinadas características de toda una población por medio de la selección de sólo algunos elementos pertenecientes al conjunto universal, llamado también *población* o *universo* y denotado por Ω. La muestra siempre será un subconjunto de la población en estudio.

7.1 FUNDAMENTOS DE MUESTREO

Ejemplo Se desea conocer el tipo de música que prefieren los estudiantes de cierta universidad.

Población = estudiantes de la universidad



Con ese fin, sólo a algunos estudiantes se les preguntará acerca de su música predilecta. Así se podrá efectuar una inferencia de la población en estudio.

Una vez esbozado el juego del muestreo en la estadística, se explicará el objetivo de un diseño de muestreo: proporcionar indicaciones para la selección de una muestra a fin de que sea representativa de la población, con lo que se obtendrá una cantidad específica de información a un costo mínimo.

En el ejemplo anterior, la función que desempeña el muestreo consiste en proporcionar el número de universitarios encuestados necesario y suficiente para realizar la inferencia respectiva. Dicho número se estima por medio de distribuciones de probabilidad y tipo de muestreo.

Técnicas de muestreo 7

PROPÓSITOS

El propósito central de este capítulo consiste en que el lector sea capaz de reconocer los procedimientos básicos de las técnicas de muestreo.

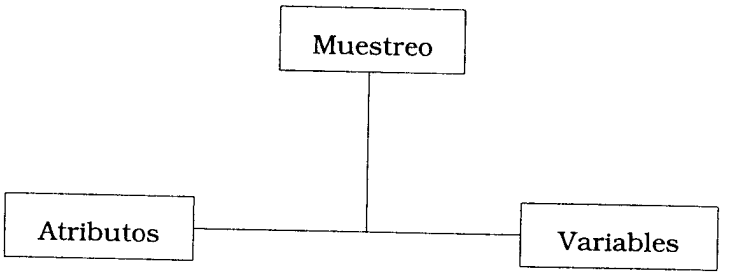
Del mismo modo, el lector logrará:

1. Conocer los fundamentos del muestreo.
2. Precisar con mayor conocimiento el objetivo de la estadística.
3. Distinguir entre las técnicas de muestreo y lo que es un censo.
4. Identificar las técnicas de muestreo ubicadas en una investigación.
5. Explicar el muestreo por atributos y variables.
6. Distinguir los dos niveles de muestreo, el determinístico y el probabilístico.
7. Desarrollar los procedimientos de cálculo para los tipos de muestreo determinístico: juicio, diseño de bola de nieve, conveniencia por cuotas (entrevista de grupo de intercepción de centros comerciales y paneles controlados).
8. Identificar y desarrollar los procedimientos de muestreo probabilístico, como el aleatorio simple, estratificado, sistemático por conglomerados, simple y por dos etapas, entre otros.
9. Interpretar los resultados del cálculo de las técnicas de muestreo.
10. Identificar la calidad del diseño de muestreo requerido para muestreo determinístico (baja) y muestreo probabilístico (alta).
11. Conocer los tipos de muestra para distribuciones muestrales.
12. Aplicar el desarrollo de estas muestras para resolver problemas relacionados con la selección de la muestra.

En el muestreo fijo, como su nombre lo indica, se establece un número fijo de tamaño de muestra y se recaban los datos antes de realizar el análisis de los resultados.

Por otra parte, en el muestreo secuencial el tamaño de muestra no se determina por una regla de decisión, ni de manera arbitraria. Se comienza con la recolección de datos y, en forma periódica, se examina si es necesario continuar con el muestreo o se llega a la conclusión de que la información recabada basta para analizar los datos y realizar las inferencias necesarias.

CASO 2

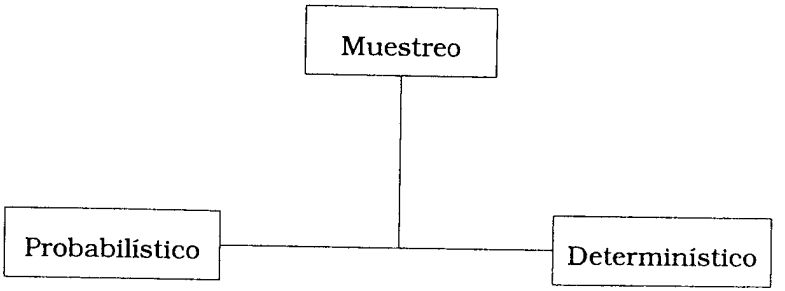


La diferencia entre el muestreo de atributos y el de variables consiste en el tipo de parámetro que se busca acerca de la población. Un atributo es una característica cualitativa o descriptiva que un integrante de la población posee o no posee; por eso siempre se hará referencia a proporciones de la población en estudio.

Por otra parte, el muestreo de variables trata de estimar una medida de la población y no una proporción de ella.

A fin de ilustrar con un ejemplo el muestreo por atributos, supóngase que se necesita conocer la opinión de pedagogos acerca de las clases que se imparten en cierta universidad. Nuestro atributo o característica lo constituirán todas las personas que cursaron la carrera de pedagogía y se titularon en dicha materia. Debido a que no todos los individuos cumplen ese requisito, se habla de una *proporción de la población total.*

CASO 3



No obstante, tal vez la muestra conste de demasiados alumnos, lo que ocasiona que el costo sea elevado y el tiempo necesario excesivo. Por consiguiente, debe reducirse el número de la muestra con la responsabilidad que implica esta decisión, ya que, conforme disminuye la muestra, decrece el grado de confiabilidad, de 99%, 95%, 90%. Esto se explicará con más amplitud en el presente capítulo.

Muestreo contra censo. Es esencial considerar las ventajas que tiene una muestra bien realizada respecto a un censo.

Disminuye el costo de la investigación y el tiempo. Si se desea conocer alguna característica o preferencia en poblaciones con un gran número de individuos, por ejemplo, la ciudad de México, Monterrey o Guadalajara, resultaría muy costoso y la obtención de la información deseada abarcaría demasiado tiempo mediante un censo; por eso, se recomienda una muestra.

Es posible obtener una información apropiada de la población a partir de la muestra, siempre que se eviten errores como la mala elaboración de la encuesta o del cuestionario o las equivocaciones de los entrevistadores. Sobre el particular, conviene impartir a éstos capacitación previa para cada proyecto que se va a realizar; este tipo de errores por lo general se llaman *sesgos*.

Otro punto valioso es que toda muestra tiene un error muestral, debe tomarse un número de elementos suficientes para poder inferir acerca de la población; cuanto menor sea el error muestral, mayor será la muestra, es decir, se tiende a encuestar al universo, y estaremos realizando un censo; lo que se busca es efectuar una muestra considerando el error y evitando sesgos en toda nuestra investigación.

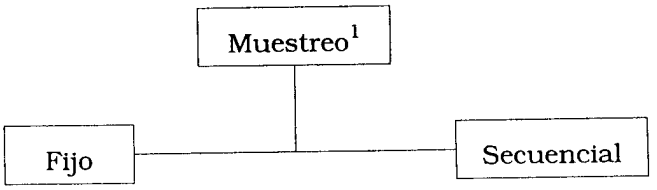
Por último, el hecho de realizar un censo no quiere decir que éste no contendrá equivocaciones. Los sesgos pueden ser de varios tipos, por eso se recomienda impartir a los encuestadores una capacitación en la que se haga hincapié en lo importante y delicado que es su trabajo.

7.2 CLASIFICACIONES DE MUESTREO

Los diseños muestrales se clasifican de diversas maneras; por ejemplo, en fijo o secuencial, por atributos y por variables, probabilístico y determinístico.

A continuación se presentan sus esquemas:

CASO 1



[1] M. Roland, Weiers *Investigación de mercados*, Prentice-Hall-Hispanoamericana, México, 1986, pp. 103-107.

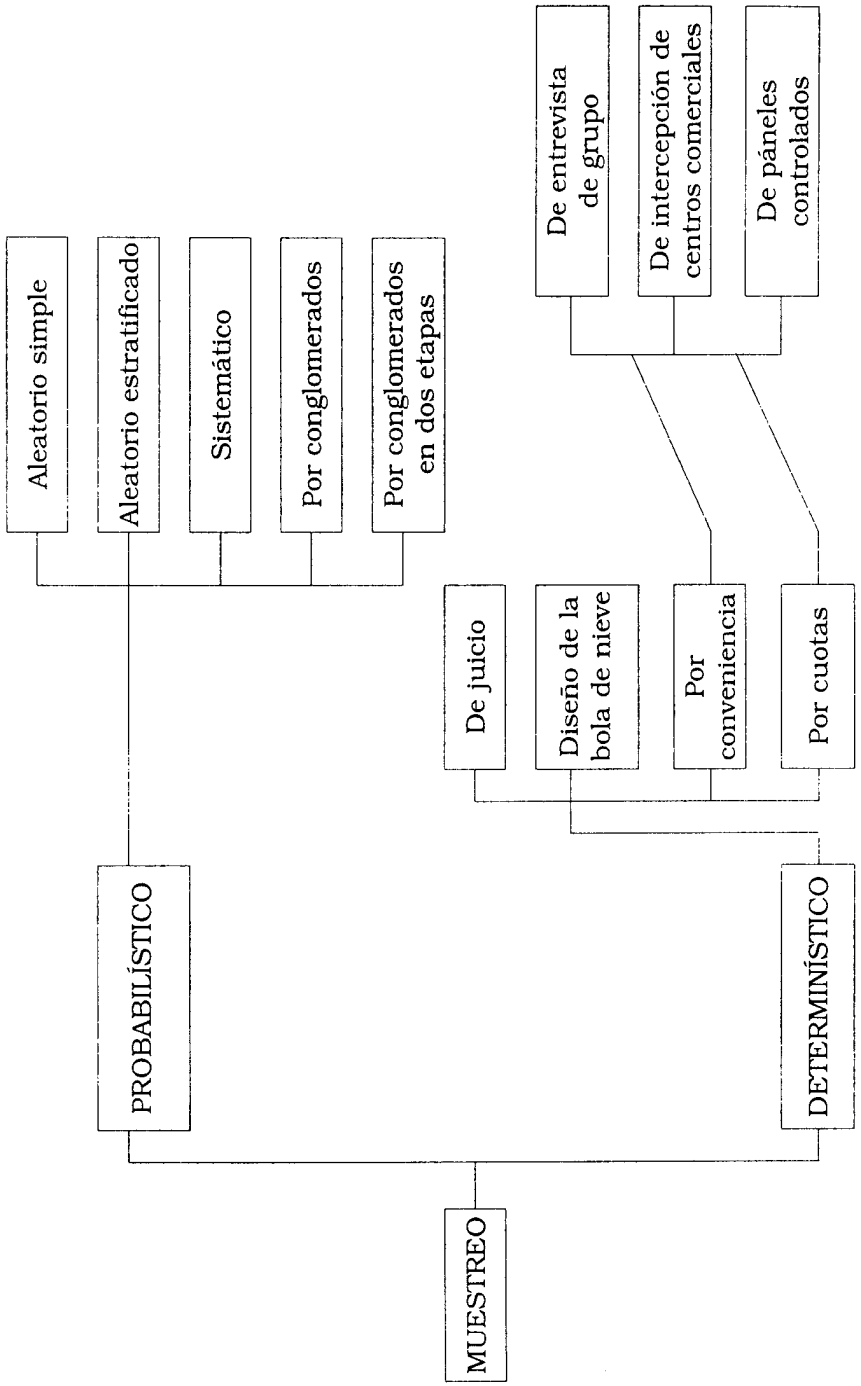

MUESTREO
PROBABILÍSTICO
DETERMINÍSTICO
Aleatorio simple
Aleatorio estratificado
Sistemático
Por conglomerados
Por conglomerados en dos etapas
De juicio
Diseño de la bola de nieve
Por conveniencia
Por cuotas
De entrevista de grupo
De intercepción de centros comerciales
De páneles controlados

Figura 7.1

Esta clasificación es la más usual y por ello se dará una explicación extensa durante el desarrollo del capítulo.

El muestreo probabilístico indica que la elección de cada integrante de la población debe ser por completo aleatorio, y que todos tendrán la misma probabilidad de ser seleccionados. Existen algunos métodos para conocer el número de muestra, como se verá más adelante (sección 4).

En el muestreo determinístico, el investigador escoge los elementos de acuerdo con sus necesidades, el número de muestra y los elementos de la población en estudio.

Por ejemplo, para establecer el número apropiado de escuelas primarias y secundarias que es necesario construir en Sonora, hace falta conocer la proporción de niños que viven en esa entidad. Sería muy costoso realizar un censo, por lo que se usará el muestreo. La pregunta inmediata es: ¿probabilístico o determinístico? En este caso conviene más el muestreo probabilístico; de lo contrario, sólo se extraerá información referente a cierto sector o colonia, la cual no será confiable para saber cuántas escuelas se requieren y la capacidad de matrículas por escuela y nivel.

Para ilustrar el muestreo determinístico, se tiene lo siguiente: un maestro de universidad desea conocer la opinión de sus alumnos respecto a calidad en clases. Él imparte la materia a cuatro grupos, pero decide encuestar sólo a dos. Es importante señalar que el maestro solamente será capaz de realizar un estudio estadístico *descriptivo.* No podrá hacer ningún tipo de inferencia al respecto.

7.2.1 Muestreo determinístico contra probabilístico

Uno de los primeros problemas que enfrenta el investigador consiste en elegir el tipo de muestreo más adecuado. Se explicarán las ventajas y desventajas de ambos a fin de brindar al investigador un panorama global, y de que esté en condiciones de determinar un modelo de muestreo acorde con sus objetivos.

En primer término se abordará el muestreo determinístico, entre cuyas ventajas está el que es económico y requiere poco tiempo; asimismo, su puesta en marcha suele ser tan sencilla como en el muestreo por conveniencia o tan compleja como en el muestreo por cuotas, ya que el investigador no necesita contar con conocimientos avanzados de estadística. Respecto a las desventajas, no es posible obtener parámetros insesgados y de medición viable y no se cuenta con un modelo para la obtención del tamaño de muestra; en la recolección de los datos existirá un sesgo intrínseco imperceptible al análisis de los mismos.

En el cuadro 7.1 se presentan los tipos de muestreos determinísticos y probabilísticos, los cuales se estudiarán con más detalle en este mismo capítulo.

de reunión. Resultaría difícil obtener una lista de vendedores ambulantes, así que una muestra de juicio es la más apropiada.

Como en todas las muestras no probabilísticas, un aspecto sobresaliente radica en la imposibilidad de asegurar que la muestra de juicio será representativa; tampoco cabe recurrir a técnicas estadísticas que permitan determinar el tamaño de la muestra adecuada para lograr cierto grado de exactitud, ni proyectar en forma estadística los resultados en un nivel de confianza para el parámetro de la población en estudio.

7.3.2 Muestreo por cuotas

El muestreo por cuotas consiste en un muestreo de juicio, con la restricción de que la muestra incluye un número mínimo de cada subgrupo específico dentro de la población.

Ejemplo Se desea conocer el nivel de ingresos de los colonos de la Delegación Venustiano Carranza; con ese fin, se realizará un muestreo por cuotas en algunas colonias. Se cuenta con los datos siguientes:

Colonia	*Porcentaje*	*Casa habitación*
Jardín Balbuena	20	47
Merced Balbuena	20	25
Puebla	20	63
Zaragoza	20	15
Moctezuma	20	41
	100	191

El porcentaje por colonia no tiene que ser necesariamente igual.

7.3.3 El diseño de bola de nieve

Este tipo de diseño resulta conveniente cuando se necesita alcanzar poblaciones pequeñas o datos especializados. Su mecanismo es el siguiente: a cada entrevistado se le pide identificar a uno o más sujetos pertenecientes al mismo campo, hasta recopilar las encuestas que el investigador crea necesarias para satisfacer los objetivos de estudio.

Ejemplo Un laboratorio de la ciudad de México desea saber si los urólogos del Instituto Mexicano del Seguro Social aceptan los medicamentos que produce. El primer paso será encontrar el primer urólogo, después éste recomendará a otro u otros, quienes proporcionarán la información buscada.

7.3.4 Muestreo por conveniencia

Este procedimiento consiste sencillamente en ponerse en contacto con unidades de muestreo apropiadas.

Por otra parte, el muestreo probabilístico tiene la ventaja de contar con modelos para calcular el tamaño de muestra requerida con una confiabilidad y un grado de error. Es viable calcular parámetros insesgados referentes al universo en estudio a partir de poca información sobre el mismo, en términos relativos como el número total de elementos. En cuanto a sus desventajas, es un muestreo de alto costo y exige más tiempo que el determinístico; para algunos investigadores resulta compleja ya que no cuentan con conocimientos de estadística inferencial.

Por último, se considerará la calidad del diseño de muestreo requerido. En la práctica el probabilístico es irrealizable en su totalidad debido a su alto costo; por eso, en ocasiones experimenta cambios (llamados *adaptaciones del muestreo probabilístico*). Éstos provocan que la calidad del diseño disminuya considerablemente; es responsabilidad del investigador decidir hasta dónde está dispuesto a sacrificar dicha calidad (fig. 7.2).

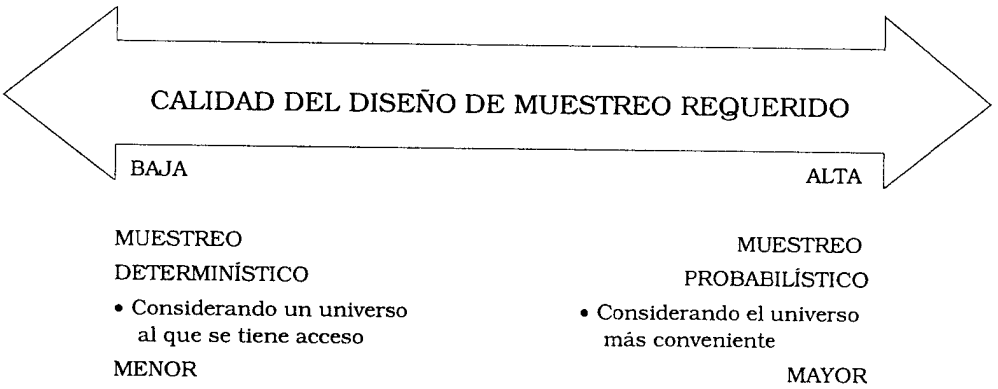


Figura 7.2

7.3 MUESTREO DETERMINÍSTICO

7.3.1 Muestreo de juicio o criterio

La característica principal de este tipo de muestreo es que tanto el tamaño de muestra como la elección de los elementos están sujetos al juicio del investigador, esto es, para realizar un estudio mediante este tipo de muestreo debe recurrirse a la experiencia que se tenga. El éxito y la eficacia dependen de la opinión del investigador que haya seleccionado los elementos.

Hay situaciones en las que el muestreo de juicio es útil y aun aconsejable: cuando el muestreo probabilístico no es factible o resulta costoso. Una muestra aleatoria de homosexuales puede ser muy difícil de obtener, por lo que se utilizará una muestra de juicio sobre los que frecuentan los bares y otros centros

Ejemplo Se desea introducir en el mercado un jabón que limpia las impurezas de la cara. Con base en su experiencia, el investigador decide aplicar directamente el jabón al público y pedir su opinión en forma escrita. Tal actividad se realizará en centros comerciales como Perisur, Plaza Satélite, Plaza Inn, Galerías Insurgentes y otros. Todo ello a *conveniencia* del investigador.

Un muestreo por conveniencia a menudo se utiliza para pruebas piloto o para probar un cuestionario.

7.4 MUESTREOS PROBABILÍSTICOS

Se estudiarán las características de los muestreos probabilísticos y se explicará cómo obtener el tamaño de muestra en cada caso. En el marco teórico del muestreo existen los estimadores de punto y los de intervalo.

7.4.1 Muestreo aleatorio simple

El muestreo aleatorio simple (MAS) o muestreo irrestricto aleatorio es el más usado por los investigadores. Es importante destacar que forma la base de la mayor parte de los diseños de muestreo, así como de las encuestas científicas que se llevan a la práctica. Según Scheaffer: "Si un tamaño de muestra n es seleccionado de una población de tamaño N, de tal manera que cada muestra posible de tamaño n tiene la misma probabilidad de ser seleccionada, el procedimiento de muestreo se denomina *irrestricto aleatorio.* A la muestra así obtenida se le llama *muestra irrestricta aleatoria.*"[2]

A continuación se estudiará cómo seleccionar una muestra irrestricta aleatoria, para lo cual algunos investigadores acostumbran utilizar una tabla de números aleatorios.

Tabla de números aleatorios

Dada una lista de los integrantes de una población, es posible enumerarlos de 1 a N, al igual que se enumera un conjunto de canicas si se les coloca en una urna, y se extraen una a una hasta obtener n canicas cuando n es el tamaño de la muestra. En ésta se incluyen los integrantes de la población a quienes les corresponden los números de las canicas muestreadas, y es posible medir o estudiar las características de estas unidades (integrantes de la población) más adecuadamente.

Si la población es muy grande, este método de selección aleatorio puede ser difícil de aplicar. Para atender este problema se utiliza la tabla de números aleatorios; los números que formarán esta tabla los genera un mecanismo en el que

---

[2] Richard. Scheaffer L., *Elementos de muestreo*, Iberoamérica, México, 1991.

cada dígito es el resultado de un ensayo, consistente en seleccionar un número de 0, 1, 2... 9 con una probabilidad igual a cada número de 1/10. Los dígitos en las diferentes posiciones son resultado de repeticiones independientes de dichos ensayos; es decir, supóngase que 10 canicas idénticas numeradas de 0 a 9 se colocan en una urna y, después de mezclarse, se extrae una y se anota el dígito correspondiente; posteriormente se regresa la canica a la urna y la operación se repite de manera sucesiva. En la actualidad, debido a consideraciones prácticas, una computadora que simula este procedimiento genera las tablas de números aleatorios, y los conjuntos de números resultantes se revisan cuidadosamente respecto a su independencia e igual probabilidad.

Uso de la tabla de números aleatorios

Para ejemplificar el uso de la tabla de números aleatorios, supóngase que hay almacenadas 40 baterías de una marca determinada y se desea obtener una muestra de $n = 4$ para conocer su estado de deterioro.

Paso 1. Se enumeran las baterías de 1 a 40; los dígitos se escogen de dos en dos, porque el tamaño de la población es $\overline{N} = 40$ (número de dos dígitos).

Paso 2. De modo arbitrario se selecciona una porción de la tabla, una columna y un renglón; supóngase que se selecciona el renglón 60 y la columna 4 y se leen los pares de dígitos en las columnas 4 y 5, lo que da 13, 02, 18, 74, 59, 13, 74, 33; se omiten tanto los números mayores de 40 como los números repetidos.

Paso 3. Se continúan leyendo los pares de dígitos hasta que se obtienen cuatro unidades diferentes, es decir, 13, 02, 18 y 33, que son las baterías seleccionadas.

Por otra parte, existen modelos que posibilitan conocer el tamaño de muestra, para cuya aplicación es necesario conocer tanto un supuesto como algunos conceptos. Éstas se refieren al uso óptimo de las fórmulas.

Resulta esencial analizar el supuesto siguiente: "La población en la que se tomará la muestra debe obedecer a una distribución." Mediante estudios previos es posible saber si lo anterior se cumple o, en su defecto, examinar si las variables permiten conocer el comportamiento de la distribución. De lo contrario, el análisis debe realizarse por medio de estadística no paramétrica; por ejemplo, en ocasiones los psicólogos desean conocer el grado de coeficiente intelectual, la intensidad de dolor de cabeza, etc., casos en los que las variables no obedecen una distribución normal. Algunas variables que se distribuyen en forma normal son el promedio de edad de determinada población, así como los gustos en materia de colores, música, comida y vestimenta, entre otros.

Los conceptos que ayudarán a determinar el tamaño de muestra son los siguientes:

E = Error muestral. Es la diferencia entre la media muestral y la media de la población aceptable.

Z = Número de unidades de desviaciones estándar en la distribución normal que producirá el nivel deseado de confianza.

σ = Desviación estándar de la población.

A partir de la definición de intervalo de confianza se descubrirá un modelo para encontrar el tamaño de muestra, el cual se estudió en el capítulo 6.

Supóngase que X es una variable aleatoria cuya ley de probabilidad depende de un parámetro desconocido θ. Dada una muestra aleatoria de $x_1, x_2, x_3, \ldots, x_n$, las dos estadísticas L_1 y L_2 conforman un intervalo de confianza del 100 $(1-\alpha)$% para θ si

$$P(L_1 \leq \theta \leq L_2) \geq 1 - \alpha$$

Supóngase que la muestra se distribuye en una forma normal 0, 1:

$$L_1 = -Z_{\alpha/2} \qquad \text{y} \qquad \theta = Z$$
$$L_2 = +Z_{\alpha/2}$$

se obtiene

$$P(-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1 - \alpha \qquad \ldots \text{(I)}$$

Se sabe que

$$Z = \frac{\overline{X} - \mu_X}{\sigma_X / \sqrt{n}} \qquad \ldots \text{(II)}$$

Se sustituye en

$$P\left(-Z_{\alpha/2} \leq \frac{\overline{X} - \mu_X}{\sigma_X / \sqrt{n}}\right) \leq Z_{\alpha/2} = 1 - \alpha$$

$$P\left(-Z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \leq \overline{X} - \mu \leq Z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\overline{X} - Z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \leq \mu_x \leq \overline{X} + Z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}\right) = 1 - \alpha$$

Como se aprecia en el intervalo de confianza, la media poblacional (μ_x) oscilará en un rango determinado por el valor de $\pm Z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}}$, al cual se le asignará el nombre de *error muestral*, es decir:

$$E = Z_{\alpha/2} \frac{\sigma_X}{\sqrt{n}} \qquad \ldots \text{(III)}$$

De la expresión III se despeja n a fin de obtener una fórmula para determinar el tamaño de muestra a partir de un intervalo de confianza.

$$\frac{E}{Z_{\alpha/2}} = \frac{\sigma_X}{\sqrt{n}}$$

$$\left(\frac{E}{Z_{\alpha/2}}\right)^2 = \frac{\sigma_X^2}{n}$$

$$n = \frac{(Z_{\alpha/2}\,\sigma_X)^2}{E^2} \qquad \ldots \text{(IV)}$$

donde E es el error de muestreo.

A la expresión IV se le denomina *caso 1*, donde la población es infinita y el parámetro que se busca acerca de la población es una variable (véanse las clasificaciones de muestreo en el caso 2).

CASO 2

La población es finita y el parámetro que se busca acerca de la población constituye una variable:

$$n = \frac{\sigma_X^2}{\left(\frac{E}{Z_{\alpha/2}}\right)^2 + \frac{\sigma_X^2}{N}}$$

donde N = tamaño de la población.

CASO 3

La población es infinita y el parámetro que se busca acerca de la población constituye un atributo o característica:

$$n = \frac{Z_{\alpha/2}^2 \;\; p\,(1-p)}{E^{\,2}}$$

donde p = proporción de la población que tiene la característica que se va a medir.

CASO 4

La población es finita y el parámetro que se busca acerca de la población constituye un atributo o característica:

$$n = \frac{p\,(1-p)}{\left(\frac{E}{Z_{\alpha/2}}\right)^2 + \frac{p\,(1-p)}{N}}$$

En la figura 7.3 se muestran los casos anteriores con una serie de pasos para el uso de las fórmulas.

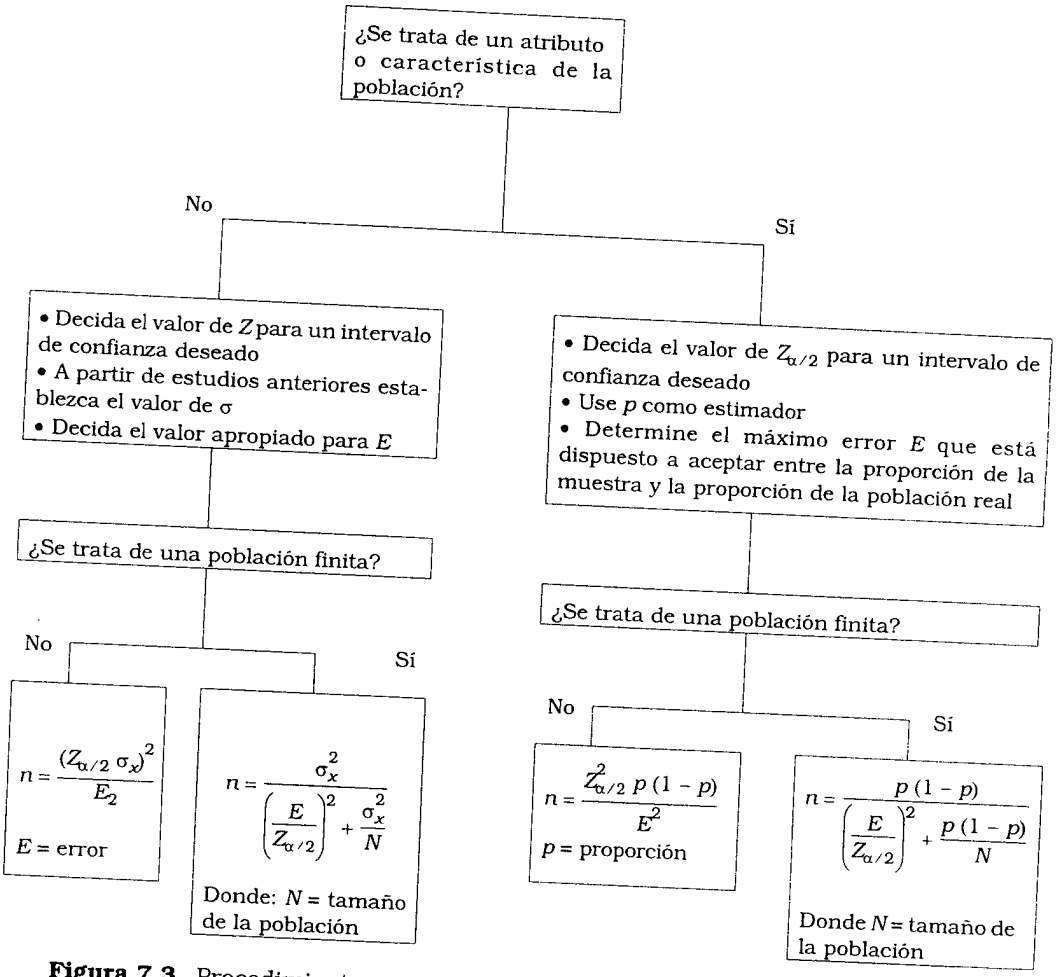


Figura 7.3 Procedimiento para encontrar el tamaño de muestra en un MAS.

Ejemplo Existe la intención de realizar una encuesta sobre elecciones presidenciales entre los alumnos de derecho de la Escuela Nacional de Estudios Profesionales (ENEP) Acatlán. Con este fin, se desea encontrar el tamaño de muestra para una población total de 12 000 estudiantes; la proporción de los alumnos de derecho respecto a la población total es de 32.76 por ciento.

Se acepta un error muestral de $E = 0.10$ y una confiabilidad de 95 por ciento:

$$E = 0.10$$

$$Z = 1.96$$

Como se desea encuestar sólo a la población de derecho, se habla de una proporción y se usará la fórmula siguiente:

$$n = \frac{p(1-p)}{\left(\frac{E}{Z_{\alpha/2}}\right)^2 + \frac{p(1-p)}{N}}$$

$$p = 0.3276,\ E = 0.10,\ N = 12\ 000$$

Se sustituyen en la fórmula los valores

$$n = \frac{0.3276\,(1-0.3276)}{\left(\frac{0.10}{1.96}\right)^2 + \frac{0.3276(1-0.3276)}{12\ 000}}$$

$$n = \frac{0.2202782}{0.0026031 + 0.0000184} = 84.02893$$

$$n = 84$$

7.4.2 Muestreo aleatorio estratificado (MAE)

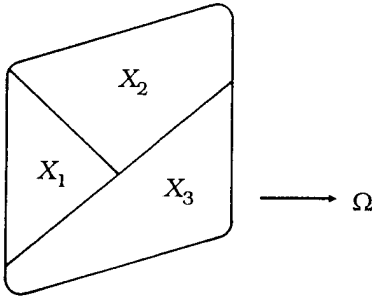
Para examinar este tipo de muestreo es necesario recordar los conceptos de conjuntos mutuamente excluyentes y el de partición. Supóngase que se desea conocer la calidad del servicio que prestan los hoteles en Acapulco por medio de muestreo aleatorio estratificado.

Sea Ω = todos los hoteles de Acapulco.
Con base en la siguiente partición, se obtienen los subconjuntos de Ω:

X_1 = todos los hoteles cinco estrellas.
X_2 = todos los hoteles cuatro estrellas.
X_3 = todos los hoteles tres estrellas.

A cada X_i $i = 1, 2, 3$ se le llamará *estrato*.
Es necesario considerar que:

a) $X_1 \cap X_2 = \varnothing$
$X_1 \cap X_3 = \varnothing$
$X_2 \cap X_3 = \varnothing$



Es decir, los subconjuntos deben ser mutuamente excluyentes, con lo que se garantiza que no se traslapen los subconjuntos:

b) $\sum_{i=1}^{3} X_i = N_\Omega = \Omega$

Es posible aplicar un MAS en cada estrato, a fin de encontrar el tamaño de muestra en cada X_i. Cada una de ellas se denotará por

$$n_1, n_2, n_3 \ldots n_i,$$

Se busca conocer el grado de satisfacción del consumidor correspondiente a cada subdivisión. Una de las ventajas que presenta el muestreo aleatorio estratificado (MAE) respecto al MAS es que con el estrato se facilita identificar las diferencias entre categorías. Por otra parte, en la estratificación existe una mayor precisión en los estimadores poblacionales.

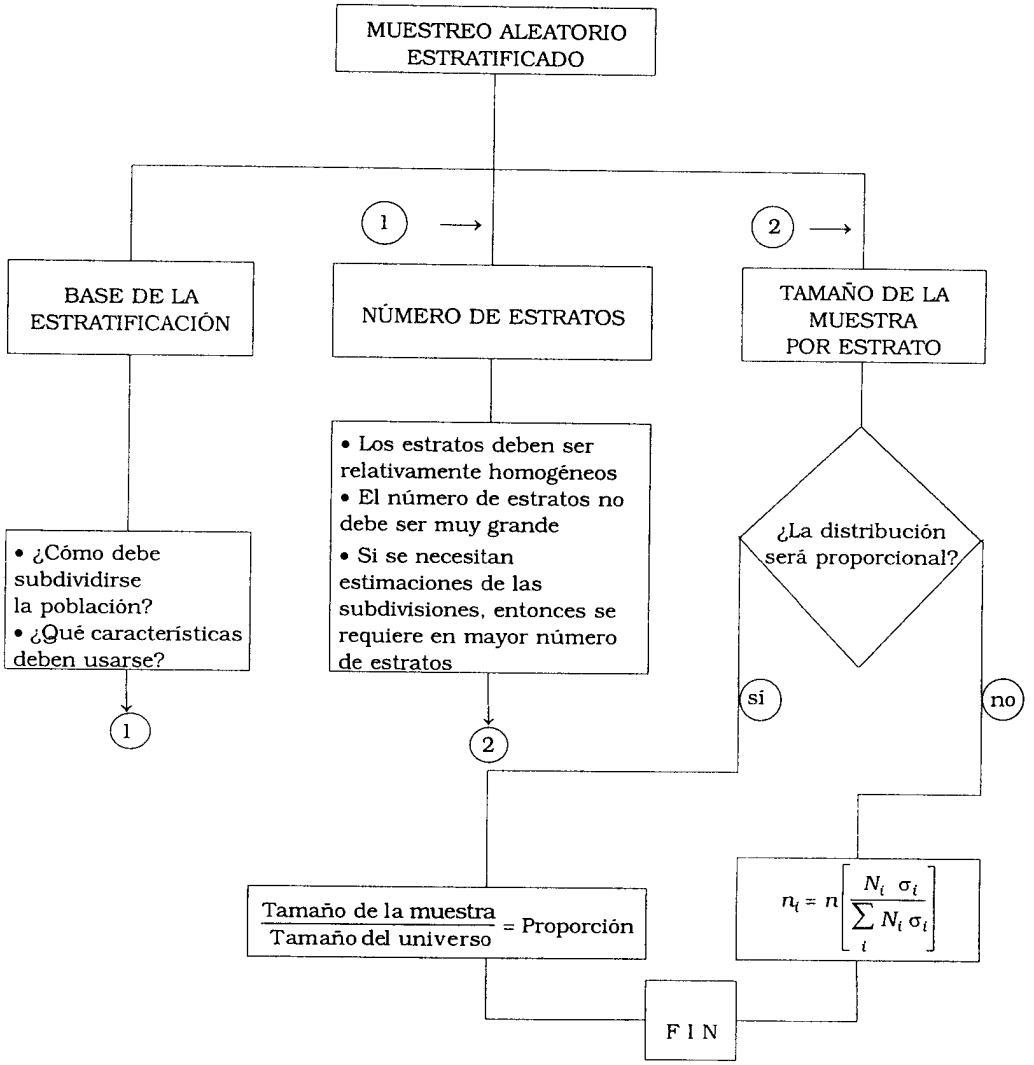


Figura 7.4

Ejemplo En la ENEP Acatlán se desea realizar una encuesta para conocer la preferencia partidista. Con este propósito, es necesario calcular el tamaño de la muestra total y por estrato.

Se aplicará un muestreo aleatorio estratificado debido a las siguientes características de la población:

- La población en estudio es relativamente homogénea en edad y nivel socioeconómico.
- Los estratos de la población se conocen: son las carreras; se requiere establecer la tendencia partidista en cada una de ellas.
- Al considerar los estratos de las carreras se minimizan los sesgos de la muestra.
- La confiabilidad es de 95%.

MUESTREO ALEATORIO ESTRATIFICADO
Datos para calcular el tamaño de muestra

Valor de $Z_{\alpha/2}$	1.96	Utilizando
Desviación estándar	25	
Valor de E	1%	$n_i = n\left(\frac{N_i}{N}\right)$
Población total (N)	12 000	
Tamaño de muestra (n)	2 000	

POBLACIÓN DE LA ENEP ACATLÁN

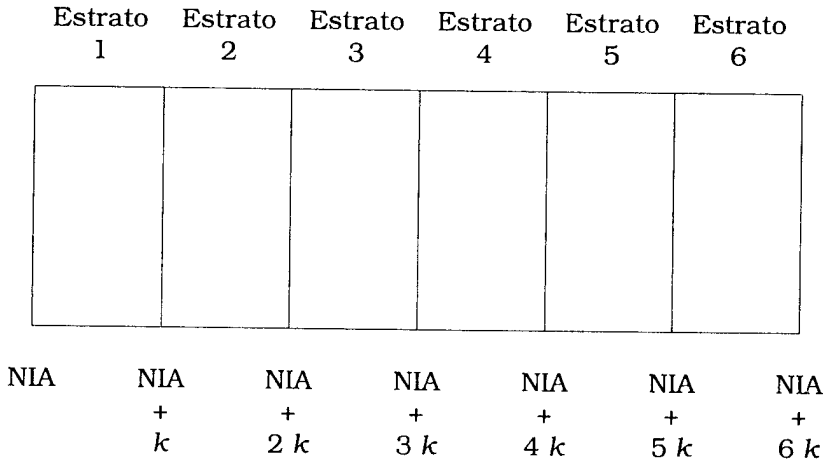
	Carrera	N_i	*Proporciones* (porcentaje)	$\left(\frac{N_i}{N}\right)$	n_i
1	Actuaría	557	4.642		93
2	Derecho	3 931	32.758		655
3	Periodismo	1 296	10.800		216
4	Arquitectura	834	6.950		139
5	Ingeniería civil	493	4.108		82
6	MAC	1 749	14.575		292
7	Diseño gráfico	572	4.767		95
8	Economía	487	4.058		81
9	Pedagogía	661	5.508		110
10	Ciencias políticas	707	5.892		118
11	Sociología	314	2.617		52
12	Filosofía	70	0.583		12
13	Letras hispánicas	75	0.625		13
14	Historia	156	1.300		26
15	Inglés	98	0.817		16
	TOTAL	12 000	100.00		2000
	Tamaño de muestra	2 000			

7.4.3 Muestreo sistemático (MS)

El muestreo sistemático es muy semejante al MAS; en lo único que difiere es en la manera como se selecciona la muestra: en el MAS se utiliza una tabla de números aleatorios, mientras que en el muestreo sistemático el procedimiento es exacto.

Antes de explicar el método, se presentan algunas ventajas del muestreo sistemático respecto al aleatorio simple:

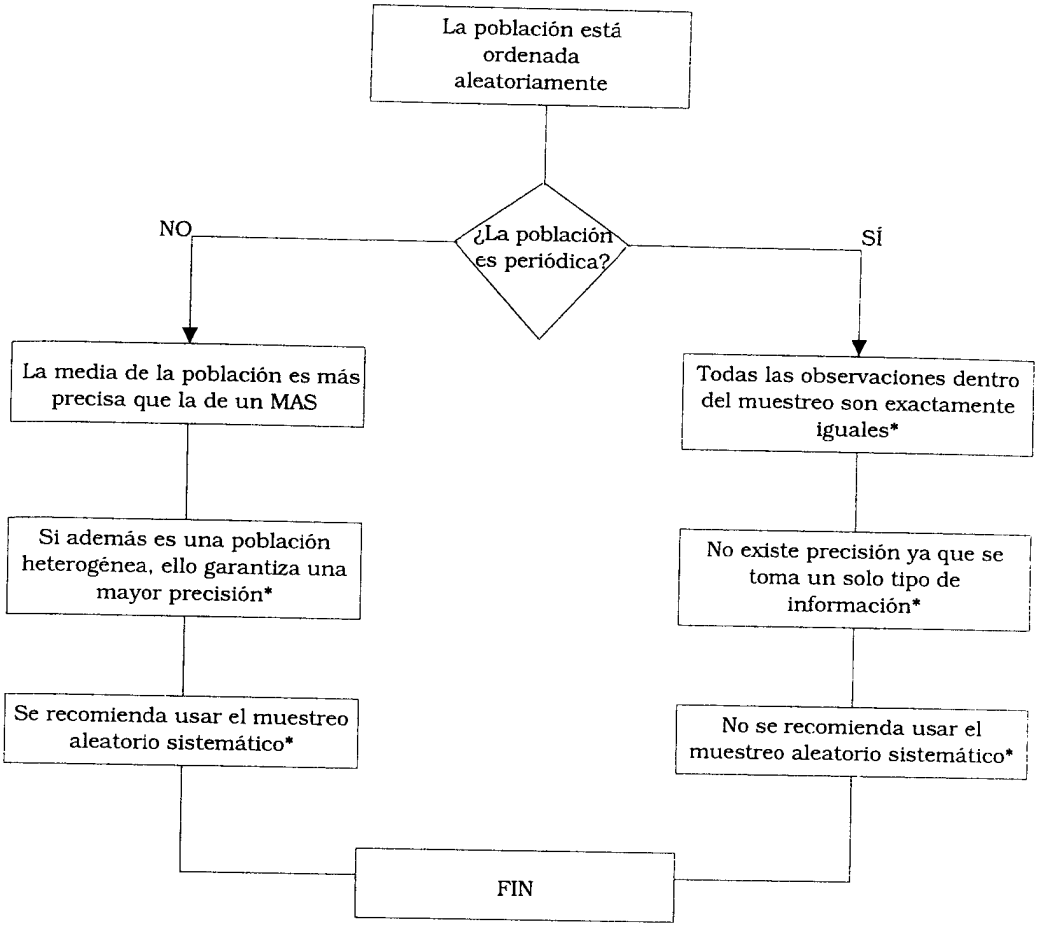
- Es más sencillo escoger los elementos de la muestra y se disminuye el grado de error.
- En el muestreo aleatorio sistemático sólo se selecciona un elemento de los estratos en que se divide la muestra, como en el caso anterior. Los estratos se dividen en grupos de k elementos; sólo se elige al k-ésimo elemento.



Por otra parte, este método es conveniente si las poblaciones se presentan en un orden aleatorio, por ejemplo, un directorio telefónico que no se encuentre ordenado alfabéticamente, sino por el primer dígito del número, o una lista que contenga las calificaciones de biología de un grupo de alumnos, es decir, cuando las unidades están ordenadas de 1 a N.

En principio, puede afirmarse que los estimadores de un muestreo aleatorio irrestricto (MAI), como se denominará al muestreo aleatorio simple, y los de un muestreo sistemático son iguales. Antes de explicar el motivo de lo anterior, se analizará otro concepto.

Definición Una población es periódica si sus elementos tienen variaciones cíclicas. Esto origina el diagrama que se muestra en la figura 7.5.



* Al muestreo aleatorio simple lo llamaremos *muestreo aleatorio irrestricto.*

Figura 7.5 Muestreo aleatorio sistemático.

Una vez que se tiene el tamaño de la población, se recurre a encontrar el tamaño de muestra n con la fórmula de un MAI.

Después se realiza el cociente:

$$k = \frac{\text{Tamaño de la población}}{\text{Tamaño de la muestra}}$$

A continuación se escoge un número aleatorio que se encuentre en el rango de $(1, k)$; a ese número se le llamará *número inicial aleatorio* (NIA). Éste es el

primer número de la sucesión, es decir, el primer elemento de la lista. Los elementos se seleccionarán según la expresión siguiente:

$$\{\text{NIA}, \text{NIA} + ik\} \; i \in N$$

La sucesión termina en n.

Para mayor claridad, en el diagrama de la figura 7.6 se muestran los pasos que deben seguirse.

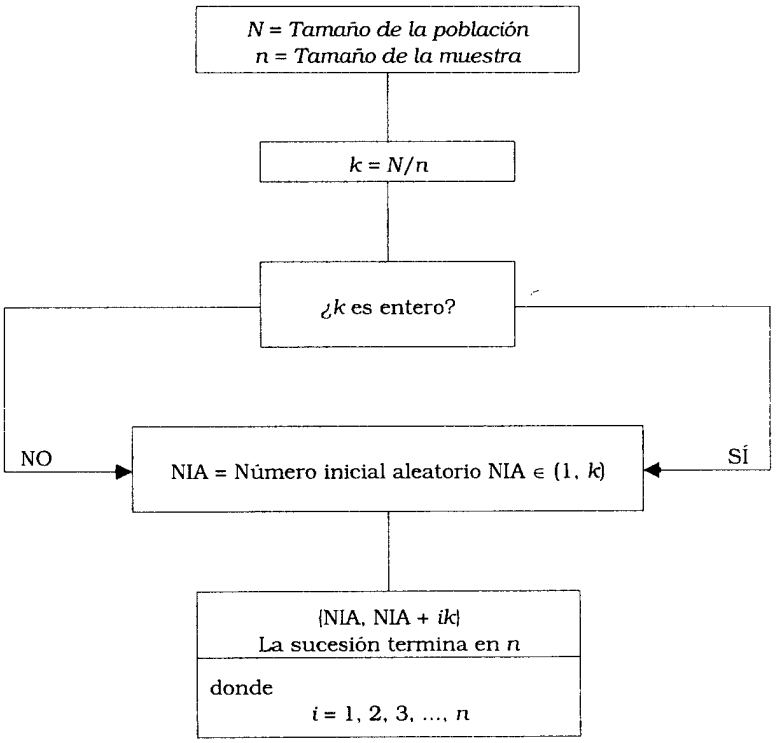


Figura 7.6 Muestreo sistemático.

Ejemplo En la Facultad de Psicología se necesita saber cuántos alumnos de quinto semestre disponen de una computadora personal. Se cuenta con los datos:

$N = 180$

$Z = 1.29 \quad \Rightarrow \quad$ 90% de confiabilidad

$\alpha = 10\%$

$E = 5\%$

$\sigma_x^2 = 0.25$

$$n = \frac{0.25}{\left(\frac{0.05}{1.29}\right)^2 + \frac{0.25}{180}} = 86.80$$

$$n = 87$$

$$k = \frac{180}{87} = 2.069 \text{ se aproxima } k = 2$$

Con base en la siguiente lista se obtendrán los elementos con un número aleatorio inicial, e incrementando *ik* para obtener los elementos de la muestra.

Lista de alumnos de la Facultad
de Psicología que cursan el quinto semestre

	Elementos de la muestra
1. Ángeles Rodríguez Eva Angelina	No
2. Ayhllon Osorio Nidya Elizabeth	Sí
3. Sandoval Miranda Lourdes	No
4. Cortez Herrera Lorena	Sí
5. Sánchez Guzmán Jéssica	No
6. Alfiz Segura Amira	Sí
7. Ruiz de Chávez Juárez Marcela	No
8. Villafuerte Cosme Selene	Sí
9. Linares Jiménez María Genoveva	No
10. Ramírez Labastida Juan Ismael	Sí
11. Reyes García Diana Angélica	No
. . .	
. . .	
. . .	
180. Urrutia Odabachián Araxi	Sí

7.4.4 Muestreo por conglomerados

Definición Una muestra por conglomerados es una muestra aleatoria, en la cual cada unidad de muestreo es una colección o conglomerado de elementos.

Es fácil confundir el muestreo por conglomerados con el aleatorio estratificado, ya que en ambos se forman subconjuntos de la población para realizar el muestreo; una diferencia radica en la recolección de la muestra.

En ocasiones se requiere obtener información que se encuentra muy dispersa, por ejemplo, el número de niños en determinada ciudad, el estado civil promedio en una comunidad. Si se quisiera aplicar un MAI, MAE o MAS sería muy costoso. Una de las características que ayudan a decidir qué tipo de muestreo se aplicará es la siguiente: "La razón principal es que por lo general no existe ninguna lista utilizable de unidades (denominada *marco*) dentro de la cual seleccionar la muestra."[3]

[3] I. Méndez, *Conceptos muy elementales del muestreo con énfasis en la determinación práctica del tamaño de la muestra*, Comunicaciones Técnicas, Serie Azul, 1980, núm. 25.

Es importante reconocer tres elementos para la aplicación correcta de este diseño:

- *Variable de interés.* Es lo que se desea conocer acerca de la población.
- *Unidades primarias.* A la primera división de la población en estudio se le denomina *unidad primaria.* Por ejemplo, para conocer el tipo y la calidad del maíz en determinada región de la República, se debe dividir en hectáreas o poblados. Éstas son las unidades primarias.
- *Unidades secundarias.* Una vez bien definidas las unidades primarias, es necesario subdividirlas para establecer cuántas unidades se someterán a observación; en el ejemplo anterior, las unidades secundarias son las milpas que se observarán.

Es primordial destacar que tanto el número de unidades primarias como el de secundarias se obtendrá por medio de un MAI.

Ejemplo Se desea conocer el consumo de electricidad en las zonas residenciales de Naucalpan; para ello, se decide aplicar un muestreo por conglomerados en colonias como San Mateo, Satélite, Lomas Verdes, entre otras.

A continuación se muestran los tres elementos antes explicados. Como se recordará, la razón principal para aplicar este tipo de muestreo es que no existe una lista utilizable de unidades (un marco) con la que pueda trabajarse. El ejemplo tiene esta característica, por lo que se recomienda usar el MAE.

En la zona residencial de Naucalpan los indicadores son los siguientes:

Variable de interés	Unidades primarias	Unidades secundarias
Consumo de electricidad	Manzanas	Casa habitación

Para saber el tamaño de muestra correspondiente a las unidades primarias, se aplica:

$$n_p = \frac{\sigma_X^2 N}{\frac{B^2}{4}(N-1) + \sigma_X^2}$$

donde

$$B = 2\sqrt{\frac{\sigma_X^2}{n}\left(\frac{N-n}{N-1}\right)}$$

donde $\frac{N-n}{N-1}$ es el factor de corrección para poblaciones finitas (fcpf).

Para las unidades secundarias se tiene:

$$n_s = \frac{\sigma_X^2 N}{\frac{B^2}{4}(N-1) + \sigma_X^2}$$

7.5 DISTRIBUCIONES MUESTRALES

Las distribuciones muestrales t, F, x^2; correspondientes a un estadístico ($\bar{x}$, σ_X^2, etc.) pueden conocerse mediante la probabilidad muestral. Sobre esta base es posible efectuar inferencias acerca de la población, ya que la distribución muestral de un estadístico es básica para hacer inferencias de su correspondiente parámetro poblacional. Por consiguiente, se considera a la distribución muestral como una distribución de probabilidad teórica, que relaciona varios valores o intervalos de valores del estadístico muestral con sus probabilidades de ocurrencia. Para conocer la distribución muestral hay que especificar la distribución poblacional.

7.6 TEOREMA CENTRAL DEL LÍMITE[4]

La distribución normal es la base de muchas distribuciones estadísticas importantes. Dicha distribución puede servir como una buena aproximación para otras distribuciones teóricas cuyas probabilidades sean muy laboriosas de obtener o imposibles de calcular.

Existe una relación cercana entre el tamaño de una muestra n, y la consideración o el supuesto de que dicha muestra se aproxime a la normal. Muchas distribuciones muestrales cuyo tamaño n es muy grande pueden aproximarse a la distribución normal, aun cuando las distribuciones poblacionales de aquéllos definitivamente no sean normales. A este principio se le conoce como el *teorema central del límite*, ya que la distribución normal es la forma límite n grande ($n \to \infty$) para una gran variedad de distribuciones muestrales.

Otra forma es ver el teorema de los grandes números. Si la probabilidad de ocurrencia de un evento E es P, y si se realizan n ensayos, independiente y exactamente dentro de las mismas condiciones, entonces la probabilidad de que la frecuencia relativa de ocurrencia de E difiera de P por alguna cantidad ε, por muy pequeña que sea ε, tiende a cero, a medida en que el número de ensayos tiende a infinito (crece indefinidamente).

[4] Teorema del Límite central.

Sea X una variable aleatoria distribuida con media μ_X y varianza σ_X^2

si $n \to \infty$

entonces $\overline{X}$ es variable aleatoria y se distribuye normalmente con media $\mu_{\overline{X}} = \mu_X$ y varianza $\sigma_{\overline{X}}^2 = \dfrac{\sigma_X^2}{n} \Rightarrow$

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma_X^2}{n}}$$

$$\frac{\sqrt{\sigma_X^2}}{\sqrt{n}}$$

$$\Rightarrow \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

En general si N; X; μ_X; σ_X^2

n; $\overline{X}$; $\mu_{\overline{X}} = \mu_X$;

$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \cdot \frac{N-n}{N-1}$$

$\dfrac{N-n}{N-1}$ se conoce como *factor de corrección para poblaciones finitas*

Si $N \to \infty$ entonces $\dfrac{N-n}{N-1} \to 1$

$\Rightarrow \dfrac{\sigma_X^2}{n} \times \dfrac{N-n}{N-1} \Rightarrow \sigma_{\overline{X}}^2 = \dfrac{\sigma_X^2}{n}$ como lo establece el teorema

Corroboración del teorema central del límite

Supóngase que de seis estudiantes, el primero tiene en moneda fraccionaria \$1; el segundo \$2, y así sucesivamente hasta el sexto, que tiene \$6.

Considérese los \$1, 2, ..., 6 como los datos de la población

X	X^2
\$1	1
2	4
3	9
4	16
5	25
6	36
21	91

$\Rightarrow N = 6; \quad \Sigma X = 21; \quad \Sigma X^2 = 91$

$$\mu_X = \frac{\Sigma X}{N} \Rightarrow \mu_X = \frac{21}{6} = 3.5 \Rightarrow \boxed{\mu_X = \frac{7}{2}}$$

$$\sigma_X^2 = E[X^2] - \mu_X^2 = \frac{\Sigma X^2}{N} - \mu_X^2$$

$$\Rightarrow \sigma_X^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = 2.916\,666\,667 = \frac{7}{2} \times \frac{5}{6}$$

$$\Rightarrow \boxed{\sigma_X^2 = \frac{7}{2} \times \frac{5}{6}}$$

Sea $n = 2$ (por ejemplo, se escoge a dos alumnos)

Muestras posibles (X_1, X_2):

\$1, 2	2, 3	3, 4	4, 5	5, 6
1, 3	2, 4	3, 5	4, 6	
1, 4	2, 5	3, 6		
1, 5	2, 6			
1, 6				

$$\Rightarrow P\left[\begin{array}{l}\text{Seleccionar una}\\ \text{cualquiera de las}\\ \text{muestras}\end{array}\right] = \frac{1}{15}$$

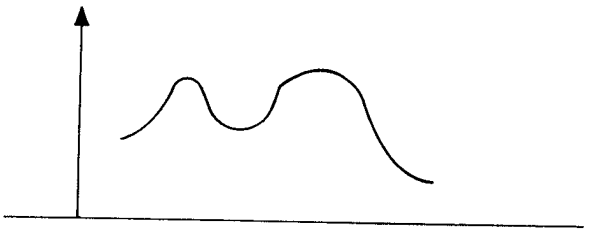
Calculemos las medias muestrales $\left(\overline{X} = \dfrac{X_1 + X_2}{2}\right)$:

\$ 1.5	2.5	3.5	4.5	5.5
2.0	3.0	4.0	5.0	
2.5	3.5	4.5		
3.0	4.0			
3.5				

$$\Rightarrow P\left[\begin{array}{l}\text{Seleccionar una}\\ \text{cualquiera de las}\\ \text{medias muestrales}\end{array}\right] = \frac{1}{15}$$

Detectemos, del caso que nos ocupa, la distribución muestral, $\overline{X} = \dfrac{X_1 + X_2}{2}$

Clase	$\overline{X}$	f	d	fd	d^2	fd^2
1	\$ 1.5	1	- 4	- 4	16	16
2	2.0	1	- 3	- 3	9	9
3	2.5	2	- 2	- 4	4	8
4	3.0	2	- 1	- 2	1	2
5	3.5	3	0	0	0	0
6	4.0	2	1	2	1	2
7	4.5	2	2	4	4	8
8	5.0	1	3	3	9	9
9	5.5	1	4	4	16	16
		15		0		70



(La distribución de la población es uniforme o plana).

$$\Rightarrow \Sigma f = 15;\ \Sigma fd = 0;\ \Sigma fd^2 = 70;\ \mu_X = 3.5,\ 5 = 0.5$$

$$\mu_{\overline{X}} = \mu_X + \left(\frac{\Sigma f d}{\Sigma f}\right) S \Rightarrow \mu_{\overline{X}} = 3.5 + \left(\frac{0}{15}\right) 0.5 = 3.5 \Rightarrow \boxed{\mu_{\overline{X}} = \frac{7}{2}}$$

$$\sigma^2_{\overline{X}} \left\{\frac{\Sigma f d^2}{\Sigma f} - \left(\frac{\Sigma f d}{\Sigma f}\right)^2\right\} S^2 \Rightarrow \sigma^2_{\overline{X}} = \left\{\frac{70}{15} - \left(\frac{0}{15}\right)^2\right\} 0.5^2$$

$$= \{4.666\,666\,667 - 0^2\}\,(0.25)$$

$$\Rightarrow \sigma^2_{\overline{X}} = 1.166\,666\,667 \qquad \Rightarrow \boxed{\sigma^2_{\overline{X}} = \frac{7}{6}}$$

resumiendo:

$$\mu_{\overline{X}} = \frac{7}{2} \quad \text{y} \quad \mu_X = \frac{7}{2} \qquad \Rightarrow \boxed{\mu_{\overline{X}} = \mu_X} \tag{1}$$

$$\sigma^2_X = \frac{7}{2} \times \frac{5}{6}$$

$$\Rightarrow \frac{\sigma^2_X}{n} \times \left(\frac{N-n}{N-1}\right) = \frac{\frac{7}{2} \times \frac{5}{6}}{\frac{2}{1}} \times \frac{6-2}{6-1} = \frac{7 \times 5}{4 \times 6} \times \frac{4}{5} = \frac{7}{6} \Rightarrow$$

$$\sigma_{\overline{X}}^2 = \frac{7}{6} \qquad \frac{\sigma_X^2}{n} \times \left(\frac{N-n}{N-1}\right) = \frac{7}{6} \qquad \Rightarrow \qquad \boxed{\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \times \left(\frac{N-n}{N-1}\right)} \qquad (2)$$

Ejemplo Se dispone de los datos de producción de trigo por hectárea de 101 chacras. La media es 15 quintales por hectárea y la desviación típica es de cuatro. Determinar la probabilidad de seleccionar una muestra de tamaño 25 con una media (muestral) menor o igual a 13.5 quintales:

$X \qquad N = 101 \qquad \mu_X = 15 \qquad \sigma_X = 4 \Rightarrow \sigma_{\overline{X}}^2 = 4^2$

$\overline{X} \qquad n = 25 \qquad \mu_{\overline{X}} = \mu_X = 15 \qquad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}\left(\frac{N-n}{N-1}\right)$

$$\Rightarrow \sigma_{\overline{X}}^2 = \frac{4^2}{25}\left(\frac{101-25}{101-1}\right) = \frac{16}{25} \times \frac{76}{100}$$

$$\Rightarrow \sigma_{\overline{X}}^2 = 0.4864$$

$$\Rightarrow \sigma_{\overline{X}} = \sqrt{0.4864}$$

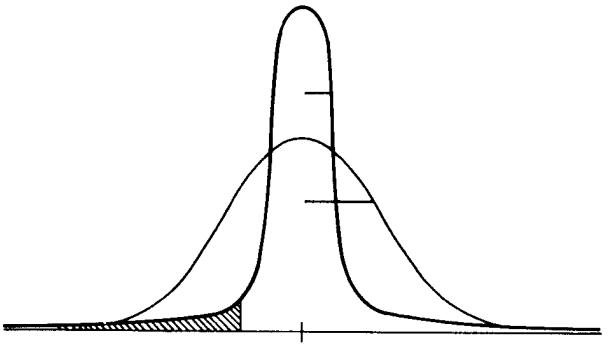
$$\Rightarrow \sigma_{\overline{X}} = 0.6974$$

$$P[\overline{X} \leq 13.5] = P\left[\frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \leq \frac{13.5 - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right]$$

$$= P\left[Z \leq \frac{13.5 - 15}{0.6974}\right]$$

$$= P[Z \leq -2.15]$$

$\Rightarrow P[\overline{X} \leq 13.5] = 0.0158$ obsérvese lo pequeño del valor;



Ejemplo Utilizando los datos anteriores, ¿cuál es la probabilidad de que:

a) la media muestral sea mayor o igual a 17 quintales?

$$a)\ P[\bar{X} \leq 12.5] = P\left[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \leq \frac{12.5 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right]$$

$$= P\left[Z \leq \frac{12.5 - 14}{0.5}\right]$$

$$= P[Z \leq -3]$$

$$= P[Z \geq 3]$$

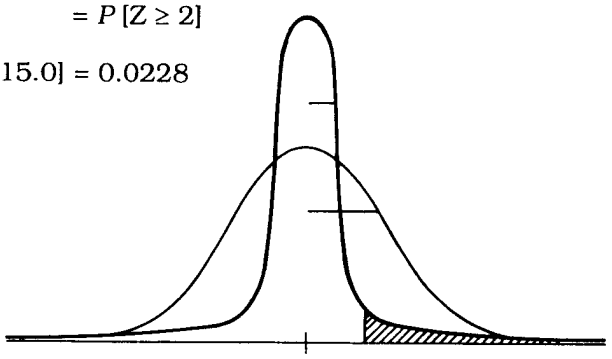
$$\Rightarrow P[\bar{X} \leq 12.5] = 0.0013$$

$$b)\ P[\bar{X} \geq 15.0] = P\left[\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \geq \frac{15.0 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right]$$

$$= P\left[Z \geq \frac{15.0 - 14}{0.5}\right]$$

$$= P[Z \geq 2]$$

$$\Rightarrow P[\bar{X} \geq 15.0] = 0.0228$$



$$c)\ P[12.5 \leq \bar{X} \leq 15.0] = P[\bar{X} \geq 12.5] - P[\bar{X} \geq 15.0]$$

$$= 1 - P[\bar{X} \leq 12.5] - P[\bar{X} \geq 15.0]$$

$$= 1 - 0.0013 - 0.0228$$

$$\Rightarrow P[12.5 \leq \bar{X} \leq 15] = 0.9759$$

Ejercicios

1. El número promedio diario de víctimas de accidentes de tránsito en una ciudad es de 120 personas y la desviación normal es de 32. Cuando se seleccionan 62 días al azar, ¿cuál es la probabilidad de que el número promedio de víctimas sea:

a) menor o igual a 111 personas?
b) mayor o igual a 130 personas?

b) la media muestral no esté entre 13.5 a 17 quintales?

a) $P[\overline{X} \geq 17] = P\left[\dfrac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \geq \dfrac{17 - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\right]$

$$= P\left[Z \geq \frac{17 - 15}{0.6974}\right]$$

$$= P[Z \geq -2.87]$$

$$P[\overline{X} \geq 17] = 0.0021$$

b) $1 - P[13.5 \leq \overline{X} \leq 17.0]$

ahora $P[13.5 \leq \overline{X} \leq 17.0] = P[\overline{X} \geq 13.5] - P[\overline{X} \geq 17.0]$

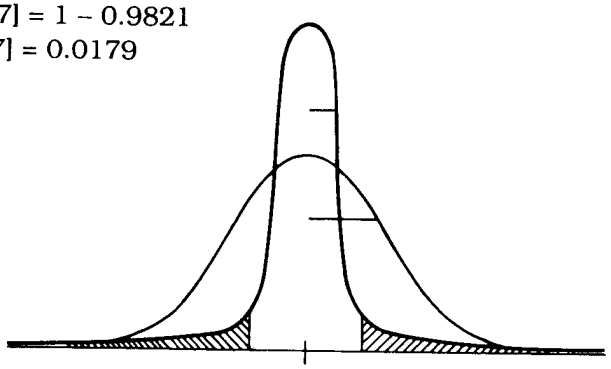
$$= 1 - P[\overline{X} \leq 13.5] - P[\overline{X} \geq 17.0]$$

$$= 1 - 0.0158 - 0.0021$$

$\Rightarrow$ $P[13.5 \leq \overline{X} \leq 17] = 0.9821$

$\Rightarrow$ $[13.5 \leq \overline{X} \leq 17] = 1 - 0.9821$

$\Rightarrow$ $[13.5 \leq X \leq 17] = 0.0179$



Ejemplo Según datos del pasado, el producto promedio de trigo por acre es de 14 bushels y la desviación normal es de cuatro. ¿Cuál es la probabilidad de seleccionar una muestra de tamaño 64 con una media muestral:

a) menor o igual a 12.5 bushels?
b) mayor o igual a 15.0 bushels?
c) entre 12.5 bushels y 15.0 bushels?

X	$N \to \infty$	$\mu_X = 14$	$\sigma_X = 4 \Rightarrow \sigma_X^2 = 4^2$
$\overline{X}$	$n = 64$	$\mu_{\overline{X}} = \mu_X = 14$	$\sigma_{\overline{X}}^2 = \dfrac{\sigma_X^2}{n}\left(\dfrac{N-n}{N-1}\right)$ como $N \Rightarrow \infty \Rightarrow \dfrac{N-n}{N-1} \to 1$

$$\sigma_{\overline{X}}^2 = \frac{4^2}{64} = \frac{4^2}{8^2} = \left(\frac{4}{8}\right)^2 = (0.5)^2$$

$$\Rightarrow \sigma_{\overline{X}}^2 = 0.5^2 \Rightarrow \sigma_{\overline{X}} = \sqrt{0.5^2}$$

$$\Rightarrow \sigma_{\overline{X}} = 0.5^2$$

c) esté entre 111 y 130 personas?
d) no esté entre 111 y 130 personas?

2. Un grupo de 100 estudiantes tiene una calificación promedio de 70 puntos con una desviación estándar de cuatro. Si se toma al azar una muestra de 16 estudiantes, ¿cuál es la probabilidad de que la media muestral sea:

a) mayor o igual a 72 puntos?
b) menor o igual a 67 puntos?
c) no esté entre 67 y 72 puntos?
d) esté entre 67 y 72 puntos?

3. Se sabe que los paquetes de azúcar marcados con 1 kg tienen un peso promedio de 1.02 kg con una desviación normal de 0.03. ¿Cuál es la probabilidad de seleccionar una muestra al azar de tamaño 20 paquetes con una media muestral de 1.03 kg o más?

4. El peso promedio de los equipajes de los pasajeros que viajan por avión es de 30 kg, con una desviación normal de 5. ¿Cuál es la probabilidd de que haya 25 personas cuyo equipaje dé un total de 800 kg o más?

5. La calificación promedio de los estudiantes de una universidad (en escala de 100) es de 69 puntos con una desviación estándar de 25. ¿Cuál es la probabilidad de seleccionar una muestra aleatoria de 25 estudiantes que tengan, en promedio, 68.2 puntos o menos.

Corroboración del teorema central del límite (*p*)

Supóngase que en una caja hay 4 perlas blancas y 6 negras $\Rightarrow$ la proporción de perlas blancas es $\pi = \frac{4}{10}$

$N = 10$

$$\mu_X = \pi \qquad \sigma_X^2 = \pi\,(1 - \pi)$$

$$\Rightarrow \mu_X = \frac{4}{10} \qquad \Rightarrow \sigma_X^2 = \frac{4}{10}\left(1 - \frac{4}{10}\right)$$

$$= \frac{4}{10}\left(\frac{6}{10}\right)$$

$$\Rightarrow \boxed{\sigma_X^2 = \frac{24}{100}} \qquad \sigma_X^2 = 8\,\frac{(3)}{100}$$

$$\sigma_X^2 = 0.08(3)$$

Sea $n = 5$

El total de muestras posibles es $\binom{10}{5} = 252$

Si queremos que 4 sean blancas y 1 negra $\Big\}$ $\binom{4}{4}\binom{6}{1} =$ 6 casos posibles

3 b, 2 n $\Big\}$ $\binom{4}{3}\binom{6}{2} = 60$

2 b, 3 n $\Big\}$ $\binom{4}{2}\binom{6}{3} = 120$

1 b, 4 n $\Big\}$ $\binom{4}{1}\binom{6}{4} = 60$

0 b, 5 n $\Big\}$ $\binom{4}{0}\binom{6}{5} = 6$

252

Muestra							
blancas	negras	p	f	pf	$p-\mu_p$	$(p-\mu_p)^2$	$(p-\mu_p)^2 f$
0	5	0.0	6	0.0	− 0.4	0.16	0.96
1	4	0.2	60	12.0	− 0.2	0.04	2.40
2	3	0.4	120	48.0	0.0	0.00	0.00
3	2	0.6	60	36.0	0.2	0.04	2.40
4	1	0.8	6	4.8	0.4	0.16	0.96
			252	100.0			6.72

$$\mu_p = \frac{\Sigma\, pf}{\Sigma f}$$

$$\Rightarrow \mu_p = \frac{100.0}{252} \Rightarrow \mu_p = 0.4 \qquad \text{por tanto,} \qquad \mu_p = \frac{4}{10}$$

$$\sigma_p^2 = \frac{\Sigma\,(p-\mu_p)^2 f}{\Sigma f}$$

$$\Rightarrow \sigma_p^2 = \frac{6.72}{252} \Rightarrow \boxed{\sigma_p^2 = \frac{0.08}{3}}$$

Resumiendo:

$N = 10$ $\qquad \mu_X = \dfrac{4}{10}$

$n = 5$ $\qquad \mu_p = \dfrac{4}{10}$ $\qquad \Rightarrow \boxed{\mu_p = \mu_X}$

pero $$\frac{\sigma_X^2}{n} \times \frac{N-n}{N-1} = \frac{0.08(3)}{5} \times \frac{10-5}{10-1}$$

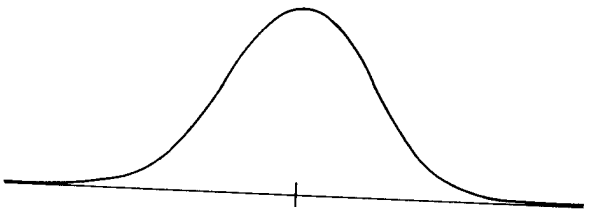
$$= \frac{0.08(3)}{5} \times \frac{5}{9}$$

$$\Rightarrow \frac{\sigma_X^2}{n} \times \frac{N-n}{N-1} = \frac{0.08}{3} \quad \text{y como} \quad \sigma_p^2 = \frac{0.08}{3}$$

Entonces

$$\boxed{\sigma_p^2 = \frac{\sigma_X^2}{n} \times \frac{N-n}{N-1}}$$

Obsérvese que la distribución de p también se acampana.



Ejemplo En una población de familias, 20% se suscribe a la revista K. ¿Cuál es la probabilidad de seleccionar una muestra aleatoria de tamaño 225 con una proporción inferior o igual a 16% (= 0.16)?

$\pi = 20\% \quad \Rightarrow \quad \pi = 0.20$ $\qquad \pi\,(1-\pi) = 0.20\,(0.80) = 0.16$

$N \to \infty \qquad \mu_X = \pi$ $\qquad \sigma_X^2 = \pi\,(1-\pi)$

$\mu_X = 0.20$ $\qquad \sigma_X^2 = 0.16$

$n = 225 \qquad \mu_p = \mu_X$ $\qquad \sigma_p^2 = \dfrac{\sigma_X^2}{n}$

recuérdese que si $N \to \infty \Rightarrow \dfrac{N-n}{N-1} \Rightarrow 1$

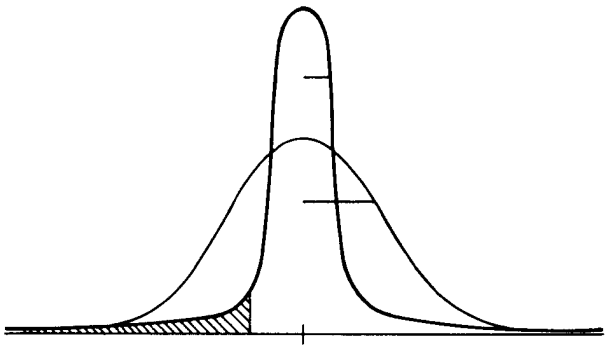
$\mu_p = 0.20$ $\qquad \sigma_p^2 = \dfrac{0.16}{225}$

$\Rightarrow \sigma_p = 0.026667$

$$P[p \leq 0.16] = P\left[\frac{p - \mu_p}{\sigma_p} \leq \frac{0.16 - 0.20}{0.026667}\right]$$

$$\Rightarrow P[p \leq 0.16] = P[Z \leq -1.50]$$

$$\Rightarrow P[p \leq 0.16] = 0.0668$$



Ejemplo De 2 000 consumidores, 40% piensa incrementar sus pedidos de lavadoras. ¿Cuál es la probabilidad de una muestra aleatoria simple de 400 consumidores con una proporción de muestra igual o mayor a 46% (= 0.46)?

$\pi = 40\%$ $\Rightarrow$ $\pi = 0.40$ $\qquad \pi(1 - \pi) = 0.40\,(0.60) = 0.24$

$N = 2\,000$ $\qquad \mu_X = \pi$ $\qquad \sigma_X^2 = \pi(1 - \pi)$

$\mu_X = 0.40$ $\qquad \sigma_X^2 = 0.24$

$n = 400$ $\qquad \mu_p = \mu_X$ $\qquad \sigma_p^2 = \dfrac{\sigma_X^2}{n}\,\dfrac{N - n}{N - 1}$

$\mu_p = 0.40$ $\qquad \sigma_p^2 = \dfrac{0.24}{400} \times \dfrac{2000 - 400}{2000 - 1}$

$$\sigma_p^2 = \frac{0.24}{400} \times \frac{1600}{1999}$$

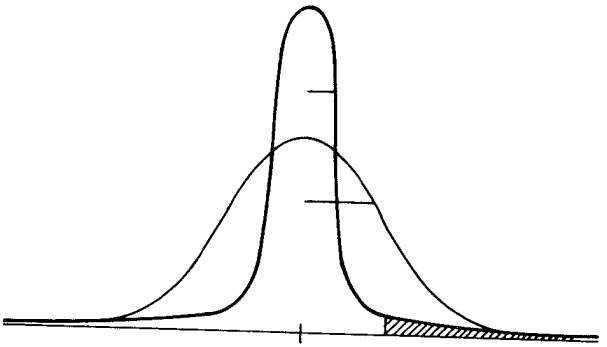
$$\sigma_p = 0.021914$$

$$P[p \geq 0.46] = P\left[\frac{p - \mu_p}{\sigma_p} \geq \frac{0.46 - \mu_p}{\sigma_p}\right]$$

$$= P\left[Z \geq \frac{0.46 - 0.40}{0.021914}\right]$$

$$= P[Z \geq 2.74]$$

$$\Rightarrow P[p \geq 0.46] = 0.0031$$



Ejemplo Se sabe que 20% de los estudiantes de una universidad está afiliado al Partido de la Revolución Democrática (PRD). Si se selecciona una muestra aleatoria de 110 estudiantes, ¿cuál es la probabilidad de encontrar 30 o más estudiantes que sean del PRD?

$\pi = 20\%$ $\Rightarrow$ $\pi = 0.20$ $\qquad \pi(1-\pi) = 0.20\,(0.80) = 0.16$

$N \to \infty$ $\qquad \mu_X = \pi$ $\qquad \sigma_X^2 = \pi(1-\pi)$

$\mu_X = 0.20$ $\qquad \sigma_X^2 = 0.16$

$n = 110$ $\qquad \mu_p = \mu_X$ $\qquad \sigma_p^2 = \dfrac{\sigma_X^2}{n}$

$\mu_p = 0.20$ $\qquad \sigma_p^2 = \dfrac{0.16}{110}$

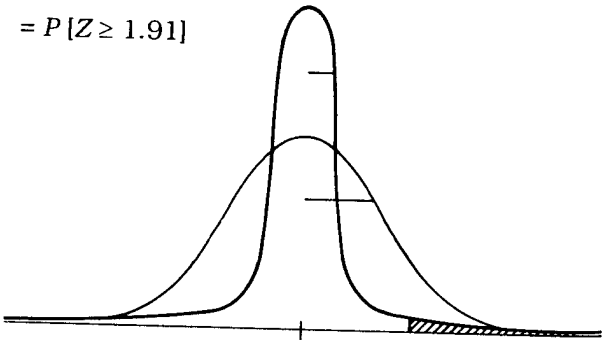
$\sigma_p = 0.038139$

$$P[110p \geq 30] = P\left[p \geq \frac{30}{110}\right]$$

$$= P[p \geq 0.272727] = P\left[\frac{p-\mu_p}{\sigma_p} \geq \frac{0.272727 - 0.20}{0.038139}\right]$$

$$= P[Z \geq 1.91]$$

$$\Rightarrow P[100p \geq 30] = 0.0281$$



Ejercicios

1. Se sabe que 40% de los estudiantes son fumadores. ¿Cuál es la probabilidad de seleccionar un grupo de 64 estudiantes con un porcentaje muestral del 35% o menos?
2. Según registros efectuados, 20% de las víctimas de accidentes de tránsito fallecen. Si hay 150 víctimas durante cierto día, ¿cuál es la probabilidad de que 39 o más víctimas mueran?
3. Se sabe que 10% de los estudiantes no aprueban el curso de estadística: ¿cuál es la probabilidad de que seis o más estudiantes de una muestra de 50 reprueben la materia?
4. Si se lanza una moneda 400 veces, ¿cuál es la probabilidad de que aparezcan 210 o más caras iguales?
5. Se sabe que en promedio 7% de los animales domésticos menores de un año mueren en cierta región. Nacieron 120 cachorros durante un cierto periodo. ¿Cuál es la probabilidad de que 12 o más de los cachorros mueran durante su primer año?

RESUMEN

La estadística tiene por objetivo el estudio de la población, no de hechos aislados. Una población puede tener un conjunto infinito de elementos, aunque en la práctica no se trata de poblaciones con número infinito de elementos, a pesar de que una población muy grande se considera infinita ($n \to \infty$).

Cuando la población es muy grande ($n \to \infty$), resulta imposible observar o medir todos y cada uno de los elementos, debido al tiempo y el costo que esto significa. Para resolver este problema se elige una muestra representativa de la población por medio de alguna de las técnicas consideradas en este capítulo. Si se realiza en forma adecuada lo anterior, los resultados obtenidos al estudiar la muestra serán más precisos que si se hubiera estudiado directamente a la población.

PROBLEMAS

7.1 ¿Cuál es la principal ventaja del muestreo aleatorio, en comparación con el muestreo determinístico?

7.2 *a)* ¿Qué es la población objetivo?
b) ¿Qué es la población adecuada?
c) ¿Qué problema ocurre si no existe una correspondencia cercana entre la población objetivo y la adecuada?

7.3 ¿En qué consiste el error de muestreo?

7.4 Un inversionista está interesado en estimar la cantidad de dinero en promedio que se gasta en diversiones en un área urbana determinada, con

el objeto de promover la construcción de un centro de diversiones. Se consideran cuatro ingresos básicos y la proporción de cada uno de ellos respecto a la población total del área urbana de que se trata:

Menos de $60 000 al mes	20%
$60 000 _____ $100 000 al mes	50%
$100 000 _____ $140 000 al mes	20%
$140 000 o más al mes	10%

Estudios similares en otras áreas urbanas han indicado que la magnitud de dispersión que describe el patrón de gastos por diversión demuestran de manera consistente lo siguiente:

	(σ) magnitud de dispersión	
Menos de $60 000	1	6 K
$60 000 _____ $100 000	1	4 K
$100 000 _____ $140 000	2	K
$140 000 o más	3	K

a) Si se hace un muestreo aleatorio y estratificado y se obtiene una muestra de 160 personas, ¿qué proporción de la muestra debe ser seleccionada para cada uno de dichos estratos?

7.5 *a)* ¿Qué es una distribución muestral de un estadístico?
b) ¿Qué información nos provee el error estándar?

7.6 Si se tiene una población de 1 000 artículos, el número de muestras diferentes de tamaño 400 que puedan seleccionarse es igual a

$$_{1\,000}C_{400} = \frac{1\,000!}{400!\,600!}$$

El número de muestras de tamaño 600 es:

$$_{1\,000}C_{600} = \frac{1\,000!}{400!\,600!}$$

Ello significa que el mismo número de medias muestrales existe para la distribución muestral de la media asociada con todas las muestras posibles, tanto de tamaño 400 como de tamaño 600. ¿Cuál de las dos distribuciones muestrales tendría la medida de dispersión más pequeña?

7.7 Cinco estudiantes tienen, respectivamente, las siguientes sumas de dinero: 2, 5, 3, 4 y 1 pesos.

a) Dibuje el polígono de frecuencias para estos valores X.
b) Considere $N = 5$ estudiantes como una población. Calcule la media poblacional.
c) Calcule la desviación normal de la población.
d) Tome una muestra aleatoria de tamaño $n = 2$ de la población $N = 5$ estudiantes. ¿Cuántas muestras pueden seleccionarse?
e) Haga una lista de todas las muestras posibles y encuentre sus medias muestrales.
f) Calcule el promedio de la media muestral:

$$\overline{X} = \frac{\overline{X}_1 + \ldots}{\binom{N}{n}}$$

y demuestre que $\overline{X} = \mu_X$.

g) Calcule la desviación normal de la media muestral usando la fórmula básica:

$$\sigma_X^2 = \frac{\Sigma\,(X - \mu_X)^2}{n - 1}$$

h) Calcule la σ_X por medio de la fórmula

$$\sigma_X = \frac{\sigma_X}{\sqrt{n}}\sqrt{\frac{N - n}{N - 1}}$$

y demuestre que la respuesta es la misma que se obtuvo en (*g*).

i) Construya una tabla de frecuencias para las medias muestrales con intervalos de clase que vayan de \$1 a \$1.99, de \$2 a \$2.99, etcétera.
j) Dibuje un histograma a partir de la tabla obtenida en (*i*) y observe cuidadosamente cómo se muestra en él la distribución muestral.

7.8 Suponga que hay seis estudiantes que tienen las siguientes cantidades de dinero: \$2, \$5, \$3, \$4, \$1 y \$9.
Tome una muestra de tamaño $n = 2$ y efectúe las 10 etapas del problema anterior.

7.9 En el problema 7.7, ¿cuál es la probabilidad de seleccionar una muestra con media muestral igual o mayor a \$4?

7.10 En el problema 7.8, ¿cuál es la probabilidad de seleccionar una muestra con media muestral igual o mayor a \$6?

7.11 Dada la población 1, 2, 6:

a) Haga una lista de todas las muestras posibles de tamaño $n = 2$, tomándolas con reemplazamiento.

b) Calcule $E(\overline{X})$ utilizando los resultados de (*a*) y observe que esto es igual a $\mu_X = 3$.

c) Calcule var $(\overline{X}) = E(\overline{X} - \mu_X)^2$ mediante los resultados de (*a*). Observe que en este caso hay N^2 valores posibles para $\overline{X}$.

d) Calcule $\dfrac{\sigma_X^2}{n}$ y observe que esto es igual a var $(\overline{X})$ obtenida en (*c*).

7.12 Dada la población 1, 2, 3, 6, ejecute las operaciones del problema anterior y observe que en este caso es sin reemplazo.

7.13 Dada la población 1, 3, 4, 8, seleccione muestras de tamaño $N = 2$. Encuentre var $(X) = \sigma_{\overline{X}}^2$ por medio de las fórmulas:

a) $\sigma_X^2 = \dfrac{(N-n)\sigma^2}{(N-1)n}$

b) $\sigma_X^2 = \dfrac{\sigma^2}{n}$

Indique cuándo deben emplearse.

7.14 Dada la población 1, 2, 3, 4, 5, seleccione muestras de tamaño $N = 3$ y demuestre que:

a) $E\left[\dfrac{n}{n-1}\, s^2\right] = \dfrac{N}{N-1}\, \sigma_X^2,$

cuando el muestreo es con remplazo;

b) $E\left[\dfrac{n}{n-1}\, s^2\right] = \sigma_X^2,$

cuando el muestreo es sin reemplazamiento.

7.15 Dada la población, 4, 9, 5, 1, 3, 7, 3, 8, 7, 1:

a) Encuentre μ_X.

b) Seleccione una muestra sin reemplazamiento de tamaño $N = 6$ y calcule $\overline{X}$.

c) Seleccione una muestra con reemplazamiento de tamaño $N = 6$ y calcule $\overline{X}$.

d) Calcule σ_X^2.

e) Utilizando la muestra de (*b*), encuentre s^2.

f) Utilizando la muestra de (*c*), encuentre s^2.

g) Explique la relación que hay entre s^2 y σ_X^2.

h) A medida que n crece, observe que s^2 tiende al valor de σ_X^2.

Parte 4

Inferencia estadística

Conceptos básicos 8

PROPÓSITOS

El propósito central de este capítulo es que el lector aprenda las nociones básicas de la inferencia estadística y las aplique a la solución de problemas concretos.

Del mismo modo, el lector podrá:

1. Describir la estimación puntual.
2. Explicar las siguientes propiedades básicas de la estimación puntual: insesgada, eficiente, mínimo error cuadrado, suficiente y consistente.
3. Desarrollar el procedimiento para obtener un estimador puntual mediante la solución de ejercicios acorde con sus propiedades básicas.
4. Precisar en qué consiste el error estándar de la media.
5. Desarrollar los procedimientos de cálculo del error estándar de la medida con un gran número de muestras escogidas al azar.
6. Identificar la función y estimar la desviación estándar de la distribución muestral.
7. Explicar tanto el intervalo de confianza como su función.
8. Reconocer la importancia del cálculo del intervalo de confianza en cuanto a precisión y confiabilidad de estimador.
9. Calcular el intervalo de confianza mediante los pasos propuestos en el texto.
10. Explicar la naturaleza de las hipótesis, así como su importancia en el quehacer científico.
11. Reconocer la diferencia entre la siguiente hipótesis estadística, la científica, la nula y la alternativa.

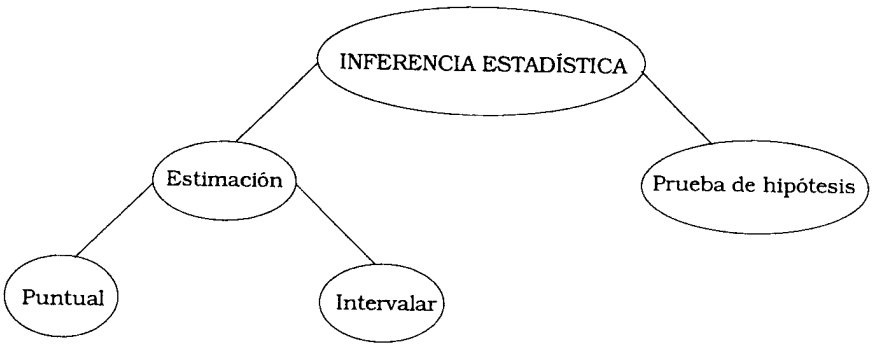
12. Comprender la lógica de la contrastación de hipótesis.
13. Reconocer el significado del concepto probabilidad de rechazo o no rechazo de una hipótesis en el proceso de investigación.
14. Explicar los errores de tipo I y II y su importancia en la aceptación o el rechazo de hipótesis.
15. Explicar las características de los dos tipos de inferencia estadística en el contexto de la distribución muestral.
16. Enumerar las características de los tipos de hipótesis que se clasifican en el texto.
17. Describir las reglas operativas para efectuar contraste de una hipótesis.
18. Ubicar en la curva característica de operación la región de rechazo y de no rechazo de H_0.
19. Aplicar la relación entre el tamaño de muestra y la prueba de hipótesis.
20. Desarrollar los procedimientos básicos en la solución de problemas específicos de investigación en la aceptación o no de hipótesis.

INTRODUCCIÓN

La estadística ha desarrollado técnicas y procedimientos para generalizar, relacionados con los parámetros de una población con base en la información contenida en una muestra representativa de dicha población.

Existen dos métodos de inferencia en el enfoque clásico: la *estimación* y la *prueba* o *contraste de hipótesis*. A su vez, la estimación es de dos formas: la *puntual* y la *intervalar*. Para realizarla, es necesario proponer estimaciones de los valores de los parámetros, debido a que resulta imposible estudiar a toda la población. Estas estimaciones estarán sujetas a un error (la diferencia entre el valor del parámetro de la población y el valor del estadístico de la muestra utilizado como estimador). La probabilidad de cometer este error puede calcularse.

La otra forma de inferir es por medio de las *pruebas de hipótesis*. En este caso se establece una hipótesis con respecto al valor de las características de los parámetros con la información generada en una muestra. Si la evidencia no es consistente con la hipótesis propuesta, ésta se rechaza.



El parámetro se da como resultado a partir de la distribución del estadístico y con base en el valor de éste en la muestra.

Tanto en la comprobación de hipótesis como en la estimación, debe conocerse la distribución del estadístico para realizar una inferencia.

La diferencia entre el contraste de hipótesis y la estimación por intervalo consiste en que en la primera se establece una hipótesis acerca del parámetro, antes de realizar el estudio; con fundamento en el resultado del estadístico muestral se rechaza o no dicha hipótesis. En cambio, en la estimación por intervalo se consideran todos los posibles valores del parámetro.

8.1 ESTIMACIÓN PUNTUAL

El estimador puntual, o de punto, es un instrumento estadístico que intenta asignar un solo valor lo más cercano posible al valor verdadero del parámetro de la población. Dicho estimador es una variable aleatoria, por lo que resulta necesario considerar las propiedades de dichas variables para obtener el mejor estimador. Algunas de estas propiedades son:

1. ***Insesgado*** o ***sin vicio.*** Cuando el estimador *es igual* al valor del parámetro se llama *insesgado*; cuando *no lo es* se denomina *estimador sesgado.*
2. ***Eficiente.*** Se preferirá al estimador cuya varianza sea la mínima con respecto a otros estimadores del mismo parámetro. Cuando dos estimadores tengan el mismo sesgo, será mejor considerar el de mínima varianza y, si ésta es la misma, se preferirá el insesgado. Cuando un estimador tiene la varianza mínima y es insesgado se le denomina *eficiente.*
3. ***Suficiente.*** Se dice que un estimador es *suficiente* con relación a un parámetro si utiliza toda la información relevante de la muestra para calcularlo.
4. ***Consistente.*** Cuando la probabilidad de que un estimador difiera del valor del parámetro, permanezca constante o sea muy pequeña a medida que el tamaño de la muestra se incremente, se dice que el estimador es *consistente.*

El procedimiento usual para obtener un estimador puntual es el siguiente:

1. Se selecciona la muestra de una población. La técnica adecuada de muestreo repercutirá en minimizar el error (la diferencia entre el valor del parámetro y el estadístico).
2. Se calcula el estadístico muestral correspondiente al parámetro poblacional que se estimará.
3. Se utiliza este estadístico como una estimación del parámetro poblacional y se verifican las propiedades anteriores que cumple.

Ejemplo 1 Si se dan los siguientes valores de una población, $X = 1, 2, 4, 4, 7$; $N = 5$

1. Obtenga μ_X.
2. Seleccionando sin reemplazo, obtenga todas las muestras posibles de tamaño 3.
3. Calcule cada una de las medias muestrales $\overline{X}$, para las muestras obtenidas anteriormente.
4. Compruebe que $\mu_{\overline{X}} = E[\overline{X}] = \mu_X$
5. Calcule el error *a)* $e = \mu_X - \overline{X}_4$
   b) $e = \mu_{\overline{X}} - \overline{X}_4$

Solución

1. $$\mu_X = \frac{\sum_{i=1}^{5} X_i}{N} = \frac{1+2+4+4+7}{5} = 3.60;\ \mu_X = 3.60$$

2. Si $N = 5$ y $n = 3$

$$_N C_n = {}_5C_3 = \frac{5!}{3!\,2!} = 10 \text{ muestras de tamaño 3}$$

a) (1, 2, 4)	e) (1, 2, 4)	*i)* (2, 4, 4)
b) (1, 2, 7)	*f)* (2, 4, 7)	*j)* (7, 4, 4)
c) (2, 4, 7)	*g)* (1, 7, 4)	
d) (1, 7, 4)	*h)* (1, 4, 4)	

3. $$\overline{X}_1 = \frac{1+2+4}{3} = \frac{7}{3} \qquad \overline{X}_6 = \frac{2+4+7}{3} = \frac{13}{3}$$

$$\overline{X}_2 = \frac{1+2+7}{3} = \frac{10}{3} \qquad \overline{X}_7 = \frac{1+7+4}{3} = \frac{12}{3}$$

$$\overline{X}_3 = \frac{2+4+7}{3} = \frac{13}{3} \qquad \overline{X}_8 = \frac{1+4+4}{3} = \frac{9}{3}$$

$$\overline{X}_4 = \frac{1+7+4}{3} = \frac{12}{3} \qquad \overline{X}_9 = \frac{2+4+4}{3} = \frac{10}{3}$$

$$\overline{X}_5 = \frac{1+2+4}{3} = \frac{7}{3} \qquad \overline{X}_{10} = \frac{7+4+4}{3} = \frac{15}{3}$$

4. Como $\mu_{\overline{X}} = E(\overline{X}) = \frac{1}{m}\left(\sum_{i=1}^{m} X_i\right)$ donde $m = 10$ son todas las muestras posibles obtenidas de la población

$$X = 1, 2, 4, 4, 7; \qquad N = 5$$

$$\mu_{\overline{X}} = E[\overline{X}] = \frac{1}{10}\left(\frac{7}{3} + \frac{10}{3} + \frac{13}{3} + \frac{12}{3} + \frac{7}{3} + \frac{13}{3} + \frac{12}{3} + \frac{9}{3} + \frac{10}{3} + \frac{15}{3}\right)$$

$$= \frac{1}{10}\left(\frac{108}{3}\right)$$

$$\mu_{\overline{X}} = E[\overline{X}] = 3.60$$

$$\therefore \mu_{\overline{X}} = E[\overline{X}] = \mu_X$$

5. *a)* $e = \mu_X - \overline{X}_4 = (3.60) - \frac{12}{3} = -0.4$

b) $e = \mu_{\overline{X}} - \mu_X = 0$

Se define a S_n^2 como $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

S^2 es un estimador sesgado de σ_X^2 porque

$$E\left[S_n^2\right] = \left(\frac{n-1}{n}\right)\sigma_X^2$$

pero si definimos a $\hat{\sigma}_X^2$ como $\hat{\sigma}_X^2 = \frac{n}{n-1} S_n^2$ entonces

$$E\left[\hat{\sigma}_X^2\right] = E\left[\frac{n}{n-1} s_n^2\right] = \frac{n}{n-1}, E\left[s_n^2\right] = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma_X^2 = \sigma_X^2$$

$\therefore E\left[\hat{\sigma}_X^2\right] = \sigma_X^2$, o sea $\hat{\sigma}_X^2$, es un estimador insesgado de σ_X^2.

Ahora la varianza de $\overline{X}$ es, en general

$$\sigma_{\overline{X}}^2 = E[(\overline{X} - \mu_{\overline{X}})^2] = \frac{\sigma_X^2}{n} \times \frac{N-n}{N-1}$$

Ejemplo 2 De la población anterior

$$X_1 = 1,\ X_2 = 2,\ X_3 = 4,\ X_4 = 4 \text{ y } X_5 = 7$$

$$N = 5 \text{ y } \mu_X = 3.6;\ n = 3 \text{ y } \mu_{\overline{X}} = 3.60$$

Calcular:

1. σ_X^2
2. $\sigma_{\overline{X}}^2$
3. Compruebe que $\sigma_{\overline{X}}^2 = \dfrac{\sigma_X^2}{n} \times \dfrac{N-n}{N-1}$
4. Las varianzas muestrales S_n^2 de cada una de las muestras.
5. Compruebe que $E\left[\hat{\sigma}_X^2\right] = \sigma_X^2$

Solución

1. $\sigma_X^2 = E\,[(X - \mu_X)^2] = \dfrac{1}{N}\sum_{i=1}^{N} (X_i - \mu_X)^2$

$$= \frac{1}{5}\Big[\,(1-3.60)^2 + (2-3.60)^2 + (4-3.60)^2 + (4-3.60)^2 + (7-3.60)^2)\,\Big]$$

$$= \frac{1}{5}\,(6.76 + 2.56 + 0.16 + 0.16 + 11.56)$$

$$= \frac{1}{5}\,(21.2)$$

$$\Rightarrow \sigma_X^2 = E\left[(X-\mu_X)^2\right] = 4.24$$

2. $\sigma_{\overline{X}}^2 = E\left[(\overline{X} - \mu_{\overline{X}})^2\right] = \dfrac{1}{m}\sum_{i=1}^{m} (\overline{X}_i - \mu_{\overline{X}})^2$ donde m es el total de muestras

$$= \frac{1}{10}\left(\left(\frac{7}{3}-3.60\right)^2 + \left(\frac{10}{3}-3.60\right)^2 + \left(\frac{13}{3}-3.60\right)^2 + \left(\frac{12}{3}-3.60\right)^2 + \left(\frac{7}{3}-3.60\right)^2 + \right.$$

$$\left.\left(\frac{13}{3}-3.60\right)^2 + \left(\frac{12}{3}-3.60\right)^2 + \left(\frac{9}{3}-3.60\right)^2 + \left(\frac{10}{3}-3.60\right)^2 + \left(\frac{15}{3}-3.60\right)^2\right)$$

$$= \frac{1}{10}\left(\frac{14.44}{9} + \frac{0.64}{9} + \frac{4.84}{9} + \frac{1.44}{9} + \frac{14.44}{9} + \frac{4.84}{9} + \frac{1.44}{9} + \frac{3.24}{9} + \frac{0.64}{9} + \frac{17.64}{9}\right)$$

$$= \frac{1}{10}\left(\frac{63.9}{9}\right)$$

$$\Rightarrow \quad \sigma_{\overline{X}}^2 = 0.7066666667$$

3. Como $\sigma_X^2 = 4.24$; $N = 5$; $n = 3 \Rightarrow \dfrac{\sigma_X^2}{n} \times \dfrac{N-n}{N-1} = \dfrac{4.24}{3} \times \dfrac{5-3}{5-1}$

$$\Rightarrow \frac{\sigma_X^2}{n}\frac{N-n}{N-1} = \frac{4.24}{3} \times \frac{5-3}{5-1} = \frac{4.24}{3} \times \frac{2}{4} = 0.706\ 666\ 666\ 7$$

$$\Rightarrow \boxed{\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}\frac{N-n}{N-1}}$$

4. $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ y $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \times \dfrac{N-1}{N}$ si la población es pequeña

5. $E\left[s_n^2\right] = \dfrac{1}{m}\sum_{i=1}^{m} S_i^2$

$$E\left[S_n^2\right] = \frac{1}{10}\left(\frac{\frac{42}{9}}{3} + \frac{\frac{18}{9}}{3} + \frac{\frac{114}{9}}{3} + \frac{\frac{162}{9}}{3} + \frac{\frac{42}{9}}{3} + \frac{\frac{114}{9}}{3} + \frac{\frac{162}{9}}{3} + \frac{\frac{54}{9}}{3} + \frac{\frac{24}{9}}{3} + \frac{\frac{54}{9}}{3}\right)$$

$$= \frac{1}{10}\left(\frac{\frac{954}{9}}{3}\right) = \frac{1}{10}\left(\frac{106}{3}\right)$$

$$\Rightarrow E\left[S_n^2\right] = \frac{106}{30} \Rightarrow E\left[S_n^2\right] = E\left[\frac{N-1}{N}S_n^2\right] = \frac{N-1}{N}E\left[S_n^2\right]$$

$$E\left[\hat{\sigma}_x^2\right]\frac{n}{n-1} \times \frac{N-1}{N} \quad E\left[S_n^2\right] = \frac{3}{3-1} \times \frac{5-1}{5}\left(\frac{106}{30}\right) = \frac{3}{2} \times \frac{4}{5}\left(\frac{106}{30}\right) = 4.24$$

$$\Rightarrow E\left[\hat{\sigma}_X^2\right] = \sigma_X^2$$

No obstante que (σ) es un estimador sesgado, resulta preferible utilizarlo en algunos análisis estadísticos.

8.1.1 Error estándar de la media muestral

Sean μ_X y σ_X la media y la desviación estándar, respectivamente, de una población cuya variable de interés es X. Entonces, la desviación estándar de la distribución muestral de la media $\overline{X}$ es $\sigma_{\overline{X}}$ (una medida de dispersión de las medias muestrales, y $E(\overline{X}) = \mu_X$).

Tanto σ_X como $\sigma_{\overline{X}}$ son desviaciones estándar de sus respectivas distribuciones y, para distinguirlas, hay que conferirles diferentes nombres. Por esto, a $\sigma_{\overline{X}}$ suele llamársele el *error estándar de la media* (debido a que la dispersión de la distribución muestral se atribuye al error de muestreo).

La relación entre $\sigma_{\overline{X}}$ y σ_X es la siguiente:

$\sigma_{\overline{X}} = \sigma_X \times \frac{1}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$ donde $\sqrt{\frac{N-n}{N-1}}$ es el factor de corrección para poblaciones finitas (fcf).

Para una población muy grande o infinita $N \to \infty$:

$\sqrt{\frac{N-n}{N-1}} = 1$ y $\sigma_{\overline{X}} = \sigma_X \times \frac{1}{\sqrt{n}}$, que se suele presentar como $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$, simplificando se tiene: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Pero para poblaciones pequeñas (finitas), de allí el nombre del factor de corrección para población finita cpf $\neq 1$ y $\sigma_{\overline{X}} = \sigma_X \times \frac{1}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$

Ejemplo (N pequeña):

si $\sigma_X = 10$
$n = 17$
al obtener
una muestra $n = 16$

calcular $\sigma_{\overline{X}}$

$$\sigma_{\overline{X}} = \sigma_X \times \frac{1}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\overline{X}} = 10 \times \frac{1}{\sqrt{16}} \sqrt{\frac{17-16}{17-1}} = 2.50 \times 0.25$$

$$\boxed{\sigma_{\overline{X}} = 0.625}$$

Ejemplo *(N grande):*

$\sigma_X = 10$

$N = 10\,001$

$n = 16$

calcular $\sigma_{\overline{X}}$

aplicando

$$\sigma_{\overline{X}} = \sigma_X \times \frac{1}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

se tiene:

$$\sigma_{\overline{X}} = 10 \times \frac{1}{\sqrt{16}} \sqrt{\frac{10\,001 - 16}{10\,001 - 1}}$$

$$\sigma_{\overline{X}} = 2.5 \sqrt{\frac{9985}{10\,000}} = 2.5\,(0.99925) \approx 2.5$$

$$\sigma_{\overline{X}} = 2.5$$

$$\therefore\ \sigma_{\overline{X}} = \sigma_X \times \frac{1}{\sqrt{n}}$$

8.2 ESTIMACIÓN INTERVALAR

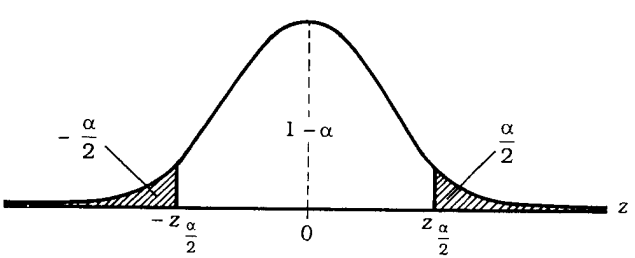
Es una forma de establecer la precisión o confiabilidad de un estimador puntual. La construcción de este intervalo a partir de la información observada y recopilada de una muestra provee una banda alrededor del parámetro estimado, asegurando con una probabilidad determinada que dicho parámetro esté ubicado dentro del intervalo. A dicho intervalo se le conoce de varias maneras: *de confianza, estimado de confianza, región de confianza, IC* (intervalo de confianza). La probabilidad determinada se denomina *coeficiente de confianza.*

Aun cuando únicamente se requiera el valor de una estimación puntual, es conveniente construir un IC.

En pocas palabras, un intervalo de confianza consiste en un par de valores ***a*** y ***b***, entre los cuales está contenido el verdadero valor del parámetro por estimar, con una probabilidad específica $(1-\alpha)$.

Si se desea construir un intervalo de confianza de la media poblacional μ_X, sería $\boldsymbol{a} \leq \mu_X \leq \boldsymbol{b}$, donde ***a*** y ***b*** dependerán de:

1. n, el tamaño de la muestra con que se cuenta.
2. El coeficiente de confianza $(1-\alpha)$ que se desee: 99%, 95%, 90%, 80%, etcétera.
3. $\overline{X}$, el valor del estimador puntual muestral de μ_X.



$$a = \overline{X} - Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \quad \text{y} \quad b = \overline{X} + Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

El intervalo es: $(\overline{X} - Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}})$

Si $\alpha = 5\% = 0.05$, entonces $1 - \alpha = 1 - 0.05 = 0.95$ o 95% y $\alpha/2 = 0.025$

Al consultar la tabla 1 del Apéndice se encuentra que $Z_{0.025} = +1.96$ y $-Z_{0.025} = -1.96$, el IC al 95%, será:

$$P\left(\overline{X} - 1.96\frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + 1.96\frac{\sigma_X}{\sqrt{n}}\right) = 0.95$$

Esto significa que la probabilidad de que μ_X se encuentre en el intervalo $\overline{X} \pm 1.96\,\sigma_{\overline{X}}$ asciende a 0.95, es decir, puede ocurrir 95 de cada 100 veces que μ_X se encuentre entre $\overline{X} - 1.96\,\sigma_{\overline{X}}$ y $\overline{X} + 1.96\,\sigma_{\overline{X}}$ (recuérdese que $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$).

Se puede afirmar que se tiene una confianza de 95%, en el sentido de que dicho intervalo comprenderá a μ_X, con un error $e = (\mu_X - \mu_{\overline{X}}$ o sea $e = Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}})$.

Ejemplo A 100 estudiantes universitarios se les aplica una prueba psicológica con la que se mide el coeficiente de inteligencia, y se obtiene un promedio muestral de 110, con una desviación estándar poblacional conocida de 20.

Construir un IC de μ_X, con el valor del promedio muestral $\overline{X}$, al 95%.

DATOS

$\overline{X} = 110$

$n = 100$

$\sigma_X = 20$

$1 - \alpha = 95\% = 0.95$

$\alpha = 5\% = 0.05$

$\alpha/2 = 0.025$

como

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

$$\sigma_{\overline{X}} = \frac{20}{\sqrt{100}} = \frac{20}{10} = 2$$

utilizando la tabla 1 del Apéndice

$Z_{0.025} = Z_{1-0.025} = |1.96| \therefore Z_{\alpha/2} = 1.96$

$\overline{X} - 1.96\,\sigma_X \leq \mu \leq \overline{X} + 1.96\,\sigma_{\overline{X}}$

$110 - 1.96\,(2) \leq \mu \leq 110 + 1.96\,(2)$

$110 - 3.92 \leq \mu \leq 110 + 3.92$

$P(106.08 \leq \mu \leq 113.92) = 0.95$

Al calcular el error e

$e = 1.96\,\sigma_{\overline{X}} = 1.96\,(2) = 3.92$

$e = 3.92$

Este error significa que se puede utilizar a $\overline{X} = 110$ como un estimador de μ_X, con una confianza de 95% de que el error e no excederá 3.92, y con un tamaño de muestra $n = 100$.

En general:

como $e = Z_{\alpha/2} \dfrac{\sigma_X}{\sqrt{n}}$, al despejar n resulta: $n = \left(\dfrac{Z_{\alpha/2}\,\sigma}{e}\right)^2$.

Esto significa que se puede calcular el tamaño de muestra mínimo necesario para un valor determinado del error.

Para el ejemplo anterior:
De qué tamaño debe ser la muestra si se quiere un error del 1.5.

DATOS:

$\sigma_X = 20$

$Z_{0.025} = 1.96$ y $e = 1.5$

sustituyendo en:

$$n = \left(\frac{Z_{\alpha/2}\,\sigma_X}{e}\right)^2$$

$$n = \left(\frac{1.96 \times 20}{1.5}\right)^2 = 683$$

El tamaño mínimo de la muestra debe ser de 683.

8.3 HIPÓTESIS ESTADÍSTICAS Y CIENTÍFICAS

La hipótesis estadística es una afirmación acerca de la distribución de cierta variable aleatoria. En dichas hipótesis se considera el valor de un parámetro correspondiente a la distribución poblacional conocida o supuestamente conocida. A su vez una prueba estadística es un procedimiento para decidir si se rechaza o no la hipótesis estadística considerando el resultado de un experimento aleatorio (con base en el valor observado de la variable aleatoria en una muestra). Lo anterior está en contraposición con los resultados de los experimentos no aleatorios, o sea, determinísticos, utilizados para comprobar una hipótesis científica.

La lógica para contrastar hipótesis científicas se muestra a continuación: al considerar hipótesis en estudio, un solo resultado específico debe ocurrir. Si la ocurrencia predecida es observada cuando el experimento crítico se ha realizado, entonces es apoyada, siempre que no exista error en las mediciones u observaciones. Este tipo de experimentos pueden repetirse en las mismas condiciones.

En el caso de una hipótesis estadística, la situación de rechazo o no rechazo (aceptación) no es muy clara. Como se señaló antes, la hipótesis estadística afirma que una variable aleatoria se distribuye en una forma particular o que un parámetro de su distribución tiene un valor específico, y como existen valores de la variable aleatoria que se observan bajo la hipótesis, ninguna observación puede conducir al rechazo de ésta con certidumbre. En suma, no existe un experimento determinante. De cualquier manera, si ciertos valores de la variable aleatoria, cuando la hipótesis es verdadera, son muy diferentes a los esperados,

ésta se rechaza. La diferencia entre las hipótesis científicas y las estadísticas no es definitiva, ya que de hecho todas las observaciones contienen incertidumbre o errores de medición.

Hipótesis nula (H_0)

La H_0 establece una afirmación acerca del valor de ciertos parámetros poblacionales y por lo general se expresa como la negación de una relación posible entre la variable independiente y la dependiente.

Si H_0 es verdadera, se niega la posibilidad de que la hipótesis de investigación también lo sea. Se supone que H_0 es cierta, a menos que los resultados de la significancia de una prueba estadística posibiliten rechazarla.

Hipótesis alternativa (H_1)

La H_1 se manifiesta acerca del valor de ciertos parámetros poblacionales y se expresa de modo que contradice la hipótesis nula. El rechazo de H_0 conduce al no rechazo de H_1, y a la posibilidad de que la hipótesis de investigación sea cierta.

En general, se propone y contrasta una hipótesis alternativa con la nula para decidir, entre dos posibles acciones, una apropiada si la nula es verdadera y otra si la nula es falsa.

La proposición de la hipótesis nula y de la alternativa determina la zona de rechazo, asignándole una posición en la cola superior, inferior o en ambas de la curva característica de operación; adicionalmente, el investigador debe precisar el tamaño de dicha región, seleccionando el valor de α, la probabilidad de rechazar la hipótesis nula cuando es verdadera (error de tipo I). Determinar el valor de α depende en parte de la confianza o creencia que el investigador tenga en la hipótesis nula. Si está firmemente convencido de la validez de ella, necesita una evidencia muy fuerte en su contra antes de rechazarla. Si, por otro lado, su creencia o confianza en la validez de la nula es débil, entonces requiere poca evidencia para rechazarla. Pero se debe reafirmar que si la nula es falsa y no se rechaza se comete el error de tipo II (β).

Por desgracia, para un tamaño determinado de la muestra, si se disminuye la probabilidad de rechazar falsamente la hipótesis nula, la probabilidad de una falsa aceptación (no rechazo) de la nula se incrementa. Se debe seleccionar α o β con mucho criterio, tomando en cuenta las repercusiones posibles de cada uno de los errores anteriores.

En conclusión, en la inferencia estadística clásica se establecen tanto una hipótesis nula como una alternativa, y el contraste de ésta consiste básicamente en decidir cuál de las dos es verdadera.

Uno de los problemas conceptuales más difíciles al contrastar hipótesis es establecer la hipótesis de una manera apropiada para el efecto. A fin de proponerla de manera correcta requiere una clara comprensión del problema, además de considerar lo siguiente:

1. Si tomara una decisión que acarreara un error muy costoso, debe ser baja la probabilidad de cometer dicho error.
2. Si los resultados obtenidos de una muestra indican un cambio con respecto a una opinión aceptada, debe ser muy baja la probabilidad de que dichos resultados conduzcan a una decisión errónea.

Estas consideraciones son necesarias no porque un investigador desee tomar una decisión incorrecta, sino porque, a causa de la naturaleza del proceso, donde se efectúan inferencias acerca de la población, a partir de una muestra existe la posibilidad de errar.

La hipótesis simple especifica en forma precisa el valor del parámetro en estudio. La compuesta establece un conjunto de valores en lugar de uno solo. Al respecto existe considerable ventaja en plantear un problema de tal manera que la hipótesis nula sea simple; usualmente, la lógica de la situación requiere que la hipótesis alternativa sea compuesta. Para formular un problema, de modo que la hipótesis nula sea simple, se necesita hipotetizar el opuesto exacto de la afirmación que se desea apoyar. Así, si un investigador desea demostrar una diferencia entre grupos, procesos, métodos o efectos, entonces le conviene hipotetizar la no diferencia (esperando rechazar dicha hipótesis nula). El término *nula* se refiere generalmente a cualquier hipótesis cuyo rechazo origine un error de tipo I. Cuando una hipótesis nula se rechaza o no, considerando la información de una muestra, existen cuatro posibles resultados:

1. No rechazar la hipótesis cuando es verdadera.
2. Rechazar la hipótesis cuando es verdadera.
3. No rechazar la hipótesis cuando es falsa.
4. Rechazar la hipótesis cuando es falsa.

La posibilidad 2 se llama *error de tipo I* (α); la 3, *error de tipo II* (β).

La investigación científica es sistemática, controlada, empírica y crítica en relación con las hipótesis de investigación propuestas que vinculan ciertos fenómenos. Estas hipótesis se pronuncian acerca de una supuesta relación entre una o varias variables independientes y una o varias variables dependientes. La hipótesis de investigación puede provenir de una teoría, basarse en observaciones anteriores o constituir un presentimiento de las expectativas del investigador, referentes a su investigación.

No es posible comprobar la hipótesis de investigación directamente, sino que debe traducirse a una estadística, a fin de constrastarla y establecer afirmaciones sobre el valor de ciertos parámetros poblacionales. Éstas son lógicamente derivadas de las de investigación, en tal forma que su rechazo o no posibilita al investigador evaluar la veracidad o falsedad de su investigación.

Por consiguiente, una tarea del investigador consiste en trasladar sus hipótesis de investigación a un conjunto dicotómico de hipótesis estadísticas, que son mutuamente excluyentes y completamente exhaustivas, siempre que la hipótesis nula sea específica o delimitada y la alternativa sea compuesta o combinada; aunque si ambas son específicas, se podrá calcular el valor del error de tipo II (β).

Expresar la hipótesis nula como simple permite una especificación exacta de α, la probabilidad de cometer el error de tipo I, pero sin olvidar que se mantiene control sobre el máximo valor que α pueda tomar, para un valor determinado de n. Esta probabilidad máxima del error de tipo I reduce el control sobre la probabilidad del error de tipo II. No obstante, es más deseable evitar el error de tipo I.

Al especificar α, (0.01, 0.05, 0.10) se representa el riesgo que se corre al rechazar la hipótesis nula cuando es verdadera. Entonces, la probabilidad de no rechazar la nula cuando es verdadera (la decisión correcta es $1 - \alpha$).

8.4 CLASIFICACIÓN DE HIPÓTESIS

A continuación se presentan consideraciones de algunos autores:

1. *Hipo* significa bajo y *tesis* es afirmación, entonces *hipótesis* es "bajo el supuesto o la afirmación".
2. *Hipótesis.* Suposición de una cosa, sea posible o imposible, para sacar de ella una consecuencia.
3. *Hipótesis de trabajo.* La que se establece provisionalmente como base de una investigación, que puede confirmar o negar la validez de aquella (definición de diccionario).
4. *Hipótesis.* Proposición para eliminar la deficiencia de conocimiento, en forma de una declaración con consecuencias deductivas verificables.
   Por su importancia distinguimos los siguientes tipos:
   a) *Hipótesis descriptiva.* Especifica forma y propiedades de los procesos y objetos.
   b) *Hipótesis explicativa.* Establece la forma de asociación o causalidad (necesaria, suficiente, contribuyente o determinística) entre dos o más propiedades de los objetos o procesos. I. Méndez, en *Ciencia*, núm. 40, 1989, págs. 39-48.
5. *Hipótesis.* Forma tentativa de llenar el conocimiento.
   a) *Hipótesis explicativa.* En términos de relaciones de causalidad (contesta a un por qué).
   b) *Hipótesis descriptiva.* En términos de ubicación temporal o espacial, determinando magnitudes (contesta dónde, cuándo, cómo, cuánto). R. Rojas Soriano, *El proceso de la investigación científica*, Trillas, México, 1983.
   c) *Hipótesis estadística.* Proposición en forma declarativa.
   Es una afirmación que se plantea tentativamente, como guía para la investigación. Las hipótesis dadas están sujetas a comprobación para ser aceptadas o rechazadas; en general se inducen de hechos u observaciones. Lo que en realidad sucede cuando se crea una hipótesis es pasar de una situación particular a una general, proceso que se conoce con el nombre de inducción; pero tal proceso no termina allí, porque una vez hecha la hipótesis supuesta general, al aplicarse a otro fenómeno similar y extraer consecuencias de ella, se regresa a una situación particular, proceso que se conoce como deducción. Esto en sí marca el método de trabajo con el que se crea la ciencia y consiste en una sucesión de procesos inductivos y deductivos.

6. *Ley.* Una hipótesis cuyas predicciones se han visto confirmadas muchas veces (significa la confianza que tenemos en ella).

   Pero conviene tener presente que todas las leyes son abstracciones de la realidad, en las que aparecen los factores relevantes de la situación o fenómeno, pero en las que se omite una variedad de otros factores que harán su aparición en cualquier situación real.
7. *Ley científica.* Es una expresión que afirma en forma cualitativa o de preferencia cuantitativa, relaciones funcionales entre dos o más variables. H, Riveros, L. Rosas, *El método científico aplicado a las ciencias experimentales,* Trillas, México, 1980.
8. *Teoría.* Serie de leyes que sirven para relacionar determinado orden de fenómenos. Hipótesis cuyas consecuencias se aplican a toda una ciencia o parte muy importante de la misma.

   El análisis estadístico nos permite detectar una hipótesis falsa y descartarla, pero no permite establecer en forma definitiva si ésta es verdadera. En la ciencia no existe una verdad final.

8.5 PRUEBA DE HIPÓTESIS

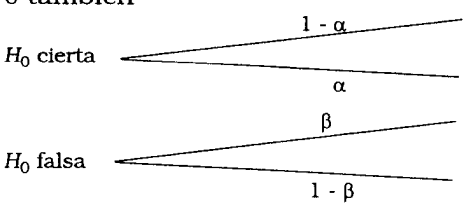
Regla convencional para comprobar o contrastar hipótesis estadísticas: establecer α (probabilidad de rechazar falsamente H_0) igual a un valor lo más pequeño posible; a continuación, de acuerdo con H_1, escoger una región de rechazo tal que la probabilidad de observar un valor muestral en esa región sea igual o menor que α cuando H_0 es cierta.

Una forma esquemática de presentar estos resultados sería la siguiente:

8.5.1 Error de tipo I y II

		Situación real	
		H_0 *cierta*	H_0 falsa
Decisión	No rechazar H_0	Decisión correcta $1-\alpha$	β Error de tipo II
	Rechazar H_0	α Error de tipo I	Decisión correcta $1-\beta$

o también



No rechazar H_0 —— DC $(1-\alpha)$[a]

Rechazar H_0 —— Error de tipo I (α)[b]

No rechazar H_0 —— Error de tipo II (β)[c]

Rechazar H_0 —— DC* $(1-\beta)$

[a] DC: Decisión correcta.
[b] También se denota E_I.
[c] También se denota E_{II}.

Ejemplo de prueba de hipótesis y error de tipo I y de tipo II.

Un individuo y la Naturaleza

Acción	Estado del Universo: W_1 Lluvioso	W_2 Despejado
A_1: Traer paraguas	✓	E_{II}
A_2: No traer paraguas	E_I	✓

Si traigo paraguas y llueve: acerté
Si traigo paraguas y no llueve: error

Si no traigo paraguas y llueve: error
Si no traigo paraguas y no llueve: acerté

Acción	Estado del Universo: W_1 Despejado	W_2 Lluvias
A_1: No traer paraguas	✓	E_{II}
A_2: Traer paraguas	E_I	✓

Si no traigo paraguas y no llueve: acerté
Si no traigo paraguas y llueve: error

Si traigo paraguas y no llueve: error
Si traigo paraguas y llueve: acerté

Un individuo y un juez

Acción	Estado del Universo: W_1 Inocente	W_2 Culpable
A_1: Liberarlo	✓	E_{II}
A_2: Encarcelarlo	E_I	✓

Si lo liberan y es inocente: acertó el juez
Si lo liberan y es culpable: error

Si lo encarcelan y es inocente: error
Si lo encarcelan y es culpable: acertó el juez

Acción	Estado del Universo: W_1	W_2
A_1:	✓	E
A_2:	E	✓

Llenar los espacios

H_0:

(Hipótesis nula)

Acción	Estado del Universo: W_1 Verdadera	W_2 Falsa
A_1: No rechazar	✓	E_{II}
A_2: Rechazar	E_I	✓

No existe el dual para este planteamiento

El error de tipo I (E_I) es rechazar la hipótesis nula (H_0) ya que es verdadera.
El error de tipo II (E_{II}) es no rechazar la hipótesis nula (H_0) ya que es falsa.

$$P\begin{bmatrix}\text{se materialice}\\ \text{el error de tipo I}\end{bmatrix} = \alpha \text{ (nivel de significación)}$$

$$P\begin{bmatrix}\text{no se materialice}\\ \text{el error de tipo I}\end{bmatrix} = 1 - \alpha \text{ (nivel de confianza)}$$

$$P\begin{bmatrix}\text{se materialice}\\ \text{el error de tipo II}\end{bmatrix} = \beta$$

$$P\begin{bmatrix}\text{no se materialice}\\ \text{el error de tipo II}\end{bmatrix} = 1 - \beta \text{ (potencia de la prueba)}$$

$\therefore \alpha = P[E_1]$

$\Rightarrow \boxed{\alpha = P[\text{Rechazar } H_0 \mid H_0 \text{ es verdadera}]}$

$1 - \alpha = 1 - P[\text{Rechazar } H_0 \mid H_0 \text{ es verdadera}]$

$\Rightarrow \boxed{1 - \alpha = P[\text{No rechazar } H_0 \mid H_0 \text{ es verdadera}]}$

$\beta = P[E_{II}]$

$\Rightarrow \boxed{\beta = P[\text{No rechazar } H_0 \mid H_0 \text{ es falsa}]}$

$1 - \beta = 1 - P[\text{No rechazar } H_0 \mid H_0 \text{ es falsa}]$

$\Rightarrow \boxed{1 - \beta = P[\text{Rechazar } H_0 \mid H_0 \text{ es falsa}]}$

	H_0:	Estado del Universo	
		W_0	W_1
	Acción	Verdadera	Falsa
A_1	No rechazar	$1 - \alpha$	β
A_2	Rechazar	α	$1 - \beta$

α puede reducirse a la mínima expresión si se incrementa el valor crítico (reduciendo el tamaño de la región de aceptación). Sin embargo, la probabilidad de cometer el error de tipo II (β), se incrementará; además, si se disminuye demasiado el valor de α también se reduce la probabilidad de rechazar tanto H_0 cierta como H_0 falsa.

En el caso extremo $\alpha = 0$, no se cometería el error de tipo I y, por consiguiente, la hipótesis nula falsa nunca sería rechazada. Obviamente, no rechazar una H_0 falsa también constituye un error.

Una forma de disminuir el valor de β es aumentando el valor de n; así también se incrementa la potencia del estadístico de prueba, pero se aumenta el costo en recursos y tiempo del estudio.

8.5.2 Curva característica de operación

Si se representa a un parámetro por θ (puede ser μ, σ^2, P, etc.) y un valor determinado del parámetro por θ_0, al establecer la hipótesis nula se tiene:

$$H_0: \theta = \theta_0$$

Pero la hipótesis alternativa podría ser:

$$a)\ H_1 : \theta > \theta_0$$

o

$$b)\ H_1 : \theta < \theta_0$$

o

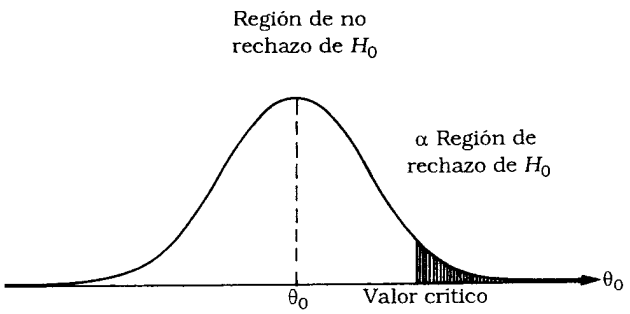
$$c)\ H_1 : \theta \neq \theta_0$$

Donde *a* y *b* suelen denominarse *hipótesis de una cola, unidireccionales* o *de una vía* (*one way*).

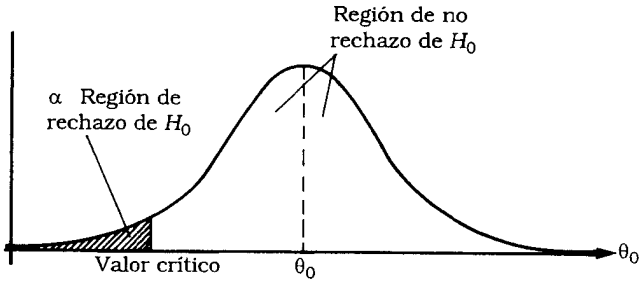
A *c* se le conoce como *hipótesis de dos colas, bidireccional* o *de dos vías* (*two ways*).

Y la curva característica de operación para cada una de las situaciones anteriores sería:

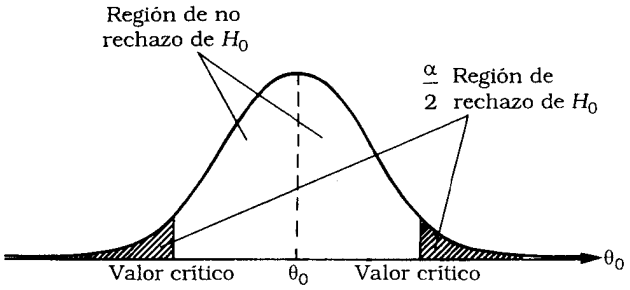
$$1)\ H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$



$$2)\ H_0 : \theta = \theta_0$$
$$H_1 : \theta < \theta_0$$



3) $H_0 : \theta = \theta_0$
$H_1 : \theta \neq \theta_0$



8.5.3 Reglas operativas para realizar un contraste de hipótesis

Paso 1. Identifique la(s) variable(s) de interés, analizando las características de los parámetros que intervienen en el estudio, investigación o experimento.
Paso 2. Proponga las hipótesis estadísticas (H_0 y H_1).
Paso 3. Establezca el valor del error (α o β que está dispuesto a aceptar).
Paso 4. Determine el modelo estadístico adecuado.
Paso 5. Obtenga mediante la tabla correspondiente el valor crítico y ubique la zona de rechazo en una gráfica.
Paso 6. Especifique la regla de decisión para rechazar H_0.
Paso 7. Calcule el valor del estadístico de prueba, utilizando el modelo propuesto.
Paso 8. Obtenga las conclusiones considerando el rechazo o no de H_0.

8.6 RELACIÓN ENTRE TAMAÑO DE MUESTRA Y PRUEBA DE HIPÓTESIS

Basándonos en el teorema central del límite, sabemos que:

1. La media de $\overline{X}$ (media muestral) es μ_X; i.e $\mu_{\overline{X}} = E[\overline{X}] = \mu_X$;
2. La varianza de $\overline{X}$ es $\frac{\sigma_X^2}{n}\frac{N-n}{N-1}$; esto es $\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \times \frac{N-n}{N-1}$

No se pierde generalidad si al factor de corrección para poblaciones finitas lo hacemos 1; $\frac{N-n}{N-1} = 1$, entonces:

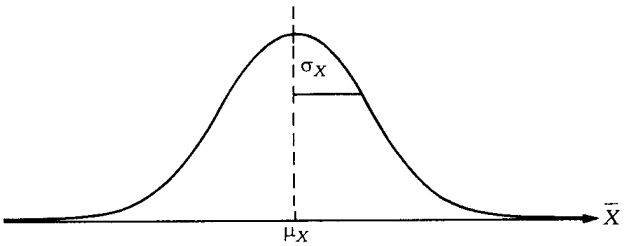
$$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} \Rightarrow \sigma_{\overline{X}} = \sqrt{\sigma_{\overline{X}}^2} = \sqrt{\frac{\sigma_X^2}{n}} = \frac{\sqrt{\sigma_X^2}}{\sqrt{n}} = \frac{\sigma_X}{\sqrt{n}} \Rightarrow \sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$$

si $n = 1 \Rightarrow \sqrt{n} = \sqrt{1} = 1$ y $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{\sigma_X}{1} = \sigma_X$

$\Rightarrow \sqrt{n} = 1$

$\Rightarrow \sigma_{\overline{X}} = \sigma_X$

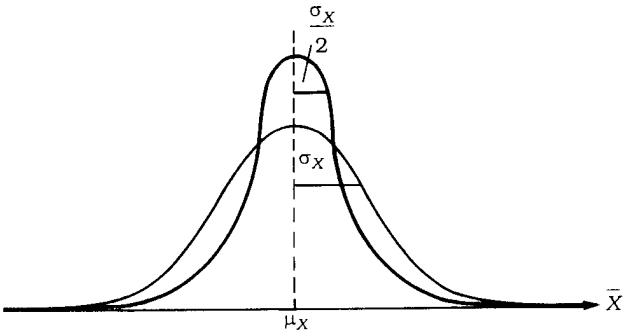
Tenemos la distribución poblacional



Si $n = 4 \Rightarrow \sqrt{n} = \sqrt{4} = 2$ y $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{\sigma_X}{2}$

$\Rightarrow \sqrt{n} = 2$

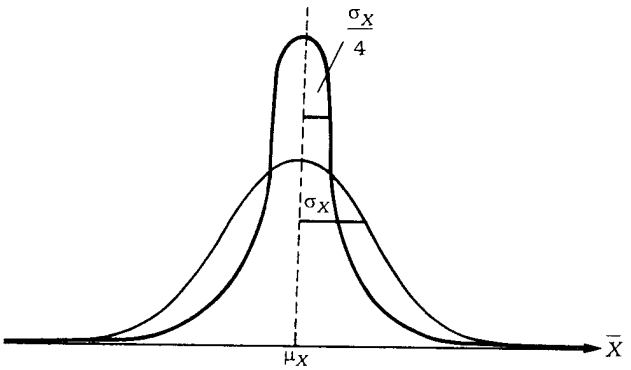
$\Rightarrow \sigma_{\overline{X}} = \frac{\sigma_X}{2}$



Si $n = 16 \Rightarrow \sqrt{n} = \sqrt{16} = 4$ y $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{\sigma_X}{4}$

$\Rightarrow \sqrt{n} = 4$

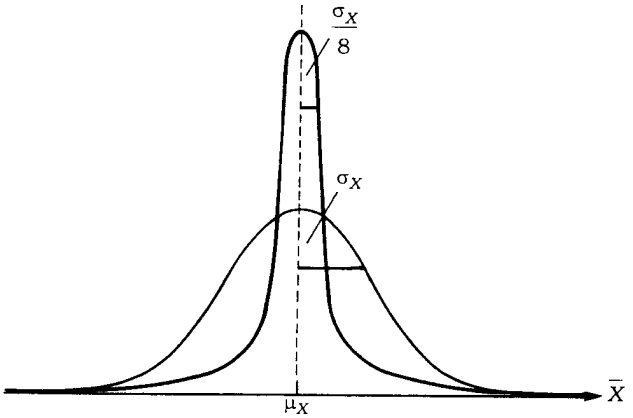
$\Rightarrow \sigma_{\bar{X}} = \dfrac{\sigma_X}{4}$



Si $n = 64 \Rightarrow \sqrt{n} = \sqrt{64} = 8$ y $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{\sigma_X}{8}$

$\Rightarrow \sqrt{n} = 8$

$\Rightarrow \sigma_{\bar{X}} = \dfrac{\sigma_X}{8}$

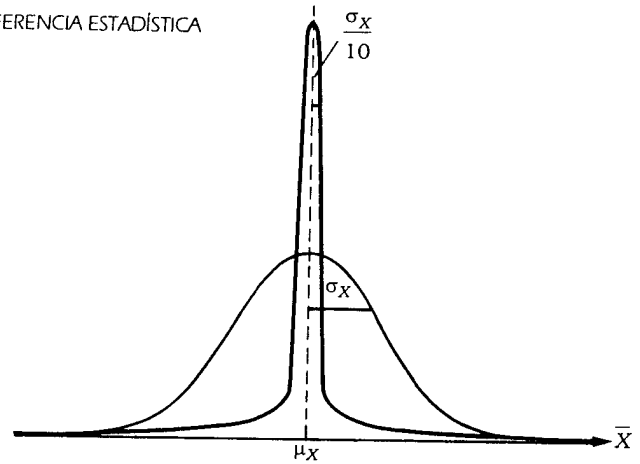


Si $n = 100 \Rightarrow \sqrt{n} = \sqrt{100} = 10$ y $\sigma_{\bar{X}} = \dfrac{\sigma_X}{\sqrt{n}} = \dfrac{\sigma_X}{10}$

$\Rightarrow \sqrt{n} = 10$

$\Rightarrow \sigma_{\bar{X}} = \dfrac{\sigma_X}{10}$

Ahora bien, si se le da un tratamiento [cambio de alguna(s) variable(s) en el procedimiento que se le aplica] a la población o a parte de ella, y si se elige una muestra de entre éstos (de tamaño mayor que 16); entonces, si el tratamiento



no es funcional la distribución de la población no cambiará, esto es $|\overline{X} - \mu_X| \to 0$ casi seguramente, pero si el tratamiento funciona (hace que cambie la distribución de la población) esto es $|\overline{X} - \mu_X| \to 0$ casi seguramente. El decisor tiene que fijar el nivel de significancia, α.

Hay tres casos

1. $H_0 : \mu_X = \mu_0$
   $H_1 : \mu_X < \mu_0$ si $\overline{X}_c < \mu_0$ (el tratamiento funcionó decrementalmente)
2. $H_0 : \mu_X = \mu_0$
   $H_1 : \mu_X > \mu_0$ si $\overline{X}_c > \mu_0$ (el tratamiento funcionó incrementalmente)
3. $H_0 : \mu_X = \mu_0$
   $H_1 : \mu_X \neq \mu_0$ si el problema es de control

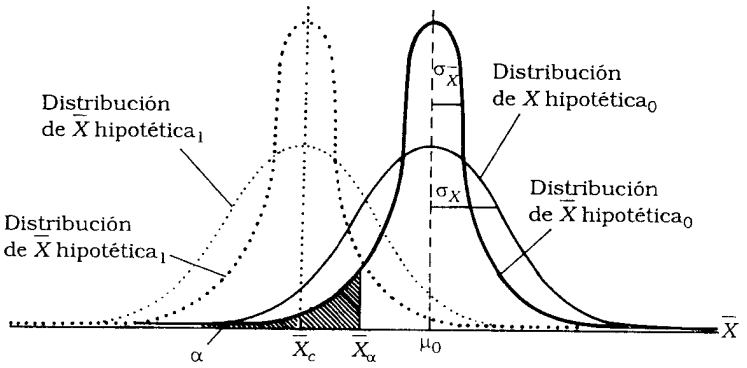
$\overline{X}_c$ se llama $\overline{X}$ *calculada*
$\overline{X}_\alpha$ se llama $\overline{X}$ *crítica*

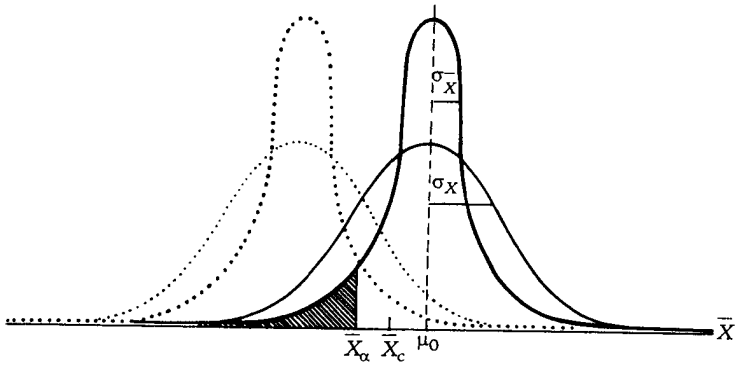
donde

$$P[\overline{X}_c \geq \overline{X}_\alpha] = \alpha$$

CASO 1

$(\overline{X}_c < \mu_0)$
$H_0 : \mu_X = \mu_0$
$H_1 : \mu_X < \mu_0$

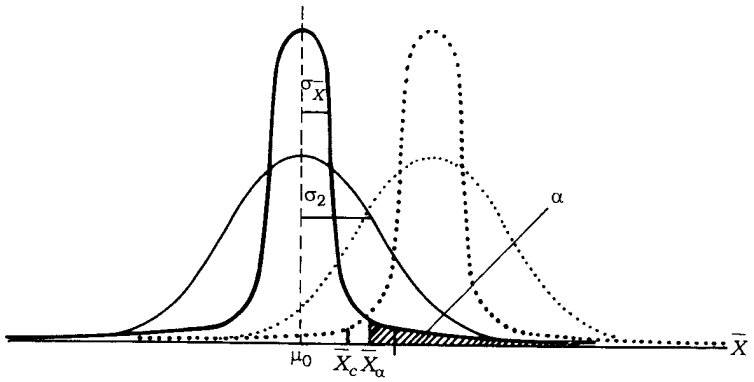




Como $\bar{X}_c \leq \mu_\alpha$ entonces se rechaza $H_0 : \mu_X = \mu_0$
El tratamiento funcionó significativamente a nivel α
Como $\bar{X}_c > \mu_\alpha$, entonces no se rechaza $H_0 : \mu_X = \mu_0$
o $\bar{X}_c > \mu_\alpha$ por azar.

CASO 2

$(\bar{X}_c > \mu_0)$
$H_0 : \mu_X = \mu_0$
$H_1 : \mu_X > \mu_0$
Como $\bar{X}_c \geq \mu_\alpha$, entonces se rechaza $H_0 : \mu_X = \mu_0$

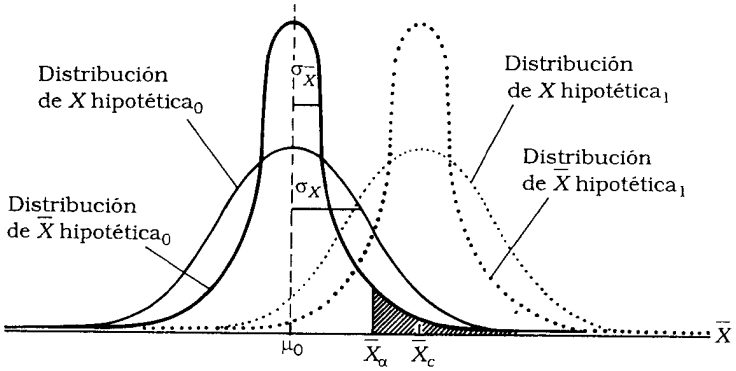


En otras palabras: el tratamiento funciona significativamente a un nivel α

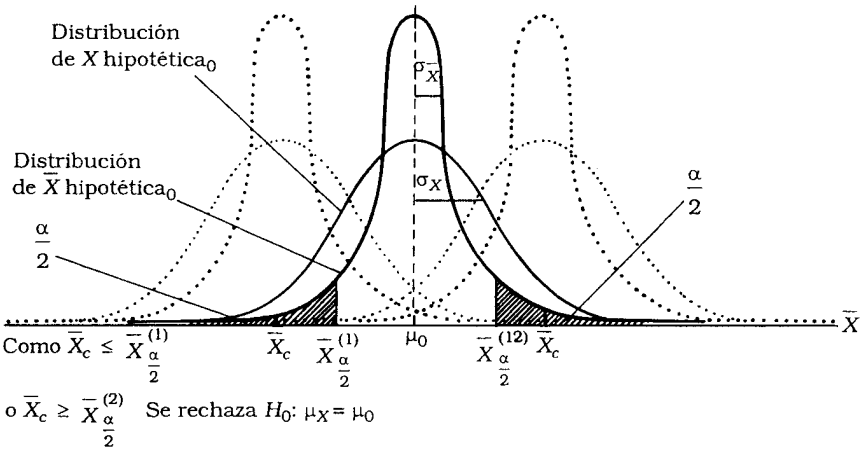
Como $\bar{X}_c < X_\alpha$ entonces no se rechaza $H_0 : \mu_X = \mu_0$
En otras palabras $\bar{X}_c > \mu_0$ por simple azar.

CASO 3

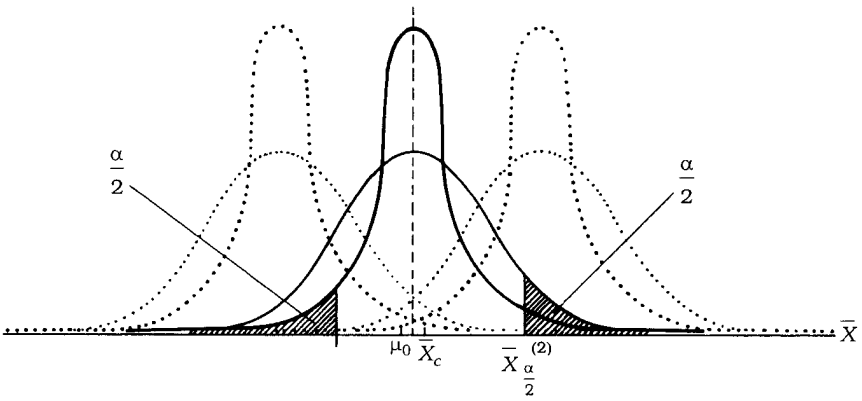
Control
$H_0 : \mu_X = \mu_0$
$H_1 : \mu_X <> \mu_0$
Como $\bar{X}_c \leq X_{\alpha/2}{}^{(1)}$ o $\bar{X}_c \leq X_{\alpha/2}{}^{(2)}$ se rechaza $H_0 : \mu_X = \mu_0$



El proceso está fuera de control.



Como $\overline{X}_{\alpha/2}{}^{(1)} < \overline{X}_c < \overline{X}_{\alpha/2}{}^{(2)}$ no se rechaza $H_0 : \mu_X = \mu_0$.
El proceso está bajo control.



Resumiendo:

Caso	Requerimiento	Planteamiento	Estadística de prueba	Regla de decisión	
1	$\overline{X}_c < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$\overline{X}_c = \frac{1}{n}\sum_{i=1}^{n} X_i$	Si $\overline{X}_c \leq \overline{X}_\alpha$ Si $\overline{X}_c > \overline{X}_\alpha$	Rechazar $H_0 : \mu_X = \mu_0$ No rechazar $H_0 : \mu_X = \mu_0$
2	$\overline{X}_c > \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$	$\overline{X}_c = \frac{1}{n}\sum_{i=1}^{n} X_i$	Si $\overline{X}_c \geq \overline{X}_\alpha$ Si $\overline{X}_c < \overline{X}_\alpha$	Rechazar $H_0 : \mu_X = \mu_0$ No rechazar $H_0 : \mu_X = \mu_0$
3	Control 1	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X <> \mu_0$	$\overline{X}_c = \frac{1}{n}\sum_{i=1}^{n} X_i$	Si $\overline{X}_c \leq \overline{X}^{(1)}_{\alpha/2}$ $\overline{X}_c \geq \overline{X}^{(2)}_{\alpha/2}$ Si $\overline{X}^{(1)}_{\alpha/2} < \overline{X}_c < \overline{X}^{(2)}_{\alpha/2}$	Rechazar $H_0 : \mu_X = \mu_0$ No rechazar $H_0 : \mu_X = \mu_0$

Si estandarizamos y usamos el caso 3, que podría decirse en la unión de los dos primeros, tendremos que

$$Z_c = \frac{\overline{X}_c - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \Rightarrow Z_c = \frac{\overline{X}_c - \mu_X}{\sigma_{\overline{X}}}$$

$$\frac{\overline{X}^{(1)}_{\alpha/2} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = -\frac{\overline{X}^{(2)}_{\alpha/2} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \quad \text{porque la distribución } Z \text{ es simétrica}$$

$$\Rightarrow Z^{(1)}_{\alpha/2} = -Z^{(2)}_{\alpha/2} \quad \text{Si} \quad Z_{\alpha/2} = Z^{(2)}_{\alpha/2} \Rightarrow -Z_{\alpha/2} = Z^{(1)}_{\alpha/2}$$

$$\Rightarrow Z^{(1)}_{\alpha/2} < Z_c < Z^{(2)}_{\alpha/2} \Rightarrow -Z_{\alpha/2} < Z_c < Z_{\alpha/2}$$

$$\Rightarrow |Z_c| < Z_{\alpha/2}$$

Por tanto, la síntesis es:

Caso	Requisito	Planteamiento	Estadística de prueba	Regla de decisión	
1	$\overline{X}_c < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$Z_c = \frac{\overline{X}_c - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$	Si $\lvert Z_c \rvert \geq Z_\alpha$ Si $\lvert Z_c \rvert < Z_\alpha$	Rechazar $H_0 : \mu_X = \mu_0$ No rechazar $H_0 : \mu_X = \mu_0$
2	$\overline{X}_c > \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$	$Z_c = \frac{\overline{X}_c - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$	Si $\lvert Z_c \rvert \geq Z_\alpha$ Si $\lvert Z_c \rvert < Z_\alpha$	Rechazar $H_0 : \mu_X = \mu_0$ No rechazar $H_0 : \mu_X = \mu_0$
3	Control	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$Z_c = \frac{\overline{X}_c - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$	Si $\lvert Z_c \rvert \geq Z_{\alpha/2}$ Si $\lvert Z_c \rvert < Z_{\alpha/2}$	Rechazar $H_0 : \mu_X = \mu_0$ No rechazar $H_0 : \mu_X = \mu_0$

Hasta aquí hemos supuesto que por estudios anteriores sabemos cuánto vale la varianza de la variable aleatoria X y es por eso que no vimos el intervalo de

confianza para la proporción poblacional, ya que en ese caso tendríamos una expresión como:

$$P - \sigma_p Z_{\alpha/2} \leq \pi \leq P + \sigma_p Z_{\alpha/2}$$

Pero $\sigma_p^2 = \dfrac{\sigma_X^2}{n}$ si $n \to \infty$ y como $\sigma_X^2 = \pi(1-\pi)$, entonces $\sigma_p^2 = \dfrac{\pi(1-\pi)}{n}$

y esto quiere decir que conocemos a π, por lo cual es innecesario estimarla en un intervalo.

Análogamente, por lo anterior no determinamos el tamaño de la muestra necesario para un error, ε previamente definido, *i e.*

$$n > \frac{\sigma_X^2 Z_{\alpha/2}^2}{\varepsilon^2} \quad \text{ya que} \quad \sigma_X^2 = \pi(1-\pi)$$

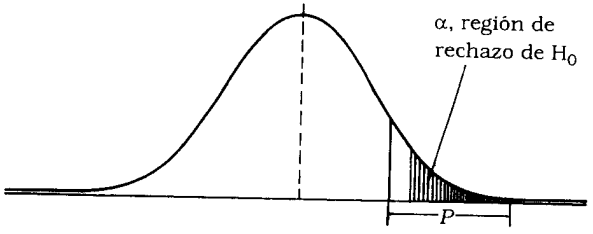
y si conocemos a π ni siquiera existe error.

RESUMEN

A veces una encuesta típica y simple es el estudio de una muestra extraída de una población real (por ejemplo, una encuesta de personas con crédito hipotecario).

Un experimento es el estudio de una población artificial creada por el experimentador. A un grupo de sujetos se les suministra cierto fármaco para saber si éste afecta el deseo de fumar. La población es ficticia, ya que la mayoría de los fumadores no ingieren dicho fármaco (la población existe únicamente para fines del estudio).

Una de las tareas del investigador es traducir sus hipótesis de investigación en un conjunto dicotómico de hipótesis estadísticas (H_0 y H_1) mutuamente excluyentes; se debe señalar si la diferencia entre la información obtenida a partir de la muestra y la información que se espera de acuerdo con H_0 tiene mucha o poca probabilidad de ocurrir, por ejemplo, $P < 0.05$.



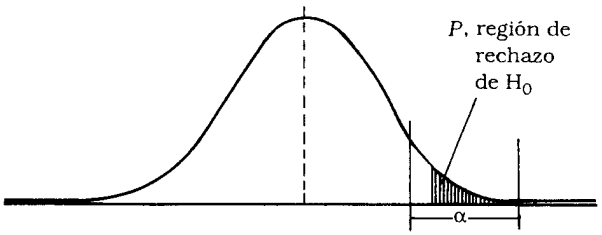
Las hipótesis científicas o de investigación explican los hechos en términos de relaciones de causalidad determinística o probabilística.

La inferencia estadística en relación con los estudios, investigaciones o experimentos se basa en el contraste o comprobación de hipótesis, cuya probabilidad de observar un hecho se evalúa suponiendo cierta la hipótesis (H_0). Si esta probabilidad es baja, entonces H_0 se rechaza y se dice que el hecho observado es estadísticamente significativo, aunque existen algunos factores externos (factores de confusión) que pueden causar inexactitud. A esto se le llama *error experimental.*

Con el objetivo de ubicar a la estadística clásica, el enfoque de la estimación y contraste de hipótesis es el de la escuela objetiva de la probabilidad, en la cual no se utiliza información anterior.

No rechazar H_0 si el estadístico de prueba no está en la región de rechazo, entonces P es mayor que α, ($P > \alpha$).

Procedimiento del valor P



La probabilidad de que el estadístico de prueba quede en la región de rechazo de H_0, si H_0 es cierta, es menor o igual a α.

El valor de P de un estadístico de prueba observado es la probabilidad de ese valor más la probabilidad de valores más extremos. Estas probabilidades se calculan a partir de la distribución H_0 del estadístico de prueba.

La regla de decisión es la siguiente: rechazar H_0 si el valor de P es menor o igual a α (área en el extremo más allá del valor observado del estadístico de prueba y menor que α), $P \leq \alpha$.

PROBLEMAS

8.1 De los siguientes enunciados, mencionar si es falso o verdadero.

a) Cuando se comete un error de tipo II la hipótesis nula es falsamente aceptada.

b) Cuando se comete un error de tipo I la hipótesis nula H_0 es falsamente rechazada.
c) Si $H_0 : \mu_1 < \mu_2$ se trata de una prueba de dos colas.
d) α es la probabilidad de cometer el error de tipo I.
e) β es la probabilidad de cometer el error de tipo II.

8.2 De los siguientes enunciados una sola opción es la verdadera.
El establecer la hipótesis alterna (H_1) depende de:

a) Lo que el investigador está interesado en probar.
b) Los datos obtenidos en el estudio.
c) La región crítica.
d) El nivel de significancia (α).
e) La potencia de la prueba.

8.3 Suponga que la región crítica para una cierta prueba de hipótesis es de la forma $|t| \geq 3.5$ y la t calculada ($t = -2.75$).
¿Cuál de las siguientes afirmaciones es correcta?

a) Rechazar H_0.
b) α es la probabilidad de que bajo H_0, $t > 2.75$ o $t < -2.75$.
c) La región de no rechazo de H_0 es $-3.5 < t < 3.5$.
d) La región de rechazo de H_0 consiste en los valores de $t > 3.5$ o $t < -3.5$.
e) Ninguna de las anteriores.

8.4 Una línea aérea desea probar la hipótesis de que 60% de sus pasajeros desaprueba que se fume dentro de los aviones. Si se establecen las hipótesis H_0 y H_1 como:

$H_0 : P = 60\%$
$H_1 : P \neq 60\%$

Explique en qué condiciones se cometería:

a) Un error de tipo I.
b) Un error de tipo II.

8.5 El departamento de investigación de un laboratorio farmacéutico, al obtener un nuevo medicamento, desea probar la hipótesis $H_0 : P = 0.90$, $H_1 : P < 0.90$. Si se realiza un estudio binomial y X es el número de aciertos observados y $n = 20$ ensayos, entonces no se rechaza H_0 si $X \geq 15$ y por consiguiente H_1 no se rechaza.

a) Establecer la región de rechazo de H_0.
b) Establecer la región de rechazo de H_1.

8.6 Se realiza un estudio entre los estudiantes en dos grupos independientes (hombres y mujeres) en una universidad, la variable de interés es la presión arterial con los siguientes resultados.

El promedio obtenido por el grupo de hombres es 125.2 y el grupo de mujeres 125.4; si la hipótesis establecida $H_0 : \mu_1 = \mu_2$ se rechaza para un $\alpha = 0.001$, ¿cuál de las siguientes conclusiones es la más razonable?

a) Existe una diferencia significativa (clínica) pero no estadística.
b) La diferencia es significativa tanto estadística como clínicamente.
c) Hay una diferencia significativa estadística pero no clínicamente.
d) No existe diferencia significativa ni estadística ni clínicamente.
e) Los tamaños de muestra son pequeños.

Estudio de una población 9

PROPÓSITOS

Al estudiar los contenidos de este capítulo, el lector estará en condiciones óptimas para utilizar el modelo estadístico adecuado; realizar un contraste de hipótesis y construir el intervalo de confianza para el parámetro poblacional de su interés.

De igual modo, el lector podrá:

1. Desarrollar el procedimiento para utilizar las tablas de distribución normal; 0, 1.
2. Dado un valor de α determinado, y utilizando las tablas mencionadas anteriormente, obtener el área bajo la curva generada por dicho nivel de significancia.
3. Conociendo la varianza σ_X^2, construir un intervalo de confianza para μ_X.
4. Conociendo la varianza σ_X^2, realizar una prueba de hipótesis, comparando la media muestral $\overline{X}$ y la media poblacional μ_X.
5. Construir un intervalo de confianza para μ_X y comparar la media muestral $\overline{X}$ y la media poblacional μ_X, si se desconoce la varianza σ_X^2.
6. Construir un intervalo de confianza para π y realizar un contraste de hipótesis entre la proporción muestral y la poblacional.
7. Construir un intervalo de confianza para σ_X^2 y realizar un contraste de hipótesis comparando la varianza muestral s^2 y la varianza poblacional σ_X^2.
8. Estimar el tamaño de muestra adecuado para cada uno de los casos anteriores.

9. Resolver problemas concernientes a las ciencias del comportamiento, utilizando los conceptos y métodos desarrollados en este capítulo.

INTRODUCCIÓN

En muchos estudios o investigaciones es necesario construir un intervalo de confianza o comprobar la hipótesis de un parámetro μ_X, σ_X^2 o π de una sola población, considerando el estadístico $\overline{X}$, s^2 o p como un estimador suficiente correspondiente de una población normal representada por la distribución muestral de $\overline{X}$.

Si se conoce la varianza poblacional, el error estándar puede ser calculado directamente y la prueba Z es la más apropiada, utilizando la tabla de distribución normal; 0, 1. En el caso más común, cuando no se conoce la varianza poblacional, el error estándar se obtiene a partir de los datos de la muestra y la prueba t es la más adecuada, utilizándose la tabla de distribución t de Student.

9.1 USO DE LAS TABLAS DE DISTRIBUCIÓN NORMAL; 0, 1

La variable aleatoria Z se define por

$$Z = \frac{X - \mu_X}{\sigma_X} \text{ y tiene por media } \mu_Z = 0 \text{ y varianza } \sigma_Z^2 = 1$$

Siempre que a una variable aleatoria, digamos X, le restamos su media, μ_X; $X - \mu_X$, y la dividimos entre su desviación estándar, σ_X, tendremos una variable aleatoria estandarizada Z o sea:

$$Z = \frac{X - \mu_X}{\sigma_X}$$

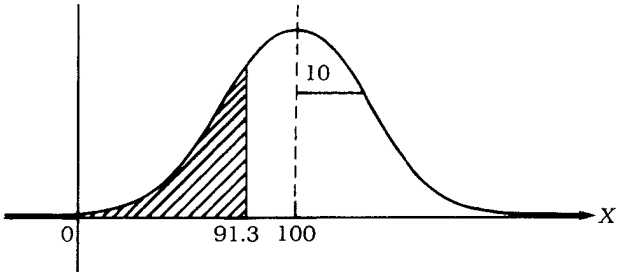
Ejemplo Supóngase que la distribución del coeficiente intelectual de los alumnos de una escuela tiene una distribución X con una media de 100 y la desviación estándar de 10. Usando una aproximación normal, calcule la probabilidad de que un alumno tenga un coeficiente intelectual:

1. Menor o igual a 91.3 puntos
2. Entre 100 y 105 puntos
3. Mayor o igual a 105.7 puntos

Solución

Como $\mu_X = 100$ y $\sigma_X = 10$

1. $P[X \leq 91.3] = ?$

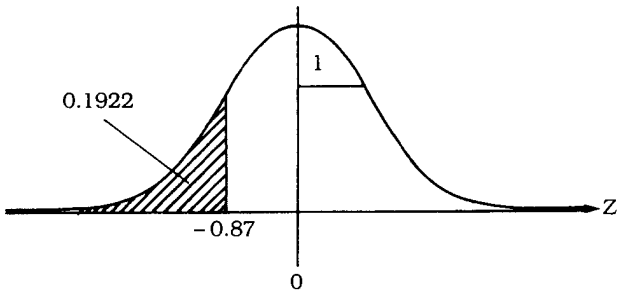


$P[X \leq 91.3] =$
$P[X - \mu_X \leq 91.3 - \mu_X]$

se le resta la media

$$= P\left[\frac{X - \mu_X}{\sigma_X} \leq \frac{91.3 - \mu_X}{\sigma_X}\right]$$

se divide entre la desviación estándar



Área bajo la curva normal

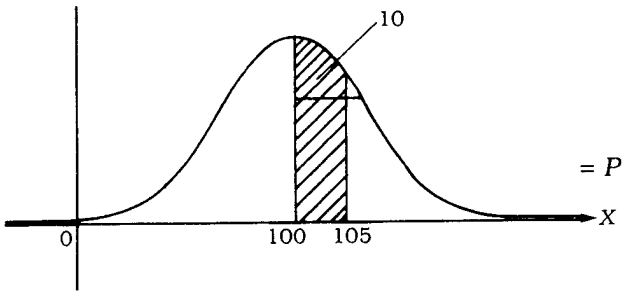
$$= P\left[Z \leq \frac{91.3 - 100}{10}\right]$$

sustituimos

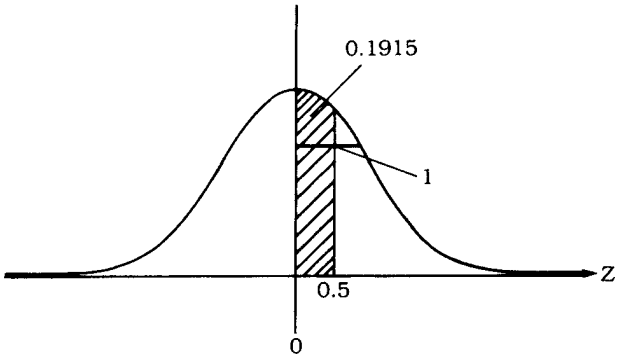
$= P[Z \leq -0.87]$
$= P[Z \geq 0.87]$

$\Rightarrow P[X \leq 91.3] = 0.1922$ buscando en la tabla el área bajo la curva normal.

2. $P[100 \leq X \leq 105] = P\left[\frac{100 - \mu_X}{\sigma_X} \leq \frac{X - \mu_X}{\sigma_X} \leq \frac{105 - \mu_X}{\sigma_X}\right]$



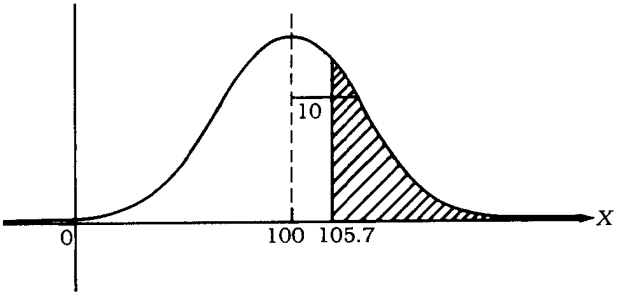
$$= P\left[\frac{100 - 100}{10} \leq Z \leq \frac{105 - 100}{10}\right]$$



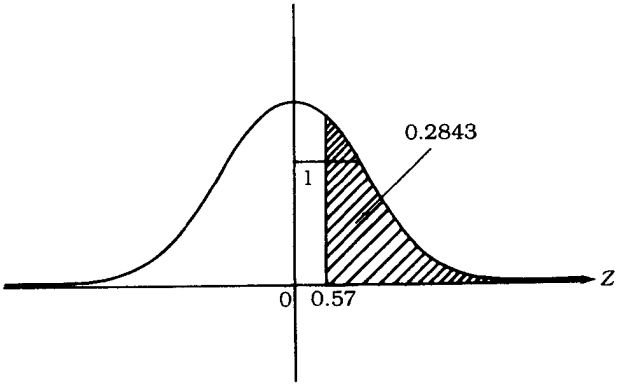
$$= P[0.00 \leq Z \leq 0.50]$$
$$= P[Z \geq 0.00] - P[Z \geq 0.50]$$
$$= 0.50 \quad - 0.3085$$
$$\Rightarrow P[100 \leq X \leq 105] = 0.1915$$

3. $P[X \geq 105.7] = P\left[\dfrac{X - \mu_X}{\sigma_X} \geq \dfrac{105.7 - \mu_X}{\sigma_X}\right]$

$$= P\left[Z \geq \frac{105.7 - 100}{10}\right]$$



$$= P[Z \geq 0.57]$$
$$\Rightarrow P[X \geq 105.7] = 0.2843$$



Si X es una variable aleatoria y ahora investigamos la proporción, entonces $\mu_X = \pi$ y $\sigma_X^2 = \pi\,(1 - \pi)$.

Ejemplo En una ciudad que tiene 20 000 habitantes se encuentra (a través de un censo) que 15 000 habitantes quieren que se materialice el proyecto de construcción de un centro cívico; entonces:

$$\pi = \frac{15\,000}{20\,000},\ \pi = 0.75, \text{o sea, } \pi = 75\%$$

$$\mu_X = \pi \Rightarrow \mu_X = 0.75$$

$$\sigma_X^2 = \pi(1-\pi) \Rightarrow \sigma_X^2 = 0.75\,(1 - 0.75) = 0.75\,(0.25) \Rightarrow \sigma_X^2 = 0.1875$$

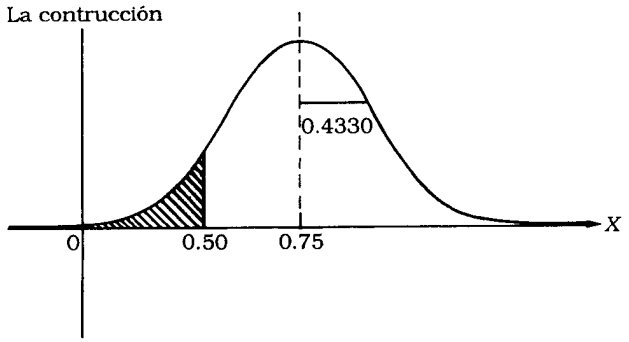
$$\sigma_X = \sqrt{\sigma_X^2} \Rightarrow \sigma_X = \sqrt{0.1875} \Rightarrow \sigma_x = 0.4330$$

Ejemplo En el ejemplo anterior calcule, usando una aproximación normal, la probabilidad de que un habitante:

1. Esté a lo más en 50% en favor de la construcción.
2. Esté a favor entre 60% y 80%.
3. Esté a favor en más de 85%.

Solución

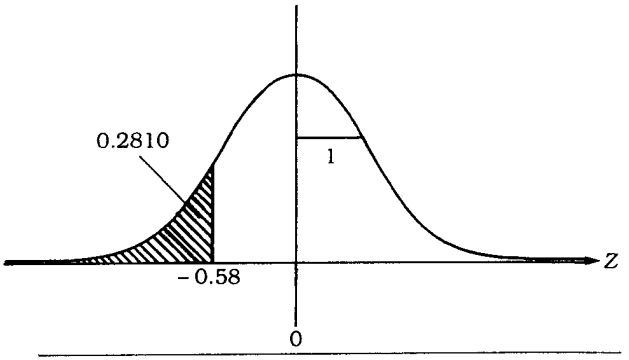
1. $P\,[X \leq 50\%] = P\,[X \leq 0.50]$



$$= P\left[\frac{X - \mu_X}{\sigma_X} \leq \frac{0.50 - \mu_X}{\sigma_X}\right]$$

$$= P\left[Z \leq \frac{0.50 - 0.75}{0.4330}\right]$$

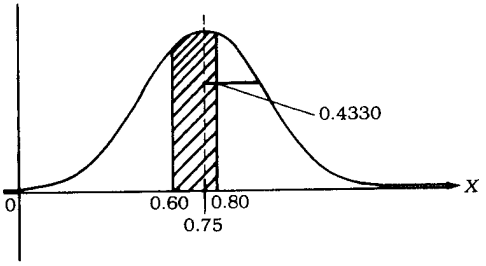
$$= P[Z \leq -0.58]$$
$$= P[Z \geq 0.58]$$
$$\Rightarrow P[X \leq 50\%] = 0.2810$$



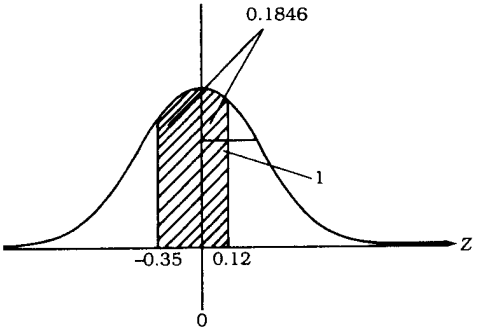
2. $P[60\% \leq X \leq 80\%] = P[0.60 \leq X \leq 0.80]$

$$= P\left[\frac{0.60 - \mu_X}{\sigma_X} \leq \frac{X - \mu_X}{\sigma_X} \leq \frac{0.80 - \mu_X}{\sigma_X}\right]$$

$$= P\left[\frac{0.60 - 0.75}{0.4330} \leq Z \leq \frac{0.80 - 0.75}{0.4330}\right]$$



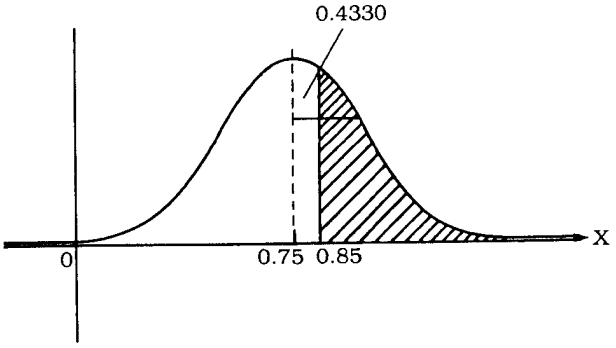
$$= P[-0.35 \leq Z \leq 0.12]$$
$$= P[Z \geq -0.35] - P[Z \geq 0.12]$$
$$= 1 - P[Z \geq 0.35] - P[Z \geq 0.12]$$
$$= (1 - 0.3632 - 0.4522)$$
$$\Rightarrow P[60\% \leq X \leq 80\%] = 0.1846$$

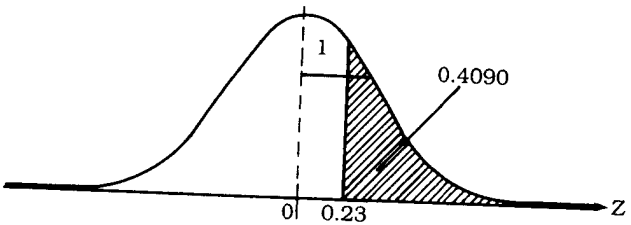


3. $P[X \geq 85\%] = P[X \geq 0.85]$

$$= P\left[\frac{X - \mu_X}{\sigma_X} \geq \frac{0.85 - \mu_X}{\sigma_X}\right]$$

$$= P\left[Z \geq \frac{0.85 - 0.75}{0.4330}\right]$$





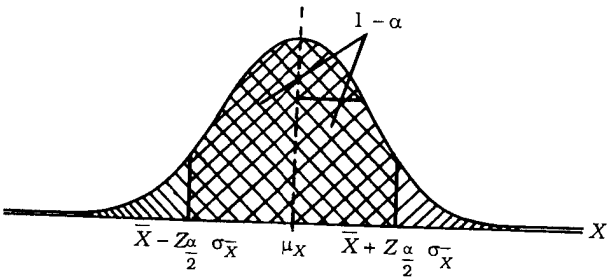
$= P[Z \geq 0.23]$
$\Rightarrow P[X \geq 85\%] = 0.4090$

9.2 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA UNA MEDIA

CASO 1. Varianza (σ_X^2 conocida)

La población de la cual se extrae la muestra debe distribuirse en forma normal, con media μ y varianza σ^2. Si la población no se distribuye en forma normal, el tamaño de la muestra deberá ser suficientemente grande ($n \to \infty$), cumpliendo así el teorema central de límite. Si se utiliza el modelo

$$P\left[-Z_{\alpha/2} \leq \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \leq Z_{\alpha/2}\right] = 1 - \alpha$$



pero $\mu_{\overline{X}} = \mu_X$ $\qquad \Rightarrow P\left[-Z_{\alpha/2} \leq \dfrac{\overline{X} - \mu_X}{\sigma_{\overline{X}}} \leq Z_{\alpha/2}\right] = 1 - \alpha$

pero $|a| = |-a|$, la anterior expresión se puede representar como:

$$\overline{X} - Z_{\alpha/2}\,\sigma_{\overline{X}} \leq \mu_X \leq \overline{X} + Z_{\alpha/2}\,\sigma_{\overline{X}}$$

o también

$$\mu_X = \overline{X} \pm Z_{\alpha/2}\,\sigma_{\overline{X}}$$

Ejemplo Un centro de copiado desea estimar la cantidad media de dinero que gasta un estudiante en época de exámenes. Se selecciona una muestra aleatoria de tamaño 36 y se encuentra que la media (muestral) es de \$1.20. Por estudios anteriores se sabe que la desviación estándar es de 0.24. Encontrar el intervalo de confianza (a un nivel de 95%) para la cantidad media de dinero, o sea, IC (95%).

$$\sigma_X = 0.24 \Rightarrow \sigma_X^2 = 0.24^2$$

$$n = 36 \qquad \overline{X} = 1.20 \qquad \mu_{\overline{X}} = \mu_X \qquad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}\frac{N-n}{N-1} \quad \text{como} \quad \frac{N-n}{N-1} \to 1$$

$$\Rightarrow \quad \sigma_{\overline{X}}^2 = \frac{0.24^2}{36} = \frac{0.24^2}{6^2} = \left(\frac{0.24}{6}\right)^2 = 0.04^2$$

$$\Rightarrow \sigma_{\overline{X}} = \sqrt{0.04^2} \Rightarrow \sigma_{\overline{X}} = 0.04$$

$$1 - \alpha = 95\% \Rightarrow \alpha = 5\% \Rightarrow Z_{\alpha/2} = 1.96$$

$$\overline{X} - \sigma_{\overline{X}}\, Z_{\alpha/2} \le \mu_X \le \overline{X} + \sigma_{\overline{X}}\, Z_{\alpha/2}$$

$$\Rightarrow 1.20 - 0.04\,(1.96) \le \mu_X = 1.20 + 0.04\,(1.96)$$

$\Rightarrow$ $1.1216 \le \mu_X \le 1.2784$, aproximando $1.12 \le \mu_X \le 1.28$ al 95%.

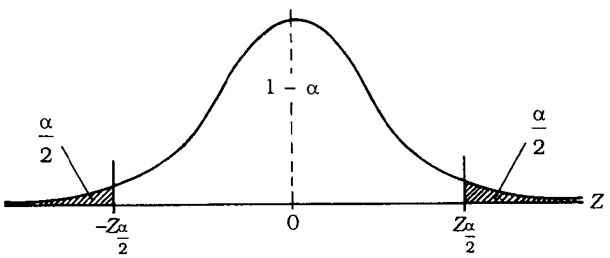
Otra forma de representar los intervalos de confianza es la siguiente:

$$\overline{X} \pm Z_{\alpha/2}\, \sigma_{\overline{X}} \text{ y su resultado } (1.12,\ 1.28) \text{ al } 95\%$$

o también así:

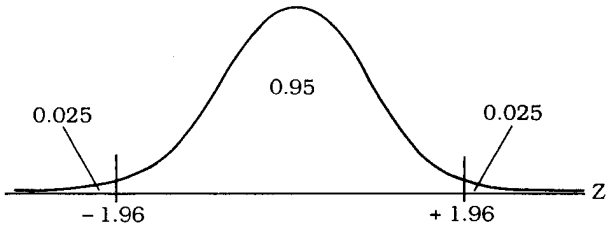
$$\overline{X} - Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} \le \mu_X \le \overline{X} + Z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}$$

Y esto se debe a que la distribución normal estándar es simétrica.



Por ejemplo, si $\alpha = 5\%$, entonces:

$-Z_{\alpha/2} = -Z_{0.025} = -1.96$ ya que está situada en la parte izquierda de la curva.
$Z_{\alpha/2} = +1.96$ ya que se ubica en el lado derecho de la curva.
En este libro se utilizará $Z_{0.025} = \pm 1.96$.



Ejemplo En un esfuerzo para mejorar su horario de citas, un médico desea estimar el tiempo promedio que dedica a cada paciente. En un periodo de tres semanas selecciona una muestra aleatoria de 49 pacientes, obteniendo su media aritmética de 30 minutos con una desviación estándar de 7 minutos.

Construir un intervalo de confianza al 95% para el tiempo promedio verdadero que el médico dedica a cada paciente; o sea, IC (95%).

Datos

$\overline{X} = 30$	$n = 49$	como	$1 - \alpha = 95\%$
$\sigma_X = 7$			$\alpha = 5\%$
			$\alpha/2 = 0.025$
		entonces	$Z_{\alpha/2} = Z_{0.025} = \pm 1.96$

$$P\left(30 - 1.96\frac{7}{\sqrt{49}} \leq \mu_X \leq 30 + 1.96\frac{7}{\sqrt{49}}\right)$$

donde el error e es $1.96\left[\frac{7}{\sqrt{49}}\right]$

$e = 1.96$
$P\,(28.04 \leq \mu_X \leq 31.96)$
$P\,(28 \leq \mu_X \leq 32)$, o también suele escribirse (28, 32) al 95%.

Ejemplo Una escuela desea estimar a un nivel de confianza de 90%, el peso medio de los estudiantes del sexto grado. Se selecciona una muestra al azar de 25 alumnos y se encuentra que el peso medio (muestral) es de 45 kg. Se sabe que la desviación estándar de la población es de 6 kg.

$N \to \infty \quad \overline{X} = \text{peso} \quad \mu_X = ? \quad \sigma_X = 6,\ \sigma_X^2 = 6^2$

$n = 25 \quad \overline{X} = 45 \quad \mu_{\overline{X}} = \mu_X \quad \sigma_{\overline{X}}^2 = \dfrac{\sigma_X^2}{n}\dfrac{N-n}{N-1} \quad \text{como} \quad \dfrac{N-n}{N-1} \to 1$

$$\Rightarrow \quad \sigma_{\overline{X}}^2 = \frac{6^2}{25} = \frac{6^2}{5^2} = \left(\frac{6}{5}\right)^2 = 1.2^2$$

$$\Rightarrow \sigma_{\overline{X}} = \sqrt{1.2^2} \quad \Rightarrow \boxed{\sigma_{\overline{X}} = 1.2}$$

$1 - \alpha = 90\% \Rightarrow \alpha = 10\% \Rightarrow \quad Z_{\alpha/2} = 1.645$

Fórmula $\overline{X} - \sigma_{\overline{X}}\ Z_{\alpha/2} \le \mu_X \le \overline{X} + \sigma_{\overline{X}}\ Z_{\alpha/2}$

$\Rightarrow \quad 45 - 1.2\ (1.645) \le \mu_X \le 45 + 1.2\ (1.645)$

$43.026 \le \mu_X \le 46.974$

$43 \le \mu_X \le 47$ al 90%

o también (43, 47).

Tamaño de la muestra para un intervalo de confianza

$$P\left[\ \left|\ \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}\ \right| \le Z_{\alpha/2}\right] = 1 - \alpha$$

Si $\overline{X} - \mu_{\overline{X}}$ existe error positivo, o sea, $\varepsilon = \overline{X} - \mu_X$

$$\Rightarrow P\left[\ \left|\ \frac{\varepsilon}{\sigma_{\overline{X}}}\ \right| > Z_{\alpha/2}\right] \qquad \text{pero} \left|\frac{\varepsilon}{\sigma_{\overline{X}}}\right| = \frac{|\varepsilon|}{|\sigma_{\overline{X}}|} = \frac{\varepsilon}{\sigma_{\overline{X}}} \quad \text{ya que} \quad \varepsilon > 0 \ \text{ y } \sigma_{\overline{X}} > 0$$

$$\Rightarrow \text{como } \frac{\varepsilon}{\sigma_{\bar{X}}} > Z_{\alpha/2}$$

$$\Rightarrow \text{entonces } \frac{1}{\sigma_{\bar{X}}} > \frac{Z_{\alpha/2}}{\varepsilon}$$

$$\Rightarrow 1 > \frac{Z_{\alpha/2}}{\varepsilon}\frac{\sigma_X}{\sqrt{n}} \Rightarrow \sqrt{n} > \frac{Z_{\alpha/2}}{\varepsilon}\sigma_X \Rightarrow n > \left(\frac{Z_{\alpha/2}}{\varepsilon}\sigma_X\right)^2$$

$$\Rightarrow n > \frac{(Z_{\alpha/2}\,\sigma_X)^2}{\varepsilon^2} \quad \Rightarrow \quad \boxed{n > \frac{Z^2_{\alpha/2}\,\sigma^2_X}{\varepsilon^2}}$$

Ejemplo Si en el ejemplo del centro de copiado que desea estimar el gasto promedio de los estudiantes en copias se quiere tener un nivel de confianza de 95%, y se desea que el error máximo sea de 0.05, ¿de qué tamaño tiene que ser la muestra?

$$\varepsilon = 0.05; \qquad \sigma_X = 0.24; \qquad Z_{\alpha/2} = 1.96$$

$$n > \frac{Z^2_{\alpha/2}\,\sigma^2_X}{\varepsilon^2}$$

$\Rightarrow n > \dfrac{1.96^2\,(0.24)^2}{0.05^2}$, $n > 88.5$, o sea, de 89 o más consumidores.

Ejemplo Si en el ejempo de la escuela que desea estimar el peso medio de los estudiantes del sexto grado se quiere tener un nivel de confianza de 90% y que el error máximo sea de 1 kg, ¿de qué tamaño tiene que ser la muestra?

$$\varepsilon = 1; \qquad \sigma_X = 6; \qquad Z_{\alpha/2} = 1.645$$

$$n > \frac{Z^2_{\alpha/2}\,\sigma^2_X}{\varepsilon^2}$$

$\Rightarrow n > \dfrac{1.645^2\,(6)^2}{1^2}$, $n \geq 97.42$, o sea, de 98 o más estudiantes.

9.3 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA LA MEDIA SI σ_X^2 NO SE CONOCE

CASO 2. Varianza (σ_X^2) desconocida

$1 - \alpha = P\left[\left| \dfrac{\overline{X} - \mu_X}{s/\sqrt{n}} \right| \leq t_{\alpha/2,\, n-1} \right]$ porque $\dfrac{\overline{X} - \mu_X}{s/\sqrt{n}} = t_{n-1}$ es la distribución t o "t de Student"

$$= P\left[\left| \frac{\mu_X - \overline{X}}{s/\sqrt{n}} \right| \leq t_{\alpha/2,\, n-1} \right]$$

$\therefore$ el intervalo para μ_X con un nivel de confianza $1 - \alpha$ es

$$\overline{X} - \frac{s}{\sqrt{n}}\, t_{\alpha/2,\, n-1} \leq \mu_X \leq \overline{X} + \frac{s}{\sqrt{n}}\, t_{\alpha/2,\, n-1}$$

Ejemplo Se desea estimar el peso promedio de las ratas que se utilizan en el laboratorio; para ello, se pesan cinco de una muestra aleatoria y se obtienen los siguientes datos:

Rata	Peso (gr)
1	200
2	203
3	206
4	209
5	212

Construir un intervalo de confianza al 90%.

Datos:

$\overline{X} = \dfrac{1030}{5} = 206$

$s = 4.24$

$\alpha = 10\%$

$_{\alpha/2} = 5\% = 0.05$

Los grados de libertad para una muestra son gl = $n - 1$

El valor de t en la tabla es:

$t_{\alpha/2,\, gl} = t_{0.05,4} = 2.132$

Se sustituyen los valores y se obtiene:

$$206 - 2.132\left(\frac{4.24}{\sqrt{5}}\right) \leq \mu_X \leq 206 + 2.132\left(\frac{4.24}{\sqrt{5}}\right)$$

$[206 - 2.132\ (1.89619) \leq \mu_X \leq 206 + 2.132\ (1.89619)]$

$P\,(206 - 4.04267 \leq \mu_X \leq + 206 + 4.04267)$

$P\,(201.95733 \leq \mu_X \leq 210.04267)$, o sea (202, 210)

Cuando no se conoce la desviación estándar de la población, ésta debe calcularse a partir de los datos muestrales, o sea, s_X es la desviación estándar de la muestra. Cuando esto sucede (en la mayor parte de las situaciones reales, se desconoce σ_X), la distribución t es la distribución de muestreo más adecuada. Se considera práctico utilizar la distribución t solamente cuando se requiera que el tamaño de la muestra sea de 30 o menor, ya que para muestras más grandes los valores de t y z son aproximadamente iguales y es posible emplear la distribución normal en lugar de la distribución t.

$$t = \frac{\overline{X} - \mu_X}{s_{\overline{X}}}$$

donde $$s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

$$gl = n - 1$$

Tamaño de la muestra para estimar μ_X si $\sigma_X^2 = ?$ (o sea, varianza desconocida).

1. Se toma una muestra piloto de tamaño n'

2. Como $E\left[\frac{n'}{n'-1} s_{n'}^2\right] = \sigma_X^2$ se calcula $s_{n'}^2 = \frac{n'}{n'-1} s_{n'}^2$

y $$n \geq \left(\frac{Z_{\alpha/2}\ s_{n'}\ \Delta\ n'}{\epsilon}\right)^2$$

9.3.1 Pruebas de hipótesis para μ_X si $\sigma_X^2 = ?$

Si $\overline{X} < \mu_0$	*Estadística de prueba*	*Regla de decisión*
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$t_{c,\,n-1} = \frac{\overline{X} - \mu_X}{s/\sqrt{n}}$	si $\lvert t_{c,\,n-1} \rvert \geq t_{\alpha,\,n-1} \Rightarrow$ Rechazar $H_0 : \mu_X = \mu_0$ si $\lvert t_{c,\,n-1} \rvert < t_{\alpha,\,n-1} \Rightarrow$ No rechazar $H_0 : \mu_X = \mu_0$
Si $\overline{X} > \mu_0$ $H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$	$t_{c,\,n-1} = \frac{\overline{X} - \mu_X}{s/\sqrt{n}}$	si $\lvert t_{c,\,n-1} \rvert \geq t_{\alpha,\,n-1} \Rightarrow$ Rechazar $H_0 : \mu_X = \mu_0$ si $\lvert t_{c,\,n-1} \rvert < t_{\alpha,\,n-1} \Rightarrow$ No rechazar $H_0 : \mu_X = \mu_0$
Control $H_0 : \mu_X = \mu_0$ $H_1 : \mu_X <> \mu_0$	$t_{c,\,n-1} = \frac{\overline{X} - \mu_X}{s/\sqrt{n}}$	si $\lvert t_{c,\,n-1} \rvert \geq t_{\alpha/2,\,n-1} \Rightarrow$ Rechazar $H_0 : \mu_X = \mu_0$ si $\lvert t_{c,\,n-1} \rvert < t_{\alpha/2,\,n-1} \Rightarrow$ No rechazar $H_0 : \mu_X = \mu_0$

Ejemplo Nueve personas han seguido una dieta especial durante dos meses. En ese tiempo, sus pérdidas de peso individuales fueron en kilogramos:

1.2
2.0
1.0
0.8
1.1
0.2
0.5
0.4
0.1

Pruebe la hipótesis nula de un promedio verdadero de pérdida de peso de 0 kg, contra la alternativa de una pérdida de peso mayor de 0 kg, utilizando $\alpha = 0.01$. Suponga que la población es normal.

Solución

Paso 1. La variable de interés es el valor medio de la pérdida de peso en las personas observadas. El parámetro que se requiere establecer es (μ_X).

Paso 2. Proponer las hipótesis estadísticas.

$$H_0 : \mu_X = 0 \text{ kg}$$
$$H_1 : \mu_X > 0 \text{ kg}$$

Paso 3. Establecer el valor del error que se está dispuesto a aceptar (α).

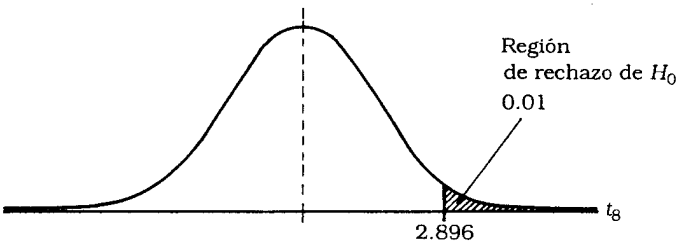
$\alpha = 0.01$ o sea $\alpha = 1\%$
confiabilidad $1 - \alpha = 99\%$

Paso 4. Determinar el modelo estadístico adecuado.

$$t_C = \frac{\overline{X} - \mu_X}{s/\sqrt{n}}, \qquad \text{gl} = n - 1, \qquad n < 30$$

Paso 5. Obtener mediante la tabla correspondiente el valor crítico y ubicar la zona de rechazo en una gráfica.

$t_{0.01,\,8} = 2.896$



Paso 6. Especificar la regla de decisión para rechazar H_0.

Si $t_C > t_{0.01,\,8} = 2.896$, entonces se rechaza H_0.

Paso 7. Calcular el valor del estadístico de prueba, utilizando el modelo propuesto.

- Se debe calcular el valor de la desviación estándar muestral s.

Se obtiene $s = \sqrt{0.3536}$

o sea $s = 0.5946$

$\overline{X} = 0.8111$

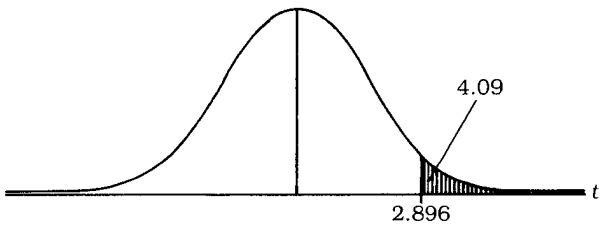
$s = 0.5946$

Se sustituyen los valores en el modelo propuesto.

$$t_c = \frac{\overline{X} - \mu_X}{s/\sqrt{n}} = \frac{0.8111 - 0}{\frac{0.5946}{\sqrt{9}}} = \frac{0.8111}{0.19822} = 4.09, \text{ o sea } t_c = 4.09$$

Paso 8. Obtener las conclusiones.

Como $4.09 > 2.896$, entonces H_0 se rechaza con una $\alpha = 0.01$.



Se dice que el método para pérdida de peso es eficaz, es decir ($\mu_X > 0$).

9.4 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA UNA PROPORCIÓN, CUANDO LA VARIANZA ES DESCONOCIDA

Esto significa $\sigma_X^2 = \pi\,(1 - \pi) = ?$

y como en una proporción $\mu_X = \pi$ y $\mu_P = \mu_X$

y $$s^2 = \frac{n}{n-1}\, p_C\,(1 - p_C)$$

Entonces

$$P_C - \frac{1}{\sqrt{n}}\sqrt{\frac{n}{n-1}\, p_C\,(1 - p_C)}\;\; t_{\alpha/2,\, n-1} \le \pi \le p_C + \frac{1}{\sqrt{n}}\sqrt{\frac{n}{n-1}\, p_C\,(1 - p_C)}\; t_{\alpha/2,\, n-1}$$

o sea: $$p_C - \sqrt{\frac{p_C\,(1 - p_C)}{n-1}}\; t_{\alpha/2,\, n-1} \le \pi \le P_C + \sqrt{\frac{p_C\,(1 - p_C)}{n-1}}\; t_{\alpha/2,\, n-1}$$

Entonces, el intervalo de confianza a nivel $1 - \alpha$ es:

$$p_C - \frac{s}{\sqrt{n}}\, t_{\alpha/2,\, n-1} \le \pi \le p_C + \frac{s}{\sqrt{n}}\, t_{\alpha/2,\, n-1}$$

donde $$s = \sqrt{\frac{n}{n-1}\, p_C\,(1 - p_C)}$$

como $$\frac{s}{\sqrt{n}} = \frac{1}{\sqrt{n}} \times \sqrt{\frac{n}{n-1}\, p_C\,(1 - p_C)}$$

$$= \frac{1}{n-1}\sqrt{p_C\,(1 - p_C)}$$

entonces $$s = \sqrt{\frac{p_C\,(1 - p_C)}{n-1}} \quad \text{y} \quad \text{si } n \to \infty \;\Rightarrow\; s = \sqrt{\frac{p_C\,(1 - p_C)}{n}}$$

$$p_C - \sqrt{\frac{p_C\,(1 - p_C)}{n-1}}\;\; t_{\alpha/2,\, n-1} \le \pi \le p_C + \sqrt{\frac{p_C\,(1 - p_C)}{n-1}}\;\; t_{\alpha/2,\, n-1}$$

Algunos autores utilizan la siguiente fórmula:

$$p_C - \sqrt{\frac{p_C(1-p_C)}{n}}\ t_{\alpha/2,\, n-1} \le \pi \le p_C + \sqrt{\frac{p_C(1-p_C)}{n}}\ t_{\alpha/2,\, n-1}$$

Cuando el tamaño de la muestra es muy grande ($n \to \infty$).

Ejemplo Un partido político desea conocer la proporción de sus partidarios en una zona de una colonia determinada; para ello toma una muestra aleatoria de 100 residentes (jefes de familia), de los cuales el 80% están afiliados a dicho partido. Construya un intervalo de confianza al 95% de la proporción de residentes de dicha zona afiliados al partido político.

Solución

Debido a que $p_C = (0.8)\ (100) = 80$ y $1 - p_C = (0.2)\ (100) = 20$, la aproximación normal de la binomial puede ser empleada

$$s = \sqrt{\frac{p_C(1-p_C)}{n}} = \sqrt{\frac{(0.8)\ (0.2)}{100}} = \sqrt{0.0016} = 0.04$$

$n = 100$ (se considera una muestra grande)

El intervalo de confianza al 95% es:

$$0.80 - (1.96)\ (0.04) \le \pi \le 0.80 + (1.96)\ (0.04)$$

donde $Z_{0.25} = t_{0.25,\, 99} = \pm\ 1.96$

entonces

$$0.72 \le \pi \le 0.88$$

Tamaño de la muestra, *n*, para estimar π

1. Se toma una muestra piloto de tamaño n'.
2. Como $E\,[s^2] = \pi\ (1 - \pi)$ se calcula

$$s^2 = \frac{n'}{n'-1}\, p_C\,(1 - p_C) \qquad \text{por lo que } n \ge \left(\frac{Z_{\alpha/2}\ s}{\varepsilon}\right)^2$$

Ejemplo Una encuesta dirigida a 900 lectores de cierta revista indica que 40% de ellos tienen por lo menos dos años de estudios universitarios. Constrúyase un intervalo de confianza de 99% para la verdadera proporción de todos los lectores de dicha revista.

Solución

El problema nos indica que: $\hat{p}$ (proporción estimada) es igual a p_C o sea $\hat{p} = p_C$

$$\hat{p} = 0.40 \qquad \hat{q} = 0.60$$

como n es grande ($n = 900$), se usará la aproximación a la normal Z.

El ejercicio indica construir un intervalo de confianza de 90%, IC (90%), entonces $\alpha = 10\%$. Se busca en las tablas el valor de $Z_{\alpha/2}$, o sea:

$$Z_{\frac{0.01}{2}} = Z_{0.05} = 1.645$$

$$\text{como} \quad p_C = 0.4$$

El intervalo de confianza queda de la siguiente forma:

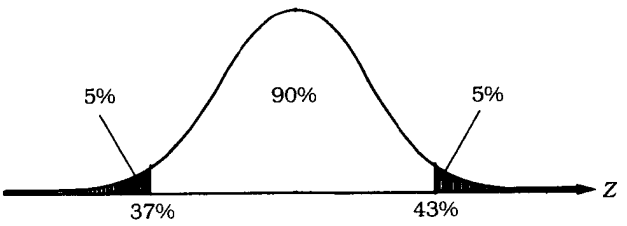
$$0.40 - 1.645\sqrt{\frac{(0.4)(0.6)}{900}} \leq \pi \leq 0.40 + 1.645\sqrt{\frac{(0.4)(0.6)}{900}}$$

$$(0.40 - 0.026781 \leq \pi \leq 0.40 + 0.026781)$$

$$(0.3732189 \leq \pi \leq 0.426781)$$

o sea $37\% \leq \pi \leq 43\%$ al 90%

Gráficamente pueden visualizarse los intervalos de confianza de la siguiente manera:



9.4.1 Resumen de pruebas de hipótesis para la proporción si $\sigma_X^2 = \pi\ (1 - \pi) = ?$

	Estadística de prueba	*Regla de decisión*
Si $p_C < \pi_0$ $H_0 : \mu_X = \pi_0$ $H_1 : \mu_X < \pi_0$	$t_C = \dfrac{P_C - \mu_p}{\sqrt{\dfrac{p_C(1-p_C)}{n-1}}}$	Si $\lvert t_{c,\,n-1} \rvert \geq t_{\alpha,\,n-1} \Rightarrow$ Rechazar $H_0 : \mu_X = \pi_0$ Si $\lvert t_{c,\,n-1} \rvert < t_{\alpha,\,n-1} \Rightarrow$ No rechazar $H_0 : \mu_X = \pi_0$
Si $p_C = \pi_0$ $H_0 : \mu_X = \pi_0$ $H_1 : \mu_X > \pi_0$	$t_C = \dfrac{P_C - \mu_p}{\sqrt{\dfrac{p_C(1-p_C)}{n-1}}}$	Si $\lvert t_{c,\,n-1} \rvert \geq t_{\alpha\,n-1} \Rightarrow$ Rechazar $H_0 : \mu_X = \pi_0$ Si $\lvert t_{c,\,n-1} \rvert < t_{\alpha\,n-1} \Rightarrow$ No rechazar $H_0 : \mu_X = \pi_0$
Control $H_0 : \mu_X = \pi_0$ $H_1 : \mu_X <> \pi_0$	$t_C = \dfrac{P_C - \mu_p}{\sqrt{\dfrac{p_C(1-p_C)}{n-1}}}$	Si $\lvert t_{c,\,n-1} \rvert \geq t_{\alpha/2n-1} \Rightarrow$ Rechazar $H_0 : \mu_X = \pi_0$ Si $\lvert t_{c,\,n-1} \rvert < t_{\alpha/2n-1} \Rightarrow$ No rechazar $H_0 : \mu_X = \pi_0$

El modelo estadístico también se considera como si n pequeña:

$$Z_C = \frac{p_C - \pi}{\sqrt{\dfrac{\pi\,(1-\pi)}{n}}} \qquad \text{si } n \to \infty$$

Ejemplo De experiencias anteriores se sabe que la proporción de ganado que tiene un defecto congénito es de 0.0005, siempre que se siga un procedimiento de apareo adecuado. Esta proporción se incrementa si se violan ciertas reglas del procedimiento de apareo. En una muestra de 10 000 animales, hay 10 con dicho defecto. ¿Esto evidencia (α = 5%) que dicho procedimiento fue violado?

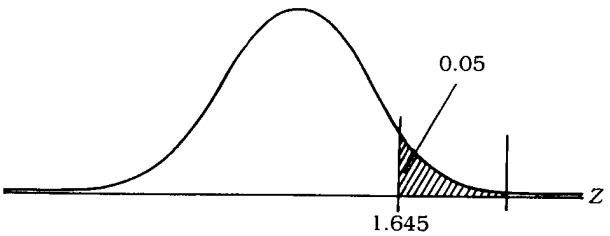
Donde

$$n = 10\,000 \quad X = 10 \qquad\qquad H_0 : \pi = 0.0005$$

$$P_C = \frac{X}{n} = 0.001 \qquad\qquad H_1 : \pi > 0.0005$$

$$Z_C = \frac{p_C - \pi}{\sqrt{\dfrac{\pi\,(1-\pi)}{n}}} = \frac{0.001 - 0.0005}{\sqrt{\dfrac{(0.0005)\,(0.9995)}{10\,000}}} = \frac{0.0005}{0.000223551} = 2.2366$$

la región crítica es $Z_{0.05} = 1.645$



con la regla de decisión si $Z_C \geq 1.645$, H_0 se rechaza.

Conclusión El procedimiento de apareo fue violado,

donde $p = \dfrac{X}{n}$ y π son los valores que se desea contrastar.

Ejemplo Un centro de salud asegura que en 90% de las mujeres de la zona se aplica un método de control natal sin supervisión médica. Este punto de vista contrasta con la hipótesis alterna, según la cual dicho porcentaje es menor de 90%. Se lleva a cabo una investigación con una muestra aleatoria formada por 1 600 mujeres adultas, de las que 1 400 responden que reciben supervisión médica. Utilizar un $\alpha = 1\%$.

Solución

Paso 1. La variable de interés es el control natal bajo supervisión médica, que constituye una característica de la población en estudio. Por ello, el parámetro que se busca es una proporción.

Paso 2. Se propone la hipótesis nula y alterna.

$$H_0 : \pi = 0.90$$
$$H_1 : \pi < 0.90$$

Paso 3. Debido a las características del problema (existe una acusación grave), el nivel de confianza debe ser lo más pequeño posible, entonces:

$$\alpha = 0.01$$

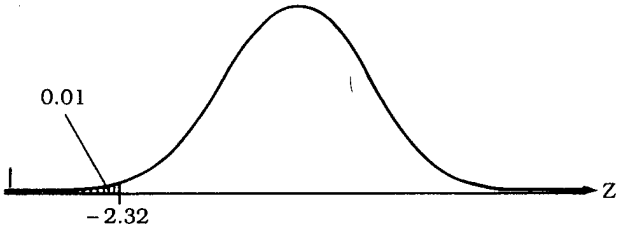
Paso 4. Se determina el modelo estadístico adecuado.

$$Z_C = \frac{p_C - \pi}{\sqrt{\frac{\pi\,(1-\pi)}{n}}}$$

Paso 5. Obtener mediante la tabla correspondiente el valor crítico y ubicar la zona de rechazo en la gráfica.

$$\alpha = 0.01 \Rightarrow Z_\alpha = 2.32$$

La zona de rechazo se presenta en la gráfica:



Como $H_1 : \pi < 0.90$ se debe considerar la parte izquierda de la curva para la región de rechazo.

Paso 6. Se especifica la regla de decisión para rechazar H_0

$$\text{si } Z_C < -2.32 \Rightarrow \text{se rechaza } H_0$$

Paso 7. Se calcula el valor del estadístico de prueba, utilizando el modelo propuesto.

$$\pi = 0.90$$

$$1 - \pi = 0.10$$

$$p_C = \frac{X}{n} = \frac{1\,400}{1\,600} = 0.875$$

$$Z_C = \frac{0.875 - 0.90}{\sqrt{\frac{(0.90)\,(0.10)}{1\,600}}}$$

$$Z_C = \frac{-0.025}{0.0075} = -3.333$$

$$Z_C = -3.333$$

Paso 8. Se establecen conclusiones:

$$\text{como } -3.333 < -2.32 \Rightarrow \text{se rechaza } H_0.$$

Puede concluirse que la declaración del centro de salud es falsa.

9.5 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA UNA VARIANZA

Para algunos investigadores resulta de mayor importancia estimar una varianza que una media; por eso, en ocasiones será necesario construir un intervalo de confianza para una varianza poblacional desconocida. Es importante recordar que en todo estudio se busca minimizar la varianza (σ_X^2) de la variable de interés. Algunos ejemplos en los que se busca estimar un intervalo de confianza para σ_X^2 son los siguientes:

- Un economista analiza el poder de compra que tiene la clase media de cierta ciudad: como él considera insuficiente para sus estudios el promedio (μ_X), decide estimar (σ_X^2). En países con problemas económicos la variabilidad tiende a ser grande.
- La gerencia de manufactura de cierto producto desea conocer cuál es la variabilidad de piezas defectuosas por mes. Como se observa, en este problema se espera que σ_X^2 sea pequeña.
- Un psicólogo desea saber la variabilidad que existe entre hombres o mujeres, casados y solteros, en una población de personas que padecen migraña.

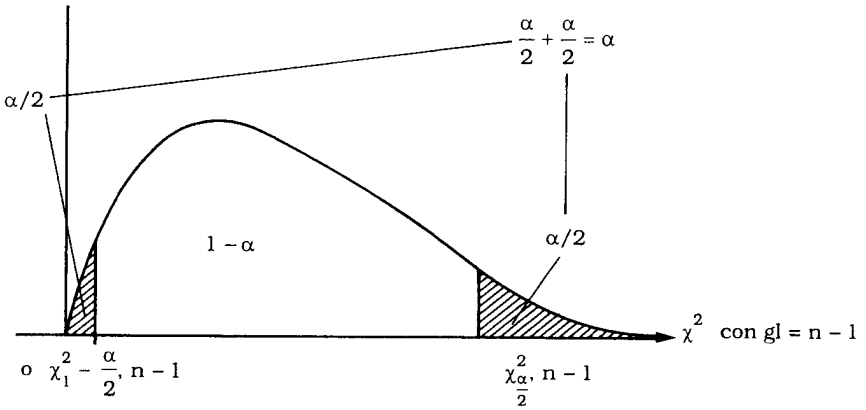
Para construir el intervalo de confianza se considera la distribución muestral S^2 cuando la población es normal.

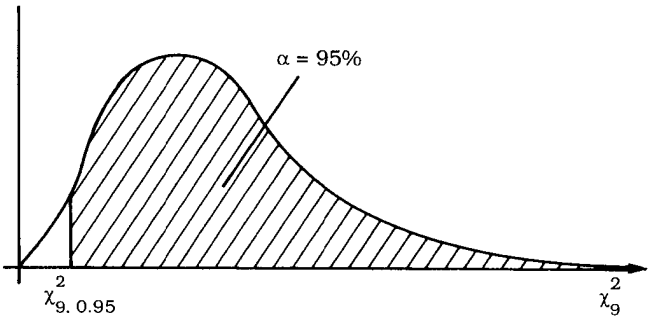
$$\chi^2_{1-\alpha/2,\, n-1} \leq \frac{(n-1)\, S^2}{\sigma_X^2} \leq \chi^2_{\alpha/2,\, n-1}$$

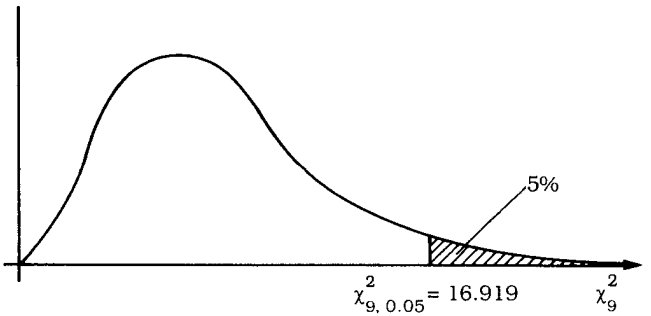
$$\frac{1}{\chi^2_{\alpha/2,\, n-1}} \leq \frac{\sigma_X^2}{(n-1)\, S^2} \leq \frac{1}{\chi^2_{1-\alpha/2,\, n-1}}$$

$$\boxed{\frac{(n-1)\, S^2}{\chi^2_{\alpha/2,\, n-1}} \leq \sigma_X^2 \leq \frac{(n-1)\, S^2}{\chi^2_{1-\alpha/2,\, n-1}}}$$

La representación gráfica se presenta a continuación:







Ejemplo Se realiza un experimento acerca del tiempo de reacción ante un suceso determinado. En la experiencia intervienen 10 adultos, es decir, $n = 10$. El suceso consiste en un estímulo visual que se asemeja a la luz roja de un semáforo, en respuesta al cual cada participante oprime un botón, lo que equivale a frenar el auto. El tiempo de reacción, medido en décimas de segundo, es el que transcurre entre el encendido de la luz y el frenar. Los resultados fueron los siguientes: 14.5, 24.4, 23.5, 13.6, 19.0, 19.2 15.4, 13.6, 17.2, 27.4.

Construya un intervalo de confianza para σ_X^2 al 95%

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

$$S^2 = \frac{215.429}{9} = 23.9366$$

A partir de la varianza muestral se construirá un intervalo de confianza para la varianza poblacional.

El problema pide construir un intervalo al 95%, entonces

$\alpha = 0.05$
$\alpha/2 = 0.025$
$1 - \alpha/2 = 0.975$

Como $n = 10$ los grados de libertad

$gl = n - 1 = 9$

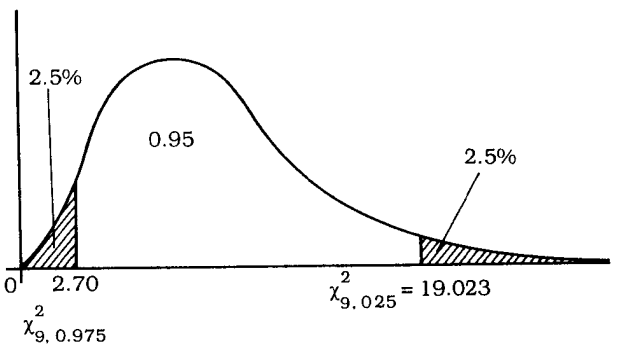
Se buscan en las tablas los valores de

$\chi^2_{0.025,9}$ y $\chi^2_{0.975,9}$

$\chi^2_{0.025,\,9} = 19.023$

$\chi^2_{0.975,\,9} = 2.70$

$$\frac{(n-1)\,S^2}{\chi^2_{0.025,9}} \leq \sigma_X^2 \leq \frac{(n-1)\,S^2}{\chi^2_{0.975,9}}$$



Al sustituir los valores en el intervalo de confianza se obtiene:

$$\left[\frac{9\,(23.9366)}{19.03} \leq \sigma_X^2 \leq \frac{9\,(23.9366)}{2.70}\right]$$

$$(11.3, \quad 79.8)$$

9.5.1 Pruebas de hipótesis para $\sigma_X^2 = \pi\ (1 - \pi) = ?$

Si $\frac{n}{n-1} S_n^2 < \sigma_0^2$	*Estadística de prueba*	*Regla de decisión*
$H_0 : \sigma_X^2 = \sigma_0^2$ $H_1 : \sigma_X^2 < \sigma_0^2$	$\chi^2_{c,\,n-1} = \frac{(n-1)\,s^2}{\sigma_0^2}$	Si $\chi^2_{c,\,n-1} \leq \chi^2_{1-\alpha,\,n-1} \Rightarrow$ Rechazar $H_0 : \sigma_X^2 = \sigma_0^2$ Si $\chi^2_{c,\,n-1} > \chi^2_{1-\alpha,\,n-1} \Rightarrow$ No rechazar $H_0 : \sigma_X^2 = \sigma_0^2$
Si $\frac{n}{n-1} S_n^2 < \sigma_0^2$ $H_0 : \sigma_X^2 = \sigma_0^2$ $H_1 : \sigma_X^2 < \sigma_0^2$	$\chi^2_{c,\,n-1} = \frac{(n-1)\,S^2}{\sigma_0^2}$	Si $\chi^2_{c,\,n-1} \geq \chi^2_{\alpha,\,n-1} \Rightarrow$ Rechazar $H_0 : \sigma_X^2 = \sigma_0^2$ Si $\chi^2_{c,\,n-1} > \chi^2_{\alpha,\,n-1} \Rightarrow$ No rechazar $H_0 : \sigma_X^2 = \sigma_0^2$
Control $H_0 : \sigma_X^2 = \sigma_0^2$ $H_1 : \sigma_X^2 <> \sigma_0^2$	$\chi^2_{c,\,n-1} = \frac{(n-1)\,S^2}{\sigma_0^2}$	Si $\chi^2_{c,\,n-1} \leq \chi^2_{1-\alpha/4,\,n-1}$ o $\chi^2_{c,\,n-1} \geq \chi^2_{\alpha/2,\,n-1} \Rightarrow$ Rechazar $H_0 : \sigma_X^2 = \sigma_0^2$ Si $\chi^2_{1-\alpha/2\,n-1} < \chi^2_{c,\,n-1} < \chi^2_{\alpha/2,\,n-1} \Rightarrow$ No rechazar $H_0 : \sigma_X^2 = \sigma_0^2$

La mayor parte de las veces un contraste de hipótesis se realiza acompañado de un intervalo de confianza, a fin de poder determinar de manera exacta el parámetro de la población en estudio.

Ejemplo Cuatro empleados tienen el siguiente salario mensual: 1 269, 1 271, 1 263 y 1 265 pesos, respectivamente. Se dice que el salario se distribuye en forma normal con una varianza igual a $2.00. Realice el contraste de hipótesis para saber si la varianza es mayor que dicha cantidad.

Solución

Paso 1. Identifique la variable de interés. Se desea conocer la variabilidad entre los salarios de los empleados de cierta empresa, por lo que el parámetro será σ_X^2.

Paso 2. Proponga las hipótesis estadísticas:

$$H_0 : \sigma_X^2 = 2$$
$$H_1 : \sigma_X^2 > 2$$

Paso 3. Establezca el error α que se está dispuesto a aceptar:

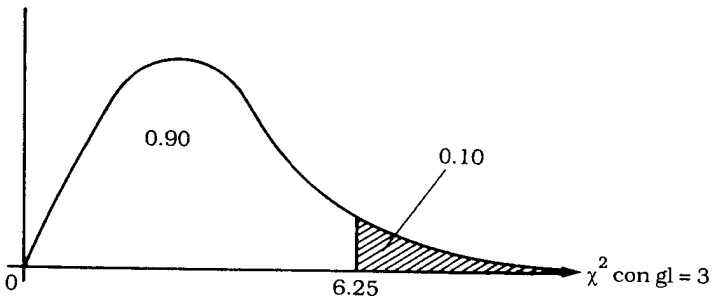
$$\alpha = 0.10$$

Paso 4. Determine el modelo estadístico adecuado:

$$\chi_C^2 = \frac{(n-1)\,S^2}{\sigma_X^2}$$

Paso 5. Obtenga, mediante la tabla correspondiente, el valor crítico y ubique la zona de rechazo en una gráfica.

$$\alpha = 0.10$$
$$1 - \alpha = 0.90$$
$$\text{gl} = n - 1 = 3$$
$$\chi^2_{0.90,\,3} = 6.25$$



Paso 6. Especifique la regla de decisión para rechazar H_0.

$$\text{Si } \chi_C^2 > \chi^2_{0.90,\,3} \qquad H_0 \text{ se rechaza}$$

es decir

$$\text{Si } \chi_C^2 > 6.25 \qquad H_0 \text{ se rechaza}$$

Paso 7. Calcule el valor del estadístico de prueba mediante el modelo propuesto.

Donde $S^2 = 13.333$

$$\chi_C^2 = \frac{(n-1)\,S^2}{\sigma_X^2} = \frac{3\,(13.333)}{2} = 20;\ \chi_C^2 = 20$$

Paso 8. Como 20 > 6.25, H_0 se rechaza, $\alpha = 10\%$.

¡La variabilidad entre los salarios de los empleados es mayor de $2.00!

Ejemplo Se desea que el contenido de un frasco de colonia sea de 100 ml. Se toma una muestra aleatoria de 144 frascos y se encuentra que el volumen medio (muestral) de colonia es de 99 ml. Si la desviación estándar es de 4 ml y el nivel de significancia es de 0.025, ¿es significativa la disminución de contenido?

$$N \to \infty \qquad \mu_X = 100 \qquad \sigma_X = 4 \qquad \sigma_X^2 = 4^2$$

$$n = 144 \qquad \overline{X}_C = 99 \qquad \mu_{\overline{X}} = \mu_X \qquad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}\frac{N-n}{N-1} \qquad \text{como } N \to \infty \Rightarrow \frac{N-n}{N-1} \to 1$$

$$\Rightarrow \sigma_{\overline{X}}^2 = 100 \Rightarrow \sigma_{\overline{X}}^2 \frac{4^2}{144} = \frac{4^2}{12^2} = \left(\frac{4}{12}\right)^2 = \left(\frac{1}{3}\right)^2 = 0.333^2$$

$$\Rightarrow \sigma_{\overline{X}} = \sqrt{0.333^2} \qquad \Rightarrow \sigma_{\overline{X}} = 0.333$$

$$\alpha = 0.025 \Rightarrow Z_\alpha = 1.960 \Rightarrow |Z_\alpha| = |1.960|$$

$$Z_C = \frac{\overline{X}_C - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \Rightarrow Z_C = \frac{99 - 100}{0.333} = -3.00 \Rightarrow |Z_C| = |-3.00| = 3.00 \Rightarrow |Z_C| = 3.00$$

Planteamiento

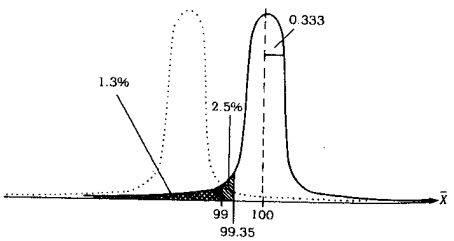
$H_0 : \mu_X = 100$
$H_1 : \mu_X < 100$ $\qquad X_C < \mu_X \qquad (99 > 100)$

Estadística de prueba:

$$Z_C = \overline{X}_C - \frac{\mu_{\overline{X}}}{\sigma_{\overline{X}}}$$

Regla de decisión:

Si $|Z_C| \geq |Z_\alpha| \Rightarrow$ Rechazar $H_0 : \mu_X = 100$
Si $|Z_C| < |Z_\alpha| \Rightarrow$ No rechazar $H_0 : \mu_X = 100$



como:

$3.00 > 1.960 \Rightarrow$ Rechazar $H_0 : \mu_X = 100$

Ejemplo Un proceso está programado para empaquetar la cantidad promedio de 6 onzas de café soluble por paquete y su desviación estándar es 0.2 onzas. Se toma una muestra aleatoria de 100 paquetes cada hora para indagar si el proceso de empaque está bajo control. A las 8 p.m. la muestra arroja una media de 6.1 onzas. ¿Está el proceso significativamente (a un nivel de 2%) fuera de control?

$$N \to \infty \qquad \mu_X = 6.0 \qquad \sigma_X = 0.2 \qquad \sigma_X^2 = 0.2^2$$

$$n = 100 \qquad \overline{X}_C = 6.1 \qquad \mu_{\overline{X}} = \mu_X \qquad \sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n}\frac{N-n}{N-1}$$

$$\Rightarrow \sigma_{\overline{X}} = \sqrt{0.02^2} \Rightarrow \sigma_{\overline{X}} = 0.02$$

Si $\alpha = 2\% \Rightarrow Z_{\alpha/2} = 2.325 \Rightarrow |Z_{\alpha/2}| = |2.325|$

$$Z_C = \frac{\overline{X}_C - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \Rightarrow Z_C = \frac{6.1 - 6.0}{0.02} = 5$$

Planteamiento

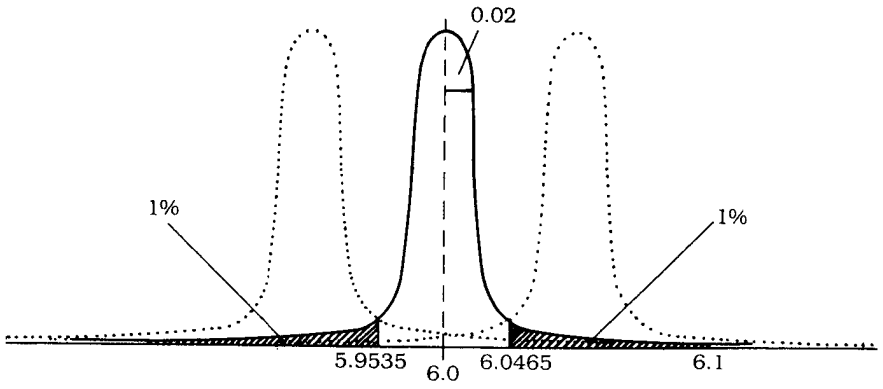
$H_0 : \mu_X = 6.0$
$H_1 : \mu_X <> 6.0$ porque es un problema de control

Regla de decisión:

Si $|Z_C| \geq |Z_{\alpha/2}| \Rightarrow$ Rechazar $H_0 : \mu_X = 6.0$

Si $|Z_C| < |Z_{\alpha/2}| \Rightarrow$ No rechazar $H_0 : \mu_X = 6.0$

como: $5 > 2.325 \Rightarrow$ Rechazar $H_0 : \mu_X = 6.0$



RESUMEN

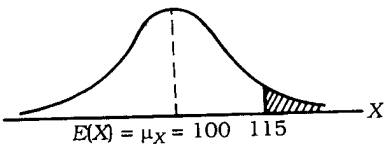
Es de vital importancia distinguir las situaciones cuando se conoce la varianza poblacional σ_X^2 y cuando se le desconoce, pues esto implica utilizar el modelo estadístico más adecuado, la Z o la t de Student. Implícitamente debe recordarse que se está basando en el teorema central del límite o en la distribución muestral de la variable aleatoria t. Sin embargo, si la muestra es grande ($n \to \infty$), S_X puede sustituir a σ_X y se utiliza el modelo Z.

PROBLEMAS

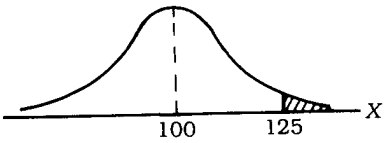
9.1 Las siguientes curvas muestran la distribución de los coeficientes de inteligencia de un grupo de estudiantes. Encontrar el porcentaje de las áreas sombreadas.

Donde $\sigma_x = 10$.

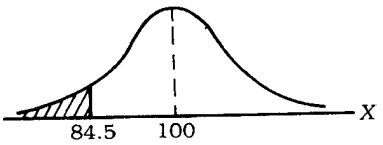
a)



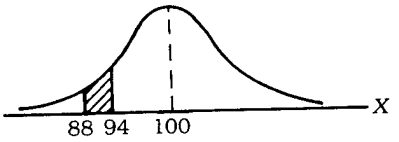
b)



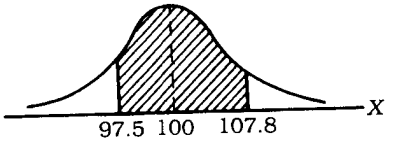
c)



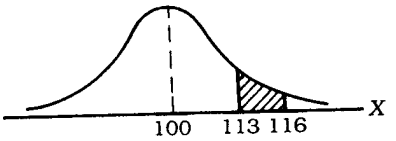
d)



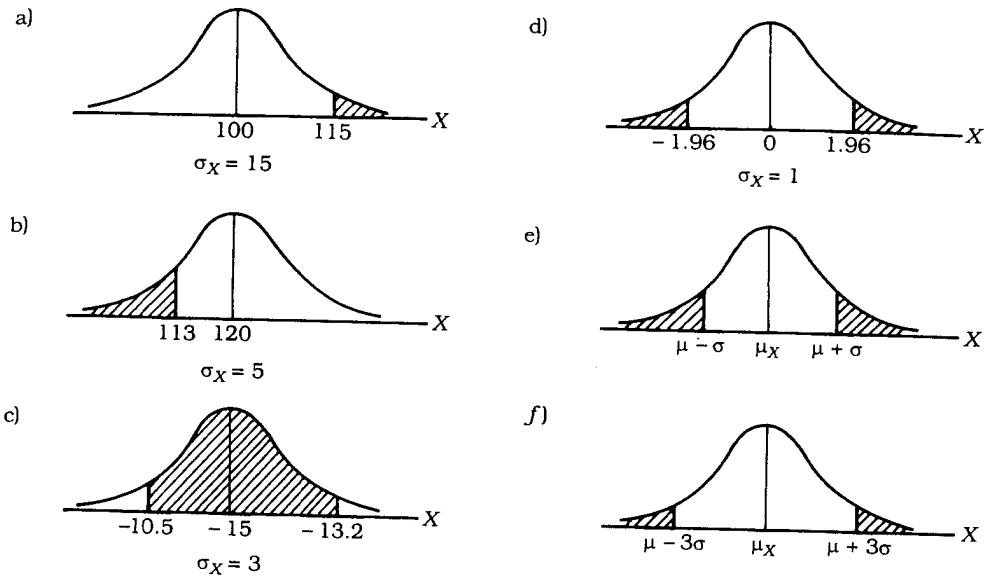
e)



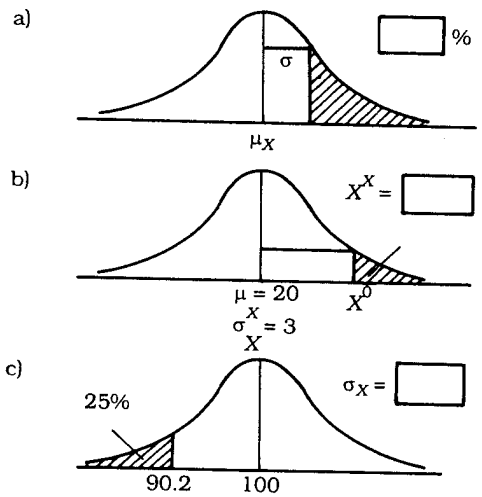
f)



9.2 Las siguientes curvas representan distribuciones normales. Encuentre los porcentajes de las áreas sombreadas.



9.3 Coloque los valores correctos en los recuadros.



9.4 Una empresa desea estimar el promedio de tiempo que necesita un empleado para llegar a su trabajo. Se selecciona al azar una muestra de 36 empleados y se encuentra que la media es de $\overline{X} = 40$ minutos. Suponiendo $\sigma_x = 12$ minutos. Construya un intervalo de confianza al 95%, de la media poblacional μ_x.

9.5 De un libro de 600 páginas se toma una muestra aleatoria de $n = 64$ páginas y se encuentra que el número promedio de palabras es $\overline{X} = 300$. La desviación estándar de las palabras por página es $\sigma_x = 40$. Construya un intervalo de confianza al 95% para el promedio del número de palabras por página.

9.6 La propaganda acerca del contenido en calorías de cierta bebida refrescante es de 50 calorías por botella. Se toma al azar una muestra de 36 botellas y se encuentra que el promedio de calorías por botella es de 49.3 calorías. Supóngase que las calorías están distribuidas de manera normal y que la varianza es conocida $\mu_X^2 = 9$.

¿Estos datos apoyan lo que se afirma? Utilice un nivel de confianza $\alpha = 5\%$. Establezca las hipótesis correspondientes.

9.7 En un autódromo se realiza una carrera de autos fórmula 1. La carrera consiste en un número determinado de kilómetros; n es el número de autos o participantes que terminaron la carrera; si el promedio fue de 20 horas y con una desviación estándar de 3 horas, construya un intervalo de confianza al 90% de la media poblacional.

9.8 Una distribuidora de agua purificada lanza al mercado una nueva presentación, cuyo envase tiene una capacidad de cinco litros en promedio. El proceso de llenado en cierta etapa muestra una desviación estándar de medio litro, 0.5l; si el Instituto del Consumidor hace un muestreo aleatorio en el mercado y obtiene una muestra $n = 50$, construya un intervalo al 95% de la media poblacional.

9.9 Una empacadora de café tiene una presentación de paquetes de un kilogramo; la desviación estándar de dichos paquetes es $\sigma_X = 0.02$ de kg. Se quiere saber si dicha máquina trabaja satisfactoriamente, se realiza un muestreo aleatorio de 25 paquetes, cuyo peso promedio en la muestra $\overline{X} = 1.014$ de kg.

Compruebe la hipótesis H_0: $\mu = 1$ kg para $\alpha = 1\%$.

9.10 (Caso 2, μ_X^2 desconocida). Supóngase que se toma una muestra $n = 36$ observaciones, cuya media es 41 200, desviación estándar muestra de 3 000; si la media poblacional $\mu_X = 40\ 000$, pero μ_X^2 es desconocida, entonces

a) contraste las siguientes tres hipótesis.

1. $H_1 : \mu_X < 40\ 000$
2. $H_1 : \mu_X > 40\ 000$
3. $H_1 : \mu_X \neq 40\ 000$

Utilice un $\alpha = 5\%$

b) Realice un intervalo de confianza al 95%.

9.11 Un especialista en control de calidad hace una auditoría en una empresa que fabrica envase tetrapac para jugos, y el fabricante asegura que 1% de los envases están defectuosos. Se hace un muestreo aleatorio y se obtiene una muestra de $n = 200$, de los cuales cuatro resultan defectuosos.

Construya un intervalo de confianza al 95% y contraste la hipótesis con un α = 1% y a = 5% para P $\neq$ = 1% (o sea para una proporción $\pi \neq 0.01$).

9.12 En un periódico de gran circulación se afirma que aproximadamente 25% de los adultos en la ciudad de México son analfabetos. Ponga a prueba esta hipótesis contra la alternativa de que el porcentaje verdadero no es 25% con un nivel de significancia de 5% de que se cometa el error de tipo I, se realiza un muestreo aleatorio y se tiene una muestra de 740 personas que indica que 20% son consideradas analfabetas; establezca un intervalo de confianza al 95%.

9.13 Una embotelladora de vinos asegura que menos de 3% de las botellas están por debajo del nivel de llenado. Se obtiene una muestra en forma aleatoria de 50 botellas, de las cuales cuatro se encuentran en esa situación. La muestra fue seleccionada de una producción de 400 botellas, establezca la hipótesis y realice el contraste al 5%.

9.14 El departamento de control de calidad de una óptica revisa las lentes de un pedido, cuyo índice de refracción debe tener una desviación estándar que no exceda de 1%. Si se toma una muestra aleatoria de 20 lentes de este pedido, cuyo índice de refracción tiene una varianza de 0.00023:

a) Establecer un intervalo de confianza de 99%.

b) Las lentes se deben aceptar o rechazar a un α = 1%.

9.15 En una ciudad se toma una muestra aleatoria de 64 familias para estimar, a un nivel de confianza de 94%, el gasto promedio semanal en alimentos. Por datos anteriores se sabe que la desviación estándar es de $10 y por la encuesta se sabe que el gasto promedio (muestral) semanal en alimentos resultó ser de $32.

9.16 Se toma una muestra aleatoria de 64 carros en cierta autopista durante cierto periodo y se encuentra que la velocidad promedio es de 60 kilómetros por hora. La desviación normal es de 15 kilómetros por hora. Encuentre un intervalo al 96% de confianza para la velocidad promedio de los carros.

9.17 Un bioquímico desea estimar el contenido vitamínico de cierto alimento. Se toma una muestra de tamaño 49 y se encuentra que el contenido promedio (muestral) de vitaminas por cada 100 gr es de 12 mg. Se estima que la desviación es de 2 mg. Encuentre el intervalo de confianza de 98% para el contenido promedio de vitamínico por cada 100 gr.

9.18 Una muestra de láminas de vidrio se toma al azar para estimar el grosor promedio. Por estudios anteriores se sabe que la desviación estándar es de 1.2 mm. Si el error debe hallarse dentro de un rango de 0.3 mm con un nivel de confianza de 96%, ¿cuál debe ser el tamaño de la muestra?

9.19 Un bioquímico desea estimar el promedio del contenido vitamínico de un alimento y sabe que por estudios anteriores se puede suponer que la desviación estándar es de 2 mg. Si el error debe ser de +/– 0.5 mg con un nivel de confianza de 94%, ¿de qué tamaño debe ser el tamaño de la muestra?

9.20 Para embarcar una fibra de algodón debe estimarse la longitud promedio. Por estudios anteriores se sabe que es posible suponer que la desviación estándar es de 50 mm. Si el error debe hallarse dentro de un rango de

15 mm, ¿qué tamaño de muestra debe tomarse para tener un nivel de confianza de 92%?

9.21 Una compañía de teléfonos sabe por su experiencia que el número promedio de llamadas mensuales por familia es de 84 y que la desviación normal es de 15 llamadas mensuales por familia. Se desarrolló una campaña publicitaria con el fin de que las familias instalaran teléfonos en sus alcobas y cocinas, y usaran el teléfono con frecuencia. Una encuesta reciente de 100 familias demostró que el número promedio (muestral) de llamadas aumentó a 89. ¿Esta muestra apoya el argumento de que la campaña publicitaria aumentó de manera significativa (a un nivel de 1%) las llamadas?

9.22 El rendimiento promedio de trigo en las granjas hace tiempo era de 22 bushels por acre con una desviación estándar de 8 bushels. Se dio un nuevo fertilizante a un grupo de 52 granjas seleccionadas al azar. Su promedio de rendimiento fue de 24 bushels por acre. ¿Puede afirmarse que hubo un incremento significativo (a un nivel de 2%) en el rendimiento?

9.23 El promedio de calificaciones del examen de admisión de 800 estudiantes en la universidad H fue de 500 puntos y la desviación normal de 40 puntos. Hubo 16 estudiantes del colegio A que obtuvieron un promedio de 485 puntos. ¿Podrían considerarse las calificaciones de los estudiantes del colegio A como significativamente (a un nivel de 10%) menores que la calificación promedio total?

9.24 Una compañía controla su proceso de producción en tal forma que se pone en bolsitas un promedio de 20 g de un producto químico, con una desviación estándar de 2 g. El proceso se detendrá cuando el promedio no sea de 20 g al tomar muestras aleatorias de tamaño 25. La última muestra aleatoria arrojó un promedio de 19 g. ¿Debe detenerse el proceso y ajustarse si el nivel de significancia es de 5%?

9.25 Investigaciones previas indican que 20% de las familias de una ciudad están suscritas al periódico K. Hay motivos para creer que en los últimos tiempos disminuyó la tasa de suscripción. Para comprobar si en realidad hubo cambio se toma una muestra al azar de 100 familias y se observa que la proporción muestral es de 0.16 (16%). ¿Ha sido significativa (a un nivel de 10%) la disminución de suscripción?

9.26 En enero, 40% de 2000 comerciantes comunican que van a aumentar su demanda de lavadoras; en marzo puede pensarse que ya realizó el incremento. Para comprobarlo se toma una muestra al azar de 400 comerciantes y se comprueba que la proporción muestral es de 46% (= 0.46). ¿Ha sido significativo (a un nivel de 1%) el aumento de la demanda?

9.27 Una compañía productora de transitores tiene un promedio de 20% de defectuosos. Cada hora se toma una muestra aleatoria de 100 transistores para llevar el control de calidad. A las 3 pm se encuentra un promedio de 28% (0.28) de defectuosos. ¿Debe pararse el proceso de producción para regularlo? El nivel de significancia es de 5%.

9.28 Una moneda se lanzó 1000 veces y apareció cara 520 veces, ¿es esta moneda una moneda regular (α = 5%.)?

9.29 Un periódico afirma que por lo menos lo leen el 30%; se toma una muestra aleatoria de 500 familias y se encuentra que 230 lo leen. ¿Este resultado confirma lo dicho? Suponga que $\alpha = 5\%$.

9.30 A una muestra aleatoria de 4 000 amas de casa que han escuchado cierto programa, se les preguntó sobre la efectividad de la propaganda, 420 amas de casa dijeron que habían sido inducidas a comprar cierto producto. ¿Este resultado da suficiente evidencia para que los patrocinadores del programa digan que la mayoría de las amas de casa serán inducidas a comprar el producto? Suponga que el nivel de significancia es de 5%.

Estudio de dos poblaciones 10

PROPÓSITOS

Con la revisión de los contenidos, así como la realización de los ejercicios, quienes estudien este capítulo serán capaces de utilizar adecuadamente los modelos estadísticos, considerando los requisitos y condiciones de uso para construir los intervalos y contrastar los parámetros que intervienen en el estudio de dos poblaciones.

Desarrollar el proceso de cálculo para construir el intervalo de confianza y realizar el contraste de hipótesis para la diferencia de:

1. Dos medias, cuando las varianzas son conocidas.
2. Dos medias, cuando las varianzas son desconocidas:

   a) muestras independientes.
   b) muestras dependientes.
3. Dos proporciones.
4. Dos varianzas.
5. Efectuar el contraste de hipótesis para más de dos varianzas.
6. Aplicar adecuadamente las pruebas de:

   a) Cochran
   b) Hartley
   c) Bartlett.

INTRODUCCIÓN

La comprobación de hipótesis estadísticas es quizá el área más importante de la inferencia estadística y, por consiguiente, de la teoría de la decisión.

Dado que una hipótesis estadística que comprende a una o más poblaciones es una suposición o afirmación establecida por el investigador, puede ser verdadera o falsa. La verdad o falsedad de una hipótesis estadística nunca se conoce con certidumbre a menos que se examine a la población total, lo cual, por supuesto, es imposible. Por lo tanto, se escoge una muestra al azar de la población en estudio y se utiliza la información contenida en dicha muestra para decidir si la hipótesis nula es falsa o verdadera. Si la información obtenida de esa manera muestra inconsistencia con la afirmación establecida se rechaza la hipótesis; si ocurre lo contrario la hipótesis no se rechaza. Debe tomarse en cuenta que una hipótesis estadística no se rechaza porque la evidencia o información es insuficiente para rechazarla, pero esto no necesariamente implica que dicha hipótesis es verdadera. Cuando las hipótesis establezcan las diferencias entre dos grupos de observaciones, deben obtenerse tales observaciones en dos muestras; en los resultados finales se compararán dos poblaciones de donde provienen las muestras. Esta situación es diferente a la de hacer inferencias a partir de una sola muestra proveniente de una sola población. Las pruebas estadísticas para llevar a cabo esto serán la Z y la *t de Student*, así como la prueba de Sandler.

El intervalo de confianza, consiste en un par de valores, entre los cuales se encuentra el parámetro por estimar, con una confianza $(1 - \alpha)$ determinada. Como el intervalo de confianza se construye a partir de los datos, nos provee información complementaria; esto hace que la inferencia estadística sea más consistente y confiable.

Puesto que en la estadística paramétrica la homogeneidad o igualdad de las varianzas, también conocida como homoscedasticidad, es un requisito muy importante para el contraste de hipótesis con las pruebas Z y t, así como en el ANOVA y la regresión lineal, en este capítulo se considerarán varias pruebas para contrastar las varianzas, como Cochran, Hartley y Bartlett.

10.1 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA DOS MEDIAS

CASO I. Varianzas conocidas

Sean dos variables aleatorias X_1, X_2, independientes con medias $\mu_{\overline{X}_1}$, $\mu_{\overline{X}_2}$ y varianzas conocidas $\sigma^2_{\overline{X}_1}$, $\sigma^2_{\overline{X}_2}$; con sus respectivas medias muestrales $\overline{X}_1$, $\overline{X}_2$ y n_1, n_2, así como S_1^2 y S_2^2, calculadas.

Entonces:

$$\mu_{\overline{X}_1} = \mu_{X_1} \; ; \quad \sigma^2_{\overline{X}_1} = \frac{\sigma^2_{\overline{X}_1}}{n_1} \quad ; \quad X_1 \sim N\,(X_1 \; ; \sigma_{\overline{X}_1} \, , \, \sigma^2_{\overline{X}_1})$$

$$\mu_{\overline{X}_2} = \mu_{X_2} \; ; \quad \sigma^2_{\overline{X}_2} = \frac{\sigma^2_{X_2}}{n_2} \quad ; \quad X_1 \sim N\,(X_2 \; ; \mu_{\overline{X}_2} \, \sigma^2_{\overline{X}_2})$$

y como

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}, \quad \text{también para las varianzas}$$

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}$$

Entonces, la diferencia de las medias se distribuye en forma normal.

$$\overline{X}_1 - \overline{X}_2 \sim N\,(\overline{X}_1 - \overline{X}_2 \; ; \mu_{\overline{X}_1} - \mu_{\overline{X}_2} \, , \quad \sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2})$$

Por ello, el intervalo de confianza a un nivel $(1 - \alpha)$, para la diferencia de medias es:

$$(\overline{X}_1 - \overline{X}_2) - \sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}} \;\; Z_{\alpha/2} \;\; \leq \mu_{X_1} - \mu_{X_2} \leq \; (\overline{X}_1 - \overline{X}_2) + \sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}} \;\; Z_{\alpha/2}$$

o

$$(\overline{X}_1 - \overline{X}_2) \; - \sqrt{\frac{\sigma^2_{\overline{X}_1}}{n_1} + \frac{\sigma^2_{\overline{X}_2}}{n_2}} \;\; Z_{\alpha/2} \;\; \leq \mu_{X_1} - \mu_{X_2} \leq (\overline{X}_1 - \overline{X}_2) + \sqrt{\frac{\sigma^2_{\overline{X}_2}}{n_1} + \frac{\sigma^2_{\overline{X}_2}}{n_2}} \;\; Z_{\alpha/2}$$

como $\sigma^2_{\overline{X}_1} = \sigma^2_{\overline{X}_2} = \sigma^2_X$, el intervalo de confianza puede quedar como un caso particular:

$$(\overline{X}_1 - \overline{X}_2) - \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \;\; \sigma_X \;\; Z_{\alpha/2} \;\; \leq \mu_{X_1} - \mu_{X_2} \leq \overline{X}_1 - \overline{X}_2 + \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \;\; \sigma_X \;\; Z_{\alpha/2}$$

Ejemplo 1 En una línea de producción de ciertos dispositivos alimentada por dos máquinas con tecnologías diferentes se obtienen los siguientes datos:

MÁQUINA 1	MÁQUINA 2
$n_1 = 25$	$n_2 = 20$
$\overline{X}_1 = 12.2$ cm	$\overline{X}_2 = 10.5$ cm
$\sigma_{X_1} = 1.5$ cm	$\sigma_{X_2} = 0.5$ cm

Construya un intervalo de confianza al 95%

como $\mu_{X_1} = 1.5$, $\sigma^2_{X_1} = 2.25$, $\mu_{X_2} = 0.5$ y $\sigma^2_{X_2} = 0.25$

son conocidas y $(1 - \alpha) = 1 - 0.05 = 0.95$, por lo que $Z_{\alpha/2} = 1.96$, se utiliza el caso:

$$(\overline{X}_1 - \overline{X}_2) - \sqrt{\frac{\sigma^2_{\overline{X}_1}}{n_1} + \frac{\sigma^2_{\overline{X}_2}}{n_2}}\, Z_{\alpha/2} \le \mu_{\overline{X}_1} - \mu_{X_2} \le (\overline{X}_1 - \overline{X}_2) + \sqrt{\frac{\sigma^2_{\overline{X}_1}}{n_1} + \frac{\sigma^2_{\overline{X}_2}}{n_2}}\, Z_{\alpha/2}$$

sustituyendo:

$$(12.2 - 10.5) - \sqrt{\frac{2.25}{25} + \frac{0.25}{20}}\,(1.96) \le \mu_{X_1} - \mu_{X_2} \le (12.2 - 10.5)$$

$$+ \sqrt{\frac{2.25}{25} + \frac{0.25}{20}}\,(1.96)$$

$$\boxed{1.07 \le \mu_{X_1} - \mu_{X_2} \le 2.33}$$

lo que significa que a una confianza de 95% la diferencia en las longitudes promedio de los dispositivos producidos por las dos máquinas está entre 1.07 y 2.33 cm.

Las pruebas de hipótesis para la igualdad de dos medias si las varianzas son conocidas, X_1 y X_2 son independientes y si $\sigma^2_{\overline{X}_1} = \sigma^2_{\overline{X}_2}$ implica que

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = 0 \text{ o también, } \mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$$

Caso	*Estadística de prueba*	*Regla de decisión*
Si $\overline{X}_1 < \overline{X}_2$		
$H_0 : \mu_{X_1} = \mu_{X_2}$	$Z_C = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma^2_{X_1}}{n_1} + \dfrac{\sigma^2_{X_2}}{n_2}}}$	Si $Z_C \geq Z_\alpha$ Rechazar H_0
$H_1 : \mu_{X_1} < \mu_{X_2}$		Si $Z_C > Z_\alpha$ No rechazar H_0
Si $\overline{X}_1 > \overline{X}_2$		
$H_0 : \mu_{X_1} = \mu_{X_2}$	$Z_C = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma^2_{X_1}}{n_1} + \dfrac{\sigma^2_{X_2}}{n_2}}}$	Si $Z_C \geq Z_\alpha$ Rechazar H_0
$H_1 : \mu_{X_1} > \mu_{X_2}$		Si $Z_C < Z_\alpha$ No rechazar H_0
Control		
$H_0 : \mu_{X_1} = \mu_{X_2}$	$Z_C = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma^2_{X_1}}{n_1} + \dfrac{\sigma^2_{X_2}}{n_2}}}$	Si $\lvert Z_C \rvert \geq Z_{\alpha/2}$ Rechazar H_0
$H_1 : \mu_{X_1} \neq \mu_{X_2}$		Si $\lvert Z_C \rvert < Z_{\alpha/2}$ No rechazar H_0

Cuando las muestras son grandes ($n \to \infty$), por medio del teorema central del límite $\sqrt{\dfrac{\sigma^2_{X_1}}{n_1} + \dfrac{\sigma^2_{X_2}}{n_2}}$ puede ser sustituido por $\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$.

Ejemplo 2 Un fabricante de cigarrillos publicita que la marca 1 no es más dañina para la salud que la marca 2 (con filtro). Suponiendo que el daño a la salud está asociado con el contenido de nicotina, el instituto del consumidor tomó al azar dos muestras de cigarrillos de la marca 1 y de la 2 y midió la cantidad de nicotina con la misma técnica, de donde surgieron los siguientes datos:

MARCA 1	MARCA 2
$n_1 = 125$	$n_2 = 180$
$\overline{X}_1 = 24.6$ mg	$\overline{X}_2 = 24.3$ mg
$S_1 = 1.4$ mg	$S_2 = 1.1$ mg
$S_1^2 = (1.4)^2 = 1.96$	$S_2^2 = (1.1)^2 = 1.21$

Establezca las hipótesis adecuadas y contrástelas utilizando un nivel de significancia de 5%.

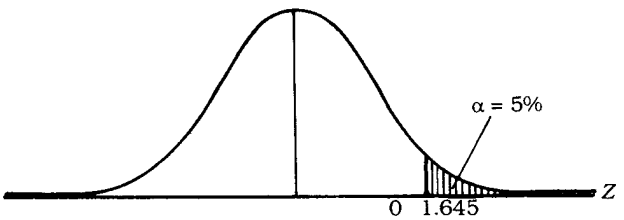
Solución

$H_0 : \mu_{X_1} = \mu_{X_2}$

$H_1 : \mu_{X_1} > \mu_{X_2}$

$\alpha = 0.05$

$$Z_C = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{24.6 - 24.3}{\sqrt{\frac{1.96}{125} + \frac{1.21}{180}}} \; ; \; Z_C = 2.00$$



como $\alpha = 5\%$, la región crítica es:
como $Z_C > Z_{0.05}$, o sea, $2.00 > 1.645$, H_0 se rechaza.

Conclusión

La marca 1 de cigarrillos es más dañina que la marca 2.

10.2 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA DOS MEDIAS (MUESTRAS INDEPENDIENTES)

CASO II. Varianzas desconocidas

Como se ha visto anteriormente:

$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2} = \mu_{\overline{X}_1 - \overline{X}_2} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$; si las varianzas son iguales pero desconocidas se puede obtener la varianza combinada.[1]

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}$$

[1] También conocida como *varianza Pooled.*

Entonces $S_p = \sqrt{S_p^2}$

$$t = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1} - \overline{X}_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

El intervalo buscado es:

$$\overline{X}_1 - \overline{X}_2 - S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2,\, n_1 + n_2 - 2} \leq \mu_{X_1} - \mu_{X_2} \leq \overline{X}_1 - \overline{X}_2 + S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2,\, n_1 + n_2 - 2}$$

donde $\overline{X}_1$ y $\overline{X}_2$ son las medias muestrales calculadas y $gl = n_1 + n_2 - 2$.

Ejemplo 3 Supongamos que en el ejemplo 1 las desviaciones estándar poblacionales de las longitudes de los dispositivos producidos por las dos máquinas no son conocidas, pero se calculan a partir de los datos de las muestras, se obtiene lo siguiente:

MÁQUINA 1	MÁQUINA 2
$n_1 = 25$	$n_2 = 20$
$\overline{X}_1 = 12.2$	$\overline{X}_2 = 10.5$
$S_1 = 1.2$	$S_2 = 0.3$
$S_1^2 = 1.44$	$S_2^2 = 0.09$

Construir un intervalo de confianza al 95% de la diferencia de dos medias.

Como $n_1 = 25$ y $n_2 = 20$ son pequeñas, la aproximación de $\sigma_{X_1}^2$ y $\sigma_{X_2}^2$ por S_1^2 y S_2^2 no se justifica.

Si suponemos que las varianzas poblacionales son iguales (aunque desconocidas), o sea $\sigma_{X_1}^2 = \sigma_{X_2}^2 = \sigma_X^2 = ?$

$$\overline{X}_1 - \overline{X}_2 - S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2,\, n_1 + n_2 - 2} \leq \mu_{X_1} - \mu_{X_2} \leq \overline{X}_1 - \overline{X}_2 + S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{\alpha/2,\, n_1 + n_2 - 2}$$

sustituyendo y efectuando los cálculos se obtiene:

$$\overline{X}_1 - \overline{X}_2 = 12.2 - 10.5 = 1.7 \quad \text{como } t_{\alpha/2,\, n_1 + n_2 - 2} = t_{0.025,43 = 1.96}$$

$$1.7 - 1.96 \sqrt{\frac{(24)(1.44) + (19)(0.09)}{43}} \sqrt{\frac{1}{25} + \frac{1}{20}} \leq \mu_{X_1} - \mu_{X_2} \leq$$

$$1.7 + 1.96 \sqrt{\frac{(24)(1.44) + (19)(0.09)}{43}} \sqrt{\frac{1}{25} + \frac{1}{20}}$$

$$1.7 - 1.96 \sqrt{0.84349}\,(0.3) \leq \mu_{X_1} - \mu_{X_2} \leq 1.7 + 1.96 \sqrt{0.84349}\,(0.3)$$

$$1.7 - 1.96\,(0.27553) \leq \mu_{X_1} - \mu_{X_2} \leq 1.7 + 1.96\,(0.27553)$$

$$\boxed{1.16 \leq \mu_{X_1} - \mu_{X_2} \leq 2.24}$$

Pruebas de hipótesis para la igualdad de dos medias si $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2_X = ?$ donde X_1 y X_2 son independientes

Caso	*Estadística de prueba*	*Regla de decisión*
$\overline{X}_1 < \overline{X}_2$ $H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} < \mu_{X_2}$	$t_C = \dfrac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	Si $t_C \geq t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$ Si $t_C < t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$
$\overline{X}_1 > \overline{X}_2$ $H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} > \mu_{X_2}$	$t_C = \dfrac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	Si $t_C \geq t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$ Si $t_C < t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$
Control $\overline{X}_1 \neq \overline{X}_2$ $H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} <> \mu_{X_2}$	$t_C = \dfrac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	Si $t_C \geq t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$ Si $t_C < t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$

$$\text{donde} \quad S_p = \sqrt{\frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}} \quad \text{y} \quad gl = n_1 + n_2 - 2$$

Ejemplo 4 Un fabricante de podadoras de césped eléctricas decide poner a prueba un modelo experimental, del que se espera que tenga un consumo mínimo de aceite; se obtienen los siguientes datos:

Modelo experimental	*Modelo tradicional*
$\overline{X}_1 = 78.2$ Ha/litro	$\overline{X}_2 = 50.6$ Ha/litro
$S_1 = 3.2$ Ha/litro	$S_2 = 2.5$ Ha/litro
$n_1 = 20$	$n_2 = 10$

¿El modelo experimental debe reemplazar al tradicional (con un nivel de significancia de 1%)?

$H_0 : \mu_{X_1} = \mu_{X_2}$

$H_1 : \mu_{X_2} > \mu_{X_2}$

Utilizando el modelo estadístico $t = \dfrac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$

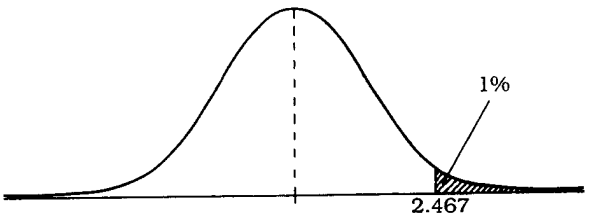
sustituyendo los valores:

$$t_C = \frac{78.2 - 50.6}{\sqrt{\dfrac{(19)(3.2)^2 + (9)(2.5)^2}{20 + 10 - 2}}\ \sqrt{\dfrac{1}{20} + \dfrac{1}{10}}} = 23.81$$

$$\therefore \boxed{t_C = 23.81}$$

$gl = 20 + 10 - 2 = 28$

Región crítica



$t_{28,\,0.01} = 2.467$
Regla de decisión
como $23.81 > 2.467$
H_0 se rechaza

Conclusión

El modelo experimental utiliza menos aceite, por lo que se justifica reemplazar el modelo tradicional.

Cuando las varianzas son desconocidas y diferentes $\sigma^2_{X_1} \neq \sigma^2_{X_2} = ?$

Entonces:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{\sigma^2_{X_1}}{n_1} + \dfrac{\sigma^2_{X_2}}{n_2}}} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}}{n_1\,n_2}}} \; \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\,n_2}}\,\sqrt{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}}} \sim Z$$

y

$$\frac{(n_1 - 1)\,S_1^2}{\sigma^2_{X_1}} + \frac{(n_2 - 1)\,S_2^2}{\sigma^2_{X_2}} \sim \chi^2_{n_1 + n_2 - 2} \qquad \text{pero}$$

$$\frac{(n_1 - 1)\,S_1^2}{\sigma^2_{X_1}} + \frac{(n_2 - 1)\,S_2^2}{\sigma^2_{X_2}} = \frac{\sigma^2_{X_2}\,(n_1 - 1)\,S_1^2 + \sigma^2_{X_1}\,(n_2 - 1)\,S_2^2}{\sigma^2_{X_1}\,\sigma^2_{X_2}} \times \frac{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}}{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}}$$

$$= \frac{(n_1 - 1)\,S_1^2\,\sigma^2_{X_2} + (n_2 - 1)\,S_2^2\,\sigma^2_{X_1}}{n_2\,\sigma^2_{X_1} + n_1\sigma^2_{X_2}} \times \frac{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}}{\sigma^2_{X_1}\,\sigma^2_{X_2}}$$

$$= \frac{(n_1 - 1)\,S_1^2\,\sigma^2_{X_2} + (n_2 - 1)\,S_2^2\,\sigma^2_{X_1}}{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}} \times \left(\frac{n_2 + n_1}{\sigma^2_{X_2}\,\sigma^2_{X_1}}\right)$$

como

$$\hat{\sigma}^2_X = S^2 = \frac{(n_1 - 1)\,S_1^2\,S_2^2 + (n_2 - 1)\,S_2^2\,S_1^2}{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}} \times \left(\frac{n_1}{S_1^2} + \frac{n_2}{S_2^2}\right)$$

$$\Rightarrow \frac{(n_1 - 1)\,S_1^2}{\sigma^2_{X_1}} + \frac{(n_2 - 1)\,S_2^2}{\sigma^2_{X_2}} \overset{\circ}{=} \frac{(n_1 + n_2 - 2)\,S_1^2\,S_2^2}{n_2\,\sigma^2_{X_1} + n_1\,\sigma^2_{X_2}} \times \left(\frac{n_1}{S_1^2} + \frac{n_2}{S_2^2}\right)$$

$$\frac{\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\, n_2}}\ \sqrt{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}}{\sqrt{\dfrac{1}{n_1 + n_2 - 2} \times \dfrac{(n_1 + n_2 - 2)\, S_1^2\, S_2^2}{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}} \times \left(\dfrac{n_1}{S_1^2} + \dfrac{n_2}{S_2^2}\right)}} = \frac{\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\, n_2}}\ \sqrt{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}}{\sqrt{\left(\dfrac{n_1}{S_1^2} + \dfrac{n_2}{S_2^2}\right) \dfrac{S_1^2\, S_2^2}{n_2 \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}}$$

$$= \frac{\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\, n_2}}\ \sqrt{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}}{\sqrt{\dfrac{n_1}{S_1^2} + \dfrac{n_2}{S_2^2}}\ \dfrac{S_1\ \ S_2}{\sqrt{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}}$$

$$\Rightarrow \frac{\dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\, n_2}}\ \sqrt{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}}{\sqrt{\dfrac{n_1}{S_1^2} + \dfrac{n_2}{S_2^2}}\ \dfrac{S_1\, S_2}{\sqrt{n_2\, \sigma^2_{X_1} + n_1\, \sigma^2_{X_2}}}} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\, n_2}}\ \sqrt{\dfrac{n_1}{S_1^2} + \dfrac{n_2}{S_2^2}}\ \sqrt{S_1^2\, S_2^2}}$$

$$= \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{1}{n_1\, n_2}\ \dfrac{n_1\, S_2^2 + n_2\, S_1^2}{S_1^2\, S_2^2}\ \dfrac{S_1^2\, S_2^2}{1}}}$$

$$\text{implica}\ \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{S_2^2}{n_1} + \dfrac{S_1^2}{n_2}}}$$

$$\therefore\ t_c = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_{\overline{X}_1 - \overline{X}_2})}{\sqrt{\dfrac{S_2^2}{n_1} + \dfrac{S_1^2}{n_2}}} \sim t_{gl}$$

$$\text{donde} \qquad gl = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 + 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 + 1}} - 2$$

Intervalo de confianza para la diferencia de dos medias si $\sigma^2_{X_1} \neq \sigma^2_{X_2}$ y X_1 y X_2 son independientes

$$\text{Como} \qquad t_c = \frac{\overline{X}_1 - \overline{X}_2 - \mu_{\overline{X}_1 - \overline{X}_2}}{\sqrt{\dfrac{S_2^2}{n_1} + \dfrac{S_1^2}{n_2}}} \sim t_{gl}$$

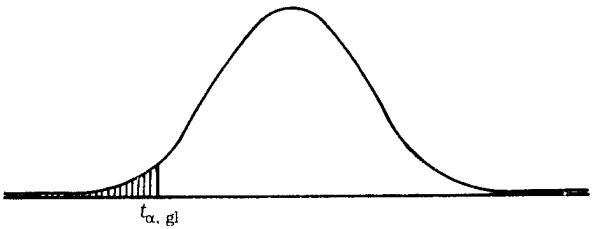
$$\text{y} \qquad gl = \frac{\left(\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{S_1^2}{n_1}\right)^2}{n_1 + 1} + \dfrac{\left(\dfrac{S_2^2}{n_2}\right)^2}{n_2 + 1}} - 2 \qquad \text{así como } \mu_{\overline{X}_1 - \overline{X}_2} = \mu_{X_1} - \mu_{X_2}$$

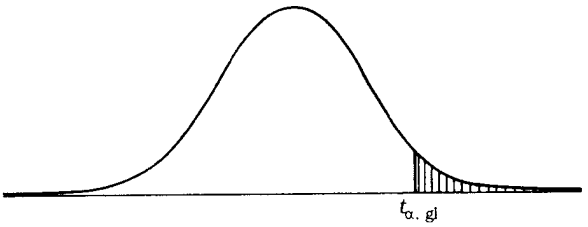
$\Rightarrow$ intervalo de confianza a un nivel $(1 - \alpha)$ será:

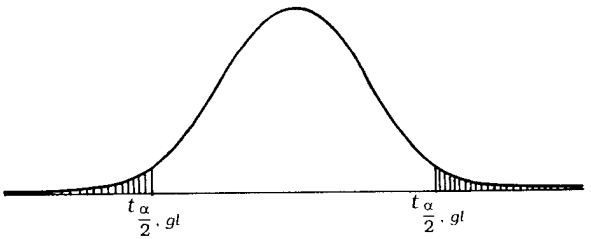
$$\left(\overline{X}_1 - \overline{X}_2 - \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\ t_{\alpha/2,\, gl} \leq \mu_{X_1} - \mu_{X_2} \leq \overline{X}_1 - \overline{X}_2 + \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\ t_{\alpha/2,\, gl}\right) = 1 - \alpha$$

Pruebas de hipótesis para la igualdad de dos medias si $\sigma^2_{\bar{X}_1} \neq \sigma^2_{\bar{X}_2} = ?$ X_1 y X_2 son independientes

Caso	*Estadística de prueba*	*Regla de decisión*
$\bar{X}_1 < \bar{X}_2$ $H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} < \mu_{X_2}$	$t_C = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	Si $t_C < -t_{\alpha,\, gl} \Rightarrow$ Rechazar H_0, o también Si $\lvert t_C \rvert \geq t_{\alpha,\, gl} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_2$ Si $t_C > t_{\alpha,\, gl} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_2$
$\bar{X}_1 > \bar{X}_2$ $H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} > \mu_{X_2}$	$t_C = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	Si $t_C \geq t_{\alpha,\, gl} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_2$ Si $t_C < t_{\alpha,\, gl} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_2$
Control		
$\bar{X}_1 \neq \bar{X}_2$ $H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} <> \mu_{X_2}$	$t_C = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	Si $t_C \geq t_{\alpha/2,\, gl} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_2$ Si $t_C < t_{\alpha/2,\, gl} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_2$







10.2.1 Intervalo de confianza y contraste de hipótesis para dos medias (muestras dependientes, relacionadas o una sola muestra medida dos veces)

Cuando se desea comparar o estimar la diferencia de dos medias muestrales dependientes, es decir, cuando de una población se extrae una muestra al azar que se sujeta a medición y proporciona los primeros datos (antes), posteriormente se le aplica una variable control (nuevo método de enseñanza-aprendizaje, un tratamiento para alguna enfermedad, dieta específica, terapia, etc.). Una vez que a dicha muestra se le aplicó la variable control determinada, se somete nuevamente a medición a fin de obtener los segundos datos (después). Algunos autores llaman a este diseño *antes-después* o *muestras dependientes*.

El método estadístico que se utilizará para este diseño también puede aplicarse al de muestras relacionadas (correlacionadas) o en pares (pareadas); o sea dos muestras de gemelos, parejas con una relación emocional, profesional de intereses o cualquier relación biunívoca de varios años. En este caso una muestra se obtiene en forma aleatoria, pero la segunda muestra está determinada por la primera.

Las varianzas poblacionales son desconocidas, la variable de interés se distribuye normalmente, por lo que en este caso se considerará la diferencia $D = X_1 - X_2$ de la variable de interés en la muestra medida dos veces o de cada una de las muestras pareadas o relacionadas.

Sea $D = X_1 - X_2$ con $X_1 \sim n\,(X_{ij}\,;\, \mu_{X_1}, \sigma^2_{X_1})$ $\quad i = 1, 2$ con $\sigma^2_{X_1} = ?$ y $\sigma^2_{X_2} = ?$

$$\Rightarrow \mu_D = \mu_{X_1} - \mu_{X_2} \Rightarrow \mu_D = \mu_{X_1} - \mu_{X_2}$$

$$\overline{D} = \frac{\sum_{j=1}^{n} Dj}{n} \Rightarrow \overline{D} = \frac{\sum_{j=1}^{n} (X_{1j} - X_{2j})}{n}$$

$$= \frac{\sum_{j=1}^{n} X_{1j} - \sum_{j=1}^{n} X_{2j}}{n}$$

$$= \frac{\sum_{j=1}^{n} X_{1j}}{n} - \frac{\sum_{j=1}^{n} X_{2j}}{n}$$

$$\Rightarrow \overline{D} = \overline{X}_1 - \overline{X}_2 \qquad \Rightarrow \qquad \overline{D} \sim n\,(\overline{d};\, \mu_{\overline{D}}, \sigma^2_{\overline{D}})$$

$$\Rightarrow \mu_{\overline{D}} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$$

$$= \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$$

$\Rightarrow \mu_{\overline{D}} = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$

$$\text{Si} \quad S^2 = \frac{\sum_{j=1}^{n} (D_j - \overline{D})^2}{n-1} \Rightarrow S^2 = \frac{\sum_{j=1}^{n} D_j^2 - n\,\overline{D}^2}{n-1}$$

$$\Rightarrow \frac{\overline{D} - \mu_{\overline{D}}}{\sqrt{\frac{1}{n} S^2}} \sim t_{n-1}$$

El intervalo de confianza a un nivel $(1 - \alpha)$ es el siguiente:

$$\overline{D} - t_{gl,\alpha/2} \frac{S}{\sqrt{n}} \leq \mu_{\overline{D}} \leq \overline{D} + t_{gl,\alpha/2} \frac{S}{\sqrt{n}}$$

donde $gl = n - 1$

Ejemplo Construir un intervalo de confianza al 95% con los resultados que un maestro de español obtuvo al aplicar un nuevo método de enseñanza en uno de sus grupos. Las puntuaciones obtenidas son:

n	Antes (1)	Después (2)	D_i	D_i^2
1	12	20	− 8	64
2	10	10	0	0
3	24	18	+ 6	36
4	12	9	+ 3	9
5	28	17	+ 11	121
6	16	9	+ 7	49
7	26	16	+ 10	100
8	18	8	+ 10	100
9	24	16	+ 8	64
10	8	4	+ 4	16
11	27	15	+ 12	144
12	4	2	+ 2	4
13	22	14	+ 8	64
14	19	13	+ 6	36
15	18	12	+ 6	36
Σ	268	183	+ 85	843

$$\overline{X}_1 = 17.86667 \qquad \overline{X}_2 = 12.2000$$

$$\overline{D}_C = \overline{X}_1 - \overline{X}_2 = 5.667, \qquad S_C^2 = \frac{843 - 15\,(5.667)}{14}$$

$$t_{14,0.025} = 2.145 \qquad S_C^2 = \frac{843\,481.72334}{14}$$

$$S_C^2 = 25.805, \quad S_C = 5.08$$

$$5.667 - (2.145)\left(\frac{5.08}{\sqrt{15}}\right) \leq \mu_{\overline{D}} \leq 5.667 + 2.145\left(\frac{5.08}{\sqrt{15}}\right)$$

$$5.6667 - 2.81349 \leq \mu_{\overline{D}} \leq 5.667 + 2.81349$$

$$\boxed{2.85 \leq \mu_{\overline{D}} \leq 8.48}$$

Prueba de hipótesis para la igualdad de dos medias por pares si $\sigma^2_{X_1}$ = ? y $\sigma^2_{X_2}$ = ? (si $\mu_{X_1} = \mu_{X_2} \Rightarrow \mu_D = \mu_{X_1} - \mu_X = 0$)

Caso	*Estadística de prueba*	*Regla de decisión*
$H_0 : \mu_{X_1} = \mu_{X_2}$ $H_1 : \mu_{X_1} <> \mu_{X_2}$	$t_C = \dfrac{\overline{D}_C}{\sqrt{\frac{1}{n}}\, S_C}$	Si $t_C \geq t_{\alpha/2,\, n-1} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$ Si $t_C < t_{\alpha/2,\, n-1} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_{X_2}$

Ejemplo La propaganda de un refresco dietético asegura que si se toma a diario y por un mes se obtendrá una pérdida de peso de dos libras; el Instituto del Consumidor sospecha que esta propaganda es falsa, por lo que realiza un estudio con 12 personas dispuestas voluntariamente a llevar a cabo dicha investigación, obteniéndose los siguientes resultados. Utilice un α = 5%.

i Sujeto	X_1 Antes	X_2 Después	D Diferencia
1	126	115	11
2	194	179	15
3	135	124	11
4	179	163	16
5	205	186	19
6	139	137	2
7	142	146	− 4
8	172	161	11
9	159	160	− 1
10	194	198	− 4
11	164	158	6
12	139	125	14

$$\sum_{i=1}^{12} D_i = 96$$

$$D_C = \frac{96}{12} = 8$$

$$S_C = 8.01$$

$$t_{5\%,\,11} = 1.80$$

$$H_0 : \mu_{X_1} - \mu_{X_2} = 2:$$
$$H_1 : \mu_{X_1} - \mu_{X_2} < 2:$$

$$t_C = \frac{\overline{D}_C - (\mu_{X_1} - \mu_{X_2})}{\frac{S_C}{\sqrt{n}}}$$

$$t_C = \frac{8 - 2}{\frac{8.01}{\sqrt{12}}} = 2.595$$

$$t_C = 2.595$$

Si $|t_C| \geq t_{11,0.025}$, entonces H_0 se rechaza; como $2.595 > 2.201$, H_0 se rechaza.

Conclusión El refresco sí es dietético; la sospecha del Instituto del Consumidor no procede.

Ejemplo Se retomará el ejemplo de la sección anterior. Además de un intervalo de confianza, el profesor de español quiere realizar un contraste de hipótesis para confirmar si existe diferencia en los datos recabados con una confiabilidad de 95% ($\alpha = 5\%$).

Paso 1. La variable de interés es saber si funciona adecuadamente el método de enseñanza que propone el maestro. Por ello recurre a comparar medias en medias dependientes, es decir, se obtienen datos antes y después.

Paso 2. Proponer la hipótesis estadística.

$$H_0 : \mu_{\overline{D}} = 0$$
$$H_1 : \mu_{\overline{D}} \neq 0$$

Paso 3. Establecer el error que se está dispuesto a aceptar.

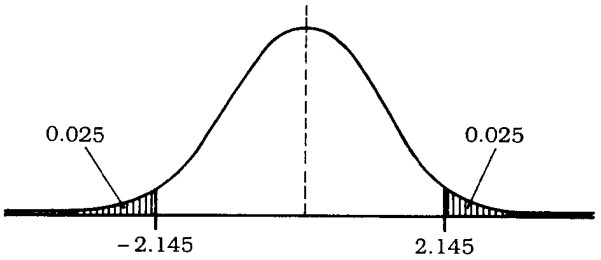
$$\alpha = 5\% \Rightarrow 1 - \alpha = 95\%$$

Paso 4. Determinar el modelo estadístico adecuado.

$$t_C = \frac{\overline{D}_C - \mu_{\overline{D}}}{\frac{S_C}{\sqrt{n}}}$$

Paso 5. Obtener mediante la tabla correspondiente el valor crítico y ubicar la zona de rechazo en una gráfica.

$\alpha = 0.05$
$\alpha/2 = 0.025$
$gl = n - 1 = 14$
$t_{14,\ 0.025} = 2.145$



Paso 6. Establecer la regla de decisión para H_0.

si $t_C < -2.145$, entonces H_0 se rechaza

si $t_C > 2.145$, entonces H_0 se rechaza

$|t_C| \geq 2.145$, entonces H_0 se rechaza.

Paso 7. Calcular el valor del estadístico de prueba utilizando el modelo propuesto.

$$t_C = \frac{\overline{D} - \mu_D}{\frac{S_C}{\sqrt{n}}} \Rightarrow t_C = \frac{5.667 - 0}{\frac{5.0805}{\sqrt{15}}}$$

$$\boxed{t_C = 4.3201}$$

Paso 8. Obtener conclusiones.

Como $4.320 > 2.145$, entonces se rechaza H_0 con una confiabilidad de 95%, es decir, sí existe diferencia en las medias de las muestras dependientes estudiadas; con $\alpha = 5\%$.

Otra alternativa para contrastar hipótesis de un grupo medido dos veces es la siguiente:

$$t_C = \frac{\Sigma X_1 - \Sigma X_2}{\sqrt{\Sigma D^2 - \frac{(\Sigma D)^2}{n} / [n(n-1)]}}$$

Ejemplo Un odontólogo desea demostrar que un nuevo aditivo químico (*fc*), presente en un dentífrico por salir al mercado, reduce la caries en los niños. Para reducir el error de muestreo decide probar dicho producto en una muestra formada por gemelos (10 pares de niños); en cada pareja un gemelo usará el dentífrico con aditivo; el otro, la misma crema dental, pero sin aditivo, X_2 y X_1 respectivamente.

La muestra aleatoria de los gemelos se extrae de una población determinada y al azar se forman los dos grupos (control y experimental). La población la constituyen niños cuyas edades fluctúan entre los ocho y 10 años de edad; la *muestra* la conforman 10 parejas de gemelos escogidos al azar de dicha población. La variable de interés es el número de nuevas caries en el periodo de un año.

Gemelos	X_1	X_2
A	4	4
B	0	1
C	2	0
D	5	3
E	3	4
F	4	2
G	5	1
H	1	1
I	4	0
J	3	1

$n = 10$

Paso 1. Se establecen las hipótesis: puesto que la hipótesis del investigador considera que el *fc* disminuye la caries, las hipótesis quedan establecidas de la siguiente forma:

$$H_0 : \mu_1 = \mu_2$$

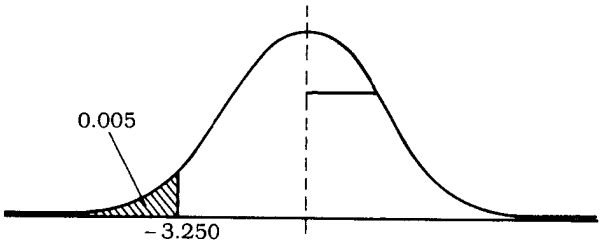
y

$$H_1 : \mu_1 < \mu_2$$

Paso 2. El nivel de confianza será el siguiente:

$\alpha = 0.005$
$gl = n - 1 = 10 - 1 = 9$
por lo que
$t_{0.005,9} = 3.250$

Paso 3. Se ubica la región de rechazo para $\alpha = 5\%$ (dos colas)



por lo que t debe cumplir con el siguiente intervalo:

$$|t| \geq 3.250$$

Paso 4. El modelo estadístico que se aplica es el siguiente:

$$t_C = \frac{\Sigma X_2 - \Sigma X_1}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{n}}{n(n-1)}}}$$

Paso 5. Se calculará: ΣX_1, ΣX_2, ΣD y ΣD^2, donde $D = X_1 - X_2$.

Gemelos	X_1	X_2	D	D^2
A	4	4	0	0
B	0	1	– 1	1
C	2	0	2	4
D	5	3	2	4
E	3	4	– 1	1
F	4	2	2	4
G	5	1	4	16
H	1	1	0	0
I	4	0	4	16
J	3	1	2	4
$n = 10$	$\Sigma X_1 = 31$	$\Sigma X_2 = 17$	$\Sigma D = 14$	$\Sigma D^2 = 50$

Sustituyendo los datos, se tiene

$$\boxed{t_C = -24.088}$$

Paso 6. Regla de decisión si $|t_C| \geq t_{(0.005,\ 9)}$; H_0 se rechaza;

como $|24.088| > 3.250$, entonces H_0 se rechaza; o también si $-24.088 < -3.260$ H_0 se rechaza.

Conclusión El aditivo (f_C) disminuye la caries.

10.2.2 Prueba de Sandler (*A*)

Es un método que puede sustituir a la prueba *t* de Student para muestras correlacionadas o dependientes.

Sandler ha derivado un valor estadístico (*A*) muy sencillo que se define como:

$$A = \frac{\Sigma D^2}{(\Sigma D)^2}$$

Para explicarla se utilizarán los datos del ejemplo anterior (el de los 10 pares de gemelos y la sustancia *fc*); así, al sustituir los datos se tiene:

$$\Sigma\, D^2 = 50$$

$$\Sigma\, (D)^2 = (14)^2$$

$$A_C = \frac{50}{(14)^2} = \frac{50}{196} = 0.2551$$

Este resultado se compara con un valor teórico que se presenta en la tabla del Apéndice. Se entrará en dicha tabla en $n - 1$ grados de libertad, y la regla de decisión es la siguiente: si el valor de A_C es igual o menor que el valor observado *A* (en la tabla), entonces A_C es significativa.

RD. Si $A_C \leq A$, entonces se rechaza H_0.

Como $0.0551 < 0.276$ ($\alpha = 0.025$, $A_{9,0.025} = 0.276$), por lo tanto se rechaza H_0.

Conclusión El dentífrico con *fc* disminuye la caries infantil.

10.3 INTERVALO DE CONFIANZA PARA DIFERENCIA ENTRE DOS PROPORCIONES

Es esencial recordar que, cuando se estudian proporciones, se hacen referencias intrínsecamente a la distribución binomial con parámetro *Px*. La construcción de intervalos de confianza para diferencias entre dos proporciones pertenece a este caso. Otra condición que no debe dejarse de señalar es que las proporciones deben ser independientes, es decir, las distribuciones binomiales serán independientes.

En algunas situaciones, el investigador requiere comparar la proporción de un grupo distinto con la de otro, en relación con otra característica en común. Por ejemplo, las autoridades de cierto país necesitan comparar los índices de divorcios que registran la zona sur y la norte del mismo. Se decide elaborar un intervalo de confianza para dicho estudio.

Sea

π_1 = Proporción (índice) de divorcios en la zona sur del país.
π_2 = Proporción (índice) de divorcios en la zona norte del país.

La diferencia de ambas proporciones es un buen instrumento para la comparación. Se considera un intervalo de confianza aproximado del $(1 - \alpha)$ 100% para $(p_1 - p_2)$, y la diferencia entre los dos parámetros binomiales (dicotómicos) la proporciona.

$$(p_{C_1} - p_{C_2}) - S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\; t_{\alpha/2,\, n_1 + n_2 - 2} \le (\pi_1 - \pi_2) \le (p_{C_1} - p_{C_2}) + S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\text{donde} \quad S = \sqrt{\frac{n_1\, p_{C_1}(1 - p_{C_1}) + n_2\, p_{C_2}(1 - p_{C_2})}{n_1 + n_2 - 2}}$$

$$\text{y } p_{C_1} = \frac{\text{número de casos favorables al evento 1}}{\text{número de casos de la muestra 1}};\; p_{C_1} = \frac{X_1}{n_1}$$

$$p_{C_2} = \frac{\text{número de casos favorables al evento 2}}{\text{número de casos de la muestra 2}};\; p_{C_2} = \frac{X_2}{n_2}$$

Ejemplo Dos grupos de pacientes (A y B) sufren cefaleas crónicas; constan de 150 y 100 individuos, respectivamente. Al primer grupo se le suministra una nueva clase de píldora contra el dolor, mientras que el segundo recibe píldoras ordinarias. Se alivian de inmediato 60 miembros del grupo A y 30 del grupo B. Constrúyase un intervalo de confianza de 95%.

Solución:

como $X_1 = 60$ y $n_1 = 150$

$p_{C_1} = \hat{p} =$ Porcentaje de pacientes con alivio inmediato del grupo *A*.

$$p_{C_1} = \frac{X_1}{n_1} = \frac{60}{150} = 0.40; \ \hat{q}_1 = 1 - p_{C_1} = 1 - 0.40 = 0.60$$

como $X_2 = 40$ y $n_2 = 100$

$p_{C_2} = p_2 =$ Porcentaje de pacientes con alivio inmediato del grupo *B*.

$$p_{C_2} = \frac{X_2}{n_2} = \frac{30}{100} = 0.30 \ ; \ \hat{q}_2 = 1 - p_{C_2} \ 1 - 0.30 = 0.70$$

IC (95%); $1 - \alpha = 0.95$; $\alpha = 5\%$, $\frac{\alpha}{2} = 0.025$ $\quad t_{\alpha/2 \ gl} = t_{0.025,248} = 1.96$

$$\text{Se calcula: } S = \sqrt{\frac{n_1 \, p_{C_1} (1 - p_{C_1}) + n_2 \, p_{C_2} (1 - p_{C_2})}{n_1 + n_2 - 2}}$$

$$S = \sqrt{\frac{150 \times 0.40 \times 0.60 + 100 \times 0.30 \times 0.70}{150 + 100 - 2}} = \sqrt{\frac{36 + 21}{248}} = 0.4794$$

$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{150} + \frac{1}{100}} = 0.1290$, se sustituyen estos resultados en el IC

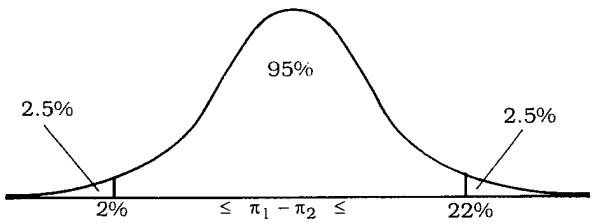
$$(\pi_1 - \pi_2) = (p_{C_1} - p_{C_2}) \pm S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \ t_{\alpha/2, \, gl}$$

$(\pi_1 - \pi_2) = (0.40 - 0.30) \geq 0.4794 \times 0.1290 \times 1.96$

$(\pi_1 - \pi_2) = 0.10 \pm 0.1212 \qquad (0.02, 0.22)$

o también $\quad 0.02 \leq (\pi_1 - \pi_2) \leq 0.22 \quad$ al 95% $\quad 2\% \leq \pi_1 - \pi_2 \leq 22\%$, con $e = 0.12$

La diferencia entre las dos proporciones está dada por el intervalo anterior con una confiabilidad de 95%. Su representación gráfica aparece a continuación:



10.3.1 Contraste de hipótesis

Nuevamente se hace referencia al parámetro binomial p. El objetivo de realizar un contraste de hipótesis es encontrar la diferencia, si existe, entre dos proporciones poblacionales p_1 y p_2.

Por ejemplo, la administración de una cervecería desea saber si su producto es más aceptado en la región 1 o en la región 2. Para establecer esto, deben compararse dos proporciones. Supóngase que:

p_1 = Proporción que consume cerveza en la región 1.
p_2 = Proporción de consumidores de cerveza en la región 2.

Una vez asignadas las proporciones en estudio, se recurre a las siguientes preguntas:

- ¿El consumo de cerveza es igual tanto en la región 1 como en la 2?
- ¿El consumo de cerveza es mayor en la región 1 que en la 2?
- ¿El consumo de cerveza es menor en la región 1 que en la 2?

Para responder alguna de las preguntas anteriores es necesario realizar un contraste, con las siguientes hipótesis estadísticas:

Pruebas de hipótesis para la igualdad de dos proporciones si $\sigma^2_{X_1} = \sigma^2_{X_2} = \pi = (1-\pi) = ?$ y si X_1 y X_2 **son independientes** $(\mu_X = \pi)$

Si $p_{C1} < p_{C2}$	*Estadística de prueba*	*Regla de decisión*
$H_0 : \mu_{X_1} = \mu_{X_2}$	$t_C = \dfrac{p_{C_1} - p_{C_2}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	Si $\lvert t_C \rvert \geq t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_2$
$H_1 : \mu_{X_1} < \mu_{X_2}$		Si $\lvert t_C \rvert < t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_2$
Si $p_{C1} > p_{C2}$		
$H_0 : \mu_{X_1} = \mu_{X_2}$	$t_C = \dfrac{p_{C_1} - p_{C_2}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	Si $\lvert t_C \rvert \geq t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_2$
$H_1 : \mu_{X_1} > \mu_{X_2}$		Si $\lvert t_C \rvert < t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_2$
Eficiencia		
$p_{C1} < p_{C2}$		
$H_0 : \mu_{X_1} = \mu_{X_2}$	$t_C = \dfrac{p_{C_1} - p_{C_2}}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	Si $\lvert t_C \rvert \geq t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ Rechazar $H_0 : \mu_{X_1} = \mu_2$
$H_1 : \mu_{X_1} <> \mu_{X_2}$		Si $\lvert t_C \rvert < t_{\alpha, n_1 + n_2 - 2} \Rightarrow$ No rechazar $H_0 : \mu_{X_1} = \mu_2$

donde

$$S = \sqrt{\frac{n_1\, p_{C1}\,(1 - p_{C1}) + n_2\, p_{C2}\,(1 - p_{C2})}{n_1 + n_2 - 2}}$$

Ejemplo Se propone instalar un hospital en los límites de un estado o en su zona centro. De los 500 jefes de manzana en la zona centro, 240 aprueban la construcción en ésta; por otra parte, de los 200 jefes de manzana en la zona limítrofe, 120 favorecen que el hospital se edifique en su jurisdicción. El hospital se erigirá en la zona donde la mayor proporción de jefes de manzana respalden la construcción. ¿Estaría usted de acuerdo en que es mayor la proporción correspondiente para la zona centro?

Solución:

Paso 1. Identifique la variable de interés. El jefe de manzana tendrá que contestar si está a favor o no de que el hospital se construya en su jurisdicción. El problema es del tipo dicotómico; en consecuencia, se recurrirá al parámetro poblacional P.

Paso 2. Proponga las hipótesis estadísticas.

$$H_0 : \pi_1 = \pi_2$$
$$H_1 : \pi_1 > \pi_2$$

Paso 3. Establezca el valor del error que se quiere aceptar (nivel de significancia).

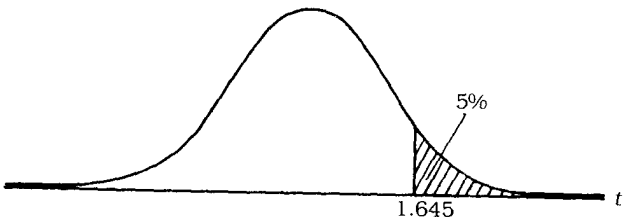
$$\alpha = 0.05$$

Paso 4. Determine el modelo estadístico adecuado. Como n_1 y n_2 son muestras grandes, se usará la aproximación a la normal, es decir, el estadístico es Z.

$$t_C = \frac{p_{C_1} - p_{C_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Paso 5. Obtenga mediante las tablas el valor crítico y ubique la zona de rechazo en una gráfica.

$\alpha = 5\%$
como $n_1 + n_2 - 2 \rightarrow \infty$
$t_{\infty,0.05} = Z_{0.05} = 1.645$



Paso 6. Especifique la regla de decisión para H_0.

Si $|t_C| > 1.645$, entonces H_0 se rechaza al 5%.

Paso 7. Calcule el valor del estadístico de prueba.

$$p_{C_1} = \hat{p}_1 = \frac{240}{500} = 0.48$$

$$S = \sqrt{\frac{n_1\, p_{C_1}(1 - p_{C_1}) + n_2\, p_{C_2}(1 - p_{C_2})}{n_1 + n_2 - 2}}$$

$$p_{C_2} = \frac{120}{200} = 0.60$$

$$1 - p_{C_1} = 0.52$$

$$1 - p_{C_2} = 0.40$$

$$\boxed{S = 0.49756}$$

Sustituyendo los datos, se tiene:

$$t_C = \frac{0.48 - 0.60}{0.49756 \times 0.08367} = -2.88; \quad |t_C| = 2.88$$

Paso 8. Como $2.88 > 1.645$ H_0 se rechaza.

Se puede concluir que la proporción de jefes de manzana que aprueban la construcción en la zona centro es mayor que su contraparte en la zona limítrofe. Por consiguiente, se construirá el hospital en la zona centro.

10.4 INTERVALO DE CONFIANZA Y CONTRASTE DE HIPÓTESIS PARA DOS VARIANZAS

Intervalo de confianza a un nivel (1 – α) para el cociente de dos varianzas si X_1 y X_2 son independientes

como
$$\frac{\dfrac{S_2^2}{\sigma_{X_2}^2}}{\dfrac{S_1^2}{\sigma_{X_1}^2}} \sim F_{n_{1-1}-n_{2-1}}$$

$$\Rightarrow 1-\alpha = P\left[F_{1-\alpha/2,\, n_{2-1},\, n_{1-1}}\right]$$

$$= P\left[\frac{S_1^2}{S_2^2} F_{1-\alpha/2,\, n_{2-1},\, n_{1-1}} \le \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} \le \frac{S_1^2}{S_2^2}\left[F_{1-\alpha/2,\, n_2-1,\, n_1-1}\right]\right.$$

∴ el intervalo de confianza deseado es

$$\frac{S_{C_1}^2}{S_{C_2}^2} F_{\alpha/2,\, n_{2-1},\, n_{1-1}} \le \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} \le \frac{S_{C_1}^2}{S_{C_2}^2}\left[F_{\alpha/2,\, n_{2-1},\, n_{1-1}}\right]$$

Ejemplo Construir un intervalo de confianza al 90% del cociente de dos varianzas.

$n_1 = 11$
$S_1^2 = 90\,000$
$S_2^2 = 40\,000$

$n_2 = 21$
$F_{0.05,\,10,\,20} = 2.35$
donde

$F_{0.05,\,20,\,10} = 2.774$

Sustituyendo los valores:

$$2.25 \times \frac{2.35}{1.94} \le \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} \le 2.25 \times 2.774$$

$$5.287 \le \frac{\sigma_{X_1}^2}{\sigma_{X_2}^2} \le 6.241$$

Pruebas de hipótesis para la igualdad de dos varianzas si X_1 y X_2 son independientes

Caso	*Estadística de prueba*	*Regla de decisión*
$S^2_{C1} < S^2_{C_2}$		
$H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$	$F_C = \dfrac{S^2_{C_1}}{S^2_{C_2}}$	Si $F_C \le F_{1-\alpha,\, gl} \Rightarrow$ Rechazar $H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$
$H_1 : \sigma^2_{X_1} < \sigma^2_{X_2}$		Si $F_C > F_{1-\alpha,\, gl} \Rightarrow$ No rechazar $H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$
$S^2_{C_1} > S^2_{C_2}$		
$H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$	$F_C = \dfrac{S^2_{C_1}}{S^2_{C_2}}$	Si $F_C \le F_{\alpha,\, gl} \Rightarrow$ Rechazar $H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$
$H_1 : \sigma^2_{X_1} < \sigma^2_{X_2}$		Si $F_C < F_{\alpha,\, gl} \Rightarrow$ No rechazar $H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$
Control		
$(S^2_{C_1} <> S^2_{C_2})$		
$H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$	$F_C = \dfrac{S^2_{C_1}}{S^2_{C_2}}$	Si $F_C \le F_{\alpha/2,\, gl} \Rightarrow$ Rechazar $H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$
$H_0 : \sigma^2_{X_1} \ne \sigma^2_{X_2}$		Si $F_C \ge F_{\alpha/2,\, gl}$
		Si $F_{1-\alpha/2,\, gl} < F_C < F_{1-\alpha,\, gl} \Rightarrow$ No rechazar $H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$

donde $gl = n_1 - 1,\ n_2 - 1$ y $S^2_{C_1}$, $S^2_{C_2}$, son las varianzas calculadas.

Si las muestras son grandes $n_1 \to \infty$ y $n_2 \to \infty$ y X_1 y X_2 son independientes, se puede utilizar la prueba Z

$$Z_C = \frac{S_1 - S_2}{S_p \sqrt{\dfrac{1}{2n_1} + \dfrac{1}{2n_1}}} \qquad S_p = \sqrt{\frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}}$$

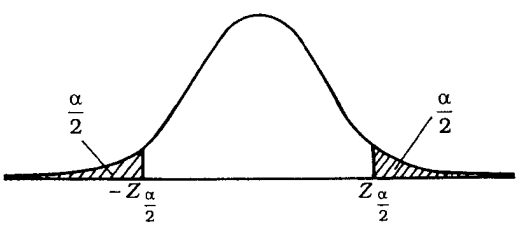
donde S_1 y S_2 son las desviaciones estándar de cada muestra.

$H_0 : \sigma^2_{X_1} = \sigma^2_{X_2}$

$H_1 : \sigma^2_{X_1} \ne \sigma^2_{X_2}$

La regla de decisión es:

a) Tanto como para una cola.
b) Como para dos colas.



Ejemplo Al poner a prueba cierto plástico, se obtuvieron los siguientes datos con respecto a dos muestras provenientes de fábricas distintas. Realice la prueba para homogeneidad de varianzas.

Fábrica	*Tamaño muestral*	*Estimación de la desviación estándar*
A	11	300
B	21	200

Paso 1. Identifique la variable de interés. Se desea saber si existe diferencia entre las varianzas de las dos muestras extraídas.

Paso 2. Proponga las hipótesis estadísticas.

$$H_0 : \sigma_A^2 = \sigma_B^2$$
$$H_1 : \sigma_A^2 > \sigma_B^2$$

Paso 3. Establezca el valor del error que se está dispuesto a aceptar (α).

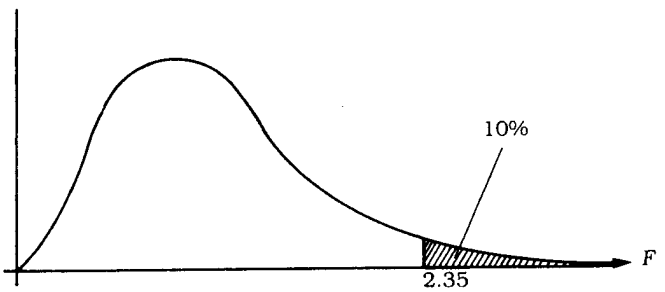
$$\alpha = 5\%$$

Paso 4. Determine el modelo estadístico adecuado.

$$F = \frac{S_1^2}{S_2^2} \text{ donde } S_1^2 > S_2^2$$

Paso 5. Obtenga, mediante la tabla correspondiente, el valor crítico y ubique la zona de rechazo en una gráfica.

$S_A^2 = 90\ 000$
$S_B^2 = 40\ 000$
$gl = 11 - 1 = 10$
$gl = 21 - 1 = 20$
$F_{0.05,\ 10,\ 20} = 2.35$



Paso 6. Especifique la regla de decisión para H_0.

$$\text{Si } F_C > 2.35 \Rightarrow H_0 \text{ se rechaza.}$$

Paso 7. Calcule el valor del estadístico de prueba mediante el modelo propuesto.

$$F_C = \left(\frac{300}{200}\right)^2 = 2.25$$

Paso 8. Obtenga las conclusiones.

Como $2.25 < 2.35 \Rightarrow H_0$ no se rechaza.
No es significativa la diferencia entre las varianzas.

Ejemplo Se quiere comparar la variabilidad de las emisiones de CO_2 que se presentan al utilizar dos tipos de gasolina (A y B). Se emplean dos automóviles en las mismas condiciones (de afinación, marca y modelo) y se obtienen los siguientes porcentajes de emisión:

	Porcentaje	*Porcentaje*	*Porcentaje*	*Porcentaje*	*Porcentaje*	*Porcentaje*
Auto A	93.12	93.57	92.81	94.32	93.77	93.52
Auto B	92.54	92.38	93.21	92.06	92.55	

Paso 1. Identifique la variable de interés. La emisión de CO_2 está en porcentaje y se busca saber si existe variabilidad entre los dos tipos de gasolina, expuestas en las mismas condiciones.

Paso 2. Proponga las hipótesis.

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Paso 3. Establezca el valor del error que se está dispuesto a aceptar (α):

$$\alpha = 10\%, \ \alpha = 0.1$$

Las tablas de los valores críticos de la prueba F son de una cola superior para $\alpha = 5\%$ o $\alpha = 1\%$. Por consiguiente, la hipótesis alterna para dicha prueba únicamente puede considerarse para una cola y $\alpha = 5\%$ o $\alpha = 1\%$; o es una hipótesis alterna para dos colas y $\alpha = 10\%$ o $\alpha = 2\%$.

$$\alpha/2 = 5\%$$

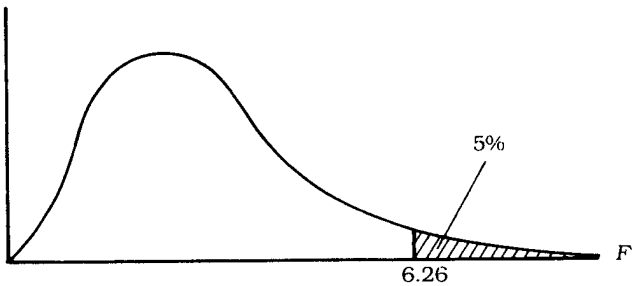
Paso 4. Determine el modelo estadístico.

$$F = \frac{S_1^2}{S_2^2} \text{ donde } S_1^2 > S_2^2$$

El grupo 1 es el que tiene la varianza mayor.

Paso 5. Obtenga mediante la tabla correspondiente el valor crítico:

$$F_{5,\,4,\,0.05} = 6.26$$



Paso 6. Establezca la regla de decisión para rechazar H_0:

$$\text{Si } F_C > 6.26 \Rightarrow H_0 \text{ se rechaza.}$$

Paso 7. Calcule el valor estadístico.

$$S_1 = 0.52327$$
$$S_1^2 = 0.27382$$
$$S_2 = 0.41973$$
$$S_2^2 = 0.17617$$

$$F = \frac{S_1^2}{S_2^2} = \frac{0.27382}{0.17617} = 1.55$$

Paso 8. Obtenga las conclusiones.

Como $1.55 < 6.26 \Rightarrow H_0$ no se rechaza. Luego, las varianzas son iguales y existe la misma variabilidad de emisión de CO_2 en ambas gasolinas.

10.5 CONTRASTE PARA MÁS DE DOS VARIANZAS

Uno de los supuestos más importantes para aplicar un análisis de varianza es la *homoscedasticidad*, o sea, la igualdad de las varianzas, también llamada *homogeneidad de varianzas*. Sin embargo, en algunas situaciones, por ejemplo, en control de calidad, la variabilidad de un producto manufacturado mediante procesos diferentes debe mantenerse igual.

Cuando las muestras independientes son del mismo tamaño, se utiliza la prueba de Cochran o la prueba de Hartley. Por otra parte, cuando las muestras independientes son de diferente tamaño, se emplea la prueba de Bartlett.

Una condición para aplicar las tres pruebas mencionadas anteriormente es que las poblaciones se distribuyan de manera normal.

10.5.1 La prueba de Cochran consiste en obtener la suma de las varianzas muestrales de las *k* muestras

$$\sum_{i=1}^{k} S_1^2$$

Por observación, se encuentra la varianza mayor $S^2_{\text{máx}}$ y se aplica.

$$R_{n,\,k} = \frac{S^2_{\text{máx}}}{\sum_{i=1}^{k} S_i^2}$$

La prueba de Hartley es parecida a la anterior: se observan las varianzas mayor y menor y se aplica la siguiente fórmula:

$$F_{\text{máx}} = \frac{S^2_{\text{máx}}}{S^2_{\text{mín}}}$$

Ejemplo Se desea saber si existe homoscedasticidad entre las varianzas de cinco muestras independientes del mismo tamaño (16 datos).

Paso 1. Se supone que las poblaciones a las cuales pertenecen estas muestras se distribuyen normalmente.

$$n = 16$$
$$k = 5$$

$$S_1^2 = 38.7,\ S_2^2 = 98.6,\ S_3^2 = 78.0,\ S_4^2 = 85.1,\ S_5^2 = 56.3$$

Paso 2.

$$H_0\text{: } \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$$

H_1: al menos una varianza es diferente.

Paso 3.

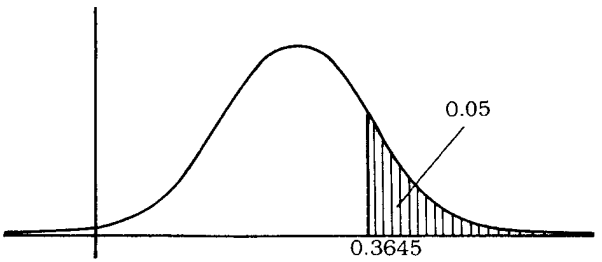
$$\alpha = 5\%$$

Paso 4. Prueba de Cochran (homogeneidad de varianzas).

$$R^*_{n,\,k} = \frac{S^2_{\text{máx}}}{\sum_{i=1}^{k} S_i^2}$$

Paso 5. Al consultar, en el Apéndice respectivo, la tabla de valor crítico, de Cochran, se obtiene:[2]

$$R_{16,\,5,\,0.05} = 0.3645$$



Paso 6.

Si $R_{16,\,5} \geq 0.3645$
H_0 se rechaza.

Paso 7.

$$R_{16,\,5} = \frac{98.6}{356.7} = 0.27642$$

$$S^2_{\text{máx}} = 98.6$$

$$\sum_{i=1}^{5} S_i^2 = 356.7$$

* En la tabla del Apéndice se usa *C* en lugar de *R*.

Paso 8. Como 0.27642 < 0.3645
H_0 no se rechaza.
En conclusión, las varianzas son iguales.
Utilizando los mismos datos de este problema, aplicar el modelo de Hartley.

10.5.2 Prueba de Hartley

Paso 1.

$$n = 16$$

$$k = 5$$

$$S_1^2 = 38.7,\ S_2^2 = 98.6,\ S_3^2 = 78.0,\ S_4^2 = 85.1,\ S_5^2 = 56.3$$

Paso 2.

$$H_0\text{: } \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$$

H_1: al menos una varianza es diferente.

Paso 3.

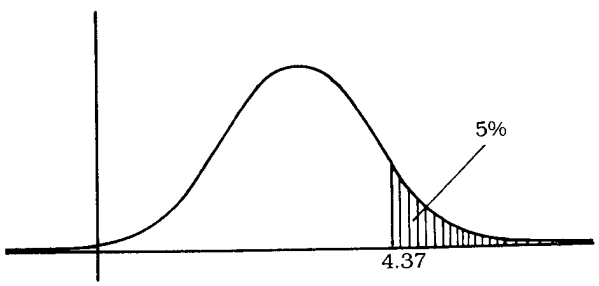
$$\alpha = 5\%$$

Paso 4. Prueba de Hartley

$$F_{máx} = \frac{S_{máx}^2}{S_{mín}^2}$$

Paso 5. Se consulta la tabla de valores críticos, de Hartley, en el Apéndice.

$$F_{16,\,5,\,0.05} = 4.37$$



Paso 6.

Si $F_{16,\,5} = 4.37$

H_0 se rechaza

Paso 7.

$$F_{16,\,5} = \frac{S^2_{máx}}{S^2_{mín}} = \frac{98.6}{38.7} = 2.5478$$

$$S^2_{máx} = 98.6$$

$$S^2_{mín} = 38.7$$

Paso 8. Como $2.5478 < 4.37$ H_0 no se rechaza.

En conclusión, las varianzas son iguales.

10.5.3 Prueba de Bartlett

Cuando las muestras no son del mismo tamaño y provienen de poblaciones normales, se aplica el siguiente modelo:

$$B = \frac{2.3026}{C} \left[(N-k)\log S^2_P - \sum_{i=1}^{k} (n_i - 1)\log S^2_i\right]$$

donde:

2.3026 es una constante

$$N = \sum_{i=1}^{k} n_i$$

log = logaritmos comunes (base 10)
n_i = el tamaño de cada muestra

S^2_p = varianza combinada; $S^2_p = \dfrac{\sum_{i=1}^{k} (n_i - 1)\, S^2_i}{N-k}$

k = número de muestras que intervienen

$$C = 1 + \frac{1}{3(k-1)}\left[\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{N-k}\right]$$

Si B está en la región de rechazo, es necesario calcular C, ya que siempre $C > 1$ y disminuirá el valor de B.

Para muestras grandes $n \to \infty$. B se aproxima a la distribución χ^2 con $k - 1$ grados de libertad.

De hecho, la región crítica de B se consulta en la tabla de la distribución χ^2.

Ejemplo De cuatro muestras independientes, que se suponen distribuidas de manera normal, se obtienen las siguientes varianzas muestrales y sus tamaños de muestra:

$S_1^2 = 103$ $\quad n_1 = 10$

$S_2^2 = 86$ $\quad n_2 = 12$

$k = 4$

$S_3^2 = 416$ $\quad n_3 = 8$

$S_4^2 = 212$ $\quad n_4 = 9$

Calcule el estadístico de Bartlett.

Paso 1. Se supone que las muestras se distribuyen normalmente.

Paso 2. H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

H_1: al menos una varianza es diferente.

Paso 3.

$\alpha = 5\%$

Paso 4. Prueba de Bartlett.

Paso 5. Al consultar la tabla de valores críticos de χ^2 para $\chi^2_{3,\,0.05}$

Se tiene $\chi^2_{3,\,0.05} = 7.815$.

Paso 6. Si $B \geq 7.815$

$N = \Sigma\, n_i = 10 + 12 + 8 + 9 = 39$

Paso 7.

$$S_p^2 = \frac{9\,(103) + 11\,(86) + 7\,(416) + 8\,(212)}{39 - 4} = 185.17143$$

$$\log\ S_p^2 = \log 185.17143 = 2.2676$$

$$(N - K)\log S_p^2 = 35\,(2.2676) = 79.3730$$

$$(n_1 - 1)\log S_1^2 = 9\,(2.0128) = 18.1152$$

$$(n_2 - 1)\log S_2^2 = 11\,(1.9345) = 21.2795$$

$$(n_3 - 1)\log S_3^2 = 7\,(2.6191) = 18.3337$$

$$(n_4 - 1)\log S_4^2 = 8\,(2.3263) = 18.6104$$

$$\sum_{i=1}^{4} (n_i - 1)\log S_1^2 = 76.3388$$

$$\text{Se calcula } C = 1 + \frac{1}{3\,(3)}\left[\frac{1}{9} + \frac{1}{11} + \frac{1}{7} + \frac{1}{8} - \frac{1}{35}\right] = 1.04902$$

$$B = \frac{2.3026}{1.04902}\,[79.3730 - 76.3388] = 6.660$$

Sin dividir entre C
$B = 6.98655$, dividiendo entre $C = 1.04902$; $\boxed{B = 6.660}$

Paso 8. Como $6.660 < 7.815$ H_0 no se rechaza. En conclusión, las varianzas son iguales.

Ejemplos Para contraste de más de dos varianzas.

- Prueba de *Cochran* de homogeneidad de varianza.

Ejemplo En 4 grupos de 6 personas cada uno, se trata de comprobar si "el repaso" es una variable decisiva para retener una lista de palabras monosílabas de igual valor asociativo, de lo cual se obtuvieron los siguientes resultados.

Datos:

Factor = repaso	$S_1^2 = 1.30$
Sujetos ($n = 6$)	$S_2^2 = 2.50$
Grupos ($k = 4$)	$S_3^2 = 1.30$
	$S_4^2 = \underline{0.70}$
	$\Sigma\ 5.80$ o sea $\Sigma S_j^2 = 5.80$

Paso 1. Calcular la suma de las varianzas ΣS_j^2.

Paso 2. Establecer las hipótesis.

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$
H_1 = al menos una varianza es diferente.
$\alpha = 5\%$

Paso 3. Calcular $R_{n,\,k} = \dfrac{S^2_{\text{máx}}}{\Sigma\, S_j^2}$

$$R_{n,\,k} = \frac{2.50}{5.80} = 0.4310$$

Paso 4. Buscar en tablas $R_{n,\,k}$ para $\alpha = 0.05$.

$$R = 6,\ 4,\ 0.05 = 0.5598$$

Paso 5. Regla de decisión:

Si $R_{n,\,k}$ calculada es > que la $R_{n,\,k}$ de tablas entonces H_0 se rechaza.

$R_{n,\,k} = 0.4310 < 0.5598$ de tablas $\therefore$ H_0 no es rechazada.

- Prueba de *Hartley* para muestras independientes del mismo tamaño.

Factor = "repaso"
Sujetos ($n = 6$)
Grupos ($k = 4$)

$S_1^2 = 1.30$
$S_2^2 = 2.50$
$S_3^2 = 1.30$
$S_4^2 = 0.70$

Paso 1. Obtener las varianzas (S^2 de cada muestra).

Paso 2. Establecer hipótesis.

$H_0 = S_1^2 = S_2^2 = S_3^2 = S_4^2$
H_1 = al menos una varianza es diferente.
$\alpha = 5\%$

Paso 3. Calcular $F_H = \dfrac{S^2_{\text{máx}}}{S^2_{\text{mín}}}$

$$F_H = \frac{2.50}{0.70} = 3.57$$

Paso 4. Tablas.

$$F_{6,\,4,\,0.05} = 10.4$$

Paso 5. Regla de decisión.

Si $F_H > F_{6,\,4,\,0.005}$ entonces H_0 se rechaza.

Conclusión. Como $F_H = 3.57$, no es mayor que F_H tablas $\therefore$ H_0 no se rechaza.

- Prueba de *Bartlett:* para grupos de diferente tamaño.

Ejemplo:

	S_i^2	n_i
S_1	1.30	$n_1 = 15$
S_2	2.50	$n_2 = 5$
S_3	1.30	$n_3 = 5$
S_4	0.70	$n_4 = 5$
		30

$k = 4$ grupos $\qquad n = 30$

$$\text{Fórmula } C = 1 + \frac{1}{3\,(k-1)}$$

Paso 1. Obtener varianza combinada.

$$S_P^2 = \frac{14\,(1.30) + 4\,(2.50) + 4\,(1.30) + 4\,(0.70)}{(26)}$$

$$S_P^2 = \frac{36.2}{26} = 1.392307692$$

Paso 2. Obtener $\log S_P^2 = 0.1437$

Paso 3. $(N - K) \log S^2p = 26 \times 0.1437 = \quad 3.7362$

$$(n_1 - 1) \log S^2_{1p} = 14 \times \log 1.30 = \quad 1.59$$
$$(n_2 - 1) \log S^2_{2p} = 4 \times \log 2.50 = \quad 1.59$$
$$(n_3 - 1) \log S^2_{3p} = 4 \times \log 1.30 = \quad 0.45$$
$$(n_4 - 1) \log S^2_{4p} = 4 \times \log 0.70 = \quad -0.61$$
$$(n_j - 1) \log S^2_{jp} = 3.02 \qquad \Sigma\, 3.02$$

Paso 4. Calcular $B = \dfrac{2.3026}{C}\ [(n - k) \log\ S^2_p - \sum_{i=1}^{k} (n_{i-1}) \log S^2_i\ ]$

$$B = 3.7362 - 3.02 = 0.7162 \qquad \boxed{B = 0.7162}$$

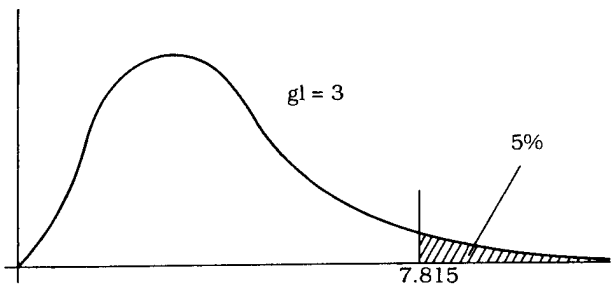
Paso 5. Calcular C.

$$C = 1 + \frac{1}{3\,(k-1)}\left(\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{N-k}\right) \quad \therefore \quad C = 1 + \frac{1}{3\,(3)}\left(\frac{1}{14} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{26}\right) =$$

$$C = 1 + 0.1111\ (0.7829) = 1 + 0.08699$$

ya que C siempre es mayor que 1

$$\boxed{C = 0.869} = 1.08699$$



Paso 6. Si $B \geq X^2_{3,0.05}$ entonces H_0 se rechaza como $0.7162 < 7.815$
H_0 se rechaza como $0.7162 < 7.815$ H_0 no se rechaza.

Conclusión: Las varianzas son homogéneas.

RESUMEN

Los conceptos, las definiciones y los requisitos de uso de los modelos estadísticos desarrollados en este capítulo, así como el contraste de hipótesis para dos o más varianzas, son la base para los temas posteriores, como análisis de varianza de uno y de dos factores, en el caso de estudios comparativos y para el capítulo de regresión lineal (estudios correlativos), la homogeneidad de las varianzas es de vital importancia.

PROBLEMAS

10.1 (Cuando se conoce σ_X^2.) Se tienen dos grupos independientes del mismo tamaño $n_1 = n_2 = 36$, donde $\overline{X} = 20$ y $\sigma_1 = 3$, $\overline{X}_2 = 18$, $\sigma_2 = 3$, por lo que $\sigma_X^2 = \sigma_1^2 = \sigma_2^2 = 9$.

a) Establezca un intervalo de confianza de 95% para la diferencia de las dos medias.

b) Contraste las hipótesis con un $\alpha = 5\%$ para la diferencia de las dos medias.

10.2 La sección maternal de una guardería decide reemplazar la leche materna por la de fórmula a los bebés de un mes de edad; en dicha guardería hay 50 niños y 100 niñas. Un mes después los pesos promedio son de 3 kg, con una desviación estándar de 0.2, 2.5 y de 0.5, kg, respectivamente.

a) Construya un intervalo de confianza de 99% de la diferencia de medias.

b) Contraste las hipótesis de la diferencia de medias a un $\alpha = 1\%$.

10.3 (Cuando no se conoce σ_X^2 y muestras grandes.) Se tienen dos grupos independientes del mismo tamaño $n_1 = n_2 = 36$, donde

$\overline{X}_1 = 5.4$ y $S_1 = 1.1$

$\overline{X}_2 = 5$ y $S_2 = 1.2$

$S_1^2 = 1.21$

$S_2^2 = 1.44$

a) Construya un intervalo de confianza de 95% de la diferencia de medias.

b) Contraste las hipótesis de la diferencia de dos medias.

10.4 (Muestras pequeñas; no se conoce σ_X^2.) Sean dos grupos independientes, donde $n_1 = 15$ y $n_2 = 3$, ambos pertenecen a poblaciones cuya distribución es normal, donde:

$\overline{X}_1 = 5.4$ $\quad$ $\overline{X}_2 = 5$

$S_1 = 1.1$ $\quad$ $S_2 = 1.2$

$S_1^2 = 1.21$ $\quad$ $S_2^2 = 1.44$

a) Construya un intervalo de confianza de 95% de la diferencia de medias.

b) Contraste un intervalo de confianza con un $\alpha = 5\%$ para la diferencia de dos medias.

10.5 Entre las políticas de una compañía de seguros para satisfacer al cliente, los ajustadores evalúan los daños en un automóvil que sufrió un siniestro y la aseguradora da opción a que los dueños pregunten presupuestos en dos talleres independientes, para comparar precios. Se hace un estudio con 15 asegurados; los datos se muestran a continuación en miles de pesos (7 600 = 7.6 miles de pesos).

Automóvil	*Taller 1*	*Taller 2*
1	7.6	7.3
2	10.2	9.1
3	9.5	8.4
4	1.3	1.5
5	3.0	2.7
6	6.3	5.8
7	5.3	4.9
8	6.2	5.3
9	2.2	2.0
10	4.8	4.2
11	11.3	11.0
12	12.1	11.0
13	6.9	6.1
14	7.6	6.7
15	8.4	7.5

a) Construya un intervalo de confianza de 95%.
b) Contraste las hipótesis con un $\alpha = 5\%$ de diferencia de las dos medias.

10.6 Un odontólogo debe demostrar que un nuevo aditivo químico (*fc*), presente en una crema dental por salir al mercado, disminuye las caries en los niños. Para reducir el error de muestreo se decide que dicha crema dental sea probada en una muestra formada por gemelos (diez pares de gemelos); en cada una de las parejas uno de los gemelos usará el dentífrico con aditivo y el otro la misma crema dental, pero sin éste.

La muestra aleatoria de los gemelos se extrae de una población determinada y al azar se escogen los dos grupos (control y experimental). La población está formada por gemelos cuyas edades están comprendidas entre los ocho y diez años de edad y la muestra la constituyen diez parejas de gemelos escogidos al azar de dicha población. La variable de interés es el número de nuevas caries en un periodo de un año.

Gemelos	X_1	X_2
A	4	4
B	0	1
C	2	0
D	5	3
E	3	4
F	4	2
G	5	1
H	1	1
I	4	0
J	3	1

a) Aplique *t* para muestras dependientes.

b) Aplique la prueba de Sandler (*A*).

10.7 A los comerciantes de dos ciudades fronterizas se les pregunta si están a favor o en contra de la propuesta 187 (ley antiinmigrante), que se encuentra en estudio en la legislatura de California. Para determinar si los comerciantes de las dos ciudades difieren en términos del porcentaje que está a favor, se toma una muestra aleatoria en cada ciudad de tamaño 100; 30 de los comerciantes están a favor en una ciudad, en tanto que 20 lo están en la otra.

a) Establezca un intervalo de confianza al 95%.

b) Realice un contraste de hipótesis con un $\alpha = 5\%$.

10.8 Se quiere saber si un recubrimiento especial que se aplica a las superficies de las matrículas de automóvil las hará más resistentes a la oxidación. En una zona costera se recubren 20 000 matrículas con dicho material y a otras 20 000 no se les aplica nada. Un año después se toma una muestra de 400 placas, de las cuales 360 están en buen estado; y de una muestra aleatoria de 225 no recubiertas, 195 están en buen estado. Establezca la hipótesis correspondiente y concluya con un $\alpha = 5\%$.

10.9 Un investigador desea aplicar una prueba *t* de Student para la diferencia entre dos medias de dos poblaciones. Se selecciona al azar una muestra de la población (1) de tamaño 15 y una varianza de 22.5. Una muestra aleatoria de la población (2) consta de 20 observaciones y una varianza de 4.2; contraste la hipótesis de las varianzas con un $\alpha = 1\%$.

10.10 Un investigador selecciona de cuatro poblaciones normales e independientes, cuatro muestras aleatorias; calcula sus varianzas y obtiene los siguientes datos.

$S_1^2 = 103$	$n_1 = 10$
$S_2^2 = 86$	$n_2 = 12$
$S_3^2 = 416$	$n_3 = 8$
$S_4^2 = 212$	$n_4 = 9$

Compruebe la homoscedasticidad de las varianzas con un $\alpha = 5\%$, aplicando:

a) Prueba de Bartlett.

b) Prueba de Hartley.

c) Prueba de Cochran.

10.11 En referencia al problema 7 del capítulo 2, al comparar la medición *A* en los cuatro métodos, se obtienen los siguientes datos:

$S_1^2 = 3.85$	$n_1 = 8$
$S_2^2 = 10.10$	$n_2 = 8$
$S_3^2 = 4.71$	$n_3 = 8$
$S_4^2 = 4.07$	$n_4 = 8$

Calcular la homoscedasticidad de las varianzas, aplicando.

a) Cochran.

b) Hartley.

10.12 Se realiza un estudio comparativo "de autoconcepto en dos grupos de adultos, morfinómanos y no morfinómanos"; se aplica la escala Tennessee de autoconcepto, que consta de nueve escalas. El grupo control o de no morfinómanos formado por 24 sujetos y el grupo de adictos al clorhidrato de morfina, formado por 20 sujetos. Los datos se muestran a continuación en cada una de las escalas del Tennessee. Establezca las hipótesis correspondientes y contrástelas al 5 y al 1% (aplique una prueba *t*).

Núm. Sujeto	*Ident.* 1	Satisf. 2	Conduc. 3	Autoc. Físico A	Autoc. Éti-mor* B	Autoc. Pers. C	Autoc. fam. D	Autoc. Social E	Autoc. Crítica A.C.
1	123	121	135	69	81	83	72	74	33
2	148	140	128	86	79	82	83	86	44
3	116	102	109	64	69	69	57	68	35
4	113	117	121	71	70	74	71	65	34
5	126	119	104	72	72	69	72	64	31
6	119	117	116	68	73	76	68	67	35
7	115	121	125	67	71	74	70	79	39
8	117	116	127	65	73	74	68	80	33
9	138	135	130	84	76	83	77	83	30
10	133	115	134	77	78	78	67	82	29
11	132	108	130	77	71	79	70	73	29
12	127	110	128	67	76	78	72	72	24
13	118	110	120	61	73	69	72	73	27
14	114	95	111	65	64	65	64	62	30
15	122	115	119	61	78	73	69	75	29
16	113	110	113	60	67	67	70	72	41
17	133	115	126	66	75	76	78	79	30
18	115	113	110	64	68	71	67	68	31
19	125	116	130	66	76	77	74	78	31
20	105	107	115	66	71	59	62	69	33
21	124	130	136	82	77	80	73	80	28
22	108	112	110	59	66	65	71	70	39
23	114	118	116	69	72	75	67	66	37
24	136	138	134	85	78	80	81	85	42

Grupo de morfinómanos

Núm Sujeto	Ident. 1	Satisf. 2	Conduc. 3	Autoc. Físico A	Autoc. Éti-mor B	Autoc. Pers. C	Autoc. fam. D	Autoc. Social E	Autoc. Crítica A.C.
1	65	62	63	36	40	36	36	42	45
2	86	84	78	47	53	49	52	47	37
3	71	70	73	43	39	48	41	43	42
4	62	65	70	36	36	36	44	45	51
5	61	60	68	35	40	34	39	41	43
6	75	74	78	37	51	45	51	43	42
7	86	80	82	48	51	52	47	50	38
8	75	68	54	36	42	39	42	38	44
9	92	84	88	53	51	54	50	56	34
10	85	83	82	46	52	56	46	50	32
11	60	63	55	33	37	31	37	40	45
12	88	87	91	52	55	54	49	56	36
13	88	99	94	58	50	54	59	60	36
14	88	88	97	47	53	58	56	59	41
15	68	73	54	33	42	37	41	42	44
16	90	94	79	55	55	52	56	48	41
17	73	65	74	41	37	45	47	42	47
18	91	83	96	55	53	58	48	56	33
19	84	90	92	49	56	52	53	56	33
20	60	61	54	31	37	30	37	40	49

* Ético-moral

11 Análisis de varianza de un factor

PROPÓSITOS

El propósito central del presente capítulo consiste en que el lector explique la importancia del análisis de varianza de un factor.

Del mismo modo, el lector podrá:

1. Identificar que el modelo estadístico por aplicar cuando se trabaja con más de dos grupos de datos es el análisis de varianza.
2. Explicar la lógica del análisis de varianza de un factor y establecer las hipótesis adecuadas.
3. Identificar las características fundamentales de la alternativa paramétrica al calcular la prueba F (Snedecor-Fisher) y contrastarla con el valor de la tabla correspondiente.
4. Explicar el diseño completamente aleatorizado de un factor.
5. Reconocer la pertinencia de este diseño experimental con respecto a la investigación, experimento o estudio.
6. Repasar estadísticas abordadas en capítulos anteriores.
7. Desarrollar el procedimiento de cálculo correspondiente a un caso particular para dos tratamientos ($k = 2$) mediante los pasos respectivos: obtención de $F = t^2$ y comprobación de la misma.
8. Describir el caso para un mismo número de sujetos ($k > 2$). Asimismo, cómo obtener la prueba F (Snedecor-Fisher).
9. Enumerar las variables de una muestra representativa que amerita el uso del procedimiento para un mismo número de sujetos.

10. Explicar la homoscedasticidad.
11. Aplicar los seis pasos y las seis etapas en el proceso de cálculo para $k > 2$.
12. Enunciar las características de la prueba Tukey-Snedecor (grupos iguales), así como el propósito de la misma.
13. Desarrollar el procedimiento de Tukey-Snedecor, mediante los pasos indicadores en el texto.
14. Encontrar los pares de medias cuyas diferencias son significativas.
15. Desarrollar las etapas de cálculo para un número diferente de sujetos por tratamiento $k > 2$, así como su gráfico.
16. Describir las características de la prueba Tukey-Snedecor para grupos desiguales.
17. Desarrollar el procedimiento de cálculo de la prueba Tukey-Snedecor para grupos desiguales.
18. Explicar los supuestos del análisis de varianza.
19. Describir las características de la prueba Dunnett para comparar el grupo control con respecto a varios grupos experimentales.
20. Desarrollar el cálculo de la prueba Dunnett mediante los pasos señalados en el texto.
21. Aplicar e interpretar los resultados obtenidos mediante el paquete MACSTAT.

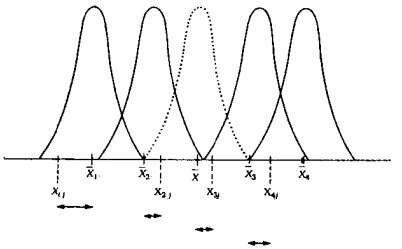
INTRODUCCIÓN

En ocasiones el investigador de ciencias del comportamiento debe comparar los datos que correspondan a más de dos grupos, por ejemplo, los de estudiantes en seis carreras diferentes. Efectuar todas las comparaciones posibles de carreras, tomadas de dos en dos (psicología frente a medicina, odontología frente a biología, etc.), y aplicar a cada combinación una prueba t de Student requerirá calcular 15 pruebas. Además de representar gran trabajo, incrementa la probabilidad de cometer el error de tipo I (rechazar la hipótesis nula, cuando debe ser aceptada). Al respecto, recuérdese que el investigador desea encontrar diferencias entre medias poblacionales que pueden ser estadísticamente significativas. Sin embargo, entre más pruebas estadísticas se realicen, será más probable obtener resultados estadísticamente significativos, lo que se debe al error de muestreo más que a una verdadera diferencia poblacional y, por ello, suele cometerse el error de tipo I.

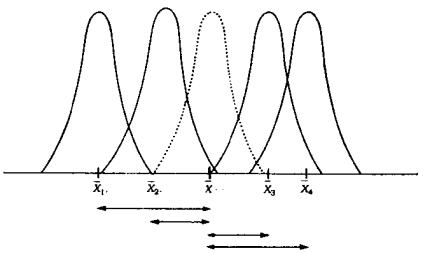
Para superar este problema y aclarar la interpretación del resultado, se necesita una prueba estadística que mantenga constante el error de tipo I al tomar una decisión global única acerca de la existencia de una diferencia estadísticamente significativa. Ésta es la prueba F de Snedecor-Fisher entre tres o más medias muestrales (también en el caso particular se puede aplicar si se quieren comparar dos grupos; así, en este caso, $F = t^2$).[1] Este método estadístico se conoce como *análisis de varianza*. Se refiere a la varianza total de un conjunto

[1] Este concepto implica que $t = \sqrt{F}$, sólo cuando las muestras son independientes; el ANOVA, debe ser el modelo completamente aleatorizado de efectos fijos de un factor.

de puntuaciones, que se divide en dos componentes: *i*) la distancia entre las puntaciones crudas y su media de grupo, conocida como *varianza dentro de los grupos* o *intragrupal*, y *ii*) la distancia entre las medias de cada grupo, también llamada *varianza entre grupos* o *intergrupal*.



Intragrupos (una de las observaciones cualquiera).



Intergrupos (una de las observaciones).

11.1 LA LÓGICA DEL ANÁLISIS DE VARIANZA

Al modelo estadístico adecuado para estimar la relación entre una o varias variables nominales independientes, con respecto a una variable continua dependiente, se le conoce como *análisis de varianza*.[2]

Las variables independientes se llaman factores; estos factores y sus niveles suelen definirse de manera experimental por el investigador. En cada factor hay al menos dos niveles o tratamientos que se medirán.

La hipótesis nula (H_0) que se desea contrastar en el diseño completamente aleatorizado de efectos fijos de un factor es $H_0 : \mu_1 = \mu_2 = \ldots \mu_K$, donde k es el número de tratamientos. En esta hipótesis los tratamientos no producen efectos (no existe diferencia estadísticamente significativa entre los grupos). El análisis de varianza presupone una relación lineal entre las variables experimentales y las observadas.

---

[2] En otras obras se les conoce como ANOVA, ANDEVA o ANVAR o A. de V.

En general, el objetivo de un experimento es investigar los efectos de las variables, tanto independientes (experimentales) como de las observadas (dependientes o de respuesta).

11.2 DISEÑO COMPLETAMENTE ALEATORIZADO

Uno de los modelos experimentales más sencillos es el diseño completamente aleatorizado de un factor (DCAlf) de efectos fijos, cuyo objetivo consiste en comparar varios niveles o tratamientos de un solo factor, que pueden ser cuantitativos o cualitativos, fijos, aleatorios o mixtos.

En este diseño se tiene un total de $kn = N$ "sujetos" asignados aleatoriamente a cada uno de los k tratamientos, de tal manera que a cada tratamiento le correspondan n sujetos; cuando el número de sujetos varía en los niveles, se dice que es *desbalanceado*. La asignación aleatoria de los sujetos a los tratamientos (grupos o muestras) es un recurso para asegurar que tanto las observaciones o mediciones como las k muestras son aleatorias e independientes, y eso sin que por necesidad tengan el mismo tamaño, aunque facilitaría el proceso si así fuera.

En resumen, el azar determina el tratamiento que se asigna a una unidad experimental (ue), es decir, a un animal de laboratorio, a un sujeto, entre otros.

Para realizar la asignación se utiliza una tabla de números aleatorios, o bien se asigna la misma condición experimental a las unidades de los diferentes tratamientos (camadas, proceso de manufactura, dosis de un fármaco, por ejemplo). La aleatorización obligada deberá ser la asignación aleatoria de las unidades experimentales a cada uno de los grupos.

Una ventaja de seleccionar un diseño experimental apropiado y sensible radica en reducir, en la medida de lo posible, el error experimental para mejorar la potencia de una prueba de significancia determinada. Por error experimental se entenderá el error de varianza utilizado para comprobar (contrastar) la significancia de las diferencias debidas a la fuente de variación involucradas en el análisis de varianza.

11.2.1 Para dos tratamientos ($k = 2$)

El análisis de varianza del diseño que se consideró en el tema anterior tiene una aplicación particular, la de dos grupos o tratamientos independientes. En este caso se puede aplicar una prueba t de Student (véase el capítulo 10) o la prueba de Snedecor-Fisher, y entonces $F = t^2$.

Para generalizar la notación que se utiliza en uno, dos o más grupos, es necesario adicionar subíndices a cada una de las mediciones u observaciones X. El primer subíndice i indica en cuál de los grupos o tratamientos se encuentra una observación o medición determinada, y el segundo subíndice j distingue a la observación de un grupo particular.

Ejemplo Se diseña un dispositivo para reducir la emisión de óxido nítrico en automóviles. Se seleccionan 10 autos al azar, según el principio de que están en condiciones óptimas y verificados, y se les equipa con este nuevo dispositivo. En paralelo, se escogen otros 10 automóviles de acuerdo con las mismas condiciones y se les equipa con el dispositivo convencional.

Se monitorea a los 20 automóviles por 30 minutos para determinar la cantidad de óxido nítrico que emiten. Se obtienen los siguientes resultados:

j Núm. de observación	X_{1j} Dispositivo A	X_{2j} Dispositivo B
1	1.01	1.35
2	0.98	1.16
3	0.95	1.23
4	1.02	1.20
5	1.05	1.32
6	0.96	1.28
7	0.99	1.21
8	0.98	1.25
9	1.01	1.17
10	1.02	1.19

En el ejemplo hay dos tratamientos (i = 2) y 10 observaciones para cada muestra

$n = 10$
$i = 1, 2$
$j = 1, 2, \ldots, 10$

Se procederá a calcular el análisis de varianza.

Paso 1. La variable factor unidireccional de interés es la emisión de óxido nítrico por los dos dispositivos (niveles o tratamientos).

Paso 2. Se establecen las hipótesis:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Paso 3. Se escoge el nivel de significancia

$$\alpha = 5\%$$

Paso 4. Realizar el análisis de varianza completamente aleatorizado de efectos fijos de un factor. Calcular el valor de F, realizando las siguientes 6 etapas.

Etapa 1

	Dispositivo A $i = 1$		Dispositivo B $i = 2$		
	$X_{1,1} = 1.01$	$X_{1,1}^2 = 1.02010$	$X_{2,1} = 1.35$	$X_{2,1}^2 = 1.82250$	
	$X_{1,2} = 0.98$	$X_{1,2}^2 = 0.96040$	$X_{2,2} = 1.16$	$X_{2,2}^2 = 1.34560$	
	$X_{1,3} = 0.95$	$X_{1,3}^2 = 0.90250$	$X_{2,3} = 1.23$	$X_{2,3}^2 = 1.51290$	
	$X_{1,4} = 1.02$	$X_{1,4}^2 = 1.04040$	$X_{2,4} = 1.20$	$X_{2,4}^2 = 1.44000$	
	$X_{1,5} = 1.05$	$X_{1,5}^2 = 1.10250$	$X_{2,5} = 1.32$	$X_{2,5}^2 = 1.74240$	
	$X_{1,6} = 0.96$	$X_{1,6}^2 = 0.92160$	$X_{2,6} = 1.28$	$X_{2,6}^2 = 1.63840$	
	$X_{1,7} = 0.99$	$X_{1,7}^2 = 0.98010$	$X_{2,7} = 1.21$	$X_{2,7}^2 = 1.46410$	
	$X_{1,8} = 0.98$	$X_{1,8}^2 = 0.96040$	$X_{2,8} = 1.25$	$X_{2,8}^2 = 1.56250$	
	$X_{1,9} = 1.01$	$X_{1,9}^2 = 1.02010$	$X_{2,9} = 1.17$	$X_{2,9}^2 = 1.36890$	
	$X_{1,10} = 1.02$	$X_{1,10}^2 = 1.04040$	$X_{2,10} = 1.19$	$X_{2,10}^2 = 1.41610$	TOTAL
ΣX	9.9700		12.3600		22.3300
ΣX^2		9.9485		15.3134	25.2619
$(\Sigma X)^2$	99.4009		152.7696		252.1705

Se observa que $n_1 = n_2 = n = 10$, $k = 2$ y $N = n_1 + n_2 = 20$

Ahora se mostrará en forma explícita la notación que se usará en las etapas posteriores, considerando también la notación utilizada por algunos otros autores:

$$\sum_{j=1}^{10} X_{1j} = 9.9700 = X_{1\cdot} = T_{1\cdot}$$

$$\sum_{j=1}^{10} X_{2j} = 12.3600 = X_{2\cdot} = T_{2\cdot}$$

$$\sum_{i=1}^{2} \sum_{j=1}^{10} X_{ij} = X_{1\cdot} + X_{2\cdot} = 22.3300 = X_{\cdot\cdot} = T_{\cdot\cdot}$$

$$\sum_{j=1}^{10} X_{1j}^2 = 9.9485$$

$$\sum_{j=1}^{10} X_{2j}^2 = 15.3134$$

De las dos sumas anteriores se forma la siguiente expresión:

$$\sum_{i=1}^{2} \sum_{j=1}^{10} X_{ij}^2 = 9.9485 + 15.3134 = 25.2619$$

$$= (1.01)^2 + (0.98)^2 + \ldots + (1.19)^2 = 25.2619$$

Es primordial señalar que la doble suma anterior no se expresa en términos de punto y que se llama *suma de cuadrados total no corregida.*

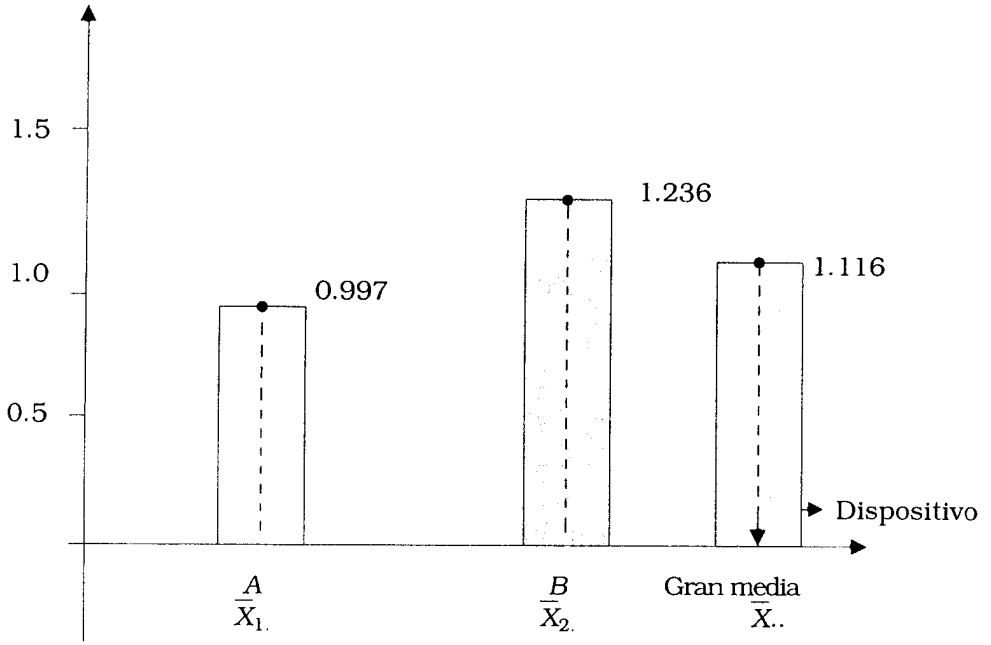
A continuación se calculan las medias aritméticas por cada grupo:

$$\overline{X}_{1\cdot} = \frac{X_{1\cdot}}{n} = \frac{9.970}{10} = 0.997$$

$$\overline{X}_{2\cdot} = \frac{X_{2\cdot}}{n} = \frac{12.3600}{10} = 1.236$$

$$\overline{X}_{\cdot\cdot} = \frac{X_{\cdot\cdot}}{N} = \frac{22.330}{20} = 1.1165$$

Posteriormente, deben graficarse las medias aritméticas obtenidas:



Etapa 2 Calcúlese la suma de cuadrados total corregida.

Al respecto, la SC_{tot} es

$$\sum_{i=1}^{2} \sum_{j=1}^{10} X_{ij}^2 = (1.01)^2 + (0.98)^2 + \ldots + (1.19)^2 = 25.2619$$

y la suma total es

$$\sum_{i=1}^{2} \sum_{j=1}^{10} X_{ij} = X_{..} = T_{..} = 22.330 = X_{1.} + X_{2.} = T_{1.} + T_{2.}$$

Por consiguiente, la SC_{tot} corregida es

$$SC_{tot} = \sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}^2 - \frac{(T_{..})^2}{N}$$

$(T_{..})^2 = (22.330)^2$

SC_{tot} (corregida) $= 25.2619 - \dfrac{(22.330)^2}{20}$

SC_{tot} (corregida) $= 0.330455$

Calcular la suma de cuadrados entre grupos SC_{eg} (también llamada *suma de cuadrados de tratamientos* SC_{trat}).

$$SC_{eg} = \frac{\sum_{i=1}^{k}\left(\sum_{j=1}^{n} X_{ij}\right)^2}{} - \frac{(T..)^2}{N}$$

Recuérdese que $\sum_{i=1}^{k}\left(\sum_{j=1}^{n} X_{ij}\right)^2 = \sum_{i=1}^{k} T_i^2$.

sustituyendo se tiene

$$SC_{eg} = \frac{(9.9700)^2 + (12.3600)^2}{10} - \frac{(22.330)^2}{20}$$

$$SC_{eg} = 25.21705 - 24.931445$$

$$SC_{eg} = 0.285605$$

Obténgase la suma de cuadrados dentro de los grupos SC_{dg} (también llamada SC_{error}), a partir de la siguiente igualdad

$$SC_{tot} = SC_{eg} + SC_{dg}$$
$$\Rightarrow SC_{dg} = SC_{tot} - SC_{eg}$$

como $SC_{tot} = 0.330455$ y $SC_{eg} = 0.285605$;

$$\begin{array}{r} 0.330455 \\ -0.285605 \\ \hline \Rightarrow SC_{dg} = 0.044850 \end{array}$$

Etapa 3 Obténgase los grados de libertad (gl), si
$k = 2$; $n = 10$; $N = kn \Rightarrow N = 2\ (10) \Rightarrow N = 20$

a) Totales(gl_{tot})
$gl_{tot} = N - 1 \Rightarrow gl_{tot} = 20 - 1 \Rightarrow gl_{tot} = 19$

b) Entregrupos(tratamientos)(gl_{eg})
$gl_{eg} = k - 1 \Rightarrow gl_{eg} = 2 - 1 \Rightarrow gl_{eg} = 1$

c) Dentro de grupos (error) (gl_{dg})
$gl_{dg} = N - k \Rightarrow gl_{eg} = 20 - 2 \Rightarrow gl_{eg} = 18$

Etapa 4 Calcúlese las medias cuadráticas o cuadrados medios (CM), también llamadas *varianzas estimadas.*

a) $CM_{eg} = CM_{trat} = \dfrac{SC_{eg}}{gl_{eg}}$

$$\Rightarrow CM_{eg} = \frac{0.285605}{1}$$

$$\Rightarrow CM_{eg} = 0.285605$$

b) Cuadrados medios dentro de grupos o cuadrado medio del error.

$$CM_{dg} = CM_{error} = \frac{SC_{dg}}{gl_{dg}}$$

$$\Rightarrow CM_{dg} = \frac{0.04485}{18}$$

$$\Rightarrow CM_{dg} = CM = 0.0024916$$

Etapa 5 Al obtener el valor de F hay que usar la fórmula $F_c = \frac{CM_{eg}}{CM_{dg}} = \frac{CM_{trat}}{CM_{error}}$.

$$\Rightarrow F_c = \frac{0.285605}{0.0024916}$$

$$\Rightarrow F_c = 114.64942$$

Etapa 6 Constrúyase la tabla del análisis de varianza.

Fuente de variación	*gl*	*SC*	*CM*	*F*
Entre grupos (tratamientos)	1	0.2856	0.2856	114.6249
Dentro de grupos (error)	18	0.04485	0.0024916	
Total	19	0.33045		

A continuación se muestra el resultado obtenido mediante el paquete MACSTAT:

TABLA DE ANÁLISIS DE VARIANZA (ANDEVA)				
Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón *F*
Tratamientos	1	0.2856	0.2856	114.6241
Error	18	0.0449	0.0025	
Total	19	0.3305		

Paso 5. Obtener la F (tablas).

$$F_{(1,\ 18,\ 0.05)} = 4.41$$

Paso 6. Si $F_c > 4.41 \Rightarrow H_0$ se rechaza.

Paso 7. Conclusiones: existe diferencia estadísticamente significativa entre el dispositivo A y el B.

Algunos autores consideran que

$$X_{1\cdot} = T_{1\cdot}$$

En nuestro caso es

$$\begin{aligned} T_{1\cdot} &= 9.970 \\ T_{2\cdot} &= 12.360 \\ X_{\cdot\cdot} = T_{\cdot\cdot} &= 22.330 \\ N &= 20 \end{aligned}$$

por lo que

$$\overline{X}_{\cdot\cdot} = \frac{T_{\cdot\cdot}}{N} \text{ es la llamada gran media; } \overline{X}_{\cdot\cdot} = 1.1165.$$

Para el ejemplo anterior, calcule la t de Student.

$$t = \frac{\overline{X}_{1\cdot} - \overline{X}_{2\cdot}}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

Se aplica este modelo, porque es el caso, cuando $\sigma_1^2 = \sigma_2^2 = \sigma^2$ pero *desconocidas.*

$$\text{donde } S_P = \sqrt{\frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2}}$$

Se sustituyen los valores en el estadístico y se obtiene

$$S_P = \sqrt{\frac{9 \times 0.00093 + 9 \times 0.00405}{18}}$$

$$S_P = \sqrt{\frac{0.00837 + 0.03645}{18}}$$

$$\Rightarrow S_P = 0.04990$$

$$t_c = \frac{0.997 - 1.236}{0.04990 \times \sqrt{\dfrac{1}{10} + \dfrac{1}{10}}} = -\frac{0.23900}{0.04990 \times 0.44721}$$

$$t_c = \frac{0.239}{0.02232} \Rightarrow |t| = 10.706 \text{ (significativo al 1\%)}$$

y se comprueba que

$$F_c = t_c^2 = 114.61^3$$

11.2.2 Caso A) Mismo número de sujetos por tratamiento ($k > 2$)

Se selecciona al azar una muestra representativa de una población, por ejemplo los estudiantes de una misma carrera universitaria. Se controla en lo posible que sean alumnos regulares, del mismo semestre y turno, así como con alto promedio de calificaciones. En primera instancia, dicha muestra puede considerarse homogénea.

La muestra se forma con 24 alumnos que cumplen los criterios de inclusión determinados anteriormente, de donde se constituyen cuatro grupos de seis alumnos cada uno, escogidos al azar. Se les asigna una tarea que consiste en resolver el mismo tipo de problemas de matemáticas.

Al grupo 1 se le advirtió que se aplicaba un examen más tarde (se le examinó una hora después). Al grupo 2 también se le informó que se aplicaba un examen más tarde (se le examinó un día después). Al grupo 3 no se le notificó sobre el examen (se le examinó una hora después). Al grupo 4 tampoco se le avisó acerca del examen (se le examinó un día después).

En este estudio se tienen cuatro grupos (cuatro tratamientos); la variable de interés es el resultado del mismo examen aplicado a cada grupo, que mide el aprendizaje según las condiciones establecidas previamente. La variable de interés (puntuaciones obtenidas en el examen) es la dependiente, cuya naturaleza es continua y numérica; la variable independiente es nominal, o sea categórica, y pertenece al grupo 1, 2, 3 o 4.

Las siguientes calificaciones se obtuvieron al aplicar el examen.

Grupo	Condición	Calificaciones
1	Aviso (examen en el corto plazo)	75, 93, 78, 71, 63, 76
2	Aviso (examen en el largo plazo)	78, 91, 97, 82, 85, 77
3	No aviso (examen en el corto plazo)	55, 66, 49, 64, 70, 68
4	No aviso (examen en el largo plazo)	64, 72, 68, 77, 56, 95

Nota: El paso cero consiste en aplicar una prueba para comprobar la homoscedasticidad.

[3] El paquete estadístico MACSTAT no calcula la t de Student, pero como se mencionó que $F = t^2$, entonces $t = \sqrt{F}$.

El objetivo de este estudio busca saber si la variabilidad de las puntuaciones obtenidas en el examen es significativa, al grado de considerar diferentes las cuatro formas de comunicación que afectan el resultado del examen.

Paso 1. La variable de interés es la calificación del examen.

Paso 2.

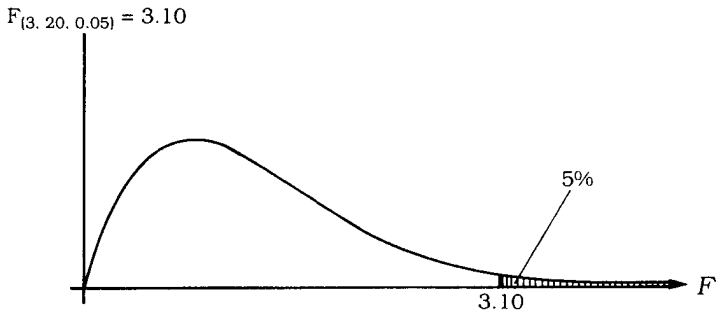
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_1 : \text{al menos una media es diferente o } H_1 : H_0 \text{ falsa.}$$

Paso 3. $\alpha = 5\%$.

Paso 4. Realice un análisis de varianza completamente aleatorizado de efectos fijos de un factor.

Calcule el valor de F para el mismo número de observaciones por grupo.

Paso 5.



Paso 6. Si $F_c > 3.10 \Rightarrow H_0$ se rechaza.

Paso 7. Para obtener el valor de F, se realizarán las siguientes etapas:

Etapa 1

a) Elabore una lista de las n observaciones X_{ij} en columna de cada tratamiento, 1, 2, 3,..., k

b) Eleve al cuadrado cada una de las observaciones X_{ij}, $i = 1, 2, \ldots, k, j = 1, 2, \ldots, n$.

c) Obtenga las sumas de las observaciones en cada tratamiento.

$$\sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij} = T..$$

Obténgase también el cuadrado de las sumas

$$(X_1.)^2, (X_2.)^2, (X_3.)^2, \ldots, (X_k)^2$$

y

$$\sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}^2 = \text{Suma de cuadrados de las puntuaciones.}$$

d) Calcule la $N_{total} = n_1 + n_2 + n_3 + \ldots + n_k = kn$.

Antes de continuar, es primordial conocer otra notación usada por algunos autores.

e) Calcule las medias aritméticas de cada grupo y grafíquelas.

	Grupo A		Grupo B		Grupo C		Grupo D		
	X_{1j}	X_{1j}^2	X_{2j}	X_{2j}^2	X_{3j}	X_{3j}^2	X_{4j}	X_{4j}^2	
	75	5 625	78	6 084	55	3 025	64	4 096	
	93	8 649	91	8 281	66	4 356	72	5 184	
	78	6 084	97	9 409	49	2 401	68	4 624	
	71	5 041	82	6 724	64	4 096	77	5 929	
	63	3 969	85	7 225	70	4 900	56	3 136	
	76	5 776	77	5 929	68	4 624	95	9 025	TOTAL
ΣX	456		510		372		432		1 770
ΣX^2		35 144		43 652		23 402		31 994	134 192
(ΣX^2)	207 936		260 100		138 384		186 624		793 044
$\bar{X}$	76		85		62		72		73.75
N	$n_1 = 6$		$n_2 = 6$		$n_3 = 6$		$n_4 = 6$		24

Sea k = Número de tratamientos, llamada también *grupos* $\Rightarrow k = 4$
n = Número de observaciones por tratamiento $\Rightarrow n = 6$
$N = kn$ $\Rightarrow N = 4\ (6)$
$\Rightarrow N = 24$

Se introduce ahora la notación (Σ) y quedan expresados de la siguiente manera.

Suma de los seis datos del tratamiento o grupo 1:

$$\sum_{j=1}^{6} X_{1j} = X_{1\cdot} = 456 \text{ para } i = 1$$

Para los siguientes tratamientos o grupos se expresan así:

$$\sum_{j=1}^{6} X_{2j} = X_{2\cdot} = 510 \quad \text{para } i = 2$$

$$\sum_{j=1}^{6} X_{3j} = X_{3\cdot} = 372 \quad \text{para } i = 3$$

$$\sum_{j=1}^{6} X_{4j} = X_{4\cdot} = 432 \quad \text{para } i = 4$$

La suma total de las puntuaciones se denota por:

$$\sum_{i=1}^{4} \sum_{j=1}^{6} X_{ij} = X_{\cdot\cdot} = 1770$$

Para la suma de los cuadrados se tiene lo siguiente:

$$\sum_{j=1}^{6} X_{1j}^2 = 35\ 144 \quad \text{para } i = 1$$

$$\sum_{j=1}^{6} X_{2j}^2 = 43\ 652 \quad \text{para } i = 2$$

$$\sum_{j=1}^{6} X_{3j}^2 = 23\ 402 \quad \text{para } i = 3$$

$$\sum_{j=1}^{6} X_{4j}^2 = 31\ 994 \quad \text{para } i = 4$$

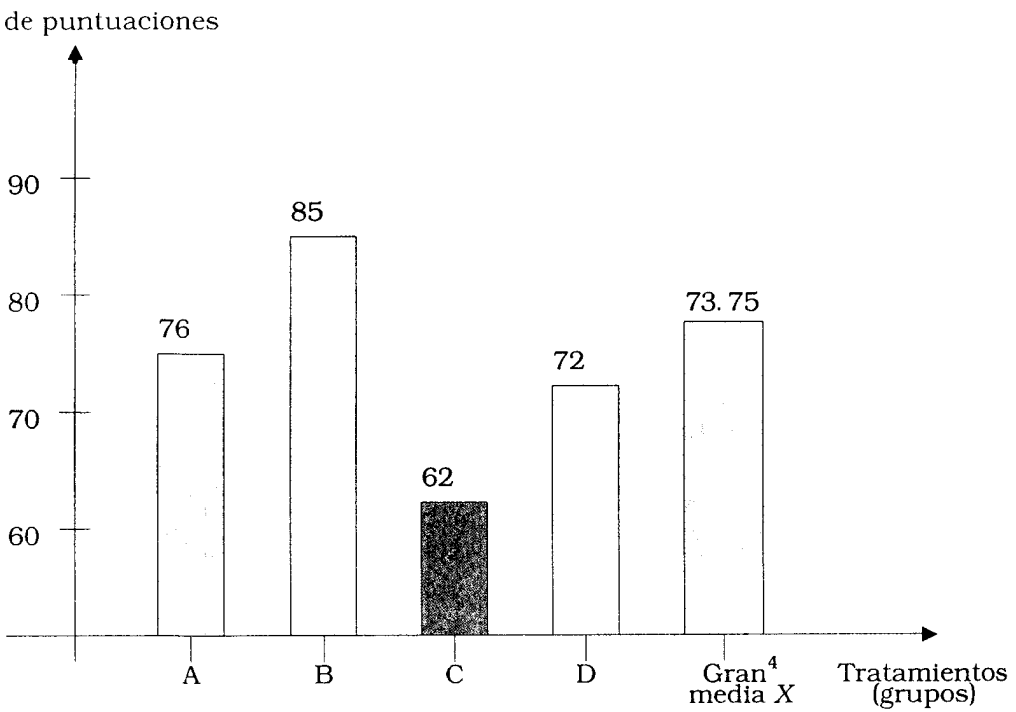
La suma total se denota por:

$$\sum_{i=1}^{4} \sum_{j=1}^{6} X_{ij}^2 = 134\ 192$$

Las sumas de las observaciones por tratamiento se elevan al cuadrado y los resultados se suman.

$$\sum_{i=1}^{4} X_{i\cdot}^2 = \sum_{i=1}^{4} \left(\sum_{j=1}^{6} X_{ij} \right)^2 = (456)^2 + (510)^2 + (372)^2 + (432)^2$$
$$= 207\ 936 + 260\ 100 + 138\ 384 + 186\ 624$$
$$= 793\ 044$$

Grafíquense las medias muestrales obtenidas.



Etapa 2 Calcule la suma de cuadrados total corregida.

$$SC_{tot} = \sum_{i=1}^{4} \sum_{j=1}^{6} X_{ij}^2 - \frac{T_{\cdot\cdot}^{\ 2}}{N}$$
$$= 134\ 192 - \frac{(1\ 770)^2}{24}$$
$$= 134\ 192 - 130\ 537.5$$

$$\boxed{SC_{tot} = 3\ 654.5}$$

[4] También llamada *media total*.

Calcule la suma de cuadrados entre grupos.

$$SC_{eg} = \frac{\sum_{i=1}^{4}\left(\sum_{j=1}^{6} X_{ij}\right)^2}{n} - \frac{(T..)^2}{N}$$

$$= \frac{(456)^2 + (510)^2 + (372)^2 + (432)^2}{6} - 130\,537.5$$

$$= 132\,174 - 130\,537.5$$

$$\boxed{SC_{eg} = 1\,636.5}$$

Obtenga la suma de cuadrados dentro de grupos.

$$SC_{dg} = SC_{tot} - SC_{eg} = 3\,654.5 - 1\,636.5$$

$$\boxed{SC_{dg} = 2\,018.0}$$

Etapa 3 Obtenga los grados de libertad (gl).

a) $gl_{tot} = N - 1 \Rightarrow gl_{tot} = 24 - 1 \Rightarrow gl_{tot} = 23$
b) $gl_{eg} = k - 1 \Rightarrow gl_{tot} = 4 - 1 \Rightarrow gl_{tot} = 3$
c) $gl_{dg} = N - k \Rightarrow gl_{tot} = 24 - 4 \Rightarrow gl_{tot} = 20$

Etapa 4 Calcule los cuadrados medios entre grupos.

$$CM_{eg} = \frac{SC_{eg}}{gl_{eg}} = \Rightarrow CM_{eg} = \frac{1\,636.5}{3}$$

$$\boxed{CM_{eg} = 545.5}$$

Cuadrados medios dentro de grupos.

$$CM_{dg} = \frac{SC_{dg}}{gl_{dg}} = \Rightarrow CM_{dg} = \frac{2\,018.0}{20} = 100.9$$

$$\boxed{CM_{dg} = 100.9}$$

Etapa 5 Obtener el valor de F.

$$F_c = \frac{CM_{eg}}{CM_{dg}} \Rightarrow F_c = \frac{545.5}{100.9}$$

$$\boxed{F_c = 5.40634}$$

Etapa 6 Construya el cuadro o tabla del análisis de varianza.

Fuente de variación	*gl*	*SC*	*CM*	*F*
Entre grupos (tratamientos)	3	1 636.5	545.5	5.40634
Dentro de grupos (error)	20	2 018.0	100.9	
Total	23	3 654.5		

Paso 8. Conclusiones: Como 5.40634 > 3.10 $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ se rechaza con $\alpha = 5\%$ (significativo al 5%)

y como $F_{(3,\ 20,\ 0.01)} = 4.94 \Rightarrow 5.40634 > 4.94$
$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ se rechaza también si $\alpha = 1\%$ (significativo al 1%).

11.2.2.1 Prueba de Tukey-Snedecor

Si la prueba F_c es mayor que la F_{tablas} (H_0 se rechaza), indica que los tratamientos experimentales asignados a los sujetos producen un efecto total significativo. Por consiguiente es necesario efectuar un análisis posterior para identificar el grupo o grupos particulares que inducen la diferencia estadísticamente significativa.

Existen varios métodos para analizar lo anterior, se les conoce como *comparaciones múltiples* o *comparaciones de medias de tratamiento*. Los principales son:

- Diferencia mínima significativa (LSD).
- Prueba nueva de rango múltiple de Duncan.
- Prueba de Student-Newman-Keuls (SNK).
- Procedimiento de Tukey modificado por Snedecor (1956).
- Prueba de Sheffé.

En este libro se utilizará el procedimiento de Tukey modificado por Snedecor (1956) (al respecto, véase Plutchik, 1983), que se desarrolla en la siguiente forma.

Paso 1. Se ubican las medias de los tratamientos, primero la de mayor valor y por último la de menor, así como las diferencias entre ellas.

Paso 2. Se calcula el error estándar de la media ($S_{\overline{X}}$), por medio de la CM_{dg} (CMe) y n el tamaño de grupo.

$$S_{\overline{X}} = \sqrt{\frac{CM_{dg}}{n}}$$

Paso 3. Determinar el valor Q en la tabla del apéndice mediante el número de tratamientos (k) y los grados de libertad dentro de grupos.

Paso 4. Se calcula D, utilizando

$$D = QS_{\overline{X}}$$

Paso 5. Se compara el valor D con las diferencias de los pares de medias de los tratamientos. La presencia de pares mayores que D significa que dichos tratamientos difieren significativamente del nivel α.

Para el ejemplo anterior, encontrar los pares de medias cuyas diferencias son significativas.

Paso 1.

Grupo	$\overline{X}$	$\overline{X} - 62$	$\overline{X} - 72$	$\overline{X} - 76$
B	85	23	13	9
A	76	14	4	
D	72	10		
C	62			

Paso 2. Como $CM_{dg} = 100.9$ y $n = 6$

$$S_{\overline{X}} = \sqrt{\frac{CM_{dg}}{n}} = \sqrt{\frac{100.9}{6}} = \sqrt{16.81667}$$

$$\boxed{S_{\overline{X}} = 4.100}$$

Paso 3. Como $k = 4$ y $gl_{dg} = 20$

$$Q_{(4,\ 20)} = 3.96$$

Paso 4. Al calcular

$$D_c = QS_{\overline{X}} = 3.96 \times 4.100$$

$$\boxed{D_c = 16.24}$$

Paso 5. Al comparar las diferencias de los pares de medias de la tabla del Apéndice (paso 1), se observa que sólo

$$\overline{X}_B - \overline{X}_C = 23$$

es mayor que 16.24.

Conclusión Sólo el grupo B con respecto al C tiene una diferencia estadísticamente significativa.

11.2.3 Caso B) Diferente número de sujetos por tratamiento (k > 2)

Este caso es desbalanceado $n_1 \neq n_2 \neq n_3$

Ejemplo

Tratamiento 1	4.79, 3.62, 3.42, 2.38, 2.15, 4.65, 3.33
Tratamiento 2	0.05, 0.57, 0.05, 0.10, 0.63
Tratamiento 3	3.61, 3.45, 2.17, 2.27

Etapa 1

	Tratamiento 1		Tratamiento 2		Tratamiento 3		
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	
	4.79	22.94	0.05	0.0025	3.61	13.0321	
	3.62	13.10	0.57	0.3249	3.45	11.9025	
	3.42	11.69	0.05	0.0025	2.17	4.7089	
	2.38	5.66	0.10	0.0100	2.27	5.1529	
	2.15	4.62	0.63	0.3969			
	4.65	21.62					
	3.33	11.08					TOTAL
ΣX	24.34		1.40		11.50		37.24
ΣX^2		90.71		0.7368		34.7964	126.243
$(\Sigma X)^2$	592.435		1.96		132.25		726.645
$\overline{X}$	3.47714		0.280		2.875		2.210
n	7		5		4		16

Nuevamente se considerará la notación punto.
Sea k = Número de tratamientos, llamado también *grupos*.
j = Número de observaciones por tratamiento.

Se introduce la notación (Σ), por lo que las sumas se expresan de la siguiente forma:

$$X_{1.} = \sum_{j=1}^{7} X_{1j} = 24.34 \quad \text{para } i = 1$$

$$X_{2.} = \sum_{j=1}^{5} X_{2j} = 1.40 \quad \text{para } i = 2$$

$$X_{3.} = \sum_{j=1}^{4} X_{3j} = 11.50 \quad \text{para } i = 3$$

La suma total es:

$$X_{..} = \sum_{i=1}^{3} \left(\sum_{j=1}^{n_i} X_{ij} \right) = 37.24$$

donde n_i = Número de observaciones por cada tratamiento, ya que éstos son diferentes.

En forma análoga para $\left[X_i^2 \right]$
o sea:

para $i = 1$

$$\sum_{j=1}^{7} X_{1j}^2 = 90.71$$

para $i = 2$

$$\sum_{j=1}^{5} X_{2j}^2 = 0.7368$$

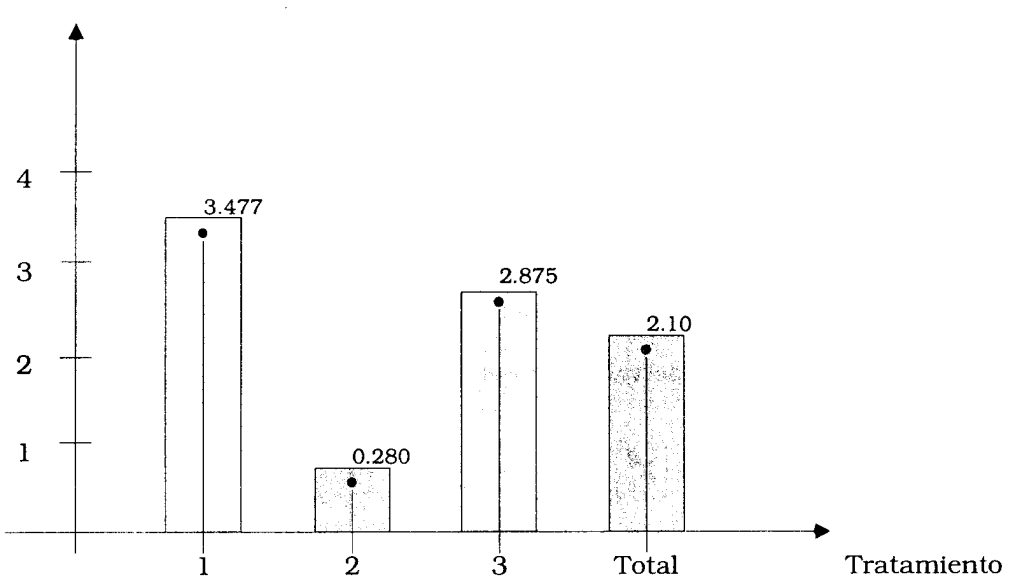
para $i = 3$

$$\sum_{j=1}^{4} X_{3j}^2 = 34.7964$$

La suma total queda expresada así

$$\sum_{i=1}^{3}\left(\sum_{j=1}^{n_i} X_{ij}^2\right) = 126.243$$

donde n_i = Número de observaciones por cada tratamiento, ya que éstos son diferentes.



Etapa 2 Calcule la suma de cuadrados total.

$$SC_{tot} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}^2 - \frac{(T..)^2}{N}$$

$$= 126.275 - \frac{(37.240)^2}{16}$$

$$= 126.275 - 86.676$$

$$\boxed{SC_{tot} = 39.599}$$

Calcule la suma de cuadrados entre grupos.

$$SC_{eg} = \sum_{i=1}^{3} \frac{(T_i)^2}{n_i} - \frac{(T..)^2}{N}$$

$$= \frac{(T_1.)^2}{n_1} + \frac{(T_2.)^2}{n_2} + \frac{(T_3.)^2}{n_3} - \frac{(T..)^2}{N}$$

$$= \frac{(24.340)^2}{7} + \frac{(1.400)^2}{5} + \frac{(11.500)^2}{4} - \frac{(37.240)^2}{16}$$

$$= 118.0875 - 86.676$$

$$\boxed{SC_{eg} = 31.4115}$$

Obtenga SC_{dg}

$$SC_{dg} = SC_{tot} - SC_{eg}$$

$$SC_{dg} = 39.599 - 31.4115$$

$$\boxed{SC_{dg} = 8.1875}$$

Etapa 3 Obtenga los grados de libertad.

a) $gl_{eg} = k - 1 = 3 - 1 = 2 \quad gl_{eg} = 2$
b) $gl_{dg} = N - k = 16 - 3 = 13 \quad gl_{dg} = 13$
c) $gl_{tot} = N - 1 = 16 - 1 = 15 \quad gl_{tot} = 15$

Etapa 4 Calcule los cuadrados medios.

$$CM_{eg} = \frac{SC_{eg}}{gl_{eg}} = \frac{31.4115}{2} = 15.705$$

$$\boxed{CM_{eg} = 15.705}$$

$$CM_{dg} = \frac{SC_{dg}}{gl_{dg}} = \frac{8.1875}{13} = 0.62981$$

$$\boxed{CM_{dg} = 0.62981}$$

Etapa 5 Obtenga F.

$$F_c = \frac{CM_{eg}}{CM_{dg}} = \frac{15.705}{0.62981}$$

$$\boxed{F_c = 24.936}$$

Etapa 6 Construya tabla de ANDEVA (análisis de varianza).

Fuente de variación	*gl*	*SC*	*CM*	*F*
Entre grupos (tratamiento)	2	31.4115	15.705	24.936*
Dentro de grupo (error)	13	8.1875	0.62981	
Total	15			

* Significativa a 1%.

debido a que $F_{(2,\ 13,\ 0.01)} = 6.70$

11.2.3.1 Prueba de Tukey-Snedecor (grupos desiguales)

Se recomienda una estrategia consistente en obtener un CM_{dg} promedio (al respecto, véase I. Méndez, 1976, pp. 31-32).

$$CM_{dg} = \frac{1}{k}\left(\frac{CM_{dg}}{n_1} + \frac{CM_{dg}}{n_2} + \ldots + \frac{CM_{dg}}{n_k}\right)$$

Ejemplo Calcular la prueba de Tukey-Snedecor.

Paso 1.

Tratamiento	$\bar{X}$	$\bar{X} - 0.28$	$\bar{X} - 2.875$
1	3.477	3.197	0.602
3	2.875	2.595	
2	0.280		

Paso 2. Como $CM_{dg} = 0.62981$

$n_1 = 7,\ n_2 = 5,\ n_3 = 4$

$$S_{\bar{X}} = \sqrt{\frac{1}{k}\left(\frac{CM_{dg}}{n_1} + \frac{CM_{dg}}{n_2} + \frac{CM_{dg}}{n_3}\right)}$$

$$S_{\bar{X}} = \sqrt{\frac{1}{3}\left(\frac{0.62981}{7} + \frac{0.62981}{5} + \frac{0.62981}{4}\right)}$$

$$S_{\bar{X}} = \sqrt{\frac{1}{3}(0.08997 + 0.12596 + 0.15745)}$$

$$S_{\bar{X}} = \sqrt{0.12446}$$

$$\boxed{S_{\bar{X}} = 0.35279}$$

Paso 3. Obtenga $Q_{(3,\ 13)} = 3.73$

Paso 4. Calcule $D = QS_{\bar{X}}$

$$D = 3.73 \times 0.35279$$

$$\boxed{D_c = 1.31591}$$

Paso 5. Como 3.197 > 1.31591, el tratamiento 1 presenta diferencia significativa con respecto al 2.

Como 2.595 > 1.31591, el tratamiento 3 manifiesta una diferencia significativa en relación con el 2.

Como 0.602 < 1.31591, entre el tratamiento 1 y el 3 no existe diferencia significativa.

11.2.4 Supuestos del análisis de varianza

1. Distribución normal

La no normalidad tiene un efecto pequeño cuando la inferencia se realiza con respecto a las medias.

2. Varianzas iguales (homoscedasticidad)

Cuando las varianzas no son iguales (no homogéneas), tienen un efecto pequeño en la inferencia de medias, si los tratamientos (grupos) son del mismo tamaño. Pero cuando los grupos son de diferente tamaño, los efectos de la no homoscedasticidad dan resultados severos.

3. Independencia de las observaciones

La no independencia de las observaciones tiene efectos severos en las inferencias acerca de las medias.

Una ventaja de realizar un experimento es que, como se controlan los fenómenos en estudio, pueden contrastarse (comparar) los resultados obtenidos. Así, el investigador selecciona las condiciones más favorables para la interpretación de dicho fenómeno y, al hacerlo, define el grado de generalidad de la población en estudio, así como la aplicabilidad de los resultados.

Para lograr lo anterior, es necesario que el diseño de la investigación garantice la comparabilidad de los tratamientos, al controlar las condiciones que producen problemas también llamadas *factores de confusión.*

11.3 PRUEBA DE DUNNETT (d)

Compara un grupo de control contra varios grupos experimentales, o varios tratamientos con respecto a un control.

Con frecuencia se requiere determinar si la respuesta de los tratamientos difiere de la correspondiente al control. Con ese propósito, se realiza un experimento en el cual las diferencias entre los grupos o tratamientos pasan a segundo término, ya que lo importante es saber si los mismos proporcionan una respuesta mayor o menor que la del control. En ocasiones, el único interés radica en establecer si existe una diferencia estadísticamente significativa entre cada uno de los grupos experimentales y el control, considerando sus respectivas medias aritméticas.

Esto es aplicable en un diseño completamente aleatorizado de efectos fijos de un factor, y su desarrollo consiste en lo siguiente:

Paso 1. Se obtienen los valores absolutos de las diferencias de las medias de cada uno de los tratamientos con el control $|\overline{X}_k - \overline{X}_c|$.

Paso 2. Se considera el número de tratamientos, incluido el control (k), los grados de libertad de CM_{error} —que se obtiene en la tabla del análisis de varianza— y los grados de libertad de CM_{error}; con esa información, se consulta la tabla del Apéndice para obtener el valor crítico t a un nivel de significancia α.

$t_{(gl,\ k,\ \alpha)}$, para una o dos colas, según sea el caso.

Paso 3. Se aplica la prueba de Dunnett (d).

$$d = t_{(gl,\ k,\ \alpha)} \sqrt{\frac{2\ CM_e}{n}}$$

donde n es el tamaño de los grupos (observaciones en los tratamientos).

Paso 4. Se aplica la siguiente regla de decisión:

Toda diferencia entre tratamiento y control, que sea mayor que d, es estadísticamente significativa.

Ejemplo Considérense los datos de la sección 11.2.1.

$$\overline{X}_A = 76,\ \overline{X}_B = 85,\ \overline{X}_C = 62,\ \overline{X}_d = 72$$

donde el grupo B es el de control.

Paso 1.

$$|\bar{X}_A - \bar{X}_B| = |76 - 85| = 9$$
$$|\bar{X}_B - \bar{X}_C| = |85 - 62| = 23$$
$$|\bar{X}_B - \bar{X}_D| = |85 - 72| = 13$$

Paso 2. Se obtiene $t_{(gl,\ k,\ \alpha)}$
donde los grados de libertad de CM_e son $gl = 20$

$$k = 4 \text{ grupos (incluido el de control)}$$

$$\alpha = 5\%$$

$$t_{(20,\ 4,\ 0.05)} = 2.19 \text{ (una cola)}$$

Paso 3. Se calcula d $(d = t_{(20,\ 4,\ 0.05)}) \sqrt{\dfrac{2\ CM_e}{n}}$

$$d = 2.19 \sqrt{\frac{2\ (100.9)}{6}} = 2.19\ (5.79943)$$

$$d = 12.70$$

Paso 4. Como $23 > 12.70$ y $13 > 12.70$

El grupo C y el D presentan una diferencia estadísticamente significativa para $\alpha = 5\%$.

Si $\alpha = 1\%$
$t_{(20,\ 4,\ 0.01)} = 2.97$
$d = 17.22$

El único grupo que tendría una diferencia estadísticamente significativa con el de control sería el C, para $\alpha = 1\%$.

Conclusión

$$H_0 : \mu_B = \mu_A = \mu_C = \mu_D$$

$$H_1 : \mu_B > \mu_A$$

$$\mu_B > \mu_C$$

$$\mu_B > \mu_D$$

El grupo B es mejor que el C y el D, o sea, el aviso y el examen de largo plazo constituyen la mejor estrategia académica.

RESUMEN

El análisis de varianza es un método para separar en componentes la suma de cuadrados total (la suma de los cuadrados de las desviaciones de las puntuaciones o mediciones con respecto a la gran media). Esta segmentación es simplemente un desarrollo algebraico, cuyo objetivo es apartar las fuentes de variación entre dichas puntuaciones individuales.

Uno de los propósitos de esta separación de las fuentes de variación, las cuales están definidas por el modelo matemático empleado, es el de contrastar la hipótesis acerca de que dos o más tratamientos son iguales. Dicha hipótesis es fundamental para una correcta decisión, debido a que las medias muestrales han sido extraídas de la misma población, o quizá representen observaciones de poblaciones con diferentes medias.

PROBLEMAS

11.1 Se realiza un experimento con 15 ratas, con las que se forman dos grupos al azar, el de control con ocho y el experimental con siete; al grupo experimental se le inyectó ARN de siete ratas que se sacrificaron después de haber sido entrenadas con éxito para desarrollar una conducta específica (la misma que se pretende obtener con las siete ratas del grupo experimental). Los ocho integrantes del grupo control, nunca han sido entrenados. Posteriormente se colocaron ambos grupos en una caja de Skinner y en cinco sesiones diarias de una hora se obtuvo el promedio de veces que mostraron la conducta esperada en cada uno de los sujetos.

Grupo experimental	*Grupo control*
1	0
10	1
7	1
9	0
8	0
10	1
3	2
	3

Realice un análisis de varianza con los resultados obtenidos.

11.2 Se lleva a cabo un estudio para comprobar la eficacia de tres métodos de enseñanza de un tema sobre ecología.

A_1 control de lectura
A_2 conferencia en el grupo
A_3 asesoría personalizada

Se seleccionan al azar 30 estudiantes de la misma escuela y el mismo grado, pero con promedio de ocho o más. Posteriormente se aplica un examen y se obtienen las siguientes puntuaciones:

A_1	A_2	A_3
4	5	3
7	6	5
9	3	1
6	8	4
9	3	4
6	2	5
5	5	7
7	6	3
7	7	5
10	5	3

Realice un análisis de varianza y concluya.

11.3 En una investigación realizada por un psicólogo acerca de la relación que existe entre la pulsión (ansiedad) de un individuo y su desempeño en la elección de puntos difíciles y fáciles en un laberinto, se encontró lo siguiente:

Se vendaban los ojos de los participantes y se midió el nivel de pulsión por medio de la administración de la tabla de ansiedad manifiesta de Taylor; se les conducía al inicio del laberinto y se pedía que lo recorrieran y lo aprendieran (con la mano), se les colocó en la mano un punzón y cuando un participante tomaba un camino equivocado se registraba el error.

El laberinto contenía 10 puntos de elección, cinco fáciles y cinco difíciles.

Se formaron dos grupos de participantes: de pulsión elevada y baja, y se registró el número de errores cometidos por ambos grupos.

PARTICIPANTES (elevada pulsión)		PARTICIPANTES (baja pulsión)	
Puntuación MAS*	Número de errores	Puntuación MAS*	Número de errores
28	21	7	20
35	18	7	18
29	13	8	14
29	12	10	12
36	11	6	10

* MAS significa la puntuación, resultante de aplicar la tabla de ansiedad manifiesta de Taylor.

Compare el grupo de elevada pulsión con el de baja pulsión en la puntuación MAS.

11.4 Realice un análisis de varianza con el número de errores, de los datos del problema 11.3.

11.5 Realice un ANOVA con los datos del problema 11.13 considerando los puntajes correspondientes a cero y negativos.

11.6 Se formaron tres grupos de 20 sujetos cada uno para llevar a cabo una investigación sobre aprendizaje programado; a cada grupo se le dio a estudiar un programa de matemáticas.

El primer grupo tomó el programa en forma normal y, por tanto, fue llamado *grupo de respuesta explícita* (RE).

El segundo grupo sólo "pensó" las respuestas sin escribirlas, *grupo de respuesta encubierta* (REn), y luego recibió información.

Un grupo control leyó solamente el programa con las respuestas ya en su lugar y tachó la porción de confirmación de respuesta, *grupo de no respuesta* (NR).

Se computó la cantidad de aprendizaje para cada estudiante, de donde se obtuvieron los siguientes resultados, en porcentajes:

Respuestas explícitas (RE)		Respuestas encubiertas (REn)		No respuesta (NR)	
0.46	0.75	0.26	0.42	0.15	0.83
0.36	0.87	0.57	0.81	0.78	0.45
0.56	0.24	0.82	0.40	0.65	0.53
0.28	0.29	0.94	0.78	0.64	0.55
0.72	0.50	0.85	0.89	0.33	0.05
0.80	0.41	0.57	0.43	0.65	0.62
0.59	0.87	0.00	0.32	0.58	0.57
0.60	0.48	0.90	0.56	0.50	0.58
0.91	0.78	0.62	0.58	0.56	0.48
0.93	0.86	0.70	0.71	0.38	0.68

Realice un análisis de varianza y concluya.

11.7 Considerando los datos del problema anterior, se aplicó la transformación de arco-seno para porcentajes en cada una de las puntuaciones; realice un análisis de varianza y concluya comparando con los resultados del problema anterior.

RE	REn	NR
42.71	30.66	45.57
36.87	49.02	62.03
48.45	64.90	53.73
31.95	75.82	53.13
58.05	67.21	35.06
63.43	49.02	53.73
50.18	0.00	49.60
50.77	71.57	45.00
72.54	51.94	48.45
74.66	56.79	38.06
60.00	40.40	65.65
68.87	64.16	42.13
29.33	39.23	46.72
32.58	62.03	47.87
45.0	70.63	12.92
39.82	40.98	51.94
68.87	34.45	49.02
43.85	48.45	49.60
62.03	49.60	43.85
68.03	57.42	55.55

11.8 Se realizó un experimento para medir la conducta verbal encubierta, utilizando los valores de electromiogramas. Los participantes eran estudiantes universitarios, a quienes se les pidió se ocuparan en cinco clases de actividades:

El grupo 1 escuchó una historia grabada.
El grupo 2 escuchó música clásica en forma similar.
El grupo 3 memorizó una parte de la misma historia presentada visualmente.
El grupo 4 leyó una sección de la historia.
El grupo 5 escuchó una cinta en blanco en la grabadora, con instrucciones de prestar atención en caso de que escucharan algo (que, por supuesto, nunca escucharon).

Antes de estas actividades todos los participantes descansaron durante un minuto, después de lo cual se ocuparon en su actividad durante cinco minutos.

Se tomaron los registros de los valores del electromiograma originados por la ligera actividad eléctrica de los músculos del mentón, en las distintas condiciones de investigación.

GRUPO				
1 (Escuchar una historia)	2 (Escuchar música)	3 (Memorizar)	4 (Leer)	5 (Nada)
1.7	- 1.2	12.4	7.1	- 2.2
- 0.6	2.0	2.4	- 2.3	0.6
28.9	0.1	- 0.8	4.3	- 0.1
1.1	0.5	0.0	0.0	0.6
0.0	- 0.3	0.0	0.6	3.1
- 0.5	3.7	2.2	2.4	2.1
2.1		0.2	6.1	4.9
			8.1	2

Realice un análisis de varianza y concluya.

11.9 Se prueban cuatro técnicas de iluminación para dibujantes que básicamente son el grado de inclinación del reflector, y se obtuvieron los siguientes datos que miden intensidad luminosa:

0°	25°	50°	75°
31	31	34	37
38	34	35	34
38	27	39	27
33	27	35	32
31	29	30	26

Cada técnica se realizó en cinco situaciones semejantes. Lleve a cabo un análisis de varianza.

11.10 Los datos de la siguiente tabla muestran el CI (Army-Beta) obtenido por 25 alumnos de primer ingreso provenientes de cinco sistemas de bachillerato diferentes. Realice un análisis de varianza y concluya.

ESCUELAS				
A	B	C	D	E
68	49	64	67	61
55	59	63	55	59
60	61	54	65	70
67	60	52	64	69
60	61	62	59	61

11.11 Supuestamente, la temperatura afecta la vida promedio de las baterías para radio; se lleva a cabo un estudio con 30 baterías de la misma marca

e igual fecha de caducidad; las temperaturas a las que se someten se muestran a continuación, con los siguientes resultados:

Vida promedio en horas				
0°	10°	20°	30°	40°
55	60	70	72	65
55	61	72	72	66
57	60	73	72	60
54	60	68	70	64
54	60	77	68	65
56	60	77	69	65

Realice un análisis de varianza.

11.12 Se empollan dos diferentes tipos de huevos de codorniz, cuyos pesos son, en gramos:

CAMADA 1	CAMADA 2
3.3	2.6
3.6	2.2
2.6	2.2
3.1	2.5
3.2	1.2
3.3	1.2
2.9	
3.4	

Realice un análisis de varianza y calcule la prueba t de Student, comprobando que $F = t^2$.

11.13 Se lleva a cabo un estudio donde participan estudiantes de nivel profesional a quienes se les detecta baja o alta susceptibilidad de ser hipnotizados. Se crean grupos de 16 estudiantes, cada uno con baja o alta susceptibilidad, de los cuales se forman al azar dos grupos de ocho integrantes, con lo que entonces se tienen cuatro grupos, dos de baja y dos de alta; a un grupo de cada categoría se les induce a dicho estado; posteriormente se les realiza una tarea y se cuenta tanto el número de errores, aciertos y no respuesta, con los siguientes resultados:

BAJA SUSCEPTIBILIDAD		ALTA SUSCEPTIBILIDAD	
Hipnotizado	No hipnotizado	Hipnotizado	No hipnotizado
0	+ 9	- 16	- 4
- 8	+ 1	0	+ 8
+1	- 5	- 20	- 10
- 20	- 14	- 41	+ 9
- 17	- 2	- 32	- 10
- 43	- 3	- 6	-23
- 4	+ 14	- 42	+ 29
- 23	+ 9	- 29	- 14

donde:
(+) son aciertos
(–) son errores
(0) no contestaron.

Realice un análisis de varianza.

Considere todas las puntuaciones (+, – y 0).

12 Análisis de varianza de dos factores

PROPÓSITOS

El propósito central del presente capítulo es que el lector explique los diseños aleatorizados y los aplique en una investigación.

De igual modo, el lector podrá:

1. Explicar el análisis de varianza con dos factores.
2. Identificar los estadísticos que conforman el ANOVA con dos factores.
3. Explicar el diseño de bloques aleatorizados.
4. Reconocer los casos en que se emplea el diseño de bloques aleatorizados, así como los criterios de uso.
5. Desarrollar el procedimiento de cálculo en el diseño de bloques aleatorizados.
6. Reconocer las limitaciones y ventajas en el uso de este diseño.
7. Explicar el diseño completamente aleatorizado de dos factores.
8. Desarrollar el procedimiento de cálculo con las etapas propuestas en el texto.
9. Enunciar el concepto de interacción de los dos factores (variables independientes) (con la variable dependiente), también llamada *variable de interés*.
10. Reconocer la existencia de interacción de dos factores y el significado de este concepto en la variable de interés.
11. Aplicar los procedimientos de cálculo en la solución de ejercicios relacionados con la eliminación de error en los tratamientos por bloques.
12. Aplicar e interpretar los resultados obtenidos mediante el uso del paquete estadístico MACSTAT, tanto para el diseño de bloques aleatorizados como para el diseño completamente aleatorizado.

INTRODUCCIÓN

Muchos estudios o investigaciones exigen mediciones repetidas en los sujetos, por lo que es necesario dividir en lotes los efectos de los sujetos, en adición a los efectos de los tratamientos, así como en el caso de que, aunque no se requieran medidas repetidas, suele considerarse más de una variable de tratamiento.

Hay dos modelos del análisis de varianza de dos factores:

1. Diseño de bloques aleatorizados.
2. Diseño completamente aleatorizado. Tanto éstos como el concepto de interacción se desarrollarán en este capítulo y constituyen la base para la comprensión y aplicación de modelos más complicados que se utilizan en el diseño de experimentos.

12.1 DISEÑO DE BLOQUES ALEATORIZADOS

En ciertos casos, las unidades experimentales de una muestra son heterogéneas, pero pueden clasificarse en grupos homogéneos llamados *bloques*. Aunque entre bloques se presentan diferencias, dentro de cada uno las unidades experimentales son homogéneas y se asignan al azar (esta característica es esencial para definir cada bloque). Así, cuando se desea compartir a métodos (tratamientos) distribuidos en b bloques, la precisión del experimento puede mejorarse considerablemente, ya que se constituyen bloques para controlar algunas fuentes de variación no asociadas con las unidades y no deseables por el investigador. Constituir bloques puede considerarse una extensión de formar pares con los tratamientos.

Este procedimiento es útil, por ejemplo, al analizar la producción de un artículo, el procesamiento de un reactivo o, en un plano más específico, el resultado de un experimento del que se sospecha hay una variación diaria, aun cuando las condiciones del laboratorio o proceso se mantienen constantes. Esta variación se eliminaría a partir de una estimación del error que considere como un bloque todas las unidades experimentales estudiadas el mismo día.

Carece de sentido proporcionar tratamiento a cada bloque, ya que las posibles diferencias serían entre éstos y no entre los tratamientos; por consiguiente, los bloques se tienen que distribuir entre los tratamientos de tal manera que cada tratamiento se aplique a cada bloque, y el valor promedio para cada tratamiento debe ser independiente de las diferencias entre los bloques.

Es necesario considerar:

- Un factor con dos o más niveles.
- Asignar a los sujetos a los bloques, de tal manera que la variabilidad entre los sujetos de cada bloque sea menor que la existente entre los bloques. El número de casos (unidades experimentales o mediciones) en cada bloque debe ser igual.

- La asignación aleatoria de los niveles de tratamiento a las unidades experimentales en cada bloque.

Si la variación entre las unidades experimentales dentro de los bloques es significativamente menor que la existente entre éstos, un diseño de bloques aleatorizados es más poderoso que uno completamente aleatorizado.[1]

Si todos los sujetos (unidades experimentales o mediciones) de la muestra son homogéneos, este diseño no es efectivo.

Ejemplo Supóngase que 16 sujetos (ue) se asignan al azar en cuatro bloques y con cuatro tratamientos.

Paso 1. Se establecen las hipótesis:

Para tratamientos

$H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4$

$H_1 : H_0$ falsa

Para bloques

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$

$H_1 : H_0$ falsa

Paso 2. Establecer el nivel de significancia.

$\alpha = 5\%$

Paso 3. Regla de decisión.

Si $F_{trat} \geq F_{tablas} \Rightarrow H_0$ se rechaza

$F_{bloq} \geq F_{tablas} \Rightarrow H_0$ se rechaza

		TRATAMIENTOS (a)				
		t_1	t_2	t_3	t_4	$T_{.j}$
BLOQUES (b)	b_1	9.8	12.1	15.8	11.7	49.4
	b_2	9.5	11.8	15.5	11.7	48.5
	b_3	11.8	15.5	17.2	13.6	58.1
	b_4	13.9	15.4	17.2	14.6	61.1
	$T_{i.}$	45.0	54.8	65.7	51.6	217.1 = $T_{..}$

Paso 1. Se obtiene el factor de corrección, que es:

$$\boxed{\frac{(T_{..})^2}{N} \quad \text{donde } N = ba}$$

[1] Roger E. Kirk, *Experimental Design: Procedures for the behavioral sciences, Wadsworth Publishing Company, EUA.* pp. 131, 132. 1968.

Como $b = 4$ y $a = 4$; $N = 16$

$$\frac{(T..)^2}{N} = \frac{(217.1)^2}{4 \times 4} = \frac{(217.1)^2}{16} = \frac{47\ 132.4}{16}$$

$$\Rightarrow \frac{(T..)^2}{N} = 2\ 945.77$$

Paso 2. Se elevan al cuadrado cada una de las observaciones X_{ij}, y se obtiene su suma total, que es:

$$\sum_i \sum_j X_{ij}^2$$

	t_1	t_2	t_3	t_4	$T_{.j}^2$
b_1	96.04	146.41	249.64	136.89	628.98
b_2	90.25	139.24	240.25	136.89	606.63
b_3	139.24	240.25	295.84	184.96	860.29
b_4	193.21	237.16	295.84	213.16	939.37
$T_{i.}^2$	518.74	763.06	1 081.57	671.90	3 035.27 = $T_{..}^2$

$$\therefore\ T_{..}^2 = \sum_i \sum_j X_{ij}^2 = 3\ 035.27$$

Paso 3. Se calculan las sumas de cuadrados (SC_{tot}).

$$SC_{tot} = \sum_i \sum_j X_{ij}^2 - \frac{(T..)^2}{n} \quad \text{o también } SC_{tot} = T_{..}^2 - \frac{(T..)^2}{N}$$

$$SC_{tot} = 3\ 035.27 - 2\ 945.77$$

$$\boxed{SC_{tot} = 89.50}$$

Por bloque (SC_b)

$$SC_b = \sum_{j=1} \frac{T_{.j}^2}{a} - \frac{(T..)^2}{N}$$

donde:

a = número de tratamientos

$T_{.j}$ = la suma de las observaciones por cada bloque [49.4 + 48.5 + 58.1 + 61.1]

$T_{.j}^2$ = la suma de las observaciones por cada bloque elevadas al cuadrado $[(49.4)^2 + \ldots (61.1)^2]$.

$$SC_b = \frac{(49.4)^2 + (48.5)^2 + (58.1)^2 + (61.1)^2}{4} - 2\,945.77$$

$$SC_b = 2\,975.3575 - 2\,945.77$$

$$\boxed{SC_b = 29.5875}$$

a) Se obtiene la de suma de cuadrados por tratamiento (SC_{trat})

$$SC_{trat} = \sum_{i=1} \frac{T_{i\cdot}^{\,2}}{b} - \frac{(T_{\cdot\cdot})^2}{N}$$

donde:

b = número de bloques
$T_{i\cdot}$ = suma de las observaciones por tratamiento
$T_{i\cdot}^{\,2}$ = suma de las observaciones por tratamiento elevadas al cuadrado.

$$SC_{trat} = \frac{(45.0)^2 + (54.8)^2 + (65.7)^2 + (51.6)^2}{4} - 2\,945.77$$

$$SC_{trat} = 3\,001.7725 - 2\,945.77$$

$$\boxed{SC_{trat} = 56.0025}$$

Paso 4. Se calcula la suma de cuadrados residual o de error (SC_e).

$$SC_e = SC_{tot} - (SC_b + SC_{trat})$$

$$SC_e = 89.50 - (29.5875 + 56.0025)$$

$$SC_e = 89.50 - 85.59$$

$$\boxed{SC_e = 3.91}$$

Paso 5. Se obtienen los grados de libertad
gl_{trat}, gl_b, gl_{tot} y gl_e

a) gl_{trat} = número de tratamientos − 1 = columnas − 1 = 4 − 1 = 3

$$\boxed{gl_{trat} = 3}$$

b) gl_b = número de bloques − 1 = renglones − 1 = 4 − 1 = 3

$$\boxed{gl_b = 3}$$

c) gl_{tot} = (bloques × tratamientos) − 1 = 4 × 4 − 1 = 16 − 1

$$\boxed{gl_{tot} = 15}$$

d) gl_e = (bloques −1) (tratamientos − 1) = 3 × 3 = 9

$$\boxed{gl_e = 9}$$

Paso 6. Se obtienen los cuadrados medios.

a) $CM_{trat} = \dfrac{SC_{\text{trat}}}{gl_{\text{trat}}} = \dfrac{56.0025}{3}$

$$\boxed{CM_{trat} = 18.6675}$$

b) $CM_b = \dfrac{SC_b}{gl_b} = \dfrac{29.5875}{3}$

$$\boxed{CM_b = 9.8625}$$

c) $CM_e = \dfrac{SC_e}{gl_e} = \dfrac{3.9100}{9}$

$$\boxed{CM_e = 0.43444}$$

Paso 7. Se calculan las *F*.

$$F_{trat} = \frac{CM_{trat}}{CM_e} = \frac{18.6675}{0.43444} = 42.9691$$

$$F_{bloques} = \frac{CM}{CM_e} = \frac{9.8625}{0.43444} = 22.7016$$

Paso 8. Se construye la tabla del análisis de varianza.

Fv	*gl*	*SC*	*CM*	*F*
Tratamientos	3	56.0025	18.6675	42.9691[a]
Bloques	3	29.5875	9.8625	22.7016[a]
Error	9	3.9100	0.43444	
Total	15	89.5000		

[a] Significativa al 1%.

Paso 9. Se obtienen:

$F_{(gl_{trat.},\ gl_{e.},\ \alpha)}$ para tratamientos

$F_{(gl_{b.},\ gl_{e.},\ \alpha)}$ para bloques

En este caso, como tratamientos es igual a bloques, se tiene:

$F_{(3,\ 9,\ 0.05)} = 3.86$

$F_{(3,\ 9,\ 0.01)} = 6.99$

Paso 10. Conclusiones

Existe diferencia estadísticamente significativa tanto entre los tratamientos como entre los bloques.

En condiciones particulares, el resultado obtenido con la prueba "*F*" (análisis de varianza de dos factores en el diseño de bloques aleatorizados), y cuando se aplica a dos grupos o tratamientos (donde uno de ellos es el grupo control), también puede obtenerse el mismo resultado utilizando la prueba *t* de Student, para una muestra medida dos veces, muestras dependientes o relacionadas. Esto es $F = t^2$; estas condiciones son:

1. Cuando cada observación está formada por un par de:

   - Animales de la misma camada.
   - Lotes del mismo terreno.
   - Hojas de la misma planta.
   - Hermanos (gemelos).
   - Frutos del mismo árbol, etcétera.

2. Cuando el interés radica en conocer la diferencia entre los elementos de cada pareja.

Por último, debe hacerse notar que la *t* de Student considerada en este caso no nos indica la diferencia entre pareja y pareja.

Si el objetivo fuese *no* considerar la fuente de variación de los sujetos, entonces en el experimento se considerarían $n_1 + n_2$ sujetos, las dos muestras son independientes y, si $n_1 = n_2 = n$, se trata de una muestra medida dos veces, muestras dependientes o relacionadas.

Si se retoma el ejemplo de la sección 11.2.1, pero ahora con diez automóviles, en lugar de los 20 tenemos una sola muestra medida dos veces (una con el dispositivo *A* y otra con el dispositivo *B*), por lo que $n = 10$.

Se aplica el análisis de varianza de bloques aleatorizados con el paquete MACSTAT y se obtiene:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón F*
Tratamientos	1	0.2856	0.2856	132.8739[a]
Bloques	9	0.0255	0.00283	1.3184
Error	9	0.0193	0.00214	
Total	19	0.3305		

[a] Significativa al 1%.

Aplicando la *t* de Student para muestras dependientes $t = \frac{\bar{d}}{S_{\bar{d}}}$, donde $S_{\bar{d}} = 0.02073$, $\bar{d} = 0.239$

entonces $t = 11.52918$

por tanto, $t = \sqrt{F}$.

12.2 DISEÑO COMPLETAMENTE ALEATORIZADO

Este es un diseño general que consiste en *a* niveles para el primer factor *A* y *b* niveles para el segundo factor *B*.

Lo primero que debe considerarse es que los efectos principales de *A*, en el ANOVA (ANDEVA), tendrán $gl = a - 1$, mientras que para *B* serán $gl = b - 1$. No obstante, como existen $a \times b$ combinaciones, los grados de libertad del error o residual $gl = ab\,(n - 1)$. Por otra parte, la interacción $A \times B$ representa un conjunto de $(a - 1)\,(b - 1)$ comparaciones independientes.

La diferencia principal con el diseño de bloques aleatorizados, que es un caso especial, consiste en que se obtienen tres valores de *F* (Snedecor-Fisher): *i*) uno para los efectos principales de *A*, *ii*) el correspondiente a los efectos principales del factor *B*, y *iii*) otro para el efecto de interacción $A \times B$.

Ejemplo

Factor A (tratamientos $a = 2$).
$i = 1$ Naloxeno
$i = 2$ Morfina

Factor B (bloques $b = 3$).
$j = 1$ Agua simple
$j = 2$ Agua con azúcar
$j = 3$ Agua con sal

Se establecen las hipótesis para tratamientos α para bloques β y la interacción $\alpha\beta$.

$H_0 : \alpha_1 = \alpha_2 = 0 \quad ; \quad H_1 : \alpha_1 \neq \alpha_2$

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad ; \quad H_1 : \beta_1 \neq \beta_2 \neq \beta_3$

$H_0: (\alpha\beta)_{11} = (\alpha\beta)_{12} = 0$

$H_1: (\alpha\beta)_{11} \neq (\alpha\beta)_{12}$

El elemento por medir es la ingesta de agua en mililitros (ml)[2]
$n = 4$ sujetos (animales de laboratorio) por casilla.
$N = abn \Rightarrow \quad N = 2 \times 3 \times 4 = 24$

FACTOR A

FACTOR B		$i = 1$	$i = 2$
	$j = 1$	10, 8, 10, 10	5, 4, 6, 5
	$j = 2$	6, 8, 7, 5	8, 6, 6, 6
	$j = 3$	4, 5, 5, 4	8, 7, 7, 6

Paso 1. Calcule las sumas totales por casilla, por tratamiento y por bloques.

	$i = 1$	$i = 2$	$T_{.j}$
$j = 1$	$T_{1,1} = 38$	$T_{2,1} = 20$	$58 = T_{.1}$
$j = 2$	$T_{1,2} = 26$	$T_{2,2} = 26$	$52 = T_{.2}$
$j = 3$	$T_{1,3} = 18$	$T_{2,3} = 28$	$46 = T_{.3}$
$T_{i.}$	$T_{1.} = 82$	$T_{2.} = 74$	$T_{..} = 156$

Nota: $T_{.j}$ = La suma de los niveles o bloques (1, 2, 3).

Paso 2. Establezca las medias aritméticas.

	$i = 1$	$i = 2$	$\bar{X}_{.j}$
$j = 1$	$\bar{X}_{1,1} = 9.5$	$\bar{X}_{2,1} = 5.0$	$\bar{X}_{.1} = 7.25$
$j = 2$	$\bar{X}_{1,2} = 6.5$	$\bar{X}_{2,2} = 6.5$	$\bar{X}_{.2} = 6.500$
$j = 3$	$\bar{X}_{1,3} = 4.5$	$\bar{X}_{2,3} = 7$	$\bar{X}_{.3} = 5.750$
	$\bar{X}_{1.} = 6.833$	$\bar{X}_{2.} = 6.166$	$\bar{X}_{..} = 6.5$

[2] Valores redondeados.

Paso 3. Determine la suma de cuadrados de cada puntuación.

		FACTOR *A*		
		$i = 1$	$i = 2$	$T_{\cdot j}^{a}$
FACTOR *B*	$j = 1$	10^2 8^2 10^2 10^2 364	5^2 4^2 6^2 52 102	$T_{\cdot 1}^{2}$ 466
	$j = 2$	6^2 8^2 7^2 5^2 174	8^2 6^2 6^2 62 172	$T_{\cdot 2}^{2}$ 346
	$j = 3$	4^2 5^2 5^2 4^2 82	8^2 7^2 7^2 62 198	$T_{\cdot 3}^{2}$ 280
	T_{ij}^{2}.	T_{1}^{2}. 620	T_{2}^{2}. 472	T^{2}.. 1 092

Suma de cuadrados no corregida.

$$\sum_{i}^{2} \sum_{j}^{3} \sum_{k}^{4} X_{ijk}^{2} = 1\ 092, \text{ suma de cuadrados no corregida}$$

donde:

$\sum_{k}^{4}$ = la suma de las puntuaciones por casilla o celda.

$\sum_{j}^{3}$ = la suma de los resultados por celda en bloques o renglón.

$\sum_{i}^{2}$ = la suma de los resultados por tratamientos o columnas.

Paso 4. Obtenga la suma de cuadrados de los totales, divididos entre *n* (número de sujetos en la casilla); a esto se le llama *factor de corrección* para la suma de cuadrados dentro de grupos.

$$\frac{\sum_{i} \sum_{j} T_{ij}^{2}}{n_{celdas}} = \frac{38^2 + 20^2 + 26^2 + 26^2 + 18^2 + 28^2}{4} = 1\ 076$$

Paso 5. Obtenga la suma de cuadrados dentro de grupos.

$$SC_{dg} = \sum_{i}^{2} \sum_{j}^{3} \sum_{k}^{4} X_{ijk}^2 - \frac{\sum_{k}^{4} \sum_{i}^{2} T_{ij}^2}{n_{celdas}}$$

$$SC_{dg} = 1\ 092 - 1\ 076 = 16$$

$$\boxed{SC_{dg} = 16}$$

Paso 6. Obtenga la suma de cuadrados totales por columna y renglón.

$$\frac{\sum_{i}^{2} (T_{i\cdot})^2}{nb} = \frac{82^2 + 74^2}{4 \times 3} = \frac{6\ 724 + 5\ 476}{12} = 1\ 016.6667$$

$$\frac{\sum_{j}^{3} T_{\cdot j}^2}{na} = \frac{58^2 + 52^2 + 46^2}{4 \times 2} = \frac{3\ 364 + 2\ 704 + 2\ 116}{8} = 1\ 023.00$$

Paso 7. Calcule el factor de corrección para la suma de cuadrados de tratamiento y suma de cuadrados de bloques.

$$\frac{(T_{\cdot\cdot})^2}{N} = \frac{(156)^2}{24} = \frac{24\ 336}{24} = 1\ 014$$

donde $N = b \times a \times n$ = (bloques) (tratamientos) (número de observaciones por celda).

Paso 8. Determine la suma de cuadrados de tratamiento (SC_A, suma de cuadrados del factor A).

$$SC_A = \frac{\sum_{i}^{2} T_{i\cdot}^2}{na} - \frac{(T_{\cdot\cdot})^2}{N}$$

$$SC_A = 1\ 016.667 - 1\ 014$$
$$SC_A = 2.6667$$

Posteriormente obtenga la suma de cuadrados de bloques (SC_B).

$$SC_B = \frac{\sum_{j}^{3} T_{\cdot j}^2}{na} - \frac{(T_{\cdot\cdot})^2}{N}$$

$$SC_B = 1\,023 - 1\,014$$

$$\boxed{SC_B = 9}$$

Paso 9. Establezca la suma de cuadrados de la interacción (SC_{AB}).

$$SC_{AB} = \frac{\sum_i^2 \sum_j^3 T_{ij}^2}{n_{celdas}} - \frac{\sum_i^2 T_{i\cdot}^2}{nb} - \frac{\sum_j^3 T_{\cdot j}^2}{na} + \frac{(T_{\cdot\cdot})^2}{N}$$

$$SC_{AB} = 1\,076 - 1\,016.6667 - 1\,023 + 1\,014$$

$$\boxed{SC_{AB} = 50.3333}$$

Paso 10. Calcule la suma de cuadrados total.

$$SC_{tot} = \sum_i^2 \sum_j^3 \sum_k^4 X_{ijk}^2 - \frac{(T_{\cdot\cdot})^2}{N}$$

$$SC_{tot} = 1\,092 - 1\,014$$

$$\boxed{SC_{tot} = 78}$$

La relación que guardan las sumas de cuadrados calculadas anteriormente es la siguiente:

$$SC_{tot} = SC_{AB} + SC_B + SC_A + SC_{dg}$$

Paso 11. Determine los grados de libertad.

$gl_{dg} = N - ij = 24 - (2) \times (3) = 18$
$gl_A = i - 1 = 2 - 1 = 1$
$gl_B = j - 1 = 3 - 1 = 2$
$gl_{AB} = (i - 1)\,(j - 1) = (2 - 1)\,(3 - 1) = 2$
$gl_{tot} = N - 1 = 24 - 1 = 23$

y se cumple

$$gl_{tot} = gl_{dg} + gl_A + gl_B + gl_{AB}$$

Paso 12. Establezca los cuadrados medios (CM).

$$CM_A = \frac{SC_A}{gl_A} = \frac{2.6667}{1} = 2.6667$$

$$CM_B = \frac{SC_B}{gl_B} = \frac{9}{2} = 4.500$$

$$CM_{AB} = \frac{SC_{AB}}{gl_{AB}} = \frac{50.333}{2} = 25.16665$$

$$CM_{dg} = \frac{SC_{dg}}{gl_{dg}} = \frac{16.00}{18} = 0.88889$$

Paso 13. Construya la tabla del análisis de varianza.

FV	*gl*	*SC*	*CM*	*F*
Factor *A* (tratamientos)	1	2.6667	2.6667	3.00 NS[a]
Factor *B* (bloques)	2	9.0000	4.500	5.0625[b]
A × *B* (interacción)	2	50.333	25.16665	28.3125[c] $P < 0.01$
Dentro de grupos (residual)	18	16.000	0.88889	
Total	23	78		

[a] No significativo.
[b] Significativa al 5% únicamente.
[c] Significativa al 1%.
$P < 0.01$ significativa para una probabilidad menor al 1%.

12.2.1 El concepto de interacción

En el diseño completamente aleatorizado de dos factores, no se pueden considerar por separado los efectos principales de cada uno de los que intervienen en dicho diseño. Por esto, el efecto de interacción de los dos factores (variables independientes) sobre un tercero (variable dependiente) es el de unir los efectos principales de cada uno de los factores.

En otras palabras, existe interacción si el efecto de un factor (tratamientos), no es el mismo para todos los niveles del otro factor (bloques).

CASO 1.
Variable de interés → sílabas recordadas (en dos situaciones factor B, aviso y no aviso, de un examen realizado en corto y largo plazos, factor *A*)

		FACTOR *A*	
		Corto	Largo
FACTOR *B*	Aviso	9, 11, 8, 7, 7	5, 7, 6, 3, 9
	No aviso	5, 9, 7, 4, 8	3, 4, 5, 1, 4

Al aplicar el paquete MACSTAT se obtiene:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón* ***F***
Tratamientos	1	39.2000	39.2000	10.8889**
Bloques	1	24.2000	24.2000	6.7222**
Interacción	1	0.8000	0.8000	0.2222+
Residuo	16	57.6000	3.6000	
Total	19	121.8000		

+ No significativo.
** Significativa al 1%.

Conclusión: En este caso no existe interacción entre el factor B y el factor A.

CASO 2.

En un estudio realizado por E. Varas y L. González en una clínica de odontopediatría de la UNAM (1998), participaron 32 niños entre tres y cuatro años de edad. La variable de interés era el número de "caries de biberón" considerando el factor A (**consistencia** líquido o sólido) y el factor B (característica **nutricional** sin azúcar o con azúcar). Los 32 niños fueron clasificados en las cuatro categorías, ocho en cada una de ellas, tomando en cuenta los hábitos de su ingesta principal, y se obtuvieron los siguientes resultados:

		FACTOR *A*	
		Líquido	*Sólido*
FACTOR *B*	Sin azúcar	8, 2, 13, 4 1, 10, 4, 7	8, 0, 6, 2 1, 8, 0, 0
	Con azúcar	11, 3, 2, 8 12, 1, 1, 6	0, 0, 3, 1 0, 1, 2, 11

Mediante el paquete MACSTAT se obtiene:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón* ***F***
Tratamientos	1	78.1250	78.1250	4.9351*
Bloques	1	4.5000	4.5000	0.2843+
Interacción	1	0.1250	0.1250	0.0079+
Residuo	28	443.2500	15.8304	
Total	31	526.0000		

+ No significativo.
* Significativa al 5%.

Conclusión: El principal efecto significativo es la consistencia del alimento.

CASO 3.

Si el resultado de un análisis de varianza de dos factores fuere el siguiente:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón* ***F***
Tratamientos	2	28.1667	14.0833	8.8947**
Bloques	1	30.0833	30.0833	19.0000**
Interacción	2	3.1667	1.5833	1.0000+
Residuo	6	9.5000	1.5833	
Total	11	70.9167		

+ No significativo.
** Significativa al 1%.

Conclusión: No existe interacción entre el factor *A* y el factor *B*. Únicamente entre los tratamientos, así como entre los bloques, existe diferencia estadísticamente significativa.

CASO 4.

Cuando los resultados del análisis de varianza de dos factores fueran de la siguiente manera:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	Cuadrados medios	*Razón* ***F***
Tratamientos	2	473.6444	236.8222	0.8570+
Bloques	2	128.8444	64.4222	0.2331+
Interacción	4	187.0222	46.7556	0.1692+
Residuo	36	9948.000	276.3444	
Total	44	10 737.9111		

+ No significativa.

Conclusión: No existe diferencia estadísticamente significativa ni entre los tratamientos, ni entre los bloques, así como tampoco existe interacción entre el factor *A* y el factor *B*.

CASO 5.

Dados los siguientes resultados:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón* ***F***
Tratamientos	2	66.8683	33.4342	124.5656[a]
Bloques	2	3.3033	1.6516	6.1535[a]
Interacción	4	24.5114	6.1278	22.8305[a]
Residuo	99	26.5722	0.2684	
Total	107	121.2552		

[a] Significativa al 1%.

Implica la diferencia estadísticamente significativa entre tratamientos, bloques y la existencia de interacción entre ambos.

CASO PARTICULAR. Cuando no existe diferencia estadísticamente significativa entre los tratamientos (factor A), ni entre los bloques (factor B), pero sí una interacción estadísticamente significativa ($A \times B$).

Ejemplo

	A_1	A_2	A_3
B_1	7, 10, 10, 11, 12	4, 6, 7, 9, 9	2, 2, 3, 7, 6
B_2	6, 5, 8, 9, 12	10, 10, 11, 11, 13	5, 4, 7, 8, 11
B_3	3, 3, 4, 8, 7	4, 6, 7, 8, 10	7, 9, 9, 10, 10

Algunos autores usan el siguiente formato para el diseño completamente aleatorizado de dos factores (*véase* A, L. Edwards, 1967).

A_1			A_2			A_3		
B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
7	6	3	4	10	4	2	5	7
10	5	3	6	10	6	2	4	9
10	8	4	7	11	7	3	7	9
11	9	8	9	11	8	7	8	10
12	12	7	9	13	10	6	11	10

Cuyo resultado es la siguiente tabla:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrados medios*	*Razón* ***F***
Tratamientos	2	21.1111	10.5556	2.2619[+]
Bloques	2	27.7778	13.8889	2.9762[+]
Interacción	4	152.2222	38.0556	8.1548[*]
Residuo	36	168.0000	4.6667	
Total	44	369.1111		

[+] No significativa.
[*] Significativa al 5%.

RESUMEN

Un diseño de bloques aleatorizados se emplea con el propósito de reducir una variación extraña, incrementando la precisión con la cual los efectos de los tratamientos se estiman.

Este capítulo es una introducción a los *experimentos factoriales*, los que se definen como un experimento en el cual los tratamientos consisten en todas las

combinaciones de los niveles de los factores. Tales experimentos pueden realizarse utilizando varios diseños:

- Bloques aleatorizados.
- Completamente aleatorizado (con su respectiva interacción).
- Cuadrados latinos.
- Cuadrados grecolatinos.
- Factorial de parcelas divididas, etcétera.

PROBLEMAS

12.1 En una institución educativa se emplearon cuatro técnicas diferentes A_1, A_2, A_3, A_4 de enseñanza de las matemáticas a nivel de bachillerato. Participan tres grupos (tres bloques) de aptitudes desiguales; B_1 = bajo rendimiento, B_2 = promedio y B_3 = alto rendimiento. Posteriormente al curso se les aplicó un examen de conocimientos en matemáticas, sobre una base de evaluación de 10 puntos, de donde se obtuvieron los siguientes resultados:

	A_1	A_2	A_3	A_4
B_1	3	3	5	6
B_2	6	7	6	8
B_3	7	8	9	9

Aplicar un análisis de varianza de bloques aleatorizados para:

a) Técnicas de enseñanza (A_1, A_2, A_3, A_4)
b) Grupos (B_1, B_2, B_3)

12.2 Se realiza una investigación de dos fármacos A_1 y A_2, ya que actualmente existe un fármaco A_3 contra la hepatitis. Esta investigación se lleva a cabo en cuatro hospitales;[3] por cada mil pacientes que se encuentran en tratamiento se obtiene el número de enfermos:

	A_1	A_2	A_3
B_1	0	7	7
B_2	1	3	8
B_3	2	4	10
B_4	0	2	7

[3] Se emplea el diseño de bloques aleatorizados para eliminar el efecto en los diferentes hospitales.

12.3 Se lleva a cabo un experimento en siete bloques y tres parcelas (tres tratamientos) que contienen cuatro, ocho y 16 unidades de un ingrediente activo diseñado para exterminar moscas adultas en una granja de pollos. Los siguientes resultados muestran el número de moscas muertas en los siete bloques y tres tratamientos:

		Tratamiento		
		4	8	16
	1	445	414	247
	2	113	127	147
	3	122	206	138
Bloques	4	227	78	148
	5	132	172	356
	6	31	45	29
	7	177	103	63

12.4 Realice un análisis de varianza para el siguiente diseño de bloques aleatorizados; establezca la hipótesis nula para bloques y para tratamientos y concluya utilizando un $\alpha = 1\%$.

		Tratamiento		
		T_1	T_2	T_3
	b_1	43.1	42.3	45.6
	b_2	27.6	29.3	35.1
Bloques	b_3	27.0	26.5	29.3
	b_4	31.0	32.3	34.5
	b_5	21.3	20.2	25.6

12.5 Realice un análisis de varianza del diseño de bloques aleatorizados con los siguientes datos:

		Tratamiento			
		T_1	T_2	T_3	T_4
	b_1	2	3	1	0
Bloques	b_2	4	-1	1	-1
	b_3	1	-2	-3	-2
	b_4	1	-5	-4	-3

12.6 Se estudian las funciones del cerebro en *macacus rhesus*, infligiendo una lesión bilateral en la parte inferior del temporal relacionada con la visión.

Las variables en estudio son tanto el efecto de la operación como la edad en que ésta se realiza.

Los siguientes datos (H. Harlow) reflejan la ejecución en una prueba de agudeza visual que exige la discriminación entre una barra vertical y una

horizontal, separadas a la mitad por 8 mm. Las mediciones se basan en el número de respuestas correctas a preguntas.

	Edad del mono (días)		
	100	300	730
Grupo control	93.90	79.94	89.91
	91.88	50.70	96.84
Grupo experimental (operados)	86.71	76.91	60.72
	68.70	51.96	93.68

Presente la tabla ANOVA e interprete los resultados de estos datos.

12.7 Se mide el tiempo de manufactura en segundos, tomando en consideración nueve máquinas y tres procesos diferentes, de donde surgen los siguientes datos:

		Máquinas								
		1	2	3	4	5	6	7	8	9
	A	18.76	20.69	19.77	19.85	22.28	20.39	24.31	22.90	19.28
		21.18	23.20	23.94	18.92	20.45	21.80	26.29	25.42	22.04
Procesos	*B*	21.28	16.85	20.75	18.72	18.97	19.52	20.08	21.27	15.67
		19.10	20.16	21.49	16.14	20.31	21.27	19.36	17.82	18.84
	C	21.74	22.68	21.90	20.28	19.89	21.12	23.02	27.48	18.70
		18.99	23.59	18.61	18.71	18.36	18.59	18.85	22.95	23.39

Realice un análisis de varianza para este diseño de 3 × 9 y dos respuestas por cada combinación.

12.8 Se estudian los efectos de la temperatura y la corriente en los filamentos de un tipo de foco.

		Corriente en el filamento (Amps)			
		100	120	140	160
	55°C	3774	4710	4176	4540
		4364	4180	4140	4530
Temperatura		4374	4514	4398	3964
	65°C	4216	3828	4122	4484
		4524	4170	4280	4332
		4136	4180	4226	4390

Realice un análisis de varianza para este diseño 2 × 4 y tres réplicas en cada combinación.

12.9 Se lleva a cabo un experimento para conocer los efectos de choque-eléctrico (*sh*) y el ruido blanco (*rb*) en la respuesta galvánica en la piel. Participan 40 sujetos en cuatro niveles de (*sh*) (0.25, 0.50, 0.75 y 1.00 mA) y dos niveles de (*rb*) (40 y 80 *dB*), con los siguientes datos:

		sh			
		0.25	0.50	0.75	1.00
	40 *dB*	3, 7.9	5, 11, 13	9, 12, 14	6, 11, 12
rb		4, 1	8, 3	11, 5	7, 4
	80 *dB*	5, 10, 10	6, 12, 15	11, 18, 15	7, 9, 15
		6, 3	9, 5	13, 9	14, 7

Realice un análisis de varianza para este diseño de 2 × 4 y cinco respuestas por cada combinación.

12.10 Efectúe el análisis de varianza apropiado, establezca las hipótesis correspondientes y haga las conclusiones pertinentes.

		B	
		b_1	b_2
	a_1	10	5
		8	4
		10	6
		10	5
	a_2	6	8
A		8	6
		7	6
		5	6
	a_3	4	8
		5	7
		5	7
		4	6

12.11 Se lleva a cabo un estudio sobre la repercusión de cuatro tipos diferentes de estrategias (descuento, rifa, muestras gratuitas y promoción) para la venta de tres productos (champú, enjuague bucal y vitaminas); se escogen al azar 48 tiendas de autoservicio de una cadena nacional y se obtienen los siguientes datos, que miden el porcentaje de cambio en el volumen de ventas (vg. "−1" significa que las ventas disminuyeron en un 10%, o no hubo variación, "2" un aumento del 20% etcétera).

Realice el análisis de varianza apropiado, para este diseño completamente aleatorizado.

		B		
		Champú	*Enjuage bucal*	*Vitaminas*
A	Descuento	−1	2	1
		−1	1	1
		−1	1	3
	Rifa	0	1	2
		0	3	5
		0	5	5
	Muestras gratis	−1	4	6
		−1	4	4
		−1	−1	0
		−2	1	0
	Promoción	−1	−1	−1
		0	−1	0
		−1	2	3
		0	2	2
		0	2	2
		1	1	1

12.12 Hacer el análisis de varianza en dos direcciones con los siguientes datos:

	Factor *A:* Ansiedad inducida		
Factor *B:* Trabajo en forma:	Alta	Media	Baja
Masiva	9, 7, 5, 8	5, 7, 6, 6	3, 5, 5, 3
Distribuida	12, 10, 11, 1	8, 8, 7, 6	7, 5, 7, 5

12.13 Se hizo un estudio sobre la forma en la que la imagen se relaciona con el procesamiento de información. Se dividieron los sujetos en tres grupos: imagen visual, imagen al hablar y control. Para todos los grupos la tarea consistía en decir el alfabeto tan rápido como fuera posible, sin omisiones. Antes del experimento, al grupo visual se le pidió que se imaginara las letras antes de decirlas, en tanto que al grupo oral se le pidió que dijera las letras en silencio antes de pronunciarlas en voz alta; el grupo control no tuvo consigna: sólo decir el abecedario.

Suponga que se desea hacer este experimento utilizando un diseño semejante con cuatro sujetos en cada grupo y que cada sujeto repita la tarea cinco veces.

	Imagen visual				Imagen oral				Control			
	Sujeto				Sujeto				Sujeto			
Ensayo	1	2	3	4	5	6	7	8	9	10	11	12
1	25	28	18	20	8	9	6	8	10	9	13	8
2	22	32	18	22	7	11	7	10	9	8	9	10
3	20	26	13	17	10	7	6	8	8	12	9	7
4	22	20	20	14	6	8	6	6	7	11	8	11
5	18	23	14	16	7	9	8	9	8	8	9	7

12.14 Se realizó un experimento para determinar el efecto de dos variables independientes: *A*) la consistencia del alimento (líquido o sólido) y *B*) la influencia de un fármaco (naloxeno o un placebo), la variable dependiente es el número de úlceras estomacales, que se les encontró a las ratas, cuando fueron sacrificadas. Se utilizaron 32 animales de laboratorio, ratas de la misma edad y peso, seleccionadas en forma aleatoria para cada una de las cuatro condiciones. A continuación se muestra el número de úlceras para cada una de las ocho ratas en cada condición del experimento.

		A (Alimento)	
		Líquido	Sólido
B (Fármaco)	Placebo	13, 1, 10	0, 0, 0, 2
		4, 8, 7, 4, 2	6, 8, 8, 1
	Naloxeno	3, 2, 8, 11	0, 0, 0, 1
		12, 1, 1, 6	1, 2, 3, 11

12.15 Supóngase que hay dos tipos de maíz A_1 y A_2, tres clases de fertilizante B_1, B_2 y B_3 y se desea comprobar si existe diferencia estadísticamente significativa entre los tipos de maíz y entre los fertilizantes; también se quiere conocer si existe interacción entre el tipo de maíz y la clase de fertilizante.

		Maíz	
		A_1	A_2
Fertilizante	B_1	1, 3	2, 1
	B_2	4, 3	5, 6
	B_3	8, 10,	9, 8

Parte 5

Asociación

Análisis de regresión lineal 13

PROPÓSITOS

El propósito central de este capítulo es que el lector comprenda, maneje y solucione situaciones relacionadas con las operaciones, procesos y modelo del análisis de regresión lineal.

De igual modo, el lector podrá:

1. Explicar el modelo de regresión lineal bivariado.
2. Identificar las características del modelo para la obtención de la recta de la regresión.
3. Desarrollar el procedimiento de cálculo del modelo de regresión lineal y relacionarlo con el diagrama de dispersión.
4. Explicar el cálculo de la recta de regresión y señalar su utilidad.
5. Explicar las características de la evaluación de la ecuación de regresión.
6. Obtener el intervalo de confianza adecuado de la ecuación, calculada de la recta de regresión, mediante los siguientes pasos:
   6.1 Intervalo de confianza para α, β, $\sigma^2_{Y \mid X}$, $\mu_{Y/X}$.
   6.2 Prueba de hipótesis para la regresión.
7. Desarrollar los procedimientos de cálculo de los estadísticos mencionados en 6.1 y 6.2 mediante la interpretación de los resultados obtenidos con el paquete estadístico MACSTAT.
8. Reconocer las limitaciones de cada método para calcular el coeficiente de correlación.
9. Aplicar el estadístico intervalo de confianza para el coeficiente de correlación, así como la prueba de hipótesis para dicho coeficiente al establecer el nivel de confianza.

10. Desarrollar el procedimiento de cálculo para los coeficientes citados en el inciso 9.
11. Explicar la prueba de hipótesis entre dos rectas de regresión e indicar su utilidad.
12. Desarrollar el cálculo de la prueba de hipótesis entre dos rectas de regresión para la toma de decisiones.
13. Describir la comparación entre dos coeficientes de correlación.
14. Considerar, con respecto al cálculo: comparación entre dos coeficientes de correlación; los aspectos básicos para evitar que ciertas variables afecten el estudio.
15. Desarrollar el cálculo de la comparación entre los coeficientes de correlación.
16. Explicar y desarrollar el cálculo de la regresión múltiple, así como de los siguientes casos:

    a) Caso particular para dos variables independientes.
    b) Relación entre el análisis de varianza y el de regresión múltiple.
    c) Evaluación de la ecuación de regresión múltiple.
    d) Prueba de hipótesis para los coeficientes de regresión.
    e) Coeficiente de correlación múltiple.

    16.1 Reconocer las limitaciones y consideraciones para su cálculo.
17. Reconocer la utilidad de los contenidos y ejercicios: para la aplicación en la solución de problemas relacionados con los análisis de regresión lineal, en el campo de trabajo y la toma de decisiones, la escuela y otras áreas de conocimiento.
18. Aplicar e interpretar los resultados obtenidos mediante el paquete estadístico MACSTAT.

INTRODUCCIÓN

En los capítulos anteriores se ha considerado sólo la medición referida a la unidad observacional o experimental (el sujeto), es decir, el análisis estadístico de una sola variable. Sin embargo, en ocasiones se requiere una tarea distinta de analizar un solo conjunto de datos (estudio descriptivo de una sola población) o de comparar dos o más conjuntos de datos (estudio comparativo entre dos o más poblaciones). Para estas situaciones, existe otro tipo de estudio, el *correlativo*, cuya utilidad radica en encontrar la posible relación entre dos o más variables, así como en medir el grado de asociación o variación mutua entre las mismas. El estudio correlativo es útil, por ejemplo, para conocer si los estudiantes notables en actividades artísticas sobresalen en forma relevante, regular, baja o no sobresalen en su formación académica; asimismo, con el fin de determinar si la edad y el aprendizaje en animales de laboratorio está relacionada y sigue algún modelo de regresión.

En este tipo de análisis debe considerarse la posible dependencia y grado de asociación de una variable con respecto a otra u otras; en el ejemplo anterior, se dice que el aprendizaje (Y) depende de la edad (X) (Y depende de X). *Debe tenerse muy claro que ni la regresión ni la correlación implican causalidad.*

13.1 EL MODELO DE REGRESIÓN

Existen dos tipos de modelos, cuyo desarrollo es el mismo, pero que varían en algunos fundamentos teóricos.

Modelo I. Cuando los valores de X están fijos y determinados por el diseño experimental, se consideran medidos sin error. Los valores de Y (variable dependiente) se determinarán por su respectiva X (variable independiente). La distribución probabilística de la población de Y se supone *normal*, y de ahí se seleccionan de manera aleatoria los valores de Y. A este modelo se le conoce como el *modelo estándar de regresión.*

Modelo II. Tanto los valores de X como los de Y son aleatorios, por lo que se debe trabajar con muestras aleatorias de una distribución poblacional bivariada. En este modelo se debe cumplir el supuesto de homoscedasticidad (la varianza de Y es la misma para toda X).[1]

Cuando el investigador supone una posible relación entre la variable independiente y la dependiente establece un modelo específico para obtener la mejor estimación ($\hat{Y}$) de Y. Con este fin, se obtiene la ecuación de regresión, que es el modelo poblacional por estimar a partir de los datos de la muestra. Posteriormente, lo más importante es evaluar dicho modelo y calcular la confianza de que esta estimación describa adecuadamente a Y.

En este apartado se desarrolla el modelo de regresión lineal (es el más sencillo, y uno de los más importantes). Como el problema consiste en el ajuste de los datos a una línea recta, se parte de la ecuación de la línea recta:

$$Y = \alpha + \beta X$$

donde

α = ordenada al origen.
β = pendiente.

En este modelo matemático es necesario incluir un componente de error ε;[2] lo anterior se debe a que, en su forma actual, el modelo no incorpora un supuesto básico del análisis de regresión: considerar Y como variable aleatoria. Así, el modelo matemático $Y = \alpha + \beta X$ se convierte en el modelo estadístico $Y = \alpha + \beta X + \varepsilon$,

[1] En este texto consideraremos el modelo II, por ser el caso más general.
[2] Este componente de error se distribuye normalmente, con media cero y varianza σ^2, o sea, $\varepsilon_i \sim N(0, \sigma^2)$; $i = 1, 2, \ldots, N$.

donde $\varepsilon = Y - (\alpha + \beta X)$ constituye una variable aleatoria que describe lo distante que está una respuesta individual de la línea de regresión poblacional. El modelo estadístico también se establece de la siguiente manera:

$$\mu_{Y|X} = \alpha + \beta X + \varepsilon$$

donde:

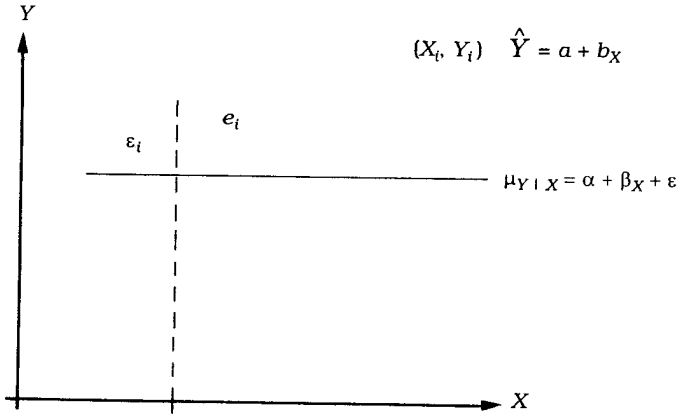
$\varepsilon = Y - (\alpha + \beta X)$, o bien, $\varepsilon = Y - \mu_{Y|X}$

$\mu_{Y|X}$ es la media de la variable aleatoria Y dado un valor fijo de X correspondiente y, por supuesto,

$\sigma^2_{Y|X}$ es la varianza de la variable aleatoria Y dado X.

El término *regresión lineal* implica que $\mu_{Y|X}$ está linealmente relacionado con X; por medio de la ecuación de regresión lineal poblacional $\mu_{Y|X} = \alpha + \beta X + \varepsilon$ se puede estimar $\mu_{Y|X}$ mediante $\hat{Y}$, que es la línea de regresión ajustada $\hat{Y} = a + bX$ (a y b son los valores muestrales para estimar α y β).

Cada par de observaciones satisface la relación $\hat{Y} = a + bX + e$, en donde $e = Y - \hat{Y}$; se le denomina *residuo* (o *residual*) y se obtiene para cada Y. Debe destacarse que $e \neq \varepsilon$, lo cual se representa en una gráfica:



X_i, Y_i es cualquier par de observaciones muestrales
ε_i = error de muestreo, y $\bar{\varepsilon} = 0$
e_i = residual = $Y_i - \hat{Y}_i$
(Error de predicción o de medición).

13.2 CÁLCULO DE LA RECTA DE REGRESIÓN

Para obtener la recta de regresión se utiliza el modelo:

$$Y = \alpha + \beta X + \varepsilon$$

con la estimación muestral,

$$\hat{Y} = a + bX$$

Para calcular a y b se aplica el siguiente desarrollo en el que se utilizan determinantes:

$$a = \frac{\begin{vmatrix} \Sigma Y & \Sigma X \\ \Sigma XY & \Sigma X^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{vmatrix}} \qquad b = \frac{\begin{vmatrix} n & \Sigma Y \\ \Sigma X & \Sigma XY \end{vmatrix}}{\begin{vmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{vmatrix}}$$

Ejemplo

Dados los siguientes datos :

X	1	2	3	4	5	6	7	8	9	10
Y	37	24	28	19	18	15	18	9	10	0

Paso 1. Se ordenan los datos en columnas y se obtienen X^2, Y^2, XY con sus sumas correspondientes: ΣX, ΣY, ΣX^2, ΣXY, ΣY^2, Y, n.

	X	Y	X^2	Y^2	XY
	1.0000	37.0000	1.0000	1 369.0000	37.0000
	2.0000	24.0000	4.0000	576.0000	48.0000
	3.0000	28.0000	9.0000	784.0000	84.0000
	4.0000	19.0000	16.0000	361.0000	76.0000
	5.0000	18.0000	25.0000	324.0000	90.0000
	6.0000	15.0000	36.0000	225.0000	90.0000
	7.0000	18.0000	49.0000	324.0000	126.0000
	8.0000	9.0000	64.0000	81.0000	72.0000
	9.0000	10.0000	81.0000	100.0000	90.0000
	10.0000	0.0000	100.0000	0.0000	0.0000
Σ	55	178	385	4 144	713
	ΣX	ΣY	ΣX^2	ΣY^2	ΣXY

$n = 10$

Paso 2. Se calculan a y b.

$$\alpha = \frac{\begin{vmatrix} \Sigma Y & \Sigma X \\ \Sigma XY & \Sigma X^2 \end{vmatrix}}{\begin{vmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{vmatrix}} = \frac{\begin{vmatrix} 178 & 55 \\ 713 & 385 \end{vmatrix}}{\begin{vmatrix} 10 & 55 \\ 55 & 385 \end{vmatrix}} = \frac{68\,530 - 39\,215}{825} = 35.5333$$

$$b = \frac{\begin{vmatrix} n & \Sigma Y \\ \Sigma X & \Sigma XY \end{vmatrix}}{\begin{vmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{vmatrix}} = \frac{\begin{vmatrix} 10 & 178 \\ 55 & 713 \end{vmatrix}}{\begin{vmatrix} 10 & 55 \\ 55 & 385 \end{vmatrix}} = \frac{7\,130 - 9\,790}{825} = -3.22424$$

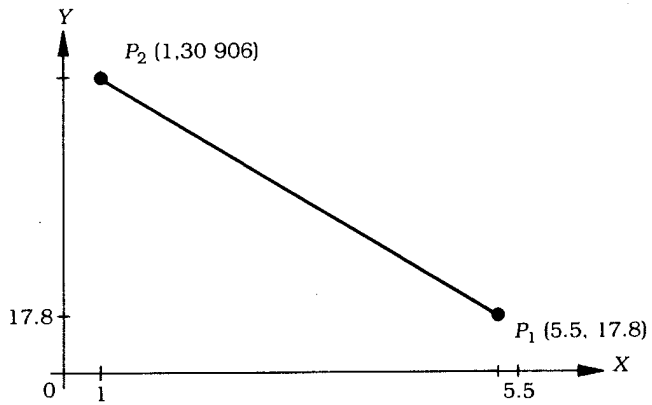
Paso 3. Se sustituye en el modelo.

$$\hat{Y} = a + bX$$

$$\hat{Y} = 35.5333 - 3.22424\,X \qquad \ldots \text{(I)}$$

Paso 4. Se obtiene la gráfica de la ecuación anterior; con este fin, se consideran dos puntos y uno lo constituyen las medias de los datos:

$$P_1\ (\overline{X}, \overline{Y})$$



Otro punto puede ser el que forma el valor más pequeño de X, el cual se sustituye en la ecuación de la recta para obtener Y:

$$P_2\ (X_{\text{mín}}, \hat{Y}_{\text{mín}})$$

donde

$$\overline{X} = \frac{\Sigma X}{n} = \frac{55}{10} = 5.5$$

$$\overline{Y} = \frac{\Sigma Y}{n} = \frac{178}{10} = 17.8$$

$$P_1 = (5.5,\ 17.8).$$

Al sustituir en la ecuación (I) el valor de $X_{\text{mín}} = 1$, se tiene:

$$\hat{Y}_{\text{mín}} = 35.5333 - 3.22424(1)$$

$$\hat{Y}_{\text{mín}} = 32.30906$$

$$P_2 = (1,\ 32.30906)$$

13.3 EVALUACIÓN DE LA ECUACIÓN DE REGRESIÓN

Dos métodos son los más adecuados para obtener un nivel de confianza satisfactorio (90, 95, 99%) en la ecuación de regresión poblacional estimada por la recta muestral que se calculó anteriormente. El primero consiste en construir los intervalos de confianza; el segundo, en realizar las pruebas de hipótesis para la regresión. A continuación se analizarán estos procedimientos.

13.3.1 Intervalos de confianza para α, β, $\sigma^2_{Y|X}$ y $\mu_{Y|X}$

Para el coeficiente de regresión α es

$(a - t_{(gl,\ \alpha/2)}\ Sa < \alpha < a + t_{(gl,\ \alpha/2)}\ Sa)$

nivel de confianza
donde: $\alpha = 1,\ 5,\ 10\%$, etcétera.
Para el coeficiente de regresión β es

$(b - t_{gl,\alpha/2}\ Sb < \beta < b + t_{gl,\alpha/2}\ Sb)$

Para la media se tiene lo siguiente.

$(\overline{Y} - t_{gl,\alpha/2}\ S_{\overline{Y}|X} < \mu_{Y|X} < \overline{Y} + t_{gl,\alpha/2}\ S_{\overline{Y}|X})$

donde $gl = n - 2$

Para la varianza

$$\frac{(n-2)\ S^2_{\overline{Y}|X}}{\chi^2_{1-\alpha/2}} < \sigma^2_{Y|X} < \frac{(n-2)\ S^2_{\overline{Y}|X}}{\chi^2_{\alpha/2}}$$

En los intervalos anteriores se considera la varianza de los errores muestrales de la estimación $S^2_{\hat{Y}|X}$ o el error estándar de la estimada $S_{\hat{Y}|X}$.

En algunos casos, para calcular la varianza de la media, que es $S^2_{\overline{Y}|X}$, o la desviación estándar de la media $S_{\overline{Y}|X}$, es necesario obtener $S^2_{\hat{Y}|X}$ o SCe, sustituyéndolo en la fórmula de $S^2_{\overline{Y}|X}$

$$S^2_{\hat{Y}|X} = \frac{S^2_{Y|X}}{n} = \frac{\dfrac{SCe}{n-2}}{n}$$

donde SCe = suma de cuadrados del error que se considerará en la tabla del análisis de varianza.

Donde $SCe = \Sigma e_i^2 = \Sigma(Y_i - \hat{Y}_i)^2$ también conocida como *suma de cuadrados de las desviaciones*, o sea $e_i = Y_i - (a + b X_i)$, e_i = error en la medición.

Para calcular $S^2_{\hat{Y}|X}$ son necesarias las varianzas muestrales de X y Y. Una forma de obtenerlas es la siguiente:

$$S_X^2 = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n-1}$$

$$S_Y^2 = \frac{\Sigma Y^2 - \dfrac{(\Sigma Y)^2}{n}}{n-1}$$

Se sustituyen en la siguiente expresión para obtener $S_{\hat{Y}|X}$:

$$S^2_{\hat{Y}|X} = \frac{n-1}{n-2}(S_Y^2 - b^2 S_X^2).$$

La anterior constituye la varianza de $\hat{Y}$, que es la suma de los cuadrados de las desviaciones divididas entre los grados de libertad (*gl*).

S_b se llama *error estándar* de b, donde b es una estimación de β, por lo que:

$$S_b^2 = \frac{S_{\hat{Y}|X}}{\Sigma (X - \overline{X})^2} = \frac{\dfrac{SCe}{n-2}}{\Sigma (X - \overline{X})^2}$$

$$S_b = \frac{S_{\hat{Y}|X}}{S_X \sqrt{n-1}}$$

Al retomar el ejemplo anterior se calcularán los intervalos de confianza de α, β, $\sigma^2_{Y|X}$ y $\mu_{Y|X}$

$$S_X^2 = \frac{\Sigma X^2 - \dfrac{(\Sigma X)^2}{n}}{n-1} = \frac{385 - \dfrac{3\,025}{10}}{9} = \frac{82.5}{9} = 9.16667$$

$$\boxed{S_X = 3.02765}$$

$$S_Y^2 = \frac{\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}}{n-1} = \frac{4\ 144 - \frac{31\ 684}{10}}{9} = \frac{975.60}{9} = 108.40$$

$$\boxed{S_Y = 10.41153}$$

Se sustituye en la siguiente expresión:

$$S_{\hat{Y}|X}^2 = \frac{n-1}{n-2}\left(S_Y^2 - b^2 S_X^2\right) = \frac{9}{8}\,(108.40 - 10.3957 \times 9.16667)$$

$$= \frac{9}{8}\,(108.40 - 95.29417) = 14.74406$$

$$\boxed{S_{\hat{Y}|X} = 3.83980}$$

Con estos resultados se calcula la varianza y el error estándar de la media.

$$S_{\bar{Y}|X}^2 = \frac{S_{\hat{Y}|X}^2}{n} = \frac{14.74406}{10} = 1.474406$$

$$\boxed{S_{\bar{Y}|X} = 1.21425}$$

Se sustituyen los valores anteriores y se obtiene

$$S_a = \frac{S_{\hat{Y}|X}\sqrt{\Sigma X^2}}{S_X\sqrt{n(n-1)}} = \frac{3.83980\ \sqrt{385}}{3.02765\ \sqrt{90}} = \frac{75.3423}{28.7281}$$

$$\boxed{S_a = 2.62308}$$

Para un intervalo de confianza de 95%
$\alpha = 5\%$, $t_{\alpha/2,\ 8} = 2.306$

como $a = 35.5333$, al sustituir en el intervalo resulta
$(35.5333 - 2.306 \times 2.62308 < \alpha < 35.5333 + 2.306 \times 2.62308)$

$$\boxed{29.5 < \alpha < 41.6}$$

Para el intervalo de confianza de β se necesita calcular:

$$S_b = \frac{S_{\hat{Y}|X}}{S_X\sqrt{n-1}}$$

De la sustitución de valores se obtiene

$$\boxed{S_b = 0.42275}$$

Para un intervalo de 95%, $t_{\alpha/2,\ 8} = 2.306$.
Se realizan las sustituciones procedentes:
$(-3.22424 - 2.306 \times 0.42275 < \beta < -3.22424 + 2.306 \times 0.42275)$

$$\boxed{-4.2 < \beta < -2.25}$$

Para construir el intervalo de $\mu_{Y|X}$ también al 95%, se sustituyen los valores en la siguiente expresión:
$\overline{Y} - t_{\alpha/2,\ gl}\ S_{\overline{Y}|X} < \mu_{Y|X} < \overline{Y} + t_{\alpha/2,\ gl}\ S_{\overline{Y}|X}$
$(17.8 - 2.306 \times 1.21425 < \mu_{Y|X} < 17.8 + 2.306 \times 1.21425)$

$$\boxed{(15.0 < \mu_{Y|X} < 20.6)}$$

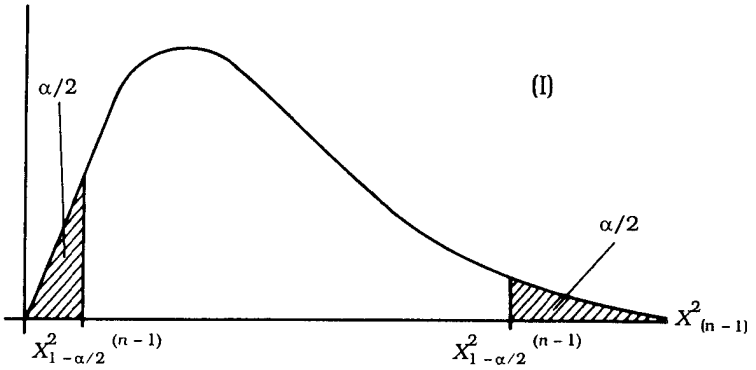
Finalmente, para calcular la varianza estimada de la media se necesita la varianza de la media que es:

$$\boxed{S^2_{\overline{Y}|X} = 1.474406}$$

y el valor de $\chi^2_{gl,\ \alpha/2}$
donde $gl = n - 2$, se consulta la tabla en el Apéndice

$\chi^2_{8,0.025} = 17.53$, $\chi^2_{8,0.975} = 2.18$

Recuérdese que la distribución χ^2 no es simétrica; por tanto, los valores de χ^2 son diferentes.



Posteriormente, se sustituyen los valores, de lo cual resulta:

$$\frac{(n-2)\,S^2_{Y|X}}{\chi^2_{1-\alpha/2,gl}} < \sigma^2_{Y|X} < \frac{(n-2)\,S^2_{Y|X}}{\chi^2_{\alpha/2,gl}}$$

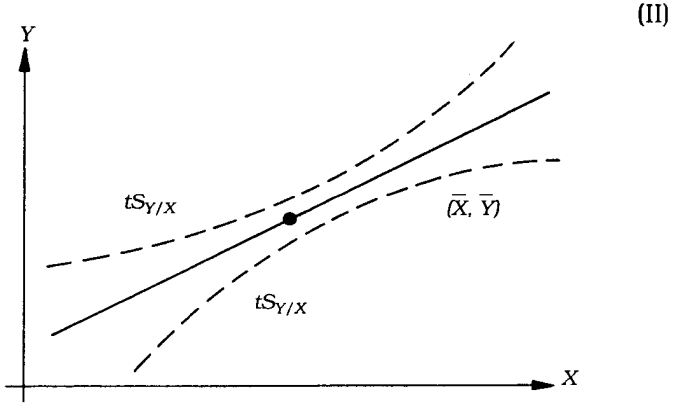
$$\frac{8\,(1.474406)}{17.53} < \sigma^2_{Y|X} < \frac{8\,(1.474406)}{2.18}$$

$$\frac{11.79525}{17.53} < \sigma^2_{Y|X} < \frac{11.79525}{2.18}$$

$$0.67286 < \sigma^2_{Y|X} < 5.41066$$

$$(0.67 < \sigma^2_{Y|X} < 5.41)$$

Una representación gráfica de los intervalos de confianza anteriores es como el de la media $\mu_{Y|X}$.



13.3.2 Pruebas de hipótesis para la regresión

El otro método para evaluar la ecuación de regresión es mediante el contraste de las hipótesis de los coeficientes de regresión α, β. En este libro se realizará la prueba correspondiente al coeficiente β, para lo cual se establece:

$H_0 : \beta = 0$
$H_1 : \beta \neq 0$ (H_0 falsa)
con un nivel de significancia α determinado.

Con este propósito se emplea el modelo de t de Student:

$$t = \frac{b}{S_b}$$

Se utilizan b y S_b del ejemplo anterior.

$$t_c = \frac{-3.22424}{0.42275}$$

$$\boxed{t_c = -7.62682}$$

$$|t_c| = 7.62682$$

Como la prueba de hipótesis es de dos colas se utiliza el valor absoluto de la t calculada.

La regla de decisión es: si $t_c \geq t_{gl, \frac{\alpha}{2}}$ entonces H_0 se rechaza para un $\alpha = 5\%$

$t_{8,\,0.025} = 2.306$, debido a que $gl = n - 2 = 10 - 2 = 8$ y que $\alpha/2 = 0.025$

Como $7.62682 > 2.306$, entonces H_0 se rechaza.[3]
Para un $\alpha = 1\%$, $t_{8,0.005} = 3.555$.
Como $7.62682 > 3.555$, entonces H_0 se rechaza a 1%.

Otra manera de contrastar estas hipótesis es mediante el análisis de varianza; para ello se necesitan: $SC_e = (n-2)\, S^2_{\hat{Y}|X}$, $SC_{reg} = (n-1)b^2\, S^2_X$
$SC_{total} = (n-1)\, S^2_Y$

donde

$$SC_{total} = SC_e + SC_{reg}$$

$$CM_{error} = CM_e = \frac{SC_e}{gl_{error}}$$

[3] Lo que significa que X, la variable independiente, contribuye a predecir a la variable dependiente Y.

donde

$gl_{error} = n - 2$
$gl_{reg} = 1$
$gl_{total} = n - 1$

$$CM_{reg} = \frac{SC_{reg}}{gl_{reg}}$$

$$\text{Finalmente, } F = \frac{CM_{reg}}{CM_e}$$

Para el ejemplo anterior resulta
$SC_{reg} = 9 \times (-3.2243)^2 \times (9.16667)$

$$\boxed{SC_{reg} = 857.67943}$$

$SC_e = 8 \times 14.74406$
$SC_e = 117.95248$
$SC_{total} = 9 \times 108.40$

$$\boxed{SC_{total} = 975.6000}$$

Como $SC_{total} = SCe + SC_{reg}$
975.6000 = 117.95248 + 857.67943
El proceso para calcular los cuadrados medios es el siguiente:

$$CM_{error} = \frac{SC_e}{gl_{error}} = \frac{117.95248}{8}; \text{ donde } CM_e = CM_{error}$$

$$\boxed{CM_{error} = 14.74406}$$

$$CM_{reg} = \frac{SC_{reg}}{gl_{reg}} = 857.67943$$

$$\boxed{CM_{reg} = 857.67943}$$

$$F = \frac{CM_{reg}}{CM_e} = \frac{857.67943}{14.74406}$$

$$\boxed{F = 58.17118}$$

$F = t^2 \Rightarrow t = 7.627,\ t^2 = 58.17113$

Análisis de varianza				
FV	gl	*SC*	*CM*	*F*
Regresión	1	$(n-1)b^2S_X^2$	$\left[\frac{(n-1)\ b^2S_X^2}{1}\right]$	CM_{reg}
Error	$n-2$	$(n-2)S_{\hat{Y}\mid X}^2$	$\left[\frac{(n-2)\ S_{\hat{Y}\mid X}^2}{n-2}\right]$	CM_{error}
Total	$n-1$	$(n-1)\ S_Y^2$		

Se sustituyen los valores calculados:

FV	*gl*	*SC*	*CM*	*F*
Regresión	1	857.67943	857.67943	58.17118[a]
Error	8	117.95248	14.74406	
Total	9	975.6000		

[a] Significativa al 1%. (*X* contribuye a predecir *Y*.)

13.4 ANÁLISIS DE CORRELACIÓN

El coeficiente de correlación es un parámetro que indica el grado de relación simultánea entre dos variables, aunque no existe distinción entre la variable independiente y la dependiente. La correlación mide la asociación, que no significa causalidad; en esto consiste la diferencia del análisis de correlación con respecto al de regresión, donde si una variable depende de otra en una muestra de sujetos promedio puede generalizarse este resultado. Para profundizar en el tema, debe distinguirse entre dos tipos de estudios: los *observacionales* y los *experimentales controlados*; en estos últimos, el investigador manipula la variable independiente y observa el efecto en la dependiente. Si se producen cambios, el investigador gradúa el rango, etapas y valores de la variable independiente. La asociación que resulta con la variable dependiente señala fuertemente una precisión.

En cambio, en un estudio observacional el investigador se limita a registrar los valores de las variables, debido a que considera a ambas independientes. Así, con respecto a un conjunto de sujetos (sin manipular los valores de variable alguna), la asociación entre las dos variables no implica causalidad; por lo general, en los estudios observacionales la asociación está en función de terceros

factores. En suma resulta muy importante clasificar de modo apropiado los estudios y tener precaución al inferir las conclusiones.

Un auxiliar para determinar la relevancia del coeficiente de regresión lo constituye el coeficiente de determinación; es decir, el cuadrado del coeficiente de correlación (r^2), que mide la proporción de la variación total en Y, considerada por la regresión de Y en X.

Hay dos métodos para calcular el coeficiente de correlación de Pearson. En el primero se utilizan los resultados del análisis de varianza y se emplea la siguiente fórmula:

$$r_{XY} = \sqrt{\frac{SC_{reg}}{SC_{total}}}$$

El segundo método conduce a la fórmula producto-momento por medio de determinantes.

Si se continúa con el ejemplo anterior, donde $SC_{reg} = 857.67943$ y $SC_{total} = 975.6000$

$$r_{XY} = \sqrt{\frac{857.67943}{975.6000}} = \sqrt{0.87913} = 0.93762$$

$$\boxed{r_{XY} = 0.93762} \text{ y } \boxed{r^2 = 0.87913}$$

Este método resulta muy sencillo de aplicar, pero no es sensible al signo de la correlación. Dicho signo puede deducirse con base en la gráfica de la recta de regresión, al considerar el signo de la pendiente de la recta.

Como el coeficiente $b = -3.2242$, entonces $\boxed{r_{XY} = -0.93762}$

El otro método por determinantes:

$$r_{XY} = \frac{\begin{vmatrix} n & \Sigma Y \\ \Sigma X & \Sigma XY \end{vmatrix}}{\sqrt{\begin{vmatrix} n & \Sigma X \\ \Sigma X & \Sigma X^2 \end{vmatrix} \begin{vmatrix} n & \Sigma Y \\ \Sigma Y & \Sigma Y^2 \end{vmatrix}}} = \frac{n\,\Sigma XY - \Sigma X\,\Sigma Y}{\sqrt{[n\,\Sigma X^2 - (\Sigma X)^2]\,[n\,\Sigma Y^2 - (\Sigma Y)^2]}} =$$

$$r_{XY} = \frac{10\,(713) - (55)\,(178)}{\sqrt{[10 \times 385 - (55)^2]\,[10 \times 4144 - (178)^2]}} = -0.93762$$

$$\boxed{r_{XY} = -0.93762}$$

13.4.1 Intervalo de confianza para el coeficiente de correlación

Para establecer la *confianza* del coeficiente de correlación muestral se tiene que señalar primero el intervalo de confianza para Z correspondiente a r. A fin de determinar lo anterior, es necesario calcular el error estándar de Z:

$$S_Z = \frac{1}{\sqrt{n-3}}$$

A continuación se especifica el nivel de confianza de Z: 5, 1%, etc. El coeficiente de correlación r se transforma en Z por medio de la tabla 5 del Apéndice. Una vez obtenido el intervalo de confianza para Z se transforma nuevamente a r, por medio de la misma tabla.

Ejemplo En una muestra de 103 sujetos se aplica un examen de conocimientos generales (X) y uno sobre coeficiente de inteligencia (Y). Al correlacionar las puntuaciones obtenidas, se encuentra que $r = 0.82$. Establecer el intervalo de confianza al 1%.

Paso 1. Se calcula el error S_Z con $n = 103$.

$$S_Z = \frac{1}{\sqrt{n-3}}$$

$$= \frac{1}{\sqrt{103-3}} = \frac{1}{\sqrt{100}} = 0.1$$

$$\therefore S_Z = 0.1$$

Paso 2. Mediante la tabla del Apéndice se transforma $r = 0.82$ en Z.

$$Z = 1.157$$

Paso 3. Por medio de la tabla de "área bajo la curva normal" se obtiene el nivel de significancia para $\alpha = 1\%$ de error y 99% de confianza.

$$Z_0 = \pm 2.58$$

Paso 4. Se sustituyen estos valores en la siguiente fórmula de intervalos de Z:

$$Z_\alpha = Z \pm Z_0 (S_Z)$$

de manera que

$$Z_{99} = 1.157 \pm 2.58(0.1)$$

$$Z'_{99} = 1.157 + 0.258(0.1) = 1.415$$

$$Z''_{99} = 1.157 - 0.258(0.1) = 0.899$$

Paso 5. Este intervalo calculado para Z se transforma por medio de la tabla del Apéndice en valores r, y resulta r_{99} = (0.715, 0.890).

13.4.2 Valor promedio del coeficiente de correlación

Algunas veces es necesario obtener la media aritmética de diversos coeficientes de correlación adjudicados a las mismas variables medidas varias veces en la misma muestra, o en muestras diferentes, pero procedentes de la misma población. Cuando la r es pequeña (± 0.30), no es conveniente considerar el promedio aritmético; cuando la r calculada es grande, se aplica un procedimiento para todas las muestras.

En relación con muestras del mismo tamaño, se transforma cada r a su valor Z. Posteriormente se utiliza la tabla del Apéndice para calcular la media aritmética de las Z y, por último, la Z promedio se restituye a r.

Ejemplo Los siguientes datos corresponden a cuatro coeficientes de correlación de cuatro muestras del mismo tamaño:

$$\begin{aligned} r_1 &= 0.90 \\ r_2 &= 0.87 \\ r_3 &= 0.77 \qquad n = 4 \\ r_4 &= 0.55 \end{aligned}$$

Paso 1. Se cambian estos valores a su correspondiente Z y luego se suman:

$$\begin{aligned} Z_1 &= 1.472 \\ Z_2 &= 1.333 \\ Z_3 &= 1.020 \\ Z_4 &= \underline{0.618} \\ & 4.443 \end{aligned}$$

$$\boxed{\Sigma Z = 4.443}$$

Paso 2. Se calcula la media de Z:

$$\bar{Z} = \frac{\Sigma Z}{n} = \frac{4.443}{4} = 1.110$$

Paso 3. El valor de la Z promedio ($\bar{Z}$) se localiza en la tabla del Apéndice y se cambia a su correspondiente r:

$$\boxed{r = 0.80}$$

Por otra parte, cuando las r se extraen de muestras de diferente tamaño el procedimiento es el mismo, excepto para calcular la Z. Ésta se obtiene al multiplicar cada Z por $(n - 3)$, donde n es el tamaño de la muestra respectiva. Como segundo paso, los productos se suman, y el total se divide entre $(n_1 - 3) + (n_2 - 3) + (n_3 - 3) + \ldots +$ (número total de muestras que intervienen). A continuación se prosigue como en el caso anteriormente presentado.

Como el coeficiente de correlación lineal obtiene valores entre -1 y $+1$, éstos pueden ubicarse en intervalos.

De ± 0.96 a ± 1.0 (correlación perfecta).
De ± 0.85 a ± 0.95 (correlación fuerte).
De ± 0.70 a ± 0.84 (correlación significativa).
De ± 0.50 a ± 0.69 (correlación moderada).
De ± 0.20 a ± 0.49 (correlación débil).
De ± 0.10 a ± 0.19 (correlación muy débil).
De ± 0.09 a ± 0.00 (correlación nula o inexistente).

13.4.3 Prueba de hipótesis para el coeficiente de correlación

Después de obtener r y de encontrar significativa la asociación entre las dos variables de interés, surgen preguntas muy importantes: ¿la correlación entre ambas variables también se presentará en la población?; o sea, ¿el grado de asociación será el mismo en la muestra que en la población? A fin de responder estas preguntas con una orientación estadística apropiada, deben formularse las *hipótesis nula* y la *alternativa*:

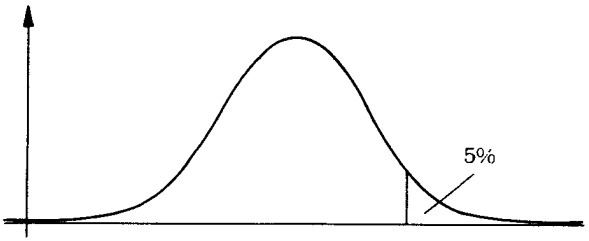
- Hipótesis nula (H_0): No existe correlación alguna en la población.
- Hipótesis alternativa (H_1): Sí hay correlación en la población.

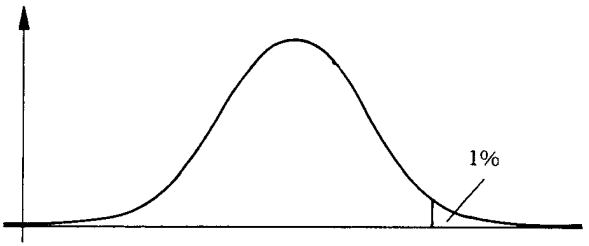
De acuerdo con la notación de estadística se tiene:

$$\text{Para } H_0 : \rho = 0$$
$$\text{Para } H_1 : \rho \neq 0$$

Al tratar de comprobar (para su aceptación o rechazo) la hipótesis nula, el investigador tiene la opción de escoger un error (a veces se establece de antemano, según las condiciones del problema, la magnitud del error). Esto significa que, si la hipótesis nula o alternativa se acepta o rechaza, existe un porcentaje de error que nos asegura lo contrario.

Los errores más comunes, que también suelen llamarse *niveles de significancia*, se denotan por α. Son de 5% y de 1%, o bien, de 0.05 y 0.01, respectivamente. Es posible presentar gráficamente estos valores de la siguiente manera:





Para aceptar o rechazar la hipótesis nula se emplea la razón t (*de Student*) definida como sigue:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

donde:
n = número de pares de valores de las variables X y Y
r = coeficiente de correlación lineal obtenido.

Una vez evaluada t, se compara con la $t_{\alpha/2,\, gl}$ que se obtiene en la tabla de valores de t, con grados de libertad $n-2$, y $\alpha/2$ nivel de significancia establecido de antemano. La t calculada debe ser mayor, o al menos igual, que la t crítica para poder rechazar la hipótesis nula (H_0). Esto se expresa de manera matemática con la siguiente regla de decisión: Si $t \geq t_{crit}$, entonces H_0 se rechaza.

Ejemplo Para el problema de la sección 13.2, se realizará la prueba de hipótesis que corresponde al coeficiente de correlación

$$r = -0.9376$$

$$n = 10$$

Se aplicará el modelo siguiente:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

donde $r^2 = 0.8791$ (coeficiente de determinación).
Se realizan las sustituciones respectivas.

$$t_c = \frac{-0.9376\sqrt{10-2}}{\sqrt{1-0.8791}}$$

$$t_c = \frac{-0.9376\sqrt{8}}{\sqrt{0.12090}}$$

$$t_c = \frac{-2.65193}{0.34771}$$

$$\boxed{t_c = -7.62685}$$

$$\boxed{|t_c| = 7.62685}$$

$t_{0.025,8} = 2.306$ como $7.62685 > 2.306$ y

$7.62685 > 3.355$

$t_{0.005,8} = 3.355$

Puede concluirse que H_0 se rechaza al 5 y 1%, respectivamente.

Este resultado es congruente con la prueba de hipótesis al coeficiente de regresión β ($t = \sqrt{F}$).

13.5 PRUEBA DE HIPÓTESIS ENTRE DOS RECTAS DE REGRESIÓN

Algunas veces se requiere conocer si existe una diferencia estadísticamente significativa entre dos rectas de regresión (lo que implica efectuar un contraste entre las dos pendientes o los coeficientes β de estas rectas).

Ejemplo A dos grupos de ratas de laboratorio se les proporcionan dietas diferentes durante diez semanas. Se desea saber si el peso inicial causa una respuesta lineal diferente en el peso ganado. El primer grupo lo constituyen ocho animales; el segundo diez.

Se calculan los coeficientes de regresión b_1 y b_2, a partir de los siguientes datos y se obtiene $b_1 = 1.64826$ y $b_2 = 1.11272$.

Grupo 1		Grupo 2	
X_1	Y_1	X_2	Y_2
63	157	72	143
74	156	51	105
62	114	81	140
69	123	70	115
55	128	55	102
58	108	74	124
48	108	61	117
54	107	67	104
		54	111
		62	121

Paso 1. Establecer las hipótesis.

$H_0 : \beta_1 = \beta_2$
$H_1 : \beta_1 \neq \beta_2$
con un $\alpha = 5\%$

Paso 2. Se aplica el siguiente modelo.

$$t = \frac{b_1 - b_2}{S_{b_1 - b_2}}$$

donde

$$S_{b_1 - b_2} = \sqrt{S_p^2 \left(\frac{1}{\sum_{i=1}^{8} (X_i - \overline{X}_1)^2} + \frac{1}{\sum_{i=1}^{10} (X_i - \overline{X}_2)^2} \right)}$$

y S_p^2 se obtiene aplicando la siguiente fórmula:

$$S_p^2 = \frac{\sum_{i=1}^{8} (Y_i - \overline{Y}_1)^2 - \frac{\left[\sum_{i=1}^{8} (X_i - \overline{X}_1)(Y_i - \overline{Y}_1) \right]^2}{\sum_{i=1}^{8} (X_i - \overline{X}_1)^2} + \sum_{i=1}^{10} (Y_i - \overline{Y}_2)^2 - \frac{\left[\sum_{i=1}^{10} (X_i - \overline{X}_2)(Y_i - \overline{Y}_2) \right]^2}{\sum_{i=1}^{10} (X_i - \overline{X}_2)^2}}{(n_1 - 2) + (n_2 - 2)}$$

Paso 3. Se efectúan las sustituciones de valores.

$$S_p^2 = \frac{3\,020.875 - \dfrac{(820.625)^2}{497.875} + 1\,833.600 - \dfrac{(952.600)^2}{856.100}}{(8-2)+(10-2)}$$

$$S_p^2 = \frac{1\,668.2757 + 773.6225}{6+8}$$

$$\boxed{S_p^2 = 174.4213}$$

$$S_{b_1 - b_2} = \sqrt{174.4213\left(\frac{1}{497.875} + \frac{1}{856.100}\right)}$$

$$S_{b_1 - b_2} = \sqrt{174.4213 \times 0.00317662}$$

$$S_{b_1 - b_2} = \sqrt{0.554070}$$

$$\boxed{S_{b_1 - b_2} = 0.7444}$$

Paso 4. Al sustituir en t se obtiene:

$$t = \frac{b_1 - b_2}{S_{b_1 - b_2}} = \frac{1.64826 - 1.11272}{0.7444}$$

$$t = \frac{0.5356}{0.7444}$$

$$\boxed{t_c = 0.72}$$

Paso 5. Regla de decisión, para $\alpha = 5\%$.

Si la t calculada es mayor o igual a t tablas, H_0 se rechaza.

$$t_{gl,\,\alpha/2} \quad \text{es} \quad t_{14,0.025} = 2.145$$

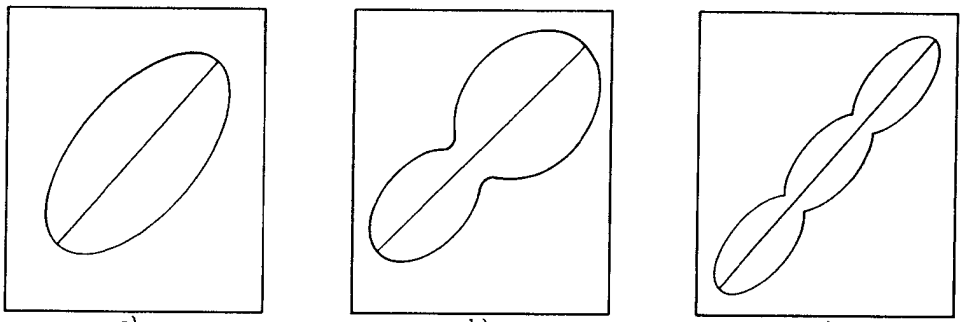
Paso 6.

Conclusión Como $0.72 < 2.145$, H_0 no se rechaza. Las pendientes de las rectas de regresión de los dos grupos son iguales; por consiguiente, *el peso inicial causa las mismas respuestas lineales.*

13.5.1 Comparación entre dos coeficientes de correlación

Para emplear correctamente el coeficiente de correlación lineal en una muestra, como medida de asociación entre las variables X y Y, deben tomarse en cuenta los siguientes aspectos:

1. Es absolutamente necesario que exista una relación lineal entre las variables X y Y. Dicha relación se confirma si, al trazar un diagrama de dispersión, todos los puntos se agrupan en torno a la recta obtenida, y unen los puntos 0, 0 y $\overline{X}$, $\overline{Y}$.
2. Todos los datos deben ser *homoscedásticos*. Esto significa que la varianza de una variable estadística es la misma para todos los valores fijos de la otra variable estadística; o sea, que debe existir la misma dispersión de los puntos respecto a la línea de regresión.



a) b) c)

Lo anterior se representa en los siguientes diagramas, donde sólo en el caso (*a*) existe homoscedasticidad.

3. Puntuaciones intervalares. Las variables X y Y deben ser intervalares (es decir, deben medir exactamente la distancia entre las puntuaciones que las representan, así como el orden en el cual se encuentran). Estas puntuaciones pueden representarse con números reales.
4. La muestra debe obtenerse en forma aleatoria. Todos los elementos de la población tendrán la misma probabilidad de ser escogidos para constituir la muestra.
5. La prueba de validez del coeficiente de correlación de Pearson requiere que tanto la variable X como la Y estén distribuidas normalmente en la población. En muestras pequeñas que no cumplan este requisito carecerá de validez el coeficiente r obtenido. Sin embargo, este aspecto no tendrá importancia cuando la muestra sea lo suficientemente grande.

En caso de utilizar el coeficiente de correlación como una medida de asociación, hay que suprimir la posibilidad de que una variable "emboscada"

afecte a las variables en estudio, ya que esta tercera variable causal desconocida originaría una correlación ficticia. Algunas veces corresponde más al sentido común que a la explicación estadística determinar si una correlación significante, tiene interpretación práctica, o si es una correlación espuria.

Es necesario realizar el análisis de regresión correspondiente a los dos grupos del problema anterior. Las ecuaciones de sus rectas de regresión son las siguientes:

$$\hat{Y}_1 = 26.71452 + 1.6483X$$
$$\hat{Y}_2 = 47.053 + 1.1127X$$

Los coeficientes de correlación para dichos grupos son:

$$r_1 = 0.6691 \quad r_2 = 0.7603$$

Una vez obtenidos, se efectuará la prueba de significancia de la diferencia entre tales coeficientes de correlación. Así, se comprobará la hipótesis nula que plantea la no existencia de diferencias estadísticamente significativas entre estos dos coeficientes de correlación; en otras palabras, ambos valores de r se basan en muestras provenientes de la misma población. Para aceptar o rechazar tal afirmación cada coeficiente r se transformará en una puntuación Z, y se aplicará la prueba de significancia a dichas puntuaciones (es decir, la prueba Fisher de las Z). La fórmula respectiva es la siguiente:

$$Z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

donde:

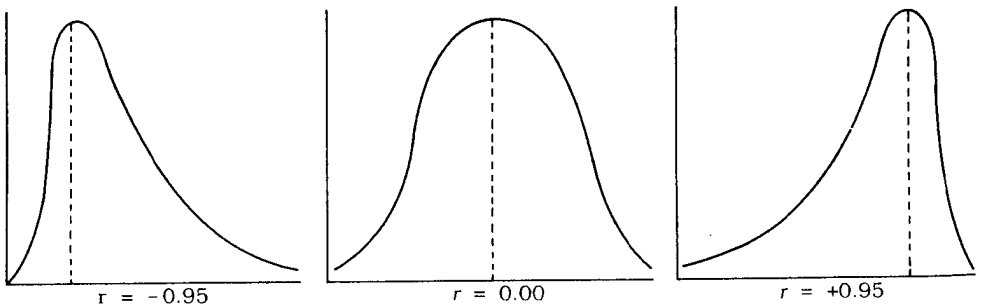
n_1 y n_2 = tamaños de las muestras, respectivamente.

Z_1 y Z_2 = puntuaciones Z provenientes de los coeficientes de correlación r_1 y r_2, respectivamente.

El valor 3 constituye una constante.

Después de aplicada la prueba y obtenidas las conclusiones correspondientes, los resultados se referirán a los coeficientes de correlación r.

Es necesario efectuar la transformación de r a Z puesto que no es normal la distribución muestral de los coeficientes de correlación. Al respecto destaca que



cuando los valores se aproximan a cero, la distribución tiende al modelo normal, y que al alejarse de cero se incrementa la asimetría en la distribución muestral de dichos coeficientes.

Ejemplo Transforme los siguientes datos en puntuaciones Z y aplique la prueba correspondiente. A continuación, refiera los resultados a los coeficientes r y compruebe la hipótesis propuesta.

Muestra I: $r_1 = 0.6691$; $n_1 = 8$
Muestra II: $r_2 = 0.7603$; $n_2 = 10$

Los valores de la Z de Fisher correspondientes a r pueden leerse directamente en la tabla del Apéndice. Los resultados correspondientes son:

$$r_1 \Rightarrow Z_1 = 0.811$$
$$r_2 \Rightarrow Z_2 = 0.996$$

Se utilizará la siguiente fórmula para comprobar la diferencia entre las dos muestras:

$$Z = \frac{Z_1 - Z_2}{\sqrt{\dfrac{1}{n_1 - 3} + \dfrac{1}{n_2 - 3}}}$$

Esta fórmula se basa en que el error estándar de Z es:

$$S_Z = \frac{1}{\sqrt{n-3}}; \text{ o sea, } S_D = \sqrt{S_{Z_1}^2 + S_{Z_2}^2}$$

Se realizan sustituciones en la fórmula:

$$Z = \frac{0.811 - 0.996}{\sqrt{\dfrac{1}{5} + \dfrac{1}{7}}} = \frac{-0.185}{\sqrt{0.2 + 0.14286}}$$

$$Z = \frac{-0.185}{0.58554}$$

$$\boxed{Z = -0.31595} \Rightarrow \boxed{|Z| = 0.31595}$$

Como $Z_{0.025} = 1.96$ y como $0.31595 < 1.96$, H_0 no se rechaza.

Este resultado es congruente con el obtenido al contrastar los coeficientes de regresión de dichas rectas.

13.6 ANÁLISIS DE REGRESIÓN MÚLTIPLE

En ciertas investigaciones, estudios o experimentos no es posible utilizar el modelo de regresión lineal simple debido a que existe la posibilidad de dos o más variables independientes; cuando éstas varían de acuerdo con algún patrón o norma, la regresión múltiple es útil para establecer el efecto de una variable independiente, en tanto que sus similares se mantienen constantes.

Una compañía de seguros lleva a cabo un estudio con el objeto de conocer la satisfacción de sus asegurados en el ramo de automóviles. Las variables que pueden intervenir son:

- Satisfacción del asegurado (*variable dependiente*).
- Prontitud de la emisión de la póliza (*variable independiente*).
- Rapidez en caso de siniestro (*variable independiente*).
- Atención del representante de la compañía (*variable independiente*).
- Lugar y hora donde ocurre el siniestro (*variable independiente*).

13.6.1 Caso particular para dos variables independientes

En este apartado se desarrollará el modelo de regresión múltiple para dos variables independientes; cuando en un estudio intervengan más de dos variables independientes se sugiere utilizar un paquete estadístico de cómputo (SYSTAT, spss, etcétera).

Ejemplo En el caso de la compañía aseguradora antes indicado, las variables por considerar son las siguientes:

X_1 = prontitud en la emisión de la póliza.

X_2 = rapidez en caso de siniestro (el tiempo que ocupa el ajustador en llegar al lugar de los hechos).

Y = satisfacción del asegurado.

Así, X_1, X_2 serán las variables independientes (pueden adoptar cualquier valor) y Y es la variable dependiente (está en función de los valores que tomen X_1, X_2).

El modelo lineal que da respuesta a Y de un conjunto de k variables independientes es de la forma:

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Si el conjunto de los datos observados considera toda la población, los coeficientes del modelo lineal múltiple serán β_0, β_1, β_2 y ε es el error o residuo.

En la mayoría de los estudios o investigaciones se realizan muestreos; para este caso el modelo se expresa como sigue:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + e$$

La compañía aseguradora realiza distintas encuestas, en las cuales considera las variables X_1 y X_2, cuyo rango muestra entre 0 y 4.

El asegurado calificó cada variable en el intervalo anterior, en el que cero es la calificación mínima y cuatro la máxima. Con respecto a la variable dependiente, el intervalo abarcó de 0 a 7, donde 0 es la mínima y 7 la máxima calificación; este número también fue asignado por el asegurado.

Es primordial señalar que se observará de manera independiente cada variable, incluida Y. Las encuestas aportaron los datos siguientes:

Y	X_1	X_2
2	1	0
3	1	1
4	1	1
5	2	2
7	2	3
7	1	2
4	3	1
5	2	2
7	3	4
6	4	4

Ejemplo[4]

Una consultoría en estudios de mercado desea estimar los gastos en alimentación con respecto al ingreso de la familia y al tamaño de la misma.

Y = gasto de la alimentación por mes.
X_1 = ingreso mensual.
X_2 = tamaño de la familia (número de hijos).

Los resultados de la muestra fueron los siguientes:

n	Y	X_1[a]	X_2
1	1	4	20
2	4	10	6
3	5	15	2
4	4	12	8
5	3	8	9
6	4	16	8
7	2	5	12
8	1	7	15
9	4	9	10
10	2	10	10

[a] X_1 = Se expresa en miles de pesos.

Con los datos de la tabla se formará el siguiente sistema de ecuaciones:

[4] Este ejemplo debe desarrollarlo el estudiante.

$$\begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 \\ \Sigma X_2 & \Sigma X_1 X_2 & \Sigma X_2^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \end{bmatrix} \text{ o sea } A^{-1}b = Y$$

A continuación se realizan los cálculos para llegar a la expresión anterior.

	Y	X_1	X_2	X_1^2	X_1X_2	X_2^2	X_1Y	X_2Y	
	2	1	0	1	0	0	2	0	
	3	1	1	1	1	1	3	3	
	4	1	1	1	1	1	4	4	
	5	2	2	4	4	4	10	10	
	7	2	3	4	6	9	14	21	
	7	1	2	1	2	4	7	14	
	4	3	1	9	3	1	12	4	
	5	2	2	4	4	4	10	10	
	7	3	4	9	12	16	21	28	
	6	4	4	16	16	16	24	24	
Σ	50	20	20	50	49	56	107	118	Sumas
	ΣY	ΣX_1	ΣX_2	ΣX_1^2	ΣX_1X_2	ΣX_2^2	ΣX_1Y	ΣX_2Y	

Una vez calculadas las sumas, se efectúan las sustituciones en el sistema.

$$\begin{bmatrix} 10 & 20 & 20 \\ 20 & 50 & 49 \\ 20 & 49 & 56 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 50 \\ 107 \\ 118 \end{bmatrix}$$

El sistema de ecuaciones se resuelve por medio de determinantes.[5]

$$\Delta = \begin{vmatrix} 10 & 20 & 20 \\ 20 & 50 & 49 \\ 20 & 49 & 56 \end{vmatrix} \quad \text{obteniendo sus menores:}$$

$$\Delta = 10 \begin{vmatrix} 50 & 49 \\ 49 & 56 \end{vmatrix} - 20 \begin{vmatrix} 20 & 49 \\ 20 & 56 \end{vmatrix} + 20 \begin{vmatrix} 20 & 50 \\ 20 & 49 \end{vmatrix}$$

$$\Delta = 790$$

$$\Delta b_0 = \begin{vmatrix} 50 & 20 & 20 \\ 107 & 50 & 49 \\ 118 & 49 & 56 \end{vmatrix} \quad \text{por lo que}$$

[5] Véase el capítulo 3.

$$\Delta b_0 = 50 \begin{vmatrix} 50 & 49 \\ 49 & 56 \end{vmatrix} - 20 \begin{vmatrix} 107 & 49 \\ 118 & 56 \end{vmatrix} + 20 \begin{vmatrix} 107 & 50 \\ 118 & 49 \end{vmatrix}$$

$\Delta b_0 = 50(399) - 20(210) + 20(-657) = 19\,950 - 4\,200 - 13\,140$
$\Delta b_0 = 2\,610$

$$\Delta b_1 = \begin{vmatrix} 10 & 50 & 20 \\ 20 & 107 & 49 \\ 20 & 118 & 56 \end{vmatrix} \quad \text{Entonces}$$

$$\Delta b_1 = 10 \begin{vmatrix} 107 & 49 \\ 118 & 56 \end{vmatrix} - 50 \begin{vmatrix} 20 & 49 \\ 20 & 56 \end{vmatrix} + 20 \begin{vmatrix} 20 & 117 \\ 20 & 118 \end{vmatrix}$$

$\Delta b_1 = 10\,(210) - 50\,(140) + 20\,(220) = 2\,100 - 7\,000 + 4\,400$
$\Delta b_1 = -500$

$$\Delta b_2 = \begin{vmatrix} 10 & 20 & 50 \\ 20 & 50 & 107 \\ 20 & 49 & 118 \end{vmatrix} \Rightarrow$$

$$\Delta b_2 = 10 \begin{vmatrix} 50 & 107 \\ 49 & 118 \end{vmatrix} - 20 \begin{vmatrix} 20 & 107 \\ 20 & 118 \end{vmatrix} + 50 \begin{vmatrix} 20 & 50 \\ 20 & 49 \end{vmatrix}$$

$\Delta b_2 = 10(657) - 20\,(220) + 50\,(-20) = 6\,570 - 4\,400 - 1\,000$
$\Delta b_2 = 1\,170$

Los coeficientes del modelo de regresión múltiple son los siguientes:

$$b_0 = \frac{2\,610}{790} = 3.304 \qquad b_1 = \frac{-500}{790} = -0.633 \qquad b_2 = \frac{1\,170}{790} = 1.481$$

Se sustituyen coeficientes en $\hat{Y}$:

$$\boxed{\hat{Y} = 3.304 - 0.633\,X_1 + 1.481\,X_2}$$

Para el ejemplo desarrollado paralelamente por el estudiante, la respuesta es:

$$\hat{Y} = 4.191 + 0.086\,X_1 - 0.202\,X_2$$

13.6.2 Relación entre el análisis de varianza y el de regresión múltiple

En esta sección se analizará la mejora introducida por el ajuste de la recta de regresión y se comprobará si ésta resulta significativa; con este fin, realizará un análisis de varianza a partir de los datos del modelo de regresión obtenido.

El primer cuadro contendrá las siguientes fórmulas:

$$SC_e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_1^2) = \text{suma de cuadrados del error.}$$

$$SC_R = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \text{regresión de la suma de cuadrados.}$$

$$SC_{tot} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \text{suma total de cuadrados.}$$

Las tres columnas del cuadro serán los Y, X_1, X_2 observados; la cuarta se formará con $\hat{Y}$, que es el resultado de sustituir el valor de X_1 y X_2 en la ecuación estimada.

Observación	X_0	X_1	X_2	Y	$\hat{Y}$	SC_e	SC_R	SC_{tot}	$X^TA^{-1}X$
1	1	1	0	2	2.671	0.450	5.424	9.0000	0.354
2	1	1	1	3	4.152	1.327	0.719	4.000	0.202
3	1	1	1	4	4.152	0.023	0.719	1.000	0.202
4	1	2	2	5	5.000	0.000	0.000	0.000	0.098
5	1	2	3	7	6.481	0.269	2.193	4.000	0.222
6	1	1	2	7	5.633	1.869	0.401	4.000	0.302
7	1	3	1	4	2.886	1.241	4.469	1.000	0.654
8	1	2	2	5	5.000	0.000	0.000	0.000	0.098
9	1	3	4	7	7.329	0.108	5.424	4.000	0.342
10	1	4	4	6	6.696	0.484	2.876	1.000	0.490
				$\overline{Y} = 5$	Suma	5 772	22 226	28 000	

Una vez que se calcularon $\hat{Y}$, SC_e, SC_R y SC_{tot} se realiza el análisis de varianza.

Fuente	*Suma de cuadrados*	*Grados de libertad*	*Cuadrados medios*	*Razón F*
Regresión	SC_R	k	SC_R/k	$\dfrac{SC_R/k}{SC_e/(n-k-1)}$
Residual	SC_e	$n-k-1$	$SC_{e/(n-k-1)}$	

La razón de F es el estadístico de prueba para las siguientes hipótesis.
$H_0 : \beta_1 = \beta_2 = 0$
H_1 : Por lo menos una β es diferente de cero.

Se sustituyen los valores de ejemplo en la tabla anterior.

Fuente	Suma de cuadrados	*Grados de libertad*	*Cuadrados medios*	*Razón F*
Regresión	22.226	2	11.113	13.478
Residual	5.772	7	0.825[a]	
[a] $CM_e = S^2$				

Como se mencionó al principio de la sección, esta prueba es un instrumento para determinar si el modelo de regresión es significativo o no.

Si H_0 se rechaza, puede concluirse que el modelo lineal de regresión es significativo.

Si no existe suficiente evidencia para rechazar H_0, hay que concluir que el modelo de regresión estimado no es significativo.

13.6.3 Evaluación de la ecuación de regresión múltiple

En esta sección se considerará primero el intervalo de confianza para la respuesta media. A continuación, se estudiará el intervalo para cada coeficiente del modelo. Respecto al primer tema, la fórmula es la siguiente:

dado: $\mu_{Y|X_1, X_2 \ldots X_k}$, el IC es:

$$\overline{Y} - t_{\alpha/2,\, gl}\ S\sqrt{X^T A^{-1} X} < \mu_{Y|X_1,\, X_2,\, \cdots X_k} < \overline{Y} + t_{\alpha/2,\, gl}\, S\sqrt{X^T A^{-1} X}$$

donde $t_{\alpha/2}$ es un valor de la distribución t con $n - k - 1$ grados de libertad.

X^T es la transpuesta de la matriz X.

Ejemplo Construya un intervalo de confianza al 95% en $X^T = [1, 1, 0]$.

$$\Rightarrow X = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

El punto se sustituye en la ecuación estimada, y se obtiene:

$$\hat{Y} = 3.304\ (1) - 0.633\ (1) + 1.481\ (0) = 2.67$$

Los grados de libertad son $gl = n - k - 1 = 7 = 10 - 2 - 1$ y $\alpha/2 = 0.025$; por consiguiente,

$$t_{\alpha/2} = 2.365$$

$$S^2 = 0.825$$

$$S = 0.9083$$

Del sistema de ecuaciones que se obtuvo en la sección anterior resulta.

$$AX = b \Leftrightarrow A^{-1}b = Y$$

Para construir el intervalo de confianza es necesario calcular la matriz inversa de A; es decir, A^{-1}.

$$A^{-1} = \begin{pmatrix} 0.506 & -0.177 & -0.025 \\ -0.177 & 0.202 & -0.114 \\ -0.025 & -0.114 & 0.126 \end{pmatrix}$$

Posteriormente se realiza la operación:

$$X^T A^{-1} X = (1,\ 1,\ 0) \begin{pmatrix} 0.506 & -0.177 & -0.025 \\ -0.177 & 0.202 & -0.114 \\ -0.025 & -0.114 & 0.126 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 0.354$$

A continuación se sustituyen los valores en el intervalo de confianza:

$$5 - 2.365\ (0.9083)\ \sqrt{0.354} < \mu_{Y|X} < 5 + 2.365\ (0.9083)\ \sqrt{0.354}$$
$$3.722 < \mu_{Y|X} < 6.278$$

Es primordial señalar que el intervalo anterior se obtuvo a partir de un punto X. A continuación se muestran los intervalos de $\mu_{Y|X}$ para cada una de las observaciones.

Observación	*Intervalo de confianza*
1	$3.722 < \mu_{Y\|X} < 6.278$
2	$4.035 < \mu_{Y\|X} < 5.965$
3	$4.035 < \mu_{Y\|X} < 5.965$
4	$4.328 < \mu_{Y\|X} < 5.672$
5	$3.988 < \mu_{Y\|X} < 6.012$
6	$3.820 < \mu_{Y\|X} < 6.180$
7	$3.263 < \mu_{Y\|X} < 6.737$
8	$4.328 < \mu_{Y\|X} < 5.672$
9	$3.744 < \mu_{Y\|X} < 6.256$
10	$3.496 < \mu_{Y\|X} < 6.504$

En segundo término, se estudiarán los intervalos de confianza para cada uno de los coeficientes del modelo de regresión (intervalos al 95%).

Para b_0 se tiene:

$$\left(b_0 - t_{\alpha/2,gl}\, S\, \sqrt{c_{11}}\, \sqrt{\frac{n}{n-k}} < \beta_0 < b_0 + t_{\alpha/2,gl}\, S\sqrt{c_{11}}\, \sqrt{\frac{n}{n-k}}\right)$$

donde

c_{11} = el primer elemento de la diagonal de la matriz inversa A^{-1}.

$gl = n - k - 1 = 7$

Al sustituir los valores del ejemplo se obtiene:

$$3.304 - 2.365\,(0.9083)\,\sqrt{0.506}\,\sqrt{\frac{10}{10-2}} < \beta_0 < 3.304 + 2.365\,(0.9083)$$

$$\sqrt{0.506}\,\sqrt{\frac{10}{10-2}}$$

$$(3.304 - 1.7086 < \beta_0 < 3.304 + 1.7086)$$

$$\boxed{(1.5954 < \beta_0 < 5.0126)}$$

Para b_1 se tiene:

$$\left(b_1 - t_{\alpha/2,\,gl}\, S\sqrt{c_{22}}\, \sqrt{\frac{n}{n-k}} < \beta_1 < b_1 + t_{\alpha/2,\,gl}\, S\sqrt{c_{22}}\, \sqrt{\frac{n}{n-k}}\right)$$

donde c_{22} es el segundo elemento de la diagonal de la matriz inversa A^{-1}.

$gl = n - k - 1 = 10 - 2 - 1 = 7$

Se realizan las correspondientes sustituciones de valores:

$$-0.633 - 2.365\,(0.9083)\,\sqrt{0.202}\,\sqrt{\frac{10}{10-2}} < \beta_1 < -0.633 + 2.365\,(0.9083)$$

$$\sqrt{0.202}\,\sqrt{\frac{10}{10-2}}$$

$$\boxed{(-1.714 < \beta_1 < 0.448)}$$

Para b_2 se tiene:

$$\left(b_2 - t_{\alpha/2,\,gl}\, S\sqrt{c_{33}}\sqrt{\frac{n}{n-k}} < \beta_2 < b_2 + t_{\alpha/2,\,gl}\, S\sqrt{c_{33}}\sqrt{\frac{n}{n-k}}\right)$$

donde c_{33} = el tercer elemento de la diagonal de la matriz inversa A^{-1}.

$$1.481 - 2.365(0.9083)\sqrt{0.126}\sqrt{\frac{10}{10-2}} < b_2 < 1.481 + 2.365\,(0.9083)$$

$$\sqrt{0.126}\sqrt{\frac{10}{10-2}}$$

$$(1.481 - 0.853 < \beta_2 < 1.481 + 0.853)$$

$$\boxed{(0.628 < \beta_2 < 2.334)}$$

13.6.4 Prueba de hipótesis para los coeficientes de regresión

Para el coeficiente β_0, las hipótesis son las siguientes:

H_0: $\beta_0 = 0$
H_1: $\beta_0 \neq 0$

Se sustituyen los valores en el estadístico de prueba:

$$t = \frac{b_0}{S\sqrt{c_{11}}}, \quad t = \frac{3.304}{0.9083 \times \sqrt{0.506}}$$

$$t = \frac{3.304}{0.64611} = 5.11370$$

$$\boxed{t_c = 5.11370}$$

Se desea efectuar la prueba de hipótesis con una confiabilidad de 95%; es decir:

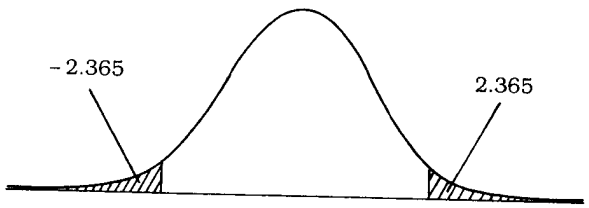
$$\alpha = 5\%$$

Se busca el valor de t en tablas con $gl = n - k - 1 = 7$, y se obtiene:

$$t_{0.025,7} = 2.365$$

La región de rechazo es:

(I)



Se concluye que b_0 es significativa, ya que 5.11370 > 2.365 con un α = 5%.

Para β_1, se plantea la siguiente hipótesis:

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$

Se sustituyen los valores en el estadístico de prueba:

$$t = \frac{b_1}{S\sqrt{c_{22}}}, \quad t_c = -\frac{0.633}{0.9083\sqrt{0.202}}$$

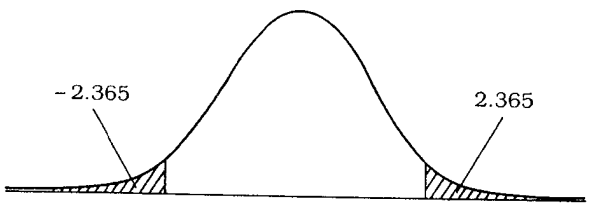
$t = -1.5506$
$|t| = 1.5506$

Debido a que se desea un nivel de significancia de 5%; es decir, una confiabilidad de 95%, nuevamente se busca el valor de t en tablas con un α = 5% y $gl = n - k - 1 = 7$. El valor es el siguiente:

$$t_{7,0.025} = 2.365$$

Se identifica la región de rechazo mediante una gráfica.

(II)



Como 1.5506 < 2.365, H_0 no se rechaza, es decir, b_1 no es significativo.

Por último, para β_2 se tienen las siguientes hipótesis:

H_0: $\beta_2 = 0$

H_1: $\beta_2 \neq 0$

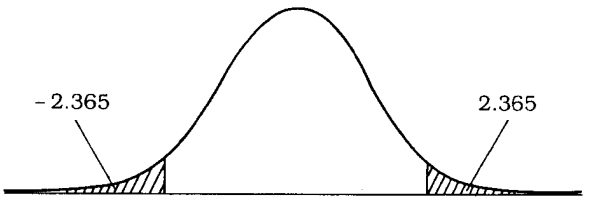
Igual que en los casos anteriores, se sustituyen los valores en el estadístico de prueba.

$$t = \frac{b_2}{S\sqrt{c_{33}}}$$

$$t = \frac{1.481}{0.9083\sqrt{0.126}}$$

$$t_c = 4.5935$$

Se presenta el mismo nivel de significancia que en los casos anteriores; por ello el valor de t es $t_{0.025,7} = 2.365$. Posteriormente, se identifica en la gráfica la región de rechazo para H_0.



Como 4.5935 > 2.365, H_0 se rechaza y resulta posible afirmar que b_2 es significativa.

13.6.5 Coeficiente de correlación múltiple

A continuación se explicará solamente el cálculo de r y r^2, ya que se consideró su interpretación en la sección sobre regresión simple. Con este fin, se utilizarán los resultados obtenidos en la sección 15.6.2, referidos al ejemplo de la aseguradora. La fórmula es la siguiente:

$$r^2 = \frac{SC_R}{SC_{tot}}$$

Al sustituir los valores del ejemplo, se obtiene:

$$r^2 = \frac{22.226}{28}$$

$$\boxed{r^2 = 0.794}$$

$$\boxed{r = 0.891}$$

RESUMEN

En este capítulo se ha analizado con gran detalle tanto el análisis de regresión lineal simple bivariado como el múltiple, debido a que son herramientas muy poderosas para cualquier tipo de investigación.

Una vez establecidos los conceptos fundamentales, el estudiante podrá utilizar el paquete estadístico MACSTAT para la solución de problemas; si bien dicho paquete en esta versión no cuenta con el módulo de regresión múltiple, es posible emplearlo para obtener otros dos tipos de regresión: la *exponencial* y la *logarítmica*.

PROBLEMAS

13.1 Un investigador en psicología experimental lleva a cabo un experimento con 10 animales de laboratorio (de la misma edad y peso). Les administra un fármaco, midiendo el tiempo en horas (X) y la concentración del fármaco en mg (Y). Se obtienen los siguientes datos:

X	Y
0.5	0.42
0.5	0.45
1.0	0.35
1.0	0.33
2.0	0.25
2.0	0.22
3.0	0.20
3.0	0.20
4.0	0.15
4.0	0.17

Realice un análisis de regresión lineal.

13.2 En un proceso de control de calidad se obtiene una muestra aleatoria de ocho frascos de un reactivo (en mg) y su correspondiente grado de pureza.

X	Y
85	2.3
65	1.2
73	1.5
90	1.9
82	1.8
80	2.0
68	1.3
88	2.1

Realice un análisis de regresión lineal.

13.3 A 10 candidatos a ingresar en un programa de doctorado en psicología, se les aplica en la Facultad de Psicología una prueba de personalidad (X) y un examen general de conocimientos (Y), con las siguientes puntuaciones:

Candidato	X	Y
A	2.96	529
B	2.46	506
C	3.36	591
D	3.40	610
E	2.43	474
F	2.12	509
G	2.85	550
H	3.12	600
I	3.20	575
J	2.75	540

Realice un análisis de regresión lineal

13.4 Se realiza un experimento para determinar la concentración de una droga determinada (X) en una solución, y se lee la fluorescencia producida (Y) al agregar ácido bórico. Se obtienen los siguientes resultados:

X	Y
0.0	1
0.25	3
0.50	6
1.0	10
2.0	15
4.0	28
6.0	43
8.0	52
10.0	70

Realice un análisis de regresión lineal.

13.5 Un investigador mide a 16 sujetos el grado de afectividad de un color (Y) relacionado con un estímulo visual (X) y su tiempo de reacción. Se registraron los siguientes datos:

X	Y
2.31	83
1.75	91
1.69	91
1.36	87
1.29	93
1.13	93
1.02	96
0.29	107
0.74	100
0.73	104
0.62	97
0.56	118
0.40	113
0.39	122
0.34	107
0.00	116

a) Realice un análisis de regresión lineal.
b) Realice un análisis de regresión exponencial.

13.6 A 21 estudiantes se les aplican tres pruebas psicológicas (X_1, X_2, X_3), de las que surgieron los siguientes datos:

Puntuación de actitud ante la agresión

Y	X_1	X_2	X_3
48	22	38	15
48	19	38	15
47	20	37	20
46	20	37	17
46	17	35	19
43	21	34	15
42	21	34	14
42	19	33	20
41	17	33	13
40	15	32	15
39	18	32	12

Puntuación de actitud ante la agresión

Y	X_1	X_2	X_3
32	11	25	15
31	17	25	9
30	16	23	9
29	15	22	13
29	15	21	9
28	16	20	11
27	16	18	11
27	13	17	10
27	12	15	8
26	12	15	8

a) Realice un análisis de regresión lineal de: Y con X_1
Y con X_2
Y con X_3

b) Realice un análisis de regresión múltiple.

c) Concluya.

13.7 Se aplicó a un grupo de 18 adolescentes sordomudos la prueba de inteligencia de Wechsler para adultos (Wais) y cuatro subtest del DAT; las puntuaciones de ambas aplicaciones son las siguientes:

	Wais	*Razonamiento mecánico*	*Razonamiento abstracto*	*Relaciones espaciales*	*Velocidad y exactitud*
	Y	X_1	X_2	X_3	X_4
1	99	15	33	16	25
2	103	24	20	20	40
3	111	17	37	13	21
4	116	5	42	8	20
5	127	9	40	6	18
6	117	5	48	3	17
7	114	14	34	7	31
8	113	13	33	13	35
9	122	24	20	16	35
10	113	15	32	12	27
11	120	14	43	10	17
12	108	12	36	11	28
13	116	20	31	12	29
14	106	20	19	19	37
15	100	32	22	18	29
16	96	25	21	16	39
17	90	20	31	19	28
18	97	23	16	21	38

a) Realice un análisis de regresión lineal de Y con cada uno de los cuatro subtest del DAT.
b) Realice un análisis de regresión múltiple.
c) Concluya

13.8 Se realiza un estudio con 16 atletas de alto rendimiento, ocho de potencia y ocho de resistencia; se les mide la extensión isométrica de la rodilla (X_1) (en kilogramos fuerza) y el porcentaje de la contracción muscular (X_2). Se obtienen con los siguientes datos; donde X_2 es la variable dependiente (Y).

Atletas de potencia		*Atletas de resistencia*	
X_1	X_2	X_1	X_2
196.0	56.0	161.0	98.4
183.0	28.8	142.0	70.8
295.0	57.2	122.5	35.4
203.0	46.0	123.0	74.5
195.0	35.5	176.0	79.5
289.0	58.6	156.0	62.1
198.0	41.4	126.0	74.3
206.9	21.6	95.0	67.7

Realice un análisis de regresión lineal para cada grupo de atletas y contraste las dos rectas de regresión y sus coeficientes de correlación.

13.9 En una empresa se analizan las ventas (Y), (X_1) los gastos en publicidad por TV y (X_2) los gastos en publicidad por el periódico durante cuatro semanas (todo en miles de pesos).

Realice una regresión múltiple con los siguientes datos:

X_1	X_2	Y
4	1	7
7	2	12
9	5	17
12	8	20

13.10 Se realiza un estudio con 10 estudiantes de posgrado en una materia determinada, donde X_1 es el número de problemas resueltos correctamente en una clase, X_2 son las puntuaciones obtenidas al aplicarles una prueba psicológica que mide autoestima y Y es el número de problemas que cada estudiante espera resolver correctamente en el examen final. Con los datos que se presentan a continuación realice un análisis de regresión múltiple.

Estudiante	X_1 *problemas resueltos en clase*	X_2 *puntuación de autoestima*	Y *problemas que espera resolver en el examen*
1	14	6	14
2	8	15	5
3	9	19	8
4	13	33	11
5	10	39	15
6	11	38	14
7	14	74	18
8	15	74	19
9	11	11	7
10	16	78	17

14

Análisis de frecuencia y estadística no paramétrica

PROPÓSITOS

El propósito central del presente capítulo consiste en que el lector explique el análisis de frecuencia y la bondad de ajuste, así como que identifique y aplique adecuadamente los modelos estadísticos no paramétricos.

De igual modo, el lector podrá:

1. Explicar en qué consiste y cuándo utilizar la prueba de *ji* cuadrada (χ^2) en su faceta de independencia, proporción (homogeneidad) y bondad de ajuste.
2. Explicar en qué consisten las diferencias entre las pruebas paramétricas y las no paramétricas.
3. Repasar cada una de las pruebas estadísticas paramétricas y las no paramétricas.
4. Diferenciar cada uno de los modelos estadísticos propuestos en este capítulo.
5. Utilizar adecuadamente cada uno de los modelos estadísticos no paramétricos considerados en este capítulo.
6. Aplicar como una opción a la correlación lineal paramétrica el modelo no paramétrico adecuado.
7. Aplicar, en las condiciones pertinentes, la prueba de Kolmogorov-Smirnov de una muestra, para la bondad de ajuste.
8. Generalizar los procedimientos de cálculo de las pruebas antes citadas, considerando sus aplicaciones y requisitos de uso.
9. Utilizar cada uno de dichos modelos y técnicas estadísticas no paramétricas en la solución de problemas concernientes a las ciencias del comportamiento.

INTRODUCCIÓN

El término *estadística no paramétrica* se refiere a un conjunto de métodos inferenciales válidos para formas muy diversas de distribución de la población. La aplicación de estos métodos no requiere modelo de población, en el sentido de un parámetro específico relacionado con la forma de la curva que representa a la población en estudio, como sí es necesario, por ejemplo, en el caso de la distribución normal. En el contraste de hipótesis, las pruebas estadísticas no paramétricas usualmente emplean algunos datos más simples de la muestra, como los signos de las mediciones, las relaciones de orden o las categorías de las frecuencias. Estos rasgos generales no requieren escalas de medición numéricas significativas. Por otra parte, aún más importante es que a estos métodos no los afecta el alargamiento o estrechamiento de la escala. Una aclaración final indispensable es que los términos *distribución libre* y *estadística no paramétrica* no son sinónimos, aunque en este texto se usarán indistintamente.

A estos procedimientos se les llama de distribución libre, por no considerar la forma como se distribuye la población. Tienen ventajas sobre las pruebas paramétricas, algunas de ellas son: *i*) implican menos requisitos de uso, *ii*) son más sencillas de entender y de aplicar, y *iii*) los procedimientos de cálculo resultan menos laboriosos. Por otra parte, los métodos no paramétricos tienen ciertas desventajas: *i*) se pierde información, *ii*) la potencia de estas pruebas es menor que la de las pruebas paramétricas, y *iii*) tienden a ser "conservadoras"; es decir, orientan hacia la aceptación de la hipótesis nula con más frecuencia de lo que deberían.

En estas circunstancias, las pruebas estadísticas paramétricas son preferibles a las no paramétricas, pero si la población no está normalmente distribuida o las varianzas poblacionales no son homogéneas o iguales, entonces puede utilizarse una prueba de distribución libre o no paramétrica como un buen sustituto de su análoga paramétrica, sobre todo cuando la muestra en estudio es pequeña.

Se discutirán métodos que únicamente requieren mediciones nominales, comparando distribuciones enteras. También se considerarán las técnicas que requieren datos ordinales. Cuando los datos son categóricos se hará un análisis estadístico, utilizando el modelo de la χ^2 (*ji* cuadrada), que si bien algunos autores la consideran no paramétrica, se ha creído oportuno que forme parte de este capítulo en sus aplicaciones más relevantes, como:

a) Independencia
b) Homogeneidad o proporción
c) Mediana.

Para la prueba de bondad de ajuste se emplearán dos técnicas:

i) La χ^2 (*ji* cuadrada) y
ii) La Kolmogorov-Smirnov.

Para el análisis de varianza se utilizarán dos modelos:

a) La prueba de Kruskal-Wallis, para muestras independientes.
b) La prueba de Friedman, para una sola muestra medida más de dos ocasiones.

Por último se encuentran los coeficientes de asociación, también llamados *de correlación:* γ_s, τ, ω, γ_{bp}.

14.1 PRUEBA DE BONDAD DE AJUSTE

En esta sección se describirán dos modelos para establecer si una variable aleatoria determinada se ajusta a una distribución de probabilidad específica; es decir, para saber si la muestra obtenida en forma aleatoria se distribuye como normal, binomial, Poisson u otra. Dichas pruebas son las siguientes:

1. χ^2 (*ji* cuadrada). Esta prueba es clásica y se emplea tanto para variables continuas como discretas.
2. *K-S* (Kolmogorov-Smirnov, para una muestra). Esta prueba se considera como no paramétrica; se utiliza cuando la muestra es pequeña y la variable de interés continua.

14.1.1 *ji* cuadrada (χ^2)

Este modelo, obtenido por K. Pearson en 1900, mide la discrepancia entre la frecuencia observada y la esperada teóricamente, con base en una distribución hipotética.

La prueba de bondad de ajuste ayuda a decidir si los resultados de un experimento concuerdan con los esperados de acuerdo con alguna ley, modelo o teoría científica.

Esto se lleva a cabo de la siguiente manera:

1. Se obtienen las frecuencias observadas y se ubican en un cuadro de contingencias (también llamada *tabla de doble entrada* o *diagrama de Carroll*) (véase sección 3.4).
2. Se construye un cuadro de frecuencias esperadas que concuerden con la distribución teórica o el modelo científico.
3. Según el número de variables de criterio que se consideren, será el cuadro de contingencia ($l \times c$) o ($r \times l$); la prueba de bondad de ajuste se empleará para una muestra y una o más variables de criterio.

Existe una familia de curvas χ^2, derivadas de una variable normalmente distribuida, cuya forma depende del tamaño de la muestra. Para muestras pequeñas (que tienen pocos grados de libertad), esta distribución se halla fuertemente sesgada en dirección positiva (todos sus valores son positivos y

varían de cero a infinito). Cuando la muestra es grande ($n \to \infty$), la χ^2 se aproxima a la distribución normal (véase la sección 7.6).

El modelo que se utilizará en esta sección es el siguiente:

$$\chi^2 = \Sigma \frac{(fo - fe)^2}{fe}$$

donde:

fo = frecuencia observada
fe = frecuencia esperada.

Para cada clase o categoría se obtiene el cuadrado de la diferencia entre la frecuencia observada y la esperada, y este resultado se divide por la misma frecuencia esperada de dicha categoría. La suma total de cada categoría proporciona el valor de χ^2.

A continuación se enumeran los requisitos de uso de la prueba de χ^2 para bondad de ajuste:

1. Se requiere un mínimo de 50 observaciones (mediciones) para que la distribución χ^2 sea una aproximación razonable de la distribución muestral esperada.
2. La frecuencia esperada para cada categoría debe ser por lo menos de 5; a fin de cumplir este requisito, combinar las categorías.
3. En el caso de la bondad de ajuste para la distribución normal, deben conocerse μ_X y σ_X o sus estimadas $\bar{x}$ y s, a fin de poder calcular las frecuencias esperadas.

Procedimiento

Paso 1. Identifique la variable de interés.

Paso 2. Establezca las hipótesis estadísticas.

H_0: las observaciones muestrales han sido extraídas de una distribución normal,

o bien

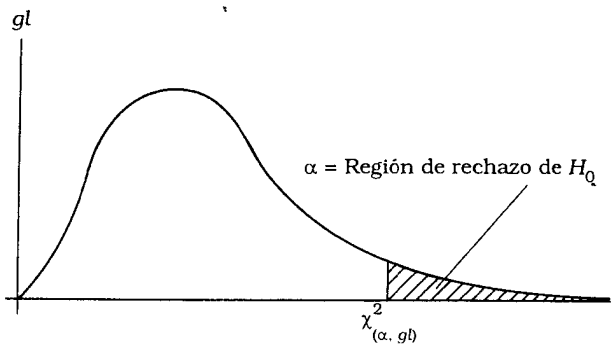
H_0: $fo = fe$ o ($fo - fe = 0$). No existe diferencia entre la frecuencia esperada y la observada.

H_1: $fo \neq fe$ o ($fo - fe \neq 0$).

Paso 3. Proponga el valor de α.

Paso 4. Determine el modelo estadístico que se utilizará. En este caso

$$\chi^2 = \Sigma \frac{(fo - fe)^2}{fe}$$



Paso 5. Obtenga mediante la tabla de χ^2 (*ji* cuadrada) del apéndice el valor crítico de χ^2 (α, *gl*) y ubique la región de rechazo en una gráfica.

Paso 6. Especifique la regla de decisión (RD) para rechazar H_0.

$$\text{Si } \chi^2_c \geq \chi^2_{(\alpha,\ gl)}, \text{ entonces } H_0 \text{ se rechaza.}$$

Paso 7. Calcule χ^2, por medio de $\chi^2 = \Sigma \dfrac{(fo - fe)^2}{fe}$, habiendo calculado previamente *fe*,

$fe = np$

donde

n = tamaño de la muestra

p = probabilidad de ocurrencia en cada categoría.

Paso 8. Obtenga las conclusiones considerando el rechazo de H_0 o de H_1. (La distribución muestral se ajusta a la distribución normal o no se ajusta a ella.)

Una variable de criterio. Cuando las categorías de la distribución de frecuencias se basan en una sola variable, constituyen un análisis de frecuencias de una variable de criterio, no obstante que existan varios resultados: lanzar una moneda (cara, cruz), tirar un dado (1, 2, 3, 4, 5 o 6).

Ejemplo Si se lanza una moneda al aire, únicamente caerá cara o cruz; pero si se lanzan 1 000 monedas, la única variable que se tomará en cuenta será la cara de la moneda que aparecerá, pero se contarán cuántas caras y cruces caerán.

Después de determinar el muestreo, las 1 000 monedas se lanzan al aire y se registra la frecuencia de caras y cruces en un cuadro de frecuencias observadas. A continuación, se construye el cuadro de frecuencias esperadas, previstas por los resultados más simples de probabilidad elemental. Para una muestra de 1 000 monedas, la hipótesis nula (H_0) implica que deben obtenerse 500 caras y 500 cruces. En este caso, la prueba de *ji* cuadrada ayudará a decidir

si la discrepancia entre los resultados observados y los esperados corresponde simplemente a la fluctuación del muestreo o indica el sesgo de la población.

Frecuencias observadas		
Cara	Cruz	Total
530	470	1 000
(a)	(b)	

Frecuencias esperadas		
Cara	Cruz	Total
500	500	1 000
(a)	(b)	

o sea:

530 / 500	470 / 500	1 000

Paso 1. Como la variable de criterio es la posición de la cara de la moneda al caer, entonces se establece la hipótesis nula y la alternativa de la siguiente manera:

H_0: cara y cruz ocurren con igual frecuencia.
H_1: cara y cruz *no* ocurren con igual frecuencia.

Paso 2. Como $gl = 1$ y $\alpha = 0.01 = 1\%$, entonces:

$$\chi^2 \text{ crítica} = \chi^2\,(1,\ 0.01) = 6.64$$

fo_1 / fe_1	fo_2 / fe_2	. . .	fo_n / fe_n

donde fo_i es la frecuencia observada i, $i = 1, 2, \ldots, n$
y fe_i es la frecuencia esperada i, $i = 1, 2, \ldots, n$

Paso 3. Se define la *población* de estudio como: todas las monedas de cierto valor y con características determinadas.

Paso 4. La muestra n son 1 000 monedas seleccionadas al azar de la población anterior.

Paso 5. El modelo estadístico por aplicar es:

$$\chi^2 = \Sigma \frac{(fe - fo)^2}{fe}$$

Paso 6. Al sustituir los datos concentrados en los cuadros anteriores y efectuar las operaciones resulta:

$$\chi_c^2 = \frac{(530 - 500)^2}{500} + \frac{(470 - 500)^2}{500} = 3.60$$

$$\boxed{\therefore\ \chi_c^2 = 3.60}$$

Paso 7. Regla de decisión (RD): si $\chi^2 \geq \chi^2_{\alpha,gl}$, se rechaza H_0. Como 3.60 < 6.64, entonces, *no* se rechaza H_0.

Paso 8. Conclusión Los resultados no son significativos. En este estudio de las monedas no se encontró indicación alguna de sesgo, debido a que no se predice respuesta (salida) específica alguna; por ejemplo, las caras ocurrirán con mayor frecuencia. Por ello, el nivel de significancia se establece relativamente bajo (0.01) por dos razones: *i*) el tamaño de la muestra es grande y se minimiza β (error tipo II) y *ii*) el error más grave sería decidir que las monedas tienen un sesgo, cuando en realidad no es así.

La regla de decisión (RD) establece que H_0 se rechazará si la *ji* cuadrada (χ^2) calculada es mayor que el valor crítico de 6.64 (los procedimientos para determinar los valores críticos se estudiarán después).

Pudo cometerse un error del tipo II; sin embargo, el interés en un posible sesgo de las monedas disminuye en gran escala, si no es que desaparece por completo.

Dos variables de criterio. En algunos problemas de bondad de ajuste se obtienen medidas de dos o más variables no controladas. Las frecuencias esperadas en cada combinación de categorías se generan a partir de algunas distribuciones teóricas; para evitar confusiones en este tipo de estudios es recomendable colocar los datos en una tabla de contingencia de doble entrada.

fo_{11} / fe_{11}	fo_{12} / fe_{12}	...	fo_{1n} / fe_{1n}
fo_{21} / fo_{21}	fo_{22} / fe_{22}	...	fo_{2n} / fe_{2n}
⋮	⋮	⋮	⋮
fo_{m_1} / fe_{m_1}	fo_{m_2} / fe_{m_2}	...	fo_{mn} / fe_{mn}

donde fo_{ij} es la frecuencia observada ij, $i = 1, 2, \ldots, n$ y $j = 1, 2, \ldots, n$
y fe_{ij} es la frecuencia esperada $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$

Ejemplo Aplicación de una prueba de bondad de ajuste para una sola muestra y dos variables de criterio.

Un genetista efectúa la cruza de una muestra de chícharos con las siguientes variables de criterio: color (verde o amarillo) y textura (liso o rugoso). Existen cuatro combinaciones posibles como resultados de las cruzas: amarillo-liso (*AL*), amarillo-rugoso (*AR*), verde-liso (*VL*) y verde-rugoso (*VR*); de acuerdo

con las leyes de Mendel, la combinación *AL* deberá ocurrir nueve veces más frecuentemente que la *VR*, mientras que las mezclas *AR* y *VL* ocurrirán tres veces con mayor frecuencia que la *VR*. Esta distribución teórica se utiliza para construir una tabla de frecuencias esperadas de (2 × 2).

Debido a que en nuestro ejemplo la muestra consta de 960 elementos (chícharos), el investigador podrá esperar que las combinaciones posibles ocurrirán aproximadamente con las siguientes frecuencias:

$$VR = 60$$

$$VL = 180$$

$$AR = 180$$

$$AL = 540$$

Las frecuencias observadas son las siguientes:

$$VR = 50$$

$$VL = 200$$

$$AR = 160$$

$$AL = 550$$

Por tanto, las frecuencias observadas y esperadas se mostrarán en las siguientes tablas de contingencia:

		Textura	
		RUGOSO	LISO
Color	VERDE	50 / 60	200 / 180
	AMARILLO	160 / 180	550 / 540

Los grados de libertad para estas tablas de contingencia son:

$gl = (r - 1)(c - 1)$
donde: r = número de renglones.
c = número de columnas.
o sea $gl = (2 - 1)(2 - 1) = 1$

aplicando el modelo $\chi^2 = \Sigma \dfrac{(fo - fe)^2}{fe}$, se tiene

$$\chi_c^2 = \frac{(50-60)^2}{60} + \frac{(200-180)^2}{180} + \frac{(160-180)^2}{180} + \frac{(550-540)^2}{540}$$

$$= \frac{100}{60} + \frac{400}{180} + \frac{400}{180} + \frac{100}{540} = 6.30; \quad \boxed{\chi_c^2 = 6.30}$$

También puede utilizarse la fórmula siguiente:

$$\chi^2 = \Sigma \frac{fo^2}{fe} - n$$

$$\chi_c^2 = \Sigma \frac{fo^2}{fe} - n = \frac{(50)^2}{60} + \frac{(200)^2}{180} + \frac{(160)^2}{180} + \frac{(550)^2}{540} - 960$$

$$\chi_c^2 = 41.667 + 222.22 + 142.22 + 560.185 - 960 = 6.30; \quad \boxed{\chi_c^2 = 6.30}$$

Supóngase que el genetista decide tomar un nivel de confianza de 10%, o sea, $\alpha = (0.10)$, y el valor crítico de χ^2, determinado en la tabla χ^2, es de 2.706. En ese caso la hipótesis nula de la distribución observada, al igual que la distribución esperada por la ley de Mendel, se rechazaría, ya que la $\chi^2 = 6.30$ es mayor que el valor crítico de $\chi_0^2 = 2.706$. Sin embargo, considérese que las probabilidades de cometer un error del tipo I son de 1 en 10, por el nivel de confianza $\alpha = 0.10$. Las conclusiones se dejan al lector.

Ejemplo 1 Se realiza una investigación con el fin de producir un nuevo tipo de insulina, la hormona que se utiliza para controlar enfermos de *diabetes mellitus*. Los siguientes criterios de inclusión se consideran en la muestra de estudio:

- $n = 400$ pacientes del mismo sexo.
- Mismo grado de evolución de la enfermedad.
- Dieta controlada.
- Supervisión médica.

En este experimento se medirá la respuesta de los pacientes a la insulina convencional, así como el porcentaje de ellos en cada una de las siguientes categorías estandarizadas:

Categoría 1. Decremento intenso en la glucosa.
Categoría 2. Decremento moderado en la glucosa.
Categoría 3. Decremento ligero en la glucosa.
Categoría 4. Decremento nulo o ligero incremento en la glucosa.

Con base en un estudio de nivel nacional, se determinó que los porcentajes por categoría son, respectivamente: 50%, 25%, 15% y 10%. Por otra parte, los

resultados de las pruebas clínicas con el nuevo tipo de insulina aplicadas en 400 pacientes son: 240, 120, 30 y 10.

Paso 1. La variable de interés es la respuesta a cada categoría.

Paso 2.

H_0: $P_1 = 50\%$, $P_2 = 25\%$, $P_3 = 15\%$, $P_4 = 10\%$.

H_1: al menos una probabilidad de una categoría es diferente del valor esperado.

Paso 3.

$\alpha = 0.005$

Paso 4. $\chi^2 = \Sigma \dfrac{(fo - fe)^2}{fe}$

Paso 5. Obtener *gl*.

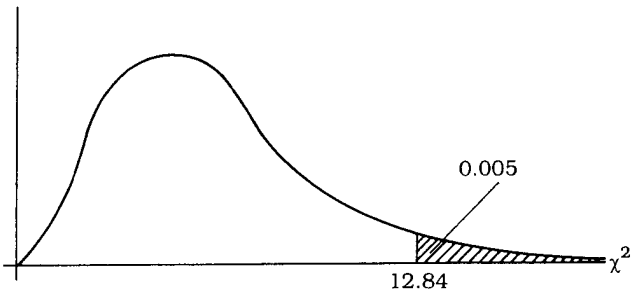
Como $\Sigma\, fo = \Sigma\, fe$ (o su equivalente *n* es la única restricción).

Entonces $gl = k - 1 = 4 - 1 = 3$

$gl = 3$

$$\chi^2_{(0.005,3)} = 12.84$$

Paso 6. Si $\chi^2_c \geq 12.84$, H_0 se rechaza.



Paso 7. Para emplear $\chi^2 = \Sigma \dfrac{(fo - fe)^2}{fe}$

Se calcula primero *fe* para cada una de las cuatro categorías. Al utilizar $fe = np$

Categoría	*n* *p* *fe*
	Porcentaje
1	400 (50) = 200
2	400 (25) = 100
3	400 (15) = 60
4	400 (10) = 40
	100% 400

Categoría	fo	fe	$\frac{(fo - fe)^2}{fe}$
1	240	200	8
2	120	100	4
3	30	60	15
4	10	40	22.5
	$\Sigma fo = 400$	$\Sigma fe = 400$	$\chi_c^2 = 49.5$

Paso 8. Si $\chi_c^2 \geq 12.84$ H_0 se rechaza.

Como 49.5 >> 12.84 H_0 se rechaza para $P < 0.005$.

Conclusiones La proporción es la misma tanto para pacientes en la *fe* (usuarios del nuevo tipo de insulina) en las categorías 3 y 4, 50 y 75%, respectivamente como en las *fo* (usuarios de insulina convencional). Por consiguiente la nueva insulina no es más efectiva que la convencional.

Ejemplo 2 Se implanta un nuevo chip para operar un programa de computadora y se instala en 200 máquinas seleccionadas al azar. Los datos son los siguientes:

Tiempo de acceso en mseg.

i	$LRI - LRS$	fo_i
1	Menos de 2.497	1
2	2.497 — 2.498	7
3	2.498 — 2.499	33
4	2.499 — 2.500	64
5	2.500 — 2.501	59
6	2.501 — 2.502	29
7	2.502 — 2.503	5
8	2.503 o más	2
		200

Calcular χ^2 prueba de bondad de ajuste de que la muestra fue extraída de una población que se distribuye normalmente con $\mu = 2.500$ mseg y $\sigma = 0.001$ mseg.

Paso 1. La variable de interés es el tiempo de acceso.

Paso 2.

H_0: $i = 1, 2, \ldots, 8$ (para todas las frecuencias).

H_1: $i = 1, 2, \ldots, 8$ (al menos una frecuencia es diferente).

Paso 3.

$\alpha = 5\%$

Paso 4.

$$\chi^2 = \Sigma \frac{(fo - fe)^2}{fe}$$

Paso 5. Obtenga los grados de libertad.

Como $\mu = 2.500$ y $\sigma = 0.001$, se especifican $gl = k - 1$ y $k = 6$.

Entonces, $\boxed{gl = 5}$

$k = 6$ debido a que en las categorías anteriores 1 y 8 $fo < 5$, se tiene que combinar, y se obtiene:

i	LRI — LRS	fo_i
1	Menos de 2.498	8
2	2.498 — 2.499	33
3	2.499 — 2.500	64
4	2.500 — 2.501	59
5	2.501 — 2.502	29
6	2.502 o más	7
		$\Sigma fo = 200$

Paso 6. Regla de decisión (RD).

$$\text{Si } \chi_c^2 \geq \chi^2_{(0.05.5)}$$

$$\chi_c^2 \geq 11.07 \quad H_0 \text{ se rechaza}$$

Paso 7. Calcule $\chi^2 = \Sigma \frac{(fo - fe)^2}{fe}$

Previamente se determina $fe = np$ y se obtiene la probabilidad P', mediante las tablas del área bajo la curva normal. Después de que se han calculado las puntuaciones z, por medio de $z = \frac{LRS - \mu}{\sigma}$, para cada una de las categorías se obtienen:

LRI — LRS	fo	z	P'
Menos de 2.498	8	-2.00	0.0228
2.498 — 2.499	33	-1.00	0.1359
2.499 — 2.500	64	0.00	0.3413
2.500 — 2.501	59	+1.00	0.3413
2.501 — 2.502	29	+2.00	0.1359
2.502 o más	7	+2.16	0.0228
	200		1.00

Se considera el área de la región más pequeña para los valores negativos de z y el área de la región más grande para los valores positivos de z.

Una vez obtenida P', se calcula $fe = np$ y la χ^2, y se comprueba $\Sigma fo = \Sigma fe = 200$

fe	$\frac{(fo-fe)^2}{fe}$
5	1.80
27	1.33
68	0.24
68	1.19
27	0.15
5	0.80
200	5.51

$$\boxed{\chi_c^2 = 5.51}$$

Paso 8. *Regla de decisión.*

Si χ_c^2 (con un valor de 5.51) es mayor que 11.07, H_0 se rechaza. Como 5.51 no es mayor que 11.07, H_0 no se rechaza (H_1 se rechaza).

Conclusión La distribución probabilística de la muestra se distribuye de manera normal.

14.1.2 Kolmogorov-Smirnov (K-S)

Se trata de una prueba alternativa a la clásica de *ji* cuadrada (χ^2) para la bondad de ajuste. Puede aplicarse a muestras pequeñas que requieran menos cálculos que la χ^2, pero mientras ésta se aplica tanto a variables continuas como discretas, la de Kolmogorov-Smirnov (*K-S*) únicamente procede para variables continuas.

En principio se supone que una población tiene una distribución determinada dividida en K intervalos de igual área o probabilidad. Posteriormente, se selecciona al azar una muestra de tamaño n de dicha población, tomando en cuenta el número de puntuaciones o mediciones correspondientes a cada intervalo K. Esto significa que la prueba *K-S* se utiliza para comparar frecuencias relativas acumuladas observadas y esperadas, así como para contrastar la hipótesis nula de que los datos observados se han recopilado de una distribución de probabilidad determinada.

Esta prueba estadística muestra cuál es la máxima diferencia absoluta ($D_{máx}$) entre cualquier par correspondiente de frecuencias relativas acumuladas observadas y esperadas.

Ejemplo Un profesor de español aplica a su grupo de 25 alumnos una prueba estandarizada de gramática, en la que obtiene los siguientes resultados:

56	58	40	77	87
75	61	70	73	71
66	69	67	68	60
72	73	61	64	66
84	72	52	65	67

El objetivo de la prueba consiste en comparar los resultados del grupo con los estándares nacionales. El manual del examen señala que los deciles de las puntuaciones, para alumnos de bachillerato son los siguientes:

Decil	1	2	3	4	5	6	7	8	9
Puntuación	45.0	56.8	62.5	66.1	68.7	71.3	74.0	78.5	84.2

De esa manera, el profesor sabe que 10% de la población tiene puntuaciones menores que 45.0, que 20% de la misma los registra inferiores a 56.8, y así sucesivamente. Para determinar la bondad de ajuste, el profesor calcula la frecuencia relativa acumulada observada (*FRAo*) inferior a cada uno de los deciles. Por ejemplo, en la muestra se observa una puntuación de 40, que se encuentra en el primer decil (45.0); así, la *FRAo* es $\frac{1}{25} = 0.04$.

En el rango inferior al segundo decil (56.8), se encuentran tres calificaciones (incluida la del primer decil). Entonces, la frecuencia relativa acumulada esperada (*FRAe*) en el intervalo inferior del segundo decil es 0.12, y así sucesivamente (todos los resultados se muestran en el cuadro 14.1). Esto se debe a que los deciles dividen la distribución de la población en 10 áreas iguales, cada una de 0.10.

El valor absoluto de la diferencia entre cada par de *FRA* (la observada y la esperada) se denota por *D* y se calcula como aparece en el cuadro antes indicado. De acuerdo con la estadística, si los datos de la muestra provienen de una población con la misma distribución, representada por la *FRA* (representada), entonces la *D* más grande ($D_{máx}$) tiene una distribución de muestreo conocida (véase el cuadro 14.1). En consecuencia, si los datos de la muestra se ajustan bien, la $D_{máx}$ será significativamente pequeña, mientras que si el ajuste es pobre, $D_{máx}$ será significativamente grande. El cuadro 14.1 contiene valores de $D_{máx}$ para distintos tamaños de muestra (n) y diferentes grados de probabilidad de que $D_{máx}$ n sea menor que el valor de la tabla (P). En el ejercicio, $D_{máx} = 0.14$ y $n = 25$; si se utiliza un nivel de confianza de 0.05, la tabla muestra que la $D_{máx}$ debe ser mayor o igual a 0.264 para rechazar la hipótesis nula; o sea RD: si $D_{máx}$ (o) $\geq D_{máx}$ (e), entonces se rechaza H_0. Como $0.14 < 0.264$, no se rechaza la hipótesis nula.

Conclusión. La distribución muestral no difiere significativamente de la poblacional. La comparación de *FRA* para el ejemplo de la prueba de gramática es $D_{máx} = 0.14$.

Cuadro 14.1

Deciles	1	2	3	4	5	6	7	8	9	*Total*
FRAo (observada)	0.04	0.12	0.28	0.44	0.56	0.68	0.84	0.92	0.96	1.0
FRAe (esperada)	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.0
D	0.06	0.08	0.02	0.04	0.06	0.08	0.14	0.12	0.06	0.00

14.2 TABLAS DE CONTINGENCIA

En general, cuando se clasifica a los sujetos en diferentes categorías y se cuentan los que forman cada una de ellas, es decir, se considera a la frecuencia en cada categoría representada por una variable categórica, nominal o cualitativa.

Estas clasificaciones cuyas categorías son mutuamente excluyentes y completamente exhaustivas implican que existen suficientes categorías para ubicar a todos los elementos de la muestra.

14.2.1 Prueba de homogeneidad

El valor estadístico χ^2 se utiliza con frecuencia para determinar si dos o más muestras, y por consiguiente las poblaciones, son *homogéneas*; es decir, si las distribuciones de los datos son similares respecto a una variable de criterio en particular.

Para ilustrar lo anterior, supóngase que se desean conocer las actitudes de la iniciativa privada y el gobierno de un estado de la República Mexicana respecto a la aprobación de un proyecto de educación especial. A fin de llevar a cabo esta investigación se seleccionan dos muestras al azar por completo separadas e independientes, una correspondiente a la iniciativa privada y la otra al sector gubernamental. La variable de criterio es la actitud, expresada en las categorías favorable o no favorable. La prueba de homogeneidad χ^2 indica que las actitudes de ambos grupos no difieren significativamente, así como que las dos poblaciones pueden considerarse homogéneas o esencialmente sensibles a diferencias entre ellas (forma, tendencia central o variabilidad) que afectan las proporciones en las categorías. En la prueba de homogeneidad no es necesario que las muestras seleccionadas de cada población tengan el mismo tamaño, pero es recomendable que sean similares en magnitud.

Las pruebas χ^2 de homogeneidad son, en ciertas condiciones, idénticas a las pruebas paramétricas *t* de Student para dos muestras, como se indicó en el capítulo 10. Por otra parte, las pruebas de homogeneidad constituyen alterna-

tivas muy útiles cuando no es posible cumplir con los requisitos de las pruebas paramétricas.

Ejemplo Aplicación de la prueba χ^2 para la homogeneidad de dos poblaciones.

Un agrónomo desea comparar la proporción del crecimiento anual de abetos en el estado de Michoacán y el de Chihuahua (véase el cuadro 14.2). Las dos poblaciones abarcan árboles que crecen tanto en una como en otra entidad; el muestreo se completa cuando se seleccionan al azar e independientemente 100 árboles de cada población.

La variable de criterio sobre la proporción de crecimiento se estableció arbitrariamente en dos categorías: menos de 10 cm y 10 cm o más. La hipótesis tiene que ser no direccional o de dos colas. La H_0 señala que ambas poblaciones son en esencia homogéneas; o sea, que no difieren respecto a la medida del criterio. Debido a que el error de tipo II es muy significativo para el investigador, éste elige $\alpha = 0.05$, cuyo valor crítico es de 3.84 para un grado de libertad.

Cabe tener en cuenta que no todos los árboles en las muestras sobrevivieron un año, por lo que A se redujo a 96 unidades y B a 94. Los árboles faltantes se clasificaron de acuerdo con la tasa de crecimiento y la observación de la tabla de frecuencias.

Para calcular el valor de la χ^2 se utilizará la fórmula siguiente:

$$\chi^2 = \frac{n\,(ad - bc)^2}{i\,jkl}$$

De esa manera se elimina el problema de calcular frecuencias esperadas. Por razones de cálculo esta fórmula se reserva para n menores que 100. Como la χ^2 es mayor que el valor crítico y, por consiguiente, se rechaza la H_0 a un nivel de 5%, los datos parecen apoyar la conclusión de que la proporción de crecimiento en el estado de Michoacán es mayor que en Chihuahua.

El agrónomo no podrá realizar esta comparación específica a un nivel conocido de probabilidad y con una prueba no direccional de dos colas de la *ji* cuadrada. Si tal hipótesis alterna pudiera haberse establecido con anterioridad y si se hubiesen cumplido otros requisitos, una prueba paramétrica de t de una cola habría aceptado esta conclusión.

Cuadro 14.2

Frecuencias observadas

	Tasa de crecimiento		
	Menos de 10 cm	10 cm o más	Total
Muestra A	9 (a)	87 (b)	96 (i)
Muestra B	23 (c)	71 (d)	94 (j)
Total	32 (k)	158 (l)	190 = n

A continuación se presenta al lector una forma desarrollada de resolver este problema:

Paso 1. La población la constituyen abetos de tres años de edad, estudiados en el periodo de un año.

Muestra *A*. Los abetos que crecen en el estado de Michoacán.

Muestra *B*. Los abetos que crecen en el estado de Chihuahua.

Paso 2. La muestra *A* se compone de 100 plantas seleccionadas al azar de la población *A*.

La muestra *B* consta de 100 plantas seleccionadas al azar de la población *B*.

Paso 3. La variable de criterio es la tasa de crecimiento durante el periodo de un año. Esta proporción será alrededor de 10 cm.

Paso 4. Modelo de decisión.

H_0: Las tasas de crecimiento de las poblaciones *A* y *B* son iguales.

H_1: Las tasas de crecimiento de las poblaciones *A* y *B* son diferentes.

Paso 5. Nivel de confianza: $\alpha = 0.05$.

Paso 6. RD: si $\chi_c^2 \geq 3.84$, se rechaza H_0.

Paso 7. Fórmula y desarrollo numérico:

$$\chi_c^2 = \frac{n\,(ad - bc)^2}{i\,jkl} = \frac{190(9 \times 71 - 87 \times 23)^2}{96 \times 94 \times 32 \times 158}$$

$$= \frac{190(639 - 2001)}{45\,625\,344} = 7.73$$

$$\boxed{\chi_c^2 = 7.73}$$

Paso 8. Debido a que $7.73 > 3.84$, se rechaza H_0.

Conclusión Según la clasificación de este estudio, la distribución de las tasas de crecimiento no es igual: en ambas difiere significativamente.

Ejemplo Prueba de homogeneidad para tres o más muestras.

Un estudio del mercado en escala nacional referente a tres estilos de sacos de gamuza para caballero se plantea determinar las distintas preferencias según regiones. Se obtienen cuatro muestras de 60 compradores, una por cada región

considerada: norte, sur, este y oeste. Por tanto, la variable de criterio es la preferencia y se obtiene al preguntar a cada comprador cuál es su estilo preferido. Así pues, la hipótesis nula establece que no existen diferencias significativas.

Las frecuencias observadas y las esperadas se muestran en el cuadro 14.3. Se obtuvieron las frecuencias esperadas mediante los totales marginales de las tablas de frecuencia observadas. En el ejemplo, las frecuencias esperadas para cada celda en una columna en particular son las mismas, debido a que todos los marginales de los renglones también son iguales. Las celdas en la columna que representa el estilo *A*, por ejemplo, proporcionan las siguientes frecuencias esperadas: (41) (60) / 240 = 10.25.

Nótese que la tabla de frecuencias esperadas es congruente con la hipótesis nula de no preferencias de estilo por renglón, aunque los estilos *B* y *C* son más populares que el *A*.

$$x = \frac{41\,(60)}{240}$$

$$x = 10.25$$

$$y = \frac{106\,(60)}{240}$$

REGIÓN	ESTILO *A*	ESTILO *B*	ESTILO *C*	
Este	6 / 10.25	32 / 26.5	22 / 23.25	60
Norte	10 x / 10.25	21 / 26.5	29 / 23.25	60
Sur	22 / 10.25	15 y / 26.5	23 / 23.25	60
Oeste	3 / 10.25	38 / 26.5	19 / 23.25	60
	41	106	93	240

calculando la χ^2, se tiene:

$$\chi_c^2 = \Sigma \frac{(fo - fe)^2}{fe} = \frac{(6.0 - 10.25)^2}{10.25} + \frac{(32.0 - 26.5)^2}{26.5} \ldots + \frac{(19.0 - 23.5)^2}{23.5}$$

$$\chi_c^2 = 34.93$$

Los grados de libertad para este problema son 6, o sea (4 – 1) (3 – 1). El investigador ha escogido el nivel de significancia $\alpha = 0.01$ (1%) y, para 6 grados de libertad, la *ji* cuadrada que se observa es 16.81. Como la *ji* cuadrada calculada es mayor, o sea $\chi_c^2 \geq \chi^2$ (crítica), entonces el investigador infiere con bastante confianza la relación de preferencia entre el estilo del saco y la región de procedencia del posible comprador.

El investigador puede necesitar información adicional del cuadro 14.3; por ejemplo, determinar si el estilo *A* es menos aceptable que los *B* y *C* en todas las regiones. Para realizar esto, se necesita un análisis más cuidadoso, que se

considerará al estudiar el análisis de varianza no paramétrico (véase también M. McSweeney, 1977).

Otra opción es extraer información nueva para el análisis de χ^2 de cada comparación o el contraste de intereses, como lo señaló la inspección en la tabla completa, o también ceder a cualquier prueba adicional de confianza y efectuar observaciones sobre cada una de las celdas o conjuntos de celdas que hayan contribuido grandemente a la obtención de la χ^2 para la tabla de contingencia anterior; por ejemplo, los valores de las celdas *g* y *h* sobresalen en particular (13.5 y 5.0, respectivamente), por lo que juntos contribuyen a más de la mitad del total de χ^2 (13.5 + 5.0 = 18.5). Con base en esto, el investigador puede inferir que es posible vender el estilo *A* con mayor éxito en las regiones del sur que el estilo *B*; pero no está en condiciones de determinar la probabilidad de verdad o error en esta decisión. Si una decisión errónea acarreara consecuencias financieras graves, se requiere una prueba estadística adicional a partir de una nueva muestra, en especial para la región sur, a fin de comprobar la hipótesis de no diferencia en la preferencia de los estilos *A* y *B*.

14.2.2 Prueba de independencia

A diferencia de la prueba de bondad de ajuste para una muestra, la prueba de independencia siempre considera medidas de dos variables para cada elemento de la muestra, y no se utiliza una distribución teórica independiente *per se* para construir frecuencias esperadas. Algunas veces se conoce como *prueba de asociación* y se relaciona con los procedimientos de correlación. La muestra para la prueba de independencia se compone de miembros seleccionados al azar de una sola población de interés.

Por consiguiente, la prueba se utiliza para determinar si las medidas en las dos variables de criterio son independientes o dependientes respecto a otra variable dentro de una población en particular. En ese sentido, se podrían investigar las posibles asociaciones entre el precio de menudeo de televisores y la calidad de su resolución; el número de libros que se leen y el nivel académico de los lectores; el peso de un mamífero y su capacidad craneal, entre otros temas.

Analíticamente, la prueba de independencia es idéntica a la de homogeneidad. Una de las distribuciones más obvias es que el investigador determina el tamaño de cada muestra en la prueba de homogeneidad; por consiguiente, la frecuencia total para el problema y los totales marginales, tanto en las columnas como en los renglones. En la prueba de independencia únicamente se conoce la frecuencia total (el tamaño de la muestra); los totales marginales de cada renglón y columna son variables aleatorias cuyas frecuencias no se determinan sino hasta que la muestra ha sido extraída y los datos recopilados. En este nivel, el lector debe ser capaz de reconocer que el problema citado en el cuadro 14.3 podría analizarse con una prueba de independencia.

Ejemplo Se analizarán las actitudes de los residentes de una ciudad ante la construcción de un nuevo centro cívico (primera variable), las cuales son inde-

pendientes de las que suscita la edificación de un nuevo palacio municipal (segunda variable). Si las dos variables se asocian positivamente, quienes están a favor del centro cívico opinarán en ese sentido respecto al palacio municipal. No obstante, si ambas variables son independientes, la gente que respalda la edificación del centro fluctuará en su posición hacia el palacio municipal: algunos aprobarán y otros no. En otros términos, cuando las actitudes son independientes, el conocimiento de una no proporcionará información respecto a la otra. Se desarrollará este ejemplo más adelante. Como se muestra en el cuadro 14.3, es posible definir una sola población y extraer una muestra; los miembros de la muestra podrían responder a dos preguntas sobre actitudes: una concerniente al centro cívico y la otra al palacio municipal.

Para elaborar la tabla de frecuencias observadas, se asignará a cada miembro una casilla o celda particular en un cuadro de doble entrada que contiene las dos respuestas posibles.

Las frecuencias observadas marginales de la tabla se utilizan para calcular las frecuencias esperadas correspondientes a estas pruebas. Nótese que los entrevistados tienden a oponerse a ambos proyectos de construcción, pero, como es proporcional en relación con las marginales, la tabla de frecuencias esperadas describe una situación congruente con H_0. De los 132 sujetos en favor del proyecto *A*, cerca de 43% apoyará el proyecto *B*; por otra parte, de los 218 sujetos contrarios al proyecto *A*, cerca de 43% estará a favor del proyecto *B*.

Cuadro 14.3

Frecuencias observadas

	Actitud B		
Actitud *A*	A favor	En contra	Total
A favor	47 (*a*)	85 (*b*)	132 (*i*)
En contra	102 (*c*)	116 (*d*)	218 (*j*)
Total	149 (*k*)	201 (*l*)	350 (*n*)

Frecuencias esperadas

	Actitud B		
Actitud *A*	A favor	En contra	Total
A favor	56.19 (*a*)	75.81 (*b*)	132 (*i*)
En contra	92.81 (*c*)	25.19 (*d*)	218 (*j*)
Total	149 (*k*)	201 (*l*)	350 = *n*

Desarrollo del ejemplo de la prueba χ^2 de independencia:

Paso 1. La población consiste en los residentes adultos de la ciudad *Z*.

Paso 2. La muestra constituye 350 individuos escogidos al azar de la población.

Paso 3. Las variables de criterio son *A*: actitud hacia la construcción del nuevo palacio municipal (a favor o en contra) y *B*: actitud hacia la construcción del nuevo centro cívico (a favor o en contra).

Paso 4. Modelo de decisión.
H_0: Las actitudes *A* y *B* son independientes. H_1: Las actitudes *A* y *B* están asociadas (son dependientes).

Paso 5. Nivel de confianza: $\alpha = 0.05$ (5%).

Paso 6. RD: si $\chi_c^2 \geq 3.84$, entonces, se rechaza H_0.

Paso 7. Fórmula y desarrollo numérico:

$$\chi_c^2 = \Sigma \frac{fo^2}{fe} - n$$

$$\chi^2 = \frac{(47)^2}{56.19} + \frac{(85)^2}{75.81} + \frac{(102)^2}{92.81} + \frac{(116)^2}{125.19} - 350$$

$$= 39.31 + 95.30 + 112.10 + 107.48 - 350 = 354.20 - 350$$

$$\boxed{\chi_c^2 = 4.20}$$

Paso 8. Debido a que $4.20 > 3.84$, se rechaza H_0.

Conclusión Las actitudes *A* y *B* están asociadas en la población definida. La fuerza de esta asociación deberá calcularse con un coeficiente de correlación no paramétrico.

14.2.3 Procedimiento *post hoc*

Las pruebas de independencia sólo indican la presencia de asociación, pero no su magnitud (valor numérico). La sensibilidad de χ^2 a cualquier grado de asociación se incrementa en función del tamaño de la muestra. Sin embargo, se obtiene un valor significativo de χ^2 con una muestra grande, aunque la relación entre las variables es tan débil que su significado práctico resulta limitado. Cuando se ha rechazado la H_0 de independencia hay que calcular un índice de

asociación para obtener una estimación numérica de la fuerza de la relación. Estos índices incluyen el coeficiente Φ, el de Kramer y el de contingencia.

Ejemplo Un investigador busca establecer si existen diferencias entre los hombres y las mujeres que presentan examen de admisión a la universidad. Supóngase que las puntuaciones de dicho examen se clasifican como aceptados o rechazados. Los datos se obtuvieron de una muestra aleatoria de 200 aspirantes de la población total que desea ingresar en dicha institución; estos datos pueden exponerse en la siguiente tabla de contingencia (2 × 2).

Si se utiliza la forma convencional, en una sola tabla de contingencia:

		ACTITUD *B*		
		A favor	En contra	
ACTITUD *A*	A favor	47 / 56.19	85 / 75.81	132
	En contra	102 / 92.81	116 / 25.19	218
		149	201	350

$$fe = 25.19 = \frac{201\ (218)}{350}$$

Con el modelo $\chi^2 = \Sigma \dfrac{(fo - fe)^2}{fe}$

se tiene

$$\boxed{\chi_c^2 = 4.20}$$

	Aceptados	*Rechazados*	*Total*
Hombres	93 (*a*)	36 (*b*)	132
Mujeres	41 (*c*)	27 (*d*)	68
Total	134	63	200

Esto indica que la muestra (n = 200) la constituían 132 hombres y 68 mujeres; en ella, 93 hombres y 41 mujeres obtuvieron puntuaciones de aceptación. A continuación se calculan las frecuencias esperadas con respecto a si el sexo del aspirante es independiente de la puntuación de aceptación. El procedimiento es el siguiente:

Paso 1. Se calcula el total de puntuaciones aceptadas: 93 + 41 = 134.

Paso 2. Se encuentra la proporción de 134 (total de aceptados) respecto a n. Así, 134/200 = 0.67.

Paso 3. Se multiplica esta proporción por el número total de hombres en la muestra: 0.67 × 132 = 88.44; ésta es la frecuencia esperada para la casilla (*a*).

Paso 4. Se resta 88.44 de 132 (132 − 88.44) = 43.56, y así se encuentra la frecuencia esperada de hombres con puntuaciones de aceptación.

Tómese en cuenta que la suma de las frecuencias esperadas para un renglón o columna será igual a la suma de las frecuencias observadas para ese renglón o columna.

Paso 5. Posteriormente, se calculan las frecuencias esperadas para las mujeres con puntuaciones de aceptación: se resta 88.44 de 134, lo que resulta 45.56, y sucesivamente se calculan todas las frecuencias esperadas. Es necesario señalar que también existe un método directo de cálculo para las frecuencias esperadas, en el cual se aplica la siguiente fórmula en cada casilla:

$$fe = \frac{RC}{n}$$

donde:

fe = Frecuencia esperada de una casilla.
R = Total marginal del renglón de dicha casilla.
C = Total marginal de la columna de dicha casilla.
n = Tamaño de la muestra.

Paso 6. Se calcula χ^2 por medio del siguiente procedimiento para cada casilla:

$$\frac{(|fo - fe| - 0.5)^2}{fe}$$

Las líneas verticales, $|fo - fe|$, indican que debe tomarse el valor absoluto de esa diferencia. Los resultados de las casillas se suman, y el total será la χ^2. La corrección de continuidad de Yates se utiliza debido a que el grado de libertad para este diseño es 1, y la distribución real es discontinua.

$$\chi_c^2 = \frac{(4.06)^2}{88.44} + \frac{(4.06)^2}{43.56} + \frac{(4.06)^2}{45.56} + \frac{(4.06)^2}{22.44}$$

$$\chi_c^2 = 0.19 + 0.38 - 0.36 + 0.73 = 1.66$$

$$\boxed{\chi_c^2 = 1.66}$$

Paso 7. Se aplica la prueba de confianza. El valor de la *ji* cuadrada teórica (χ^2), al nivel de 5%, con un grado de libertad es de 3.84. Como 1.66 es menor que 3.84, se rechaza la hipótesis alterna (H_1), de acuerdo con la regla de decisión de la prueba *ji* cuadrada.

Conclusión El sexo de los aspirantes y las puntuaciones obtenidas en el examen de admisión son independientes entre sí; en otras palabras: no existe diferencia significativa entre los hombres y las mujeres respecto al examen de admisión.

Requisitos para usar la prueba *ji* cuadrada (χ^2)

1. *Aleatoriedad.* Los elementos que constituyen la muestra deben haber sido extraídos de manera aleatoria de la población de estudio.
2. *Independencia de las submuestras.* Las observaciones dentro de las muestras y entre las mismas (submuestras) deben ser mutuamente independientes.
3. *Distribución regular.* Dentro de cada categoría, las observaciones repetidas a través de todos los posibles valores se distribuyen normalmente alrededor del valor esperado para esa categoría. Así, el valor esperado es un tipo de media poblacional para los datos de frecuencia.
4. *Aleatoriedad.* Sólo se requieren frecuencias de casos o respuestas.
5. *Independencia de las muestras.* Esto requiere la presencia de un cuadro 2×2 como mínimo. La suposición de independencia indica que la χ^2 no puede aplicarse a una sola muestra (de panel antes-después); deben obtenerse por lo menos dos muestras de entrevistados.
6. Algunos estadísticos sugieren que no es adecuado el uso de la χ^2 cuando las frecuencias esperadas sean menores de cinco, aunque este requisito no constituye una restricción.

14.2.3.1 Coeficiente *fi* (Φ)

Cuando ambas variables son nominales y dicotómicas, es posible determinar el grado de asociación entre las variables de interés. Dicho coeficiente (Φ) también constituye un caso particular del coeficiente de correlación de Pearson, y se utiliza con cierta frecuencia, aunque no necesariamente con este propósito, en la elaboración y el análisis de pruebas. En secciones posteriores se considerará la independencia o dependencia de dos variables en una muestra determinada; cuando se realice este análisis, a partir de las hipótesis establecidas, si la conclusión indica que existe dependencia el análisis estadístico más pertinente es conocer el grado de asociación que implica la dependencia entre las variables o las muestras. Para determinar esto, se necesita un número que se establecerá a partir del coeficiente de correlación Φ (*fi*); sin embargo, se debe recordar que dicho número está supeditado al diseño 2×2, al tamaño de muestra y a la

proporción de las variables dicotomizadas. Cuando el número de casos en una variable es igual al de la otra variable, el coeficiente Φ tendrá el valor máximo de 1; cuando los totales marginales son diferentes no se alcanzará ese valor.

Ejemplo Se desea establecer una escala de medición de cierto rasgo de agresión en adultos. La interrogante al respecto consiste en determinar si existe relación entre el sexo de la persona y la respuesta (dicotómica) ante una situación que provoque agresividad. Así, se realiza un estudio con 400 personas, 200 del sexo masculino y 200 del femenino. Los datos son los siguientes: 160 personas del sexo masculino están de acuerdo y 40 en desacuerdo; por parte del sexo femenino, 40 están de acuerdo y 160 no lo están. Calcule el coeficiente Φ de correlación y concluya, con base en los resultados obtenidos.

Paso 1. Se presentan los datos obtenidos en una tabla de doble entrada de dos renglones y dos columnas (2 × 2).

	Acuerdo	*Desacuerdo*	*Total*
Masculino	160 a	40 b	200 $a + b$
Femenino	40 c	160 d	200 $c + d$
Total	200	200	400
	$a + c$	$b + d$	

Paso 2. Se aplica la definición de Φ.

$$\Phi = \frac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}}$$

donde:

a = las puntuaciones del primer renglón y la primera columna (en este caso, masculino y de acuerdo = 160).

b = las puntuaciones del primer renglón y segunda columna (masculino y en desacuerdo = 40).

c = las puntuaciones del segundo renglón y la primera columna (femenino y de acuerdo = 40).

d = Segundo renglón y segunda columna (femenino y en desacuerdo = 160).

Paso 3. Se sustituyen los datos anteriores en la fórmula anterior y se efectúan las operaciones.

$$\Phi = \frac{(160)(160) - (40)(40)}{\sqrt{(200)(200)(200)(200)}}$$

$$\Phi = \frac{25\ 600 - 1\ 600}{\sqrt{1\ 600\ 000\ 000}}$$

$$\Phi = \frac{24\ 000}{40\ 000} = 0.60$$

$$\boxed{\therefore\ \Phi = 0.60}$$

Paso 4. Se concluye teóricamente que el grado de asociación entre el sexo y la agresión es del 60%.

Prueba de significancia de *fi* (Φ)

Para poder comprobar la significancia de dicho coeficiente se utiliza la siguiente definición:

$$\chi^2 = n\Phi^2$$

donde:

n = Número total de casos (suma de los totales marginales, tanto de las columnas como de los renglones); en el ejemplo $n = 400$.
Φ = coeficiente *fi* = 0.60.
χ^2 = *ji* cuadrada, que guarda la siguiente relación:

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

Al sustituir los valores anteriores se obtiene:

$$\chi_c^2 = (400)(0.60)^2$$

$$= (400)(0.36) = 144$$

$$\boxed{\therefore\ \chi_c^2 = 144}$$

Este resultado se contrasta con el valor crítico de *ji* cuadrada, calculado mediante la siguiente regla de decisión RD:

$$\text{Si } \chi_c^2 \geq \chi^2_{crit}, \text{ entonces } \Phi \text{ es significativa.}$$

El valor crítico de la χ^2 está en función de los grados de libertad y el nivel de significancia (α), establecido de antemano. Para un diseño 2 × 2, los grados de libertad serán siempre igual a 1, pero, en general, para cualquier diseño diferente del mencionado se calcularán mediante la siguiente definición:

$$gl = (r - 1)\,(c - 1)$$

donde:

gl = grados de libertad.
r = número de renglones de la tabla de contingencia.
c = número de columnas de la tabla de frecuencias observadas.
En este caso se tiene $r = 2$ y $c = 2$ (dos renglones y dos columnas), o sea:
$gl = (2 - 1)(2 - 1) = 1 \times 1 = 1$.

Se compara el valor calculado de χ^2 con el valor crítico (véase la tabla χ^2 del Apéndice), así como con $gl = 1$, y se establecen los niveles de significancia α = 5% y α = 1%, cuyos valores son $\chi^2_{(5\%)} = 3.84$ y $\chi^2_{(1\%)} = 6.63$, respectivamente. La χ^2 calculada es lo bastante grande para ambos valores, lo que significa que el coeficiente Φ = 60% es significativo.

Requisitos de uso de Φ

A fin de utilizar adecuadamente el coeficiente Φ como medida de asociación entre las variables x y y, dicotomizadas, tienen que tomarse en cuenta las siguientes condiciones:

1. *Datos nominales.* Las variables x y y deben ser nominales y dicotomizables, ya que únicamente se requerirán las frecuencias observadas (el número de veces que ocurren en cierta nominación).
2. *Tabla de contingencia* 2 × 2. Los datos deben colocarse en un diseño de dos renglones por dos columnas. Es inadecuado aplicar el coeficiente Φ a diseños mayores del indicado, donde se comparan varios grupos o categorías.
3. *Muestreo aleatorio.* Para poder comprobar la significancia y validez de Φ, la muestra en estudio debe extraerse en forma aleatoria (todos los elementos de la población deben tener la misma posibilidad de ser escogidos).
4. *Un caso particular.* Cuando la muestra en estudio es pequeña se utilizará la definición de χ^2, pero con la corrección de Yates, también llamada *de Pirie-Handem.* Un criterio para definir la pequeñez es que la frecuencia Φ observada, en dos o más casillas, sea menor que 10. La definición de χ^2 consiste en lo siguiente:

$$\chi^2 = \frac{n\,[\,|ad - bc| - 0.5]^2}{(a+b)(a+c)(b+d)(c+b)}$$

donde:

$|ad - bc|$ = valor absoluto de la diferencia entre ad y bc;

y

$$\Phi = \sqrt{\frac{\chi_c^2}{n}}$$

14.2.3.2 Coeficiente de contingencia (*C*)

Es un coeficiente de correlación para datos nominales colocados en una tabla de contingencia con un diseño mayor que 2 × 2. Con éste se trata de determinar el grado de asociación, al comparar varios grupos o categorías, y puede calcularse mediante la siguiente definición:

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

donde:

C = Coeficiente de contingencia.
χ_c^2 = Valor calculado de *ji* cuadrada.
n = Número total de casos (tamaño de la muestra).

Ejemplo Se investiga la posible relación entre cierto tipo de defecto genético en recién nacidos y la edad de la madre al ocurrir el parto. Se necesita encontrar el grado de asociación, si existe dependencia entre ambas variables. El defecto genético se clasifica en cuatro tipos: *A*, *B*, *C* y *D*; la edad de la madre se clasificará en tres intervalos: menor de 25, entre 25 y 40 y mayor de 40 años. Para llevar a cabo este estudio se consideran 309 niños (n = 309) en los que se han diagnosticado defectos genéticos.

Paso 1. Los datos obtenidos al realizar este estudio se concentran en una tabla de contingencia:

	Padecimiento genético				
Edad materna	*A*	*B*	*C*	*D*	*Total*
Menor de 25 años	51	46	25	15	137
Entre 25 y 40 años	33	17	49	20	119
Mayor de 40 años	4	11	35	3	53
Total	88	74	109	38	309

Si las dos clasificaciones son independientes entre sí, la posibilidad (proporción relativa) de una casilla determinada será igual al producto de su total marginal del renglón por el total marginal de su columna, dividido entre n. A eso se le llama *frecuencia esperada* (fe).

Para calcular la χ^2 puede utilizarse la siguiente fórmula:

$$\chi^2 = \sum \frac{(fo - fe)^2}{fe}$$

donde:

fo = Frecuencia observada (se mide directamente).

fe = Frecuencia esperada (producto total marginal del renglón por el total marginal de la columna, dividida entre n).

Paso 2. Se calculan las frecuencias esperadas:

		1 *A*	2 *B*	3 *C*	4 *D*	*Total*
1	Menor de 25 años	51/39	46/32.8	25/48.3	15/16.8	137
2	Entre 25 y 40 años	33/33.9	17/28.5	49/42	20/14.6	119
3	Mayor de 40 años	4/15.1	11/12.7	35/18.7	3/6.5	53
	Total	88	74	109	38	309

Por ejemplo:

$$fe_{11} = \frac{137 \times 88}{309} = 39$$

$$fe_{23} = \frac{119 \times 109}{309} = 42$$

Sucesivamente se obtienen las demás frecuencias, hasta $fe\ (3) = \frac{53 \times 38}{309} = 6.5$

Paso 3. Para cada casilla se obtiene $fo - fe$, lo cual se eleva después al cuadrado en cada casilla a fin de establecer $(fo - fe)^2$.

De este modo, las casillas quedan en particular:

$$(fo_{11} - fe_{11})^2 = (51 - 39)^2 = (12)^2 = 144$$

$$(fo_{23} - fe_{11})^2 = (49 - 42)^2 = 49$$

Paso 4. Se divide cada uno de los resultados anteriores entre su respectiva frecuencia esperada (fe), y se suman los mismos.

$$\frac{144}{39} + \frac{174.24}{32.8} + \frac{542.89}{48.3} + \frac{3.24}{16.8} + \frac{0.81}{33.9} + \frac{132.25}{28.5} +$$

$$\frac{49}{42} + \frac{29.16}{14.6} + \frac{123.21}{15.1} + \frac{2.89}{12.7} + \frac{265.69}{18.7} + \frac{12.25}{6.5}$$

A continuación:

$$3.7 + 5.31 + 11.24 + 0.19 + 0.02 + 4.64 + 1.16 +$$

$$3.40 + 8.16 + 0.227 + 14.20 + 1.88 = 54.127$$

$$\boxed{\therefore \chi_c^2 = 54.127}$$

Paso 5. Se calcula el coeficiente C mediante la fórmula:

$$C = \sqrt{\frac{\chi_c^2}{n + \chi^2}}$$

$$C = \sqrt{\frac{54.127}{309 + 54.127}}$$

$$C = \sqrt{\frac{54.127}{363.127}}$$

$$C = \sqrt{0.149 = 0.386}$$

$$\boxed{\therefore C = 0.386}$$

Paso 6. Teóricamente el coeficiente de contingencia de la correlación entre el defecto genético del recién nacido y la edad de la madre es de 38.6%.

Prueba de significancia

La significancia estadística del coeficiente de contingencia se consigue a partir de la magnitud de la χ^2 obtenida con la siguiente regla de decisión: si $\chi_c^2 \geq \chi^2_{crit}$, entonces

C es significativo. El valor de la χ^2 crítica (χ^2_{crit}) se presenta en la tabla χ^2 del Apéndice, y se consideran los grados de libertad de la siguiente manera:

$$gl = (r - 1)\ (c - 1)$$

donde:

r = Número de renglones.
c = Número de columnas.

En el ejemplo:

$$gl = (3 - 1)\ (4 - 1) = 2 \times 3 = 6;\ gl = 6.$$

Así, el valor de χ^2_{crit} al nivel de significancia de 5% es 12.6 y al de 1% es 16.81, pero el valor de la χ^2 calculada es de 54.127, es decir, mayor que cualesquiera de los dos valores anteriores. En consecuencia, el coeficiente de contingencia calculado es estadísticamente significativo, por lo que se rechaza la hipótesis nula: los tipos de defectos congénitos dependen de la edad de la madre.

Requisitos para el uso del coeficiente de contingencia:

1. *Datos nominales.* Sólo se requieren datos de las variables dicotomizadas, las que se colocan en una tabla de contingencia $n \times m$.
2. *Muestreo aleatorio.* Con la finalidad de comprobar la significancia estadística del coeficiente de contingencia, la muestra se obtuvo en forma aleatoria.

14.2.3.3 Coeficiente *V* de Kramer

El coeficiente C presenta una desventaja: el número de correspondientes a la tabla de contingencia influirá en el valor obtenido de C. Esto significa que el coeficiente de contingencia no siempre variará entre 0 y 1 (por supuesto, nunca excederá de 1). En ciertas condiciones, el máximo valor de C podrá ser de 0.94 pero, en general, es de 0.89.

Para resolver este problema se emplea otro coeficiente de correlación, el V de Kramer, que expresa el grado de asociación entre variables nominales, pero en tablas mayores de 2×2. Dicho coeficiente se define de la siguiente manera:

$$V = \sqrt{\frac{\chi^2}{n\,(L - 1)}}$$

donde:

n = Total de casos (tamaño de la muestra).
L = Columna o renglón más pequeño de los dos.

Este coeficiente, a diferencia del C, tiene 1 como límite superior.

Como prueba de significancia, cuando la χ^2 que se utiliza al calcular V es estadísticamente significativa, dicho coeficiente también lo es.

Ejemplo A 350 estudiantes de una escuela preparatoria se les aplica una prueba que mide la habilidad verbal; los resultados de la anterior se correlacionarán con los correspondientes a historia de México, que se llevará a cabo al final del curso. Los resultados se muestran en la tabla de contingencias.

		Calificaciones en historia de México			
		10 – 8.1	8 – 6	Inferior a 6	*Total*
Habilidad verbal	Superior a 600	100	15	5	120
	Entre 500 y 599	50	85	15	150
	Inferior a 500	10	20	50	80
	Total	160	120	70	350

Paso 1. Se calculan las frecuencias esperadas (*fe*) mediante sus correspondientes totales marginales, que luego se dividen entre el total *n*: (54.9), (41.1), (24),..., (16).

100 (54.9)	15 (41.1)	5 (24)	120
50 (68.6)	85 (51.4)	15 (30)	150
10 (36.6)	20 (27.4)	50 (16)	80
160	120	70	350

Paso 2. Se calcula el valor de χ^2 mediante la fórmula conocida:

$$\chi^2 = \Sigma \frac{(f_e - f_e)^2}{f_e}$$

$$\chi_c^2 = \frac{(100 - 54.9)^2}{54.9} + \frac{(15 - 41.1)^2}{41.1} + \frac{(5 - 24)^2}{24} + \frac{(50 - 68.6)^2}{68.6} + \frac{(85 - 51.4)^2}{51.4} +$$

$$\frac{(15 - 30)^2}{30} + \frac{(10 - 36.6)^2}{36.6} + \frac{(20 - 27.4)^2}{27.4} + \frac{(50 - 16)^2}{16} = 196.6$$

$$\therefore \chi_c^2 = 196.6$$

Paso 3. Se sustituyen los valores de $n = 350$, $L = 3$, así como de la χ_c^2 obtenida anteriormente, en la definición de V.

$$V = \sqrt{\frac{\chi_c^2}{n(L-1)}}$$

$$V = \sqrt{\frac{196.6}{350(3-1)}}$$

$$V = \sqrt{\frac{196.6}{700}} = 0.53$$

$$\boxed{\therefore\ V = 0.53}$$

Paso 4. Teóricamente el coeficiente de Kramer de la correlación entre la habilidad verbal y las calificaciones que obtuvieron cada uno de los 350 estudiantes en historia de México es del 53%.

Prueba de significancia de *V*

De la misma manera que C, en el coeficiente V de Kramer se obtiene la prueba de significancia, o sea, si $\chi_c^2 \geq \chi^2_{crít}$, H_0 se rechaza y el coeficiente V es significativo.

Para $\alpha = 5\%$, $\alpha = 1\%$ y $\alpha = 0.005$ con $gl = (3-1)\ (3-1) = 4$, se obtienen:

$$\chi^2(4,\ 0.05) = 9.49$$

$$\chi^2(4,\ 0.01) = 13.28$$

$$\chi^2(4,\ 0.005) = 14.86$$

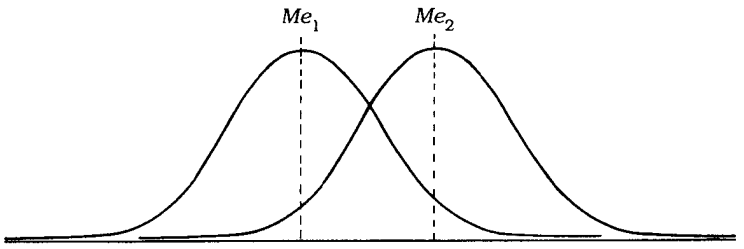
Como 196.6 >> 14.86, el coeficiente $V = 0.53$ es significativo.

14.3 PRUEBA DE LA *U* DE MANN-WHITNEY

Es posible emplear esta prueba como una alternativa de la paramétrica t de Student para comprobar la diferencia entre dos medias en dos muestras independientes. Las puntuaciones que representan mediciones, observaciones

o datos en general deben ser mutuamente independientes dentro de la misma muestra a la que pertenecen con respecto a otra; no es necesario que sean del mismo tamaño, pero sí se requiere que las variables que representan a dichas puntuaciones sean continuas. El uso de esta prueba se considera cuando las muestras son pequeñas y los requisitos de uso para la t resultan dudosos. La hipótesis nula establece la analogía (homogeneidad) de las dos distribuciones poblacionales, y en cierta manera la igualdad de las dos medias o medianas. No obstante, si las dos distribuciones poblacionales son más o menos similares, tanto en forma como en variabilidad, la U es una prueba excelente de la tendencia central. Pero debido a que la prueba se basa en categorías o clases, la medida de tendencia central más adecuada es la *mediana* (*Me*). Entonces, cuando se consideran los resultados de dicha prueba, lo primero en que se piensa es la comparación de las dos medianas.

Supóngase que hay dos grupos cuyas distribuciones se sobreponen, como se muestra en la siguiente figura:



Una vez jerarquizadas las puntuaciones combinadas de las distribuciones anteriores, asignando el rango 1 (uno) a la puntuación más baja y el rango más alto a la puntuación $n_x + n_y$, se obtienen a continuación la suma de las puntuaciones tanto de la distribución x como de la y. En este caso particular, la diferencia entre ambas sumas es pequeña.

Por consiguiente, la prueba de la U de Mann-Whitney se basa en esta diferencia. Se deduce de la probabilidad de obtener una suma de categorías de una distribución que difiere de la suma de rangos esperada conforme la hipótesis de igualdad de las dos distribuciones.

Ejemplo En dos grupos de estudiantes, cada profesor aplica una prueba de comprensión de lectura después de haber entrenado a los alumnos en dos técnicas diferentes. El tamaño de los grupos es diferente, $n_1 = 14$ y $n_2 = 13$, y los resultados obtenidos son los siguientes:

Grupo I	*Grupo II*
54	
50	45
48	40
47	32
46	30
45	30
43	30
42	29
38	28
26	28
25	38
22	31
21	28
26	27
$n_1 = 14$	$n_2 = 13$

Paso 1. Se etiquetan los dos grupos x y y; si uno contiene menos casos que el otro, debe llamarse x.

Paso 2. Se ordenan en forma creciente o decreciente las puntuaciones en una sola distribución de $n_x + n_y$ casos, jerarquizándolos a continuación. Se asigna el intervalo 1 a la puntuación menor, 2 al que sigue en forma creciente, y así sucesivamente hasta el caso $n_x + n_y$.

y *Grupo I*	x *Grupo II*	R_y	R_x
54		27	
50	45	26	21.5
48	40	25	18
47	38	24	16.5
46	32	23	15
45	31	21.5	14
43	30	20	12
42	30	19	12
38	30	16.5	12
26	29	4.5	10
26	28	4.5	8
25	28	3	8
22	28	2	8
21	27	1	6
$n_y = 14$	$n_x = 13$	$\Sigma R_y = 217$	$\Sigma R_x = 161$

Paso 3. Se comprueba si no se cometió algún error aplicando:

$$\frac{n(n+1)}{2} = \Sigma R_x + \Sigma R_y$$

donde $n = n_x + n_y = 13 + 14 = 27$; al sustituir los valores, se tiene:

$$\frac{27(28)}{2} = 161 + 217$$

$$\boxed{\therefore\ 378 = 378}$$

Paso 4. Se calcula la suma de los rangos de la distribución x, o sea ΣR_x.

Paso 5. Localice en la tabla el Apéndice de la U de Mann-Whitney, tanto el renglón y la columna que corresponden a $n_x = 13$, como a $n_y = 14$. Ése es el valor crítico de ΣR_x para probabilidades de 0.005, 0.01, 0.025 y 0.05 unidireccionales (una cola), para valores tanto de n_x como de n_y que van desde tres a 15.

Paso 6. Aplique uno de los tres siguientes criterios, según la hipótesis alternativa que haya establecido:

a) H_1: $x \neq y$. Encuentre un par de números correspondientes a la columna de $\alpha/2$ (si $\alpha = 0.05$, utilice la columna de 0.025). Rechace H_0 si ΣR_x es igual o menor que el número más pequeño o también si es igual o mayor que el número más grande.

b) H_1: $x < y$. Encuentre un par de números correspondiente a la columna de α (si $\alpha = 0.05$, entonces use esa misma columna). Rechace H_0 si ΣR_x igual o menor que el número más pequeño.

c) H_1: $x \geq y$. Encuentre el par de números correspondiente a la columna de α (si $\alpha = 0.05$, utilice esa misma columna). Rechace H_0 si ΣR_x es igual o superior al número más grande.

Paso 7. En este problema, para ejemplificar los tres criterios anteriores, se establecerá tanto la hipótesis nula como la alterna para *a)*, *b)* y *c)*, con $\alpha = 0.05$.

a) 141 y 213
H_0: $x = y$
H_1: $x \neq y$

Como $\Sigma R_x = 161$ no es menor ni igual que 141, ni mayor ni igual que 213, entonces no se rechaza H_0.

b) H_0 : $x \geq y$
H_1 : $x < y$
$\alpha = 0.05$, (147, 217)
$\Sigma R_x = 161$

Como ΣR_x no es igual ni menor que 147, entonces no rechazar H_0.

c) H_0: $x \leq y$
H_1: $x > y$
$\alpha = 0.05$, (147, 217)
ΣR_x no es igual ni menor que 217, entonces no rechazar H_0.

Conclusión Las dos distribuciones poblacionales son análogas.

Requisitos de uso de la U de Mann-Whitney

Los requisitos básicos para esta prueba son que la obtención de las muestras que representan a cada una de las poblaciones haya sido aleatoria y que no existan empates en los intervalos jerarquizados, aunque un número moderado de ellos no altera el resultado.

Prueba U de Mann-Whitney para muestras grandes

Por el teorema central del límite, una muestra grande tiende a distribuirse en forma normal, por lo que podrá utilizar el estadístico Z, definido por ΣR_x y n_x, n_y, y un factor de corrección (0.5):

$$Z = \frac{\Sigma R_x - 0.5\,[n_x\,(n_x + n_y + 1)]}{\sqrt{\dfrac{n_x\,n_y\,(n_x + n_y + 1)}{12}}}$$

- Valores positivos de Z implican que $x > y$.
- Valores negativos de Z implican que $x < y$.

Para propósitos de demostración, se aplicará la fórmula anterior con los datos del ejemplo anterior:

Datos:

$$n_x = 13.$$
$$n_y = 14.$$
$$\Sigma R_x = 161.$$

Al sustituir resulta:

$$Z_c = \frac{161 - 0.5\,[13\,(13 + 14 + 1)]}{\sqrt{\dfrac{13 \times 14\,(13 + 14 + 1)}{12}}}$$

$$Z_c = \frac{161 - 0.5\,[364]}{\sqrt{\dfrac{182\,(28)}{12}}} = \frac{161 - 182}{\sqrt{424.6}} = \frac{-21}{20.61}$$

$$\boxed{\therefore\ Z_c = -1.02}$$

RD: Puesto que $Z_{0.05} = \pm 1.96$, si $Z_c \geq Z_{0.05}$, entonces se rechaza H_0; o también si $-Z_c \leq Z_{0.05}$. Como -1.02 no es menor que -1.96, no se rechaza H_0.

Conclusión Las dos distribuciones poblacionales son análogas.

Ejemplo A dos grupos de sujetos, grupos $n_x = 23$ (fumadores de mariguana) y $n_y = 24$ (no fumadores de mariguana), se les aplica una prueba que mide las habilidades en la conducción de automóviles (hábitos de conducir, velocidad, estado de alerta, entre otras), la que arroja los siguientes resultados. Compruebe si existen diferencias significativas entre ambos grupos.

x Grupo I	*y* Grupo II
	42
37	41
35	41
35	40
34	39
34	38
30	37
29	37
28	36
27	35
27	35
26	33
25	32
25	32
24	31
23	30
22	29
21	28
20	27
19	26
19	25
18	24
14	22
14	20
$n_x = 23$	$n_y = 24$

Paso 1. Se jerarquizan las puntuaciones en una sola distribución, después se separan en las dos originales y se obtiene la suma de los rangos de cada una de ellas, o sea ΣR_x y ΣR_y.

R_x	R_y
	47.0
40.0	45.5
37.5	45.5
35.0	44.0
33.0	43.0
28.5	42.0
26.5	40.0
24.5	40.0
22.5	37.5
20.0	35.0
20.0	35.0
17.5	32.0
15.0	30.5
15.0	30.5
12.5	28.5
11.0	26.5
9.5	24.5
8.0	22.5
7.0	20.0
5.5	17.5
5.5	15.0
3.5	12.5
2.0	9.5
1.0	3.5
$\Sigma R_x = 400.5$	$\Sigma R_y = 727.5$

Como es una muestra grande:

$$n_x = 23$$
$$n_y = 24$$
$$\Sigma R_x = 400.5$$

Se aplica

$$Z_c = \frac{\Sigma R_x - 0.5\,[n_x\,(n_x + n_y + 1)]}{\sqrt{\dfrac{n_x n_y\,(n_x + n_y + 1)}{12}}}$$

$$Z_c = \frac{400.5 - 0.5\,[23\,(23 + 24 + 1)]}{\sqrt{\dfrac{23 \times 24\,(23 + 24 + 1)}{12}}}$$

$$Z_c = \frac{400.5 - 552}{\sqrt{2208}} = \frac{-151.5}{47} = -3.22$$

$$\boxed{\therefore\ Z_c = -3.22}$$

RD: Si $Z_c \leq Z_{0.01}$, entonces H_0 se rechaza, o también si $|Z_c| \geq Z_{0.01}$.
Como $-3.22 < -2.58$, se rechaza H_0. Como Z es negativo, entonces $x < y$.

Conclusión Existen diferencias estadísticamente significativas entre el grupo I y el grupo II.

14.4 PRUEBA DE RANGOS CON SIGNOS EN PARES DE WILCOXON

(The Wilcoxon's matched-pair signed-ranks)

Esta prueba se utiliza para comparar (contrastar) dos muestras dependientes o relacionadas.

Se obtienen las diferencias absolutas, restando la puntuación del grupo 1 con la correspondiente al grupo 2 para cada uno de los sujetos. Cuando se realiza en una sola muestra, la diferencia es de la puntuación antes con la de después (antes-después). Se ordenan dichas diferencias; a la mínima se le asigna el primer lugar y a la máxima el último. Cuando se obtienen dos o más diferencias iguales, se promedian sus rangos (denominados *empates*). Cuando ya están jerarquizadas las diferencias, se les adjudica su signo algebraico (+ o −), y se obtienen sus sumas ΣR_+ y ΣR_-. Si se cumple $\Sigma R_+ = \Sigma R_-$ y además es igual a $\frac{n(n+1)}{2}$, la jerarquización se ha realizado correctamente.

El valor menor de ΣR_+ o ΣR_- se compara con el valor crítico de la tabla (ver Apéndice), con la siguiente regla de decisión.

Si $\Sigma R_{i\,min} \leq$ valor tablas, H_0 se rechaza.

Otro método es el siguiente: cuando $n > 8$, se utiliza la aproximación normal.

$$Z = \frac{\Sigma R_i - \mu_w}{\sigma_w}$$

donde:

ΣR_i = Suma del rango + o −.
μ_w = Media aritmética de los rangos.
σ_w = Desviación estándar de los rangos.
Los pasos que se deben seguir son los siguientes:

Paso 1. Se identifica la variable de interés (x). Si no se distribuye en forma normal (no cumple los requisitos para utilizar una prueba paramétrica) tal como

la de Sandler o t de Student para muestras dependientes o una sola muestra medida dos veces, se utilizará una prueba no paramétrica.

Paso 2. Se proponen las hipótesis nulas y sus correspondientes alternativas:

H_0: $x_1 = x_2$	Al emplear los rangos:
H_1:	H_0: $\Sigma R_+ = \Sigma R_-$
a) $x_1 > x_2$	H_1:
b) $x_1 < x_2$	*a)* $\Sigma R_+ > \Sigma R_-$
c) $x_1 \neq x_2$	*b)* $\Sigma R_+ < \Sigma R_-$
	c) $\Sigma R_+ \neq \Sigma R_-$

Paso 3. Se establece el nivel de significancia α: 5%, 1%, ...

Paso 4. Si se cumplen los requisitos, se utiliza la prueba no paramétrica de rangos de signos en pares de Wilcoxon.

a) Si $n > 8$, se empleará: $Z = \dfrac{\Sigma R_i - \mu_w}{\sigma_w}$

b) Si $n < 8$, se empleará: $\Sigma R_i = \dfrac{n(n+1)}{2}$

Paso 5. Consultar la tabla (Z) o tabla 2 del Apéndice, según sea el caso, para establecer la región crítica de rechazo de H_0.

Paso 6. Si $n < 8$, la regla de decisión es la siguiente:

Se compara el valor absoluto de la menor suma de rangos $\Sigma R_{i\,mín}$, con el valor crítico (tabla correspondiente).

Si $\Sigma R_{i\,mín} \leq$ valor tablas, entonces H_0 se rechaza.

Paso 7. Sustituir los valores y calcular el valor de $\Sigma R_{i\,mín}$ o de Z, según sea el tamaño de n.

Paso 8. Obtener conclusiones.

Ejemplo En una institución educativa se seleccionan al azar los estudiantes del mismo grado escolar que obtuvieron igual promedio de calificaciones en matemáticas. De este grupo, se forman n pares de estudiantes y se ubican en forma aleatoria en dos grupos, a los que se enseñará un tema nuevo de matemáticas con el método A y con el método B, respectivamente. Al evaluar el aprendizaje del tema se obtienen las puntuaciones siguientes:

	Método	
Par	*A*	*B*
1	20	25
2	26	29
3	31	28
4	42	37
5	35	40
6	19	29
7	33	41
8	38	43
9	29	21
10	27	35
11	40	47
12	37	41

Paso 1. No se tienen bases para considerar que las puntuaciones que miden el aprendizaje de un tema específico se distribuyen de manera normal. Por ese motivo se utilizará la prueba no paramétrica de rangos de signos —en pares— de Wilcoxon.

Paso 2. H_0: $\Sigma R_+ = \Sigma R_-$
H_1: $\Sigma R_+ \neq \Sigma R_-$

o sea:

H_0: No existe diferencia en el aprendizaje de este tema de matemáticas con el método *A* y el método *B*.
H_1: H_0 es falsa.

o sea

H_0: No existe diferencia en el aprendizaje de este tema de matemáticas con el método *A* y el *B*.

H_1: H_0 es falsa.

Paso 3. $\alpha = 5\%$

Paso 4. Como $n > 8$ ($n = 12$ pares), se utilizará:

$$\text{(I) } \Sigma R_i = \frac{n(n+1)}{2}, \text{ para fines didácticos también se utilizará}$$

$$\text{(II) } Z = \frac{\Sigma R_{i\,mín} - \mu_w}{\sigma_w}$$

Paso 5. Para (I) $R(n, \alpha) = R_{(12,\ 0.05)} = 14$

$$\text{y para (II)} \ \left|Z_{0.025}\right| = 1.96$$

Paso 6. Regla de decisión (RD) para:

$$\text{(I) Si } \Sigma R_i < 14,\ H_0 \text{ se rechaza.}$$

$$\text{(II) Si } |\ Z\ | \geq 1.96,\ H_0 \text{ se rechaza.}$$

Paso 7. Para (I):

a) Se obtienen las diferencias de las puntuaciones del método *A* y del *B*.
b) Se jerarquizan dichas diferencias, para lo cual se considera su valor absoluto, pero se conserva su signo original en los rangos obtenidos.
c) Se obtienen ΣR_+ y ΣR_-

d) Se calcula $\dfrac{n(n+1)}{2}$

Para (II): se calculan μ_w, σ_w y Z

$$\sigma_w \sqrt{\frac{n(n+1)(2n+1)}{24}}, \qquad \mu_w = \frac{n(n+1)}{4} \qquad \text{y} \qquad Z = \frac{\Sigma R_{i\ min} - \mu_w}{\sigma_w}$$

donde $\Sigma R_{i\ min}$, puede ser ΣR_+ o ΣR_-; y se obtiene para *a)* y *b)*

Método				
Par	*A*	*B*	*Diferencia*	*Rango*
1	20	25	−5	−5.5
2	26	29	−3	−1.5[a]
3	31	28	3	1.5[a]
4	42	37	5	5.5
5	35	40	−5	−5.5
6	19	29	−10	−12.0
7	33	41	−8	−10.0
8	38	43	−5	−5.5
9	29	21	8	10.0
10	27	35	−8	−10.0
11	40	47	−7	−8.0
12	37	41	−4	−3.0

[a] La diferencia absoluta es 3 en ambos casos, y es la mínima. Por eso les corresponde el rango 1 y 2; al promediar estos valores se tiene $\frac{1+2}{2} = \frac{3}{2} = 1.5$ y se les antepone el signo correspondiente.

e) Se obtienen:

$$\Sigma R_- = 61$$

y

$$\Sigma R_+ = 17$$

Paso 8.
CASO I. Aplicar la regla de decisión para el valor más pequeño de ΣR_+ o ΣR_- de la siguiente manera: como $\Sigma R_{i\,mín} = 17$ y $17 > 14$, H_0 no se rechaza.

CASO II. Como $n = 12$ ($n > 8$), se puede aplicar la aproximación Z, por lo que:

$$\mu_w = \frac{n(n+1)}{4} = \frac{12 \times 13}{4} = 39$$

$$\mu_w = 39$$

$$\sigma_w = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{12 \times 13 \times 25}{24}} = \sqrt{162.5}$$

$$\boxed{\sigma_w = 12.74755}$$

Sustituyendo en $Z = \dfrac{\Sigma R_i - \mu_w}{\sigma_w}$

Si utilizamos $\Sigma R_- = 61$; $Z = \dfrac{61 - 39}{12.74755} = 1.72582$

Si utilizamos $\Sigma R_+ = 17$; $Z = \dfrac{17 - 39}{12.74755} = -1.72582$

CASO III. Si $|Z_c| \geq Z_{0.025}$, H_0 se rechaza.
como $\alpha = 5\%$, $Z_\alpha/2 = Z_{0.025} = 1.96$, así $|Z_c| = 1.72582$
$1.72582 < 1.96$, H_0 no se rechaza.
En los resultados obtenidos al aplicar tanto el caso I como el II, H_1 se rechaza (H_0 no se rechaza).

Conclusión Los métodos de enseñanza *A* y *B* conducen al mismo resultado; no existe diferencia estadísticamente significativa entre ambos.

14.5 PRUEBA DE McNEMAR

Para el caso de una misma muestra, esta prueba mide una situación antes y después de posibles cambios. También se aplica el modelo para dos muestras

dependientes o correlacionadas, obtenidas en forma aleatoria de una misma población, midiendo de igual manera los posibles cambios antes y después.

Ejemplo Para un grupo de pacientes que recibe determinado tratamiento (el mismo grupo de pacientes es su control), la variable de interés sería actitud positiva o negativa, bienestar o malestar, acuerdo o desacuerdo, alivio o no alivio, cumplimiento de la norma establecida o no, entre otros aspectos.

En el caso más simple se establecería la hipótesis nula (H_0) en términos de igualdad de proporciones correlacionadas; el grupo se mide antes del tratamiento y después del mismo.

La variable es cualitativa (nominal) y dicotómica. La proporción (número de sujetos), que pertenecen a una categoría son mutuamente excluyentes con la proporción de las otras categorías; dichas categorías pueden representarse en una tabla de contingencia (doble entrada) como la siguiente:

		Situación 2		
		+	−	*Total*
Situación 1	+	a	b	$a + b$
	−	c	d	$c + d$
	Total	$a + c$	$b + d$	N

En el anterior, a, b, c, d son las probabilidades de intersección (*joint*) y $a + b$, $c + d$, $a + c$ y $b + d$ son las probabilidades marginales. El modelo estadístico por emplear, si el número de casos de $c + b$ es mayor que 20, será una aproximación a la prueba z.

$$z = \frac{c - b \pm 1}{\sqrt{c + b}}, \quad \text{si:}$$

1) $c > b$, el signo es negativo (–).
2) $c \leq b$, el signo es positivo (+).

Si $c + b \leq 20$, entonces el valor de b se compara con un valor crítico en la tabla.

Las hipótesis se establecen en la siguiente forma:

$H_0 : c = b$

Para la hipótesis alterna:

1) $H_1 : c > b$
2) $H_1 : c < b$
3) $H_1 : c \neq b$

El valor crítico se obtiene en la tabla Z, con la siguiente regla de decisión:

Si $z_{calc} \geq z_{tablas}$, H_0 se rechaza.

Ejemplo Unas flores recibieron un nuevo tipo de mantenimiento, que se evaluará mediante un experimento antes-después en el que se aplicará la prueba de Mc Nemar. Las observaciones proporcionaron los siguientes resultados:

		Después		
		Rosa	Clavel	
Antes	Rosa	20 (a)	18 (b)	38
	Clavel	6 (c)	38 (d)	44
		26	56	82

Las hipótesis estadísticas son las siguientes:

H_0: el mantenimiento de la rosa es igual antes y después.
H_1: el mantenimiento de la rosa es diferente antes y después.
Se desea realizar la prueba con un $\alpha = 5\%$.
La región de rechazo para H_0 es:

Es decir, si $|z_c| \geq z_{(0.025)} = 1.96$, H_0 se rechaza.
Como $b + c > 20$ y $b > c$ se utiliza la siguiente fórmula:

$$z = \frac{(c - b) + 1}{\sqrt{c + b}}$$

Se sustituye:

$$z = \frac{(6 - 18) + 1}{\sqrt{24}}$$

$$z_c = \frac{-12 + 1}{\sqrt{24}} = \frac{-11}{4.9}$$

$$\boxed{z_c = -2.24}$$

Se aplica posteriormente la regla de decisión:
$z_c = |-2.24| = 2.24$
$2.24 > 1.96 = z_{0.025}$
Entonces, H_0 se rechaza.

Conclusión El mantenimiento de la rosa es diferente antes y después.

14.6 PRUEBA DE LA MEDIANA

Cuando se necesita comparar dos o más poblaciones independientes y no es posible utilizar las medias aritméticas, existe la opción de emplear la mediana, otra medida de tendencia central.

A fin de aplicar esta prueba, que contrastaría la hipótesis nula de que la mediana poblacional es la misma para todas las poblaciones que intervienen, resulta indispensable que dichas poblaciones tengan una distribución continua y que las muestras que las representan se hayan selecionado de manera aleatoria.

Al aplicar los ocho pasos, resulta:

Paso 1. Se agrupan los datos de los k grupos en uno solo y se calcula la mediana (Me).

Paso 2. Se proponen las siguientes hipótesis:

H_0: $Me_1 = Me_2 = \ldots = Me_k = Me_0$
H_1: Al menos dos poblaciones tienen mediana diferente.

Paso 3. Se establece el valor de α.

Paso 4. Una vez obtenida la mediana, la frecuencia de los datos se expone en una tabla de contingencia de la siguiente manera:

	Muestras (k)				*Total*
Datos menores o iguales a la *Me*	$f_{o_{11}}$	$f_{o_{12}}$	. . .	$f_{o_{1k}}$	a
Datos mayores que la *Me*	$f_{o_{21}}$	$f_{o_{22}}$	. . .	$f_{o_{2k}}$	b
TOTAL	n_1	n_2	. . .	n_k	N

Se aplica el modelo estadístico χ^2

$$\chi^2 = \Sigma \frac{(f_{oi} - f_{ei})^2}{f_{ei}} \quad \text{si } a \approx b$$

Paso 5. Se obtienen los grados de libertad, considerando el número de renglones y de columnas de la tabla de contingencia, y se lee el valor crítico de χ^2, según α y gl.

Paso 6. Se considera la siguiente regla de decisión si $\chi^2_{calc} \geq \chi^2$ (α, gl), entonces H_0 se rechaza.

Paso 7. Se calcula el valor de χ^2.
Paso 8. Si no se rechaza H_0, entonces las poblaciones anteriores tienen la misma mediana.

Cuando son dos poblaciones, en el paso 4 se puede utilizar el siguiente modelo:

$$\chi^2 = \frac{N\left[\,|ad - bc| - \frac{N}{2}\right]^2}{klmn}$$

El correspondiente cuadro de contingencia sería el siguiente:

	Población		
	1	2	*Total*
Menor o igual a la *Me*	a	*b*	*k*
Mayor que la *Me*	*c*	*d*	*l*
Total	*m*	*n*	*N*

Ejemplo A dos grupos de estudiantes de la misma escuela y del mismo grado (referidos como $n_1 = 20$ y $n_2 = 17$), se les enseña una técnica odontológica con dos métodos diferentes (I y II). Ambos grupos se evalúan y se obtienen las siguientes puntuaciones:

n_1 *Método I*		n_2 *Método II*	
48	27	54	36
47	27		32
46	26	50	30
46	25	48	25
44	24	47	25
40	18	46	22
38	18	45	20
32	16	45	14
30	12	43	42
28	10	$n_2 = 17$	
$n_1 = 20$			

Paso 1. Se combinan todas las puntuaciones en una sola distribución y se ordenan en forma creciente o decreciente.

$$n = n_1 + n_2 = 20 + 17 = 37$$

Paso 2. Se calcula la mediana mediante la fórmula de datos no agrupados (véase el capítulo 2). En el ejemplo: $Me = 32$.

Paso 3. Para cada distribución (métodos I y II), se cuenta el número de casos que se sitúan en el rango superior e inferior de la mediana, y se presentan los valores en un cuadro de contingencia.

	Método I	*Método II*	*Total*
Arriba de la mediana	8 (*a*)	11 (*b*)	19
Abajo de la mediana	12 (*c*)	6 (*d*)	18
	20	17	37

Paso 4. Se calcula χ^2 mediante la siguiente fórmula:

$$\chi_c^2 = \frac{n\,[\,|ad - bc| - 0.5]^2}{(a+b)(a+c)(b+d)(c+d)}$$

$$= \frac{37\,[\,|(8)(6) - (11)(12)| - 0.5]^2}{(19)(20)(17)(18)}$$

$$= \frac{37[(84) - 0.5]^2}{116\,280} = \frac{257\,979.25}{116\,280} = 2.22$$

$$\boxed{\therefore\ \chi_c^2 = 2.22}$$

Paso 5. Se compara este valor con el observado en la tabla de χ^2 (véase la pág. 644) con $gl = 1$, y a un nivel de significación de 5%, y como $2.22 < 3.841$, se rechaza la hipótesis alternativa en el sentido de que las dos medianas provienen de pobla- ciones diferentes.

Conclusión Como la χ^2 no es significativa (ni siquiera en el nivel de 10%, donde $\chi_t^2 = 2.706$), se acepta la hipótesis de que las medianas proceden de la misma población. En otros términos, los dos métodos de enseñanza son iguales. Esta prueba puede aplicarse a más de dos grupos al utilizar tablas de contingencia de tipo 2×2.

Cuando el número de frecuencias esperadas es pequeño (menor que 2), no es conveniente utilizar la χ^2, sino la prueba exacta de Fisher (véase Ferguson,

1971, pp. 340-342). Es necesario considerar que la potencia de la prueba de la mediana es baja, por lo que se desarrolla una prueba más potente.

14.7 PRUEBA DE KRUSKAL-WALLIS (*H*)

Cuando se desea comparar más de dos grupos independientes se utiliza una extensión de la *U* de Mann-Whitney: el análisis de varianza en una dirección de Kruskal-Wallis. Este procedimiento es la contraparte no paramétrica de la prueba *F* (análisis de varianza paramétrico), la cual, aunque muy sólida, presenta muchas dificultades para llevarla a la práctica si las condiciones de aplicación no se cumplen en su totalidad. Por eso, si no se cumplen los requisitos de distribución para la prueba *F*, la alternativa más adecuada es la prueba *H*. Los requisitos básicos para aplicarla son:

1. La variable dependiente es numérica por naturaleza.
2. La variable dependiente está distribuida continuamente.
3. Hay legitimidad en la independencia de las mediciones. (Tanto los métodos no paramétricos como los paramétricos son sensibles a la violación de este último requisito.)

En el supuesto de que todos los requisitos se cumplen y el investigador quiere contrastar la hipótesis nula que establece que las *n* muestras provienen de una misma población, se demuestra que, si proviene de la misma población, la prueba *H* sigue el modelo (*ji* cuadrada) de χ^2 con $gl = k - 1$. La prueba de Kruskal-Wallis es un diseño completamente aleatorizado, debido a que las unidades experimentales son asignadas en forma completamente aleatoria a los *k* tratamientos.

Ejemplo Cuatro grupos de estudiantes de secundaria, del mismo grado académico, cursan matemáticas con diferente profesor y tienen un promedio alto de calificación en la materia (por lo que las muestras no necesariamente deben ser del mismo tamaño: $n_1 = 5$, $n_2 = 7$, $n_3 = 7$ y $n_4 = 10$). A dichos grupos se les aplica una prueba psicológica de aptitudes con respecto al aprendizaje de las matemáticas, cuyos resultados son los siguientes:

n_1	n_2	n_3	n_4
12	14	18	14
8	12	15	13
6	12	12	12
4	9	10	12
2	7	8	7
	6	8	6
	2	7	6
			5
			3
			1

Para comprobar las hipótesis:

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ Provienen de la misma población.*

H_1: al menos una μ es diferente (H_0 falsa). Provienen de poblaciones diferentes.

Así, se efectuarán los siguientes pasos:

Paso 1. Se jerarquizan todos los datos dentro de una sola distribución, y se asigna a la puntuación más baja el rango de 1; después se ubican de nuevo los valores jerarquizados en sus respectivos grupos, y se suman los rangos en cada grupo.

Rangos

R_1	R_2	R_3	R_4
21.5	26.5	29.0	26.5
15.0	21.5	28.0	25.0
8.5	21.5	21.5	21.5
5.0	17.0	18.0	21.5
2.5	12.0	15.0	12.0
$\Sigma R_1 = 52.5$	8.1	15.0	8.5
	2.5	12.0	8.5
	$\Sigma R_2 = 109.5$	12.0	6.0
		$\Sigma R_3 = 138.5$	4.0
			1.0
			$\Sigma R_4 = 134.5$

Paso 2. Se aplica la siguiente fórmula:

$$H = \frac{12}{n\,(n+1)} \left[\frac{\Sigma R_j^2}{n_k} \right] - 3\,(n+1)$$

donde:

ΣR_j^2 = Suma de los rangos de cada grupo, elevados al cuadrado.

n_k = Número de mediciones (elementos) de cada grupo.

n = Suma total de n grupos.

12 = Constante.

3 = Constante.

En el ejemplo los datos son:

$n_1 = 5,\ n_2 = 7,\ n_3 = 7,\ n_4 = 10;\ \therefore\ n = 29$

Al sustituir los resultados de los pasos 1 y 2 en la fórmula, se tiene:

$$H = \frac{12}{29\,(29+1)} \left[\frac{(52.5)^2}{5} + \frac{(109.5)^2}{7} + \frac{(138.5)^2}{7} + \frac{(134.5)^2}{10} \right] - 3\,(30)$$

* Algunos autores consideran la comparación entre las medianas de las poblaciones.

$$H = \frac{12}{870}\left[\frac{2\,756.25}{5} + \frac{11\,990.25}{7} + \frac{19\,182.25}{7} + \frac{18\,090.2}{10}\right] - 90$$

$$H = 0.01379\left[551.25 + 1\,712.8 + 2\,740.3 + 1\,809.0\right]^{5} - 90$$

$$H = 0.01379\,[6813.485] - 90;$$

$$H_c = 93.9579 - 90 = 3.9579;\ H_c = 3.9579$$

Paso 3. Se interpreta H como una *ji* cuadrada (χ^2), con grados de libertad iguales al número de muestras menos 1 ($gl = k - 1$). Como $gl = 3$, a un nivel de confianza de 5%, se observa que $\chi^2_{3,(0.05)} = 7.815$.

RD: Si $H_c \geq \chi^2_{(gl,\alpha)}$, se rechaza H_0.

Como $3.9579 < 7.815$, *no* se rechaza H_0 y los grupos de estudiantes provienen de la misma población (no hay diferencia estadísticamente significativa en las medias aritméticas).

Ejemplo Suponga que se estudiarán tres tipos de estilos (*A*, *B* y *C*), por medio de su elección y que la calificación de elección es la que se registra en la siguiente tabla. Las diferencias se examinarán de acuerdo con $\alpha = 0.01$.

Tipos de estilo

A		*B*		*C*	
Calif.	*Rango*	*Calif.*	*Rango*	*Calif.*	*Rango*
22	3	26	8	23	4
24	5	36	16	30	12
21	2	37	10	29	11
17	1	37	17	26	8
26	8	33	15	31	13
25	6	32	14		
$n_i = 6$ $\Sigma R_i = 25$		$n_i = 6$ $\Sigma R_i = 80$		$n_i = 5$ $\Sigma R_i = 48$	

$$\Sigma R = 25 + 80 + 48 = 153$$

$$\Sigma R = \Sigma n_i(\Sigma n_i + 1)/2 = 17(17 + 1)/2 = 153$$

$$H_c = -3\,(\Sigma n_i + 1) + \frac{12}{\Sigma n_i\,(\Sigma n_i + 1)}\;\frac{\Sigma R_1^2}{n_i} + \frac{\Sigma R_2^2}{n_2} + \ldots$$

$$= -3\,(17 + 1) + \frac{12}{17\,(17 + 1)}\;\frac{(25)^2}{6} + \frac{(80)^2}{6} + \frac{(48)^2}{5}$$

$$= -54 + 0.0392\,[1631.63]$$

$$\boxed{\therefore\; H_c = 9.96}$$

$gl = k - 1 = 3 - 1 = 2.\chi_c^2 = 9.21$, *o sea* $\chi^2(2,0.01) = 9.21$. Si $H_c > \chi_c^2$, rechazar H_0; como $9.96 > 9.21$, H_0 se rechaza al 1%.

14.8 PRUEBA DE FRIEDMAN

Este análisis de varianza no paramétrico es una extensión de la prueba no paramétrica de rangos de signos en pares o pareados, de Wilcoxon, que se utiliza cuando se comparan dos muestras dependientes, correlacionadas, o una sola muestra medida dos veces. También se le conoce como *de dos direcciones por rangos*.

El término *dos direcciones* se refiere a los niveles del factor tratamiento (k) y a las situaciones diferentes a las que se someten los sujetos.

La prueba de Friedman es un diseño de bloques aleatorios, donde las kn unidades experimentales pueden dividirse en n grupos homogéneos (bloques), en el que cada bloque contiene k unidades; entonces $N = kn$.

Existen dos situaciones en las que se aplica esta prueba:

1. Para una misma muestra medida más de dos veces.
2. Una sola medición de más de dos muestras dependientes o correlacionadas.

En el caso 1, k significa el número de tratamientos o las diferentes situaciones en las que miden a los sujetos, mientras que en el 2, k es el número de muestras o grupos correlacionados que participan en el estudio.

En el caso 1, cada sujeto sirve como su propio control, incrementando la posibilidad de observación de las diferencias significativas entre los distintos tratamientos; por ello, la hipótesis nula se plantea así:

H_0: las puntuaciones obtenidas en cada medición son iguales a las que se registran a las evaluaciones.

Ejemplo Se realiza un estudio con un grupo de niños con la misma escolaridad, escogidos al azar de una misma escuela (población); se les aplica una prueba psicológica que mide hostilidad en tres situaciones diferentes (programas de

televisión con niveles de violencia baja, moderada y alta). Las puntuaciones de hostilidad fluctúan desde 20 hasta 60; las más altas representan mayor hostilidad. Los resultados de la prueba son los siguientes:.

Violencia en un programa de televisión

Niño	*Baja*	*Moderada*	*Alta*
A	23	30	32
B	41	45	43
C	36	35	39
D	28	29	35
E	39	41	47
F	25	28	27
G	38	46	51
H	40	47	49
I	45	46	42
J	29	34	38

Aplicamos los ocho pasos y resulta:

Paso 1. La variable de interés es la hostilidad (h).

Paso 2.

$$H_0 : h_B = h_M = h_A$$

$$H_1 : h_B \neq h_M \neq h_A$$

Paso 3.

$$\alpha = 5\%$$

Paso 4. Análisis de varianza no paramétrico (prueba de Friedman).

$$\chi_r^2 = \frac{12}{kn(k+1)} \left[\sum_{i=1}^{k} (\Sigma R_i)^2 \right] - 3n(k+1)$$

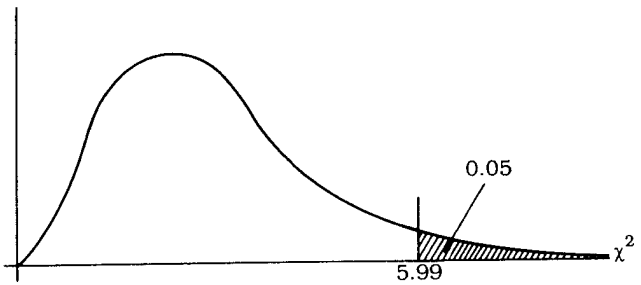
donde:

n = número de sujetos
k = número de tratamientos (mediciones en cada sujeto)
ΣR_i = suma de los rangos de cada tratamiento
12 = una constante.

Paso 5. El estadístico χ_r^2 se contrasta con la $\chi^2_{(gl,\alpha)}$ tabla correspondiente, donde $gl = k - 1$; para este ejemplo, se tiene:

$$k = 3 \text{ y } \alpha = 5\%$$

$$\chi^2_{(2,0.05)} = 5.99$$



Paso 6. La regla de decisión (RD) es:

$$\text{si } \chi_r^2 \geq 5.99 \Rightarrow H_0 \text{ se rechaza.}$$

Paso 7. Para obtener χ_r^2, se realizan las siguientes etapas.

Etapa 1 Se jerarquizan las puntuaciones (se obtienen los rangos), por sujeto; al valor más alto se le asigna 1; al moderado, 2; y al bajo, 3. Se calculan ΣR_B, ΣR_M, y ΣR_A.

Niño	R_B	R_M	R_A
A	3	2	1
B	3	1	2
C	2	3	1
D	3	2	1
E	3	2	1
F	3	1	2
G	3	2	1
H	3	2	1
I	2	1	3
J	3	2	1
ΣR	28	18	14
$(\Sigma R)^2$	784	324	196

Etapa 2 Sustituya los valores anteriores, así como $n = 10$ y $k = 3$

$$\chi_r^2 = \frac{12}{kn(k+1)} [\Sigma (\Sigma R_i)^2] - 3n(k+1)$$

$$\chi_r^2 = \frac{12}{3 \times 10 \times 4} \left[784 + 324 + 196\right] - 3 \times 10 (4)$$

$$\chi_r^2 = \frac{1}{10} [1\ 304] - 120 = 130.4 - 120$$

$$\boxed{\chi_r^2 = 10.40}$$

Paso 8. Como 10.40 > 5.991 y
10.40 > 9.210 ⇒
H_0 se rechaza para $\alpha = 5\%$ y $\alpha = 1\%$.

Conclusión Existen diferencias estadísticamente significativas respecto a la hostilidad, de acuerdo con el nivel de violencia.

14.8.1 Prueba de Nemenyi

Cuando la prueba de Friedman (χ_r^2) resulta significativa, se necesita efectuar comparaciones por pares de los k niveles de la variable independiente (tratamiento) para conocer cuál es el más efectivo.

Existen ${}_kC_2$ comparaciones posibles (pares) entre la medida de los rangos de los tratamientos.

$$\overline{R}_1 - \overline{R}_2,\ \overline{R}_1 - \overline{R}_3, \ldots$$

A continuación se obtiene la diferencia mínima (*DM*) por medio de:

$$DM = \sqrt{\frac{x_{(gl,\alpha)}\,[k(k+1)]}{6\,n}}$$

donde $\chi(gl, \alpha)^2$ es el valor crítico de tablas cuando χ_r^2 es significativa, k es el número de tratamientos, n de sujetos y el 6 es una constante, cuando la diferencia entre dos tratamientos (representada por la diferencia de sus rangos promediados) es mayor que la *DM*, entonces el tratamiento que hace la diferencia es el más efectivo.

En el ejemplo anterior se tiene:

Paso 1. Calcule todas las parejas posibles de los promedios de tratamiento, así como sus valores.

$$_kC_2 = {}_3C_2 = \frac{3\,!}{2\,!\,1\,!} = 3$$

$$\bar{R}_B - \bar{R}_A,\ \bar{R}_M - \bar{R}_A \text{ y } \bar{R}_B - \bar{R}_M$$

$$\bar{R}_A = \frac{\Sigma R_A}{n} = \frac{14}{10} = 1.40 \qquad \bar{R}_B - \bar{R}_A = 2.80 - 1.40 = 1.40$$

$$\bar{R}_B = \frac{\Sigma R_B}{n} = \frac{28}{10} = 2.80 \qquad \bar{R}_M - \bar{R}_A = 1.80 - 1.40 = 0.40$$

$$\bar{R}_M = \frac{\Sigma R_M}{n} = \frac{18}{10} = 1.80 \qquad \bar{R}_B - \bar{R}_M = 2.80 - 1.80 = 1.0$$

Paso 2. Se obtiene la *DM*, para $\chi^2_{(2.0.05)} = 5.991$

$$DM = \sqrt{\frac{x^2\,(gl,\alpha)\,[k\,(k+1)]}{6\,n}}$$

$$DM = \sqrt{\frac{5.991\,[3 \times 4]}{6 \times 10}}$$

$$DM = \sqrt{\frac{5.991 \times 2}{10}} = \sqrt{1.1982}$$

$$\boxed{DM = 1.09462}$$

Paso 3. Como $1.4 > 1.09462$, que corresponde a la pareja $\bar{R}_B - \bar{R}_A$; por consiguiente el nivel $\bar{R}_B$ es el que hace la diferencia estadísticamente significativa. Para $\chi^2_{(2.0.01)} = 9.210$, la diferencia mínima (*DM*) = 1.3572 es menor que 1.4, por lo que R_B también lo cumple al 1%.

14.9 COEFICIENTE DE SPEARMAN (r_s)

Este coeficiente de correlación, también conocido como *de rangos ordenados*, es de los que más se aplican. Destaca su utilidad cuando el número de pares de puntaciones (n) que se desea asociar es pequeño (menor que 30). Si el número de dichos pares es muy grande, se emplea un modelo paramétrico ya que, por el teorema central del límite, la condición de normalidad se minimiza.

Por otra parte, cuando las puntuaciones se jerarquizan (o se ponen en correspondencia biunívoca con el conjunto de números ordinales), es muy probable que se presenten muchos "empates": la coincidencia de puntuaciones con el mismo número ordinal. Si estos dos hechos ocurrieran, lo más adecuado sería utilizar el coeficiente de correlación de Pearson que, como se recordará, es:

$$r_{xy} = \frac{n[\Sigma xy] - (\Sigma x)(\Sigma y)}{\sqrt{[n(\Sigma x^2) - (\Sigma x)^2][n(\Sigma y^2) - (\Sigma y)^2]}}$$

No obstante, si el número de puntuaciones que resulta se desea correlacionar es menor que 30, y hay pocos empates, el coeficiente de Spearman resulta el más apropiado. Esto se debe a que dichas variables representan ciertas observaciones y pueden ser ordenadas; al mismo tiempo, es deseable comprobar la hipótesis de independencia entre ambas variables aleatorias. Este coeficiente es una variedad particular del de Pearson, por lo que se define como sigue:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} [R(x_i) - R(y_i)]^2}{n(n^2 - 1)}$$

o bien,

$$r_s = 1 - \frac{6T}{n(n^2 - 1)}$$

donde:

$$T = \sum_{i=1}^{n} [R(x_i) - R(y_i)]^2$$

n = tamaño de la muestra.
$R(x_i)$ = variables x jerarquizadas (ordenadas).
$R(y_i)$ = variables y jerarquizadas (ordenadas).
1 y 6 = constantes numéricas.

Además de obtener el grado de asociación entre ambas variables con r_s, es posible determinar la dependencia o independencia de dos variables aleatorias.

Para la prueba de independencia en la que se utiliza r_s existen tres casos:

1. Prueba bilateral.
   H_0: Las x_i y las y_i son mutuamente independientes.
   H_1: *a*) Cuando los valores altos de x tienden a ser pareados con los valores altos de y.
   b) Cuando los valores bajos (o pequeños) de x tienden a ser pareados con los valores altos (o grandes) de y.
   Si H_i no se rechaza, tanto para *a*) como *b*), x y y son dependientes.
2. Prueba unilateral, correlación positiva.
   H_0: Las x_i y las y_i son mutuamente independientes.
   H_1: Cuando los valores altos (o grandes) de x y de y tienden a ser pareados al mismo tiempo. Entonces x y y son dependientes.
3. Prueba unilateral, correlación negativa.
   H_0: Las x_i y las y_i son mutuamente independientes.
   H_1: Cuando los valores pequeños o bajos de x tienden a ser pareados con los valores altos o grandes de y; también se prevé el caso inverso.
   x y y se consideran dependientes.

No obstante que la hipótesis nula de no correlación entre x y y es más precisa que el concepto de independencia entre x y y mencionado anteriormente, y que implica la existencia, como indican las hipótesis alternas de correlación entre x y y, en este texto se usará el concepto de independencia debido a que es más fácil de interpretar y tiene un uso más amplio.

Ejemplo A un grupo de 10 estudiantes de una escuela superior se les aplicó una prueba de conocimientos en matemáticas (x) y una de lógica (y). Se obtuvieron las siguientes puntuaciones:

Estudiante	x	*y*
A	84	52*
B	75	39
C	98*	48
D	70**	32**
E	75	40
F	80	36
G	83	38
H	75	37
I	84	50
J	90	46

* Calificación más alta.
** Calificación más baja.

a) Se desea determinar el grado de semejanza entre las calificaciones obtenidas por los estudiantes en las pruebas x y y.

b) H_0: Las calificaciones obtenidas en matemáticas son mutuamente independientes de las obtenidas en lógica por los 10 estudiantes, contra la alternativa bilateral (dos colas), al 0.05 de nivel de significancia.
H_1: Existe una correlación positiva o negativa entre las calificaciones obtenidas en ambas pruebas (dependencia).

Solución para a)

Paso 1. Se ubican las calificaciones más altas y las más bajas para cada variable. En el caso de la variable x, el estudiante C tiene la calificación más alta, por lo que le corresponde el lugar 1. En cambio, el estudiante D obtuvo la calificación más baja, por lo que le corresponde el lugar 10.

Para la variable y, el estudiante A tiene la calificación más alta, por lo que le corresponde el lugar 1; de manera consecuente, el estudiante D tiene la calificación más baja, por lo que su lugar es el 10.

Paso 2. Se obtienen y se ubican las puntuaciones jerarquizadas (ordenadas):

Estudiante	*R(x)*	*Estudiante*	*R(y)*
A	3*	A	1
B	7**	B	6
C	1	C	3
D	10	D	10
E	8**	E	5
F	6	F	9
G	5	G	7
H	9**	H	8
I	4*	I	2
J	2	J	4

* Promedio = 3.5.
** Promedio = 8.

Paso 3. Se consideran las puntuaciones que obtuvieron el mismo orden al ser jerarquizadas (o sea, las puntuaciones empatadas) y se establecen sus verdaderos lugares al promediar las puntuaciones con empate. Por ejemplo, las puntuaciones de los estudiantes A e I en la variable x son ambas 84, y los estudiantes B, E y H tienen la misma puntuación de 75. Por lo anterior, al promediar las puntuaciones iguales, pero ya jerarquizadas, se determina su lugar exacto. A los estudiantes A e I les corresponden los lugares 3 y 4 (en forma indistinta de A o I).

$$\frac{3+4}{2} = \frac{7}{2} = 3.5 \text{ (orden verdadero)}$$

Lo mismo ocurre para los estudiantes B, E y H.

$$\frac{7+8+9}{3} = \frac{24}{3} = 8 \text{ (Puntuación ordenada verdadera).}$$

Paso 4. Se calcula T de la siguiente manera: se restan las puntuaciones jerarquizadas para cada estudiante; posteriormente, se elevan al cuadrado cada una de las restas y se obtiene la suma total:

$$T = \sum_{i=1}^{10} [R(x_i) - R(y_i)]^2$$

Estudiante	$R(x)$	$R(y)$	$R(x) - R(y)$	$[R(x) - R(y)]^2$
A	3.5	1.0	2.5	6.25
B	8.0	6.0	2.0	4.0
C	1.0	3.0	–2.0	4.0
D	10.0	10.0	0.0	0.0
E	8.0	5.0	3.0	9.0
F	6.0	9.0	–3.0	9.0
G	5.0	7.0	–2.0	4.0
H	8.0	8.0	0.0	0.0
I	3.5	2.0	1.5	2.25
J	2.0	4.0	–2.0	4.0

$$T = \sum_{i=1}^{10} [R(x_i) - R(y_i)]^2 \therefore T = 42.50$$

Paso 5. Los resultados anteriores se sustituyen en la ecuación.

$$r_s = 1 - \frac{6T}{n(n^2 - 1)}$$

y así

$$r_s = 1 - \frac{6\,(42.50)}{10(100 - 1)}$$

$$r_s = 1 - \frac{255}{990} = 1 - 0.26$$

$$\therefore r_s = 0.74$$

Solución para b)

La hipótesis propuesta anteriormente acerca de la independencia o dependencia entre las calificaciones obtenidas por los estudiantes de este ejemplo se comprobará a 0.05 de nivel de significancia y dos colas. Al respecto, existen dos formas de proceder.

Paso 1. Con el valor calculado de $T = 42.50$ y la tabla correspondiente del Apéndice se aplicará la prueba de Hotelling y Pabst, la cual destaca la naturaleza no paramétrica de r_s.

Como el nivel de significancia es de 0.05 y la prueba es de dos colas, entonces $\alpha = 0.025$ (la mitad de 0.05), al buscar el cuantil de $T = 42.50$ para $n = 10$ y $\alpha = 0.025$, en la tabla del apéndice.

$$\omega_T = \omega_{0.025} = 60$$

Puesto que 42.50 es menor que 60, la hipótesis nula se rechaza.

Paso 2. Como r_s ha sido calculado ($r_s = 0.74$), será más fácil usarlo con el fin de comprobar la hipótesis propuesta; para esto $1 - 0.025 = 0.975$, que en la tabla respectiva del Apéndice corresponde a 0.6364. Debido a que la regla de decisión indica que si r_s es mayor que el cuantil expuesto en la tabla, la hipótesis nula se rechaza.

Como se aprecia, ambos resultados son congruentes, lo que incrementa ligeramente la confiabilidad en las conclusiones.

Interpretación de los resultados

El coeficiente de correlación de Spearman es un caso particular del coeficiente de correlación lineal de Pearson; por eso, para fines prácticos, r_s puede interpretarse como r_{xy}, aunque si existieran muchos empates en los datos ordenados r_{xy} y r_s discreparán. En el ejemplo desarrollado, 0.74 como resultado de r_s, indica una correlación significativa.

Prueba de significancia de r_s

Cuando el tamaño de la muestra es menor que 10, se consultará la tabla especial del Apéndice, ya que cuando n es pequeña, el valor de r_s debe ser muy grande para que sea significativo. Si $n > 10$, se podrá utilizar la siguiente fórmula (igual que en el caso del coeficiente de correlación de Pearson):

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$$

Al sustituir los valores anteriores se obtiene:

$$t = \frac{0.74\sqrt{10-2}}{\sqrt{1-(0.74)^2}}$$

$$= \frac{0.74\sqrt{8}}{\sqrt{1-0.5476}}$$

$$= \frac{2.093}{0.6726} = 3.11$$

$$\boxed{\therefore\ t = 3.11}$$

Este resultado se contrasta con la tabla del Apéndice de la razón t de Student, y se concluye lo siguiente:

$$gl = n - 2 = 10 - 2 = 8; \quad \alpha = 0.05$$

Para una prueba bilateral, la t crítica será:

$$t_8\ (0.05) = 2.36$$

Por consiguiente, r_s es significativo, ya que la regla de decisión es: si $t \geq t_{crít}$, H_0 se rechaza, y como $3.11 > 2.36$, la hipótesis nula (H_0) también se descarta.

Se concluye que

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

donde ρ significa la correlación no paramétrica en la población.

14.10 COEFICIENTE TAU (τ) DE KENDALL

Esta medida de correlación se basa en intervalos jerarquizados de las observaciones, más que en los números mismos; lo anterior tiene la ventaja de que la distribución de dicho coeficiente no depende de la que presentan de x y y, siempre que las observaciones representadas por x y y sean independientes y continuas. Algunos investigadores prefieren este coeficiente, desarrollado por Kendall (1938), al de Spearman, aunque τ es ligeramente más difícil de calcular que r_s en especial cuando existen empates. No obstante esto, el coeficiente de Kendall cuenta con una gran ventaja: su distribución tiende a lo normal más rápidamente que el de Spearman. Este "ajuste" a lo normal es mejor para el τ de Kendall, siempre y cuando la hipótesis nula (H_0) de independencia entre x y y sea cierta.

Ejemplo Se cataloga a un grupo de nueve niños como los más agresivos en el rango de cinco años de edad de la guardería donde acuden con regularidad. Para confirmar el grado de agresividad, se aplicó a los infantes una prueba: durante una semana se efectuaron registros observacionales, diarios y de acuerdo con ciertas condiciones. Los resultados promedio fueron los siguientes: registros observacionales en la guardería (RO_y), registros observacionales en los hogares (RO_x); (R_x), rangos o intervalos en el hogar (R_x), así como rangos o intervalos en la guardería (RO_y). Dichos registros se presentan en la tabla siguiente.

Niños	RO_x	RO_y	R_x	R_y
A	84	60	1	4
B	80	64	2	2
C	78	71	3	1
D	76	61	4	3
E	70	58	5	5
F	64	57	6	6
G	62	54	7	8
H	50	55	8	7
I	47	52	9	9

A continuación se calcula el coeficiente de correlación de Kendall (τ) entre lo detectado por los padres de los niños y lo que se advierte en la guardería.

Paso 1. Cada distribución de puntaciones, que representa a la variable x o a la y, se jerarquiza casi de la misma manera que cuando se calcula el coeficiente de Spearman para obtener R_x y R_y; la modificación consiste en que un conjunto de rangos (x o y) debe estar ordenado en una secuencia natural y creciente. El objetivo de este paso es tener una referencia que se utilizará posteriormente.

Paso 2. Se obtiene la columna de rangos más altos (P) y la de rangos menores (Q) que tengan como referencia la columna R_y. Con ese fin se procede así: se considera el valor numérico correspondiente al primer niño (4, en la columna R_y del ejemplo) y se cuenta hacia abajo cuántos valores numéricos son menores que él (en este caso son 2, 1 y 3); a continuación, se cuentan cuántos valores son mayores que él (5, 6, 8, 7 y 9).

R_x	R_y		*Cantidad de rangos más altos*	*Cantidad de rangos más bajos*
1	4	Primer sujeto	5	3
2	2			
3	1	Tres rangos		
4	3	más bajos que		
5	5	el primer sujeto		
6	6			
7	8	Cinco rangos		
8	7	más altos que		
9	9	el primer sujeto		

El procedimiento se aplica de nuevo para obtener el segundo sujeto (2 en el ejemplo); así, hay un sujeto más bajo que el 2 (1) y seis más altos (3, 5, 6, 8, 7 y 9).

R_x	R_y
1	4
2	2
3	1
4	3
5	5
6	6
7	8
8	7
9	9

Segundo sujeto
Un rango
más bajo que el
segundo sujeto

Seis rangos más
altos que el
segundo sujeto

Cantidad de rangos más altos	*Cantidad de rangos más bajos*
5	3
6	1

A fin de determinar el tercer sujeto (1), se excluyen los sujetos anteriores a él y se cuenta hacia abajo cuántos son menores y cuántos mayores. En este caso, no existe valor numérico menor; los más altos son seis (3, 5, 6, 8, 7 y 9).

R_x	R_y
1	4
2	2
3	1
4	3
5	5
6	6
7	8
8	7
9	9

Tercer sujeto
(cero rangos más
bajos que él)

Seis rangos
más altos que
el tercer sujeto

Rangos más altos	*Rangos más bajos*
5	3
6	1
6	0

El método se aplicará sucesivamente hasta llegar al último sujeto, que siempre tendrá cero rangos más altos y cero más bajos.

Paso 3. Una vez determinadas las anteriores columnas, se obtiene la sumatoria correspondiente a la columna de rangos más altos (P), así como la relativa a la columna de rangos más bajos (Q).

Sujetos	RO_x	RO_y	R_x	R_y	Rangos más altos P	*Rangos más bajos* Q
A	84	60	1	4	5	3
B	80	64	2	2	6	1
C	78	71	3	1	6	0
D	76	61	4	3	5	0
E	70	58	5	5	4	0
F	64	57	6	6	3	0
G	62	54	7	8	1	1
H	50	55	8	7	1	0
I	47	52	9	9	0	0
					31	5

Paso 4. El resultado se sustituye en la fórmula *tau* de Kendall.

$$\tau = \frac{P - Q}{\frac{n(n-1)}{2}}$$

donde:

n = número de casos o sujetos; en el ejemplo $n = 9$.
P = suma de rangos más altos ($P = 31$).
Q = suma de rangos más bajos ($Q = 5$).

Así:

$$\tau = \frac{P - Q}{\frac{n(n-1)}{2}}$$

$$= \frac{31 - 5}{\frac{72}{2}}$$

$$= \frac{26}{36} = 0.72$$

$$\boxed{\therefore \tau = 0.72}$$

14.11 COEFICIENTE DE CONCORDANCIA, ω, DE KENDALL

Es posible considerar este coeficiente como el promedio de un grupo de coeficientes de Spearman; o sea, ω es una medida del grado de acuerdo o concordancia entre m conjuntos de n rangos. En otros términos, para un grupo de n objetos que m jueces jerarquizan o evalúan, ω provee información sobre el grado de acuerdo existente entre los m conjuntos de rangos o jerarquizaciones que otorgan los jueces.

Una diferencia entre r_s (Spearman) y ω (Kendall) es que éste siempre será positivo, por lo que su intervalo de valores fluctúa de 0 a 1. De esa manera, si la evaluación que otorga cada juez a los n objetos es la misma, entonces $\omega = 1.0$; en cambio, si existe un total desacuerdo entre ellos, $\omega = 0$.

Sin demérito de lo anterior, hay que destacar que lo medido por ω es el acuerdo (concordancia) entre los jueces con respecto a la materia evaluada; por tanto, aunque ω sea grande (muy cercana a 1.0), no siempre significa que los jueces valoren correctamente, ya que pueden coincidir por completo en una evaluación incorrecta en términos de un criterio externo. Tampoco implica que los jueces

utilizaron los mismos criterios de evaluación o los mismos estándares, independientemente de si son los más adecuados.

Por otra parte, si los jueces no se ponen de acuerdo ($\omega = 0$), quizá se deba a que los atributos por evaluar son ambiguos o están pobremente definidos, o también a que dichos objetos no difieren significativamente en la característica o atributo medido. En consecuencia, la discriminación no es posible, ni puede ser concordante la opinión de los jueces.

Ejemplo A cuatro catadores especializados ($m = 4$) se les presenta un grupo de cinco vinos diferentes ($n = 5$). La información obtenida se expresa de la siguiente manera:

Vino	*Opinión de los catadores*	*Sumas de los rangos (R)*
	1 2 3 4	
1	5 4 5 5	19
2	3 3 2 3	11
3	1 2 1 2	6
4	2 1 3 1	7
5	4 5 4 4	17
		$\Sigma R = 60$

Calcule la concordancia entre las opiniones otorgadas por los cuatro catadores, que son los jueces. La mayor puntuación (5) significa que el vino reúne varias características de calidad, y éstas descienden sucesivamente hasta llegar a 1, donde el vino es de baja calidad.

Paso 1. Las jerarquizaciones otorgadas por los cuatro jueces a cada vino se suman y se colocan en una columna, como se ve en la tabla anterior.

Paso 2. En caso de no existir relación entre la jerarquización otorgada por los jueces ($\omega = 0$), la suma de dichas jerarquizaciones para cada uno de los n objetos (vino) será la misma. Si esto hubiese ocurrido en el ejemplo, la suma sería 12; o sea, la suma total entre el número de casos ($60/5 = 12$).

Paso 3. Se obtiene la diferencia (D) entre la suma de cada grupo de jerarquizaciones y el promedio de 12, y se colocan los resultados (D) en la columna 2.

n *Vino*	*m* *Rango de los catadores*	(1) *Suma de los rangos*	(2) *Diferencia de rangos*
	1 2 3 4	R	$D = \Sigma R - 12$
1	5 4 5 5	19	7
2	3 3 2 3	11	1
3	1 2 1 2	6	6
4	2 1 3 1	7	5
5	4 5 4 4	17	5
		$\Sigma R = 60$	

Paso 4. Se elevan al cuadrado estas diferencias (D), se colocan los resultados en la columna 3 y se obtiene su suma.

n Vino	m Rango de los catadores	(1) Suma de los rangos	(2) Diferencia de rangos	(3)
	1 2 3 4	R	D	D^2
1	5 4 5 5	19	7	49
2	3 3 2 3	11	1	1
3	1 2 1 2	6	6	36
4	2 1 3 1	7	5	25
5	4 5 4 4	17	5	25
		$\Sigma R = 60$		$\Sigma D^2 = 136$

Paso 5. Se calcula ω, mediante la siguiente fórmula:

$$\omega = \frac{12\ \Sigma D^2}{m^2\ n(n^2 - 1)}$$

$$\omega = \frac{12\ (136)}{(4^2)(5)(5^2 - 1)}$$

$$\omega = \frac{1\ 632}{(16)(5)(24)} = \frac{1\ 632}{1\ 920} \therefore \omega = 0.85$$

Con los resultados obtenidos anteriormente, se comprueba que la suma de las jerarquizaciones (ΣR) es igual al producto del número de jueces ($m = 4$) por el de objetos evaluados ($n = 5$) y por ($n + 1$). Todo lo anterior se divide entre 2; o sea, que debe obtenerse 60 = 60.

Esto es:

$$\Sigma R = \frac{mn(n + 1)}{2}$$

Como $\Sigma R = 60$;

$$60 = \frac{4 \cdot 5 \cdot 6}{2}$$

$$60 = \frac{120}{2}$$

$$60 = 60$$

Prueba de significancia de ω

La significancia estadística de ω se evalúa por medio de la tabla del Apéndice. Los datos para consultar esta tabla son $m = 4$ (jueces) y $n = 5$ (vinos), que

constituyen los objetos jerarquizados o calificados. Se observa que con un nivel de significancia de 1%, el valor de ω_0 es de 0.67 y, dado que la ω calculada asciende a 0.85 y supera el valor de 0.67 (el de la tabla), puede concluirse que este coeficiente de concordancia ω es significativo a partir de 1% y, por supuesto, de 5% ($\omega_0 = 0.54$). Por consiguiente, los catadores realizaron un trabajo correcto al evaluar estos cinco vinos.

En general, la regla de decisión para el coeficiente de Kendall es la siguiente:

Si $\omega \geq \omega_0$, entonces ω es significativo.

Conclusión $\omega = 0.85$ indica que existe un alto grado de acuerdo (concordancia) entre los cuatro jueces (los catadores) en la evaluación de las bondades de los cinco vinos.

14.12 COEFICIENTE DE CORRELACIÓN (r_{bp}) BISERIAL DE PUNTO

La correlación biserial, en general, es una medida de asociación entre dos variables continuas que se distribuyen de esa misma forma, en donde, para facilitar el proceso, una de ellas se ha dicotomizado (sí-no; falso-verdadero; masculino-femenino; casado-soltero; bueno-malo, entre otros). La variable dicotomizada se supone discreta o discontinua cuando se trata de relacionar con la que permanece continua; el coeficiente más apropiado para este propósito es el de correlación biserial puntual (r_{bp}), que se define en la siguiente forma:

$$r_{bp} = \frac{n(\Sigma fcx) - nc(\Sigma fx)}{\sqrt{(nc\ n_i)\ [n\Sigma fx^2 - (\Sigma fx)^2]}}$$

donde:

n = número de sujetos.
f = frecuencia de ocurrencia de las puntuaciones obtenidas por los sujetos.
x = puntuaciones obtenidas por los sujetos.
fx = producto de la frecuencia por las puntuaciones.
fc = número de sujetos que obtuvieron exactamente las puntuaciones x.
nc = Σfc = total de sujetos que obtuvieron exactamente las puntuaciones x.
f_i = número de sujetos que no obtuvieron exactamente las puntuaciones x.
n_i = Σf_i = número total de sujetos que no obtuvieron las puntuaciones x.
fcx = producto del número de sujetos que obtuvieron exactamente las puntuaciones x por la puntuación x de cada sujeto representativo.
Σfx^2 = suma de los productos de la frecuencia absoluta por las puntuaciones x al cuadrado.
$(\Sigma fx)^2$ = cuadrado de la suma del producto de la frecuencia por las puntuaciones x.

Ejemplo A 100 estudiantes de una escuela preparatoria se les aplica una prueba de conocimiento x, que consta de 40 reactivos, así como la frecuencia con que se distribuyen los 100 estudiantes. Se obtienen los siguientes resultados:

x	f
40	2
38	4
37	6
36	12
32	12
31	10
30	12
28	10
27	10
25	4
24	4
22	3
20	3
18	3
16	2
12	2
10	1
	$n = 100$

Cuando se elaboró la prueba se hizo especial hincapié en el reactivo número 33; el criterio para evaluarla será incorrecto o correcto. Por consiguiente, se calculará el coeficiente r_{bp} mediante las puntuaciones obtenidas por los 100 estudiantes en la prueba x y la respuesta al reactivo 33, que se dicotomizó artificialmente como "incorrecto-correcto".

Paso 1. Se concentra en un cuadro la información sobre los resultados de la prueba y sobre la respuesta de los 100 estudiantes al reactivo 33. En la columna 1 se presentan las puntuaciones; en la columna 2, la frecuencia con que ocurren éstos; en la 3, el número de estudiantes que respondieron de manera correcta el reactivo. Por último, en la columna 4, al restar las puntuaciones de la columna 3 a la 2, se obtiene la frecuencia de los estudiantes que respondieron incorrectamente al reactivo 33.

(1)	(2)	(3)	(4)
x	f	fc	f_i
40	2	2	0
38	4	4	0
37	6	5	1
36	12	10	2
32	12	9	3
31	10	8	2
30	12	7	5
28	10	6	4
27	10	7	3
25	4	1	3
24	4	1	3
22	3	1	2
20	3	1	2
18	3	0	3
16	2	1	1
12	2	0	2
10	1	0	1

Paso 2. Se obtienen los productos de los valores de la columna (1) multiplicada por la columna (2). O sea fx, lo cual forma la columna (5). Por último, se suman estos resultados para obtener Σfx.

(5)
(fx)
80
152
222
432
384
310
360
280
270
100
96
66
60
54
32
24
10
$\Sigma fx = 2\ 932$

Paso 3. Se elevan al cuadrado las puntuaciones x de la columna 1 y este resultado se multiplica por cada valor f de la columna 2; esta operación origina la columna (6). Por último, se suman todos los valores y se obtiene Σfx^2.

Paso 4. Se obtiene la columna (7) al multiplicar las puntuaciones x de la columna 1 con los valores fc de la columna (3); después se suman todos estos resultados y se obtiene Σfcx.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
x	f	fc	f_i	fx	fx^2	fcx
40	2	2	0	80	3 200	80
38	4	4	0	152	5 776	152
37	6	5	1	222	8 214	185
36	12	10	2	432	15 552	360
32	12	9	3	384	12 288	288
31	10	8	2	310	9 610	248
30	12	7	5	360	10 800	210
28	10	6	4	280	7 840	168
27	10	7	3	270	7 290	189
25	4	1	3	100	2 500	25
24	4	1	3	96	2 304	24
22	3	1	2	66	1 452	22
20	3	1	2	60	1 200	20
18	3	0	3	54	972	0
16	2	1	1	32	512	16
12	2	0	2	24	288	0
10	1	0	1	10	100	0
	100	63	37	2 932	89 898	1 987
	Σf	nc	n_i	Σfx	Σfx^2	Σfcx

Paso 5. Los resultados obtenidos en la tabla anterior se sustituyen en la definición del coeficiente de corrección biserial-puntual.

$$r_{bp} = \frac{n(\Sigma fcx) - nc(\Sigma fx)}{\sqrt{(nc\ n_i)[n\ \Sigma fx^2 - (\Sigma fx)^2]}}$$

$$r_{bp} = \frac{(100)(1\ 987) - 63(2\ 932)}{\sqrt{(63)(37)[100(89\ 898) - (2\ 932)^2]}}$$

$$r_{bp} = \frac{13\ 984}{\sqrt{(2\ 331)(393\ 176)}}$$

$$r_{bp} = \frac{13\ 984}{30\ 274} = 0.46$$

$$\boxed{r_{bp} = 0.46}$$

Prueba de significancia de r_{bp}

Debido a que el coeficiente de correlación biserial de punto, o puntual, es un caso particular del coeficiente de correlación lineal de Pearson, su significancia y su relación en la población se llevan a cabo de la siguiente manera: en primer término, se establecen las hipótesis acerca de la existencia y no existencia de correlación en la población. La notación estadística es:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

También se establece el nivel de significancia (α) con una probabilidad de error de 0.05,0.01, 0.001 y demás. Los grados de libertad se determinan de la siguiente forma: $gl = n - 2$; con éstos se compara el coeficiente r_{bp}, calculado con el coeficiente crítico que se obtiene por medio de la tabla del Apéndice con la siguiente regla de decisión:

Si $r_{bp} \geq r_{bpcrit}$, entonces H_0 se rechaza.

Para el ejemplo desarrollado, resulta lo siguiente:

$$gl = 98, \text{ debido a que } n = 100$$

$$\alpha = 0.001$$

$r_{bp(crit)} = r_{bp(gl,\alpha)} = r_{bp(98,\ 0.001)} = 0.3211$. Este valor es para una prueba bilateral (de dos colas).

Como 0.46 > 0.3211, entonces H_0 se rechaza, lo que significa que sí existe correlación biserial puntual en la población y que el valor de $r_{bp} = 0.46$ es significativo. Debido a que las correlaciones biseriales-puntuales se usan frecuentemente como índices de asociación entre una pregunta en particular y la puntuación total de la prueba, son muy útiles para seleccionar preguntas (rubros) durante el proceso de diseño y construcción de pruebas, a fin de incluir en las mismas ciertas características específicas que se necesita medir.

Por último, se toma en consideración que r_{bp} presenta limitaciones cuando la variable continua se distribuye normalmente y la variable dicotómica tiene una probabilidad de ocurrencia de 0.5. La limitación proviene de que cada vector de la variable dicotómica (falso-verdadero, correcto-incorrecto) tiene la misma posibilidad de ser escogido, por lo que el máximo valor que r_{bp} adquiriría es 0.80; pero si la variable no dicotomizada es bimodal, el máximo valor de r_{bp} asciende a 0.90.

Finalmente, al interpretar el resultado de r_{bp} en el ejemplo de los 100 estudiantes a los que se aplica la prueba con un reactivo especial, se concluye que el coeficiente no es alto (0.46), pero existe correlación significativa (dadas las limitaciones del biserial-puntual de 0.80). Adicionalmente, destaca que la correlación es positiva; esto significa que los estudiantes con las mejores puntuaciones en la prueba tienden a responder correctamente al reactivo 33 y que, por el contrario, quienes obtuvieron puntuaciones bajas tienden a contestar incorrectamente el reactivo. Entonces, la pregunta 33 realiza lo que se supone constituye el propósito de la prueba: separar a los estudiantes buenos de los mediocres.

RESUMEN

Los datos categóricos obtenidos a partir de una sola variable de interés ocurren en un sinnúmero de situaciones reales. Se examinó la prueba de bondad de ajuste que se utiliza para comprobar si las frecuencias muestrales asociadas con las categorías de una variable coinciden con lo esperado en una situación hipotética.

Los problemas que consideran dos variables categóricas se resolvieron mediante la prueba de *ji* cuadrada que contrasta independencia para una sola muestra y dos variables de criterio.

Existen varias alternativas no paramétricas para el coeficiente de correlación lineal de Pearson, a fin de determinar la relación entre las variables distintas de las correspondientes a los intervalos: ordinales, dicotómicas o nominales.

Cuando el tipo de variables que intervienen en el análisis estadístico son ordinales, y la muestra es pequeña, así como la distribución de éstas no sigue el modelo normal, el coeficiente de correlación por intervalos jerarquizados (r_s) de Spearman o el coeficiente τ de Kendall, que en ciertas situaciones se prefiere a r_s, son los adecuados para encontrar la relación y el grado de asociación de este tipo de variables; o bien, puede obtenerse un valor estadístico (el coeficiente de concordancia ω) que, además de emplear la función de coeficiente promedio de un grupo de coeficientes de correlación de Spearman, suele medir el grado de concordancia existente entre la opinión o calificación que según determinado criterio tienen un grupo de personas (jueces) ante una situación de tipo subjetiva. En el caso de enfrentarse a un evento en el que una variable es continua y la otra dicotómica, el coeficiente biserial puntual (r_{bp}) es el apropiado para determinar las posibles relaciones entre estas dos variables.

El análisis de frecuencias observadas, variables nominales, de características en las cuales no es posible obtener un valor numérico, y que al mismo tiempo se clasifican en situaciones excluyentes en las tablas de doble entrada (tablas de contingencia), representa uno de los procesos más antiguos y útiles empleados por los investigadores de las ciencias sociales y del comportamiento. Para este análisis es posible obtener el coeficiente C de contingencia o el V de Kramer, cuya interpretación es la misma, pero los requisitos de uso varían según el tamaño (número de renglones por número de columnas) que dicha tabla de contingencia tenga, lo cual repercute en el número de casillas o clasificaciones del problema por resolver.

PROBLEMAS

14.1 Se comparan dos tipos de adhesivos para materiales dentales midiendo la presión a la cual hay ruptura en megapascales. Aplique la prueba de:

a) U de Mann-Whitney

b) t de Student

Grupo 1	*Grupo 2*
3.2	4.1
2.5	3.9
4.0	3.6
3.1	4.7
3.5	4.5
2.9	3.4
3.7	4.8
3.9	4.4
3.3	3.3

14.2 Se mide a dos grupos de 24 ratas bajo dos diferentes tipos de droga experimental. El número de respuestas ante dos condiciones diferentes, grupo 1 con ruido y grupo 2 sin ruido, es el siguiente:

Grupo 1 (con ruido)	*Grupo 2 (sin ruido)*
18	23
32	23
36	25
39	35
25	28
33	25
18	17
39	42
38	43
25	15
25	15
40	37
26	29
35	18
32	27
34	25
15	13
29	33
17	11
22	38
20	17
25	16
33	21
19	16

Aplique la prueba de la U de Mann-Whitney para muestras grandes.

14.3 Se realiza un experimento con cinco personas en el que se mide la relación del tamaño de la pupila cuando se les presentan en forma verbal cuatro tipos de problemas de diferente grado de dificultad. Se obtuvieron los siguientes resultados.

Persona	*Problema 1*	*Problema 2*	*Problema 3*	*Problema 4*
A	15.2	15.8	20.2	22.9
B	9.8	14.1	24.9	21.2
C	10.0	8.9	13.5	23.1
D	4.0	8.8	7.8	11.6
E	16.2	9.1	25.1	29.6

Realice un análisis de varianza no paramétrico (Prueba de Friedman).

14.4 Se realiza una encuesta con un grupo de 100 pacientes con cáncer y a 100 personas normales, para conocer su conducta frente a la bebida.

Ingestión de alcohol	*Con cáncer*	*Sin cáncer*	
Abstemio	37	39	
Moderada	34	30	
Excesiva	15	21	
Adicción	14	10	
Total	100	100	200

Aplique la prueba de χ^2 y concluya.

14.5 En un estudio donde participaron 8 766 personas, entre las que había 6 087 sanas, 883 con un cáncer incipiente en el estómago y 796 con úlcera duodenal, se aplicó una clasificación a partir del tipo de sangre: O, A y B.

	GRUPOS			
Tipo de sangre	*Úlcera*	*Cáncer*	*Sanos*	*Total*
O	983	393	2842	4758
A	679	416	2625	3720
B	134	84	570	788
Total	793	883	6057	8766

Aplique la prueba de independencia.

14.6 A 10 cigarrillos de la misma marca, pero de 10 diferentes paquetes seleccionados al azar, se les mide la cantidad de alquitrán y de nicotina, de donde surgen los siguientes datos:

Cigarrillo	*Alquitrán*	*Nicotina*
A	14	0.9
B	17	1.1
C	28	1.6
D	17	1.3
E	16	1.0
F	13	0.3
G	24	1.5
H	25	1.4
I	18	1.2
J	31	2.0

Obtenga una correlación de Spearman.

14.7

	-1- *control*	-2- *grupo* 1	-3- *grupo* 2	-4- *grupo* 3	-5- *grupo* 4	-6- *grupo* 5
1	69.38	63.58	79.53	90.00	60.07	59.60
2	30.66	31.18	34.27	39.64	29.06	90.00
3	37.58	39.87	41.21	50.89	40.11	54.70
4	21.89	18.44	22.87	24.58	21.05	24.95
5	16.85	14.30	31.95	44.43	21.64	27.28
6	44.25	13.05	43.62	54.15	47.18	34.02
7	35.37	39.11	38.53	50.24	26.99	37.76
8	22.71	23.26	22.06	27.90	37.82	19.37
9	22.79	24.58	23.26	28.25	17.66	34.51
10	16.85	16.54	16.54	17.85	25.33	28.29
11	23.11	21.05	21.56	25.33	21.56	30.85
12	21.13	20.79	20.79	22.79	23.50	28.59
13	21.81	21.97	23.89	23.26	24.35	26.99
14	23.42	24.80	25.55	26.28	28.04	38.06
15	20.62	20.62	20.88	24.27	22.63	30.26
16	20.96	24.58	26.64	31.63	24.12	27.56

Los datos anteriores muestran tensiones que se soportan antes de la ruptura. Realice un análisis de varianza en rangos (Prueba de Kruskal-Wallis).

14.8

	-1- *control*	-2- *grupo* 1	-3- *grupo* 2	-4- *grupo* 3	-5- *grupo* 4	-6- *grupo* 5
1	0.88	0.80	0.97	1.23	0.75	0.74
2	0.26	0.27	0.32	0.41	0.24	1.12
3	0.37	0.41	0.43	0.60	0.41	0.67
4	0.14	0.10	0.15	0.17	0.13	0.18
5	0.08	0.06	0.28	0.49	0.14	0.21
6	0.49	0.05	0.48	0.66	0.54	0.31
7	0.34	0.40	0.39	0.59	0.21	0.37
8	0.15	0.16	0.14	0.22	0.38	0.11
9	0.15	0.17	0.16	0.22	0.09	0.32
10	0.08	0.08	0.08	0.09	0.18	0.23
11	0.15	0.13	0.13	0.18	0.14	0.26
12	0.13	0.13	0.13	0.15	0.16	0.23
13	0.14	0.14	0.16	0.16	0.17	0.21
14	0.16	0.18	0.19	0.20	0.22	0.38
15	0.12	0.12	0.13	0.17	0.15	0.25
16	0.13	0.17	0.20	0.28	0.17	0.21

Los datos anteriores muestran el porcentaje de la tensión antes de la ruptura. Aplique la prueba de Kruskal-Wallis.

14.9 A 10 pacientes diagnosticados con ansiedad severa se les aplica una terapia que consiste en tres sesiones. De ella se obtiene lo siguiente:

Paciente	*I*	*II*	*III*
A	21	23	15
B	29	30	21
C	16	19	18
D	20	19	18
E	13	10	14
F	5	12	6
G	18	18	12
H	26	32	21
I	17	20	9
J	4	10	2

Calcule la prueba de Friedman.

14.10 Se aplica a 12 personas tres tratamientos nuevos que se comparan con el tratamiento tradicional, con los siguientes datos:

	Tratamiento			
Sujetos	*(Control)* *A*	*B*	*C*	*D*
1	8	21	23	15
2	10	29	30	21
3	20	16	19	18
4	15	20	19	18
5	12	13	10	14
6	10	15	12	17
7	9	5	12	6
8	15	18	18	12
9	17	26	32	21
10	12	17	20	9
11	10	4	10	2
12	6	12	15	4

Calcule la prueba de Friedman.

Bibliografía

Aaker, D., *Investigación de mercados*, McGraw-Hill, México, 1994.

Benzecri, J. P. *et al.*, *L'analyse des données, vol. I, La Taxonomie, vol. II, L'analyse des correspondances*, Dunod, París, 1973.

Boom, A. R. y D. de L. T. Boom, *Análisis de decisiones y computación en medicina*, JGH Editores, México, 1997.

Box, G. E., W., Hunter y J. S. Hunter, *Statistics for experimenters*, Wiley, Nueva York, 1978.

Britton, J. e I. Bello, *Matemáticas contemporáneas*, Harla, México, 1982.

Cochran, W., *Técnicas de muestreo*, CECSA, México, 1980.

Conover, W., *Practical Nonparametric Statistics*, Wiley, 1971.

Chalmers, A. F., *¿Qué es esa cosa llamada ciencia?*, 14a. ed., Siglo XXI, México, 1996.

Damond, R. y W. Harvey, *Experimental design. Anova and Regression*, Harper and Row, Nueva York, 1987.

Draper, N. R. y H. Smith, *Applied Regression Analysis*, 2a. ed., Wiley, Nueva York, 1981.

Hamburger, J., *Los límites del conocimiento*, Fondo de Cultura Económica, México, 1986.

Harris, B., *Theory of Probability*, Addison-Wesley, 1966.

Hasser, N. y J. Lassalle, *Análisis matemático*, vol. 1, Trillas, México, 1970.

Hawkins, C. y C. Webern, *Statistical Analysis*, Harper and Row, Nueva York, 1980.

Herzberg, Paul, *Principles of Statistics*, Wiley, Nueva York, 1983.

Hines, W. y D. Montgomery, *Probabilidad y estadística para ingeniería y administración*, CECSA, México, 1993.

Infante Gil, Said y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpresión, Trillas, México, 1996.

Isaac, S. y W. Michael, *Handbook in Research and Evaluation*, 2a. ed., Edits Publishers, 1981.

Kennedy, J. y A. Neville, *Estadística para ciencias e ingeniería*, 2a. ed., Harla, México, 1982.

Kirk, R. E., *Experimental Design Procedures for the Behavioral Sciences*, Wadswarth, 1968.

Kleinbaum, D. y L. Kupper, *Applied Regression Analysis and other Multivariable Methods*, Duxbury Press, 1978.

Kuhn, T. S., *La estructura de las revoluciones científicas*, 6a. reimpresión, Fondo de Cultura Económica, México, 1985.

Lakatos, I., *La metodología de los programas de investigación científica*, Alianza Universidad, Madrid, 1983.

Leithold, L., *Matemáticas previas al cálculo*, Oxford University Press-Harla, México, 1996.

Lovaglia, F., *Álgebra*, Harla, México, 1972.

Marascuilo, L. y M. McSweeney, *Nonparametric and Distribution-Free Methods for the Social Science*, Wadsworth, 1997.

Méndez, I. *et al.*, *El protocolo de investigación*, 2a. ed., Trillas, México, 1990.

Montgomery, D. C., *Diseño y análisis de experimentos*, Iberoamericana, México, 1991.

Murphy, *Introducción al análisis estadístico*, Addison-Wesley, México, 1987.

Myers, J. L., *Fundamentals of Experimental Design*, 3a ed., Allyn and Bacun, 1979.

Neale, J. M. y R. M. Liebert, *Science and Behavior*, Prentice Hall, N. J., 1980.

Parsons, R., *Statistical Analysis: A Decision Making Approach*, Harper and Row, Nueva York, 1983.

Plutchik, R., *Foundations of Experimental Research*, 3a. ed., Harper and Row, Nueva York, 1983.

Rojas, S. R., *El proceso de la investigación científica*, 3a. ed., Trillas, México, 1983.

Ross, S., *Introduction to Probability Models*, Academic Press, 1972.

Scheaffer, R., W. Mendelhall y L. Ott, *Elementos de muestreo*, 3a ed., Iberoamericana, México, 1987.

Snedecor, G. y W. Cochran, *Statistical Methods*, 6a. ed., Iowa State University Press, Iowa, 1979.

Stevenson, W. J., *Estadística para administración y economía*, Harla, México, 1981.

Swokoswski, E., *Álgebra y trigonometría*, 2a. ed., Iberoamericana, México, 1988.

Winkler, R. y W. Hays, *Statistics*, 2a. ed., Holt, Reinhardt and Wiston, Nueva York, 1975.

Yamane, T., *Estadística*, 3a. ed., Harla, México, 1979.

Anexo 1

Tablas

Probabilidades acumuladas de la distribución binomial

		p									
	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 1	0	0.9500	0.9000	0.8500	0.8000	0.7500	0.7000	0.6500	0.6000	0.5500	0.5000
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9687
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0467	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9841	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6563
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

* Si $X \sim Bin\ (n, p)$, la tabla de valores de $P(X \leqq c)$, $c = 0, 1, \ldots, n$, $P(X \leqq c), = \sum_{x=0}^{c} \binom{n}{x} p^x (1-p)^{n-x}$

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a, reimpr., Trillas, México, 1996.

Probabilidades acumuladas de la distribución binomial (*Continuación*)

		p								
	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
$n = 1$	0	0.4500	0.4000	0.3500	0.3000	0.2500	0.2000	0.1500	0.1000	0.0500
	1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 2$	0	0.2025	0.1600	0.1225	0.0900	0.0625	0.0400	0.0225	0.0100	0.0025
	1	0.6975	0.6400	0.5775	0.5100	0.4375	0.3600	0.2775	0.1900	0.0975
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 3$	0	0.0911	0.0640	0.0429	0.0270	0.0156	0.0080	0.0034	0.0010	0.0001
	1	0.4253	0.3520	0.2818	0.2160	0.1563	0.1040	0.0608	0.0280	0.0073
	2	0.8336	0.7840	0.7254	0.6570	0.5781	0.4880	0.3859	0.2710	0.1426
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 4$	0	0.0410	0.0256	0.0150	0.0081	0.0039	0.0016	0.0005	0.0001	0.0000
	1	0.2415	0.1792	0.1265	0.0837	0.0508	0.0272	0.0120	0.0037	0.0005
	2	0.6090	0.5248	0.4370	0.3483	0.2617	0.1808	0.1095	0.0523	0.0140
	3	0.9085	0.8704	0.8215	0.7599	0.6836	0.5904	0.4780	0.3439	0.1855
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 5$	0	0.0185	0.0102	0.0053	0.0024	0.0010	0.0003	0.0001	0.0000	0.0000
	1	0.1312	0.0870	0.0540	0.0308	0.0156	0.0067	0.0022	0.0005	0.0000
	2	0.4069	0.3174	0.2352	0.1631	0.1035	0.0579	0.0266	0.0086	0.0012
	3	0.7438	0.6630	0.5716	0.4718	0.3672	0.2627	0.1648	0.0815	0.0226
	4	0.9497	0.9222	0.8840	0.8319	0.7627	0.6723	0.5563	0.4095	0.2262
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 6$	0	0.0083	0.0041	0.0018	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000
	1	0.0692	0.0410	0.0223	0.0109	0.0046	0.0016	0.0004	0.0001	0.0000
	2	0.2553	0.1792	0.1174	0.0705	0.0376	0.0170	0.0059	0.0013	0.0001
	3	0.5585	0.4557	0.3529	0.2557	0.1694	0.0989	0.0473	0.0159	0.0022
	4	0.8364	0.7667	0.6809	0.5798	0.4661	0.3446	0.2235	0.1143	0.0328
	5	0.9723	0.9533	0.9246	0.8824	0.8220	0.7379	0.6229	0.4686	0.2649
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 7$	0	0.0037	0.0016	0.0006	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.0357	0.0188	0.0090	0.0038	0.0013	0.0004	0.0001	0.0000	0.0000
	2	0.1529	0.0963	0.0556	0.0288	0.0129	0.0047	0.0012	0.0002	0.0000
	3	0.3917	0.2898	0.1998	0.1260	0.0706	0.0333	0.0121	0.0027	0.0002
	4	0.6836	0.5801	0.4677	0.3529	0.2436	0.1480	0.0738	0.0257	0.0038
	5	0.8976	0.8414	0.7662	0.6706	0.5551	0.4233	0.2834	0.1497	0.0444
	6	0.9848	0.9720	0.9510	0.9176	0.8665	0.7903	0.6794	0.5217	0.3017
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 8$	0	0.0017	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0181	0.0085	0.0036	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000
	2	0.0885	0.0498	0.0253	0.0113	0.0042	0.0012	0.0002	0.0000	0.0000
	3	0.2604	0.1737	0.1061	0.0580	0.0273	0.0104	0.0029	0.0004	0.0000
	4	0.5230	0.4059	0.2936	0.1941	0.1138	0.0563	0.0214	0.0050	0.0004
	5	0.7799	0.6846	0.5722	0.4482	0.3215	0.2031	0.1052	0.0381	0.0058
	6	0.9368	0.8936	0.8309	0.7447	0.6329	0.4967	0.3428	0.1869	0.0572
	7	0.9916	0.9832	0.9681	0.9424	0.8999	0.8322	0.7275	0.5695	0.3366
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Probabilidades acumuladas de la distribución binomial (*Continuación*)

		p									
	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0088	0.0036	0.0014	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0606	0.0302	0.0139	0.0059
	2	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.2001	0.1189	0.0652	0.0327
	3	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.4256	0.2963	0.1911	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.6683	0.5328	0.3971	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.8513	0.7535	0.6331	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9499	0.9006	0.8262	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9878	0.9707	0.9390	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9980	0.9941	0.9852	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9993	0.9978	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9995
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 13	0	0.5133	0.2542	0.1209	0.0550	0.0238	0.0097	0.0037	0.0013	0.0004	0.0001
	1	0.8646	0.6213	0.3983	0.2336	0.1267	0.0637	0.0296	0.0126	0.0049	0.0017
	2	0.9755	0.8661	0.6920	0.5017	0.3326	0.2025	0.1132	0.0579	0.0269	0.0112
	3	0.9969	0.9658	0.8820	0.7473	0.5843	0.4206	0.2783	0.1686	0.0929	0.0461
	4	0.9997	0.9935	0.9658	0.9009	0.7940	0.6543	0.5005	0.3530	0.2279	0.1334

Probabilidades acumuladas de la distribución binomial (*Continuación*)

		p								
	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
n = 9	0	0.0008	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0091	0.0038	0.0014	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
	2	0.0498	0.0250	0.0112	0.0043	0.0013	0.0003	0.0000	0.0000	0.0000
	3	0.1658	0.0994	0.0536	0.0253	0.0100	0.0031	0.0006	0.0001	0.0000
	4	0.3786	0.2666	0.1717	0.0988	0.0489	0.0196	0.0056	0.0009	0.0000
	5	0.6386	0.5174	0.3911	0.2703	0.1657	0.0856	0.0339	0.0083	0.0006
	6	0.8505	0.7682	0.6627	0.5372	0.3993	0.2618	0.1409	0.0530	0.0084
	7	0.9615	0.9295	0.8789	0.8040	0.6997	0.5638	0.4005	0.2252	0.0712
	8	0.9954	0.9899	0.9793	0.9596	0.9249	0.8658	0.7684	0.6126	0.3698
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 10	0	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0045	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0274	0.0123	0.0048	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
	3	0.1020	0.0548	0.0260	0.0106	0.0085	0.0009	0.0001	0.0000	0.0000
	4	0.2616	0.1662	0.0949	0.0473	0.0197	0.0064	0.0014	0.0001	0.0000
	5	0.4956	0.3669	0.2485	0.1503	0.0781	0.0328	0.0099	0.0016	0.0001
	6	0.7340	0.6177	0.4862	0.3504	0.2241	0.1209	0.0500	0.0128	0.0010
	7	0.9004	0.8327	0.7384	0.6172	0.4744	0.3222	0.1798	0.0702	0.0115
	8	0.9767	0.9536	0.9140	0.8507	0.7560	0.6242	0.4557	0.2639	0.0861
	9	0.9975	0.9940	0.9865	0.9718	0.9437	0.8926	0.8031	0.6513	0.4013
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 11	0	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0022	0.0007	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0148	0.0059	0.0020	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	3	0.0610	0.0293	0.0122	0.0043	0.0012	0.0002	0.0000	0.0000	0.0000
	4	0.1738	0.0994	0.0501	0.0216	0.0076	0.0020	0.0003	0.0000	0.0000
	5	0.3669	0.2465	0.1487	0.0782	0.0343	0.0117	0.0027	0.0003	0.0000
	6	0.6029	0.4672	0.3317	0.2103	0.1146	0.0504	0.0159	0.0028	0.0001
	7	0.8089	0.7037	0.5744	0.4304	0.2867	0.1611	0.0694	0.0185	0.0016
	8	0.9348	0.8811	0.7999	0.6873	0.5448	0.3826	0.2212	0.0896	0.0152
	9	0.9861	0.9698	0.9394	0.8870	0.8029	0.6779	0.5078	0.3026	0.1019
	10	0.9986	0.9964	0.9912	0.9802	0.9578	0.9141	0.8327	0.6862	0.4312
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 12	0	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0079	0.0028	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0356	0.0153	0.0056	0.0017	0.0004	0.0001	0.0000	0.0000	0.0000
	4	0.1117	0.0573	0.0255	0.0095	0.0028	0.0006	0.0001	0.0000	0.0000
	5	0.2607	0.1582	0.0846	0.0386	0.0143	0.0039	0.0007	0.0001	0.0000
	6	0.4731	0.3348	0.2127	0.1178	0.0544	0.0194	0.0046	0.0005	0.0000
	7	0.6956	0.5618	0.4167	0.2763	0.1576	0.0726	0.0239	0.0043	0.0002
	8	0.8655	0.7747	0.6533	0.5075	0.3512	0.2054	0.0922	0.0256	0.0022
	9	0.9579	0.9166	0.8487	0.7472	0.6093	0.4417	0.2642	0.1109	0.0196
	10	0.9917	0.9804	0.9576	0.9150	0.8416	0.7251	0.5565	0.3410	0.1184
	11	0.9992	0.9978	0.9943	0.9862	0.9683	0.9313	0.8578	0.7176	0.4596
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 13	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0041	0.0013	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0203	0.0078	0.0025	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	4	0.0698	0.0321	0.0126	0.0040	0.0010	0.0002	0.0000	0.0000	0.0000

Probabilidades acumuladas de la distribución binomial (*Continuación*)

		p									
	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 13	5	1.0000	0.9991	0.9925	0.9700	0.9198	0.8346	0.7159	0.5744	0.4268	0.2905
	6	1.0000	0.9999	0.9987	0.9930	0.9757	0.9376	0.8705	0.7712	0.6437	0.5000
	7	1.0000	1.0000	0.9998	0.9988	0.9944	0.9818	0.9538	0.9023	0.8212	0.7095
	8	1.0000	1.0000	1.0000	0.9998	0.9990	0.9960	0.9874	0.9679	0.9302	0.8666
	9	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9975	0.9922	0.9797	0.9539
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9987	0.9959	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	13	1.0000	1 0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 14	0	0.4877	0.2288	0.1028	0.0440	0.0178	0.0068	0.0024	0.0008	0.0002	0.0001
	1	0.8470	0.5846	0.3567	0.1979	0.1010	0.0475	0.0205	0.0081	0.0029	0.0009
	2	0.9699	0.8416	0.6479	0.4481	0.2811	0.1608	0.0839	0.0398	0.0170	0.0065
	3	0.9958	0.9559	0.8535	0.6982	0.5213	0.3552	0.2205	0.1243	0.0632	0.0287
	4	0.9996	0.9908	0.9533	0.8702	0.7415	0.5842	0.4227	0.2793	0.1672	0.0898
	5	1.0000	0.9985	0.9885	0.9561	0.8883	0.7805	0.6405	0.4859	0.3373	0.2120
	6	1.0000	0.9998	0.9978	0.9884	0.9617	0.9067	0.8164	0.6925	0.5461	0.3953
	7	1.0000	1.0000	0.9997	0.9976	0.9897	0.9685	0.9247	0.8499	0.7414	0.6047
	8	1.0000	1.0000	1.0000	0.9996	0.9978	0.9917	0.9757	0.9417	0.8811	0.7880
	9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9983	0.9940	0.9825	0.9574	0.9102
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9989	0.9961	0.9886	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9978	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	0.9994	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0010	0.0003	0.0001	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0098	0.0033	0.0010	0.0003
	2	0.9571	0.7892	0.5614	0.3518	0.1971	0.0994	0.0451	0.0183	0.0066	0.0021
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.1339	0.0651	0.0281	0.0106
	4	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.2892	0.1666	0.0853	0.0384
	5	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.4900	0.3288	0.1976	0.1051
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.6881	0.5272	0.3660	0.2272
	7	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.8406	0.7161	0.5629	0.4018
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.9329	0.8577	0.7441	0.5982
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9771	0.9417	0.8759	0.7728
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9938	0.9809	0.9514	0.8949

Probabilidades acumuladas de la distribución binomial *(Continuación)*

	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
						p				
n = 13	5	0.1788	0.0977	0.0462	0.0182	0.0056	0.0012	0.0002	0.0000	0.0000
	6	0.3563	0.2288	0.1295	0.0624	0.0243	0.0070	0.0013	0.0001	0.0000
	7	0.5732	0.4256	0.2841	0.1654	0.0802	0.0300	0.0075	0.0009	0.0000
	8	0.7721	0.6470	0.4995	0.3457	0.2060	0.0991	0.0342	0.0065	0.0003
	9	0.9071	0.8314	0.7217	0.5794	0.4157	0.2527	0.1180	0.0342	0.0031
	10	0.9731	0.9421	0.8868	0.7975	0.6674	0.4983	0.3080	0.1339	0.0245
	11	0.9951	0.9874	0.9704	0.9363	0.8733	0.7664	0.6017	0.3787	0.1354
	12	0.9996	0.9987	0.9963	0.9903	0.9762	0.9450	0.8791	0.7458	0.4867
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 14	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0022	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0114	0.0039	0.0011	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0426	0.0175	0.0060	0.0017	0.0003	0.0000	0.0000	0.0000	0.0000
	5	0.1189	0.0583	0.0243	0.0083	0.0022	0.0004	0.0000	0.0000	0.0000
	6	0.2586	0.1501	0.0753	0.0315	0.0103	0.0024	0.0003	0.0000	0.0000
	7	0.4539	0.3075	0.1836	0.0933	0.0383	0.0116	0.0022	0.0002	0.0000
	8	0.6627	0.5141	0.3595	0.2195	0.1117	0.0439	0.0115	0.0015	0.0000
	9	0.8328	0.7207	0.5773	0.4158	0.2585	0.1298	0.0467	0.0092	0.0004
	10	0.9368	0.8757	0.7795	0.6448	0.4787	0.3018	0.1465	0.0441	0.0042
	11	0.9830	0.9602	0.9161	0.8392	0.7189	0.5519	0.3521	0.1584	0.0301
	12	0.9971	0.9919	0.9795	0.9525	0.8990	0.8021	0.6433	0.4154	0.1530
	13	0.9998	0.9992	0.9976	0.9932	0.9822	0.9560	0.8972	0.7712	0.5123
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 15	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0063	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0255	0.0093	0.0028	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	5	0.0769	0.0338	0.0124	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000
	6	0.1818	0.0950	0.0422	0.0152	0.0042	0.0008	0.0001	0.0000	0.0000
	7	0.3465	0.2131	0.1132	0.0500	0.0173	0.0042	0.0006	0.0000	0.0000
	8	0.5478	0.3902	0.2452	0.1311	0.0566	0.0181	0.0036	0.0003	0.0000
	9	0.7392	0.5968	0.4357	0.2784	0.1484	0.0611	0.0168	0.0022	0.0001
	10	0.8796	0.7827	0.6481	0.4845	0.3135	0.1642	0.0617	0.0127	0.0006
	11	0.9576	0.9095	0.8273	0.7031	0.5387	0.3518	0.1773	0.0556	0.0055
	12	0.9893	0.9729	0.9383	0.8732	0.7639	0.6020	0.3958	0.1841	0.0362
	13	0.9983	0.9948	0.9858	0.9647	0.9198	0.8329	0.6814	0.4510	0.1710
	14	0.9999	0.9995	0.9984	0.9953	0.9866	0.9648	0.9126	0.7941	0.5367
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 16	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0035	0.0009	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0149	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0486	0.0191	0.0062	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000
	6	0.1241	0.0583	0.0229	0.0071	0.0016	0.0002	0.0000	0.0000	0.0000
	7	0.2559	0.1423	0.0671	0.0257	0.0075	0.0015	0.0002	0.0000	0.0000
	8	0.4371	0.2839	0.1594	0.0744	0.0271	0.0070	0.0011	0.0001	0.0000
	9	0.6340	0.4728	0.3119	0.1753	0.0796	0.0267	0.0056	0.0005	0.0000
	10	0.8024	0.6712	0.5100	0.3402	0.1897	0.0817	0.0235	0.0033	0.0001

Probabilidades acumuladas de la distribución binomial *(Continuación)*

		p									
	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 16	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9851	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9991	0.9965	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 17	0	0.4181	0.1668	0.0631	0.0225	0.0075	0.0023	0.0007	0.0002	0.0000	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0501	0.0193	0.0067	0.0021	0.0006	0.0001
	2	0.9497	0.7618	0.5198	0.3096	0.1637	0.0774	0.0327	0.0123	0.0041	0.0012
	3	0.9912	0.9174	0.7556	0.5489	0.3530	0.2019	0.1028	0.0464	0.0184	0.0064
	4	0.9988	0.9779	0.9013	0.7582	0.5739	0.3887	0.2348	0.1260	0.0596	0.0245
	5	0.9999	0.9953	0.9681	0.8943	0.7653	0.5968	0.4197	0.2639	0.1471	0.0717
	6	1.0000	0.9992	0.9917	0.9623	0.8929	0.7752	0.6188	0.4478	0.2902	0.1662
	7	1.0000	0.9999	0.9983	0.9891	0.9598	0.8954	0.7872	0.6405	0.4743	0.3145
	8	1.0000	1.0000	0.9997	0.9974	0.9876	0.9597	0.9006	0.8011	0.6626	0.5000
	9	1.0000	1.0000	1.0000	0.9995	0.9969	0.9873	0.9617	0.9081	0.8166	0.6855
	10	1.0000	1.0000	1.0000	0.9999	0.9994	0.9968	0.9880	0.9652	0.9174	0.8338
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9970	0.9894	0.9699	0.9283
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9975	0.9914	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9931	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 18	0	0.3972	0.1501	0.0536	0.0180	0.0056	0.0016	0.0004	0.0001	0.0000	0.0000
	1	0.7735	0.4503	0.2241	0.0991	0.0395	0.0142	0.0046	0.0013	0.0003	0.0001
	2	0.9419	0.7338	0.4797	0.2713	0.1353	0.0600	0.0236	0.0082	0.0025	0.0007
	3	0.9891	0.9018	0.7202	0.5010	0.3057	0.1646	0.0783	0.0328	0.0120	0.0038
	4	0.9985	0.9718	0.8794	0.7164	0.5187	0.3327	0.1886	0.0942	0.0411	0.0154
	5	0.9998	0.9936	0.9581	0.8671	0.7175	0.5344	0.3550	0.2088	0.1077	0.0481
	6	1.0000	0.9988	0.9882	0.9487	0.8610	0.7217	0.5491	0.3743	0.2258	0.1189
	7	1.0000	0.9998	0.9973	0.9837	0.9431	0.8593	0.7283	0.5634	0.3915	0.2403
	8	1.0000	1.0000	0.9995	0.9957	0.9807	0.9404	0.8609	0.7368	0.5778	0.4073
	9	1.0000	1.0000	0.9999	0.9991	0.9946	0.9790	0.9403	0.8653	0.7473	0.5927
	10	1.0000	1.0000	1.0000	0.9998	0.9988	0.9939	0.9788	0.9424	0.8720	0.7597
	11	1.0000	1.0000	1.0000	1.0000	0.9998	0.9986	0.9938	0.9797	0.9463	0.8811
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9986	0.9942	0.9817	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	0.9951	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 19	0	0.3774	0.1351	0.0456	0.0144	0.0042	0.0011	0.0003	0.0001	0.0000	0.0000
	1	0.7547	0.4203	0.1985	0.0829	0.0310	0.0104	0.0031	0.0008	0.0002	0.0000
	2	0.9335	0.7054	0.4413	0.2369	0.1113	0.0462	0.0170	0.0055	0.0015	0.0004
	3	0.9868	0.8850	0.6841	0.4551	0.2631	0.1332	0.0591	0.0230	0.0077	0.0022
	4	0.9980	0.9648	0.8556	0.6733	0.4654	0.2822	0.1500	0.0696	0.0280	0.0096
	5	0.9998	0.9914	0.9463	0.8369	0.6678	0.4739	0.2968	0.1629	0.0777	0.0318
	6	1.0000	0.9983	0.9837	0.9324	0.8251	0.6655	0.4812	0.3081	0.1727	0.0835
	7	1.0000	0.9997	0.9959	0.9767	0.9225	0.8180	0.6656	0.4878	0.3169	0.1796

Probabilidades acumuladas de la distribución binomial *(Continuación)*

		p								
	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
n = 16	11	0.9147	0.8334	0.7108	0.5501	0.3698	0.2018	0.0791	0.0170	0.0009
	12	0.9719	0.9349	0.8661	0.7541	0.5950	0.4019	0.2101	0.0684	0.0070
	13	0.9934	0.9817	0.9549	0.9006	0.8029	0.6482	0.4386	0.2108	0.0429
	14	0.9990	0.9967	0.9902	0.9739	0.9365	0.8593	0.7161	0.4853	0.1892
	15	0.9999	0.9997	0.9990	0.9967	0.9900	0.9719	0.9257	0.8147	0.5599
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 17	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0086	0.0025	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0301	0.0106	0.0030	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	6	0.0826	0.0348	0.0120	0.0032	0.0006	0.0001	0.0000	0.0000	0.0000
	7	0.1834	0.0919	0.0383	0.0127	0.0031	0.0005	0.0000	0.0000	0.0000
	8	0.3374	0.1989	0.0994	0.0403	0.0124	0.0026	0.0003	0.0000	0.0000
	9	0.5257	0.3595	0.2128	0.1046	0.0402	0.0109	0.0017	0.0001	0.0000
	10	0.7098	0.5522	0.3812	0.2248	0.1071	0.0377	0.0083	0.0008	0.0000
	11	0.8529	0.7361	0.5803	0.4032	0.2347	0.1057	0.0319	0.0047	0.0001
	12	0.9404	0.8740	0.7652	0.6113	0.4261	0.2418	0.0987	0.0221	0.0012
	13	0.9816	0.9536	0.8972	0.7981	0.6470	0.4511	0.2444	0.0826	0.0088
	14	0.9959	0.9877	0.9673	0.9226	0.8363	0.6904	0.4802	0.2382	0.0503
	15	0.9994	0.9979	0.9933	0.9807	0.9499	0.8818	0.7475	0.5182	0.2078
	16	1.0000	0.9998	0.9993	0.9977	0.9925	0.9775	0.9369	0.8332	0.5819
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 18	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0010	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0049	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0183	0.0058	0.0014	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0537	0.0203	0.0062	0.0014	0.0002	0.0000	0.0000	0.0000	0.0000
	7	0.1280	0.0576	0.0212	0.0061	0.0012	0.0002	0.0000	0.0000	0.0000
	8	0.2527	0.1347	0.0597	0.0210	0.0054	0.0009	0.0001	0.0000	0.0000
	9	0.4222	0.2632	0.1391	0.0596	0.0193	0.0043	0.0005	0.0000	0.0000
	10	0.6085	0.4366	0.2717	0.1407	0.0569	0.0163	0.0027	0.0002	0.0000
	11	0.7742	0.6257	0.4509	0.2783	0.1390	0.0513	0.0118	0.0012	0.0000
	12	0.8923	0.7912	0.6450	0.4656	0.2825	0.1329	0.0419	0.0064	0.0002
	13	0.9589	0.9058	0.8114	0.6673	0.4813	0.2836	0.1206	0.0282	0.0015
	14	0.9880	0.9672	0.9217	0.8354	0.6943	0.4990	0.2798	0.0982	0.0109
	15	0.9975	0.9918	0.9764	0.9400	0.8647	0.7287	0.5203	0.2662	0.0581
	16	0.9997	0.9987	0.9954	0.9858	0.9605	0.9009	0.7759	0.5497	0.2265
	17	1.0000	0.9999	0.9996	0.9984	0.9944	0.9820	0.9464	0.8499	0.6028
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 19	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0028	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0109	0.0031	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0342	0.0116	0.0031	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	7	0.0871	0.0352	0.0114	0.0028	0.0005	0.0000	0.0000	0.0000	0.0000

Probabilidades acumuladas de la distribución binomial *(Continuación)*

		p									
	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 19	8	1.0000	1.0000	0.9992	0.9933	0.9713	0.9161	0.8145	0.6675	0.4940	0.3238
	9	1.0000	1.0000	0.9999	0.9984	0.9911	0.9674	0.9125	0.8139	0.6710	0.5000
	10	1.0000	1.0000	1.0000	0.9997	0.9977	0.9895	0.9653	0.9115	0.8159	0.6762
	11	1.0000	1.0000	1.0000	1.0000	0.9995	0.9972	0.9886	0.9648	0.9129	0.8204
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9884	0.9658	0.9165
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9969	0.9891	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9972	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1,0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0,0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 21	0	0.3406	0.1094	0.0329	0.0092	0.0024	0.0006	0.0001	0.0000	0.0000	0.0000
	1	0.7170	0.3647	0.1550	0.0576	0,0190	0.0056	0.0014	0.0003	0.0001	0.0000
	2	0.9151	0.6484	0.3705	0.1787	0.0745	0.0271	0.0086	0.0024	0.0006	0.0001
	3	0.9811	0.8480	0.6113	0.3704	0.1917	0.0856	0.0331	0.0110	0.0031	0.0007
	4	0.9968	0.9478	0.8025	0.5860	0.3674	0.1984	0.0924	0.0370	0.0126	0.0036
	5	0.9996	0.9856	0.9173	0.7693	0.5666	0.3627	0.2009	0.0957	0.0389	0.0133
	6	1.0000	0.9967	0.9713	0.8915	0.7436	0.5505	0.3567	0.2002	0.0964	0.0392
	7	1.0000	0.9994	0.9917	0.9569	0.8701	0.7230	0.5365	0.3495	0.1971	0.0946
	8	1.0000	0.9999	0.9980	0.9856	0.9439	0.8523	0.7059	0.5237	0.3413	0.1917
	9	1.0000	1.0000	0.9996	0.9959	0.9794	0.9324	0.8377	0.6914	0.5117	0.3318
	10	1.0000	1.0000	0.9999	0.9990	0.9936	0.9736	0.9228	0.8256	0.6790	0.5000
	11	1.0000	1.0000	1.0000	0.9998	0.9983	0.9913	0.9687	0.9151	0.8159	0.6682
	12	1.0000	1.0000	1.0000	1.0000	0.9996	0.9976	0.9892	0.9648	0.9092	0.8083
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9969	0.9877	0.9621	0.9054
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9964	0.9868	0.9608
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9867
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9964
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

Probabilidades acumuladas de la distribución binomial *(Continuación)*

		p								
	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
n = 19	8	0.1841	0.0885	0.0347	0.0105	0.0023	0.0003	0.0000	0.0000	0.0000
	9	0.3290	0.1861	0.0875	0.0326	0.0089	0.0016	0.0001	0.0000	0.0000
	10	0.5060	0.3325	0.1855	0.0839	0.0287	0.0067	0.0008	0.0000	0.0000
	11	0.6831	0.5122	0.3344	0.1820	0.0775	0.0233	0.0041	0.0003	0.0000
	12	0.8273	0.6919	0.5188	0.3345	0.1749	0.0676	0.0163	0.0017	0.0000
	13	0.9223	0.8371	0.7032	0.5261	0.3322	0.1631	0.0537	0.0086	0.0002
	14	0.9720	0.9304	0.8500	0.7178	0.5346	0.3267	0.1444	0.0352	0.0020
	15	0.9923	0.9770	0.9409	0.8668	0.7369	0.5449	0.3159	0.1150	0.0132
	16	0.9985	0.9945	0.9830	0.9538	0.8887	0.7631	0.5587	0.2946	0.0665
	17	0.9998	0.9992	0.9969	0.9896	0.9690	0.9171	0.8015	0.5797	0.2453
	18	1.0000	0.9999	0.9997	0.9989	0.9958	0.9856	0.9544	0.8649	0.6226
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 20	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0015	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0064	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0214	0.0065	0.0015	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0580	0.0210	0.0060	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000
	8	0.1308	0.0565	0.0196	0.0051	0.0009	0.0001	0.0000	0.0000	0.0000
	9	0.2493	0.1275	0.0532	0.0171	0.0039	0.0006	0.0000	0.0000	0.0000
	10	0.4086	0.2447	0.1218	0.0480	0.0139	0.0026	0.0002	0.0000	0.0000
	11	0.5857	0.4044	0.2376	0.1133	0.0409	0.0100	0.0013	0.0001	0.0000
	12	0.7480	0.5841	0.3990	0.2277	0.1018	0.0321	0.0059	0.0004	0.0000
	13	0.8701	0.7500	0.5834	0.3920	0.2142	0.0867	0.0219	0.0024	0.0000
	14	0.9447	0.8744	0.7546	0.5836	0.3828	0.1958	0.0673	0.0113	0.0003
	15	0.9811	0.9490	0.8818	0.7625	0.5852	0.3704	0.1702	0.0432	0.0026
	16	0.9951	0.9840	0.9556	0.8929	0.7748	0.5886	0.3523	0.1330	0.0159
	17	0.9991	0.9964	0.9879	0.9645	0.9087	0.7939	0.5951	0.3231	0.0755
	18	0.9999	0.9995	0.9979	0.9924	0.9757	0.9308	0.8244	0.6083	0.2642
	19	1.0000	1.0000	0.9998	0.9992	0.9968	0.9885	0.9612	0.8784	0.6415
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 21	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0008	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0132	0.0036	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0379	0.0123	0.0031	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000
	8	0.0908	0.0352	0.0108	0.0024	0.0004	0.0000	0.0000	0.0000	0.0000
	9	0.1841	0.0849	0.0313	0.0087	0.0017	0.0002	0.0000	0.0000	0.0000
	10	0.3210	0.1744	0.0772	0.0264	0.0064	0.0010	0.0001	0.0000	0.0000
	11	0.4883	0.3086	0.1623	0.0676	0.0206	0.0041	0.0004	0.0000	0.0000
	12	0.6587	0.4763	0.2941	0.1477	0.0561	0.0144	0.0020	0.0001	0.0000
	13	0.8029	0.6505	0.4635	0.2770	0.1299	0.0431	0.0083	0.0006	0.0000
	14	0.9036	0.7998	0.6433	0.4495	0.2564	0.1085	0.0287	0.0033	0.0000
	15	0.9611	0.9043	0.7991	0.6373	0.4334	0.2307	0.0827	0.0144	0.0004
	16	0.9874	0.9630	0.9076	0.8016	0.6326	0.4140	0.1975	0.0522	0.0032
	17	0.9969	0.9890	0.9669	0.9144	0.8083	0.6296	0.3887	0.1520	0.0189
	18	0.9994	0.9976	0.9914	0.9729	0.9255	0.8213	0.6295	0.3516	0.0849

Probabilidades acumuladas de la distribución binomial (*Continuación*)

						p					
	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
n = 21	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 22	0	0.3235	0.0985	0.0280	0.0074	0.0018	0.0004	0.0001	0.0000	0.0000	0.0000
	1	0.6982	0.3392	0.1367	0.0480	0.0149	0.0041	0.0010	0.0002	0.0000	0.0000
	2	0.9052	0.6200	0.3382	0.1545	0.0606	0.0207	0.0061	0.0016	0.0003	0.0001
	3	0.9778	0.8281	0.5752	0.3320	0.1624	0.0681	0.0245	0.0076	0.0020	0.0004
	4	0.9960	0.9379	0.7738	0.5429	0.3235	0.1645	0.0716	0.0266	0.0083	0.0022
	5	0.9994	0.9818	0.9001	0.7326	0.5168	0.3134	0.1629	0.0722	0.0271	0.0085
	6	0.9999	0.9956	0.9632	0.8670	0.6994	0.4942	0.3022	0.1584	0.0705	0.0262
	7	1.0000	0.9991	0.9886	0.9439	0.8385	0.6713	0.4736	0.2898	0.1518	0.0669
	8	1.0000	0.9999	0.9970	0.9799	0.9254	0.8135	0.6466	0.4540	0.2764	0.1431
	9	1.0000	1.0000	0.9993	0.9939	0.9705	0.9084	0.7916	0.6244	0.4350	0.2617
	10	1.0000	1.0000	0.9999	0.9984	0.9900	0.9613	0.8930	0.7720	0.6037	0.4159
	11	1.0000	1.0000	1.0000	0.9997	0.9971	0.9860	0.9526	0.8793	0.7543	0.5841
	12	1.0000	1.0000	1.0000	0.9999	0.9993	0.9957	0.9820	0.9449	0.8672	0.7383
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9989	0.9942	0.9785	0.9383	0.8569
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9930	0.9757	0.9331
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9981	0.9920	0.9738
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9979	0.9915
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9978
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 23	0	0.3074	0.0886	0.0238	0.0059	0.0013	0.0003	0.0000	0.0000	0.0000	0.0000
	1	0.6794	0.3151	0.1204	0.0398	0.0116	0.0030	0.0007	0.0001	0.0000	0.0000
	2	0.8948	0.5920	0.3080	0.1332	0.0492	0.0157	0.0043	0.0010	0.0002	0.0000
	3	0.9742	0.8073	0.5396	0.2965	0.1370	0.0538	0.0181	0.0052	0.0012	0.0002
	4	0.9951	0.9269	0.7440	0.5007	0.2832	0.1356	0.0551	0.0190	0.0055	0.0013
	5	0.9992	0.9774	0.8811	0.6947	0.4685	0.2688	0.1309	0.0540	0.0186	0.0053
	6	0.9999	0.9942	0.9537	0.8402	0.6537	0.4399	0.2534	0.1240	0.0510	0.0173
	7	1.0000	0.9988	0.9848	0.9285	0.8037	0.6181	0.4136	0.2373	0.1152	0.0466
	8	1.0000	0.9998	0.9958	0.9727	0.9037	0.7709	0.5860	0.3884	0.2203	0.1050
	9	1.0000	1.0000	0.9990	0.9911	0.9592	0.8799	0.7408	0.5562	0.3636	0.2024
	10	1.0000	1.0000	0.9998	0.9975	0.9851	0.9454	0.8575	0.7129	0.5278	0.3388
	11	1.0000	1.0000	1.0000	0.9994	0.9954	0.9786	0.9318	0.8364	0.6865	0.5000
	12	1.0000	1.0000	1.0000	0.9999	0.9988	0.9928	0.9717	0.9187	0.8164	0.6612
	13	1.0000	1.0000	1.0000	1.0000	0.9997	0.9979	0.9900	0.9651	0.9063	0.7976
	14	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9970	0.9872	0.9589	0.8950
	15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9960	0.9847	0.9534
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9990	0.9952	0.9827
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9947
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Probabilidades acumuladas de la distribución binomial (*Continuación*)

						p				
	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
n = 21	19	0.9999	0.9997	0.9986	0.9944	0.9810	0.9424	0.8450	0.6353	0.2830
	20	1.0000	1.0000	0.9999	0.9994	0.9976	0.9908	0.9671	0.8906	0.6594
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 22	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0021	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0080	0.0019	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0243	0.0070	0.0016	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	8	0.0617	0.0215	0.0058	0.0011	0.0001	0.0000	0.0000	0.0000	0.0000
	9	0.1328	0.0551	0.0180	0.0043	0.0007	0.0001	0.0000	0.0000	0.0000
	10	0.2457	0.1207	0.0474	0.0140	0.0029	0.0003	0.0000	0.0000	0.0000
	11	0.3963	0.2280	0.1070	0.0387	0.0100	0.0016	0.0001	0.0000	0.0000
	12	0.5650	0.3756	0.2084	0.0916	0.0295	0.0061	0.0007	0.0000	0.0000
	13	0.7236	0.5460	0.3534	0.1865	0.0746	0.0201	0.0030	0.0001	0.0000
	14	0.8482	0.7102	0.5264	0.3287	0.1615	0.0561	0.0114	0.0009	0.0000
	15	0.9295	0.8416	0.6978	0.5058	0.3006	0.1330	0.0368	0.0044	0.0001
	16	0.9729	0.9278	0.8371	0.6866	0.4832	0.2674	0.0999	0.0182	0.0006
	17	0.9917	0.9734	0.9284	0.8355	0.6765	0.4571	0.2262	0.0621	0.0040
	18	0.9980	0.9924	0.9755	0.9319	0.8376	0.6680	0.4248	0.1719	0.0222
	19	0.9997	0.9984	0.9939	0.9793	0.9394	0.8455	0.6618	0.3800	0.0948
	20	1.0000	0.9998	0.9990	0.9959	0.9851	0.9520	0.8633	0.6608	0.3018
	21	1.0000	1.0000	0.9999	0.9996	0.9982	0.9926	0.9720	0.9015	0.6765
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 23	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0048	0.0010	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0153	0.0040	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	8	0.0411	0.0128	0.0030	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000
	9	0.0937	0.0349	0.0100	0.0021	0.0003	0.0000	0.0000	0.0000	0.0000
	10	0.1836	0.0813	0.0283	0.0072	0.0012	0.0001	0.0000	0.0000	0.0000
	11	0.3135	0.1636	0.0682	0.0214	0.0046	0.0006	0.0000	0.0000	0.0000
	12	0.4722	0.2871	0.1425	0.0546	0.0149	0.0025	0.0002	0.0000	0.0000
	13	0.6364	0.4438	0.2592	0.1201	0.0408	0.0089	0.0010	0.0000	0.0000
	14	0.7797	0.6116	0.4140	0.2291	0.0963	0.0273	0.0042	0.0002	0.0000
	15	0.8848	0.7627	0.5864	0.3819	0.1963	0.0715	0.0152	0.0012	0.0000
	16	0.9490	0.8760	0.7466	0.5601	0.3463	0.1598	0.0463	0.0058	0.0001
	17	0.9814	0.9460	0.8691	0.7312	0.5315	0.3053	0.1189	0.0226	0.0008
	18	0.9945	0.9810	0.9449	0.8644	0.7168	0.4993	0.2560	0.0731	0.0049
	19	0.9988	0.9948	0.9819	0.9462	0.8630	0.7035	0.4604	0.1927	0.0258
	20	0.9998	0.9990	0.9957	0.9843	0.9508	0.8668	0.6920	0.4080	0.1052
	21	1.0000	0.9999	0.9993	0.9970	0.9884	0.9602	0.8796	0.6849	0.3206
	22	1.0000	1.0000	1.0000	0.9997	0.9987	0.9941	0.9762	0.9114	0.6926
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Probabilidades acumuladas de la distribución binomial (*Continuación*)

	c	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$n = 24$	0	0.2920	0.0798	0.0202	0.0047	0.0010	0.0002	0.0000	0.0000	0.0000	0.0000
	1	0.6608	0.2925	0.1059	0.0331	0.0090	0.0022	0.0005	0.0001	0.0000	0.0000
	2	0.8841	0.5643	0.2798	0.1145	0.0398	0.0119	0.0030	0.0007	0.0001	0.0000
	3	0.9702	0.7857	0.5049	0.2639	0.1150	0.0424	0.0133	0.0035	0.0008	0.0001
	4	0.9940	0.9149	0.7134	0.4599	0.2466	0.1111	0.0422	0.0134	0.0036	0.0008
	5	0.9990	0.9723	0.8606	0.6559	0.4222	0.2288	0.1044	0.0400	0.0127	0.0033
	6	0.9999	0.9925	0.9428	0.8111	0.6074	0.3886	0.2106	0.0960	0.0364	0.0113
	7	1.0000	0.9983	0.9801	0.9108	0.7662	0.5647	0.3575	0.1919	0.0863	0.0320
	8	1.0000	0.9997	0.9941	0.9638	0.8787	0.7250	0.5257	0.3279	0.1730	0.0758
	9	1.0000	0.9999	0.9985	0.9874	0.9453	0.8472	0.6866	0.4891	0.2991	0.1537
	10	1.0000	1.0000	0.9997	0.9962	0.9787	0.9258	0.8167	0.6502	0.4539	0.2706
	11	1.0000	1.0000	0.9999	0.9990	0.9928	0.9686	0.9058	0.7870	0.6151	0.4194
	12	1.0000	1.0000	1.0000	0.9998	0.9979	0.9885	0.9577	0.8857	0.7580	0.5806
	13	1.0000	1.0000	1.0000	1.0000	0.9995	0.9964	0.9836	0.9465	0.8659	0.7294
	14	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9945	0.9783	0.9352	0.8463
	15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9925	0.9731	0.9242
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9978	0.9905	0.9680
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9972	0.9887
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9967
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n = 25$	0	0.2774	0.0718	0.0172	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
	1	0.6424	0.2712	0.0931	0.0274	0.0070	0.0016	0.0003	0.0001	0.0000	0.0000
	2	0.8729	0.5371	0.2537	0.0982	0.0321	0.0090	0.0021	0.0004	0.0001	0.0000
	3	0.9659	0.7636	0.4711	0.2340	0.0962	0.0332	0.0097	0.0024	0.0005	0.0001
	4	0.9928	0.9020	0.6821	0.4207	0.2137	0.0905	0.0320	0.0095	0.0023	0.0005
	5	0.9988	0.9666	0.8385	0.6167	0.3783	0.1935	0.0826	0.0294	0.0086	0.0020
	6	0.9998	0.9905	0.9305	0.7800	0.5611	0.3407	0.1734	0.0736	0.0258	0.0073
	7	1.0000	0.9977	0.9745	0.8909	0.7265	0.5118	0.3061	0.1536	0.0639	0.0216
	8	1.0000	0.9995	0.9920	0.9532	0.8506	0.6769	0.4668	0.2735	0.1340	0.0539
	9	1.0000	0.9999	0.9979	0.9827	0.9287	0.8106	0.6303	0.4246	0.2424	0.1148
	10	1.0000	1.0000	0.9995	0.9944	0.9703	0.9022	0.7712	0.5858	0.3843	0.2122
	11	1.0000	1.0000	0.9999	0.9985	0.9893	0.9558	0.8746	0.7323	0.5426	0.3450
	12	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.9396	0.8462	0.6937	0.5000
	13	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9745	0.9222	0.8173	0.6550
	14	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9907	0.9656	0.9040	0.7878
	15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9971	0.9868	0.9560	0.8852
	16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9957	0.9826	0.9461
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9988	0.9942	0.9784
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9927
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9980
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1 0000	1.0000	1.0000	0.9999	0.9995
	21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	23	1.0000	1.0000	1.0000	1.0000	1.0000	1 0000	1.0000	1.0000	1.0000	1.0000
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Probabilidades acumuladas de la distribución binomial (*Continuación*)

		p								
	c	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
n = 24	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0028	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0095	0.0022	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	8	0.0269	0.0075	0.0016	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	9	0.0648	0.0217	0.0055	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000
	10	0.1341	0.0535	0.0164	0.0036	0.0005	0.0000	0.0000	0.0000	0.0000
	11	0.2420	0.1143	0.0423	0.0115	0.0021	0.0002	0.0000	0.0000	0.0000
	12	0.3849	0.2130	0.0942	0.0314	0.0072	0.0010	0.0001	0.0000	0.0000
	13	0.5461	0.3498	0.1833	0.0742	0.0213	0.0038	0.0003	0.0000	0.0000
	14	0.7009	0.5109	0.3134	0.1528	0.0547	0.0126	0.0015	0.0001	0.0000
	15	0.8270	0.6721	0.4743	0.2750	0.1213	0.0362	0.0059	0.0003	0.0000
	16	0.9137	0.8081	0.6425	0.4353	0.2338	0.0892	0.0199	0.0017	0.0000
	17	0.9636	0.9040	0.7894	0.6114	0.3926	0.1889	0.0572	0.0075	0.0001
	18	0.9873	0.9600	0.8956	0.7712	0.5778	0 3441	0.1394	0.0277	0.0010
	19	0.9964	0.9866	0.9578	0.8889	0.7534	0.5401	0.2866	0.0851	0.0060
	20	0.9992	0.9965	0.9867	0.9576	0.8850	0.7361	0.4951	0.2143	0.0298
	21	0.9999	0.9993	0.9970	0.9881	0.9602	0.8855	0.7202	0.4357	0.1159
	22	1.0000	0.9999	0.9995	0.9978	0.9910	0.9669	0.8941	0.7075	0.3392
	23	1.0000	1.0000	1.0000	0.9998	0.9990	0.9953	0.9798	0.9202	0.7080
	24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n = 25	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.0016	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	7	0.0058	0.0012	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	8	0.0174	0.0043	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	9	0.0440	0.0132	0.0029	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
	10	0.0960	0.0344	0.0093	0.0018	0.0002	0.0000	0.0000	0.0000	0.0000
	11	0.1827	0.0778	0.0255	0.0060	0.0009	0.0001	0.0000	0.0000	0.0000
	12	0.3063	0.1538	0.0604	0.0175	0.0034	0.0004	0.0000	0.0000	0.0000
	13	0.4574	0.2677	0.1254	0.0442	0.0107	0.0015	0.0001	0.0000	0.0000
	14	0.6157	0.4142	0.2288	0.0978	0.0297	0.0056	0.0005	0.0000	0.0000
	15	0.7576	0.5754	0.3697	0.1894	0.0713	0.0173	0.0021	0.0001	0.0000
	16	0.8660	0.7265	0.5332	0.3231	0.1494	0.0468	0.0080	0.0005	0.0000
	17	0.9361	0.8464	0.6939	0.4882	0.2735	0.1091	0.0255	0.0023	0.0000
	18	0.9742	0.9264	0.8266	0.6593	0.4389	0.2200	0.0695	0.0095	0.0002
	19	0.9914	0.9706	0.9174	0.8065	0.6217	0.3833	0.1615	0.0334	0.0012
	20	0.9977	0.9905	0.9680	0.9095	0.7863	0.5793	0.3179	0.0980	0.0072
	21	0.9995	0.9976	0.9903	0.9668	0.9038	0.7660	0.5289	0.2364	0.0341
	22	0.9999	0.9996	0.9979	0.9910	0.9679	0.9018	0.7463	0.4629	0.1271
	23	1.0000	0.9999	0.9997	0.9984	0.9930	0.9726	0.9069	0.7288	0.3576
	24	1.0000	1.0000	1.0000	0.9999	0.9992	0.9962	0.9828	0.9282	0.7226
	25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Probabilidades acumuladas de la distribución de Poisson

	λ									
c	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5
0	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067	0.0041
1	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404	0.0266
2	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247	0.0884
3	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650	0.2017
4	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405	0.3575
5	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160	0.5289
6	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622	0.6860
7	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666	0.8095
8	1.0000	1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319	0.8944
9	1.0000	1.0000	1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682	0.9462
10	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9972	0.9933	0.9863	0.9747
11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9991	0.9976	0.9945	0.9890
12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980	0.9955
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9993	0.9983
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	λ									
c	11.0	11.5	12.0	12.5	13.0	13.5	14.0	14.5	15.0	15.5
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0002	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0012	0.0008	0.0005	0.0003	0.0002	0.0001	0.0001	0.0001	0.0000	0.0000
3	0.0049	0.0034	0.0023	0.0016	0.0011	0.0007	0.0005	0.0003	0.0002	0.0001
4	0.0151	0.0107	0.0076	0.0053	0.0037	0.0026	0.0018	0.0012	0.0009	0.0006
5	0.0375	0.0277	0.0203	0.0148	0.0107	0.0077	0.0055	0.0039	0.0028	0.0020
6	0.0786	0.0603	0.0458	0.0346	0.0259	0.0193	0.0142	0.0105	0.0076	0.0055
7	0.1432	0.1137	0.0895	0.0698	0.0540	0.0415	0.0316	0.0239	0.0180	0.0135
8	0.2320	0.1906	0.1550	0.1249	0.0998	0.0790	0.0621	0.0484	0.0374	0.0288
9	0.3405	0.2888	0.2424	0.2014	0.1658	0.1353	0.1094	0.0878	0.0699	0.0552
10	0.4599	0.4017	0.3472	0.2971	0.2517	0.2112	0.1757	0.1449	0.1185	0.0961
11	0.5793	0.5198	0.4616	0.4058	0.3532	0.3045	0.2600	0.2201	0.1848	0.1538
12	0.6887	0.6329	0.5760	0.5190	0.4631	0.4093	0.3585	0.3111	0.2676	0.2283
13	0.7813	0.7330	0.6815	0.6278	0.5730	0.5182	0.4644	0.4125	0.3632	0.3171
14	0.8540	0.8153	0.7720	0.7250	0.6751	0.6233	0.5704	0.5176	0.4657	0.4154
15	0.9074	0.8783	0.8444	0.8060	0.7636	0.7178	0.6694	0.6192	0.5681	0.5170
16	0.9441	0.9236	0.8987	0.8693	0.8355	0.7975	0.7559	0.7112	0.6641	0.6154
17	0.9678	0.9542	0.9370	0.9158	0.8905	0.8609	0.8272	0.7897	0.7489	0.7052
18	0.9823	0.9738	0.9626	0.9481	0.9302	0.9084	0.8826	0.8530	0.8195	0.7825
19	0.9907	0.9857	0.9787	0.9694	0.9573	0.9421	0.9235	0.9012	0.8752	0.8455
20	0.9953	0.9925	0.9884	0.9827	0.9750	0.9649	0.9521	0.9362	0.9170	0.8944

* Si $X \sim P(\lambda)$, la tabla de valores de $P(X \leq c)$, $c = 0, 1, \ldots, 20$. $P(X \leq c) = \sum_{x=0}^{c} \frac{e^{-\lambda}\lambda^{x}}{x!}$

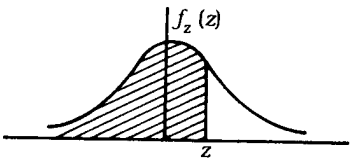
Probabilidades acumuladas de la distribución de Poisson *(Continuación)*

	λ									
c	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0	10.5
0	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000
1	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005	0.0003
2	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0018
3	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0071
4	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293	0.0211
5	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0504
6	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301	0.1016
7	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202	0.1785
8	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328	0.2794
9	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579	0.3971
10	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830	0.5207
11	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968	0.6387
12	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916	0.7420
13	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645	0.8253
14	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165	0.8879
15	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513	0.9317
16	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730	0.9604
17	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911	0.9857	0.9781
18	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957	0.9928	0.9885
19	1.0000	1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980	0.9965	0.9942
20	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9972

	λ								
c	16.0	16.5	17.0	17.5	18.0	18.5	19.0	19.5	20.0
0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0004	0.0003	0.0002	0.0001	0.0001	0.0001	0.0000	0.0000	0.0000
5	0.0014	0.0010	0.0007	0.0005	0.0003	0.0002	0.0002	0.0001	0.0001
6	0.0040	0.0029	0.0021	0.0015	0.0010	0.0007	0.0005	0.0004	0.0003
7	0.0100	0.0074	0.0054	0.0040	0.0029	0.0021	0.0015	0.0011	0.0008
8	0.0220	0.0167	0.0126	0.0095	0.0071	0.0052	0.0039	0.0028	0.0021
9	0.0433	0.0337	0.0261	0.0201	0.0154	0.0117	0.0089	0.0067	0.0050
10	0.0774	0.0619	0.0491	0.0387	0.0304	0.0237	0.0183	0.0141	0.0108
11	0.1270	0.1041	0.0847	0.0684	0.0549	0.0438	0.0347	0.0273	0.0214
12	0.1931	0.1621	0.1350	0.1116	0.0917	0.0748	0.0606	0.0488	0.0390
13	0.2745	0.2357	0.2009	0.1699	0.1426	0.1189	0.0984	0.0809	0.0661
14	0.3675	0.3225	0.2808	0.2426	0.2081	0.1771	0.1497	0.1257	0.1049
15	0.4667	0.4180	0.3715	0.3275	0.2867	0.2490	0.2148	0.1840	0.1565
16	0.5660	0.5165	0.4677	0.4204	0.3751	0.3321	0.2920	0.2550	0.2211
17	0.6593	0.6120	0.5640	0.5160	0.4686	0.4226	0.3784	0.3364	0.2970
18	0.7423	0.6996	0.6550	0.6089	0.5622	0.5156	0.4695	0.4246	0.3814
19	0.8122	0.7757	0.7363	0.6945	0.6509	0.6061	0.5606	0.5151	0.4703
20	0.8682	0.8385	0.8055	0.7694	0.7307	0.6898	0.6472	0.6034	0.5591

Probabilidades acumuladas de la distribución normal estándar

La tabla da el área a la izquierda de un valor de Z o sea $\int_{-\infty}^{Z} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt.$



Z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed. 3a. reimpr., Trillas, México, 1996.

Z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed. 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de $\chi^2_\alpha(\nu)$

En la distribución *Ji*-cuadrada con ν grados de libertad la tabla proporciona el valor $\chi^2_\alpha(\nu)$ tal que $P(\chi^2_\alpha \geq \chi^2_\alpha(\nu)) = \alpha$.

ν \ α	0.001	0.005	0.010	0.025	0.050	0.950	0.975	0.990	0.995	0.999
1	10.8276	7.8794	6.6349	5.0239	3.8415	0.0039	0.0010	0.0002	0.0000	0.0000
2	13.8155	10.5966	9.2103	7.3778	5.9915	0.1026	0.0506	0.0201	0.0100	0.0020
3	16.2662	12.8382	11.3449	9.3484	7.8147	0.3518	0.2158	0.1148	0.0717	0.0243
4	18.4668	14.8603	13.2767	11.1433	9.4877	0.7107	0.4844	0.2971	0.2070	0.0908
5	20.5150	16.7496	15.0863	12.8325	11.0705	1.1455	0.8312	0.5543	0.4117	0.2102
6	22.4577	18.5476	16.8119	14.4494	12.5916	1.6354	1.2373	0.8721	0.6757	0.3811
7	24.3219	20.2777	18.4753	16.0128	14.0671	2.1673	1.6899	1.2390	0.9893	0.5985
8	26.1244	21.9550	20.0902	17.5345	15.5073	2.7326	2.1797	1.6465	1.3444	0.8571
9	27.8771	23.5893	21.6660	19.0228	16.9190	3.3251	2.7004	2.0879	1.7349	1.1519
10	29.5883	25.1882	23.2092	20.4832	18.3070	3.9403	3.2470	2.5582	2.1559	1.4787
11	31.2641	26.7568	24.7250	21.9200	19.6751	4.5748	3.8157	3.0535	2.6032	1.8339
12	32.9095	28.2995	26.2170	23.3367	21.0261	5.2260	4.4038	3.5706	3.0738	2.2142
13	34.5282	29.8195	27.6882	24.7356	22.3620	5.8919	5.0088	4.1069	3.5650	2.6172
14	36.1233	31.3193	29.1412	26.1189	23.6848	6.5706	5.6287	4.6604	4.0747	3.0407
15	37.6973	32.8013	30.5779	27.4884	24.9958	7.2609	6.2621	5.2293	4.6009	3.4827
16	39.2523	34.2672	31.9999	28.8453	26.2962	7.9616	6.9077	5.8122	5.1422	3.9416
17	40.7902	35.7185	33.4087	30.1910	27.5871	8.6718	7.5642	6.4078	5.6972	4.4161
18	42.3124	37.1564	34.8053	31.5264	28.8693	9.3905	8.2307	7.0149	6.2648	4.9048
19	43.8202	38.5823	36.1909	32.8523	30.1435	10.1170	8.9065	7.6327	6.8440	5.4068
20	45.3147	39.9968	37.5662	34.1696	31.4104	10.8508	9.5908	8.2604	7.4338	5.9210
21	46.7970	41.4011	38.9322	35.4789	32.6706	11.5913	10.2829	8.8972	8.0337	6.4467
22	48.2679	42.7956	40.2894	36.7807	33.9244	12.3380	10.9823	9.5425	8.6427	6.9830
23	49.7282	44.1813	41.6384	38.0756	35.1725	13.0905	11.6886	10.1957	9.2604	7.5292
24	51.1786	45.5585	42.9798	39.3641	36.4150	13.8484	12.4012	10.8564	9.8862	8.0849
25	52.6196	46.9279	44.3141	40.6465	37.6525	14.6114	13.1197	11.5240	10.5197	8.6493
26	54.0519	48.2899	45.6417	41.9232	38.8851	15.3792	13.8439	12.1981	11.1602	9.2221
27	55.4760	49.6449	46.9629	43.1945	40.1133	16.1514	14.5734	12.8785	11.8076	9.8028
28	56.8922	50.9934	48.2782	44.4608	41.3371	16.9279	15.3079	13.5647	12.4613	10.3909
29	58.3011	52.3356	49.5879	45.7223	42.5570	17.7084	16.0471	14.2565	13.1211	10.9861
30	59.7030	53.6720	50.8922	46.9792	43.7730	18.4927	16.7908	14.9535	13.7867	11.5880
31	61.0983	55.0027	52.1914	48.2319	44.9853	19.2806	17.5387	15.6555	14.4578	12.1963
32	62.4872	56.3281	53.4858	49.4804	46.1943	20.0719	18.2908	16.3622	15.1340	12.8107
33	63.8701	57.6484	54.7755	50.7251	47.3999	20.8665	19.0467	17.0735	15.8153	13.4309
34	65.2472	58.9639	56.0609	51.9660	48.6024	21.6643	19.8063	17.7891	16.5013	14.0567
35	66.6188	60.2748	57.3421	53.2033	49.8018	22.4650	20.5694	18.5089	17.1918	14.6878
36	67.9851	61.5812	58.6192	54.4373	50.9985	23.2686	21.3359	19.2327	17.8867	15.3241
37	69.3464	62.8833	59.8925	55.6680	52.1923	24.0749	22.1056	19.9602	18.5858	15.9653
38	70.7028	64.1814	61.1621	56.8955	53.3835	24.8839	22.8785	20.6914	19.2889	16.6112
39	72.0546	65.4756	62.4281	58.1201	54.5722	25.6954	23.6543	21.4262	19.9959	17.2616
40	73.4019	66.7660	63.6907	59.3417	55.7585	26.5093	24.4330	22.1643	20.7065	17.9164
41	74.7449	68.0527	64.9501	60.5606	56.9424	27.3256	25.2145	22.9056	21.4208	18.5754
42	76.0837	69.3360	66.2062	61.7768	58.1240	28.1440	25.9987	23.6501	22.1385	19.2385
43	77.4186	70.6159	67.4593	62.9904	59.3035	28.9647	26.7854	24.3976	22.8595	19.9055
44	78.7495	71.8925	68.7095	64.2015	60.4809	29.7875	27.5746	25.1480	23.5837	20.5763
45	80.0767	73.1661	69.9568	65.4102	61.6562	30.6123	28.3662	25.9013	24.3110	21.2507
46	81.4003	74.4365	71.2014	66.6165	62.8296	31.4390	29.1601	26.6572	25.0413	21.9287
47	82.7204	75.7041	72.4433	67.8206	64.0011	32.2676	29.9562	27.4158	25.7746	22.6101
48	84.0371	76.9688	73.6826	69.0226	65.1708	33.0981	30.7545	28.1770	26.5106	23.2949
49	85.3505	78.2307	74.9195	70.2224	66.3386	33.9303	31.5549	28.9406	27.2493	23.9828
50	86.6608	79.4900	76.1539	71.4202	67.5048	34.7643	32.3574	29.7067	27.9907	24.6739

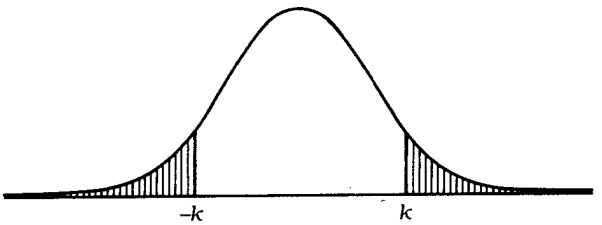
Valores seleccionados de *ji*-cuadrada *(Continuación)*

60	99.6072	91.9517	88.3794	83.2977	79.0819	43.1880	40.4817	37.4849	35.5345	31.7383
70	112.3169	104.2149	100.4252	95.0232	90.5312	51.7393	48.7576	45.4417	43.2752	39.0364
80	124.8392	116.3211	112.3288	106.6286	101.8795	60.3915	57.1532	53.5401	51.1719	46.5199
90	137.2084	128.2989	124.1163	118.1359	113.1453	69.1260	65.6466	61.7541	59.1963	54.1552
100	149.4493	140.1695	135.8067	129.5612	124.3421	77.9295	74.2219	70.0649	67.3276	61.9179

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed. 3a. reimpr., Trillas, México, 1996.

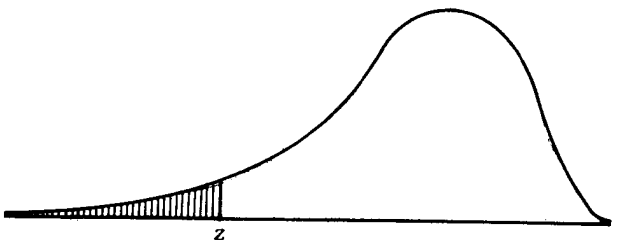
Porcentiles de la distribución normal estandarizada

$P(Z \leq z)$	z
.001	−3.0902
.005	−2.5758
.01	−2.3263
.02	−2.0537
.03	−1.8808
.04	−1.7507
.05	−1.6449
.10	−1.2816
.15	−1.0364
.20	−0.8416
.30	−0.5244
.40	−0.2533
.50	0
.60	0.2533
.70	0.5244
.80	0.8414
.85	1.0364
.90	1.2816
.95	1.6449
.96	1.7507
.97	1.8808
.98	2.0537
.99	2.3263
.995	2.5758
.999	3.0902



Valores críticos correspondientes a los niveles de α, de ambos extremos, de la distribución normal estandarizada.

$P(\lvert Z\rvert > K)$	K
.001	3.2905
.002	3.0902
.005	2.80703
.01	2.5758
.02	2.3263
.03	2.1701
.04	2.0537
.05	1.9600
.06	1.8808
.08	1.7507
.10	1.6449
.15	1.4395
.20	1.2816
.30	1.0364



Valores seleccionados de $t_\alpha(\nu)$.

En la distribución t de Student con ν grados de libertad la tabla proporciona el valor $t_\alpha(\nu)$ tal que $P(t_\nu \geq t_\alpha(\nu)) = \alpha$.

ν \ α	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005	0.00025
1	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	127.3213	318.3088	636.6192	1273.2393
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	14.0890	22.3271	31.5991	44.7046
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409	7.4533	10.2145	12.9240	16.3263
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041	5.5976	7.1732	8.6103	10.3063
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	4.7733	5.8934	6.8688	7.9757
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	4.3168	5.2076	5.9588	6.7883
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	4.0293	4.7853	5.4079	6.0818
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	3.8325	4.5008	5.0413	5.6174
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	3.6897	4.2968	4.7809	5.2907
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814	4.1437	4.5869	5.0490
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	3.4966	4.0247	4.4370	4.8633
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	3.4285	3.9296	4.3178	4.7165
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	3.3725	3.8520	4.2208	4.5975
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	3.3257	3.7874	4.1405	4.4992
15	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	3.2860	3.7329	4.0728	4.4166
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	3.2520	3.6862	4.0150	4.3463
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.2225	3.6458	3.9652	4.2858
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.1966	3.6105	3.9217	4.2332
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.1737	3.5794	3.8834	4.1870
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.1534	3.5518	3.8495	4.1461

Valores seleccionados de $t_\alpha(\nu)$ *(Continuación)*

$\nu \backslash \alpha$	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005	0.00025
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314	3.1352	3.5272	3.8193	4.1096
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3 1188	3.5050	3.7922	4.0770
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.1040	3.4850	3.7677	4.0475
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969	3.0905	3.4668	3.7454	4.0208
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.0782	3.4502	3.7252	3.9965
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	3.0669	3.4350	3.7066	3.9743
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.0565	3.4210	3.6896	3.9539
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.0469	3.4082	3.6739	3.9351
29	0.6830	1.3114	1 6991	2.0452	2.4620	2.7564	3.0380	3.3963	3.6594	3.9178
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.0298	3.3852	3.6460	3.9017
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440	3.0221	3.3749	3.6335	3.8868
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385	3.0150	3.3653	3.6218	3.8728
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333	3.0082	3.3563	3.6109	3.8598
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284	3.0020	3.3479	3.6007	3.8477
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238	2.9960	3.3401	3.5912	3.8363
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195	2.9905	3.3326	3.5822	3.8255
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154	2.9852	3.3256	3.5737	3.8155
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116	2.9803	3.3190	3.5657	3.8059
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079	2.9756	3.3128	3.5581	3.7969
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045	2.9712	3.3069	3.5510	3.7884
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012	2.9670	3.3013	3.5442	3.7804
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981	2.9630	3.2960	3.5378	3.7727
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951	2.9592	3.2909	3.5316	3.7654
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923	2.9555	3.2861	3.5258	3.7585
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896	2.9521	3.2815	3.5203	3.7519
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870	2.9488	3.2771	3.5150	3.7456
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846	2.9456	3.2729	3.5099	3.7397
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822	2.9426	3.2689	3.5051	3.7339
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800	2.9397	3.2651	3.5005	3.7284
50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778	2.9370	3.2614	3.4960	3.7232
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603	2.9146	3.2317	3.4602	3.6807
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479	2.8987	3.2108	3.4350	3.6509
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387	2.8870	3.1953	3.4163	3.6289
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316	2.8779	3.1833	3.4019	3.6119
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259	2.8707	3.1737	3.3905	3.5984
00	0.6740	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291	3.480

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario.* 2a. ed., 3a. reimpr., Trillas, México, 1996

Valores seleccionados de $F^{m}_{n,\,0.01}$.

En la distribución de F con m grados de libertad en el numerador y n grados de libertad en el denominador, la tabla da el valor $F^{m}_{n,\,0.01}$, tal que $P(F^{m}_{n} \geq F^{m}_{n,\,0.01}) = 0.01$.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10	11
1	4052.181	4999.500	5403.352	5624.583	5763.650	5858.986	5928.356	5981.070	6022.473	6055.847	6083.317
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.408
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.133
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.452
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.963
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.790
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.538
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.734
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.178
10	10.044	7 559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.772
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.462
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.220
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	4.025
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.864
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.730
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.616
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.519
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.434
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.360
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.294
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.236
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.184
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.137
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.094
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	3.056
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	3.021
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.988
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.959
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.931
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	3.067	2.979	2.906
31	7.530	5.362	4.484	3.993	3.675	3.449	3.281	3.149	3.043	2.955	2.882
32	7.499	5.336	4.459	3.969	3.652	3.427	3.258	3.127	3.021	2.934	2.860
33	7.471	5.312	4.437	3.948	3.630	3.406	3.238	3.106	3.000	2.913	2.840
34	7.444	5.289	4.416	3.927	3.611	3.386	3.218	3.087	2.981	2.894	2.821
35	7.419	5.268	4.396	3.908	3.592	3.368	3.200	3.069	2.963	2.876	2.803
36	7.396	5.248	4.377	3.890	3.574	3.351	3.183	3.052	2.946	2.859	2.786
37	7.373	5.229	4.360	3.873	3.558	3.334	3.167	3.036	2.930	2.843	2.770
38	7.353	5.211	4.343	3.858	3.542	3.319	3.152	3.021	2.915	2.828	2.755
39	7.333	5.194	4.327	3.843	3.528	3.305	3.137	3.006	2.901	2.814	2.741
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.727
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.785	2.698	2.625
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.559
70	7.011	4.922	4.074	3.600	3.291	3.071	2.906	2.777	2.672	2.585	2.512
80	6.963	4.881	4.036	3.563	3.255	3.036	2.871	2.742	2.637	2.551	2.478
90	6.925	4.849	4.007	3.535	3.228	3.009	2.845	2.715	2.611	2.524	2.451
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.590	2.503	2.430
110	6.871	4.803	3.965	3.495	3.188	2.970	2.806	2.677	2.573	2.486	2.413
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.399
200	6.763	4.713	3.881	3.414	3.110	2.893	2.730	2.601	2.497	2.411	2.338
500	6.686	4.648	3.821	3.357	3.054	2.838	2.675	2.547	2.443	2.356	2.283

n \ m	12	13	14	15	20	24	30	40	60	120	500
1	6106.321	6125.865	6142.674	6157.285	6208.730	6234.631	6260.649	6286.782	6313.030	6339.391	6359.501
2	99.416	99.422	99.428	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.497
3	27.052	26.983	26.924	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.148
4	14.374	14.307	14.249	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.486
5	9.888	9.825	9.770	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.042
6	7.718	7.657	7.605	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.902
7	6.469	6.410	6.359	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.671
8	5.667	5.609	5.559	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.880
9	5.111	5.055	5.005	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.332
10	4.706	4.650	4.601	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.930
11	4.397	4.342	4.293	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.624
12	4.155	4.100	4.052	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.382
13	3.960	3.905	3.857	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.187
14	3.800	3.745	3.698	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.026
15	3.666	3.612	3.564	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.891
16	3.553	3.498	3.451	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.775
17	3.455	3.401	3.353	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.676
18	3.371	3.316	3.269	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.589
19	3.297	3.242	3.195	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.512
20	3.231	3.177	3.130	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.445
21	3.173	3.119	3.072	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.384
22	3.121	3.067	3.019	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.329
23	3.074	3.020	2.973	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.280
24	3.032	2.977	2.930	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.235
25	2.993	2.939	2.892	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.194
26	2.958	2.904	2.857	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.156
27	2.926	2.871	2.824	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.122
28	2.896	2.842	2.795	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.090
29	2.868	2.814	2.767	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.060
30	2.843	2.789	2.742	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.032
31	2.820	2.765	2.718	2.677	2.525	2.445	2.362	2.275	2.183	2.086	2.006
32	2.798	2.744	2.696	2.655	2.503	2.423	2.340	2.252	2.160	2.062	1.982
33	2.777	2.723	2.676	2.634	2.482	2.402	2.319	2.231	2.139	2.040	1.959
34	2.758	2.704	2.657	2.615	2.463	2.383	2.299	2.211	2.118	2.019	1.938
35	2.740	2.686	2.639	2.597	2.445	2.364	2.281	2.193	2.099	2.000	1.918
36	2.723	2.669	2.622	2.580	2.428	2.347	2.263	2.175	2.082	1.981	1.899
37	2.707	2.653	2.606	2.564	2.412	2.331	2.247	2.159	2.065	1.964	1.881
38	2.692	2.638	2.591	2.549	2.397	2.316	2.232	2.143	2.049	1.947	1.864
39	2.678	2.624	2.577	2.535	2.382	2.302	2.217	2.128	2.034	1.932	1.848
40	2.665	2.611	2.563	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.833
50	2.562	2.508	2.461	2.419	2.265	2.183	2.098	2.007	1.909	1.803	1.713
60	2.496	2.442	2.394	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.633
70	2.450	2.395	2.348	2.306	2.150	2.067	1.980	1.886	1.785	1.672	1.574
80	2.415	2.361	2.313	2.271	2.115	2.032	1.944	1.849	1.746	1.630	1.530
90	2.389	2.334	2.286	2.244	2.088	2.004	1.916	1.820	1.716	1.598	1.494
100	2.368	2.313	2.265	2.223	2.067	1.983	1.893	1.797	1.692	1.572	1.466
110	2.350	2.296	2.248	2.206	2.049	1.965	1.875	1.778	1.672	1.551	1.442
120	2.336	2.282	2.234	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.421
200	2.275	2.220	2.172	2.129	1.971	1.886	1.794	1.694	1.583	1.453	1.328
500	2.220	2.166	2.117	2.075	1.915	1.829	1.735	1.633	1.517	1.377	1.232

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de $F^{m}_{n,\,0.025}$.

En la distribución de F con m grados de libertad en el numerador y n grados de libertad en el denominador, la tabla proporciona valores $F^{m}_{n,\,0.025}$, tales que $P(Fmn \geq F^{m}_{n,\,0.025}) = 0.025$.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10	11
1	647.789	799.500	864.163	899.583	921.848	937.111	948.217	956.656	963.285	968.627	973.025
2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398	39.407
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374
4	12.218	10.649	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844	8.794
5	10.007	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.568
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.410
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.709
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.243
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.912
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.665
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.474
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.321
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.197
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.095
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	3.008
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.934
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.870
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.814
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.765
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.721
21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735	2.682
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700	2.647
23	5.750	4.349	3.750	3.408	3.183	3.023	2.902	2.808	2.731	2.668	2.615
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.703	2.640	2.586
25	5.686	4.291	3.694	3.353	3.129	2.969	2.848	2.753	2.677	2.613	2.560
26	5.659	4.265	3.670	3.329	3.105	2.945	2.824	2.729	2.653	2.590	2.536
27	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568	2.514
28	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.547	2.494
29	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.669	2.592	2.529	2.475
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.575	2.511	2.458
31	5.549	4.165	3.573	3.234	3.010	2.851	2.730	2.635	2.558	2.495	2.442
32	5.531	4.149	3.557	3.218	2.995	2.836	2.715	2.620	2.543	2.480	2.426
33	5.515	4.134	3.543	3.204	2.981	2.822	2.701	2.606	2.529	2.466	2.412
34	5.499	4.120	3.529	3.191	2.968	2.808	2.688	2.593	2.516	2.453	2.399
35	5.485	4.106	3.517	3.179	2.956	2.796	2.676	2.581	2.504	2.440	2.387
36	5.471	4.094	3.505	3.167	2.944	2.785	2.664	2.569	2.492	2.429	2.375
37	5.458	4.082	3.493	3.156	2.933	2.774	2.653	2.558	2.481	2.418	2.364
38	5.446	4.071	3.483	3.145	2.923	2.763	2.643	2.548	2.471	2.407	2.353
39	5.435	4.061	3.473	3.135	2.913	2.754	2.633	2.538	2.461	2.397	2.344
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.334
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.381	2.317	2.263
60	5.286	3.925	3.343	3.008	2.786	2.627	2.507	2.412	2.334	2.270	2.216
70	5.247	3.890	3.309	2.975	2.754	2.595	2.474	2.379	2.302	2.237	2.183
80	5.218	3.864	3.284	2.950	2.730	2.571	2.450	2.355	2.277	2.213	2.158
90	5.196	3.844	3.265	2.932	2.711	2.552	2.432	2.336	2.259	2.194	2.140
100	5.179	3.828	3.250	2.917	2.696	2.537	2.417	2.321	2.244	2.179	2.124
110	5.164	3.815	3.237	2.904	2.684	2.525	2.405	2.309	2.232	2.167	2.112
120	5.152	3.805	3.227	2.894	2.674	2.515	2.395	2.299	2.222	2.157	2.102
200	5.100	3.758	3.182	2.850	2.630	2.472	2.351	2.256	2.178	2.113	2.058
500	5.054	3.716	3.142	2.811	2.592	2.434	2.313	2.217	2.139	2.074	2.019

Valores seleccionados de $F^{m}_{n,\,0.025}$ *(Continuación).*

$n \backslash m$	12	13	14	15	20	24	30	40	60	120	500
1	976.708	979.837	982.528	984.867	993.103	997.249	1001.414	1005.598	1009.800	1014.020	1017.240
2	39.415	39.421	39.427	39.431	39.448	39.456	39.465	39.473	39.481	39.490	39.496
3	14.337	14.304	14.277	14.253	14.167	14.124	14.081	14.037	13.992	13.947	13.913
4	8.751	8.715	8.684	8.657	8.560	8.511	8.461	8.411	8.360	8.309	8.270
5	6.525	6.488	6.456	6.428	6.329	6.278	6.227	6.175	6.123	6.069	6.028
6	5.366	5.329	5.297	5.269	5.168	5.117	5.065	5.012	4.959	4.904	4.862
7	4.666	4.628	4.596	4.568	4.467	4.415	4.362	4.309	4.254	4.199	4.156
8	4.200	4.162	4.130	4.101	3.999	3.947	3.894	3.840	3.784	3.728	3.684
9	3.868	3.831	3.798	3.769	3.667	3.614	3.560	3.505	3.449	3.392	3.347
10	3.621	3.583	3.550	3.522	3.419	3.365	3.311	3.255	3.498	3.140	3.094
11	3.430	3.392	3.359	3.330	3.226	3.173	3.118	3.061	3.004	2.944	2.898
12	3.277	3.239	3.206	3.177	3.073	3.019	2.963	2.906	2.848	2.787	2.740
13	3.153	3.115	3.082	3.053	2.948	2.893	2.837	2.780	2.720	2.659	2.611
14	3.050	3.012	2.979	2.949	2.844	2.789	2.732	2.674	2.614	2.552	2.503
15	2.963	2.925	2.891	2.862	2.756	2.701	2.644	2.585	2.524	2.461	2.411
16	2.889	2.851	2.817	2.788	2.681	2.625	2.568	2.509	2.447	2.383	2.333
17	2.825	2.786	2.753	2.723	2.616	2.560	2.502	2.442	2.380	2.315	2.264
18	2.769	2.730	2.696	2.667	2.559	2.503	2.445	2.384	2.321	2.256	2.204
19	2.720	2.681	2.647	2.617	2.509	2.452	2.394	2.333	2.270	2.203	2.150
20	2.676	2.637	2.603	2.573	2.464	2.408	2.349	2.287	2.223	2.156	2.103
21	2.637	2.598	2.564	2.534	2.425	2.368	2.308	2.246	2.182	2.114	2.060
22	2.602	2.563	2.528	2.498	2.389	2.331	2.272	2.210	2.145	2.076	2.021
23	2.570	2.531	2.497	2.466	2.357	2.299	2.239	2.176	2.111	2.041	1.986
24	2.541	2.502	2.468	2.437	2.327	2.269	2.209	2.146	2.080	2.010	1.954
25	2.515	2.476	2.441	2.411	2.300	2.242	2.182	2.118	2.052	1.981	1.924
26	2.491	2.451	2.417	2.387	2.276	2.217	2.157	2.093	2.026	1.954	1.897
27	2.469	2.429	2.395	2.364	2.253	2.195	2.133	2.069	2.002	1.930	1.872
28	2.448	2.409	2.374	2.344	2.232	2.174	2.112	2.048	1.980	1.907	1.848
29	2.430	2.390	2.355	2.325	2.213	2.154	2.092	2.028	1.959	1.886	1.827
30	2.412	2.372	2.338	2.307	2.195	2.136	2.074	2.009	1.940	1.866	1.806
31	2.396	2.356	2.321	2.291	2.178	2.119	2.057	1.991	1.922	1.848	1.788
32	2.381	2.341	2.306	2.275	2.163	2.103	2.041	1.975	1.905	1.831	1.770
33	2.366	2.327	2.292	2.261	2.148	2.088	2.026	1.960	1.890	1.815	1.753
34	2.353	2.313	2.278	2.248	2.135	2.075	2.012	1.946	1.875	1.799	1.737
35	2.341	2.301	2.266	2.235	2.122	2.062	1.999	1.932	1.861	1.785	1.722
36	2.329	2.289	2.254	2.223	2.110	2.049	1.986	1.919	1.848	1.772	1.708
37	2.318	2.278	2.243	2.212	2.098	2.038	1.974	1.907	1.836	1.759	1.695
38	2.307	2.267	2.232	2.201	2.088	2.027	1.963	1.896	1.824	1.747	1.682
39	2.298	2.257	2.222	2.191	2.077	2.017	1.953	1.885	1.813	1.735	1.670
40	2.288	2.248	2.213	2.182	2.068	2.007	1.943	1.875	1.803	1.724	1.659
50	2.216	2.176	2.140	2.109	1.993	1.931	1.866	1.796	1.721	1.639	1.569
60	2.169	2.129	2.093	2.061	1.944	1.882	1.815	1.744	1.667	1.581	1.507
70	2.136	2.095	2.059	2.028	1.910	1.847	1.779	1.707	1.628	1.539	1.463
80	2.111	2.071	2.035	2.003	1.884	1.820	1.752	1.679	1.599	1.508	1.428
90	2.092	2.051	2.015	1.983	1.864	1.800	1.731	1.657	1.576	1.483	1.401
100	2.077	2.036	2.000	1.968	1.849	1.784	1.715	1.640	1.558	1.463	1.378
110	2.065	2.024	1.988	1.955	1.836	1.771	1.701	1.626	1.543	1.447	1.359
120	2.055	2.014	1.977	1.945	1.825	1.760	1.690	1.614	1.530	1.433	1.343
200	2.010	1.969	1.932	1.900	1.778	1.712	1.640	1.562	1.474	1.370	1.269
500	1.971	1.929	1.892	1.859	1.736	1.669	1.596	1.515	1.423	1.311	1.192

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de $F^{m}_{n,\,0.05}$.

Para la distribución de F con m grados de libertad en el numerador y n grados de libertad en el denominador, la tabla da el valor $F^{m}_{n,\,0.05}$ tal que $P(F^{m}_{n} \geq F^{m}_{n,\,0.05}) = 0.05$.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10	11
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882	242.983
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.405
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.763
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.313
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.216
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236	2.198
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220	2.181
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204	2.166
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.151
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.138
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.126
31	4.160	3.305	2.911	2.679	2.523	2.409	2.323	2.255	2.199	2.153	2.114
32	4.149	3.295	2.901	2.668	2.512	2.399	2.313	2.244	2.189	2.142	2.103
33	4.139	3.285	2.892	2.659	2.503	2.389	2.303	2.235	2.179	2.133	2.093
34	4.130	3.276	2.883	2.650	2.494	2.380	2.294	2.225	2.170	2.123	2.084
35	4.121	3.267	2.874	2.641	2.485	2.372	2.285	2.217	2.161	2.114	2.075
36	4.113	3.259	2.866	2.634	2.477	2.364	2.277	2.209	2.153	2.106	2.067
37	4.105	3.252	2.859	2.626	2.470	2.356	2.270	2.201	2.145	2.098	2.059
38	4.098	3.245	2.852	2.619	2.463	2.349	2.262	2.194	2.138	2.091	2.051
39	4.091	3.238	2.845	2.612	2.456	2.342	2.255	2.187	2.131	2.084	2.044
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077	2.038
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026	1.986
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.952
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969	1.928
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.910
90	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043	1.986	1.938	1.897
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.886
110	3.927	3.079	2.687	2.454	2.297	2.182	2.094	2.024	1.966	1.918	1.877
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910	1.869
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878	1.837
500	3.860	3.014	2.623	2.390	2.232	2.117	2.028	1.957	1.899	1.850	1.808

Valores seleccionados de $F^{m}_{n,\,0.05}$ *(Continuación)*

n \ m	12	13	14	15	20	24	30	40	60	120	500
1	243.906	244.690	245.364	245.950	248.013	249.052	250.095	251.143	252.196	253.253	254.059
2	19.413	19.419	19.424	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.494
3	8.745	8.729	8.715	8.703	8.660	8.639	8.617	8.594	8.572	8.549	8.532
4	5.912	5.891	5.873	5.858	5.803	5.774	5.746	5.717	5.688	5.658	5.635
5	4.678	4.655	4.636	4.619	4.558	4.527	4.496	4.464	4.431	4.398	4.373
6	4.000	3.976	3.956	3.938	3.874	3.841	3.808	3.774	3.740	3.705	3.678
7	3.575	3.550	3.529	3.511	3.445	3.410	3.376	3.340	3.304	3.267	3.239
8	3.284	3.259	3.237	3.218	3.150	3.115	3.079	3.043	3.005	2.967	2.937
9	3.073	3.048	3.025	3.006	2.936	2.900	2.864	2.826	2.787	2.748	2.717
10	2.913	2.887	2.865	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.548
11	2.788	2.761	2.739	2.719	2.646	2.609	2.570	2.531	2.490	2.448	2.415
12	2.687	2.660	2.637	2.617	2.544	2.505	2.466	2.426	2.384	2.341	2.307
13	2.604	2.577	2.554	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.218
14	2.534	2.507	2.484	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.142
15	2.475	2.448	2.424	2.403	2.328	2.288	2.247	2.204	2.160	2.114	2.078
16	2.425	2.397	2.373	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.022
17	2.381	2.353	2.329	2.308	2.230	2.190	2.148	2.104	2.058	2.011	1.973
18	2.342	2.314	2.290	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.929
19	2.308	2.280	2.256	2.234	2.155	2.114	2.071	2.026	1.980	1.930	1.891
20	2.278	2.250	2.225	2.203	2.124	2.082	2.039	1.994	1.946	1.896	1.856
21	2.250	2.222	2.197	2.176	2.096	2.054	2.010	1.965	1.916	1.866	1.825
22	2.226	2.198	2.173	2.151	2.071	2.028	1.984	1.938	1.889	1.838	1.797
23	2.204	2.175	2.150	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.771
24	2.183	2.155	2.130	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.747
25	2.165	2.136	2.111	2.089	2.007	1.964	1.919	1.872	1.822	1.768	1.725
26	2.148	2.119	2.094	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.705
27	2.132	2.103	2.078	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.686
28	2.118	2.089	2.064	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.669
29	2.104	2.075	2.050	2.027	1.945	1.901	1.854	1.806	1.754	1.698	1.653
30	2.092	2.063	2.037	2.015	1.932	1.887	1.841	1.792	1.740	1.683	1.637
31	2.080	2.051	2.026	2.003	1.920	1.875	1.828	1.779	1.726	1.670	1.623
32	2.070	2.040	2.015	1.992	1.908	1.864	1.817	1.767	1.714	1.657	1.610
33	2.060	2.030	2.004	1.982	1.898	1.853	1.806	1.756	1.702	1.645	1.597
34	2.050	2.021	1.995	1.972	1.888	1.843	1.795	1.745	1.691	1.633	1.585
35	2.041	2.012	1.986	1.963	1.878	1.833	1.786	1.735	1.681	1.623	1.574
36	2.033	2.003	1.977	1.954	1.870	1.824	1.776	1.726	1.671	1.612	1.564
37	2.025	1.995	1.969	1.946	1.861	1.816	1.768	1.717	1.662	1.603	1.553
38	2.017	1.988	1.962	1.939	1.853	1.808	1.760	1.708	1.653	1.594	1.544
39	2.010	1.981	1.954	1.931	1.846	1.800	1.752	1.700	1.645	1.585	1.535
40	2.003	1.974	1.948	1.924	1.839	1.793	1.744	1.693	1.637	1.577	1.526
50	1.952	1.921	1.895	1.871	1.784	1.737	1.687	1.634	1.576	1.511	1.457
60	1.917	1.887	1.860	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.409
70	1.893	1.863	1.836	1.812	1.722	1.674	1.622	1.566	1.505	1.435	1.374
80	1.875	1.845	1.817	1.793	1.703	1.654	1.602	1.545	1.482	1.411	1.347
90	1.861	1.830	1.803	1.779	1.688	1.639	1.586	1.528	1.465	1.391	1.326
100	1.850	1.819	1.792	1.768	1.676	1.627	1.573	1.515	1.450	1.376	1.308
110	1.841	1.810	1.783	1.758	1.667	1.617	1.563	1.504	1.439	1.363	1.293
120	1.834	1.803	1.775	1.750	1.659	1.608	1.554	1.495	1.429	1.352	1.280
200	1.801	1.769	1.742	1.717	1.623	1.572	1.516	1.455	1.386	1.302	1.221
500	1.772	1.740	1.712	1.686	1.592	1.539	1.482	1.419	1.345	1.255	1.159

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de $F^{m}_{n,\,0.10}$.

En la distribución de F con m grados de libertad en el numerador y n grados de libertad en el denominador, la tabla proporciona valores $F^{m}_{n,\,0.10}$, tales que $P(F^{m}_{n} \geq F^{m}_{n,\,0.10}) = 0.10$.

$n \backslash m$	1	2	3	4	5	6	7	8	9	10	11
1	39.863	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.858	60.195	60.473
2	8.526	9.000	9.162	9.243	9.293	9.326	9.349	9.367	9.381	9.392	9.401
3	5.538	5.462	5.391	5.343	5.309	5.285	5.266	5.252	5.240	5.230	5.222
4	4.545	4.325	4.191	4.107	4.051	4.010	3.979	3.955	3.936	3.920	3.907
5	4.060	3.780	3.619	3.520	3.453	3.405	3.368	3.339	3.316	3.297	3.282
6	3.776	3.463	3.289	3.181	3.108	3.055	3.014	2.983	2.958	2.937	2.920
7	3.589	3.257	3.074	2.961	2.883	2.827	2.785	2.752	2.725	2.703	2.684
8	3.458	3.113	2.924	2.806	2.726	2.668	2.624	2.589	2.561	2.538	2.519
9	3.360	3.006	2.813	2.693	2.611	2.551	2.505	2.469	2.440	2.416	2.396
10	3.285	2.924	2.728	2.605	2.522	2.461	2.414	2.377	2.347	2.323	2.302
11	3.225	2.860	2.660	2.536	2.451	2.389	2.342	2.304	2.274	2.248	2.227
12	3.177	2.807	2.606	2.480	2.394	2.331	2.283	2.245	2.214	2.188	2.166
13	3.136	2.763	2.560	2.434	2.347	2.283	2.234	2.195	2.164	2.138	2.116
14	3.102	2.726	2.522	2.395	2.307	2.243	2.193	2.154	2.122	2.095	2.073
15	3.073	2.695	2.490	2.361	2.273	2.208	2.158	2.119	2.086	2.059	2.037
16	3.048	2.668	2.462	2.333	2.244	2.178	2.128	2.088	2.055	2.028	2.005
17	3.026	2.645	2.437	2.308	2.218	2.152	2.102	2.061	2.028	2.001	1.978
18	3.007	2.624	2.416	2.286	2.196	2.130	2.079	2.038	2.005	1.977	1.954
19	2.990	2.606	2.397	2.266	2.176	2.109	2.058	2.017	1.984	1.956	1.932
20	2.975	2.589	2.380	2.249	2.158	2.091	2.040	1.999	1.965	1.937	1.913
21	2.961	2.575	2.365	2.233	2.142	2.075	2.023	1.982	1.948	1.920	1.896
22	2.949	2.561	2.351	2.219	2.128	2.060	2.008	1.967	1.933	1.904	1.880
23	2.937	2.549	2.339	2.207	2.115	2.047	1.995	1.953	1.919	1.890	1.866
24	2.927	2.538	2.327	2.195	2.103	2.035	1.983	1.941	1.906	1.877	1.853
25	2.918	2.528	2.317	2.184	2.092	2.024	1.971	1.929	1.895	1.866	1.841
26	2.909	2.519	2.307	2.174	2.082	2.014	1.961	1.919	1.884	1.855	1.830
27	2.901	2.511	2.299	2.165	2.703	2.005	1.952	1.909	1.874	1.845	1.820
28	2.894	2.503	2.291	2.157	2.064	1.996	1.943	1.900	1.865	1.836	1.811
29	2.887	2.495	2.283	2.149	2.057	1.988	1.935	1.892	1.857	1.827	1.802
30	2.881	2.489	2.276	2.142	2.049	1.980	1.927	1.884	1.849	1.819	1.794
31	2.875	2.482	2.270	2.136	2.042	1.973	1.920	1.877	1.842	1.812	1.787
32	2.869	2.477	2.263	2.129	2.036	1.967	1.913	1.870	1.835	1.805	1.780
33	2.864	2.471	2.258	2.123	2.030	1.961	1.907	1.864	1.828	1.799	1.773
34	2.859	2.466	2.252	2.118	2.024	1.955	1.901	1.858	1.822	1.793	1.767
35	2.855	2.461	2.247	2.113	2.019	1.950	1.896	1.852	1.817	1.787	1.761
36	2.850	2.456	2.243	2.108	2.014	1.945	1.891	1.847	1.811	1.781	1.756
37	2.846	2.452	2.238	2.103	2.009	1.940	1.886	1.842	1.806	1.776	1.751
38	2.842	2.448	2.234	2.099	2.005	1.935	1.881	1.838	1.802	1.772	1.746
39	2.839	2.444	2.230	2.095	2.001	1.931	1.877	1.833	1.797	1.767	1.741
40	2.835	2.440	2.226	2.091	1.997	1.927	1.873	1.829	1.793	1.763	1.737
50	2.809	2.412	2.197	2.061	1.966	1.895	1.840	1.796	1.760	1.729	1.703
60	2.791	2.393	2.177	2.041	1.946	1.875	1.819	1.775	1.738	1.707	1.680
70	2.779	2.380	2.164	2.027	1.931	1.860	1.804	1.760	1.723	1.691	1.665
80	2.769	2.370	2.154	2.016	1.921	1.849	1.793	1.748	1.711	1.680	1.653
90	2.762	2.363	2.146	2.008	1.912	1.841	1.785	1.739	1.702	1.670	1.643
100	2.756	2.356	2.139	2.002	1.906	1.834	1.778	1.732	1.695	1.663	1.636
110	2.752	2.351	2.134	1.997	1.900	1.828	1.772	1.727	1.689	1.657	1.630
120	2.748	2.347	2.130	1.992	1.896	1.824	1.767	1.722	1.684	1.652	1.625
200	2.731	2.329	2.111	1.973	1.876	1.804	1.747	1.701	1.663	1.631	1.603
500	2.716	2.313	2.095	1.956	1.859	1.786	1.729	1.683	1.644	1.612	1.583

Valores seleccionados de $F^{m}_{n,\,0.10}$ *(Continuación)*

$n \backslash m$	12	13	14	15	20	24	30	40	60	120	500
1	60.705	60.903	61.073	61.220	61.740	62.002	62.265	62.529	62.794	63.061	63.264
2	9.408	9.415	9.420	9.425	9.441	9.450	9.458	9.466	9.475	9.483	9.489
3	5.216	5.210	5.205	5.200	5.184	5.176	5.168	5.160	5.151	5.143	5.136
4	3.896	3.886	3.878	3.870	3.844	3.831	3.817	3.804	3.790	3.775	3.764
5	3.268	3.257	3.247	3.238	3.207	3.191	3.174	3.157	3.140	3.123	3.109
6	2.905	2.892	2.881	2.871	2.836	2.818	2.800	2.781	2.762	2.742	2.727
7	2.668	2.654	2.643	2.632	2.595	2.575	2.555	2.535	2.514	2.493	2.476
8	2.502	2.488	2.475	2.464	2.425	2.404	2.383	2.361	2.339	2.316	2.298
9	2.379	2.364	2.351	2.340	2.298	2.277	2.255	2.232	2.208	2.184	2.165
10	2.284	2.269	2.255	2.244	2.201	2.178	2.155	2.132	2.107	2.082	2.062
11	2.209	2.193	2.179	2.167	2.123	2.100	2.076	2.052	2.026	2.000	1.979
12	2.147	2.131	2.117	2.105	2.060	2.036	2.011	1.986	1.960	1.932	1.911
13	2.097	2.080	2.066	2.053	2.007	1.983	1.958	1.931	1.904	1.876	1.853
14	2.054	2.037	2.022	2.010	1.962	1.938	1.912	1.885	1.857	1.828	1.805
15	2.017	2.000	1.985	1.972	1.924	1.899	1.873	1.845	1.817	1.787	1.763
16	1.985	1.968	1.953	1.940	1.891	1.866	1.839	1.811	1.782	1.751	1.726
17	1.958	1.940	1.925	1.912	1.862	1.836	1.809	1.781	1.751	1.719	1.694
18	1.933	1.916	1.900	1.887	1.837	1.810	1.783	1.754	1.723	1.691	1.665
19	1.912	1.894	1.878	1.865	1.814	1.787	1.759	1.730	1.699	1.666	1.639
20	1.892	1.875	1.859	1.845	1.794	1.767	1.738	1.708	1.677	1.643	1.616
21	1.875	1.857	1.841	1.827	1.776	1.748	1.719	1.689	1.657	1.623	1.595
22	1.859	1.841	1.825	1.811	1.759	1.731	1.702	1.671	1.639	1.604	1.576
23	1.845	1.827	1.811	1.796	1.744	1.716	1.686	1.655	1.622	1.587	1.558
24	1.832	1.814	1.797	1.783	1.730	1.702	1.672	1.641	1.607	1.571	1.542
25	1.820	1.802	1.785	1.771	1.718	1.689	1.659	1.627	1.593	1.557	1.527
26	1.809	1.790	1.774	1.760	1.706	1.677	1.647	1.615	1.581	1.544	1.514
27	1.799	1.780	1.764	1.749	1.695	1.666	1.636	1.603	1.569	1.531	1.501
28	1.790	1.771	1.754	1.740	1.685	1.656	1.625	1.592	1.558	1.520	1.489
29	1.781	1.762	1.745	1.731	1.676	1.647	1.616	1.583	1.547	1.509	1.478
30	1.773	1.754	1.737	1.722	1.667	1.638	1.606	1.573	1.538	1.499	1.467
31	1.765	1.746	1.729	1.714	1.659	1.630	1.598	1.565	1.529	1.489	1.457
32	1.758	1.739	1.722	1.707	1.652	1.622	1.590	1.556	1.520	1.481	1.448
33	1.751	1.732	1.715	1.700	1.645	1.615	1.583	1.549	1.512	1.472	1.439
34	1.745	1.726	1.709	1.694	1.638	1.608	1.576	1.541	1.505	1.464	1.431
35	1.739	1.720	1.703	1.688	1.632	1.601	1.569	1.535	1.497	1.457	1.423
36	1.734	1.715	1.697	1.682	1.626	1.595	1.563	1.528	1.491	1.450	1.415
37	1.729	1.709	1.692	1.677	1.620	1.590	1.557	1.522	1.484	1.443	1.408
38	1.724	1.704	1.687	1.672	1.615	1.584	1.551	1.516	1.478	1.437	1.402
39	1.719	1.700	1.682	1.667	1.610	1.579	1.546	1.511	1.473	1.431	1.395
40	1.715	1.695	1.678	1.662	1.605	1.574	1.541	1.506	1.467	1.425	1.389
50	1.680	1.660	1.643	1.627	1.568	1.536	1.502	1.465	1.424	1.379	1.340
60	1.657	1.637	1.619	1.603	1.543	1.511	1.476	1.437	1.395	1.348	1.306
70	1.641	1.621	1.603	1.587	1.526	1.493	1.457	1.418	1.374	1.325	1.281
80	1.629	1.609	1.590	1.574	1.513	1.479	1.443	1.403	1.358	1.307	1.261
90	1.620	1.599	1.581	1.564	1.503	1.468	1.432	1.391	1.346	1.293	1.245
100	1.612	1.592	1.573	1.557	1.494	1.460	1.423	1.382	1.336	1.282	1.232
110	1.606	1.585	1.567	1.550	1.488	1.453	1.415	1.374	1.32i	1.272	1.221
120	1.601	1.580	1.562	1.545	1.482	1.447	1.409	1.368	1.320	1.265	1.212
200	1.579	1.558	1.539	1.522	1.458	1.422	1.383	1.339	1.289	1.228	1.168
500	1.559	1.537	1.518	1.501	1.435	1.399	1.358	1.313	1.260	1.194	1.122

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de la distribución de la estadística de Mann Whitney.

Para $T_+ = \sum_{i=1}^{n} R(X_i) - \frac{n(n+1)}{2}$, la tabla de valores $T_\alpha(n, m)$ tales que: $P(T_+ > T_\alpha(n, m)) = \alpha$; ($\alpha$ = 0.001, 0.005, 0.01, 0.025, 0.05, 0.10) para diferentes valores de n y m, los tamaños de las muestras. Para obtener valores en la cola izquierda use: $T_{1-\alpha}(n, m) = nm - T_\alpha(n, m)$

		m																		
n	α	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	0.001	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40
	0.005	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	37	39
2	0.010	4	6	8	10	12	14	16	18	20	22	24	25	27	29	31	33	35	36	38
	0.025	4	6	8	10	12	14	15	17	19	21	22	24	26	28	30	31	33	35	37
	0.050	4	6	8	9	11	13	14	16	18	20	21	23	24	26	28	30	31	33	35
	0.100	4	5	7	8	10	12	13	15	16	18	19	21	23	24	26	27	29	30	32
	0.001	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	50	53	56	59
	0.005	6	9	12	15	18	21	24	26	29	32	34	37	40	42	45	48	51	53	56
3	0.010	6	9	12	15	18	20	23	25	28	31	33	36	39	41	44	46	49	52	54
	0.025	6	9	12	14	16	19	21	24	26	29	31	34	36	39	41	44	46	49	51
	0.050	6	8	11	13	15	18	20	22	25	27	30	32	34	37	39	41	44	46	48
	0.100	5	7	10	12	14	16	18	21	23	25	27	29	31	34	36	38	40	42	44
	0.001	8	12	16	20	24	28	32	36	39	43	47	50	54	58	61	65	68	72	76
	0.005	8	12	16	20	23	27	30	34	37	41	44	48	51	54	58	61	65	68	71
4	0.010	8	12	16	19	22	26	29	32	36	39	42	46	49	51	56	59	62	66	69
	0.025	8	12	15	18	21	24	27	31	34	37	40	43	46	49	52	56	59	62	65
	0.050	8	11	14	17	20	23	26	29	32	35	38	41	44	47	49	52	55	58	61
	0.100	7	10	12	15	18	21	24	26	29	32	35	38	40	43	46	49	51	54	57
	0.001	10	15	20	25	30	35	39	43	48	52	57	61	66	70	74	79	83	87	92
	0.005	10	15	20	24	28	33	37	41	45	49	53	57	62	66	70	74	78	82	86
5	0.010	10	15	19	23	27	31	35	39	43	47	51	55	59	63	67	71	75	79	83
	0.025	10	14	18	22	26	29	33	37	41	45	48	52	56	60	64	67	71	75	79
	0.050	9	13	17	20	24	28	31	35	38	42	46	49	53	56	60	64	67	71	74
	0.100	8	12	15	19	22	26	29	32	36	39	42	46	49	52	56	59	62	66	69

Fuente: Said Infante Gil y Guillermo P. Zárate, *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de la distribución de la estadística de Mann Whitney *(Continuación)*

6	0.001	12	18	24	30	36	42	46	51	56	61	67	72	77	82	87	92	97	102	107
	0.005	12	18	23	28	33	38	43	48	53	58	62	67	72	77	82	86	91	96	101
	0.010	12	18	22	27	32	37	41	46	51	56	60	65	70	74	79	83	88	93	97
	0.025	12	16	21	26	30	35	39	43	48	52	57	61	68	70	74	79	83	88	92
	0.050	11	15	20	24	28	33	37	41	45	49	54	58	62	66	70	75	79	83	87
	0.100	10	14	18	22	26	30	34	38	42	46	50	54	58	62	66	70	73	77	81
7	0.001	14	21	28	35	41	47	53	59	64	70	76	82	88	94	100	105	111	117	123
	0.005	14	21	27	33	38	44	49	55	60	66	71	77	82	88	93	99	104	110	115
	0.010	14	20	26	31	37	42	48	53	58	64	69	74	80	85	90	95	101	106	111
	0.025	14	19	24	29	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105
	0.050	13	18	23	28	33	37	42	47	52	57	62	66	71	76	81	85	90	95	100
	0.100	12	16	21	26	30	35	39	44	48	53	57	62	66	71	75	80	84	89	93
8	0.001	16	24	32	39	46	53	59	66	73	79	86	92	99	105	112	118	125	131	138
	0.005	16	24	30	37	43	49	56	62	68	74	80	86	93	99	105	111	117	123	129
	0.010	16	23	29	35	41	48	54	60	66	72	78	83	89	95	101	107	113	119	125
	0.025	15	21	27	33	39	45	50	56	62	68	73	79	85	90	96	101	107	113	118
	0.050	14	20	26	31	37	42	48	53	59	64	69	75	80	86	91	96	102	107	112
	0.100	13	18	24	29	34	39	44	49	55	60	65	70	75	80	85	90	95	100	105
9	0.001	18	27	36	43	51	59	66	73	81	88	95	102	110	117	124	131	138	145	153
	0.005	18	26	34	41	48	55	62	69	76	82	89	96	103	110	116	123	130	137	143
	0.010	18	25	32	39	46	53	60	66	73	80	86	93	99	106	112	119	125	132	139
	0.025	17	24	31	37	43	50	56	63	69	75	81	88	94	100	106	113	119	125	131
	0.050	16	22	29	35	41	47	53	59	65	71	77	83	89	95	101	107	113	119	125
	0.100	15	21	26	32	38	44	49	55	61	67	72	78	84	89	95	100	106	112	117
10	0.001	20	30	39	48	56	64	73	81	89	97	105	112	120	128	136	144	152	160	167
	0.005	20	29	37	45	53	60	68	76	83	91	98	105	113	120	128	135	142	150	157
	0.010	20	28	36	43	51	58	66	73	80	87	95	102	109	116	123	131	138	145	152
	0.025	19	26	34	41	48	55	62	69	76	83	90	96	103	110	117	124	131	137	144
	0.050	18	25	32	38	45	52	59	65	72	78	85	92	98	105	111	118	124	131	137
	0.100	16	23	29	36	42	48	55	61	67	73	80	86	92	98	105	111	117	123	129

Valores seleccionados de la distribución de la estadística de Mann Whitney *(Continuación)*

n	α	m = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
11	0.001	22	33	43	52	61	70	79	88	97	105	114	122	131	140	148	157	165	174	182
	0.005	22	32	41	49	58	66	74	82	91	99	107	115	123	131	139	147	155	163	171
	0.010	22	31	39	47	56	64	72	80	87	95	103	111	119	127	134	142	150	158	166
	0.025	21	29	37	45	52	60	68	75	83	90	98	105	113	120	128	135	142	150	157
	0.050	20	27	35	42	49	57	64	71	78	86	93	100	107	114	121	129	136	143	150
	0.100	18	25	32	39	46	53	60	67	73	80	87	94	101	107	114	121	128	135	141
12	0.001	24	36	47	57	67	76	86	95	105	114	123	132	142	151	160	169	178	187	197
	0.005	24	34	44	53	62	71	80	89	98	107	116	124	133	142	150	159	168	176	185
	0.010	24	33	42	51	60	69	78	86	95	103	112	120	129	137	145	154	162	171	179
	0.025	22	31	40	48	57	65	73	81	90	98	106	114	122	130	138	146	154	162	170
	0.050	21	30	38	46	54	62	69	77	85	93	101	108	116	124	131	139	147	155	162
	0.100	19	27	35	42	50	57	65	72	80	87	94	102	109	116	124	131	138	146	153
13	0.001	26	39	50	61	72	82	92	102	112	122	132	142	152	162	172	182	192	201	211
	0.005	26	37	48	57	67	77	86	96	105	115	124	134	143	152	162	171	180	189	199
	0.010	25	36	46	55	65	74	83	93	102	111	120	129	138	147	156	165	174	183	192
	0.025	24	34	43	52	61	70	79	88	96	105	114	123	131	140	148	157	166	174	183
	0.050	23	32	41	49	58	66	75	83	92	100	108	117	125	133	142	150	158	166	175
	0.100	21	29	38	46	54	62	70	78	86	94	102	110	118	126	133	141	149	157	165
14	0.001	28	42	54	66	77	88	99	110	120	131	142	152	163	173	184	194	205	215	225
	0.005	28	40	51	62	72	82	93	103	113	123	133	143	153	163	173	183	193	202	212
	0.010	27	39	49	59	70	80	89	99	109	119	129	138	148	158	167	177	186	196	206
	0.025	26	36	46	56	66	75	85	94	103	113	122	131	140	150	159	168	177	187	196
	0.050	24	34	44	53	62	71	80	89	98	107	116	125	134	143	152	160	169	178	187
	0.100	22	31	40	49	58	66	75	84	92	101	109	118	126	135	143	152	160	168	177
15	0.001	30	45	58	70	82	94	105	117	128	140	151	162	173	184	196	207	218	229	240
	0.005	30	42	54	66	77	88	99	110	120	131	142	152	163	173	184	194	205	215	226
	0.010	29	41	52	63	74	85	95	106	116	127	137	147	158	168	178	188	199	209	219
	0.025	28	39	49	60	70	80	90	100	110	120	130	140	150	160	169	179	189	199	209
	0.050	26	37	47	56	66	76	86	95	105	114	124	133	143	152	162	171	181	190	199
	0.100	24	34	43	52	62	71	80	89	98	107	116	126	135	144	153	162	171	180	189

Valores seleccionados de la distribución de la estadística de Mann Whitney *(Continuación)*

16	0.001	32	48	61	77	87	100	112	124	136	148	160	172	184	196	207	219	231	243	254
	0.005	32	45	58	74	82	93	105	116	128	139	150	162	173	184	195	206	217	229	240
	0.010	31	44	56	72	79	90	101	112	123	134	145	156	167	178	189	200	211	221	232
	0.025	30	41	52	68	74	85	96	106	117	128	138	148	159	169	180	190	201	211	221
	0.050	28	39	49	65	70	81	91	101	111	121	131	142	152	162	172	182	192	202	212
	0.100	26	36	46	62	66	75	85	95	105	114	124	133	143	153	162	172	181	191	200
17	0.001	34	50	65	79	92	105	118	131	144	157	169	182	194	207	219	231	244	256	269
	0.005	34	48	61	74	86	99	111	123	135	147	159	171	183	194	206	218	230	241	253
	0.010	33	46	59	71	83	95	107	119	131	142	154	165	177	188	200	211	223	234	246
	0.025	31	44	56	67	79	90	101	113	124	135	146	157	168	179	190	201	212	223	234
	0.050	30	41	52	64	75	85	96	107	118	129	139	150	160	171	182	192	203	213	224
	0.100	27	38	49	59	70	80	90	100	111	121	131	141	152	162	172	182	192	202	212
18	0.001	36	53	68	83	97	111	125	138	152	165	178	191	205	218	231	244	257	270	283
	0.005	36	51	65	78	91	104	117	130	142	155	168	180	193	205	217	230	242	254	267
	0.010	35	49	62	75	88	101	113	125	138	150	162	174	186	199	211	223	235	247	259
	0.025	33	46	59	71	83	95	107	119	131	142	154	166	177	189	201	212	224	235	247
	0.050	31	44	55	67	79	90	102	113	124	136	147	158	169	181	192	203	214	225	236
	0.100	29	40	51	62	73	84	95	106	117	128	138	149	160	171	181	192	203	213	224
19	0.001	38	56	72	87	102	117	131	145	160	174	187	201	215	229	243	256	270	283	297
	0.005	37	53	68	82	96	110	123	137	150	163	176	189	202	215	229	241	254	267	280
	0.010	36	52	66	79	93	106	119	132	145	158	171	183	196	209	221	234	247	259	272
	0.025	35	49	62	75	88	100	113	125	137	150	162	174	187	199	211	223	235	247	260
	0.050	33	46	58	71	83	95	107	119	131	143	155	166	178	190	202	213	225	237	249
	0.100	30	42	54	66	77	89	100	112	123	135	146	157	168	180	191	202	213	225	236
20	0.001	40	59	76	92	107	123	138	153	167	182	197	211	225	240	254	269	283	297	311
	0.005	39	56	71	86	101	115	129	143	157	171	185	199	212	226	240	253	267	280	294
	0.010	38	54	69	83	97	111	125	139	152	166	179	192	206	219	232	246	259	272	285
	0.025	37	51	65	79	92	105	118	131	144	157	170	183	196	209	221	234	247	260	272
	0.050	35	48	61	74	87	100	112	125	137	150	162	175	187	199	212	224	236	249	261
	0.100	32	44	57	69	81	93	105	117	129	141	153	165	177	189	200	212	224	236	248

Valores seleccionados de la distribución de la estadística de Kruskal Wallis para 3 tratamientos y tamaños de muestra o iguales que cinco.

Para n_1, n_2 y n_3 dados, la tabla de la probabilidd (p) de un valor mayor que T(la estadística de Kruskal Wallis) si la hipótesis nula es cierta.

n_1	n_2	n_3	T	p
2	1	1	2.700	0.500
			1.800	0.833
2	2	1	3.600	0.200
			3.000	0.333
			2.400	0.467
2	2	2	4.571	0.067
			3.714	0.200
			3.429	0.333
3	1	1	3.200	0.300
			2.133	0.700
3	2	1	4.286	0.100
			3.857	0.133
			3.524	0.200
			3.095	0.283
			2.381	0.433
3	2	2	5.357	0.029
			4.714	0.048
			4.500	0.067
			4.464	0.105
3	3	1	5.143	0.043
			4.571	0.100
			4.000	0.129
3	3	2	6.250	0.011
			5.556	0.025
			5.361	0.032
			5.139	0.061
			5.000	0.075
			4.556	0.100
			4.250	0.121
			4.111	0.129
3	3	3	7.200	0.004
			6.489	0.011
			5.956	0.025
			5.600	0.050
			5.422	0.071
			4.622	0.100
			4.267	0.139
4	1	1	3.571	0.200
			2.500	0.467
4	2	1	4.821	0.057
			4.500	0.076
			4.018	0.114
			3.750	0.133
4	2	2	6.000	0.014
			5.500	0.024
			5.333	0.033
			5.125	0.052
			4.458	0.100
			4.167	0.105
4	3	1	5.833	0.021
			5.389	0.036
			5.208	0.050
			4.764	0.071
			4.097	0.086
			4.056	0.093
			3.889	0.129
			3.764	0.136
4	3	2	7.000	0.005
			6.444	0.008
			6.300	0.011
			6.000	0.024
			5.400	0.051
			4.811	0.076
			4.444	0.102
			4.278	0.124
			4.000	0.149
4	3	3	8.018	0.001
			7.000	0.006
			6.746	0.010
			6.154	0.025
			5.727	0.050
			5.000	0.074
			4.700	0.101
			4.091	0.126
			3.386	0.150
4	4	1	6.667	0.010
			6.167	0.022
			4.969	0.048
			4.867	0.054
			4.267	0.070
			4.167	0.083
			4.067	0.102
			3.900	0.108
4	4	2	7.855	0.002
			7.036	0.006
			6.873	0.011
			6.082	0.025
			5.454	0.046
			5.236	0.052
			4.773	0.075
			4.554	0.098
			4.446	0.103
			4.418	0.120
			4.009	0.142
4	4	3	8.909	0.001
			7.598	0.004
			7.477	0.006
			7.144	0.010
			6.394	0.025
			5.598	0.049
			5.576	0.051
			5.144	0.073
			4.546	0.099
			4.477	0.102
			4.296	0.125
			3.962	0.140
			3.848	0.150
4	4	4	9.846	0.000
			8.654	0.001
			8.000	0.005
			7.538	0.011
			6.615	0.024
			5.692	0.049
			5.654	0.055
			5.115	0.074
			4.654	0.097
			4.500	0.104
			4.269	0.122
			3.962	0.145
5	1	1	3.857	0.143
			2.829	0.333
5	2	1	5.250	0.036
			5.000	0.048
			4.450	0.071
			4.200	0.095
			4.050	0.119
			3.783	0.131

Fuente: Said Infante Gil y Guillermo P. Zárate. *Métodos estadísticos. Un enfoque interdisciplinario.* 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de la distribución de la estadística de Kruskal Wallis *(Continuación)*

n_1	n_2	n_3	T	p
5	2	2	6.533	0.008
			6.133	0.013
			6.000	0.019
			5.693	0.029
			5.160	0.034
			5.040	0.056
			4.373	0.090
			4.293	0.122
			4.093	0.148
5	3	1	6.400	0.012
			6.044	0.020
			5.760	0.028
			4.960	0.048
			4.871	0.052
			4.551	0.075
			4.018	0.095
			3.840	0.123
			3.804	0.131
			3.484	0.135
			3.378	0.143
5	3	2	7.636	0.002
			6.949	0.006
			6.822	0.010
			6.004	0.025
			5.251	0.049
			5.106	0.052
			4.727	0.085
			4.494	0.101
			3.942	0.146
5	3	3	8.727	0.001
			7.515	0.005
			7.079	0.009
			6.982	0.011
			6.303	0.026
			5.648	0.049
			5.515	0.051
			4.533	0.097
			4.412	0.109
			4.194	0.126
			3.927	0.149
5	4	1	7.364	0.005
			6.840	0.011
			5.776	0.025
			4.986	0.044
			4.860	0.055
			3.987	0.098
			3.960	0.102
			3.524	0.146
5	4	2	8.114	0.001
			7.573	0.005
			7.118	0.010
			6.041	0.025
			5.268	0.050
			4.541	0.098
			4.518	0.101
			3.818	0.148
5	4	3	8.503	0.001
			7.906	0.005
			7.445	0.010
			6.410	0.025
			5.631	0.050
			4.549	0.099
			4.523	0.103
			3.831	0.150
5	4	4	8.997	0.001
			8.140	0.005
			7.760	0.009
			7.714	0.011
			5.618	0.050
			4.619	0.100
			4.170	0.125
			3.910	0.146
5	5	1	8.182	0.002
			7.746	0.005
			7.309	0.009
			6.836	0.011
			5.127	0.046
			4.909	0.053
			4.109	0.086
			4.036	0.105
5	5	2	8.685	0.001
			8.131	0.005
			7.269	0.010
			5.338	0.047
			5.246	0.051
			4.508	0.100
			4.231	0.124
			3.862	0.150
5	5	3	9.055	0.001
			8.237	0.005
			7.543	0.010
			6.488	0.025
			5.706	0.046
			5.626	0.051
			4.545	0.100
			3.807	0.147
5	5	4	9.323	0.001
			8.371	0.005
			7.766	0.010
			6.671	0.025
			5.643	0.050
			5.023	0.075
			4.520	0.101
			3.883	0.148
5	5	5	9.680	0.001
			8.720	0.005
			7.980	0.011
			6.740	0.025
			5.780	0.049
			5.660	0.051
			4.560	0.100
			3.860	0.150

Valores seleccionados de la distribución de la estadística de rangos con signo de Wilcoxon. Para α = 0.005, 0.01, 0.025, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, la tabla de valores $T_\alpha(n)$ tales que, bajo la hipótesis nula:

$$P(T > T_\alpha(n)) = \alpha,$$

donde T es la estadística de Wilcoxon.

Para valores en la cola izquierda de la distribución use la ecuación:

$$T_{1-\alpha}(n) = \frac{n(n+1)}{2} - T_\alpha(n)$$

n \ α	0.005	0.010	0.025	0.050	0.100	0.200	0.300	0.400	0.500
3	6	6	6	6	6	5	4	3	2
4	10	10	10	10	9	7	7	6	5
5	15	15	15	14	12	11	10	9	7.5
6	21	21	20	18	17	15	13	12	10.5
7	28	27	25	24	22	19	17	16	14
8	35	34	32	30	27	24	22	20	18
9	43	41	39	36	34	30	27	25	22.5
10	51	49	46	44	40	36	33	30	27.5
11	60	58	55	52	48	43	39	36	33
12	70	68	64	60	56	50	46	42	39
13	81	78	73	69	64	58	53	49	45.5
14	92	89	83	79	73	66	61	57	52.5
15	104	100	94	89	83	75	69	65	60
16	116	112	106	100	93	85	78	73	68
17	129	125	118	111	104	95	88	82	76.5
18	143	138	130	123	115	105	98	91	85.5
19	157	152	143	136	127	116	108	101	95
20	172	166	157	149	140	128	119	112	105

Fuente: Said Infante Gil y Guillermo P. Zárate. *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores seleccionados de la distribución de la estadística de Friedman.
La tabla presenta, para *t* tratamientos y *b* bloques, la probabilidad (*p*) de un valor mayor o igual que *T* (la estadística de Friedman) si la hipótesis nula es cierta.

t	*b*	*T*	*p*
3	2	4.000	0.167
		3.000	0.500
3	3	6.000	0.028
		4.667	0.194
		2.667	0.361
3	4	8.000	0.005
		6.500	0.042
		6.000	0.069
		4.500	0.125
		3.500	0.273
3	5	10.000	0.001
		8.400	0.008
		7.600	0.024
		6.400	0.039
		5.200	0.093
		4.800	0.124
		3.600	0.182
		2.800	0.367
3	6	10.330	0.002
		9.000	0.008
		8.330	0.012
		7.000	0.029
		6.330	0.052
		5.330	0.072
		4.330	0.142
3	7	11.143	0.001
		10.286	0.004
		8.857	0.008
		8.000	0.016
		6.000	0.051
		4.571	0.112
3	8	12.000	0.001
		9.750	0.005
		9.000	0.010
		6.250	0.047
		5.250	0.079
		4.750	0.120
3	9	11.556	0.001
		9.556	0.006
		8.667	0.010
		6.222	0.048
		6.000	0.057
		5.556	0.069
		4.667	0.107
4	2	6.000	0.042
		5.400	0.167
		4.800	0.208
4	3	9.000	0.002
		8.200	0.017
		7.000	0.054
		6.600	0.075
		5.800	0.148
4	4	11.100	0.001
		9.900	0.006
		9.300	0.012
		7.500	0.052
		6.300	0.094
		6.000	0.105
		5.700	0.141

Fuente: Said Infante Gil y Guillermo P. Zárate. *Métodos estadísticos. Un enfoque interdisciplinario*, 2a. ed., 3a. reimpr., Trillas, México, 1996.

Valores de Z en términos de r coeficiente de correlación $Z = \frac{1}{2}\left(L_n\left(\frac{1+r}{1+r}\right)\right)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0 0599	0.0699	0.0798	0.0898
0.1	0.0997	0.1096	0.1194	0.1293	0.1391	0.1489	0.1586	0.1684	0.1781	0.1877
0.2	0.1974	0.2070	0.2165	0.2260	0.2355	0.2449	0.2543	0.2636	0.2729	0.2821
0.3	0.2913	0.3004	0.3095	0.3185	0.3275	0.3364	0.3452	0.3540	0.3627	0.3714
0.4	0.3799	0.3885	0.3969	0.4053	0.4136	0.4219	0.4301	0.4382	0.4462	0.4542
0.5	0.4621	0.4699	0.4777	0.4854	0.4930	0.5005	0.5080	0.5154	0.5227	0.5299
0.6	0.5370	0.5441	0.5511	0.5581	0.5649	0.5717	0.5784	0.5850	0.5915	0.5980
0.7	0.6044	0.6107	0.6169	0.6231	0.6291	0.6351	0.6411	0.6469	0.6527	0.6584
0.8	0.6640	0.6696	0.6751	0.6805	0.6858	0.6911	0.6963	0.7014	0.7064	0.7114
0.9	0.7163	0.7211	0.7259	0.7306	0.7352	0.7398	0.7443	0.7487	0.7531	0.7574
1.0	0.7616	0.7658	0.7699	0.7739	0.7779	0.7818	0.7857	0.7895	0.7932	0.7969
1.1	0.8005	0.8041	0.8076	0.8110	0.8144	0.8178	0.8210	0.8243	0.8275	0.8306
1.2	0.8337	0.8367	0.8397	0.8426	0.8455	0.8483	0.8511	0.8538	0.8565	0.8591
1.3	0.8617	0.8643	0.8668	0.8692	0.87]7	0.8741	0.8764	0.8787	0.8810	0.8832
1.4	0.8854	0.8875	0.8896	0.8917	0.8937	0.8957	0.8977	0.8996	0.9015	0.9033
1.5	0.9051	0.9069	0.9087	0.9104	0.9121	0.9138	0.9154	0.9170	0.9186	0.9201
1.6	0.9217	0.9232	0.9246	0.9261	0.9275	0.9289	0.9302	0.9316	0.9329	0.9341
1.7	0.9354	().9366	0.9379	0.9391	0.9402	0.9414	0.9425	0.9436	0.9447	0.9458
1.8	0.9468	0.9478	0.9488	0.9498	0.9508	0.9517	0.9527	0.9536	0.9545	0.9554
1.9	0.9562	0.9571	0.9579	0.9587	0.9595	0.9603	0.9611	0.9618	0.9626	0.9633
2.0	0.9640	0.9647	0.9654	0.9661	0.9667	0.9674	0.9680	0.9687	0.9693	0.9699
2.1	0.9705	0.9710	0.9716	0.9721	0.9727	0.9732	0.9737	0.9743	0.9748	0.9753
2.2	0.9757	0.9762	0.9767	0.9771	0.9776	0.9780	0.9785	0.9789	0.9793	0.9797
2.3	0.9801	0.9805	0.9809	0.9812	0.98 i 6	0.9820	0.9823	0.9827	0.9830	0.9833
2.4	0.9837	0.9840	0.9843	0.9846	0.9849	0.9852	0.9855	0 9858	0.9861	0.9863
2.5	0.9866	0.9869	0.9871	0.9874	0.9876	0.9879	0.9881	0.9884	0.9886	0.9888
2.6	0.9890	0.9892	0.9895	0 9897	0.9899	0.9901	0.9903	0.9905	0.9906	0.9908
2.7	0.9910	0.9912	0.9914	0.9915	0.9917	0.9919	0.9920	0.9922	0.9923	0.9925
2.8	0.9926	0.9928	0.9929	0.9931	0.9932	0.9933	0.9935	0.9936	0.9937	0.9938
2.9	0.9940	0.9941	0.9942	0.9943	0.9944	0.9945	0.9946	0.9947	0.9949	0.9950

Tabla generada por J. Carlos Medina (MACSTAT)

Valores de *r* en términos de *Z*.

$$r = (e^{2Z} - 1)/(e^{2Z} + 1)$$

r	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0100	0.0200	0.0300	0 0400	0.0500	0.0601	0.0701	0.0802	0.0902
0.1	0.1003	0.1104	0.1206	0.1307	0.1409	0.1511	0.1614	0.1717	0.1820	0,1923
0.2	0.2027	0.2132	0.2237	0.2342	0.2448	0.2554	0.2661	0.2769	0.2877	0.2986
0.3	0.3095	0.3205	0.3316	0.3428	0.3541	0.3654	0.3769	0.3884	0.4001	0.4118
0.4	0,4236	0.4356	0.4477	0.4599	0.4722	0.4847	0.4973	0.5101	0.5230	0.5361
0.5	0.5493	0.5627	0.5763	0.5901	0.6042	0.6184	0.6328	0.6475	0.6625	0.6777
0.6	0.6931	0.7089	0.7250	0.7414	0.7582	0.7753	0.7928	0.8107	0.8291	0.8480
0.7	0.8673	0.8872	0.9076	0.9287	0.9505	0.9730	0.9962	1.0203	1.0454	1.0714
0.8	1.0986	1.1270	1.1568	1.1881	1.2212	1.2562	1.2933	1.3331	1.3758	1.4219

r	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.90	1.4722	1.4775	1.4828	1.4882	1.4937	1.4992	1.5047	1.5103	1.5160	1.5217
0.91	1 5275	1.5334	1.5393	1.5453	1.5513	1.5574	1.5636	1.5698	1.5762	1.5826
0.92	1.5890	1.5956	1.6022	1.6089	1.6157	1.6226	1.6296	1.6366	1.6438	1.6510
0.93	1.6584	1.6658	1.6734	1.6811	1.6888	1.6967	1.7047	1.7129	1.7211	1.7295
0.94	1.7380	1.7467	1.7555	1.7645	1.7736	1.7828	1.7923	1.8019	1.8117	1.8216
0.95	1.8318	1.8421	1.8527	1.8635	1.8745	].8857	].8972	1.9090	1.9210	1.9333
0.96	1.9459	1.9588	1.9721	1.9857	1.9996	2.0139	2.0287	2.0439	2.0595	2.0756
0.97	2.0923	2.1095	2.1273	2. 1457	2.1649	2.1847	2.2054	2.2269	2.2494	2.2729
0.98	2.2976	2.3235	2.3507	2.3796	2.4101	2.4427	2.4774	2.5147	2.5550	2.5987
0.99	2.6467	2.6996	2.7587	2.8257	2.9031	2.9945	3.1063	3.2504	3.4534	3.8002

Tabla generada por J. Carlos Medina (MACSTAT)

Ángulos correspondientes a los porcentajes ANG = ARCO Seno $\sqrt{\text{Porcentaje}}$

%	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.5730	0.8103	0.9924	1.1460	1.2813	1.4036	1.5161	1.6208	1.7191
0.1	1.8122	1.9006	1.9852	2.0663	2.1443	2.2196	2.2924	2.3630	2.4316	2.4983
0.2	2.5632	2.6265	2.6884	2.7489	2.8080	2.8660	2.9228	2.9785	3.0332	3.0870
0.3	3.1398	3.1917	3.2429	3.2932	3.3428	3.3916	3.4398	3.4873	3.5342	3.5805
0.4	3.6261	3.6712	3.7158	3.7598	3.8034	3.8464	3.8890	3.9311	3.9728	4.0140
0.5	4.0548	4.0952	4.1352	4.1749	4.2142	4.2531	4.2916	4.3299	4.3677	4.4053
0.6	4.4426	4.4795	4.5161	4.5525	4.5886	4.6244	4.6599	4.6951	4.7301	4.7648
0.7	4.7993	4.8336	4.8676	4.9013	4.9349	4.9682	5.0013	5.0342	5.0668	5.0993
0.8	5.1315	5.1636	5.1955	5.2271	5.2586	5.2899	5.3210	5.3520	5.3827	5.4133
0.9	5.4437	5.4740	5.5041	5.5340	5.5638	5.5934	5.6228	5.6521	5.6813	5.7103
1	5.7392	6.0203	6.2891	6.5470	6.7952	7.0349	7.2669	7.4918	7.7103	7.9229
2	8.1301	8.3323	8.5298	8.7230	8.9121	9.0974	9.2792	9.4575	9.6327	9.8049
3	9.9742	10.1408	10.3048	10.4664	10.6256	10.7826	10.9374	11.0902	11.2410	11.3899
4	11.5370	11 6823	11.8259	11.9679	12.1084	12.2473	12.3848	12.5208	12.6555	12.7889
5	12.9210	13.0518	13.1814	13.3099	13.4372	13.5634	13.6885	13.8126	13.9356	14.0577
6	14.1788	14.2990	14.4182	14.5366	14.6541	14.7707	14.8865	15.0015	15.1156	15.2291
7	15.3417	15.4536	15.5648	15.6753	15.7851	15.8942	16.0026	16.1104	16.2175	16.3240
8	16.4299	16.5352	16.6399	16.7441	16.8476	16.9506	17.0531	17.1550	17.2564	17.3572
9	17.4576	17.5575	17.6568	17.7557	17.8541	17.9520	18.0495	18.1465	18.2431	18.3392
10	18.4349	18.5302	18.6251	18.7195	18.8136	18.9072	19.0005	19.0934	19.1859	19.2780
11	19.3697	19.4611	19.5521	19.6428	19.7331	19.8231	19.9127	20.0020	20.0909	20.1796
12	20.2679	20.3559	20.4436	20.5310	20.6180	20.7048	20.7913	20.8775	20.9634	21.0490
13	21.1343	21.2193	21.3041	21.3886	21.4728	21.5568	21.6405	21.7240	21.8071	21.8901
14	21.9728	22.0552	22.1374	22.2193	22.3011	22.3825	22.4638	22.5448	22.6256	22.7062
15	22.7865	22.8666	22.9465	23.0262	23.1057	23.1849	23.2640	23.3428	23.4215	23.4999
16	23.5782	23.6562	23.7341	23.8117	23.8892	23.9665	24.0436	24.1205	24.1972	24.2737
17	24.3501	24.4262	24.5022	24.5781	24.6537	24.7292	24.8045	24.8797	24.9546	25.0294
18	25.1041	25.1786	25.2529	25.3271	25.4011	25.4749	25.5486	25.6222	25.6956	25.7688
19	25.8419	25.9149	25.9877	26.0604	26.1329	26.2052	26.2775	26.3496	26.4215	26.4934
20	26.5651	26.6366	26.7080	26.7793	26.8505	26.9215	26.9924	27.0632	27.1338	27.2043
21	27.2747	27.3450	27.4152	27.4852	27.5551	27.6249	27.6946	27.7641	27.8336	27.9029
22	27.9721	28.0412	28.1102	28.1791	28.2478	28.3165	28.3850	28.4535	28.5218	28.5901
23	28.6582	28.7262	28.7941	28.8619	28.9297	28.9973	29.0648	29.1322	29.1995	29.2667
24	29.3339	29.4009	29.4678	29.5347	29.6014	29.6681	29.7346	29.8011	29.8675	29.9338

Tabla generada por J. Carlos Medina (MACSTAT)

Ángulos correspondientes a los porcentajes *(Continuación)*

25	30.0000	30.0661	30.1321	30.1981	30.2639	30.3297	30.3954	30.4610	30.5265	30.5919
26	30.6573	30.7226	30.7878	30.8529	30.9179	30.9829	31.0477	31.1125	31.1772	31.2419
27	31.3064	31.3709	31.4354	31.4997	31.5640	31.6282	31.6923	31.7563	31.8203	31.8842
28	31.9481	32.0118	32.0755	32.1392	32.2027	32.2662	32.3296	32.3930	32.4563	32.5195
29	32.5827	32.6458	32.7088	32.7718	32.8347	32.8976	32.9604	33.0231	33.0858	33.1484
30	33.2109	33.2734	33.3358	33.3982	33.4605	33.5228	33.5849	33.6471	33.7092	33.7712
31	33.8332	33.8951	33.9569	34.0187	34.0805	34.1422	34.2038	34.2654	34.3270	34.3885
32	34.4499	34.5113	34.5726	34 6339	34.6952	34.7563	34.8175	34.8786	34.9396	35.0006
33	35.0616	35.1225	35.1833	35.2441	35.3049	35.3656	35.4263	35.4869	35.5475	35.6080
34	35.6685	35.7290	35.7894	35.8498	35.9101	35.9704	36.0306	36.0908	36.1510	36.21 11
35	36.2712	36.3312	36.3912	36.4512	36.5111	36.5710	36.6309	36.6907	36.7505	36.8102
36	36.8699	36.9296	36.9892	37.0488	37.1083	37.1679	37.2274	37.2868	37.3462	37.4056
37	37.4650	37.5243	37.5836	37.6428	37.7021	37.7612	37.8204	37.8795	37.9386	37.9977
38	38.0567	38.1157	38.1747	38.2337	38.2926	38.3515	38.4103	38.4692	38.5280	38.5867
39	38.6455	38 7042	38.7629	38.8216	38.8802	38.9388	38.9974	39.0560	39.1145	39.1730
40	39.2315	39.2900	39.3484	39.4068	39.4652	39.5236	39.5820	39.6403	39.6986	39.7569
41	39.8151	39.8734	39.9316	39.9898	40.0479	40.1061	40.1642	40.2223	40.2804	40.3385
42	40.3966	40.4546	40.5126	40.5706	40.6286	40.6865	40.7445	40.8024	40.8603	40.9182
43	40.9761	41.0339	41.0918	41.1496	41.2074	41.2652	41.3230	41.3807	41.4385	41.4962
44	41.5539	41.6117	41.6693	41.7270	41.7847	41.8423	41.9000	41.9576	42.0152	42.0728
45	42.1304	42.1880	42.2456	42.3031	42.3607	42.4182	42.4757	42.5332	42.5907	42.6482
46	42.7057	42.7632	42.8207	42.8781	42.9356	42.9930	43.0504	43.1079	43.1653	43.2227
47	43.2801	43.3375	43.3949	43.4523	43.5096	43.5670	43.6244	43.6817	43.7391	43.7964
48	43.8538	43.9111	43.9685	44.0258	44.0831	44.1404	44.1978	44.2551	44.3124	44.3697
49	44.4270	44.4843	44.5416	44.5989	44.6562	44.7135	44.7708	44.8281	44.8854	44.9427
50	45.0000	45.0573	45.1146	45.1719	45.2292	45.2865	45.3438	45.4011	45.4584	45.5157
51	45.5730	45.6303	45.6876	45.7449	45.8022	45.8596	45.9169	45.9742	46.0315	46.0889
52	46.1462	46.2036	46.2609	46.3183	46.3756	46.4330	46.4904	46.5477	46.6051	46.6625
53	46.7199	46.7773	46.8347	46.8921	46.9496	47.0070	47.0644	47.1219	47.1793	47.2368
54	47.2943	47.3518	47.4093	47.4668	47.5243	47.5818	47.6393	47.6969	47.7544	47.8120
55	47.8696	47.9272	47.9848	48.0424	48.1000	48.1577	48.2153	48.2730	48.3307	48.3883
56	48.4461	48.5038	48.5615	48.6193	48.6770	48.7348	48.7926	48.8504	48.9082	48.9661
57	49.0239	49.0818	49.1397	49.1976	49.2555	49.3135	49.3714	49.4294	49.4874	49.5454
58	49.6034	49.6615	49.7196	49.7777	49.8358	49.8939	49.9521	50.0102	50.0684	50.1266
59	50.1849	50.2431	50.3014	50.3597	50.4180	50.4764	50.5348	50.5932	50.6516	50.7100
60	50.7685	50.8270	50.8855	50.9440	51.0026	51.0612	51.1198	51.1784	51.2371	51.2958
61	51.3545	51.4133	51.4720	51.5308	51.5897	51.6485	51.7074	51.7663	51.8253	51.8843

Ángulos correspondientes a los porcentajes *(Continuación)*

62	51.9433	52.0023	52.0614	52.1205	52.1796	52.2388	52.2979	52.3572	52.4164	52.4757
63	52.5350	52.5944	52.6538	52.7132	52.7726	52.8321	52.8917	52.9512	53.0108	53.0704
64	53.1301	53.1898	53.2495	53.3093	53.3691	53.4290	53.4889	53.5488	53.6088	53.6688
65	53.7288	53.7889	53.8490	53.9092	53.9694	54.0296	54.0899	54.1502	54.2106	54.2710
66	54.3315	54.3920	54.4525	54.5131	54.5737	54.6344	54.6951	54.7559	54.8167	54.8775
67	54.9384	54.9994	55.0604	55.1214	55.1825	55.2437	55.3048	55.3661	55.4274	55.4887
68	55.5501	55.6115	55.6730	55.7346	55.7962	55.8578	55.9195	55.9813	56.0431	56.1049
69	56.1668	56.2288	56.2908	56.3529	56.4151	56.4772	56.5395	56.6018	56.6642	56.7266
70	56.7891	56.8516	56.9142	56.9769	57.0396	57.1024	57.1653	57.2282	57.2912	57.3542
71	57.4173	57.4805	57.5437	57.6070	57.6704	57.7338	57.7973	57.8608	57.9245	57.9882
72	58.0519	58.1158	58.1797	58.2437	58 3077	58.3718	58.4360	58.5003	58.5646	58.6291
73	58.6936	58.7581	58.8228	58.8875	58.9523	59.0171	59.0821	59.1471	59.2122	59.2774
74	59.3427	59.4081	59.4735	59.5390	59.6046	59.6703	59.7361	59.8019	59.8679	59.9339
75	60.0000	60.0662	60.1325	60.1989	60.2654	60.3319	60.3986	60.4653	60.5322	60.5991
76	60.6661	60.7333	60.8005	60.8678	60.9352	61.0027	61.0703	61.1381	61.2059	61.2738
77	61.3418	61.4099	61.4782	61.5465	61.6150	61.6835	61.7522	61.8209	61.8898	61.9588
78	62.0279	62.0971	62.1664	62.2359	62.3054	62.3751	62.4449	62.5148	62.5848	62.6550
79	62.7253	62.7957	62.8662	62.9368	63.0076	63.0785	63.1495	63.2207	63.2920	63.3634
80	63.4349	63.5066	63.5785	63.6504	63.7225	63.7948	63.8671	63.9396	64.0123	64.0851
81	64.1581	64.2312	64.3044	64.3778	64.4514	64.5251	64.5989	64.6729	64.7471	64.8214
82	64.8959	64.9706	65.0454	65.1203	65.1955	65.2708	65.3463	65.4219	65.4978	65.5738
83	65.6499	65.7263	65.8028	65.8795	65.9564	66.0335	66.1108	66.1883	66.2659	66.3438
84	66.4218	66.5001	66.5785	66.6572	66.7360	66.8151	66.8943	66.9738	67.0535	67.1334
85	67.2135	67.2938	67.3744	67.4552	67.5362	67.6175	67.6989	67.7807	67.8626	67.9448
86	68.0272	68.1099	68.1929	68.2760	68.3595	68.4432	68.5272	68.6114	68.6959	68.7807
87	68.8657	68.9510	69.0366	69.1225	69.2087	69.2952	69.3820	69.4690	69.5564	69.6441
88	69.7321	69.8204	69.9091	69.9980	70.0873	70.1769	70.2669	70.3572	70.4479	70.5389
89	70.6303	70.7220	70.8141	70.9066	70.9995	71.0928	71.1864	71.2805	71.3749	71.4698
90	71.5651	71.6608	71.7569	71.8535	71.9505	72.0480	72.1459	72.2443	72.3432	72.4425
91	72.5424	72.6428	72.7436	72.8450	72.9469	73.0494	73.1524	73.2559	73.3601	73.4648
92	73.5701	73.6760	73.7825	73.8896	73.9974	74.1058	74.2149	74.3247	74.4352	74.5464
93	74.6583	74.7709	74.8844	74.9985	75.1 135	75.2293	75.3459	75.4634	75.5818	75.7010
94	75.8212	75.9423	76.0644	76.1874	76.3115	76 4366	76.5628	76.6901	76.8186	76.9482
95	77.0790	77.2111	77.3445	77.4792	77.6152	77.7527	77.8916	78.0321	78.1741	78.3177
96	78.4630	78.6101	78.7590	78.9098	79.0626	79.2174	79.3744	79.5336	79.6952	79.8592
97	80.0258	80.1951	80.3673	80.5425	80.7208	80.9026	81.0879	81.2770	81.4702	81.6677
98	81.8699	82.0771	82.2897	82.5082	82.7331	82.9651	83.2048	83.4530	83.7109	83.9797

Ángulos correspondientes a los porcentajes *(Continuación)*

99.0	84.2608	84.2897	84.3187	84.3479	84.3772	84.4066	84.4362	84.4660	84.4959	84.5260
99.1	84.5563	84.5867	84.6173	84.6480	84.6790	84.7101	84.7414	84.7729	84.8045	84.8364
99.2	84.8685	84.9007	84.9332	84.9658	84.9987	85.0318	85.0651	85.0987	85.1324	85.1664
99.3	85.2007	85.2352	85.2699	85.3049	85.3401	85.3756	85.4114	85.4475	85.4839	85.5205
99.4	85.5574	85.5947	85.6323	85.6701	85.7084	85.7469	85.7858	85.8251	85.8648	85.9048
99.5	85.9452	85.9860	86.0272	86.0689	86.1110	86.1536	86.1966	86.2402	86.2842	86.3288
99.6	86.3739	86.4195	86.4658	86.5127	86.5602	86.6084	86.6572	86.7068	86.7571	86.8083
99.7	86.8602	86.9130	86.9668	87.0215	87.0772	87.1340	87.1920	87.2511	87.3116	87.3735
99.8	87.4368	87.5017	87.5684	87.6370	87.7076	87.7804	87.8557	87.9337	88.0148	88.0994
99.9	88.1878	88.2809	88.3792	88.4839	88.5964	88.7187	88.8540	89.0076	89.1897	89.4270
100.0	90.0000	–	–	–	–	–	–	–	–	–

Valores críticos de A (prueba de Sandler)

Para cualquier valor de N-1, la tabla muestra los valores de A, correspondientes a varios niveles de probabilidad R.D. Si $A_{c(\text{calculada})} \leq A_{\alpha(\text{tabla})}$. Entonces A_c es significativa.

	Niveles de significancia para una prueba direccional					
	.05	.025	.01	.005	.0005	
	Niveles de significancia para una prueba bidireccional					
N-1*	.10	.05	.02	.01	.001	N-1
1	0.5125	0.5031	0.50049	0.50012	0.5000012	1
2	0.412	0.369	0.347	0.340	0.334	2
3	0.385	0.324	0.286	0.272	0.254	3
4	0.376	0.304	0.257	0.238	0.211	4
5	0.372	0.293	0.240	0.218	0.184	5
6	0.370	0.286	0.230	0.205	0.167	6
7	0.369	0.281	0.222	0.196	0.155	7
8	0.368	0.278	0.217	0.190	0.146	8
9	0.368	0.276	0.213	0.185	0.139	9
10	0.368	0.274	0.210	0.181	0.134	10
11	0 368	0.273	0.207	0.178	0.130	11
12	0.368	0.271	0.205	0.176	0.126	12
13	0.368	0.270	0.204	0.174	0.124	13
14	0.368	0.270	0.202	0.172	0.121	14
15	0.368	0.269	0.201	0.170	0 119	15
16	0.368	0.268	0.200	0.169	0.117	16
17	0.368	0.268	0.199	0.168	0.116	17
18	0.368	0.267	0.198	0.167	0.114	18
19	0.368	0.267	0.197	0.166	0.113	19
20	0.368	0.266	0.197	0.165	0.112	20
21	0.368	0.266	0.196	0.165	0.111	21
22	0.368	0.266	0.196	0.164	0.110	22
23	0.368	0.266	0.195	0.163	0.109	23
24	0.368	0.265	0.195	0.163	0.108	24
25	0.368	0.265	0.194	0.161	0.108	25
26	0.368	0.265	0.194	0.162	0.107	26
27	0.368	0.265	0.193	0.161	0.107	27
28	0.368	0.265	0.193	0.161	0.106	28
29	0.368	0.264	0.193	0.161	0.106	29
30	0.368	0.264	0.193	0.160	0.105	30
40	0.368	0.263	0.191	0.158	0.102	40
60	0.369	0.262	0.189	0.155	0.099	60
120	0.369	0.261	0.187	0.153	0.095	120
∞	0.370	0.260	0.185	0.151	0.092	∞

* N = el número de pares.

Tabla generada por J. Carlos Medina (MACSTAT)

Tabla de Dunnet 1*a* para *p* = 95%

d.f.	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
24	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

Fuente: Journal of the American Statistical Association, vol. 50, pp. 1096-1121.

Tabla de Dunnet 1*b* para *p* = 99%

d.f.	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

Tabla de Dunnet 2a para p = 95%

d.f.	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
6	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
7	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3.86
8	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
9	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
10	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
11	2.20	2.53	2.76	2.92	3.05	3.15	3.24	3.31	3.38
12	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
13	2.16	2.48	2.69	2.84	2.96	3.06	3.14	3.21	3.27
14	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
15	2.13	2.44	2.64	2.79	2.90	2.99	3.07	3.13	3.19
16	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
17	2.11	2.41	2.61	2.75	2.85	2.94	3.01	3.08	3.13
18	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
19	2.09	2.39	2.58	2.72	2.82	2.90	2.97	3.04	3.09
20	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
24	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
30	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
40	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
60	2.00	2.27	2.43	2.55	2.63	2.70	2.76	2.81	2.85
120	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
∞	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75

Tabla de Dunnet 2*b* para *p* = 99%

d.f.	1	2	3	4	5	6	7	8	9
5	4.03	4.63	5.09	5.44	5.73	5.97	6.18	6.36	6.53
6	3.71	4.22	4.60	4.88	5.11	5.30	5.47	5.61	5.74
7	3.50	3.95	4.28	4.52	4.71	4.87	5.01	5.13	5.24
8	3.36	3.77	4.06	4.27	4.44	4.58	4.70	4.81	4.90
9	3.25	3.63	3.90	4.09	4.24	4.37	4.48	4.57	4.65
10	3.17	3.53	3.78	3.95	4.10	4.21	4.31	4.40	4.47
11	3.11	3.45	3.68	3.85	3.98	4.09	4.18	4.26	4.33
12	3.05	3.39	3.61	3.76	3.89	3.99	4.08	4.15	4.22
13	3.01	3.33	3.54	3.69	3.81	3.91	3.99	4.06	4.13
14	2.98	3.29	3.49	3.64	3.75	3.84	3.92	3.99	4.05
15	2.95	3.25	3.45	3.59	3.70	3.79	3.86	3.93	3.99
16	2.92	3.22	3.41	3.55	3.65	3.74	3.82	3.88	3.93
17	2.90	3.19	3.38	3.51	3.62	3.70	3.77	3.83	3.89
18	2.88	3.17	3.35	3.48	3.58	3.67	3.74	3.80	3.85
19	2.86	3.15	3.33	3.46	3.55	3.64	3.70	3.76	3.81
20	2.85	3.13	3.31	3.43	3.53	3.61	3.67	3.73	3.78
24	2.80	3.07	3.24	3.36	3.45	3.52	3.58	3.64	3.69
30	2.75	3.01	3.17	3.28	3.37	3.44	3.50	3.55	3.59
40	2.70	2.95	3.10	3.21	3.29	3.36	3.41	3.46	3.50
60	2.66	2.90	3.04	3.14	3.22	3.28	3.33	3.38	3.42
120	2.62	2.84	2.98	3.08	3.15	3.21	3.25	3.30	3.33
∞	2.58	2.79	2.92	3.01	3.08	3.14	3.18	3.22	3.25

Porcentiles de la prueba de Bartlett

	Número de poblaciones								
n	2	3	4	5	6	7	8	9	10
				1% puntos					
3	.1411	.1672	*	*	*	*	*	*	*
4	.2843	.3165	.3475	.3729	.3937	.4110	*	*	*
5	.3984	.4304	.4607	.4850	.5046	.5207	.5343	.5458	.5558
6	.4850	.5149	.5430	.5653	.5832	.5978	.6100	.6204	.6293
7	.5512	.5787	.6045	.6248	.6410	.6542	.6652	.6744	.6824
8	.6031	.6282	.6518	.6704	.6851	.6970	.7069	.7153	.7225
9	.6445	.6676	.6892	.7062	.7197	.7305	.7395	.7471	.7536
10	.6783	.6996	.7195	.7352	.7475	.7575	.7657	.7726	.7786
11	.7063	.7260	.7445	.7590	.7703	.7795	.7871	.7935	.7990
12	.7299	.7483	.7654	.7789	.7894	.7980	.8050	.8109	.8160
13	.7501	.7672	.7832	.7958	.8056	.8135	.8201	.8256	.8303
14	.7674	.7835	.7985	.8103	.8195	.8269	.8330	.8382	.8426
15	.7825	.7977	.8118	.8229	.8315	.8385	.8443	.8491	.8532
16	.7958	.8101	.8235	.8339	.8421	.8486	.8541	.8586	.8625
17	.8076	.8211	.8338	.8436	.8514	.8576	.8627	.8670	.8707
18	.8181	.8309	.8429	.8523	.8596	.8655	.8704	.8745	.8780
19	.8275	.8397	.8512	.8601	.8670	.8727	.8773	.8811	.8845
20	.8360	.8476	.8586	.8671	.8737	.8791	.8835	.8871	.8903
21	.8437	.8548	.8653	.8734	.8797	.8848	.8890	.8926	.8956
22	.8507	.8614	.8714	.8791	.8852	.8901	.8941	.8975	.9004
23	.8571	.8673	.8769	.8844	.8902	.8949	.8988	.9020	.9047
24	.8630	.8728	.8820	.8892	.8948	.8993	.9030	.9061	.9087
25	.8684	.8779	.8867	.8936	.8990	.9034	.9069	.9099	.9124
26	.8734	.8825	.8911	.8977	.9029	.9071	.9105	.9134	.9158
27	.8781	.8869	.8951	.9015	.9065	.9105	.9138	.9166	.9190
28	.8824	.8909	.8988	.9050	.9099	.9138	.9169	.9196	.9219
29	.8864	.8946	.9023	.9083	.9130	.9167	.9198	.9224	.9246
30	.8902	.8981	.9056	.9114	.9159	.9195	.9225	.9250	.9271
40	.9175	.9235	.9291	.9335	.9370	.9397	.9420	.9439	.9455
50	.9339	.9387	.9433	.9468	.9496	.9518	.9536	.9551	.9564
60	.9449	.9489	.9527	.9557	.9580	.9599	.9614	.9626	.9637
80	.9586	.9617	.9646	.9668	.9685	.9699	.9711	.9720	.9728
100	.9669	.9693	.9716	.9734	.9748	.9759	.9769	.9776	.9783

Fuente: Journal of the American Statistical Association, vol. 75, 1980, pp. 314-317.

Porcentiles de la prueba de Bartlett *(Continuación)*

	Número de poblaciones								
n	2	3	4	5	6	7	8	9	10
				5% puntos					
3	.3123	.3058	.3173	.3299	*	*	*	*	*
4	.4780	.4699	.4803	.4921	.5028	.5122	.5204	.5277	.5341
5	.5845	.5762	.5850	.5952	.6045	.6126	.6197	.6260	.6315
6	.6563	.6483	.6559	.6646	.6727	.6798	.6860	.6914	.6961
7	.7075	.7000	.7065	.7142	.7213	.7275	.7329	.7376	.7418
8	.7456	.7387	.7444	.7512	.7574	.7629	.7677	.7719	.7757
9	.7751	.7686	.7737	.7798	.7854	.7903	.7946	.7984	.8017
10	.7984	.7924	.7970	.8025	.8076	.8121	.8160	.8194	.8224
11	.8175	.8118	.8160	.8210	.8257	.8298	.8333	.8365	.8392
12	.8332	.8280	.8317	.8364	.8407	.8444	.8477	.8506	.8531
13	.8465	.8415	.8450	.8493	.8533	.8568	.8598	.8625	.8648
14	.8578	.8532	.8564	.8604	.8641	.8673	.8701	.8726	.8748
15	.8676	.8632	.8662	.8699	.8734	.8764	.8790	.8814	.8834
16	.8761	.8719	.8747	.8782	.8815	.8843	.8868	.8890	.8909
17	.8836	.8796	.8823	.8856	.8886	.8913	.8936	.8957	.8975
18	.8902	.8865	.8890	.8921	.8949	.8975	.8997	.9016	.9033
19	.8961	.8926	.8949	.8979	.9006	.9030	.9051	.9069	.9086
20	.9015	.8980	.9003	.9031	.9057	.9080	.9100	.9117	.9132
21	.9063	.9030	.9051	.9078	.9103	.9124	.9143	.9160	.9175
22	.9106	.9075	.9095	.9120	.9144	.9165	.9183	.9199	.9213
23	.9146	.9116	.9135	.9159	.9182	.9202	.9219	.9235	.9248
24	.9182	.9153	.9172	.9195	.9217	.9236	.9253	.9267	.9280
25	.9216	.9187	.9205	.9228	.9249	.9267	.9283	.9297	.9309
26	.9246	.9219	.9236	.9258	.9278	.9296	.9311	.9325	.9336
27	.9275	.9249	.9265	.9286	.9305	.9322	.9337	.9350	.9361
28	.9301	.9276	.9292	.9312	.9330	.9347	.9361	.9374	.9385
29	.9326	.9301	.9316	.9336	.9354	.9370	.9383	.9396	.9406
30	.9348	.9325	.9340	.9358	.9376	.9391	.9404	.9416	.9426
40	.9513	.9495	.9506	.9520	.9533	.9545	.9555	.9564	.9572
50	.9612	.9597	.9606	.9617	.9628	.9637	.9645	.9652	.9658
60	.9677	.9665	.9672	.9681	.9690	.9698	.9705	.9710	.9716
80	.9758	.9749	.9754	.9761	.9768	.9774	.9779	.9783	.9787
100	.9807	.9799	.9804	.9809	.9815	.9819	.9823	.9827	.9830

Porcentiles de la prueba de Bartlett *(Continuación)*

	Número de poblaciones								
n	2	3	4	5	6	7	8	9	10
					10% puntos				
3	.4359	.3991	.3966	.4006	.4061	.4116	*	*	*
4	.5928	.5583	.5551	.5582	.5626	.5673	.5717	.5759	.5797
5	.6842	.6539	.6507	.6530	.6566	.6605	.6642	.6676	.6708
6	.7429	.7163	.7133	.7151	.7182	.7214	.7245	.7274	.7301
7	.7834	.7600	.7572	.7587	.7612	.7640	.7667	.7692	.7716
8	.8130	.7921	.7895	.7908	.7930	.7955	.7978	.8000	.8021
9	.8356	.8168	.8143	.8154	.8174	.8196	.8217	.8236	.8254
10	.8533	.8362	.8339	.8349	.8367	.8386	.8405	.8423	.8439
11	.8676	.8519	.8498	.8507	.8523	.8540	.8557	.8574	.8589
12	.8794	.8649	.8629	.8637	.8652	.8668	.8683	.8698	.8712
13	.8892	.8758	.8740	.8746	.8760	.8775	.8789	.8803	.8816
14	.8976	.8851	.8833	.8840	.8852	.8866	.8879	.8892	.8904
15	.9048	.8931	.8914	.8920	.8932	.8944	.8957	.8969	.8980
16	.9110	.9000	.8985	.8990	.9001	.9013	.9025	.9036	.9046
17	.9165	.9061	.9046	.9051	.9062	.9073	.9084	.9094	.9104
18	.9214	.9115	.9101	.9106	.9115	.9126	.9137	.9146	.9156
19	.9257	.9163	.9150	.9154	.9163	.9174	.9183	.9193	.9201
20	.9295	.9206	.9194	.9198	.9207	.9216	.9226	.9234	.9243
21	.9330	.9245	.9233	.9237	.9245	.9255	.9263	.9272	.9280
22	.9362	.9281	.9269	.9273	.9281	.9289	.9298	.9306	.9313
23	.9390	.9313	.9302	.9305	.9313	.9321	.9329	.9337	.9344
24	.9417	.9342	.9332	.9335	.9342	.9350	.9358	.9365	.9372
25	.9441	.9369	.9359	.9362	.9369	.9377	.9384	.9391	.9398
26	.9463	.9394	.9384	.9387	.9394	.9401	.9408	.9415	.9421
27	.9484	.9417	.9408	.9410	.9417	.9424	.9431	.9437	.9443
28	.9503	.9439	.9429	.9432	.9438	.9445	.9452	.9458	.9464
29	.9520	.9458	.9449	.9452	.9458	.9464	.9471	.9477	.9483
30	.9537	.9477	.9468	.9471	.9476	.9483	.9489	.9495	.9500
40	.9655	.9610	.9603	.9605	.9609	.9614	.9619	.9623	.9627
50	.9725	.9689	.9683	.9685	.9688	.9692	.9696	.9699	.9703
60	.9771	.9741	.9737	.9738	.9741	.9744	.9747	.9750	.9753
80	.9829	.9806	.9803	.9804	.9806	.9808	.9811	.9813	.9815
100	.9864	.9845	.9843	.9843	.9845	.9847	.9849	.9851	.9852

Porcentiles de la prueba de Bartlett *(Continuación)*

	Número de poblaciones								
n	2	3	4	5	6	7	8	9	10
				25% puntos					
3	.6614	.5711	.5411	.5279	.5212	.5176	.5156	.5146	*
4	.7728	.7025	.6779	.6667	.6609	.6577	.6559	.6549	.6544
5	.8299	.7737	.7534	.7440	.7391	.7363	.7347	.7338	.7333
6	.8644	.8177	.8006	.7926	.7884	.7860	.7846	.7838	.7833
7	.8873	.8475	.8327	.8258	.8221	.8200	.8188	.8181	.8177
8	.9036	.8690	.8560	.8499	.8467	.8448	.8437	.8431	.8427
9	.9158	.8852	.8737	.8682	.8653	.8636	.8626	.8620	.8617
10	.9253	.8978	.8875	.8825	.8799	.8784	.8775	.8769	.8766
11	.9329	.9080	.8985	.8940	.8916	.8902	.8894	.8889	.8887
12	.9390	.9163	.9076	.9035	.9013	.9000	.8993	.8988	.8986
13	.9442	.9232	.9152	.9114	.9094	.9082	.9075	.9071	.9068
14	.9485	.9291	.9217	.9181	.9162	.9151	.9145	.9141	.9139
15	.9522	.9342	.9272	.9239	.9221	.9211	.9205	.9201	.9199
16	.9554	.9385	.9320	.9289	.9273	.9263	.9257	.9254	.9252
17	.9582	.9424	.9363	.9333	.9317	.9308	.9303	.9300	.9298
18	.9607	.9457	.9400	.9372	.9357	.9349	.9343	.9340	.9339
19	.9629	.9487	.9433	.9407	.9393	.9384	.9379	.9377	.9375
20	.9649	.9514	.9463	.9438	.9424	.9416	.9412	.9409	.9407
21	.9667	.9539	.9489	.9466	.9453	.9445	.9441	.9438	.9437
22	.9683	.9561	.9513	.9491	.9479	.9471	.9467	.9465	.9463
23	.9697	.9580	.9535	.9514	.9502	.9495	.9491	.9489	.9487
24	.9710	.9599	.9556	.9535	.9523	.9517	.9513	.9511	.9509
25	.9722	.9615	.9574	.9554	.9543	.9537	.9533	.9531	.9530
26	.9734	.9631	.9591	.9572	.9561	.9555	.9552	.9550	.9548
27	.9744	.9645	.9607	.9588	.9578	.9572	.9569	.9567	.9566
28	.9754	.9658	.9621	.9603	.9594	.9588	.9585	.9583	.9581
29	.9762	.9670	.9635	.9617	.9608	.9603	.9599	.9597	.9596
30	.9771	.9682	.9647	.9630	.9621	.9616	.9613	.9611	.9610
40	.9830	.9763	.9737	.9725	.9718	.9714	.9712	.9710	.9710
50	.9865	.9811	.9791	.9781	.9775	.9772	.9770	.9769	.9769
60	.9888	.9843	.9826	.9818	.9813	.9811	.9809	.9808	.9808
80	.9916	.9883	.9870	.9864	.9861	.9859	.9857	.9857	.9856
100	.9933	.9907	.9896	.9891	.9889	.9887	.9886	.9886	.9885

Rango studentizado o (Tukey-Snedecor) para comparaciones múltiples, después del ANOVA.

gl	2	3	4	5	6	7	8	9	10
1	8.93	13.44	16.36	18.49	20.15	21.51	22.64	23.62	24.48
2	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9. 41	9.72
3	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29
4	3.01	3.98	4.59	5.03	5.39	6.68	5.93	6.14	6.33
5	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82
6	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50
7	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28
8	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4 99	5.13
9	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01
10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91
11	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84
12	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78
13	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72
14	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68
15	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64
16	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61
17	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58
18	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13

gl	11	12	13	14	15	16	17	18	19	20
1	25.24	25.92	26.54	27.10	27.62	28.10	28.54	28.96	29.35	29.71
2	10.01	10.26	10.49	10.70	10.89	11.07	11.24	11.39	11.54	11.68
3	7.49	7.67	7.83	7.98	8.12	8.25	8.37	8.48	8.58	8.68
4	6.49	6.65	6.78	6.91	7.02	7.13	7.23	7.33	7.41	7.50
5	5.97	6.10	6.22	6.34	6.44	6.54	6.63	6.71	6.79	6.86
6	5.64	5.76	5.87	5.98	6.07	6.16	6.25	6.32	6.40	6.47
7	5.41	5.53	5.64	5.74	5.83	5.91	5.99	6.06	6.13	6.19
8	5.25	5.36	5.46	5.56	5.64	5.72	5.80	5.87	5.93	6.00
9	5.13	5.23	5.33	5.42	5.51	5.58	5.66	5.72	5.79	5.85
10	5.03	5.13	5.23	5.32	5.40	5.47	5.54	5.61	5.67	5.73
11	4.95	5.05	5.15	5.23	5.31	5.38	5.45	5.51	6.57	5.63
12	4.89	4.99	5.08	5.16	5.24	5.31	5.37	5.44	5.49	5.55
13	4.83	4.93	5.02	5.10	5.18	5.25	5.31	5.37	5.43	5.48
14	4.79	4.88	4.97	5.05	5.12	5.19	5.26	5.32	5.37	5.43
15	4.75	4.84	4.93	5.01	5.08	5.15	5.21	5.27	5.32	5.38
16	4.71	4.81	4.89	4.97	5.04	5.11	5.17	5.23	5.28	5.33
17	4.68	4.77	4.86	4.93	5.01	5.07	5.13	5.19	5.24	5.30
18	4.65	4.75	4.83	4.90	4.98	5.04	5.10	5.16	5.21	5.26
19	4.63	4.72	4.80	4.88	4.95	5.01	5.07	5.13	5.18	5.23
20	4.61	4.70	4.78	4.85	4.92	4.99	5.05	5.10	5.16	5.20
24	4.54	4.63	4.71	4.78	4.85	4.91	4.97	5.02	5.07	5.12
30	4.47	4.56	4.64	4.71	4.77	4.83	4.89	4.94	4.99	5.03
40	4.41	4.49	4.56	4.63	4.69	4.75	4.81	4.86	4.90	4.95
60	4.34	4.42	4.49	4.56	4.62	4.67	4.73	4.78	4.82	4.86
120	4.28	4.35	4.42	4.48	4.54	4.60	4.65	4.69	4.74	4.78
∞	4.21	4.28	4.35	4.41	4.47	4.52	4.57	4.61	4.65	4.69

Fuente: Biometrika Tables for Statisticians, Vols. I y II, editado por E. S. Pearson y H. O. Hartley, Biometrika Trust, Oxford University Press, 1976.

Continuación

gl	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

gl	11	12	13	14	15	16	17	18	19	20
1	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83	59.56
2	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
3	9.72	9.95	10.15	10.35	10.52	10.69	10.84	10.98	11.11	11.24
4	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	5.61	5.71	5.81	5.90	5.98	6.00	6.13	6.20	6.27	6.33
12	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.69
30	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

Continuación

gl	2	3	4	5	6	7	8	9	10
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
2	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69
3	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69
4	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27
5	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24
6	5 24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37
8	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
11	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99
12	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
15	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44
16	4.13	4.79	5.19	3.49	5.72	5.92	6.08	6.22	6.35
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14
20	4.02	4.64	5.02	5.29	5.51	5.69	6.84	5.97	6.09
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
40	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

gl	11	12	13	14	15	16	17	18	19	20
1	233.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3	298.0
2	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50	37.95
3	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55	19.77
4	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24	14.40
5	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	9.30	9.4S	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	7.65	7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57
10	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
11	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7 73
13	6.79	6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55
14	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7 39
15	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7 05
18	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
19	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	6.19	6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82
24	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21
60	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01
120	6.37	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

Valores críticos del coeficiente de Spearman

n-4		n-5		n-6		n-7		n-8		n-9		n – 10	
S_r	Q	S_r	Q	S_r	Q	S_r	Q	S_r	Q	S_r	Q	S_r	Q
12	0.458	22	0.475	50	0.210	74	0.249	108	0.250	156	0.218	208	0.235
14	0.375	24	0.392	52	0.178	78	.198	114	.195	164	.168	218	.184
16	0.208	26	0.342	54	0.149	82	.151	120	.150	172	.125	228	.139
18	0.167	28	0.258	56	0.121	86	.118	126	.108	180	.080	238	.102
20	0.042	30	0.225	58	0.088	90	.083	132	.076	188	.060	248	.072
		32	0.175	60	0.068	94	0.055	138	0.048	196	0.038	258	0.048
		34	0.117	62	0.051	98	.033	144	.029	204	.022	268	.030
		36	0.067	64	0.029	102	.017	150	.014	212	.011	278	.017
		38	0.042	66	0.017	106	.0062	156	.0054	220	.0041	288	.0087
		40	0.0083	68	0.0083	110	.0014	162	.0011	228	.0010	298	.0036
				70	0.0014							303	0.0011
20		40		70		112		168		240		330	

S_t	Valores de n				S_t	Valores de n		
	4	5	8	9		6	7	10
0	0.625	0.592	0.548	0.510	1	0 500	0.500	0.500
2	0.375	0.408	0.452	0.460	3	0.360	0.386	0.431
4	0.167	0.242	0.360	0.381	5	0.235	0.281	0.364
6	0.042	0.117	0.274	0.306	7	0.136	0.191	0.300
8		0.042	0.199	0.238	9	0.068	0.119	0.242
10		0.0083	0.138	0.179	11	0.028	0.068	0.190
12			0.089	0.130	13	0.0083	0.035	0.146
14			0.054	0.090	15	0.0014	0.015	0.108
16			0.031	0.060	17		0.0054	0.078
18			0.016	0.038	19		0.0014	0.054
20			0.0071	0.022	21		0.0002	0.036
22			0.0028	0.012	23			0.023
24			0.0009	0.0063	25			0.014
26			0.0002	0.0029	27			0.0083
28				0.0012	29			0.0046
30				0.0004	31			0.0023
					33			0.0011
					35			0.0005

Fuente: Biometrika Tables for Statisticians, Vols. I y II, editado por E. S. Pearson y H. O. Hartley, Biometrika Trust, Oxford University Press, 1976.

Valores críticos coeficiente de concordancia ω

(a) Caso $n = 3$

$m = 3$		$m = 5$		$m = 6$		$m = 7$		$m = 8$		$m = 9$		$m = 10$	
S_w	Q	S_w	Q	S_w	Q	S_w	Q	S_w	Q	S_w	Q	S_w	Q
6	0.528	14	0.367	18	0.252	24	0.237	26	0.236	32	0.187	32	0.222
8	0.361	18	0.182	24	0.184	26	0.192	32	0.149	38	0.154	42	0.135
14	0.194	24	0.124	26	0.142	32	0.112	38	0.120	42	0.107	50	0.092
18	0.028	26	0.093	32	0.072	38	0.085	42	0.079	50	0.069	56	0.066
		32	0.039	38	0.052	42	0.051	50	0.047	56	0.048	62	0.046
		38	0.024	42	0.029	50	0.027	56	0.030	62	0.031	74	0.026
		42	.0085	50	0.012	56	0.016	72	0.0099	78	0.010	86	0.012
		50	.0008	54	0.0081	62	0.0084	78	0.0048	86	0.0060	96	0.0075
				56	0.0055	72	0.0036	86	0.0024	98	0.0029	104	0.0034
				62	0.0017	78	0.0012	98	0.0009	104	0.0013	122	0.0013
				72	0.0001	96	0.0003			114	0.0007	126	0.0008

$m = 4$

S_w	Q
8	0.431
14	.273
18	.125
24	.069
26	.042
32	.0046

(b) Caso $n = 4$

$m = 3$		$m = 4$		$m = 5$		$m = 6$	
S_w	Q	S_w	Q	S_w	Q	S_w	Q
19	0.342	32	0.200	41	0.210	46	0.218
21	0.300	36	0.158	43	0.162	52	0.163
25	0.207	40	0.105	51	0.107	62	0.108
27	0.175	46	0.068	57	0.075	68	0.073
29	0.148	50	0.052	61	0.055	74	0.056
33	0.075	54	0.033	67	0.034	80	0.37
35	0.054	62	0.012	81	0.012	100	0.010
37	0.033	66	0.0062	85	0.0067	108	0.0061
41	0.017	70	0.0027	93	0.0023	118	0.0028
45	0.0017	74	0.0009	101	0.014	128	0.0009
				105	0.0006		

(c) Caso $n = 3$

$m = 3$	
S_w	Q
46	.0213
50	0.163
56	0.096
60	0.063
62	0.056
66	0.038
74	0.015
78	0.0053
82	0.0028
86	0.0009

Valores críticos de *Q* (Tukey-Snedecor) para comparaciones múltiples

gl grados de libertad	Números de grupos							
	2	3	4	5	6	7	8	9
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36
2	6.08	8.28	9.80	10.89	11.73	12.43	13.03	13.54
3	4.50	5.91	6.83	7.51	8.04	8.47	8.85	9.18
4	3.93	5.04	5.76	6.29	6.70	7.06	7.35	7.60
5	3.64	4.60	5.22	5.67	5.93	6.38	6.58	6.80
6	3.46	4.34	4.90	5.31	5.63	5.89	6.12	6.32
7	3.34	4.16	4.68	5.06	5.35	5.59	5.82	5.99
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.60
10	3.15	3.88	4.33	4.66	4.91	5.12	5.30	5.46
11	3.11	3.82	4.26	4.58	4.82	5.03	5.20	5.35
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27
13	3.06	3.73	4.15	4.46	4.69	4.88	5.05	5.19
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08
16	3.00	3.65	4.05	4.34	4.56	4.74	4.90	5.03
17	2.98	3.62	4.02	4.31	4.52	4.70	4.86	4.99
18	2.97	3.61	4.00	4.28	4.49	4.67	4.83	4.96
19	2.96	3.59	3.98	4.26	4.47	4.64	4.79	4.92
20	2.95	3.58	3.96	4.24	4.45	4.62	4.77	4.90
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81
30	2.89	3.48	3.84	4.11	4.30	4.46	4.60	4.72
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55
120	2.80	3.36	3.69	3.92	4.10	4.24	4.36	4.47
∞	2.77	3.32	3.63	3.86	4.03	4.17	4.29	4.39

Fuente: Biometrika Tables for Statisticians, vols. I y II, editado por E.S. Pearson y H. O. Hartley, Biometrika Trust, Oxford University Press, 1976.

Valores críticos de la prueba de Fmáx = $\frac{s^2_{máx}}{s^2_{min}}$ (prueba de Hartley de homogeneidad de varianzas)

df para $\hat{\sigma}^2_j$	α	k = número de varianzas										
		2	3	4	5	6	7	8	9	10	11	12
4	.05	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.4	44.6	48.0	51.4
	.01	23.2	37.	49.	59.	69.	79.	89.	97.	106.	113.	120.
5	.05	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
	.01	14.9	22.	28.	33.	38.	42.	46.	50.	54.	57.	60.
6	.05	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
	.01	11.1	15.5	19.1	22.	25.	27.	30.	32.	34.	36.	37.
7	.05	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
	.01	8.89	12.1	14.5	16.5	18.4	20.	22.	23.	24.	26.	27.
8	.05	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
	.01	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21.
9	.05	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
	.01	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	.05	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
	.01	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	.05	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
	.01	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	.05	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
	.01	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	.05	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
	.01	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	.05	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
	.01	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	.05	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
	.01	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
∞	.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Fuente: Biometrika Tables for Statisticians, vols. I y II, editado por E.S. Pearson y H. O. Hartley, Biometrika Trust, Oxford University Press, 1976.

Valores críticos para la prueba de homogeneidad de varianzas de Cochran
C = (la s^2 más grande $\sum S_i^s$

gl para s_j^2	$1-\alpha$	*k* = número de varianzas										
		2	3	4	5	6	7	8	9	10	15	20
1	.95	.9985	.9669	.9065	.8412	.7808	.7271	.6798	.6385	.6020	.4709	.3894
	.99	.9999	.9933	.9676	.9279	.8828	.8376	.7945	.7544	.7175	.5747	.4799
2	.95	.9750	.8709	.7679	.6838	.6161	.5612	.5157	.4775	.4450	.3346	.2705
	.99	.9950	.9423	.8643	.7885	.7218	.6644	.6152	.5727	.5358	.4069	.3297
3	.95	.9392	.7977	.6841	.5981	.5321	.4800	.4377	.4027	.3733	.2758	.2205
	.99	.9794	.8831	.7814	.6957	.6258	.5685	.5209	.4810	.4469	.3317	.2654
4	.95	9057	.7457	.6287	.5441	.4803	.4307	.3910	.3584	.3311	.2419	.1921
	.99	.9586	.8335	.7212	.6329	.5635	.5080	.4627	.4251	.3934	.2882	.2288
5	.95	.8772	.7071	.4447	.5895	.5065	.3974	.3595	.3286	.3029	.2195	.1735
	.99	.9373	.7933	.6761	.5875	.5195	.4559	.4226	.3870	.3572	.2593	.2048
6	.95	.8534	.6771	.5598	.4783	.4184	.3726	.3362	.3067	.2823	.2034	.1602
	.99	.9172	.7606	.6410	.5531	.4866	.4347	.3932	.3592	.3308	.2386	.1877
7	.95	.8332	.6530	.5365	.4564	.3980	.3535	.3185	.2901	.2666	.1911	.1501
	.99	.8988	.7335	.6129	.5259	.4608	.4105	.3704	.3378	.3106	.2228	.1748
8	.95	.8159	.6333	.5175	.4387	.3817	.3384	.3043	.2768	.2541	.1815	.1422
	.99	.8823	.7107	.5897	.5037	.4401	.3911	.3522	.3207	.2945	.2104	.1646
9	.95	.8010	.6167	.5017	.4241	.3682	.3259	.2926	.2659	.2439	.1736	.1357
	.99	.8674	.6912	.5702	.4854	.4229	.3751	.3373	.3067	.2813	.2002	.1567
16	.95	.7341	.5466	.4366	.3645	.3135	.2756	.2462	.2226	.2032	.1429	.1108
	.99	.7949	.6059	.4884	.4094	.3529	.3105	.2779	.2514	.2297	.1612	.1248
36	.95	.6602	.4748	.3720	.3066	.2612	.2278	.2022	.1820	.1655	.1144	.0879
	.99	.7067	.5153	.4057	.3351	.2858	.2494	.2214	.1992	.1811	.1251	.0960
144	.95	.5813	.4031	.3093	.2513	.2119	.1833	.1616	.1446	.1308	.0889	.0675
	.99	.6062	.4230	.3251	.2644	.2229	.1929	.1700	.1521	.1376	.0934	.0709

Fuente: Journal of the American Statistical Association, vol. 75, 1980.

Valores de $|n|$ para la prueba de McNemar.

	Nivel de significancia a una cola					
	0.100	0.050	0.025	0.010	0.005	0.0005
	Nivel de significancia a dos colas					
n	0.200	0.100	0.050	0.020	0.010	0.001
4	0					
5	0	0				
6	0	0	0			
7	1	0	0	0		
8	1	1	0	0	0	
9	2	1	1	0	0	
10	2	1	1	0	0	
11	2	2	1	1	0	0
12	3	2	2	1	1	0
13	3	3	2	1	1	0
14	4	3	2	2	1	0
15	4	3	3	2	2	1
16	4	4	3	2	2	1
17	5	4	4	3	2	1
18	5	5	4	3	3	1
19	6	5	4	4	3	2
20	6	5	5	4	3	2
21	7	6	5	4	4	2
22	7	6	5	5	4	3
23	7	7	6	5	4	3
24	8	7	6	5	5	3
25	8	7	7	6	5	4
26	9	8	7	6	6	4
27	9	8	7	7	6	4
28	10	9	8	7	6	5
29	10	9	8	7	7	5
30	10	10	9	8	7	5
31	11	10	9	8	7	6
32	11	10	9	8	8	6
33	12	11	10	9	8	6
34	12	11	10	9	9	7
35	13	12	11	10	9	7
36	13	12	11	10	9	7
37	14	13	12	10	10	8
38	14	13	12	11	10	8
39	15	13	12	11	11	8
40	15	14	13	12	11	9
41	15	14	13	12	11	9
42	16	15	14	13	12	10
43	16	15	14	13	12	10
45	17	16	15	14	13	11
46	17	16	15	14	13	11
47	18	17	16	15	14	11
48	19	17	16	15	14	12
49	19	18	17	15	15	12
50	19	18	17	16	15	13

Fuente: Ignacio Méndez Ramírez *et al.*, *El protocolo de la investigación. Lineamientos para su elaboración y análisis*, 2a ed., 5a. reimpr., Trillas, México, 1997.

Ábacos para calcular el tamaño de una muestra

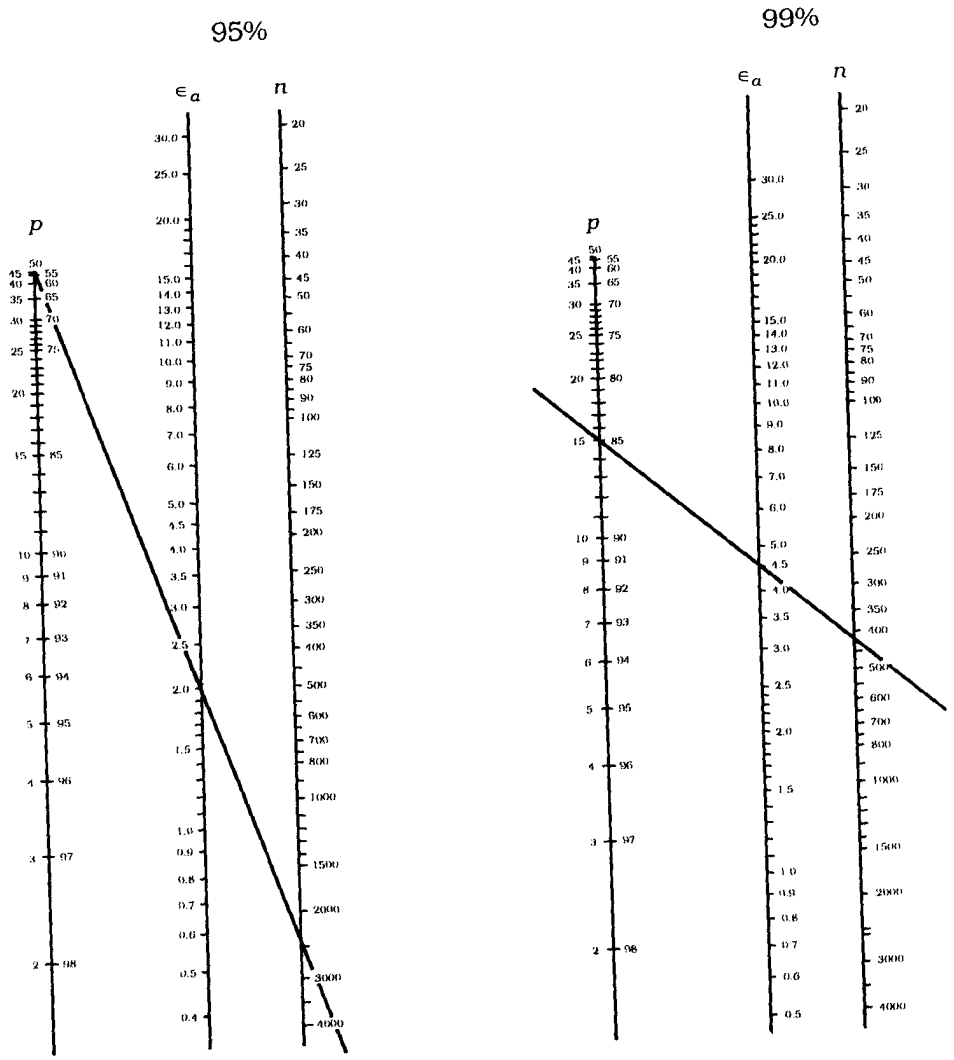


Tabla generada por Adip Sabag.

Anexo 2

Soluciones a los problemas

Capítulo 2

2.1 *a)* $Y_1 + Y_2 + Y_3 + Y_4 + Y_5$
b) $X_3^2 + X_4^2 + X_5^2 + X_6^2$
c) $X_1Y_1 + X_2Y_2 + X_3Y_3 + X_4Y_4 + X_5Y_5 + X_6Y_6 + X_7Y_7$
d) $(2X_1 + 5) + (2X_2 + 5) + (2X_3 + 5)$
e) $X_2^2\, l_2 + X_3^2\, l_3 + X_4^2\, l_4$

2.2 *a)* $\sum_{i=1}^{17} Y_i^2$ *b)* $\sum_{i=1}^{100} 2Z_i$

c) $\sum_{i=2}^{5} X_iY_i^2$ *d)* $\sum_{i=1}^{3} (X_i^2 + 5)^3$

2.3 *a)* 15 *b)* 3 *c)* 35

2.4 *a)* 21 *b)* 20 *c)* 21 *d)* 6

2.5 409

2.6 *a)* $\overline{X} = 55.6$ $Me = 56$ Mo = No hay.
b) $S = 5.4129$ $S^2 = 29.3$
c) $CV = 9.73\ \%$

2.7 Método 1.

$\overline{X}_A = 8.31250$	$\overline{X}_B = 8.31250$
$SA = 1.962824$	$S_B = 1.06695$
$S_A^2 = 3.85268$	$S_B^2 = 1.13839$
$CV_A = 23.613\ \%$	$CV_B = 12.836\ \%$

Método 2.

$\overline{X}_A = 6.43750$	$\overline{X}_B = 6.43750$
$SA = 3.\ 17847$	$S_B = 1.01550$
$S_A^2 = 10.10268$	$S_B^2 = 1.03125$
$CV_A = 49.374\ \%$	$CV_B = 15.775\ \%$

Método 3.

$\overline{X}_A = 4.000$	$\overline{X}_B = 4.000$
$SA = 2.17124$	$S_B = 0.84515$
$S_A^2 = 4.71429$	$S_B^2 = 0.71429$
$CV_A = 54.281\ \%$	$CV_B = 21.129\ \%$

Método 4.

$\overline{X}_A = 2.50$ $\qquad$ $\overline{X}_B = 2.500$

SA = 2.01778 $\qquad$ $S_B = 0.96362$

$S_A^2 = 4.07143$ $\qquad$ $S_B^2 = 0.92857$

$CV_A = 80.711\ \%$ $\qquad$ $CV_B = 38.545\ \%$

2.8 $\overline{X} = 4.82$ min.
Mediana = 1. 5 min.
Moda = 1.0 min.
con una $S = 9.47$ min. que implica un coeficiente de variación de 196. 5% de "*Error*". Esto se debe a que a una rata le tomó 30 minutos recorrer el laberinto y representa una magnitud muy grande con respecto a la media; por ello la mediana es la mejor medida de tendencia central en este caso, ya que ésta no es sensible a los extremos. Lo recomendable es excluir a este sujeto.

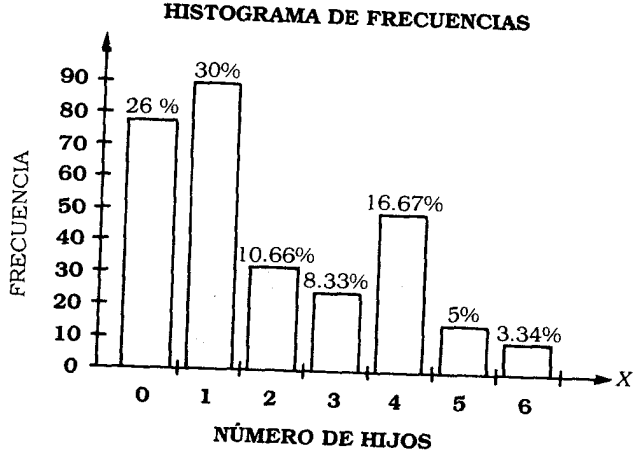
2.9 $\overline{X} = 1.075$ $\qquad$ $CV = 47.0350\ \%$

$S = 0.5056$ $\qquad$ Confiabilidad = 52.965 %

$S^2 = 0.2556$

2.10 Tenemos que, como se trata de una variable aleatoria discreta, sólo podremos presentar su histograma de frecuencias. Esto es:

$n = 300$

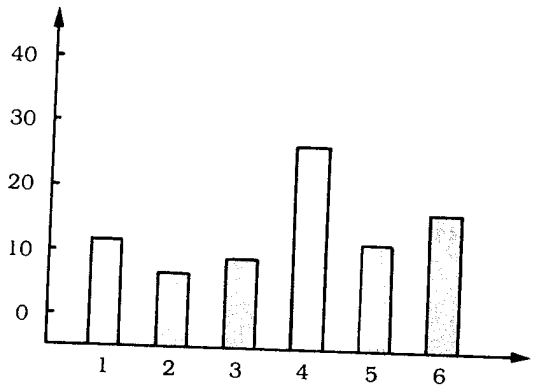


X	F	
0	78	26%
1	90	30%
2	32	10.66%
3	25	8.33%
4	50	16.67%
5	15	5.00%
6	10	3.34%
		100%

2.11 $\overline{X} = 716.614$
$S = 167.0527$
$S^2 = 27\ 906.606$
$CV = 23.31\ \%$
$Me = 710.284$

Mo = 660.6112
Confiabilidad = 76. 69 %

2.12



Análisis de: GRANDE. *MS*1
Datos: 50
Media = 4.9600
Desviación estándar (*S*) = 2.5790
Varianza (S^2) = 6.6514
C.V. = 51.9967 %
Límite inferior de datos 0.0000
Límite superior de datos 9.0000

CLASES	*LRI*	*LRS*	Frec.	Frec *R.*	Frec. *A.*	Frec. *R. A.*
1	[0,0000	1.5000]	7	14.000 %	7	14.000 %
2	[1.5000	3.0000]	5	10,000 %	12	24.000 %
3	[3.0000	4.5000]	6	12.000 %	18	36.000 %
4	[4.5000	6.0000]	15	30.000 %	33	66.000 %
5	[6.0000	7.5000]	7	14.000 %	40	80.000 %
6	[7.5000	9.0000]	10	20.000 %	50	100.000 %

Moda : 15

2.13 $\overline{X}$ = 0.5344
S = 0.0047
S^2 = 0.0000221
CV = 0.87%
Me = 0.5343
Mo = 0.5337
As = + 0.0504
Confiabilidad = 99%

2.14 $\overline{X}$ = 199.3073
S = 32.3468
S^2 = 1 046.3123
CV = 16.23%
Me = 192.7283
Mo = 178.6111
As = + 0.61
Confiabilidad = 83.77%

2.15 $\overline{X}$ = 13 024.41
S = 2 457.39

S^2 = 6 038 765.6
CV = 18.87%
Me = 13 604.04
Mo = 14 681.32
As = – 0.7076
Confiabilidad = 81.13%
n = 239

Lecturas, estadísticas e histograma

#1: 0.4	#2: 0.42	#3: 0.42	#4: 0.43	#5: 0.43		
#6: 0.43	#7: 0.46	#8: 0.46	#9: 0.47	#10: 0.47		
#11: 0.48		#12: 0.48		#13: 0.48	#14: 0.48	#15: 0.48
#16: 0.49		#17: 0.49		#18: 0.49	#19: 0.49	#20: 0.49
#21: 0.49		#22: 0.5	#23: 0.5	#24: 0.5 #25: 0.5		
#26: 0.51		#27: 0.51		#28: 0.52	#29: 0.52	#30: 0.52
#31: 0.53		#32: 0.53		#33: 0.53	#34: 0.55	

2.16 *a*) Para datos no agrupados
Media = 0.4838235
Desviación estándar de la muestra = 0.0360147
Varianza = 1.297059E-03
Sesgo = – 0.5212529
Curtosis = 2.580929
El valor máximo es: 0.55
El valor mínimo es: 0.4
El rango es: 0.15
Mediana = 0.49
CV = 7.44%

b) Para datos agrupados
$\overline{X}$ = 0.4831
S = 0.0345
S^2 = 0.0012
CV = 7.147%
Me = 0.4867
Mo = 0.4875
As = – 0.3109
Confiabilidad = 92.853%
n = 34

LÍMITES	PUNTO MEDIO	FRECUENCIA		%
0.385 / 0.415	.4	1	x	2.941176
0.415 / 0.445	.43	5	xxxx+	14.70588
0.445 / 0.475	.46	4	xxxx	11.76471
0.475 / 0.505	.49	15	xxxx+xxxx+xxxx+	44.11765
0.505 / 0.535	.52	8	xxxx+xxx	23.52941
0.535 / 0.565	.55	1	x	2.941176

Suma de frecuencias: 34

2.17 *a*) Datos no agrupados

# 1 : 2.2	# 2 : 3.4	# 3 : 2.5	# 4 : 3.3	# 5 : 4.7
# 6 : 4.1	# 7 : 1.6	# 8 : 4.3	# 9 : 3.1	# 10 : 3.8
# 11 : 3.5	# 12 : 3.1	# 13 : 3.4	# 14 : 3.7	# 15 : 3.2
# 16 : 4.5	# 17 : 3.3	# 18 : 3.6	# 19 : 4.4	# 20 : 2.6
# 21 : 3.2	# 22 : 3.8	# 23 : 2.9	# 24 : 3.2	# 25 : 3.9
# 26 : 3.7	# 27 : 3.1	# 28 : 3.3	# 29 : 4.1	# 30 : 3
# 31 : 3	# 32 : 4.7	# 33 : 3.9	# 34 : 1.9	# 35 : 4.2
# 36 : 2.6	# 37 : 3.7	# 38 : 3.1	# 39 : 3.4	# 40 : 3.5

Media : 3.4125
Desviación estándar de la muestra = 0.7028103
Varianza = 0.4939424
Sesgo = – 0.3368486
Curtosis = 3.012605
El valor máximo es: 4.7
El valor mínimo es: 1.6
El rango es: 3.1
Mediana = 3.4

b) Datos agrupados

LÍMITES	PUNTO MEDIO	FRECUENCIA		%
1.35 / 1.85	1.6	1	x	2.5
1.85 / 2.35	2.1	2	xx	5
2.35 / 2.85	2.6	3	xxx	7.5
2.85 / 3.35	3.1	13	xxxx+xxxx+xxx	32.5
3.35 / 3.85	3.6	11	xxxx+xxxx+x	27.5
3.85 / 4.35	4.1	6	xxxx+x	15
4.35 / 4.85	4.6	4	xxxx	10

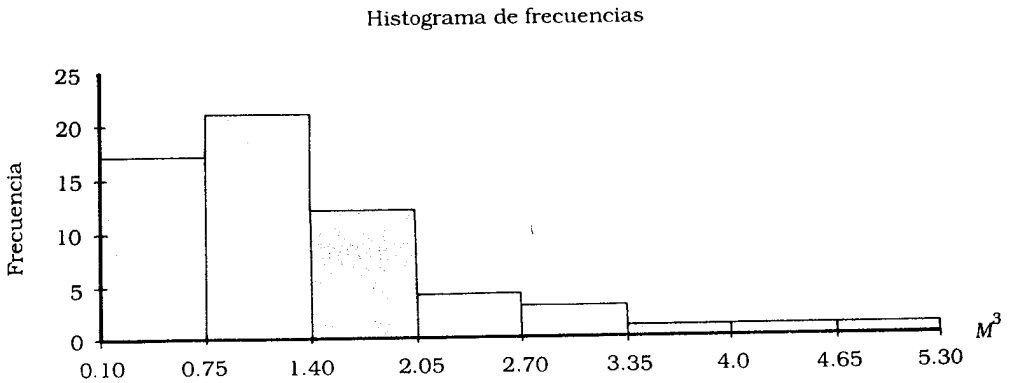
Suma de frecuencias: 40

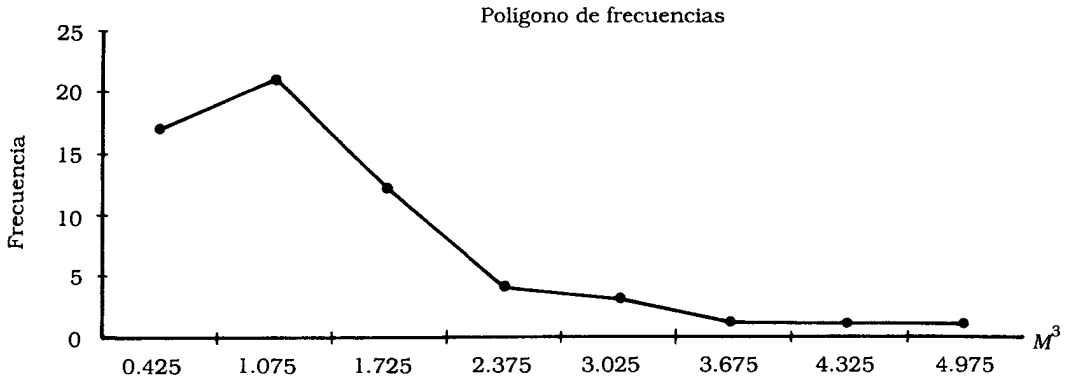
$n = 40$
$R = 3.1$
$R^* = 3.2$
$a = 0.4$
$\Delta = 0.1$
$\Delta / 2 = 0.05$
$LRS = 4.75$
$LRI = 1.55$

$\overline{X} = 135.6 / 40 = 3.39$
$S = 0.700366$
$S^2 = 0.4905$
$CV = (0.7/3.39)(100) = 20.64\%$
Confiabilidad = 79.36%
$Me = 3.15 + [(20-13)/11](0.4) = 3.40$
$Mo = 3.15 + [8/15](0.4) = 3.15 + 0.2133 = 3.363$
$AS = (3.39 - 3.363)/0.700366 = 0.03855$

2.18 $n = 60$
Val. máximo = 5.2
Val. mínimo = 0.2
Rango = $R = 5$
$R^* = 5.2$
$K = 8$
$a = 0.65$
$\Delta = 0.2$
$\Delta/2 = 0.1$
$LRI = 0.1$
$LRS = 5.3$
$\overline{X} = 82.05/60 = 1.3675$
$S = (57.8191/59)^{1/2} = 0.989942$
$S^2 = 0.979985$
$Me = 0.75 + [13/21](0.65) = 1.15238$
$Mo = 0.75 + [12/29](0.65) = 1.01896$
$CV = [0.989942/1.3675](100) = 72.39\%$
Confiabilidad = 100% – 72.39% = 27.61%
Asimetría = (1.3675 – 1.01896)/0.989942 = 0.352076

Y así construimos lo siguiente:





2.19 $n = 28$
$\overline{X} = 80.679$
$S = 11.902$
$S^2 = 141.65$
$CV = 14.75\%$
Confiabilidad = 85.25%

2.20

Tiempo	Tiempo transformado
2.50	2.833
2.55	2.917
3.00	3.00
3.00	3.00
3.10	3.167
3.05	3.001
3.07	3.001
3.20	3.333
3.22	3.367
3.30	3.500
3.38	3.633
3.50	3.833
4.00	4.000
4.05	4.001
4.30	4.500
4.15	4.250
3.18	3.300
3.25	3.417
2.58	2.967
3.19	3.317
3.24	3.400
3.25	3.417
3.32	3.533
3.22	3.367
3.22	3.367

Nota: El tiempo se transforma de la siguiente manera:

para 2.50

$$2.50 = 2 + \frac{50}{60} = 2 + 0.833$$

$$\boxed{2.50 = 2.833}$$

para 4.30

$$4.30 = 4 + \frac{30}{60} = 4 + 0.5$$

$$\boxed{4.30 = 4.5}$$

Con los tiempos transformados se obtiene $\overline{X}$ y S.

$$\boxed{\overline{X} = 3.417} \quad \boxed{S = 0.42702}$$

Se regresan al tiempo original

$$3.417 = 3 + 0.417 = 3 + (0.417 \times 60 \text{ min})$$

$$\boxed{\overline{X} = 3.25}$$

$$0.42702 \times 60 = 25.621 = 25 \text{ min}$$

$$0.621 \times 60 = 37 \text{ seg}$$

$$\boxed{S = 25 \text{ min con } 37 \text{ seg}}$$

2.21

MEDIA	1.58		
DESV. E	0.47		
VARIANZA	0.22	C.V.	29.86
MÍNIMO	0.65		
MÁXIMO	2.55		
AMPLITUD	1.90		
REDONDEO	2.00		
$K = 8$		MODA	1.72
INTERVALO	0.25	MEDIANA	1.65
$K = 10$		MODA	1.69
INTERVALO	0.20	MEDIANA	1.64
$K = 12$		MODA	1.70
INTERVALO	0.17	MEDIANA	1.65

2.22 Edad: $\overline{X} = 29.7$; $s = 7.54$
Sexo: hombres, 55%; mujeres, 45%
Raza: caucásico 95%; color, 5%
Localización de la lesión: lengua, 15%; paladar, 25%; piso de boca 20%; mucosa yugal, 20% labios, 10% encías, 10%
Lesiones extraorales: zonas cutáneas, 40%; mucosas 60%
Frecuencia de recidiva: no hubo, 30%; sí hubo, 70%

2.23 Edad: $\overline{X} = 34.75$ años; $s = 15.24$; $s^2 = 232.3$
Sexo: hombres, 60%; mujeres, 40%
Diente por tratar: anteriores, 35%; premolares, 30%; molares, 35%
Número de conductos: 1, 50%; 2, 20%; 3, 30%
Vitalidad: vital, 45%; no vital, 55%
Antecedentes de caries, 15%; resina, 35%; amalgama, 20%; incrustración, 25%, otros, 5%
Tratamiento: recubrimiento, 15%; de conductos, 45%; cirugía periapical, 20%; otros, 20%

2.24 Esta media aritmética no es representativa
$\overline{X} = 4$
$S = 6.708$ $Me = 1$
$S^2 = 45$ $Mo =$ No hay
$CV = 167\%$

2.25 $\overline{G} = 35.38\%$
La media geométrica es la medida de tendencia central más adecuada que la media aritmética cuando se promedian tasas de cambio. Esta medida no puede utilizarse cuando haya valores negativos o ceros.

2.26 Costo promedio $= \dfrac{\text{Costo total}}{\text{cantidad total}} = \dfrac{4\,000}{1\,000 + 800 + 625 + 500} = \dfrac{4\,000}{2\,925} = \$1.37$

donde

$$1\,000 = \frac{1}{1}(1\,000),\ 800 = \left(\frac{1}{1.25}\right) 1\,000,\ 625 = \left(\frac{1}{1.60}\right) 1\,000$$

Otra manera es

$$\overline{A} = \frac{4}{\dfrac{1}{\$\,1.0} + \dfrac{1}{\$\,1.25} + \dfrac{1}{\$\,1.60} + \dfrac{1}{\$\,2.0}} = \$1.37$$

2.27

$$\overline{X}p = \frac{(1 \times 6.5) + (1 \times 7) + (1 \times 8.5) + (2 \times 9)}{1 + 1 + 1 + 2}$$

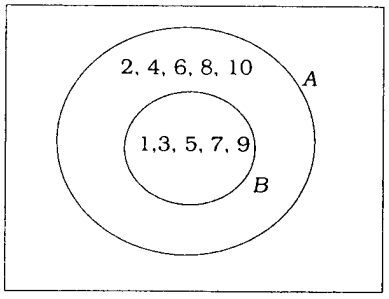
$$\overline{X}p = \frac{40}{5} = 8.0$$

Capítulo 3

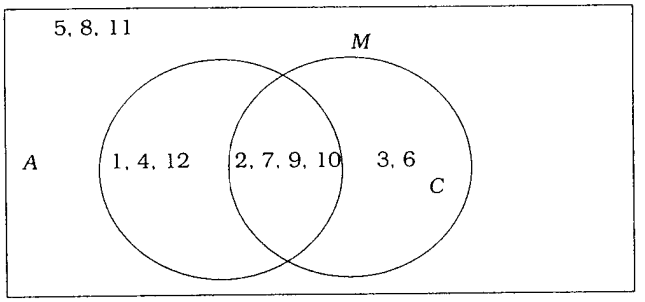
3.1 *a*) *c*), *d*), *e*), *f*)

3.2 *a*) {*a*, *b*, *c*}, {*a*, *b*}, {*a*, *c*}, {*b*, *c*}, {*a*}. {*b*}, {*c*}, Ø, Esto es 2^n, o sea $2^3 = 8$ subconjuntos.
b) {*a*, *b*, *c*}

3.3 $A' = \{2, 4, 6. 8, 10\}$



C = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}



$C = (A \cup M)'$

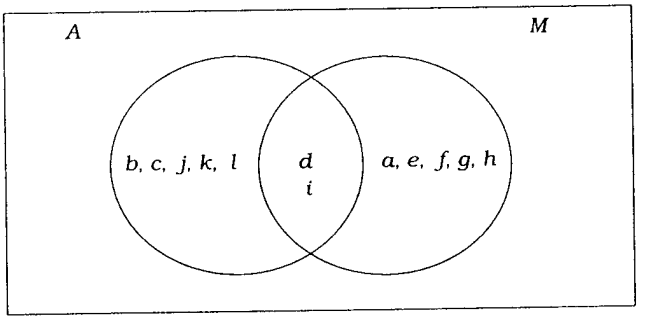
3.6 $A = \{x \mid x$ utilizan autobús$\}$
$M = \{x \mid x$ utilizan metro$\}$
$C = \{x \mid$ x caminan a su trabajo o utilizan automóvil propio$\}$

3.7

1) V	5) F	9) V
2) F	6) F	10) V
3) V	7) V	
4) V	8) F	

3.8

$C \approx D$	$G \neq H$
$C \neq D$	$C \approx H$
$C \neq E$	$C \neq G$
$C = F$	$D \approx F$
	$D \approx H$

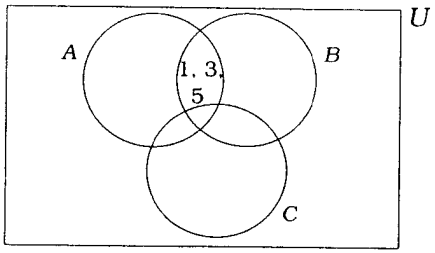


3.10

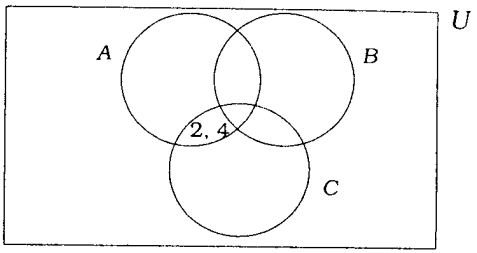
	F	$\overline{F}$	Total
H	45	15	60
M	30	10	40
	75	25	100

3.11 *a)* {1, 3, 5}
b) {2, 4}
c) { } = Ø
d) { } = Ø
e) $A \cup B$ = {1, 2, 3, 4, 5, 7, 9}
f) B = {1, 3, 5, 7, 9}
g) {2, 4, 6, 7, 8, 9, 10}
h) {1, 2, 3, 4, 5, 7, 9}
i) { } = Ø
j) {1, 2, 3, 4, 5} = A
k) {2, 4}
l) {1, 2, 3, 4, 5, 7, 9}
m) {1, 2, 3, 4, 5, 6, 8, 10}
n) {6, 8, 10}
o) { } = Ø

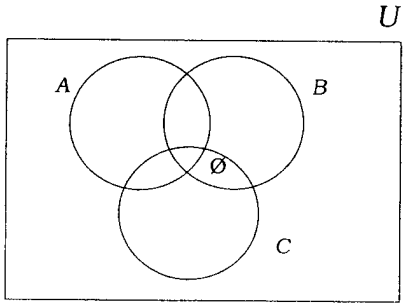
3.12 *a)*



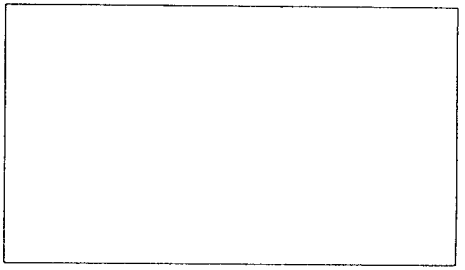
b)



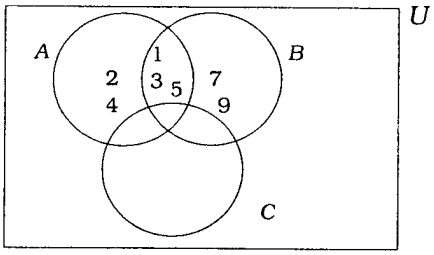
c)



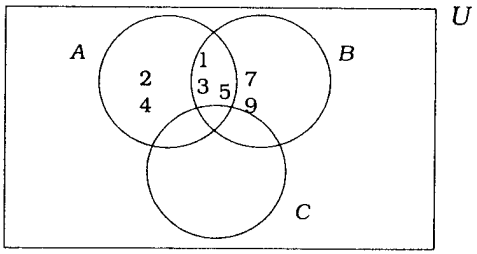
d) Conjunto vacío



e)

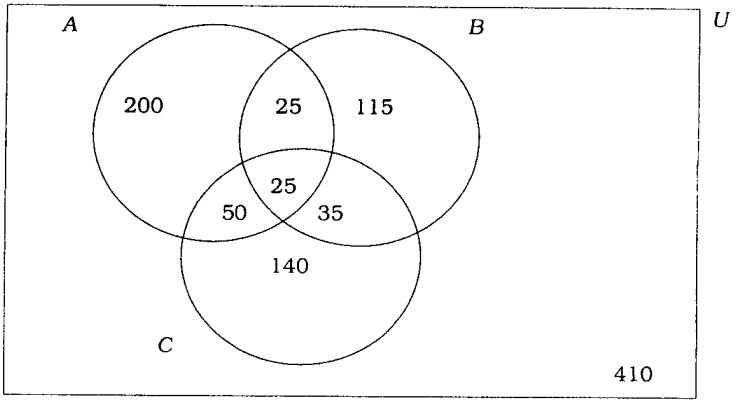


f)

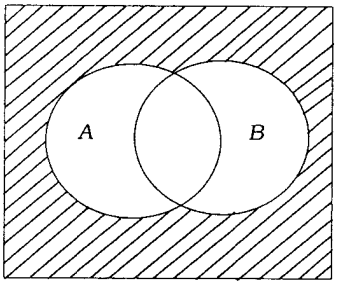


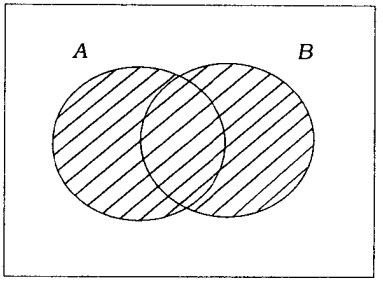
3.13 *a)* {−5, −4, −3, −2, −1, 0, 1, 2, 3, 4, 5}
b) {0, 1, 2, 3, 4, 5, −2, −4}
c) {1, 2, 3, 4, 5, −2, −4, 0}
d) {1, 2, 3, 4, 5}

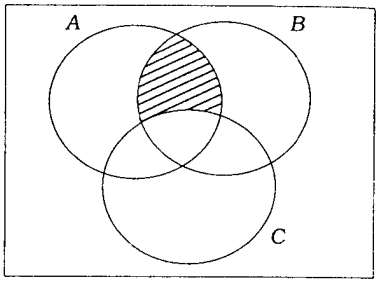
3.14 *a*) 410
b) 200
c) 35



3.15 *a*) *b*) *c*)

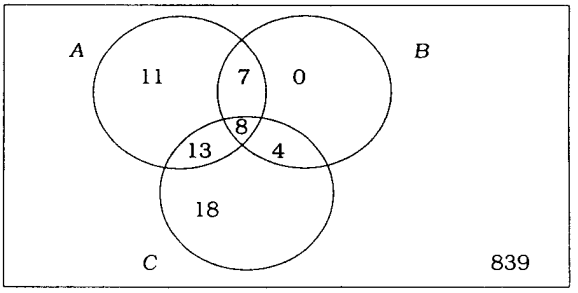






donde: *a*) $(A \cup B)'$ es lo sombreado *b*) $b(\overline{A' \cap B'})'$ *c*) $A \cap B - (A \cap B \cap C)$ es lo sombreado

3.16 *a*) 4 artículos $n\,[(B \cap C) - (A \cap B \cap C)]$
b) $\emptyset$, $n\,[B - (A \cap B \cup B \cap C) \cup A \cap B \cap C]$
c) 839 artículos, $n\,(A \cup B \cup C)'$
d) 49 artículos, $n\,[(A \cup B \cup C) - (B \cap C)]$
e) 29 artículos, $n\,\{(A \cup B \cup C) - (A \cap B) \cup [(A \cap C) - (A \cap B \cap C)] \cup [(B \cap C) - (A \cap B \cap C)]\}$
f) 24 artículos, $n\,[(A \cap B) - (A \cup B \cup C)] \cup [(A \cup C) - (A \cap B \cap C)] \cup [(B \cap C) - (A \cap B \cap C)]$



g) 61 artículos, $n\,(A \cup B \cup C)$
h) 32 artículos, $n\,\{(A \cap B) \cup [(A \cap C) - (A \cap B \cap C)] \cup [(B \cap C) - (A \cap B \cap C)]\}$
i) 868 artículos, $n\,\{(A \cap B) \cup [(A \cap C) - (A \cap B \cap C)] \cup [(B \cap C) - (A \cap B \cap C)]'\}$
j) 892 artículos, $n\,(A \cap B \cap C)'$

3.17

	H	M	C_u	Total
A	18	24	9	51
B	11	8	19	38
C	9	10	12	31
Total	38	42	40	120

1) *a)* 69
b) 31
c) 42
d) 19
e) 9
f) 42

2) 38 hondureños
42 mexicanas
40 cubanos

3.18

	V	R	A	a	Total
m	4	10	5	3	22
f	1	6	11	12	30
p	1	3	4	1	9
Total	6	19	20	16	61

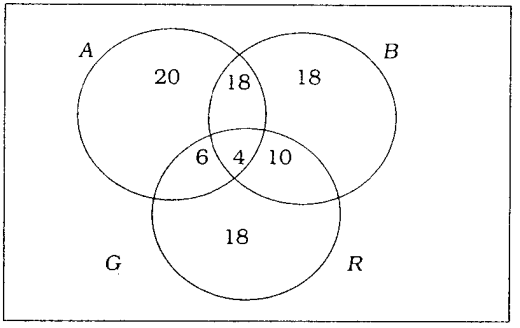
1) *a)* 6
b) 39
c) 35
d) 5
e) 60
f) 57
g) 26

2) *a)* 10
b) 6
c) 10
d) 5
e) 3
f) 4
g) 11
h) 12
i) 1
j) 4
k) 1
l) 1

3.19 *a)* $n(A \cup B) = n(A) + n(B)$
b) $n(A \cap B) = \emptyset$

3.20 *a)* 8
b) A^+, B^+, AB^+, O^+, A^-, B^-, AB^-, O^-

3.21 *a)* 18 personas
b) 6 personas



$U = 100$

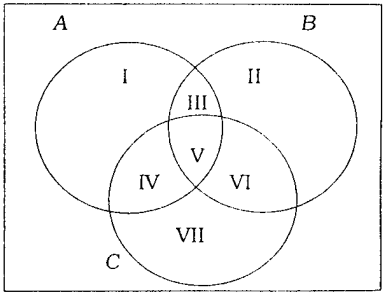
3.22

	Mexicanos	Extranjeros	Total
Hombres	8	170	178
Muchachos	2	16	18
Muchachas	16	14	30
Total	26	200	226

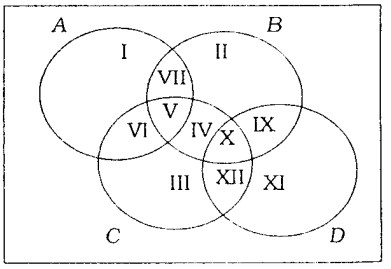
3.23 *a)*
I. $A - [(A \cap B) \cup (A \cap C)]$
II. $B - [(A \cap B) \cup (B \cap C)]$
III. $(A \cap B) - (A \cap B \cap C)$
IV. $(A \cap C) - (A \cap B \cap C)$
V. $A \cap B \cap C$
VI. $(B \cap C) - (A \cap B \cap C)$
VII. $C - [(A \cap C) \cup (B \cap C)]$
VIII. $(A \cup B \cup C)'$

b)
I. $A - [(A \cap B) \cup (A \cap C)]$
II. $B - [(B \cap D) \cup (B \cap C) \cup (B \cap A)]$
III. $C - [(A \cap C) \cup (B \cap C) \cup (C \cap D)]$
IV. $B \cap C - [(A \cap B \cap C) \cup (B \cap D \cap C)]$
V. $A \cap B \cap C$
VI. $(A \cap C) - (A \cap B \cap C)$
VII. $(A \cap B) - (A \cap B \cap C)$
VIII. $(A \cup B \cup C \cup D)'$
IX. $(B \cap D) - (B \cap C \cap D)$

a)



b)



X. $B \cap C \cap D$
XI. $D - [(C \cap D) \cup (B \cap D)]$
XII. $C \cap D - (B \cap D \cap C)$

3.24 I
a) 97
b) 97
c) 17
d) 68
e) 64
f) 75
g) 21
h) 97

	$H \cap L$	$P \cap L$	$T \cap H$	$P \cap T$	Total
A	22	13	11	4	50
B	17	18	7	5	47
	39	31	18	9	97

II
a) Conjunto de altas o bajas
b) Conjunto de muchachas hermosas o pedantes
c) Conjunto de muchachas altas y pedantes
d) Conjunto de todas las altas o pedantes y listas
e) Conjunto de todas excepto hermosas y listas así como hermosas y tontas.
f) Conjunto de todas excepto las (bajas hermosas y pedantes) y (hermosas, listas y bajas)
g) Conjunto (hermosas listas y bajas) o (pedantes, altas y tontas)

3.25
a) $A \cup B = A = B$
b) A o B
c) $\{x : x = 0, 1, 2 \ldots 10\}$
d) $\{1, 2, 3, 4, 5, 6\}$
e) $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30\}$
f) $\{10\}$
g) $\{x \mid x \varepsilon\ 0 \leq x \leq 10\}$
h) $\{1, 2, 3, 4, 5, 6\}$
i) $\{x \mid x = 20, 30$ o $0 \leq x \leq 10\}$
j) $\{10\}$
k) $\{0, 1, 2, 3, 4, 5, 6, 10, 20, 30\}$
l) $\{0\}$
m) $\{x \mid x = 0 \leq x \leq 10\}$
n) $\{1, 2, 3, 4, 5, 6, 7, 8, 9\ 10\}$
o) A
p) $\{1, 2, 3, 4, 5, 6\}$
q) $\{0, 1, 2, 3, 4, 5, 6, 10\}$
r) B

3.26
a) $A \cup B$
b) $A \cap B$
c) $B' \cup A$
d) $U - (A - B)$

3.27

	Sistema		
	Trimestral	Semestral	Total
Licenciatura	27	23	50
Posgrado	6	29	35
Total	33	52	85

a) 52
b) 6

3.28 Puesto que 70 compran *El Imparcial*

$n\{E \cap (\overline{M} \cup \overline{A})\} = 70 - 14 - 13 - 3 = 40$

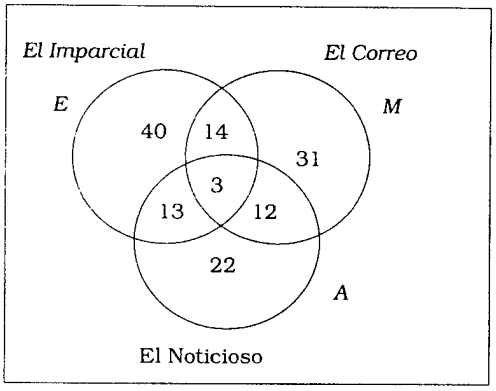
40 compran solamente El Imparcial

$n\{M \cap (\overline{A} \cup \overline{E})\} = 60 - 12 - 14 - 3 = 31$

31 compran solamente El Correo

$n\{A \cap (\overline{E} \cup \overline{M}\} = 50 - 12 - 13 - 3 = 22$

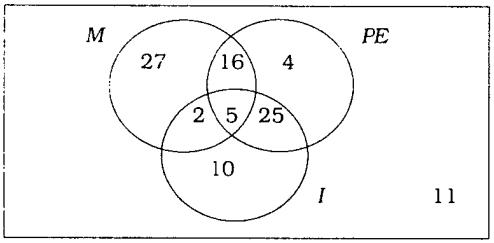
22 compran solamente El Noticioso



40 + 14 + 31 + 12 + 3 + 13 + 22 lectores

Podemos ver que la agencia noticiosa tiene: 135 lectores.

3.29 *a)*



b) *i)* $n(PE \cap H \cap I) = 25$
ii) $n(H \cap D) = 15$

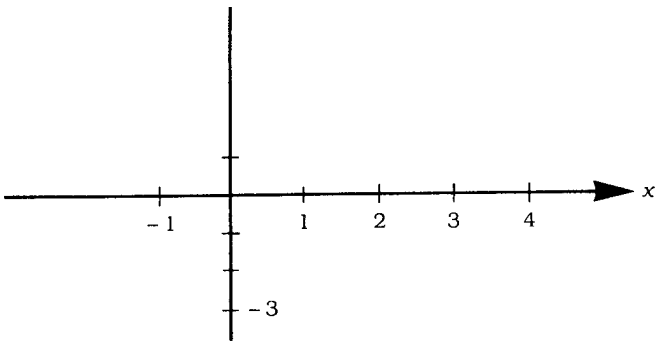
c)

	$E \cap I$	$\overline{E} \cap \overline{I}$	$E \cap \overline{I}$	$\overline{E} \cap I$	Total
M	5	27	16	2	50
H	25	11	4	10	50
					100

50 machos
50 previamente entrenados
42 entrenados tomaron a la izquierda en la primera oportunidad
21 machos entrenados previamente

7 machos tomaron a la izquierda
30 ratones previamente entrenados tomaron a la izquierda
5 machos previamente entrenados tomaron a la izquierda

3.30 Relación



3.31 $\{(3, 5), (2, 5)\}$ $D = \{3, 2\}$ $CD = \{5\}$ *Función*

3.32 *a*) función
b) función
c) función
d) función
e) función

3.33 $A(6, -3)$ y $B(-2, 3)$

$$y - y_1 = m(x - x_1)$$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{3 - (-3)}{-2 - (6)} = \frac{6}{-8} = -\frac{3}{4} \qquad y - (-3) = -\frac{3}{4}(x - 6)$$

$$y - \frac{3}{4}x + \frac{3}{2} \qquad 3x + 4y = 6$$

3.34 $m = 4$
$A(2, -3)$

$y - y_1 = m(x - x_1)$
$y - (-3) = 4(x - 2)$
$y = 4x - 11$
$4x - y = 11$

3.35 $y = 4x - 3$
$m = 4,\ b = -3$

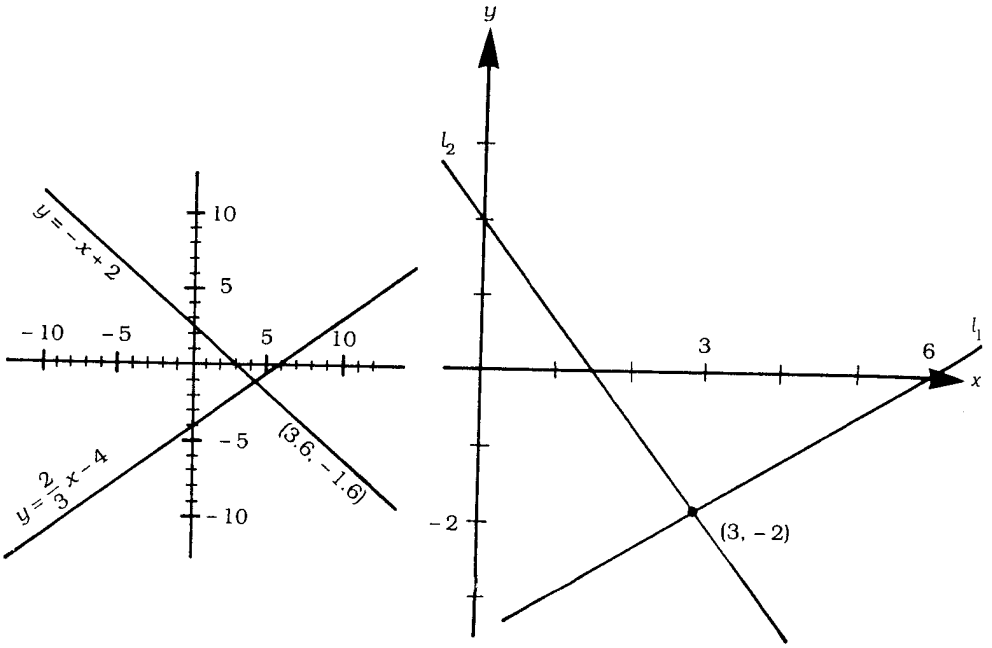
3.36 $m_1 = \dfrac{-6-2}{3-1}$ $\qquad m_2 = \dfrac{7-(-5)}{-1-2}$

$m_1 = \dfrac{-8}{2}$ $\qquad m_2 = \dfrac{12}{-3}$

$\boxed{m_1 = -4}$ $\qquad \boxed{m_2 = -4}$

Como $m_1 = m_2$, se deduce que l_1 y l_2 son paralelas.

3.37 *a)*



b) $y = \dfrac{2}{3}x - 4$

$y = -\dfrac{4}{3}x + 2$

c) En el punto de intersección las coordenadas y son iguales y, por tanto, en este punto los segundos miembros de las ecuaciones también son iguales. Al igualar estas dos expresiones resulta:

$$\frac{2}{3}x - 4 = -\frac{4}{3}x + 2$$

$$\frac{2}{3}x + \frac{4}{3}x = 2 + 4$$

$$\frac{6}{3}x = 6$$

$$x = \frac{18}{6} = 3$$

Este valor de x es la abscisa del punto de intersección. Para determinar la ordenada sustituimos de la siguiente manera.

$$y = \frac{2}{3}x - 4$$

$$y = \frac{2}{3}(3) - 4$$

$$y = -2$$

Por tanto, el punto de intersección es $(3, -2)$.

3.38 Si m_1 es la pendiente que pasa por A y B, y m_2 es la pendiente que pasa por B y C, entonces:

$$m_1 = \frac{-1 - (-4)}{2 - (-3)}$$

$$m_1 = \frac{3}{5}$$

$$m_2 = \frac{2 - (-1)}{7 - 2}$$

$$m_2 = \frac{3}{5}$$

Así que $m_1 = m_2$. Por lo tanto, la recta que pasa por A y B, y la que pasa por B y C tienen la misma pendiente y el punto común B. Así pues, ambas son la misma recta y por ello, A, B y C son colineales.

3.39 $A^2 = I$ Es una matriz idempotente.

3.40 $A^3 = O$ La matriz se llama *matriz nilpotente*.

3.41 A es una matriz simétrica.

3.42 $AB = (BA)^T$

3.43 $A^{-1} = \begin{pmatrix} \frac{-7}{4} & 2 & \frac{3}{8} \\ \frac{5}{4} & -1 & \frac{-1}{8} \\ \frac{-1}{4} & 0 & \frac{1}{8} \end{pmatrix}$ $\quad |A| = -8$ $\quad$ A es **no** singular, si tiene matriz inversa.

3.44 Como $|B| = 0$, es una matriz singular y no tiene inversa.

3.45 No tiene solución el sistema.

3.46 $(x, y) = (0, 0)$ $\quad$ det $|A| = -11$

3.47 $x = \frac{-13}{5}$ $\quad$ $y = -\frac{11}{5}$

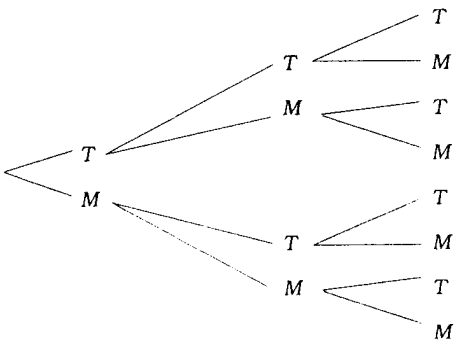
3.48 $x = 2$
$y = -3$
$z = 1$

3.49 $x_1 = 3, \quad x_2 = 0, \quad x_3 = 0$

3.50 $|D| = 0$ $\quad$ El sistema no tiene *solución*.

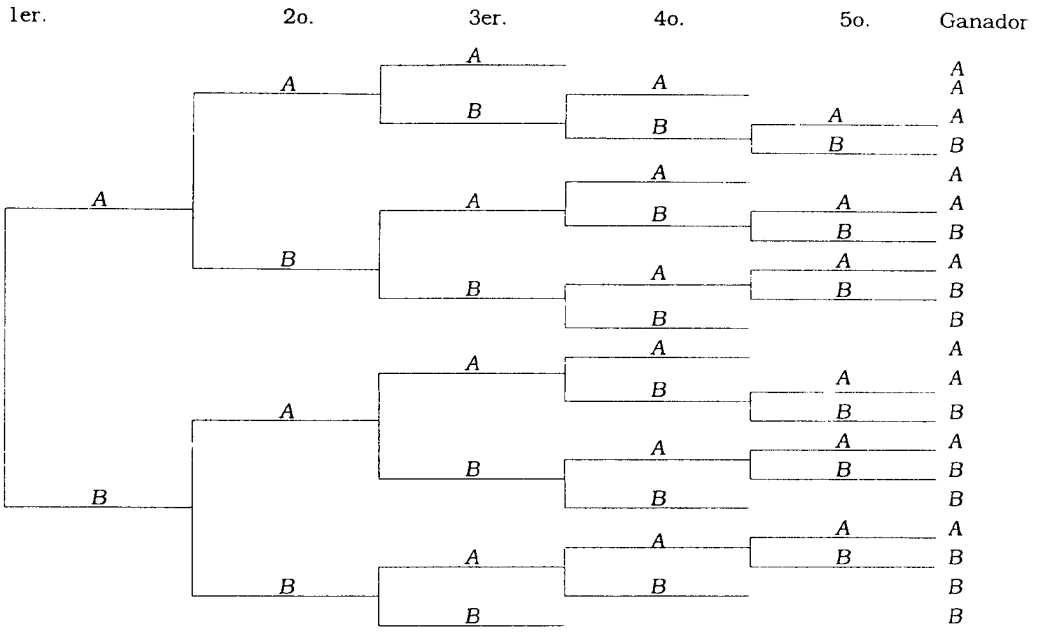
Capítulo 4

4.1 *a)* T = Trigo, M = Maíz



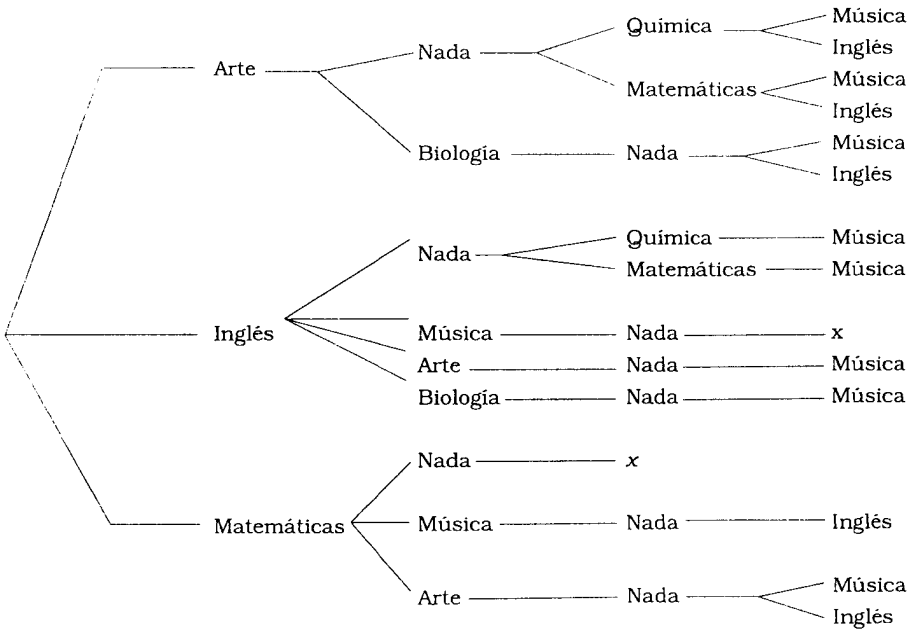
b) {*TTT, TTM, TMT, TMM, MTT, MTM, MMT, MMM*}

4.2



n = 20 (número de combinaciones)

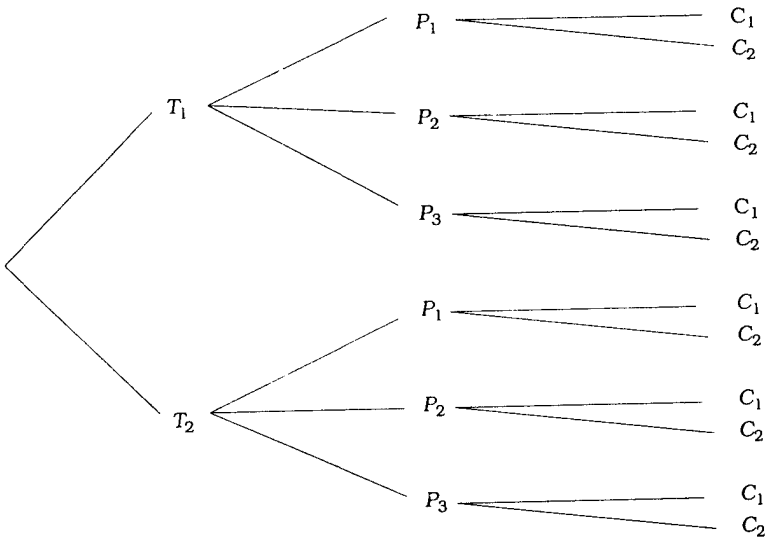
4.3



4.4 Dos postres para que el letrero sea cierto y se obtengan 216 comidas diferentes.

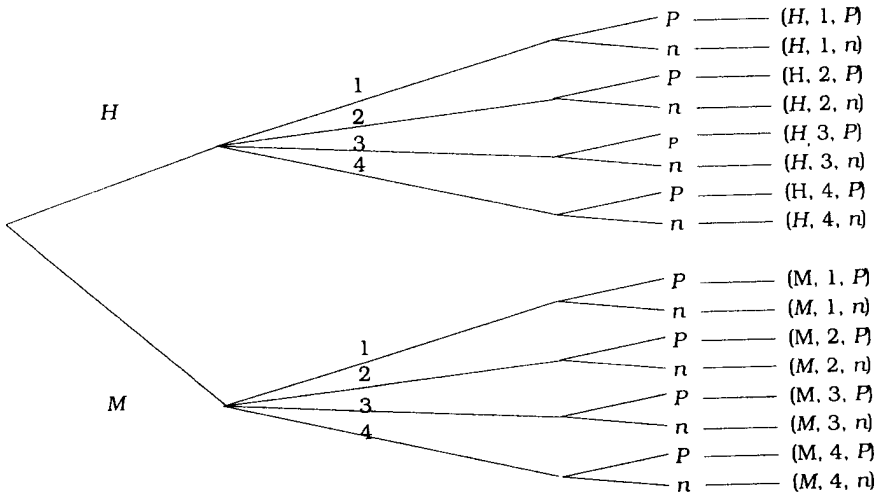
4.5 30 240 listas.

4.6



12 experimentos diferentes.

4.7 *a*)



b) 16

4.8 294 autos diferentes.

4.9 1 680. Aplique el principio de multiplicación.

4.10 *a)* 496, no importa el orden.
b) 992, sí importa el orden.

4.11 *a)* 11 232 000
b) 17 576 000
c) 22 464 000 y 35 152 000 juegos de placas, respectivamente.

4.12 7 059 052 combinaciones diferentes.

4.13 $(n - 1)! = > (12 - 1)! = 11! = 39\ 916\ 800$, formas diferentes.

4.14 $(4 - 1)! = (3!) = 6$ collares diferentes.

4.15 $C^6_6 = 1$ $C^6_5 = 6$ $C^6_4 = 15$ $C^6_3 = 20$ $C^6_2 = 15$ $C^6_1 = 6$ $C^6_0 = 1$

4.16 $7 \times 5 = 35$ parejas diferentes.

4.17 92 000 parejas diferentes.

4.18 1 365 equipos diferentes.

4.19 210 exámenes diferentes

4.20 *a)* 60 nombres diferentes.
b) 12 nombres diferentes.
c) 20 160 nombres diferentes.
d) 1) Kalimán 2 520 nombres diferentes.
2) Kalimán 5 040 nombres diferentes.
e) 720 nombres diferentes.

4.21 *a)* r^3
b) $2 + 2r + 2r^2 + 2r^3$
c) 302 integrantes.

Capítulo 5

5.1 *a)* $n = 20$ *b)* $P(R) = \frac{1}{2}$ *c)* $P(B) = \frac{1}{4}$ *d)* $P(A) = \frac{1}{4}$

5.2 *a)* $P(A) = \frac{2}{5}$ *b)* $P(\overline{A}) = \frac{3}{5}$ *c)* $P(A) + P(\overline{A}) = \frac{2}{5} + \frac{3}{5} = \frac{5}{5} = 1$

5.3 *a)* $N = 50$ *b)* $n(m) = 20$ *c)* $n(H) = 30$ *d)* $P(M) = \frac{2}{5}$ *e)* $P(H) = \frac{3}{5}$

5.4 *a)* $\overline{E}_1$, (cruz en la moneda y cualquier número en el dado)
b) 1, 2, 3, 4, o 5 en el dado y cualquier lado en la moneda.
c) Probabilidad de cara en la moneda y 1, 2, 3, 4 o 5 en el dado.

5.5 *a)* $S = \{A, B, C, D\}$

$$P(A) = \frac{1}{4}, \quad P(B) = \frac{1}{4}, \quad P(C) = \frac{1}{4}, \quad P(D) = \frac{1}{4}$$

b) $\begin{pmatrix} AB & BC \\ AC & BD \\ AD & CD \end{pmatrix}$ para cualquier pareja, por ejemplo: $P(AB) = \frac{1}{6}$

c)

AB	BC	AA	CC
AC	BD	BB	
AD	CD	CC	

para cualquier pareja $P(AB) = \frac{1}{10}$

5.6 *a)* $P(C) = \frac{1}{3}$ *d)* $P(\overline{A} \cap \overline{B}) = \frac{1}{3}$

b) $P(A \cup B) = \frac{2}{3}$ *e)* $P(\overline{A} \cup \overline{B}) = 1$

c) $P(\overline{A}) = \frac{2}{3}$ *f)* $P(B \cup C) = \frac{2}{3}$

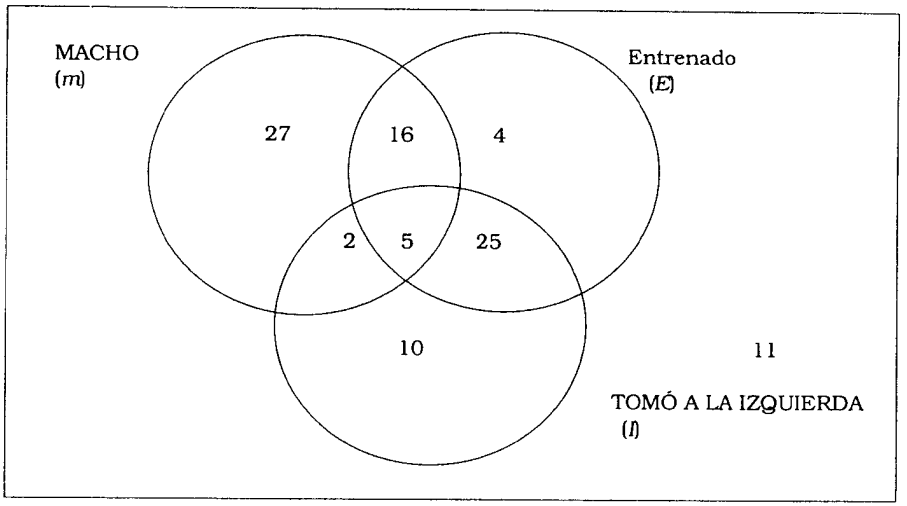
5.7 *a)* $P(\overline{A}) = \frac{1}{2}$ *e)* $P(\overline{A} \cup \overline{B}) = \frac{2}{3}$

b) $P(\overline{B}) = \frac{1}{2}$ *f)* $P(A \cap \overline{B}) = \frac{1}{6}$

c) $P(A \cap B) = \frac{1}{3}$ *g)* $P(\overline{A \cup B}) = \frac{1}{3}$

d) $P(\overline{A} \cap \overline{B}) = \frac{1}{3}$ *h)* $P(\overline{A} \cup B) = \frac{5}{6}$

5.8 *a)* Diagrama de Venn Euler



$n = 100$

b)

Diagrama de Carroll

	$E \cap I$	$\bar{E} \cap \bar{I}$	$E \cap \bar{I}$	$\bar{E} \cap I$	Total
M	5	27	16	2	50
H	25	11	4	10	50
					100

c) $\frac{25}{100}$ *e)* $\frac{4}{100}$

d) $\frac{2}{100}$

5.9 *a)* $\frac{55}{885} = 0.06215 \Rightarrow 6.21\%$

b) $\frac{195}{885} = 0.2203 \Rightarrow 22.03\%$

c) $\frac{0}{885} = 0$ Recuerde que todas al menos ven uno de estos programas.

d) $\frac{190}{885} = 0.21469 \Rightarrow 21.47\%$

e) $\frac{140}{885} + \frac{350}{885} = 0.15819 + 0.39548 = 0.55367 \Rightarrow 55.37\%$

5.10 *a)* $P(H \cap M) = P(H) \cdot P(M) = 0.12$

b) Como $P(H) = 0.3$, $P(\overline{H}) = 0.7$

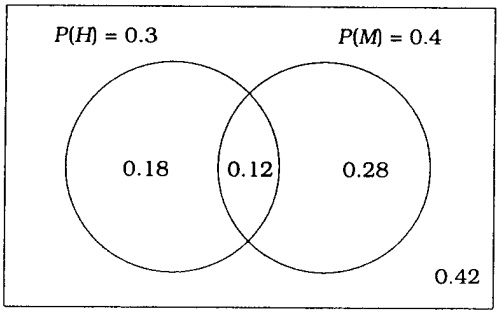
c) Como $P(M) = 0.4$, $P(\overline{M}) = 0.6$

d) $P(\overline{H} \cap \overline{M}) = 0.42$

e) $P(H \cup M) = 0.58$

f) $P(\overline{H \cap M}) = 1 - P(H \cap M) = 1 - 0.12 = 0.88$

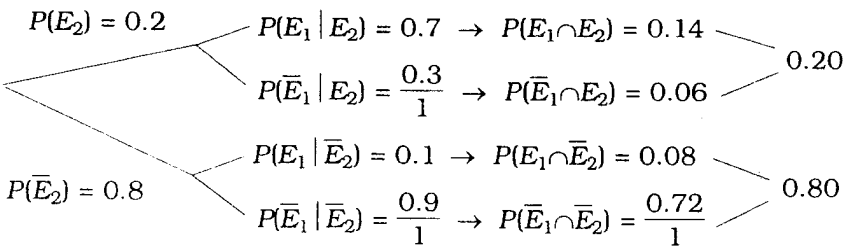
g) Diagrama de Venn



5.11 *a)*

	Llueva mañana E_1	No llueva mañana $\overline{E}_1$	
Llueve hoy E_2	$E_1 \cap E_2$ 0.14	$\overline{E}_1 \cap E_2$ 0.06	0.20
No llueve hoy $\overline{E}_2$	$E_1 \cap \overline{E}_2$ 0.08	$\overline{E}_2 \cap \overline{E}_1$ 0.72	0.80
	0.22	0.78	1

b)



c) $P(E_1) = P(E_1 \cap E_2) + P(E_1 \cap \overline{E}_2) = 0.14 + 0.08$
$P(E_1) = 0.22$

5.12 a) $P(E_1) = \dfrac{25}{300}$

b) $P(E_1 \cup E_6) = \dfrac{45}{300}$

c) $P(E_1 \cap E_6) = \dfrac{20}{300}$

d) $P(E_1 \mid \Omega) = P(E_1)$

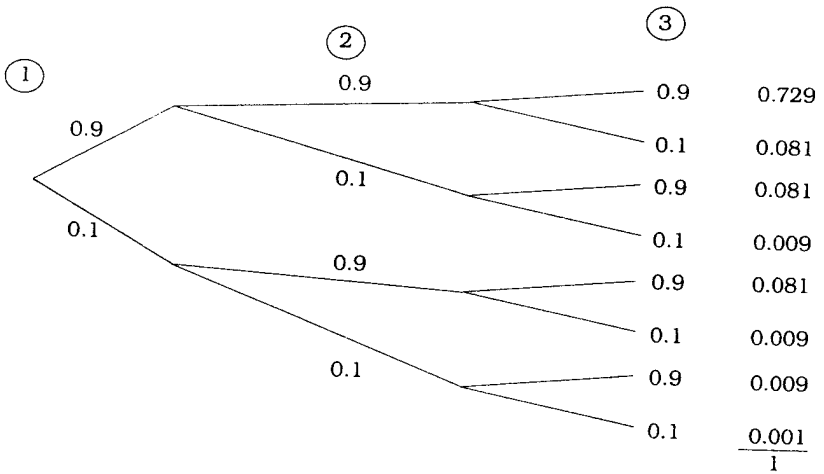
e) $P(E_1 \cup E_6 \mid \Omega) = \dfrac{45}{300}$

f) $P(E_1 \mid E_6) = \dfrac{20}{40}$

g) $P(E_6 \mid E_1) = \dfrac{20}{25}$

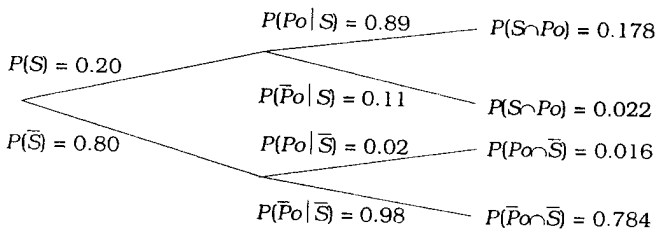
5.13 a) 0.001
b) 0.081
c) 0.009
d) 0.243
e) 0.027
f) 0.271
g) 0.028

Diagrama de árbol



5.14 a) 0.488
b) 0.104

5.15



	Po	$\overline{P}o$	Total
S	0.178	0.022	0.2
$\overline{S}$	0.016	0.784	0.8
Total	0.194	0.806	1

S = con sida

$\overline{S}$ = sin sida

P = reacción positiva

$\overline{P}o$ = reacción negativa

$$P(S \mid Po) = \frac{0.178}{0.178 + 0.016} = 0.9227$$

$$P(S \mid \overline{P}o) = \frac{0.022}{0.022 + 0.784} = 0.0273$$

$$P(\overline{S} \mid Po) = \frac{0.016}{0.178 + 0.016} = 0.0825$$

$$P(\overline{S} \mid \overline{P}o) = \frac{0.784}{0.022 + 0.784} = 0.9727$$

$$P(Po \mid S) = \frac{0.178}{0.178 + 0.022} = 0.89$$

$$P(Po \mid \overline{S}) = \frac{0.016}{0.016 + 0.784} = 0.020$$

$$P(\overline{P}o \mid S) = \frac{0.022}{0.178 + 0.022} = 0.110$$

$$P(\overline{P}o \mid \overline{S}) = \frac{0.784}{0.016 + 0.784} = 0.98$$

5.16 *a)*

	B	$\overline{B}$	
A	$P(A \cap B)$	$P(A \cap \overline{B})$	$P(A)$
$\overline{A}$	$P(\overline{A} \cap B)$	$P(\overline{A} \cap \overline{B})$	$P(\overline{A})$
	$P(B)$	$P(\overline{B})$	1

b) $$P(B \mid A) = \frac{P(A \cap B)}{P(\overline{A})} \Rightarrow P(A \cap B) = P(B \mid A)\, P(A)$$

$$P(B \mid \overline{A}) = \frac{P(\overline{A} \cap B)}{P(\overline{A})} \Rightarrow P(\overline{A} \cap B) = P(B \mid \overline{A})\, P(\overline{A})$$

c) Como

$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$
$$= P(B \mid A)\, P(A) + P(B \mid \bar{A})\, P(\bar{A}) \quad \ldots \text{(I)}.$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \ldots \text{(II)}$$

Se sustituye (I) en (II) obteniendo:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B \mid A)\, P(A) + P(B \mid \bar{A})\, P(\bar{A})}$$

Teorema de Bayes

$$\boxed{P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B \mid A)\, P(A) + P(B \mid \bar{A})\, P(\bar{A})}}$$

5.17 a)

	$C \cap P$	$C \cap \bar{P}$	$S \cap P$	$S \cap \bar{P}$	Total
H	125	72	83	37	317
M	94	25	15	49	183
Total	219	97	98	86	500

b) $\dfrac{219}{500} + \dfrac{93}{500} = \dfrac{312}{500}$

$P(P) = 62.4\%$

5.18 a) La probabilidad de que un paciente tenga cáncer, si cree tenerlo, es igual a 1.
b) La probabilidad de que tenga cáncer, si no cree tenerlo, es igual a 0.2.
c) La probabilidad de que crea tener cáncer y no lo tenga es igual a 9/17.
d) La probabilidad de que crea que tiene cáncer y sí lo tenga es igual a 1/3.

5.19 a) $P(C \mid P) = 0.095$
b) $P(C' \mid P) = 0.000264$
c) $P(C' \mid P') = 0.91$
d) $P(C \mid P') = 0.999754$

5.20

	S	$\bar{S}$	Total
H	0.08	0.22	0.3
$\bar{H}$	0.32	0.38	0.7
Total	0.4	0.6	1

H = gane hoy

$\bar{H}$ = no gane hoy

S = gane la siguiente semana

$\bar{S}$ = no gane la siguiente semana

5.21

	P	N	Total
I	0.0285	0.0015	0.03
$\bar{I}$	0.0194	0.9506	0.97
Total	0.0479	0.9521	1

I = con insomnio

$\bar{I}$ = sin insomnio

P = resultado positivo

N = resultado negativo

5.22

	H	$\bar{H}$	Total
S	0.08	0.32	0.4
$\bar{S}$	0.22	0.38	0.6
Total	0.3	0.7	1

S = gane la siguiente semana

$\bar{S}$ = no gane la siguiente semana

H = gane hoy

$\bar{H}$ = no gane hoy

5.23

	P	$\bar{P}$	Total
E	0.48	0.20	0.68
$\bar{E}$	0.12	0.20	0.32
Total	0.6	0.4	1.0

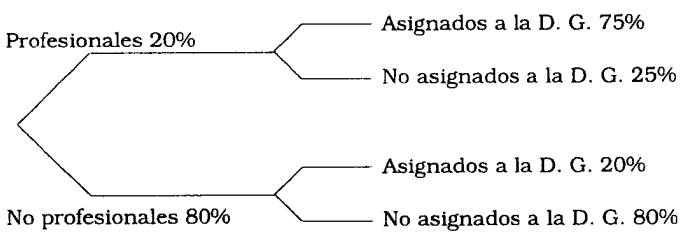
P = pasan la prueba

$\bar{P}$ = no pasan la prueba

E = terminan exitosamente

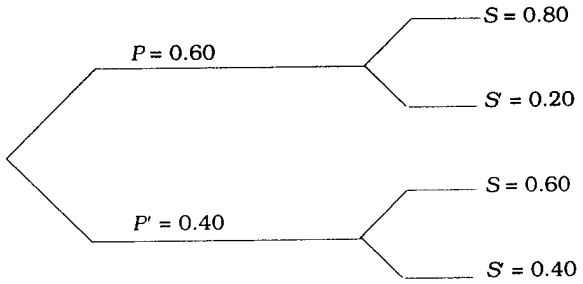
$\bar{E}$ = no terminan exitosamente

5.24 *a*)



b) $P(P \cap A) = P(P) \cdot P(A) = 0.15$
c) $P(P \cap A') = 0.05$
d) $P(P' \cap A) = 0.16$
e) $P(P \mid A) = 0.48$
f) $P(A \mid P) = 0.75$
g) $P(A' \mid P) = 0.25$
h) $P(P' \mid A) = 0.51$
i) $P(A \mid P') = 0.21$

5.25 *a)* Sea
P = Pase el examen
P' = No pase el examen
S = Trabajo satisfactorio
S' = Trabajo no satisfactorio



b) $P(P \mid S) = 0.66$
c) $P(P \mid S') = 0.42$
d) $P(S \mid P) = 0.80$
e) $P(S' \mid P) = 0.20$
f) $P(S \mid P') = 0.60$
g) $P(S' \mid P') = 0.40$

5.26 *a)*

	$H \cap L$	$P \cap L$	$H \cap T$	$T \cap P$	Total
A	22	13	11	4	50
B	17	18	7	5	47
Total	39	31	18	9	97

b) i) La probabilidad de que sea alta = 0.515
ii) La probabilidad de que sea baja = 0.485
iii) La probabilidad de que sea lista = 0.40
iv) La probabilidad de que sea pedante y lista = 31/97 = 0.31

v) La probabilidad de que sea alta dado que es hermosa y lista = 0.56
vi) La probabilidad de que sea alta dado que es pedante y lista = 0.41

Nota: los eventos que falta calcular se dejan al lector.

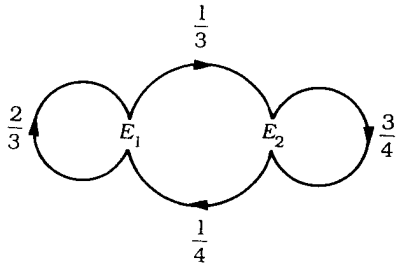
5.27 *a)*

	Mexicanos	Extranjeros	Total
Hombres	8	170	178
Muchachos	2	16	18
Muchachas	16	14	30
Total	26	200	226

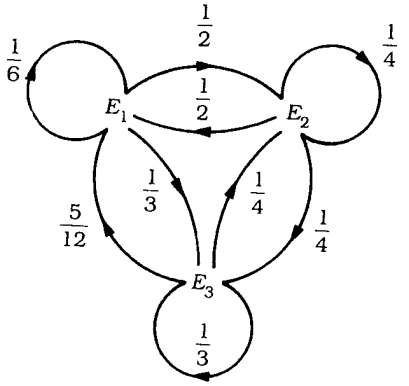
5.28 *a)* $\frac{31}{1045}$ *b)* $\frac{23}{1045}$ *c)* $\frac{31}{1045} + \frac{23}{1045} = \frac{54}{1045}$

5.29 *a)* $\frac{750}{1000}$ *b)* $\frac{420}{1000}$ *c)* 0.7 *d)* $\frac{110}{1000}$

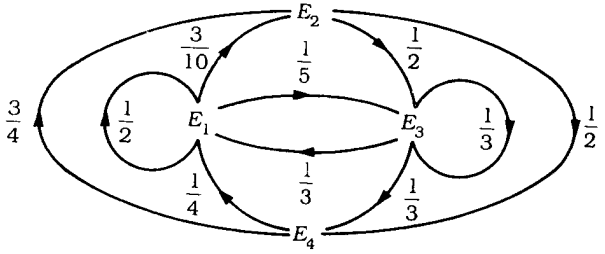
5.30



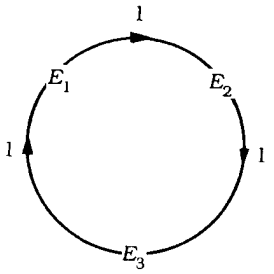
a)



b)



c)



d)

5.31 *a)*

$$\begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

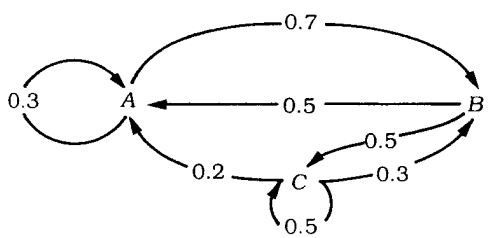
b)

$$\begin{array}{c c} & \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{array}{c} a_1 \\ a_2 \\ a_3 \end{array} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \end{array}$$

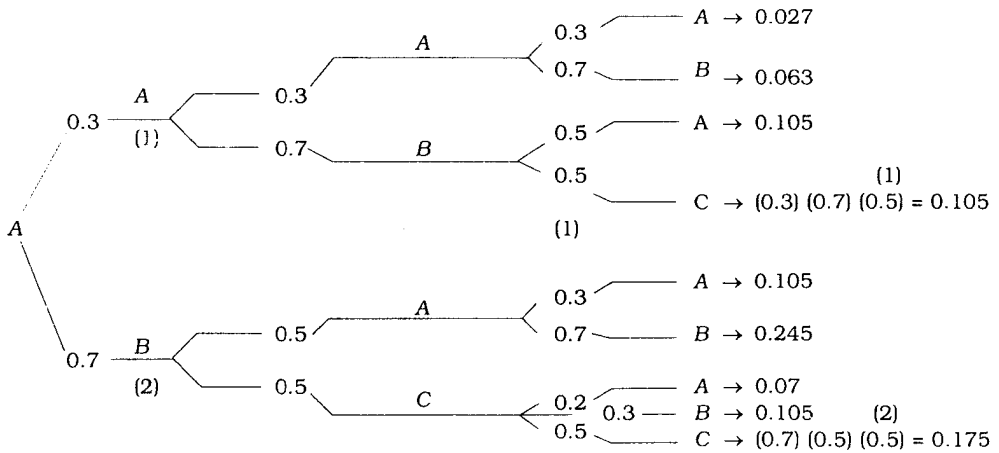
5.32 *a)*

$$\begin{array}{c c} & \begin{array}{ccc} A & B & C \end{array} \\ \begin{array}{c} A \\ B \\ C \end{array} & \begin{pmatrix} 0.3 & 0.7 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} \end{array}$$

b)



c)



5.33 *a)* $P(G_R) = 0.5 \Rightarrow 50\%$

b) $$T = \begin{array}{c} \\ G \\ P \end{array} \begin{array}{c} \begin{array}{cc} G & P \end{array} \\ \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} \end{array}$$

c) Como el vector inicial de probabilidad para ambas jugadoras es (0.5, 0.5), la probabilidad de que Rosalba gane el 2o. juego es:

$$(0.5,\ 0.5)\begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} = (0.45,\ 0.55)$$

La probabilidad de que gane el 2o. juego

$$\boxed{P(G_R) = 0.45}$$

d) Para calcular la probabilidad de que gane el 3er. juego se necesita T^2 y se obtiene

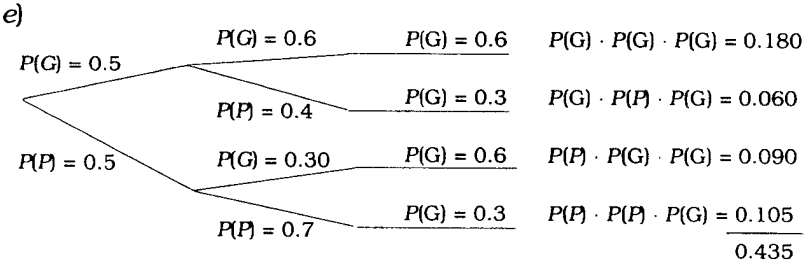
$$T^2 = \begin{array}{c} \\ G \\ P \end{array} \begin{array}{c} \begin{array}{cc} G & P \end{array} \\ \begin{pmatrix} 0.48 & 0.52 \\ 0.39 & 0.61 \end{pmatrix} \end{array}$$

La probabilidad de que gane Rosalba en el 3er. juego.
$P(G_R)$

$$(0.5,\ 0.5)\begin{pmatrix} 0.48 & 0.52 \\ 0.39 & 0.61 \end{pmatrix} = (0.435,\ 0.565)$$

$$\boxed{P(G_R) = 0.435}$$

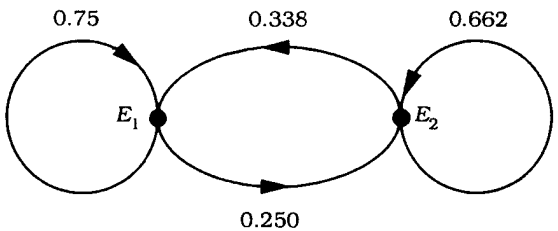
e)



5.34 *a)*

$$\begin{array}{cc} & \begin{array}{cc} E_1 & E_2 \end{array} \\ \begin{array}{c} E_1 \\ E_2 \end{array} & \begin{pmatrix} 0.750 & 0.250 \\ 0.338 & 0.662 \end{pmatrix} \end{array}$$

b)



5.35 *a)*

$$T = \begin{array}{cc} & \begin{array}{cc} \text{Cd. Méx y Z. M.} & \text{Provincia} \end{array} \\ \begin{array}{c} \text{Cd. Méx y Z. M.} \\ \text{Provincia} \end{array} & \begin{pmatrix} 0.999 & 0.001 \\ 0.008 & 0.992 \end{pmatrix} \end{array}$$

b) $A \cdot T^3 = \begin{array}{cc} \text{C. M. y Z. M.} & \text{Provincia} \\ \begin{pmatrix} 0.26710 & 0.73290 \end{pmatrix} \end{array}$, el 26.7% vivirá en la Ciudad de México y Z.M. y el 73. 29% en provincia ya que:

$$T^3 = \begin{pmatrix} 0.9970 & 0.0030 \\ 0.0238 & 0.9762 \end{pmatrix}$$

c) Como T^{10} es:

$$T^{10} = \begin{pmatrix} 0.9904 & 0.0096 \\ 0.0768 & 0.9232 \end{pmatrix}$$

entonces, $A \cdot T^{10} = (0.30520, 0.69480)$
El 30.52% vivirá en la Ciudad de México y Zona Metropolitana.
El 69.48% vivirá en provincia.

5.36 *a)* 0.667 = 66.7%
b) (0.667, 0.333)
c) Sí

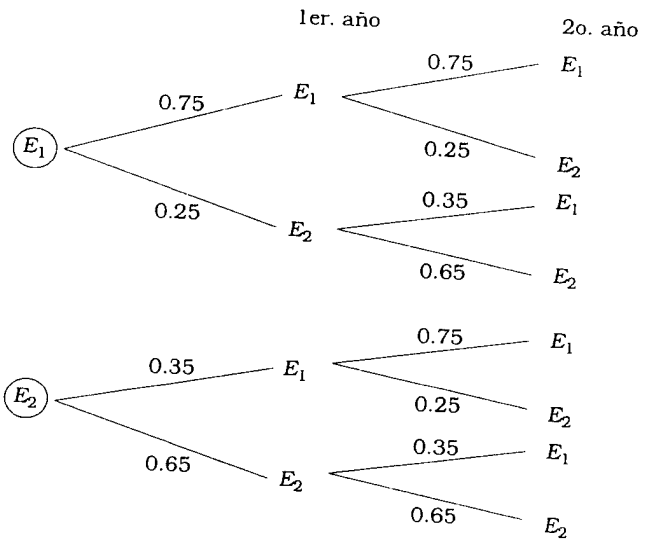
5.37 *a)*

$$P = \begin{array}{c} \\ \text{Modelo } X\,(E_1) \\ \text{Otros modelos } (E_2) \end{array} \begin{array}{c} \begin{array}{cc} \text{Modelo } X\,(E_1) & \text{Otros modelos } (E_2) \end{array} \\ \begin{pmatrix} 0.75 & 0.25 \\ 0.35 & 0.65 \end{pmatrix} \end{array}$$

$$P^2 = \begin{pmatrix} 0.75 & 0.25 \\ 0.35 & 0.65 \end{pmatrix} \begin{pmatrix} 0.75 & 0.25 \\ 0.35 & 0.65 \end{pmatrix} = \begin{pmatrix} 0.65 & 0.35 \\ 0.49 & 0.51 \end{pmatrix}$$

$$\begin{array}{c} \\ E_1 \\ E_2 \end{array} \begin{array}{c} \begin{array}{cc} E_1 & E_2 \end{array} \\ \begin{pmatrix} 0.65 & 0.35 \\ 0.49 & 0.51 \end{pmatrix} \end{array}$$

b)



$P(E_1 \cap E_1) = (0.75)(0.75) = 0.5625$

$P(E_1 \cap E_2) = (0.75)(0.25) = 0.1875$

$P(E_2 \cap E_1) = (0.25)(0.35) = 0.0875$

$P(E_2 \cap E_1) = (0.25)(0.75) = \dfrac{0.1625}{1.00}$

$P(E_1 \cap E_1) = (0.35)\,(0.75) = 0.2625$

$P(E_1 \cap E_2) = (0.35)(0.25) = 0.0875$

$P(E_2 \cap E_1) = (0.65)(0.35) = 0.2275$

$P(E_2 \cap E_2) = (0.65)(0.65) = \dfrac{0.4225}{1.00}$

c) La probabilidad de dos etapas:

i) de E_1 a E_1 a E_1 es $P_{11}^{(2)} = (0.75)(0.75) + (0.25)(0.35) = 0.5625 + 00.875$
$= 0.65$

ii) de E_1 a E_2 es $P_{12}^{(2)} = (0.75)(0.25) + (0.25)(0.65) = 0.1875 + 0.1625$
$= \dfrac{0.35}{1.00}$

iii) de E_2 a E_1 es $P_{21}^{(2)} = (0.35)(0.75) + (0.65)(0.35) = 0.2625 + 0.2275$
$= 0.49$

iv) de E_2 a E_2 es $P_{22}^{(2)} = (0.35)(0.25) + (0.65)(0.65) = 0.0875 + 0.4225$
$= \dfrac{0.51}{1.00}$

Entonces:

$$(0.50,\ 0.50)\begin{pmatrix} 0.65 & 0.35 \\ 0.49 & 0.51 \end{pmatrix} =$$

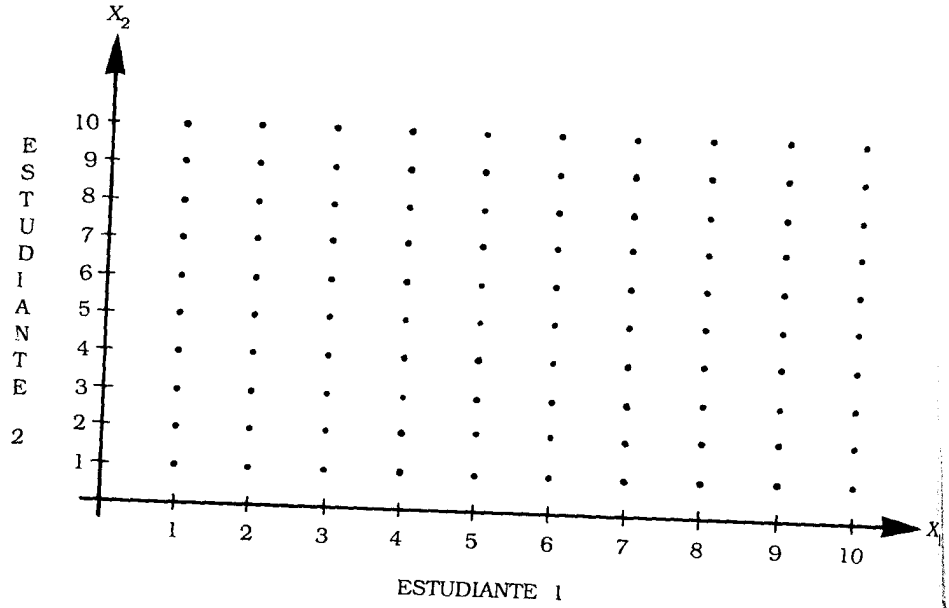
$$A^{(0)} \quad P^2 = \begin{array}{cc} E_1 & E_2 \\ \end{array}$$
$$\begin{pmatrix} 0.57 & 0.43 \end{pmatrix}$$

5.38 *a)* T^2
b) T^5 utilizar MACSTAT.

5.39 *a)* 0.1
b) 0.02
c) 0.20
d) 0.1

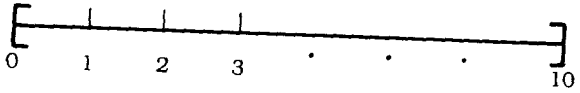
Capítulo 6

6.1 *a)*



La variable aleatoria es $\overline{X}$

b)



La variable aleatoria $\overline{X}$ se encuentra en el intervalo de cero a 10, de los números reales.

6.2 *a)* Continua
b) Continua
c) Discreta
d) Continua
e) Discreta

6.3 $P[|X-\mu| \le K\sigma_X] \ge 1 - \dfrac{1}{K^2}$, $\mu = 8$ $\sigma = 3$

$$\begin{aligned} P[-4 \le X \le 20] &= P[-4-\mu \le X-\mu \le 20-\mu] \\ &= P[-4-8 \le X-\mu \le 20-8] \\ &= P[-12 \le X-\mu \le 12] \\ &= P[|X-\mu| \le 12] \\ &= P[|X-\mu| \le K\sigma] \end{aligned}$$

$$= P[\,|X-\mu| \le 4\,(3)] \ge 1 - \frac{1}{16} = \frac{15}{16}$$

$$\frac{1}{K^2} = \frac{1}{16}$$

6.4 $P = 0.3$
$n = 16$ $\quad P(K = 4) = {}_{16}C_{4}\,(0.3)^4\,(0.7)^{12} = 0.2040$ aprox. 20.4%
$K = 4$

6.5 $P = 0.2$
$n = 20$ $\quad P(K = 5) = {}_{20}C_{5}\,(0.2)^5\,(0.8)^{15} = 0.1746$ aprox. 17.5%
$K = 5$

6.6 $P = 0.7$
$n = 16$ $\quad P(X = 9) = {}_{16}C_{9}\,(0.7)^9\,(0.3)^7 = 0.101$ aprox. 10%
$X = 9$

6.7 Como
$n = 300$
$P = 0.02$
$X = 4$

al aplicar
$P(X = 4) = {}_{300}C_{4}\,(0.02)^4\,(0.98)^{296}$

Resulta muy laborioso, tardado y con la posibilidad de cometer varios errores.

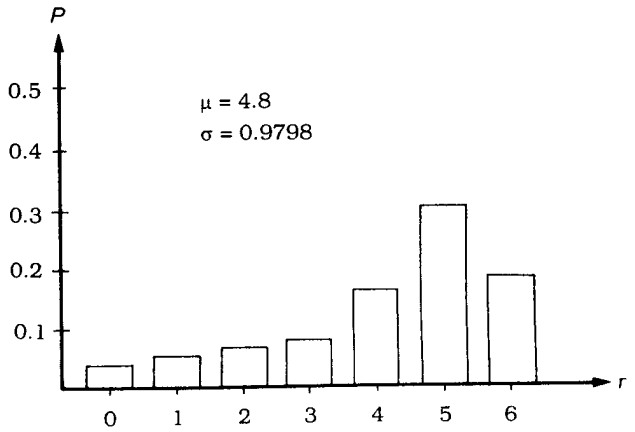
Utilizando la aproximación de Poisson, se tiene:

$$\mu = nP = 300\,(0.02) = 6$$

$$P(X = 4) = \frac{\mu^x\, e^{-\mu}}{x\,!} = \frac{6^4\, e^{-6}}{4\,!} = 0.135 \text{ aprox. } 13.5\%$$

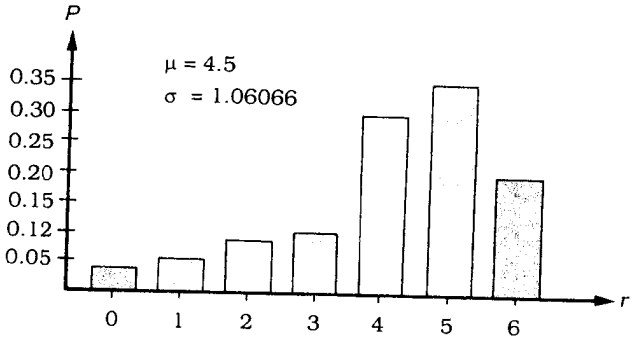
6.8

$P(r = 0)$	=	0.000064
$P(r = 1)$	=	0.001536
$P(r = 2)$	=	0.01536
$P(r = 3)$	=	0.08192
$P(r = 4)$	=	0.24576
$P(r = 5)$	=	0.393216
$P(r = 6)$	=	0.262144
		1.0000

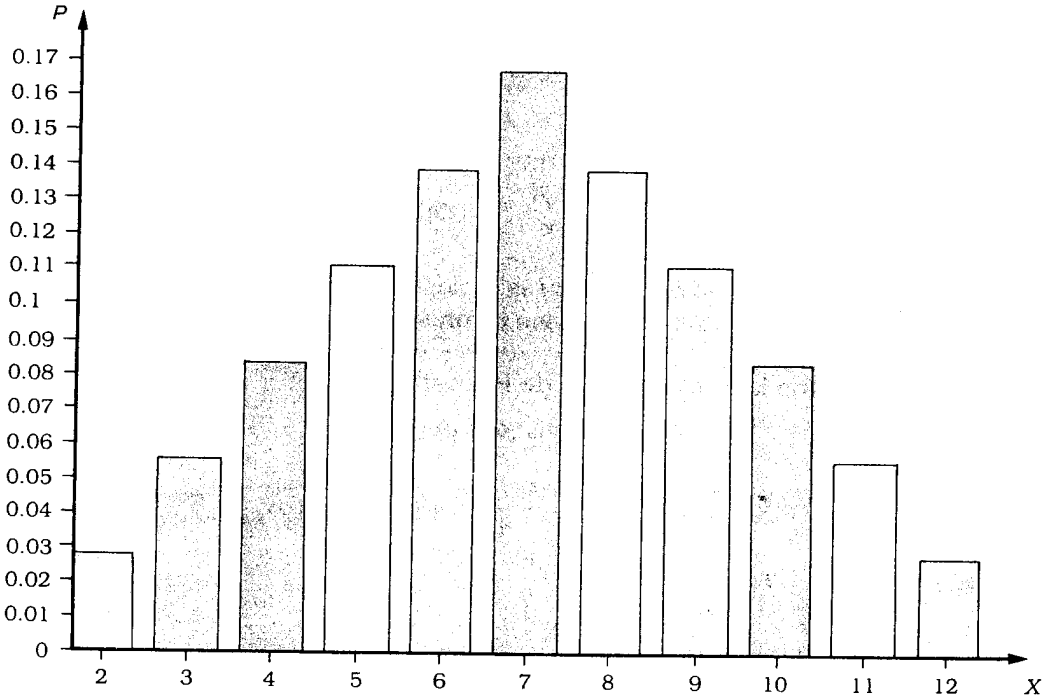


6.9

$P(r = 0)$	=	0.00024414
$P(r = 1)$	=	0.004392
$P(r = 2)$	=	0.032956
$P(r = 3)$	=	0.1318125
$P(r = 4)$	=	0.296625
$P(r = 5)$	=	0.35595
$P(r = 6)$	=	0.17798
		1.0000



6.10 *c)*



a)

X = 2	3	4	5	6	7	8	9	10	11	12
(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)	(2, 6)	(3, 6)	(4, 6)	(5, 6)	(6, 6)
	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(3, 5)	(4, 5)	(5, 5)	(6, 5)	
		(3, 1)	(3, 2)	(3, 3)	(3, 4)	(4, 4)	(5, 4)	(6, 4)		
			(4, 1)	(4, 2)	(4, 3)	(5, 3)	(6, 3)			
				(5, 1)	(5, 2)	(6, 2)				
					(6, 1)					

$$P = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = 1.0$$

b) $\mu = 0.30556 \qquad \sigma^2 = 0.29707 \qquad \sigma = 0.54504$

6.11 $n = 15$
$X = 0, 1, \ldots, 15$
$P = 0.88$
$q = 0.12$

a) Menos de 6 tengan que repetir el curso

$$P(X < 6) = P\left(\sum_{X=0}^{5} P(X, 15, 0.12)\right) = 0.98602$$

b) Exactamente 10 aprueben el curso

$$P(X = 10) = P\left(C_{15\ 10}\,(0.88)^{10}\,(0.12)^{5}\right) = 0.02081077$$

c) Más de 12 aprueben el curso

$$P(X > 12) = P\left(\sum_{X=13}^{15} P(X, 15, 0.12)\right) = 0.734565$$

d) $\mu = np = 13.20$

$\boxed{\mu = 13.20}$ $\boxed{\sigma = 1.25857}$

6.12 $n = 13$
$P = 0.44$
$q = 0.56$
$X = 0, 1, \ldots, 13$

a) $P(X \geq 6) = 0.5440572$

b) $P(X \leq 7) = 0.82518$

c) $\boxed{\mu = 5.72}$ $\boxed{\sigma = 1.78975}$

6.13 $n = 1\,000, \quad P = 0.005, \quad \lambda = 5$

a) $\boxed{P(X = 0) = 0.00674}$

b) $\boxed{P(X = 10) = 0.01814}$

c) $\boxed{P(X = 15) = 0.000157}$

6.14 $n = 200, \quad P = 0.03, \quad \lambda = 6$

a) $\boxed{P(X = 10) = 0.0413030}$

b) $\boxed{P(X = 5) = 0.160623}$

6.15 $n = 10\,000, \quad P = 0.001, \quad \mu = \lambda = 10, \quad X = 0, 1, 2, \ldots, 10\,000$

a) $P(6 \leq X \leq 8) = 0.2657$

b) $\sigma = 3.16228$

6.16 $n = 2\,000, \quad P = 0.002$

a) $P(X < 5) = 0.62893$

b) $\mu = 4, \quad \sigma = 2$

6.17 $n_1 = 60$

$n_2 = 40$

$X = 4$ de 75 watts

$r = 10$

$r - X = 10 - 4 = 6$ de 100 watts

$\begin{bmatrix} n_1 \\ X \end{bmatrix} \quad \begin{bmatrix} n_2 \\ r - X \end{bmatrix}$

son las formas de elegir $r - X = 6$ focos de 100 watts.

$$\therefore \binom{n_1}{X}\binom{n_2}{r - X} \Rightarrow \binom{n}{r}\binom{100}{10}$$

$$f(X) = \frac{\binom{n_1}{X}\binom{n_2}{r - X}}{\binom{n}{r}} = \frac{\binom{60}{4}\binom{40}{10 - 4}}{\binom{100}{10}} = 0.108 \therefore \text{ la probabilidad es } 0.108$$

6.18 $$f(X) = \frac{\binom{n_1}{X}\binom{n_2}{r - X}}{\binom{n_1 + n_2}{r}} = \frac{\binom{6}{3}\binom{4}{5 - 3}}{\binom{6 + 4}{5}} = 0.48$$

$$\therefore \boxed{\text{Probabilidad} = 0.48}$$

6.19 *a)* $Z = -1.97$, $Z = -0.76$, $Z = 0.36$
$Z = 1.21$, $Z = 1.86$, $Z = 2.33$ respectivamente
b) $X = -4.5$, $X = -3.2$, $X = -2.0$
$X = -1.1$, $X = -0.4$, $X = 0.1$ respectivamente

6.20 *a)* 10.64% *d)* 34.13% *g)* 28.81%
b) 49.32% *e)* 24.22% *h)* 49.93%
c) 49.87% *f)* 1.99%

6.21 *a)* 2.28% *c)* 9.18%
b) 0.04% *d)* 49.08%

6.22 *a)* 68.26% *c)* 66.78% *e)* 28.53%
b) 95.44% *d)* 12.10%

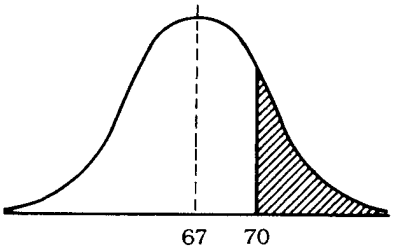
6.23 *a)* 29.57% *c)* 25.18% *e)* 4.56%
b) 29.57% *d)* 76.53%

6.24 *a)* 15.87% *d)* 50% *g)* 44.95%
b) 2.28% *e)* 39.97%
c) 0.13% *f)* 89.97%

6.25 *a)* $Z = 0.5$ *c)* $Z = -0.83$
b) $Z = 1.16$ *d)* $Z = -1.5$

6.26 *a)* $Z = 1.93$ *d)* $Z = 0.71$ *g)* $Z = 2.33$
b) $Z = -0.36$ *e)* $Z = 0.92$ *h)* $Z = 2.58$
c) $Z = -1.74$ *f)* $Z = 1.96$

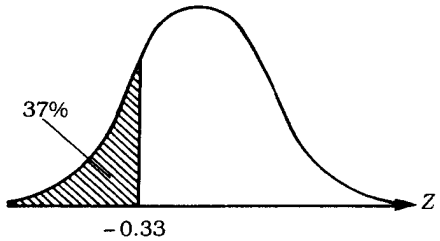
6.27



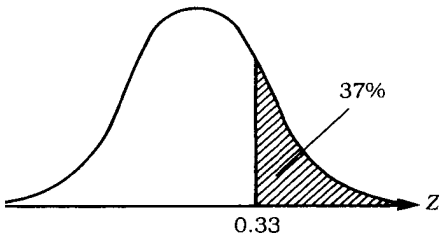
$P(X \geq 70) = P(Z \geq 0.25) = 0.40$
aproximadamente 40%
(40 profesores)

6.28 *a)* 14 injertos
b) 7 injertos
c) 205 injertos

6.29 *a)* $P(X \leq 5) \approx P(X \leq 5.5) = P\left(\frac{X - \mu}{\sigma} \leq \frac{5.5 - 6}{1.5}\right)$; $P(Z \leq -0.33) = 37\%$



b) $P(X \geq 6.5) = P\left(Z \geq \frac{0.5}{1.5}\right) = P(Z \geq 0.33) = 37\%$



6.30 Utilizando la aproximación normal de la binomial

$$n = 80 \qquad \mu = \mu P = 80\left(\frac{1}{4}\right) = 20; \qquad \boxed{\mu = 20}$$

$$P = \frac{1}{4} \qquad \sigma = \sqrt{npq} = \sqrt{80\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)} = \sqrt{15} = 3.87 \qquad \boxed{\sigma = 3.87}$$

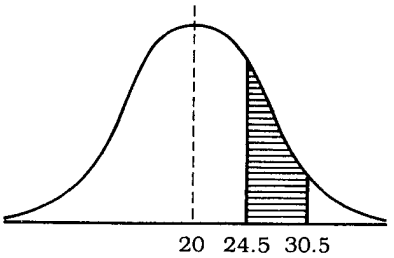
Como se desea calcular el intervalo $P(25 \leq X \leq 30)$ aplicando la corrección de continuidad $P(24.5 \leq X \leq 30.5)$.

Por lo que:

$$P\left(\frac{24.5 - 20}{3.87} \leq \frac{X - 20}{3.87} \leq \frac{30.5 - 20}{3.87}\right)$$

$$P(1.163 \leq Z \leq 2.713) = 0.9966 - 0.8776 = 0.1190$$

$$\boxed{P(24.5 \leq X \leq 30.5) = 0.1190}$$



6.31 $P(X \leq 1000) = F(1000) = 1 - e^{-\frac{1000}{1000}} = 1 - e^{-1}$

$$= 1 - \frac{1}{e} = 0.6321$$

$$\boxed{P(X \leq 1000) = 0.6321}$$

Capítulo 7

7.1 La principal ventaja del *muestreo aleatorio* es que no solamente provee una estimación adecuada del parámetro desconocido, sino que en adición ayuda al investigador para construir una afirmación que indique el grado de confianza de que dicho parámetro estimado es el correcto.

7.2 *a)* *Insesgado*:
Un estimador es insesgado cuando el valor esperado del estadístico muestral (el estimador) es igual al parámetro que se está estimando.

b) *Eficiente*:
La eficiencia puede ser definida tanto considerando el tamaño de la muestra como el costo que implica dicha muestra. Esto significa que la muestra estimada debe concentrarse lo más cercanamente posible al parámetro que se está estimando, para un tamaño determinado de muestra y con el menor costo económico.

c) *Consistente*:
Un estimador se considera consistente si la muestra estimada se concentra más y más cercanamente alrededor del parámetro que se está estimando, a medida que el tamaño de la muestra se incrementa.

7.3 *Error de muestreo* se refiere a la diferencia entre el valor de la media muestral y la media poblacional, que puede atribuirse al hecho de que uno de los cálculos se basa en una muestra extraída de la población objetivo, mientras que la otra se considera del universo.

7.4 *Estrato*

a) menos de 60 000

Determinación de la proporción a seleccionarse

$$\frac{n_1\,\sigma_1}{n_1\,\sigma_1 + n_2\,\sigma_2 + n_3\,\sigma_3 + n_4\,\sigma_4} =$$

$$= \frac{(0.2)\,(K)}{(0.2)\,(K) + (0.5)\,(1.4K) + (0.2)\,(2K) + (0.1)(3K)}$$

$$= \frac{0.2\,K}{0.2\,K + 0.7\,K + 0.4\,K + 0.3\,K}$$

$$= \frac{2\,K}{16\,K} = \frac{1}{8}$$

Por lo que la proporción a seleccionarse del estrato *a)* menos de 60 000, es de $\frac{1}{8}$ de 160, o sea, 20 personas.

b) 60 000 —— 100 000

$$\frac{n_2\,\sigma_2}{n_1\,\sigma_1 + n_2\,\sigma_2 + n_3\,\sigma_3 + n_4\,\sigma_4} = \frac{7\,K}{16\,K} = 0.4375 \approx 43.75\%$$

aproximadamente 70 personas.

c) 100 000 —— 140 000

$$\frac{n_3\,\sigma_3}{n_1\,\sigma_1 + n_2\,\sigma_2 + n_3\,\sigma_3 + n_4\,\sigma_4} = \frac{4\,K}{16\,K} = 0.25 \approx 25\%$$

o sea 40 personas.

d) 140 000 o más

$$\frac{n_4\,\sigma_4}{n_1\,\sigma_1 + n_2\,\sigma_2 + n_3\,\sigma_3 + n_4\,\sigma_4} = \frac{3\,K}{16\,K} = 0.1875 \approx 18.75\%$$

o sea 30 personas.

Conclusión:
El conocimiento de las características de gastos de diversión de varios estratos, podría identificar las características de ingreso que se deben buscar para asegurar el éxito de dicho centro.

7.5 *a)* Supóngase que todas las muestras de un tamaño específico de un universo o población, se determinan y el valor estadístico descriptivo para cada una de ellas (la media aritmética puede cumplir este papel). Si todos estos estadísticos muestrales se agrupan en una distribución probabilística, la distribución resultante sería una distribución muestral del estadístico, que enlista los diferentes valores que puede tomar.

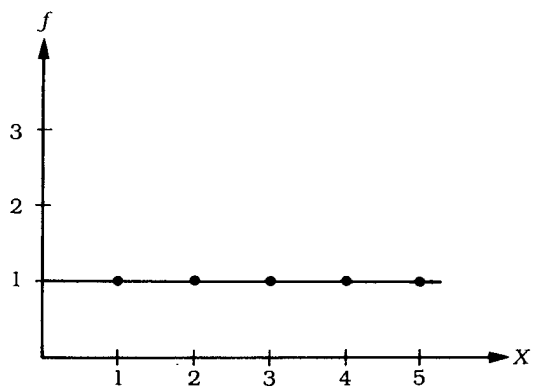
b) El error estándar es una medida de dispersión que describe la variabilidad de los valores de un estadístico muestral, debido a la ocurrencia del "error de muestreo".

7.6 $$C_{1000\ 400} = \frac{1000!}{400!\ 600!}$$

$$C_{1000\ 600} = \frac{1000!}{600!\ 400!}$$

Por lo que son iguales.

7.7 *a)*



b) $\mu_x = \frac{15}{5} = 3 \therefore \boxed{\mu_x = 3}$

c) $\sigma_x^2 = \frac{10}{5} = 2 \therefore \boxed{\sigma_x = \sqrt{2}}$

d) $C_{N\ n} = C_{5\ 2} = 10$

e)

CASO	X_1, X_2	$\overline{X}$
1	1, 2	1.5
2	1, 3	2.0
3	1, 4	2.5
4	1, 5	3.0
5	2, 3	2.5
6	2, 4	3.0
7	2, 5	3.5
8	3, 4	3.5
9	3, 5	4.0
10	4, 5	4.5

f) $$\mu_{\bar{x}} = \frac{1.5 + 2.0 + 2.5 + \ldots + 4.5}{10} = 3.0$$

$$\therefore \boxed{\mu_{\bar{x}} = \mu_x}$$

g)

$\overline{X}$	$\overline{X} - \mu$	$(\overline{X} - \mu_x)^2$
1.5	−1.5	2.25
2.0	−1.0	1.00
2.5	−0.5	0.25
3.0	0.0	0.00
2.5	−0.5	0.25
3.0	0.0	0.00
3.5	0.5	0.25
3.5	0.5	0.25
4.0	1.0	1.00
4.5	1.5	2.25
		7.50

o sea $\Sigma(\overline{X} - \mu_x)^2 = 7.50$

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma\left(\overline{X} - \mu_x\right)^2}{\underset{N\ n}{C}}} = \sqrt{\frac{7.5}{10}} = \sqrt{0.75} = 0.8660$$

$$\boxed{\sigma_{\bar{x}} = 0.8660}$$

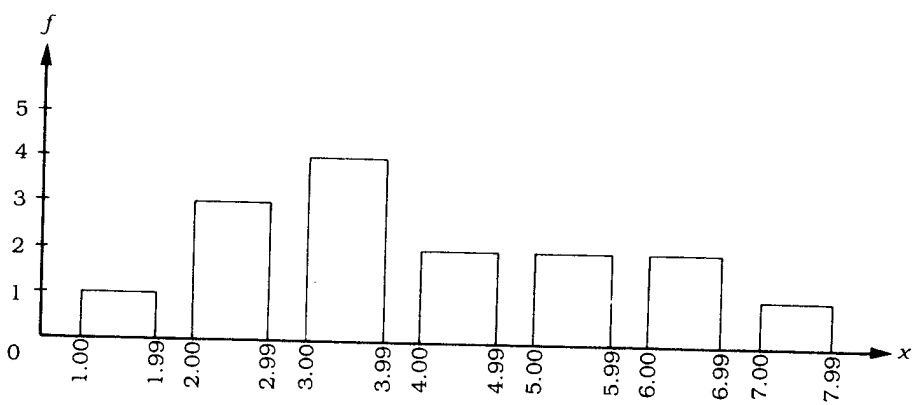
h) $$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \frac{\sqrt{2}}{\sqrt{2}} \times \sqrt{\frac{3}{4}} = \sqrt{0.75} = 0.8660$$

$$\therefore \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

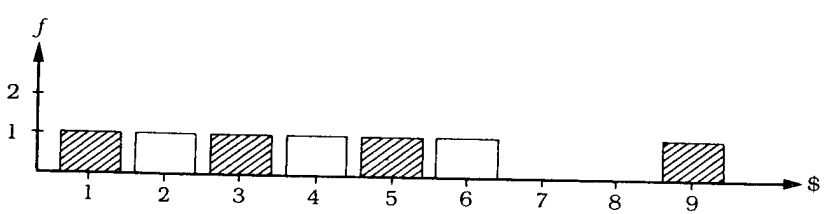
i)

Li – Ls	*f*
1.00 – 1.99	1
2.00 – 2.99	3
3.00 – 3.99	4
4.00 – 4.99	2
5.00 – 5.99	2
6.00 – 6.99	2
7.00 – 7.99	1
	15

j)



7.8 *a)*



b) $\mu_x = \dfrac{2+5+3+4+1+9}{6} = 4$ $\boxed{\mu_x = 4}$

c) $\sigma_x^2 = \dfrac{\sum_{i=1}^{6}\left(\overline{X}_i - \mu_x\right)^2}{N} = \dfrac{40}{6} = 6.667$

$\boxed{\sigma_x^2 = 6.667}$ $\boxed{\sigma_x = 2.58}$

d) $C_{6\,2} = \dfrac{6\,!}{2\,!\,(6-2)\,!} = 15$ maneras diferentes

e)

Muestra	$\overline{X}$	Muestra	$\overline{X}$	Muestra	$\overline{X}$	Muestra	$\overline{X}$	Muestra	$\overline{X}$
1, 2	1.5	2, 3	2.5	3, 4	3.5	4, 5	4.5	5, 9	7.0
1, 3	2.0	2, 4	3.0	3, 5	4.0	4, 9	6.5		
1, 4	2.5	2, 5	3.5	3, 9	6.0				
1, 5	3.0	2, 9	5.5						
1, 9	5.0								

f) $\mu_{\overline{x}} = \dfrac{1.5 + 2.0 + \ldots + 6.5 + 7.0}{15} = \dfrac{60}{15} = 4$

$\boxed{\mu_{\overline{x}} = 4}$

g)

$\overline{X}_i$	$\overline{X}_i - \mu_x$	$(\overline{X}_i - \mu_x)^2$
1.5	-2.5	6.25
2.0	-2.0	4.00
2.5	-1.5	2.25
2.5	-1.5	2.25
3.0	-1.0	1.00
3.0	-1.0	1.00
3.5	-0.5	0.25
3.5	-0.5	0.25
4.0	0.0	0.0
4.5	0.5	0.25
5.0	1.0	1.00
5.5	1.5	2.25
6.0	2.0	4.00
6.5	2.5	6.25
7.0	3.0	9.00

$\therefore \Sigma\,(\overline{X}_i - \mu_x)^2 = 40$

h) $$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{15}\left(X_i - \mu_x\right)^2}{N}} = \sqrt{\frac{40}{15}} = 1.633$$

$$\boxed{\sigma_{\bar{x}} = 1.633}$$

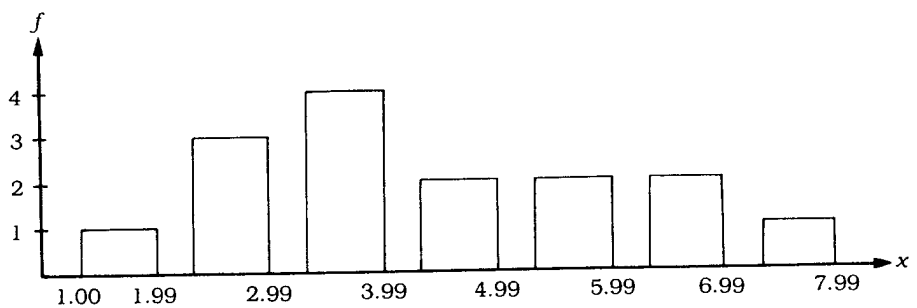
i) $$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{\frac{20}{3}}{2}} \sqrt{\frac{4}{5}} = 1.633$$

$$\boxed{\sigma_{\bar{x}} = 1.633}$$

j)

Li - Ls	*f*
1.00 - 1.99	1
2.00 - 2.99	3
3.00 - 3.99	4
4.00 - 4.99	2
5.00 - 5.99	2
6.00 - 6.99	2
7.00 - 7.99	1
	15

k)



7.9 $\mu_{\bar{x}} = 3$

$\sigma_{\bar{x}} = 0.866$

$$P[\bar{X} \geq 4] = P\left[\frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \geq \frac{4 - \mu_{\bar{x}}}{\sigma_{\bar{x}}}\right]$$

$$= P\left[Z \geq \frac{4-3}{0.866}\right] = P[Z \geq 1.15] = 0.1251$$

$$\therefore \boxed{P(\bar{X} \geq 4) = 12.5\%}$$

7.10 $\mu_{\bar{x}} = 4$

$\sigma_{\bar{x}} = 1.633$

$$P[\bar{X} \geq 6] = P\left[Z \geq \frac{6-4}{0.1.633}\right] = P[Z \geq 1.22] = 0.1112$$

$$\therefore \boxed{P(\bar{X} \geq 6) = 11.12\%}$$

7.11 *a)*

Muestra	$\bar{X}$	Muestra	$\bar{X}$	Muestra	$\bar{X}$
1, 1	1.0	2, 1	1.5	6, 1	3.5
1, 2	1.5	2, 2	2.0	6, 2	4.0
1, 6	3.5	2, 6	4.0	6, 6	6.0

b) $$\mu_{\bar{x}} = \frac{2}{2} \times \frac{1}{9} + \frac{3}{2} \times \frac{1}{9} + \frac{7}{2} \times \frac{1}{9} + \frac{3}{2} \times \frac{1}{9} + \frac{4}{2} \times \frac{1}{9}$$

$$+ \frac{8}{2} \times \frac{1}{9} + \frac{7}{2} \times \frac{1}{9} + \frac{8}{2} \times \frac{1}{9} + \frac{12}{2} \times \frac{1}{9} = \frac{54}{18} = 3 \quad \therefore \boxed{\mu_{\bar{x}} = 3}$$

c) $$\sigma^2_{\bar{x}} = \frac{\sum_{i=1}^{9} (\bar{X}_i - \mu_x)^2}{9} = \frac{21}{9} = \frac{7}{3}; \qquad \boxed{\sigma^2_{\bar{x}} = \text{Var}(\bar{X}) = \frac{7}{3}}$$

$\bar{X}$	$\bar{X} - \mu_x$	$(\bar{X} - \mu_x)^2$
1.0	−2.0	4.0
1.5	−1.5	2.25
1.5	−1.5	2.25
2.0	−1.0	1.00
3.5	0.5	0.25
3.5	0.5	0.25
4.0	1.0	1.00
4.0	1.0	1.00
6.0	3.0	9.00
		21.0

d) $E(\overline{X} - \mu_x)^2 = Var\,(\overline{X}) = \sigma^2_{\overline{x}} = \dfrac{\sum_{i=1}^{6} \left(\overline{X}_i - \mu_{\overline{x}}\right)^2}{6} = \dfrac{7}{6}$

$\overline{X}$	$\overline{X} - \mu_x$	$(\overline{X}_i - \mu_x)^2$
1.5	– 1.5	2.25
2.0	– 1.0	1.00
2.5	– 0.5	0.25
3.5	0.5	0.25
4.0	1.0	1.00
4.5	1.5	2.25
		7.0

7.12 *a)*

Muestra	$\overline{X}$	Muestra	$\overline{X}$	Muestra	$\overline{X}$	Muestra	$\overline{X}$
1, 1	1.0	2, 1	1.5	3, 1	2.0	6, 1	3.5
1, 2	1.5	2, 2	2.0	3, 2	2.5	6, 2	4.0
1, 3	2.0	2, 3	2.5	3, 3	3.0	6, 3	4.5
1, 6	3.5	2, 6	4.0	3, 6	4.5	6, 6	6.0

b) $\mu_{\overline{x}} = \dfrac{2}{2} \times \dfrac{1}{16} + \dfrac{3}{2} \times \dfrac{1}{16} \times 2 + \dfrac{4}{2} \times \dfrac{1}{16} \times 3 + \dfrac{5}{2} \times \dfrac{1}{16} \times 2 + \dfrac{6}{2} \times \dfrac{1}{16}$

$+ \dfrac{7}{2} \times \dfrac{1}{16} \times 2 + \dfrac{8}{2} \times \dfrac{1}{16} \times 2 + \dfrac{9}{2} \times \dfrac{1}{16} \times 2 + \dfrac{12}{2} \times \dfrac{1}{16} =$

$= \dfrac{1}{32} \times 96 = 3 \quad \therefore \quad \boxed{\mu_{\overline{x}} = 3}$

c) $\sigma^2_{\overline{x}} = \dfrac{\sum_{i=1}^{16} \left(\overline{X}_i - \mu_{\overline{x}}\right)^2}{16} = \dfrac{28}{16} = 1.75\,; \quad \boxed{\sigma^2_{\overline{x}} = 1.75}$

d) $\sigma^2_{\overline{x}} = \dfrac{14}{4} \quad \therefore \quad \dfrac{\sigma^2_x}{n} = \dfrac{\frac{14}{4}}{2} = \dfrac{7}{4} = 1.75$

$\therefore \quad \boxed{\sigma^2_{\overline{x}} = \dfrac{N - n}{n - 1} \times \dfrac{\sigma^2_x}{n}}$

7.13 $$\sigma_x^2 = \frac{\sum_{i=1}^{4}\left(\overline{X}_i - \mu_x\right)^2}{4} = \frac{26}{4} = \frac{13}{2}: \quad \boxed{\sigma_x^2 = 6.5}$$

$$\mu_x = \frac{1+3+4+8}{4} = \frac{16}{4}; \quad \boxed{\mu_x = 4}$$

a) Sin reemplazo:

$$\sigma_{\bar{x}}^2 = \frac{N-n}{N-1} \times \frac{\sigma_x^2}{n} = \frac{4-2}{4-3} \times \frac{\frac{13}{2}}{2} = \frac{13}{2}$$

$$\therefore \quad \boxed{\sigma_{\bar{x}}^2 = 6.5}$$

b) Con reemplazo:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} = \frac{\frac{13}{2}}{2} = \frac{13}{4} \quad \therefore \quad \boxed{\sigma_{\bar{x}}^2 = 3.25}$$

7.14 Sean: $X_1 = 1,\ X_2 = 2,\ X_3 = 3,\ X_4 = 4,\ X_5 = 5$

$$\mu_x = \frac{1+2+3+4+5}{5} = 3; \quad \boxed{\mu_x = 3}$$

$$\sigma_x^2 = \text{Var}(X) = \frac{\sum_{i=1}^{5}\left(\overline{X}_i - \mu_x\right)^2}{5} = \frac{10}{5} = 2; \quad \boxed{\sigma_x^2 = 2}$$

a) $$\overline{X}_1 = \frac{X_1 + X_2 + X_3}{3} = 2; \quad S_1^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = \frac{2}{3}$$

$$\overline{X}_2 = \frac{X_1 + X_2 + X_4}{3} = \frac{7}{3}; \quad S_2^2 = \frac{\left(1-\frac{7}{3}\right)^2 + \left(2-\frac{7}{3}\right)^2 + \left(4-\frac{7}{3}\right)^2}{3} = \frac{42}{27}$$

$$\overline{X}_3 = \frac{X_1 + X_2 + X_5}{3} = \frac{8}{3}; \quad S_3^2 = \frac{\left(1-\frac{8}{3}\right)^2 + \left(2-\frac{8}{3}\right)^2 + \left(5-\frac{8}{3}\right)^2}{3} = \frac{88}{27}$$

$$\overline{X}_4 = \frac{X_1 + X_2 + X_4}{3} = \frac{8}{3}; \quad S_4^2 = \frac{\left(1-\frac{8}{3}\right)^2 + \left(3-\frac{8}{3}\right)^2 + \left(4-\frac{8}{3}\right)^2}{3} = \frac{42}{27}$$

Obteniendo en la misma forma

$$\overline{X}_5 = \frac{9}{3} \; ; \quad S_5^2 = \frac{8}{3}$$

$$\overline{X}_6 = \frac{10}{3} \; ; \quad S_6^2 = \frac{78}{27}$$

$$\overline{X}_7 = 3 \; ; \quad S_7^2 = \frac{2}{3}$$

$$\overline{X}_8 = \frac{10}{3} \; ; \quad S_8^2 = \frac{42}{27}$$

$$\overline{X}_9 = \frac{11}{3} \; ; \quad S_9^2 = \frac{32}{27}$$

$$\overline{X}_{10} = 4 \; ; \quad S_{10}^2 = \frac{2}{3}$$

$$E\left(\frac{n}{n-1} \quad S^2\right) = \frac{3}{2} \times \frac{1}{10} \quad (S_1^2 + S_2^2 + \ldots + S_{10}^2) = \frac{5}{2}$$

$$\frac{N}{N-1}\,\sigma_x^2 = \frac{5}{4} \times 2 = \frac{5}{2} \qquad \therefore \quad \boxed{E\left(\frac{n}{n-1} \quad S^2\right) = \frac{N}{N-1} \quad \sigma_x^2}$$

7.15 *a)* $\boxed{\mu_x = 4.8}$

b) Para una muestra $n = 6$, sin reemplazo, por ejemplo: 4, 9, 5, 1, 3, 8; $\boxed{\overline{X}_b = 5}$

c) Para una muestra de tamaño $n = 6$, con reemplazo,

por ejemplo: 5, 3, 7, 3, 8, 7 ; $\boxed{\overline{X}_c = 5.5}$

d) $\sigma_x^2 = \dfrac{73.4}{10} = 7.34$; $\boxed{\sigma_x^2 = 7.34}$

e) $S_b^2 = \dfrac{46}{6}$; $\boxed{S_b^2 = 7.667}$

f) $S_c^2 = \dfrac{19.5}{6}$; $\boxed{S_c^2 = 3.23}$

Capítulo 8

8.1 *a)* Verdadero
b) Verdadero
c) Falso
d) Verdadero
e) Verdadero

8.2 *a)* Lo que el investigador está interesado en probar.

8.3 *d)* La región de rechazo de *Ho* consiste en aquellos valores de + > 3.5 o − < 3.5.

8.4 *a)* Cuando el porcentaje es diferente al 60%.
b) Cuando es el 60% el que está en desacuerdo.

8.5 *a)* $X = 15, 16, 17, 18, 19$ y 20
b) $X = 0, 1, 2, 3 \ldots, 14$

8.6 *c)*

Capítulo 9

9.1 *a)* $P(X \geq 115) = 0.0668 \approx 6.7\%$
b) $P(X \geq 125) = 0.0062 \approx 0.62\%$
c) $P(X \leq 84.5) = 0.0606 \approx 6.06\%$
d) $P(88 \leq X \leq 94) = 0.1592 \approx 16\%$
e) $P(97.5 \leq X \leq 107.8) = 0.3810 \approx 38\%$
f) $P(113 \leq X \leq 116) = 0.042 \approx 4.2\%$

9.2 *a)* $P(X \geq 115) = 0.1587 \approx 16\%$
b) $P(X \leq 113) = 0.0808 \approx 8.1\%$
c) $P(-10.5 \leq X \leq -13.2) = 0.6247 \approx 62.5\%$
d) $P(-1.96 \geq X \geq 1.96) = 0.05 \approx 5\%$
e) $P(\mu - \sigma \geq X \geq \mu + \sigma) = 0.3174 \approx 32\%$
f) $P(\mu - 3\sigma \geq X \geq \mu + 3\sigma) = 0.0026 \approx 0.26\%$

9.3 *a)* 15.87%
b) $X = 24.935$
c) $\sigma = 1.3$

9.4 $n = 36$

$\overline{X} = 40$ $\quad 40 - 1.96 \times \dfrac{12}{\sqrt{36}} \leq \mu_x \leq 40 + 1.96 \times \dfrac{12}{\sqrt{36}}$

$\sigma_x = 12$

$Z_{0.025} = \pm 1.96$ $\quad 36.08 \leq \mu_x \leq 43.92$

$36 \leq \mu_x \leq 44$ al 95%

9.5 $n = 64$
$\overline{X} = 600$ $\quad 290 \leq \mu_x \leq 310$ $\quad$ al 95%
$\sigma_x = 40$
$Z_{0.025} = \pm 1.96$

9.6 $Ho: \mu_x = 50$ cal.
$H_i: \mu_x < 50$ cal.
$n = 36$
$\sigma^2 = 9$ $\quad$ Entonces $\quad \sigma_{\overline{x}} = \dfrac{3}{\sqrt{36}} = 0.5$
$\overline{X} = 49.3$ cal.
$\alpha = 5\%$
$Z_{0.05} = \pm 1.64$ $\quad Z_{calc} = \dfrac{\overline{X} - \mu_x}{\sigma_{\overline{x}}} = \dfrac{49.3 - 50}{0.5} = 1.4$
Como $Z_{calc} \not> 1.64$ $\quad$ *Ho no* se rechaza.

Conclusión: Los datos apoyan la propaganda acerca del contenido en calorías.

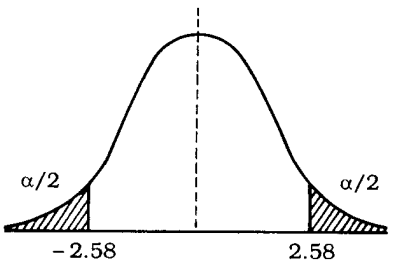
9.7 $P\left(20 - \dfrac{4.94}{\sqrt{n}} < \mu_x < 20 + \dfrac{4.94}{\sqrt{n}}\right)$

9.8 $P(4.86 < \mu_x < 5.14)$

$Ho: \mu_x = 1$

$H_i: \mu_x \neq 1$

9.9 $Z = \dfrac{\overline{x} - \mu_x}{\sigma_x / \sqrt{n}} = \dfrac{1.014 - 1}{0.02 / \sqrt{25}} = 3.5$



Como $3.5 > 2.58$ *Ho* se rechaza con un $\alpha = 1\%$. La máquina no está bajo control.

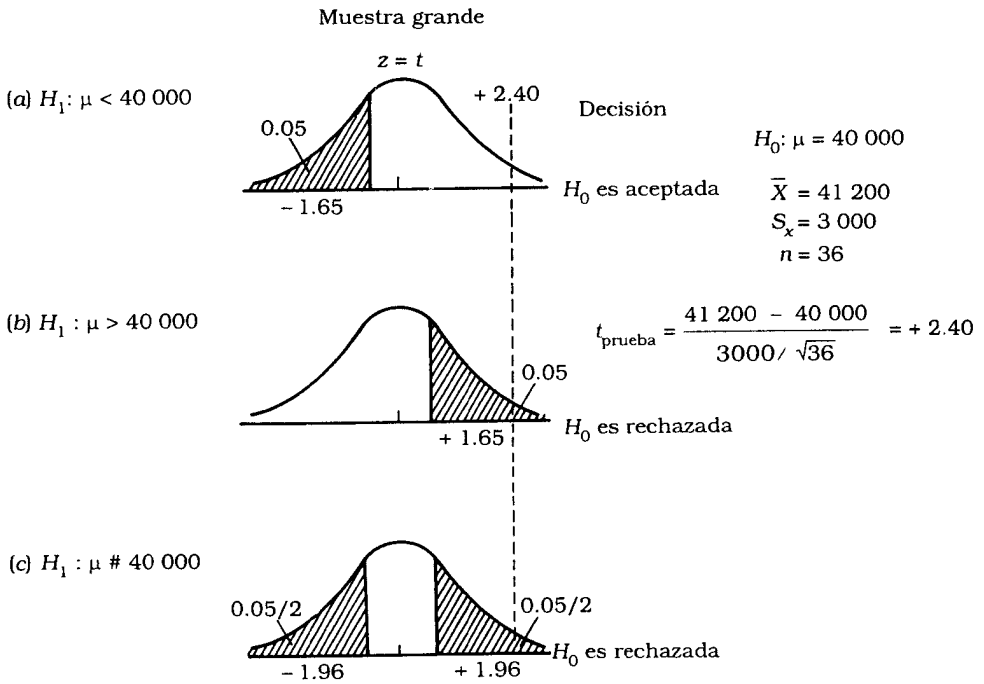
9.10 1. $H_1: \mu_x < 40\,000$.
2. $H_1: \mu_x > 40\,000$.
3. $H_1: \mu_x \neq 40\,000$.

9.11

Muestra grande: Supóngase que se toma una muestra de 36 observaciones, y que la media de la muestra resultante es 41 200 y la desviación estándar de 3 000. Considérese también $\alpha = 0.05$. Ahora, como n es mayor que 30, el valor de z, tomado a partir de una tabla normal, puede utilizarse para aproximar el valor de t. De ahí que el valor crítico sea + 1.65, o bien, − 1.65 en el caso de una prueba de una cola, y ± 1.96 en el de una de dos colas. Para cualquier caso, el valor estadístico de prueba es:

$$t_{prueba} = \frac{\overline{X} - \mu_0}{S_x / \sqrt{n}} = \frac{41\,200 - 40\,000}{3\,000/\sqrt{36}} = +2.4$$

Las tres hipótesis alternativas posibles se comprueban en la figura:



Comparación entre las tres alternativas en términos de la evaluación de los datos muestrales en el caso de una muestra grande.

Cuando la varianza poblacional (σ_x^2) es desconocida, el intervalo de confianza de μ_x es el siguiente:

$$\overline{X} - t_{\alpha/2,\, n-1}\, s/\sqrt{n} \le \mu_x \le \overline{X} + t_{\alpha/2,\, n-1}\, S/\sqrt{n}$$

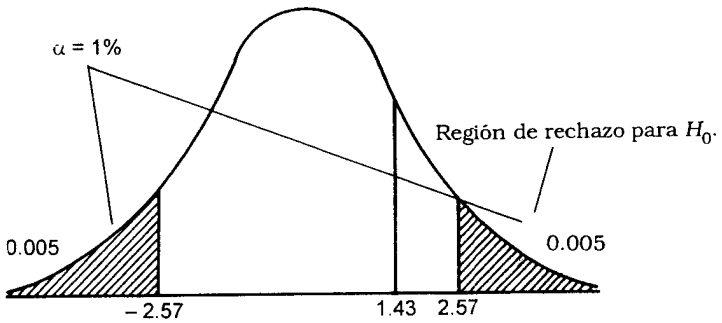
9.11 Intervalo de confianza, al 95%

$$(0.0063,\ 0.034)$$

Contraste de hipótesis

$$|Z| = 1.43$$

$$Z_{0.005} = \pm\ 257$$



H_0 no se rechaza al $\alpha = 5\%$ ni $\alpha = 1\%$

Sean: $H_0 : P = 25\%$

$H_1 : P \neq 25\%$

9.12 $\sigma_P = \sqrt{\dfrac{P(1 - P)}{n}}$

$$\sigma_P = \sqrt{\frac{0.25\ (0.75)}{740}} = 0.016$$

$$Z = \frac{0.20 - 0.25}{0.016} = \frac{-0.05}{0.016} = -3.1 \qquad Z_{\alpha/2} = 1.96$$

$|Z| = 3.1$

Entonces.
H_0 se rechaza con $\alpha = 5\%$.
Intervalo de confianza al 95%

$$(0.22,\ 0.28)$$

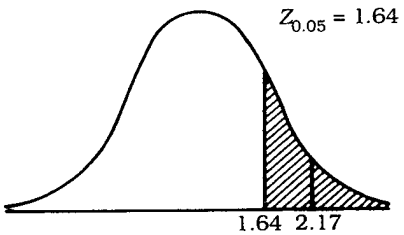
$H_0 : P = 5.5$
$H_1 : P < 5.5$

9.13 $\sigma_p = \sqrt{\dfrac{P(1-P)}{n}}\ \sqrt{\dfrac{N-n}{N-1}} = \sqrt{\dfrac{0.03\,(0.97)}{50}}\ \sqrt{\dfrac{350}{399}}$

$$\boxed{\sigma_p = 0.023}$$

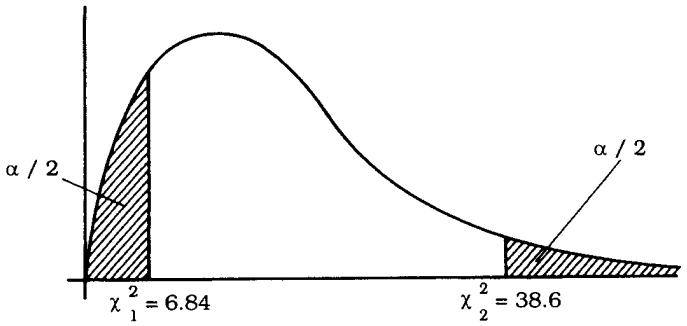
$$Z = \frac{(X/n) - P}{\sigma_p} = \frac{4/50 - 0.03}{0.023} = \frac{0.08 - 0.03}{0.023}$$

$$\boxed{Z = 2.17}$$



H_0 se rechaza con un $\alpha = 5\%$.

9.14 *a*)



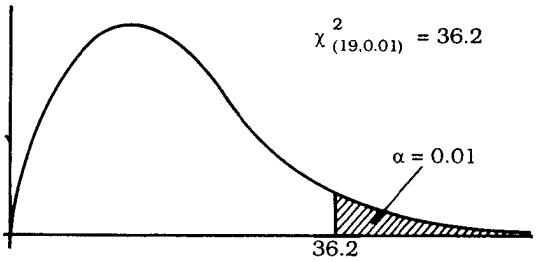
$gl = n - 1 = 19$

$$P(0.00011 \le \sigma_x^2 \le 0.00064) = 99\%$$

b) $H_0 : \sigma_x^2 = 0.0001$

$H_1 : \sigma_x^2 > 0.0001$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_x^2} = \frac{(19)\,(0.00023)}{(0.01)^2} = 43.7$$



Como 43.7 > 36.2 H_0 se rechaza. La compañía debe rechazar los lentes.

9.15 Como $n = 64$, $\overline{X} = 32$, $\sigma_x = 10$, $Z_{0.025} = \pm\ 1.96$
$29.6 < \mu_x < 34.5$, al 95% o (30, 35) redondeando.

9.16 $n = 64$, $\overline{X} = 60$, $\sigma_x = 15$, $Z_{0.025} = \pm\ 1.96$
$56 < \mu_x < 64$, al 95% o (56, 64)

9.17 $n = 49$, $\overline{X} = 12$ $\sigma_x = 2$, $Z_{0.025} = \pm\ 1.96$
$11.5 < \mu_x < 12.6$ o (11.5, 12.6) al 95%

9.18 $\sigma_x = 1.2$, $e = 0.3$, $Z_{0.025} = \pm\ 1.96$

$$n = \frac{Z^2\sigma_x^2}{e^2} = \frac{(1.96)^2 \times (1.2)^2}{(0.3)^2} = \frac{3.841 \times 1.44}{0.09} = 61.46$$

El tamaño de la muestra debe ser $n = 62$.

9.19 $e = 0.5$, $\sigma_x = 2$, $Z_{0.025} = \pm\ 1.96$

$$n = \frac{(1.96)^2 \times (2.0)^2}{(0.5)^2} = 61.5$$

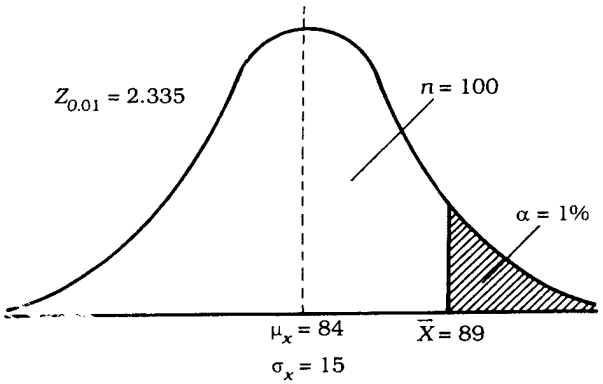
La muestra seleccionada debe ser $n = 62$.

9.20 $e = 15$, $\sigma_x = 50$, $Z_{0.025} = \pm\ 1.96$

$$n = \frac{(1.96)^2 \times (50)^2}{(15)^2} = 42.7$$

Se necesita una muestra $n = 43$.

9.21 $H_0 : \mu_x = 84$
$H_1 = \mu_x > 84$



Bajo la hipótesis nula H_0, se calcula la probabilidad de seleccionar muestras con medias muestrales iguales o mayores a 89.

Como: $\sigma_{\bar{x}} = \dfrac{\sigma_x}{\sqrt{n}} \sqrt{\dfrac{N - n}{N - 1}}$ y f.p.c. = 1, o sea $\sqrt{\dfrac{N - n}{N - 1}} = 1$

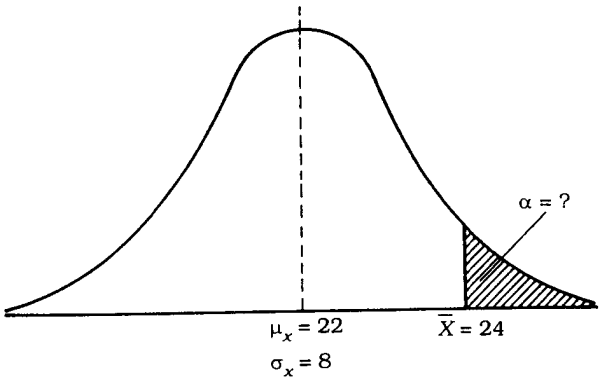
entonces $\sigma_{\bar{x}} = \dfrac{\sigma_x}{\sqrt{n}} = \dfrac{15}{\sqrt{100}} = 1.5 \quad \therefore \quad Z_c = \dfrac{\bar{x} - \mu_x}{\sigma_{\bar{x}}}$

$$Z_c = \frac{\bar{x} - \mu_x}{\sigma_{\bar{x}}} = \frac{89 - 84}{1.5} = 3.34$$

$P\{\bar{X} \geq 89 \mid H_0 \text{ cierta}\} = 0.001 < 0.005$

Esto es porque $3.34 > 3.30$, donde $Z_{0.0005} = 3.30$
La campaña publicitaria sí aumentó significativamente las llamadas telefónicas.

9.22 $H_0 : \mu_x = 22$
$H_1 : \mu_x > 22$

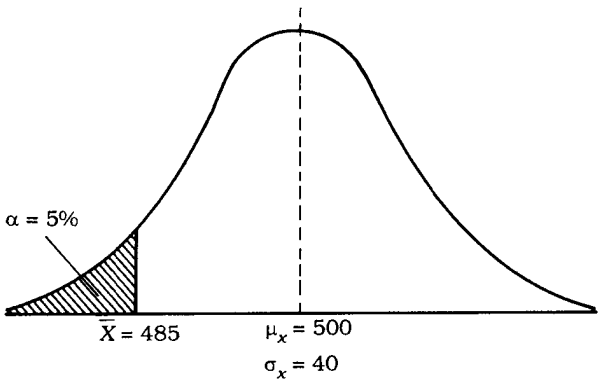


El razonamiento es igual que el problema anterior.
$Z_c = 1.801$ es mayor que 1.80, donde
$Z_{0.0359} = 1.80$, entonces H_0 se rechaza.
Hubo un incremento en la cosecha.

9.23 $H_0 : \mu_x = 500$

$H_1 : \mu_x < 500$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{40}{4} = 10$$



$$Z_c = -1.5$$

$$P\{\bar{X} \leq 485 \mid \mu_x = 500\} = 0.0668$$

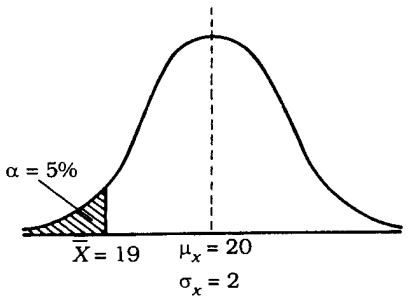
$$\text{como } \alpha = 5\% = 0.05$$

No se rechaza H_0, el promedio de calificaciones del colegio A no puede considerarse estadísticamente bajo con respecto a la calificación promedio.

9.24 $H_0 : \mu_x = 20$ f.p.c = 1
$H_1 : \mu_x \neq 20$ $\alpha = 5\%$
$n = 16$

$\sigma_{\bar{x}} = 0.5$

$Z_c = 2$



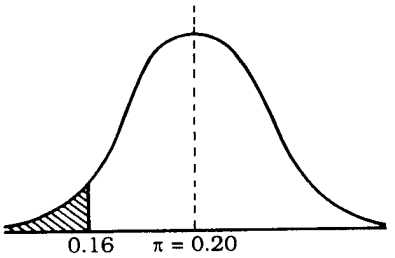
$P\{\bar{X} < 19 \mid \mu_x = 20\} = 0.0228 < 0.025$, se rechaza H_0
El proceso de producción debe detenerse y ajustarse.

9.25 $H_0 : \pi = 20\%$ $P = 16\%$

$H_1 : \pi < 20\%$ $Z_c = \dfrac{P - \pi}{\sigma_p} = \dfrac{0.16 - 0.20}{0.04} = -1.0$

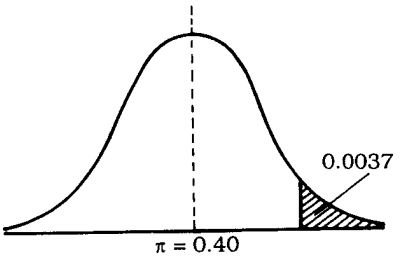
$$\sigma_p^2 = \frac{0.16}{100}; \quad \sigma_p = 0.04 \quad \text{o} \quad \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.04$$

f.p.c. = 1
$P[P \leq 0.16 \mid \pi = 0.20] = 0.1908$
H_0 no se rechaza



9.26 $H_0 : \pi = 0.40$ $\quad Z = 2.68$
$H_1 : \pi > 0.40$ $\quad \sigma_p = 0.0219$
$P = 0.46$ $\quad Z_{tablas} = 2.68$

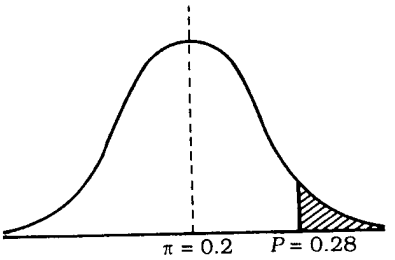
$P[P \geq 0.46 \mid \pi = 0.40] = 0.0037$



o sea que 3.7 veces en 1000 se podrán seleccionar muestras con proporciones iguales o mayores a 0.46 de entre una población con $\pi = 0.40$

con $\alpha = 5\%$ H_0 se rechaza

9.27 $H_0 : \pi = 0.2$
$H_1 : \pi > 0.2$



$$\sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}} = 0.04$$

$$Z_c = \frac{P - \pi}{\sigma_p} = 1.625$$

$$P\{P \geq 0.28 \mid \pi = 0.2\} = 0.052 > 0.05$$

H_0 no se rechaza, no existe diferencia estadísticamente significativa entre $P = 0.28$ y $\pi = 0.2$. El proceso de producción no se debe parar ni ajustar.

9.28 $H_0 : \pi = 0.5$ (o sea el 50%)
$H_1 : \pi \neq 0.5$ (diferente al 50%)
$n = 1\,000$

$\alpha = 5\%$ $\quad P = \dfrac{520}{1\,000} = 0.52$

$$\sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}} = \sqrt{\frac{0.5}{1\,000}} = 0.0158$$

$$Z_c = \frac{P - \pi}{\sigma_p} = \frac{0.52 - 0.5}{0.0158} = 1.266$$

$$P\{P \geq 0.52 \mid \pi = 0.5\} = 0.10275 > 0.025$$

H_0 no se rechaza $\therefore$ la moneda no es honesta.

9.29 $H_0 : \pi = 0.3$

$H_1 : \pi < 0.3$ $\quad P = \dfrac{100}{400} = 0.25$

$n = 400$ $\quad \sigma_p = \sqrt{\dfrac{0.3 \times 0.7}{400}} = 0.0225$

$\alpha = 5\%$

$Z_c = \dfrac{0.25 - 0.3}{0.0225} = -2.22$

$P\{P \leq 0.25 \mid \pi = 0.3\} = 0.0132 < 0.05$
H_0 se rechaza, el resultado no confirma lo dicho.

9.30 $H_0 : \pi = 0.5$

$H_1 : \pi < 0.5$ $\quad P = \dfrac{420}{1000} = 0.42$

$n = 1\,000$ $\quad \sigma_p = \sqrt{\dfrac{(0.5)^2}{1\,000}} = 0.0158$
$\alpha = 5\%$

$$Z_c = \frac{0.42 - 0.5}{0.0158} = -5.06$$

H_0 se rechaza.
No existe suficiente evidencia.

Capítulo 10

10.1 *a)* Como $n_1 + n_2 = 72 > 30$

$$Z_c = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\dfrac{\sigma^2_{x_1}}{n_1} + \dfrac{\sigma^2_{x_2}}{n_2}}}, \qquad Z_{0.025} = \pm 1.96$$

IC (95%) : $0.61 \leq \mu_{x_1} - \mu_{x_2} \leq 3.39$, redondeando (0.6, 3.4) con un error $e = 1.386$

b) $Z_c = 2.828^{**}$, H_0 se rechaza, $\alpha = 1\%$

10.2 *a)* $0.352 \leq \mu_{x_1} - \mu_{x_2} \leq 0.648$; *IC* (99%) o (0.35, 0.65)

b) $$t_c = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} = \frac{0.5}{0.05795} = 8.70^{**}$$ H_0 se rechaza $\alpha = 1\%$

10.3 *a)* $-0.13 \leq \mu_{x_1} - \mu_{x_2} \leq 0.93$ para un *IC* (95%) o también (−0.1, 0.9) con $e = 0.5$

b) $t_c = 1.48$ n.s. H_0 *no* se rechaza (no existe diferencia estadísticamente significativa)

10.4 Como $n_1 + n_2 < 30$ el modelo es

$$t_c = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\left(\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\right)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

a) *IC* (95%) $1.092 \leq \mu_{x_1} - \mu_{x_2} \leq 1.892$ o (1.09, 1.9)

b) $t_c = 0.56818$ n.s. H_0 *no* se rechaza.
$t_{16,\,0.025} = 2.12$

10.5 $t = \dfrac{\bar{d}}{\dfrac{S_d}{\sqrt{n}}}$, *donde* $S_d = \sqrt{\dfrac{1}{n-1}\left(\sum_i d_i^2 - \dfrac{(\sum_i d_i)^2}{n}\right)}$

$$\text{y } \bar{d} = \frac{\Sigma(X_1 - X_2)}{n}$$

$\bar{d} = 0.6134$
$S_d^2 = 0.156, \quad S_d = 0.394$

a) *IC* (95%) (0.38, 0.82) con $e = 0.2187$

b) $t_c = 6.03$ ** H_0 se rechaza, $\alpha = 1\%$

10.6 El modelo estadístico para muestras dependientes también se encuentra como

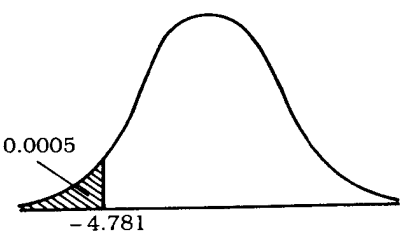
$$t_c = \frac{\left|\Sigma\bar{X}_1 - \Sigma\bar{X}_2\right|}{\sqrt{\dfrac{\Sigma D^2 - \dfrac{(\Sigma D)^2}{n}}{n(n-1)}}}$$

donde $D = X_1 - X_2$
$t_c = 5.81$ si $\Sigma X_1 - \Sigma X_2$
$t_c = -5.81$ *si* $\Sigma X_2 - \Sigma X_1$

a) $H_0 : \mu_{x_1} - \mu_{x_2}$
$H_1 : \mu_{x_2} < \mu_{x_1}$

Se supone que f_c disminuye las caries.
$\alpha = 0.0005$

$t_{9,0.0005} = 4.781$
como $-5.81 < -4.781$ H_0 se rechaza
¡f_c disminuye las caries!
$\alpha = 0.0005$



b) $A_c = \dfrac{\Sigma D^2}{(\Sigma D)^2} = \dfrac{50}{(14)^2} = 0.2551$

R.D. Si $A_c \leq A_{tablas}$ H_0 se rechaza
$A_{0.025} = 0.276$ F_c disminuye las caries $\alpha = 0.025$

10.7 *a)* $\sigma_{p_1 - p_2} = 0.06124$ $-0.02 < p_1 - p_2 < 0.22$ al 95% con $e = 0.12$
b) $Z_c = 1.63$ n.s. H_0 *no* se rechaza

10.8 H_0: $P_R = P_{NR}$ $Z_c = 1.28$ n.s. El recubrimiento *no* es efectivo.
H_1: $P_R > P_{NR}$ $Z_{0.05} = 1.645$

10.9 H_0: $\sigma_1^2 = \sigma_2^2$ $F_c = 5.36$ ** H_0 se rechaza
H_1 : $\sigma_1^2 \neq \sigma_2^2$ $F_{14,\,19,\,0.01} = 3.19$ Las varianzas no son iguales, no existe homoscedasticidad.

$\alpha = 1\%$

10.10 *a)* $B = 6.66$ n.s. H_0 *no* se rechaza, las varianzas son iguales.
$\chi^2_{3,0.05} = 7.815$
b) y *c)* no se puede aplicar Cochran ni Hartley, los grupos no son del mismo tamaño.

10.11 *a)* $C = 0.444$ n. s. H_0 *no* se rechaza las varianzas son iguales.
$C_{4,\,7,\,0.05} = 0.5365$
b) $F_H = 2.62$ n. s. H_0 *no* se rechaza las varianzas son iguales.
$F_{n-1,\,k,\,0.05} = F_{7,\,4,\,0.05} = 8.44$

10.12

ESCALA	*tc*	
1	13.51	**
2	11.56	**
3	12.53	**
A	10.40	**
B	14.80	**
C	11.86	**
D	12.78	**
E	12.22	**
AC	4.72	**

** significativas al 1%

∴ Existe una diferencia estadísticamente significativa entre el grupo de morfinómanos y de no morfinómanos.

Capítulo 11

11.1 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	1	128.0762	128.0762	20.0947
Error	13	82.8571	6.3736	
Total	14	210.9333		

Calculando $t = 4.49$

11.2 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	2	46.6667	23.3333	7.1591
Error	27	88.000	3.2593	
Total	29	134.6667		

11.3 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	1	1 416.1000	1416.100	170.6145
Error	8	66.400	8.300	
Total	9	1 482.5000		

11.4 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	1	0.1000	0.1000	0.0056
Error	8	142.800	17.8500	
Total	9	142.900		

11.5 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	3	893.819	297.94	1.849 n.s.
Error	20	3 222.014	161.101	
Total	23	4 115.833		

11.6 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	2	0.0895	0.0448	0.9021
Error	57	2.8286	0.0496	
Total	59	2.9181		

11.7 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	2	263.6066	131.8033	0.6196
Error	57	12 124.8381	212.7165	
Total	59	12 388.4447		

11.8 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	4	65.5648	16.3912	0.5277
Error	31	962.9752	31.0637	
Total	35	1028.5400		

11.9 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	3	86.8000	28.9333	2.1632
Error	16	214.000	13.3750	
Total	19	300.800		

11.10 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	4	120.0000	30.0000	1.0989
Error	20	546.0000	21.4590	
Total	24	666.0000		

11.11 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	4	1268.5333	317.1333	70.2659
Error	25	112.8333	4.5133	
Total	29	1381.3667		

11.12 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	1	4.8688	4.8688	22.10 *
Error	12	2.6433	0.2203	
Total	13	7.5121		

* Calculando $t = 4.70 \therefore F = t^2$

11.13 Tabla de análisis de varianza

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Razón F
Tratamientos	3	3 061.1250	1 020.3750	5.0276
Error	28	5 682.7500	202.9554	
Total	31	8 743.8750		

Capítulo 12

12.1 $F_{trat.} = 5.35$ $p < 0.01$
$F_{bloq.} = 29.400$ $p < 0.01$

12.2 $F_{trat.} = 22.8795$ $p < 0.001$
$F_{bloq.} = 1.2892$ n. s.

12.3 $F_{trat.} = 0.0979$ n. s.
$F_{bloq.} = 5.4167$ $p < 0.01$

12.4 $F_{trat.} = 18.1122$ $p < 0.001$
$F_{bloq.} = 129.0623$ $p < 0.0001$

12.5 $F_{trat.} = 5.5833$ $p < 0.01$
$F_{bloq.} = 7.7500$ $p < 0.01$
A = tratamientos B = bloques AB = interacción

12.6 $F_A = 2.2679$ n. s.
$F_B = 1.4145$ n. s.
$F_{AB} = 0.2988$ n. s.

12.7 $F_A = F_{trat.} = 2.7488$ n. s.
$F_B = F_{bloq.} = 7.8721$ $p < 0.01$
$F_{AB} = F_{interacción} = 1.2764$ n. s.

12.8 $F_A = 0.5555$ n. s.
$F_B = 0.4866$ n. s.
$F_{AB} = 1.6397$ n. s.

12.9 $F_A = 4.4879$ $p < 0.01$
$F_B = 3.7088$ $p < 0.01$
$F_{AB} = 0.0868$ n. s.

12.10 $F_A = F_{trat.} = 5.0625$ $p < 0.01$
$F_B = F_{bloq.} = 3.000$ $p < 0.05$
$F_{AB} = F_{trat.\ x\ bloq} = 28.3125$ $p < 0.001$

12.11 $F_A = 43.5823$ $p < 0.001$
$F_B = 59.5823$ $p < 0.001$
$F_{AB} = 8.4937$ $p < 0.01$

12.12 $F_A = 3.0381$ n. s.
$F_B = 2.4670$ n. s.
$F_{AB} = 0.0685$ n. s.

12.13 $F_A = 103.9467$ $p < 0.001$
$F_B = 1.6402$ n. s.
$F_{AB} = 0.8475$ n. s.

12.14 $F_A = 4.9351$ $p < 0.05$
$F_B = 0.2843$ n. s.
$F_{AB} = 0.0079$ n. s.

12.15 $F_A = 0.3333$ n. s.
$F_B = 49.75$ $p < 0.001$
$F_{AB} = 2.0833$ n. s.

Capítulo 13

13.1 $\hat{y} = 0.4301 - 0.0743\,x$
$r = -0.9432,\ r^2 = 0.8897$
$t = -8.0332 \qquad p < 0.001$
$F = 64.5316 \qquad p < 0.001$

13.2 $\hat{y} = -1.2123 + 0.0377\,x$
$r = 0.8901,\ r^2 = 0.7923$
$t = 4.7847$ * *
$F = 22.8932$ * *

13.3 $\hat{y} = 279.0716 + 94.0064\,x$
$r = 0.8894,\ r^2 = 0.7910$
$t = 5.5022$ * *
$F = 30.2745$ * *

13.4 $\hat{y} = 1.9209 + 6.6366\,x$
$r = 0.9982,\ r^2 = 0.9963$
$t = 43.6299 \qquad p < 0.00001$
$F = 1903.5896 \qquad p < 0.00001$

13.5 *a)* $\hat{y} = 115.7110 - 15.9628\,x$
$r = -0.8607,\ r^2 = 0.7408$
$t = -6.3248$
$F = 40.0030$

b) $\hat{y} = 116.3042 + e^{-0.1599}\,x$
$r = -0.8787 \qquad r^2 = 0.7722$
$t = -6.8880$
$F = 47.4445$

13.6 $\hat{y} = 8.017 + 0.0375\,x_1 + 0.911\,x_2 + 0.195\,x_3$
$r = 0.983,\ r^2 = 0.966$
Error estándar de la estimada = 1.647
$F = 160.762,\ P < 0.001$

Y puede predecirse por x_1, x_2, x_3 con una P.

	P
X_1	0.859
X_2	< 0.001
X_3	0.265

13.7 $\hat{y} = 198.215 - 0.297\,x_1 - 1.158\,x_2 - 2.229\,x_3 - 0.638\,x_4$
$r^2 = 0.628$, r = 0.792
Error estándar de la estimada = 7.053
$F = 5.475$, $p < 0.008$

Coeficiente	Error estándar	t	P
198.215	45.463	4.360	< 0.001
X_1	- 0.297	- 0.556	0.588
X_2	- 1.158	- 1.583	0.137
X_3	- 2.229	- 3.186	0.007
X_4	- 0.638	- 1.142	0.274

13.8

Atletas de potencia	Atletas de resistencia
$\hat{y} = -1.4353 + 0.2019\,X_1$	$\hat{y} = 30.157 + 0.292\,X_1$
$r = 0.65$, $r^2 = 0.4224$	$r = 0.430$ $r^2 = 0.185$
$t = 2.0949$	$t = 0.861$
$F = 4.389$	$F = 1.359$

13.9 $\hat{y} = 0.644 + 1.661\,X_1 + 0.017\,X_2$
$r = 0.987$ $r_2 = 0.975$
$F = 19.576$ (n. s), $P = 0.158$
Error estándar de la estimada = 1.562

Variable	Coeficiente	Error estándar	t	
Constante	0.644	4.524	0.142	n. s.
X_1	1.661	1.114	1.491	n. s.
X_2	0.017	1.186	0.014	n. s.

Los datos no se distribuyen en forma normal.
Las varianzas no son homogéneas.

13.10 $\hat{y} = -0.372526 + 0.797523\, x_1 + 0.091025\, x_2$
$r = 0.877$, $r^2 = 0.768$
$F = 11.61$ $p < 0.006$
Error estándar de la estimada = 2.646

Variable	Coeficiente	Error estándar	t	
Constante	−0.372526	4.480	−0.0831	n. s.
x_1	0.797523	0.440	1.811	n. s.
x_2	0.091025	0.0428	2.125	n. s.

Capítulo 14

14.1 *a*) $U = 13$
b) $t = 3.217$

14.2 $Z = -1.4981$

14.3 $\chi^2 = 10.920$ $p < 0.012$

14.4 $\chi^2 = 3.949$ n.s.

14.5 $\chi^2 = 1191.8$ $C = 0.346$ $p < 0.001$

14.6 $r_s = 0.967$

14.7 $H = 13.333$ $p < 0.020$

14.8 $H = 13.352$ $p < 0.020$

14.9 $\chi^2 = 8.051$ $p < 0.018$

14.10 $\chi^2 = 11.873$ $p < 0.008$

Índice analítico

A. de V. *Véase* análisis de varianza
ábaco, 689
adhesivos para materiales dentales, 5, 618
adición, regla general de, 211
agresividad, 607
ajuste
  a una línea recta, 505
  bondad de, 545
    para la distribución normal, 548
álgebra
  de conjuntos, 106
  de funciones, 146
  matricial, 146, 151-156
alternativa (H_1), 348
  de dos colas o bidireccional o de dos vías (*two ways*), 354
  de investigación, 349
  de trabajo, 350
  de una cola o unidireccional o de una vía (*one way*), 354
  descriptiva, 350
  explicativa, 350
  para la regresión, 509, 514-516
  prueba de, 351-355
    para el coeficiente de correlación, 520-522
    para la regresión, 509, 514-516, 522-524, 536-538
  simple, 349
ámbito. *Véase* contradominio
American Statistical Association, 676, 687
amplitud (A), 39, 44, 77
análisis
  combinatorio, 177
  de correlación, 516-517
  de covarianza, 16
  de regresión múltiple, 528-513
  de varianza, 445
    con dos factores, 481, 487
    de bloques aleatorizados, 488
    en una dirección, 594
    lógica del, 447-448
    no paramétrico, 563
    supuestos del, 469-470
    tabla del, 454
    y su relación con el de regresión múltiple, 532-533
  estadístico, 32
Anaxágoras, 6
ANDEVA. *Véase* análisis de varianza
ángulos, 667-670
ANOVA. *Véase* análisis de varianza
ansiedad, 473
  tabla de, de Taylor, 473n
ANVAR. *Véase* análisis de varianza
Aristarco, 6
Aristóteles, 6, 26
Army-Beta, prueba de, 476
Arquímedes, 6
asimetría (As), 69
asociación
  prueba de, 563
  significativa, 520

asociativa, propiedad, 113
atletas, 543
autoconcepto, 443
autoestima, 543
Avenarius, 7
axiomas de probabilidad, 203
Ayer, 8

Bacon, Francis, 7
Bartlett, prueba de, 401, 431, 676-679
Bayes, teorema de, 224-233
Berkeley, 7
Bernoulli
  distribución, 266-268
  proceso de, 266-267
binomio, teorema del. *Véase también* distribución binomial, 189-191, 270
Biometrika Tables for Statisticians, 680, 683, 685-686
Boyle, Robert, 5

cadena de Markov, 234-235
cálculo de posibilidades, 177
cardinalidad, 100, 114
"caries de biberón", 494
Carnap, Rudolf, 8
catadores, 611
causalidad, 505
censo, 22
censo, contra muestreo, 298
centil o centila, 54, 56, 60, 62-64
certeza, búsqueda de la, 8
Chebyshev
  desigualdad de, 261
  teorema de, 265
ciencia
  aspectos fundamentales de la, 3-4, 6
  empírica, 8
  griegos y, 6
clases, 44
Cochran, prueba de, 401, 431, 436, 687
coeficiente. *Véase también* prueba de hipótesis para
  de asimetría o sesgo, 70
  de concordancia, ω, de Kendall, 610-613
  de confianza, 345
  de contingencia, 572-575
  de correlación, 517-519, 601
    biserial de punto (r*bp*), 613-616
    comparación entre, 525-527
    de Pearson, 517, 568, 601
    múltiple, 538-539
    prueba de hipótesis para el, 520-522
    valor promedio del, 519-520
    valores de Z y, 665
  de determinación, 517
  de rangos ordenados. *Véase* coeficiente de correlación
  de regresión, 509, 517, 522
  de Spearmon (rs), 601-607, 683-684
  de variación (cv), 68, 81-82
  fi (Φ), 568-572
  multinomial, 188-189
  tau (τ) de Kendall, 607-610
  V de kramer, 575-577
cofactor, 150
colas, teoría de, 274
combinación, 187-188
comparaciones
  de medias de tratamiento, 462
  entre medianas, 595n
  múltiples, 462
complementación, propiedades de, 113
complemento relativo, 109
Comte, Auguste, 7
concordancia, ω, 618
  de Kendall, 610-613
condominio. *Véase* contradominio
conectivo lógico, 140
confianza
  coeficiente de, 345
  estimado de, 345
  intervalo de, 306, 345, 387, 401-422, 426-430, 509-513, 518-519
  región de, 345
confusión, factor de, 363, 470
conjunto(s), 98-100
  ajenos, 102
  álgebra de, 106
  complemento, 106
  complementos de, 104
  diferencia de dos, 109
  finitos, 100
  iguales, 102-103
  infinitos, 100
  propiedades de los, 113
  similares, 103
  universal, 102-103
  vacío, 102, 111
conmutativa, propiedad, 113
conteo, 179-185
  de elementos, 113
  patrón de, 182
  principio del, 114
contingencia
  coeficiente de, 572-575
  tablas de, 559 *ss.*
continuidad, corrección de, 285-287
contradominio, 123-126, 137
contraste para más de dos varianzas, 431-439

control
de calidad, 540
grupo de, 470
Copérnico, 7
corrección
de continuidad de Yates, 567, 571
de Pirie-Handem. *Véase* corrección de continuidad de Yates
factor de, 356
correlación
análisis de, 516-517
biserial de punto (r_{bp}), 613-616
coeficiente de, 517, 538-539
comparación entre, 525-527
intervalo de confianza para el, 518-519
prueba de hipótesis para el, 520-522
valor promedio del, 519-520
y valores de Z, 665
de Pearson, 517, 568, 601
débil, 520
espuria, 526
ficticia, 526
fuerte, 520
moderada, 520
muy débil, 520
nula o inexistente, 520
perfecta, 520
significativa, 520
Cramer, método de, 163
cuadrados
grecolatinos, 497
latinos, 497
medios. *Véase* medias cuadráticas
cuantil, 54, 56
cuantilas, 54-65
cuartil o cuartila, 54, 56
cuestionario, 19, 24. *Véase también* investigación, instrumentos de
curtosis, 70-71
curva
aguda o leptocúrtica, 71
aplanada o platocúrtica, 71
de De Moivre, 143
de Gauss, 143
de la campana, 143
normal, 69
semiaguda o mesocúrtica, 71

Dalton, 5
datos
agrupados, 31
categóricos, 546
no agrupados, 31, 82
De Anima (Aristóteles), 26
de tipo II (β), 349, 351-354
de predicción o de medición, 506
de varianza, 448
experimental, 363, 448
decil o decila, 54, 56, 558
defecto
genético, 572
Demócrito, 5-6
desigualdad, 127
de Chebyshev, 261
desviación
cuadrada media, 80
estándar (*s* o σ), 14, 78-80
media (*Dm*), 66, 77
determinación, coeficiente de, 517
determinantes, 149-151
Dewey, John, 8
diabetes mellitus, 553
diagrama
de árbol, 179-182, 231
de cajas, 64
de Carrioll, 103, 547
de Galton. *Véase* ojiva
de Venn-Euler, 103
diferencia
entre dos proporciones, 421-422
estadísticamente significativa, 446
mínima significativa (LSD), 462
o resta aritmética (*A-B*), 109
de dos conjuntos, 109
simétrica de conjuntos, 111
diseño
aleatorizado, 447-448, 488-493
de bloques, 481
de dos factores, 481
antes-después, 413
cuasiexperimental, 16
de control, 16
experimental, 15-16
factorial, 16
factorial de parcelas divididas, 497
y construcción de pruebas, 617
distribución(es)
Bernoulli, 262, 266-268
binomial, 261-262, 268-273, 276
probabilidades acumuladas de la, 627-640
de F, 652-656
de la estadística
de Friedman, 664
de Kruskal Wallis, 661-663
de Mann-Whitney, 657-660
de la estadística de rangos con signo de Wilcoxon, 663
de Poisson, 261-262, 274-276
probabilidades acumuladas de la, 641-642

de probabilidad, 547
exponencial, 287-289
hipergeométrica, 261-262, 277-278
ji cuadrada, 644-645
libre, 546
muestrales, 317
normal, 268-274, 276
normal estándar, 142, 278-280, 284-285
parcentiles de la, 646
probabilidades acumuladas de la, 643
valores críticos de la, 646
poblacionales, 505
probabilística, 14, 144, 199, 262-263
proceso de, 261
simétrica, 73
t de Student, 650-651
Z, 280-285
dominio, 123-126
Dunnett
prueba de, 470-471
tabla de, 672-675

ecología, 472
ecuación de regresión, 509
evaluación de la, 533-536
y coeficiente de regresión, 509
y media, 509
y varianza, 509
ecuaciones lineales, 156-166
efecto(s)
de interacción A × B, 488
principales, 493
electromiogramas, 475
elementos, conteo de, 113
empates, 584
empiricismo, 4
empirismo, 7
encuestas, 19
entrevista, 24. *Véase también* investigación, instrumentos de
epistemología, 6
error
aceptable, 22
aleatorio, 10-11
de medida, 10-11
de muestreo, 22
de tipo I (α), 349, 351-354
estadístico, 21
estándar, 344-345, 367, 463, 509-511
experimental, 10
frecuentes, 23
muestral, 305-306
sistemático, 10-11
escala
de intervalo, 13
de razón, 13
nominal, 12
original, 12
escalar, 149, 152
escuela
bayesiana o subjetivista, 198
clásica u objetivista, 198
espacio
de eventos, 183, 199
muestral, 183, 202
estadística, 10, 12
de Friedman, 664
de Kruskal Wallis, 661-663
de Mann Whitney, 657-660
de rangos con signo de Wilcoxon, 663
e informe científico, 17
multidimensional, 26
no paramétrica, 546
estado absorbente, 234
estimación, 338, 505
intervalar, 338
puntual, 338
estimador, 339
consistente, 339
eficiente, 339
insesgado o sin vicio, 80, 339
poblacional, 310
puntual, 339
suficiente, 339
estudio(s)
correlativo, 504
de mercado, 529
experimentales controlados, 516
observacionales, 516
ética, 19
evaluación de la ecuación de regresión, 509
eventos
compuestos, 204
espacio de, 183, 199
independientes, 215-219
mutuamente excluyentes, 206-212, 216-219
simples, 203
experimentación, 24, 178, 202. *Véase también* investigación, instrumentos de
experimento(s)
factoriales, 496
objetivo de un, 448

factor, 447
de confusión, 363
de corrección, 356, 483, 490
factorial, 185
Feigel, 8
Ferguson, 593-594
Feyerabend, 4, 8

filosofía analítica, 8
Fisher, prueba de, 593
frecuencias, 34, 37
distribución de, 34-44
polígono de, 35-36, 42
relativa acumulada, 43, 56
relativa, 43, 200
tablas de, 86
Friedman
distribución de la estadística de, 664
prueba de, 547, 597-600
función, 98, 125
algebraica, 129, 140, 146
constante, 129
cuadrática, 129, 135
cúbica, 129, 136
valor absoluto, 129
contradominio de la, 125
de probabilidad, 263n
dominio de la, 125
trascendente, 129, 140
exponencial, 129, 140
logarítmica, 129, 140
trigonométrica, 129
exponencial de base, 140
hiperbólica, 129

Galileo, 7
época de, 15
Gauss, Carl Friedrich, 143
glucosa, 553
gráfica
circular o de sectores, 33
de barras, 33
gran media, 544
grandes números, ley de los, 265-266
grupo(s)
de control, 470
experimentales, 470

Hartley, prueba de, 401, 431
hepatitis, 497
hipótesis, 3, 19, 350
alterna, 14
científicas, 9, 347
contraste, 338, 355, 401 *ss.*
de investigación, 15
estadística, 347, 350
nula (H_0), 348
histograma, 35
Hobbes, Thomas, 7
homogeneidad, 21-23
de dos poblaciones, 560
de varianzas, 431, 436
prueba de, 559-563
homoscedasticidad, 401, 431, 469, 525
Hume, David, 4, 7-8

idempotencia, propiedades de, 113
identidad, propiedades de, 113
igualdad, 127
relación de, 113
imagen. *Véase* contradominio
inclusión, relación de, 113
independencia, prueba de, 563-565
inducción, 4, 8
como método lógico, 7
probabilística, 8
proceso de, 4
Infante Gil, Said, 663-664
inferencia, 7
estadística, 14
clásica, 348
informe científico y estadística, 17
ingesta de agua, 489
insulina, 553
interacción, concepto de, 482, 493-496
interés compuesto, 141
intersección, 108
intervalo(s)
abierto, 127
cerrado, 127
de confianza, 306, 345, 401-422, 426-430, 509-513
para el coeficiente de correlación, 518-519
híbrido, 127
jerarquizados, 581
modal, 40-41
investigación, 6, 19
científica, 10
crítica, 7
de mercados, 21
instrumentos de, 24

Journal of the American Statistical Association, 672
juego del melate, 193
justificacionismo, 8
justificacionistas, 5

Kendall
coeficiente *tau* (τ) de, 607-610
concordancia, ω, de, 610-613
diferencia con Spearman (r_s), 610
Kepler, Johannes, 7
Kirk, Roger E., 483
Experimental Design: Procedures for the behavioral sciences, 483n
Kneller, 4

Kolmogorov-Smirnov
modelo, 14
prueba de, 546, 557-559
Kramer, coeficiente V de, 575-577
Kruskal-Wallis
distribución de la estadística de, 661-663
prueba de, 547, 594-597
Kuhn, 4, 8

Lakatos, 4, 8-9
Laplace, Pierre Simon, marqués de, 143, 201
Leucipo, 5
ley
científica, 4, 351
de De Morgan, 112
de los grandes números, 265-266
de Mendel, 552
de Newton, 179
idempotentes, 112
libertad, grados de, 453, 463
lingüística, 8
Locke, John, 7
logaritmo natural (L_n), 145
lógica, 8

Mach, Ernst, 7
MACSTAT, 163n, 454, 488, 494, 539, 665-671
Mann-Whitney
distribución de la estadística de, 657-660
prueba de la *U* de, 577-584
para muestras grandes, 581-584
marca de clase, 40-41
Markov, cadena de, 234-235
masa de probabilidad, 263n, 279
matriz, 98, 146-156
álgebra de, 146, 151-156
cuadrada, 148-150
de cofactores, 155
determinante de, 150
diagonal, 148
escalar, 148-149
identidad o unitaria, 149
igualdad de, 149
inversa de una, 154, 535
no singular, 154
nula, 149
operaciones con, 152
multiplicación de, 152-154
resta de, 153
suma de, 152
opuesta, 153
orden de, 147
renglón, 147-148
representación de, 236-248
simétrica, 151
singular, 154
transpuesta de una, 151
Maxwell, 4
Mc Nemar, prueba de, 588-590, 688
media(s), 14
aritmética (x), 49-52, 74-75
armónica (h), 84
confiabilidad de la, 68
cuadrática(s) o cuadrados
geométrica (x_g), 84-85
gran, 455
medios, 453
del error, 454
dentro de grupos, 454
ponderada (x_p), 83
y ecuación de regresión, 509
mediana (*Me*), 49-50, 56, 60, 75, 578
comparación entre, 595n
prueba de la, 591-594
medición, 3
error de, 10
escala de intervalo de, 13
escala de razón de, 13
escala nominal de, 12
escala ordinal de, 12
medida
de dispersión o variabilidad, 66, 76-83
de tendencia central, 34, 49-53, 74-76
Medina, J. Carlos, 665-671
megapascales, 618
Méndez, I., 315, 350, 688
Mercado, 6
metateoría, 8
método
de Cramer, 163
empírico, 7
experimental, 7
lógico, 7
moda (*Mo*), 49-50, 75
modalidad, 71
modelo(s)
aleatorios, 178
de la χ^2 (ji cuadrada), 14, 546
de regresión múltiple, 531
determinísticos, 178
estándar de regresión, 505
estocásticos, 178
Kolmogorov-Smirnov, 14
no determinísticos, 178
no paramétricos, 14
paramétricos, 13
probabilísticos, 178
U de Mann-Whitney, 14
Moivre, Abraham de, 143
Moore, George Edward, 8

morfina, 489
Morgan, propiedad de, 113
muestra, 297
ábacos para calcular, 689
aleatoria, 23
binomial, 547
correlacionada, 413
de cuotas, 22
de testigos, 21
dependiente, 413-420
espacio de, 183, 202
homogénea, 456
independiente, 405-412
irrestricta aleatoria, 304
normal, 547
Poisson, 547
seguridad de una, 22
tamaño de la, para estimar π, 382-383
muestreo, 297
censo contra, 298
clasificaciones de, 298
aleatorio estratificado, 309-311
aleatorio simple, 304-309
aleatorio sistemático, 312
de atributos, 299
de bola de nieve, 303
de juicio o criterio, 302-303
de variables, 299
determinístico, 299-303
fijo, 298-299
irrestricto aleatorio, 304
por conglomerados, 315-317
por conveniencia, 303-304
por cuotas, 303
probabilístico, 299-302, 304-317
secuencial, 298-299
sistemático, 312-315

naloxeno, 489, 502
Nemenyi, prueba de, 600-601
Newton, Isaac, 7
segunda ley de, 179
no simetría, propiedad de, 113
notación punto, 465
número(s)
aleatorios, tabla de, 304-309, 448
desbalanceado, 448
reales, 130

observación, 24. *Véase también* investigación, instrumentos de
ojiva o diagrama de Galton, 35, 43, 56
creciente o positiva, 44
decreciente o negativa, 44
ordenada al origen, 134
óxido nítrico, emisión de, 449

partición, 204-212. *Véase también* coeficiente multinomial
Pascal, triángulo de, 190
Pavlov, Iván, 178
Pearson, Karl, 7
coeficiente de correlación de, 517, 568, 601
pendiente, 134
permutaciones (Pn, r), 185-187
circulares, 186-187
plano cartesiano, 123
Platón, 6
Plutchik, 462
población, 73
población finita, 318, 356. *Véase también* universo
Poincaré, Henri, 7
Poisson, distribución de, 261-262, 274-276
Polanyi, 9
Popper, K. R., 4-5, 8
Porcentil(es), 54
de la distribución normal estandarizada, 646
de la prueba de Bartlett, 676-679
posibilidades, cálculo de, 177
positivismo, 4
clásico, 7
lógico, 8
probabilidad, 8, 198
distribución de, 14, 144, 199, 262-263
axiomas de, 203
condicional, 212-214
de transición, 234
marginal, 219
masa de, 263n, 279
objetiva, 201
simple o no condicionada, 215
subjetiva, 201
teoría clásica de, 201
total, teorema de la, 219-224
uso de, 199
vector de, 234
procedimiento de Tukey, 462
proceso
aleatorio, 199
científico, 4
de Bernoulli, 266-267
de inducción, 4
de las distribuciones, 261
estacionario, 267
estocástico, 199, 233-248
producto
cartesiano, 119-121
lógico de dos conjuntos, 108
propiedad
asociativa, 113
conmutativa, 113

de complementación, 113
de De Morgan, 113
de idempotencia, 113
de identidad, 113
de no simetría, 113
de reflexión, 113
de simetría, 113
de transitividad, 113
prueba(s)
de Army-Beta, 476
de asociación, 563
de Bartlett, 401, 431, 676-679
de bondad de ajuste, 546
de dos direcciones por rangos. *Véase* prueba de Friedman
de Cochran, 401, 431
de homogeneidad de varianzas, 687
de Dunnett, 470-471
de Friedman, 547, 597-600
de Hartley, 401, 431
de homogeneidad de varianzas, 686
de hipótesis
para el coeficiente de correlación, 520-522
para la regresión, 509, 514-516, 522-524, 536-538
de homogeneidad, 559-563
de independencia, 563-565
de inteligencia de Wechsler (Wais), 542
de Kolmogorov-Smirnov, 546, 557-559
de Kruskal-Wallis, 547, 594-597
de la mediana, 591-594
de la *U* de Mann-Whitney, 577-584
para muestras grandes, 581-584
de McNemar, 588-590, 688
de Nemenyi, 600-601
de rangos con signos en pares de Wilcoxon, 584-588
de Sheffé, 462
de significancia
de *fi* (ϕ), 570
de ω, 612-613
de r_{bp}, 616-617
de r_s, 606-607
de Student-Newman-Keuls (SNK), 462
de Tukey-Snedecor, 462-469
diseño y construcción de, 617
estadísticas, 401
de Sandler, 401, 420, 585, 671
t de Student, 401, 420, 585
Z, 401
exacta de Fisher, 593
F (Snedecor-Fischer), 445, 448, 487
nueva de rango múltiple de Duncan, 462
prueba de inteligencia (WISC), 282
psicología, 12
estadísticas en, 12
experimental, 539
psicometría, 12
punto
coeficiente de correlación biserial de (*rbp*), 613-616
medio, 41

rango. *Véase también* contradominio
centílico, 62
incluyente, 45
prueba de, 584-588
realismo crítico, 8
rechazo, región de, 351
recta de regresión, 506-509
reflexión, propiedad de, 113
región de rechazo, 351
regresión. *Véase también* prueba de hipótesis para la
coeficiente de, 509, 517, 522
ecuación de, 509
exponencial, 539
hipótesis para la, 509, 514-516
lineal poblacional, 506
logarítmica, 539
modelo estándar de, 505
múltiple, 528-531
ecuación de, 533-536
recta de, 506-509

relación
binaria, 122
de igualdad, 113
de inclusión, 113
residual. *Véase* residuo
residuo, 506
respuesta
encubierta (REn), 474
explícita (RE), 474
galvánica, 500
Riveros, H., 351
Rojas Soriano, R., 350
Rosas, L., 351
ruido blanco, 500
Russell, Bertrand, 8

Sandler, prueba de, 401, 420, 585, 671
Scheaffer, Richard L., 304
sentidos, 8
serendipity, 10
Sheffé, prueba de, 462
simetría, propiedad de, 113
sistema de ecuaciones lineales, 156-166
sociología, 7

Spearman
coeficiente de (r_s), 601-607, 683-684
diferencia con Kendall, 610
prueba de significancia de, 606-607
spss, 528
Student-Newman-Keuls, prueba de, 462
subconjuntos, 101, 137
suma de cuadrados
de la interacción, 492
de las desviaciones, 510
de tratamientos, 452
dentro de los grupos, 453
total corregida, 452
total no corregida, 451
SYSTAT, 528
suma lógica, 106
sumatorias, 47

t de Student, 14
distribución, 650-651
prueba, 401, 420, 585
tabla de, 607
tabla
de contingencia o doble entrada, 105
de contingencia, 559ss
de doble entrada o diagrama de Carroll, 547
de Dunnet, 672-675
de frecuencias, 86
de números aleatorios, 304-309, 448
de *t* de Student, 607
del análisis de varianza, 454
Tarsky, Alfred, 8
teorema
central del límite, 265-266, 317-329, 356, 581
corroboración del, 319-329
de Bayes, 224-233
de Chebyschev, 265
de la probabilidad total, 219-224
del binomio, 189-191, 270
teoría(s), 5-6
contraste entre, 5
de colas, 274
explicativa universal, 4
fenómenos y, 5
leyes y, 5
principios y 5
rivales, 9
Tomás, Santo, 26
transición, 234
probabilidad de, 234
transitividad, propiedad de, 113
triángulo de Pascal, 190
Tukey, procedimiento de, 462
prueba de, 462-469
Snedecor, 680-682
valores críticos de, 685

U de Mann-Whitney, modelo, 14
úlcera estomacal, 502
unidad experimental, 448
unión, 106, 140
universo o población, 23, 297

valores
críticos
de A (prueba de Sandler), 671
de homogeneidad de varianzas, 687
de la distribución normal estandarizada, 646
de la prueba de Cochran
de la prueba de Hortley de homogeneidad de varianzas, 686
de Q (Tukey-Snedecor), 685
del coeficiente de Spearman, 683-684
de *r*, en términos de *Z*, 666
de Z, en términos de *r*, 665
Varas, E., 494
variable(s)
activas, 25
aleatoria, 262
binomial, 284
aleatoria, 339, 367
continua dependiente, 447
continuas, 32
cualitativa (nominal), 589
de interés, 316
dependientes, 15-16, 25, 126-127
dicotomizadas, 569, 589
discretas, 32
"emboscada", 525-526
experimentales, 25
extrañas, 9, 15-16
independientes, 15-16, 126-127
nominal independiente, 447
variación, fuentes de, 472
varianza(s)
análisis de, 445
con dos factores, 481, 487
de bloques aleatorizados, 488
lógica del, 447-448
supuestos del, 469-470
tabla del, 454
de un factor, 445
error de, 448
estimada, 453
homogeneidad de, 431, 436
iguales (homoscedasticidad), 469
intergrupal, 447
intragrupal, 447
(s^2 o σ^2), 14, 80

y ecuación de regresión, 509
y error de la media, 511
vinos, 611
violencia en T.V., 598

Wais, 542
Weimer, 4, 7-9
Weschsler (Wais), prueba de, 542
Whitehead, Alfred North, 8
Wilcoxon, prueba de rangos con signos en pares de, 584-588
WISC, 282
Wisdom, 8
Wittgenstein, 8

Yates, continuidad de, 567

Sí, envíeme el catálogo de las novedades de OXFORD en

- ❒ Español
- ❒ Inglés
- ❒ Texto universitario
- ❒ Área profesional
- ❒ Ciencia/Tecnología
- ❒ Derecho
- ❒ Economía/Negocios
- ❒ Informática
- ❒ Ingeniería
- ❒ Otros

Nombre . . . . . .
Institución . . . . . .
Dirección . . . . . .
Ciudad, Estado o Departamento. . . . . .
. . . . . .

Cargo . . . . . .
Departamento . . . . . .
Código postal . . . . . .
País . . . . . .
Teléfono . . . . . .

Temas que le gustaría fueran tratados
en futuros libros de OXFORD: . . . . . .
. . . . . .
. . . . . .
. . . . . .
Título de la obra . . . . . .
. . . . . .
. . . . . .

¿Por qué elegí este libro?
- ❒ Prestigio del autor
- ❒ Prestigio OXFORD
- ❒ Reseña Revista . . . . . .
- ❒ Catálogo OXFORD
- ❒ Buscando en librería . . . . . .
- ❒ Requerido como texto
- ❒ Precio

Comentarios . . . . . .
. . . . . .
. . . . . .
. . . . . .

Este libro me ha parecido:
- ❒ Malo
- ❒ Bueno
- ❒ Excelente

OXFORD
UNIVERSITY PRESS

Por favor, llene este cupón y envíelo por fax al (01) 5705 3738
o por e-mail a: oxford@oupmex.com.mx

OFICINAS OXFORD UNIVERSITY PRESS

México
Oxford University Press México, S.A. de C.V.
Antonio Caso 142, Col. San Rafael, 06470,
México, D.F., tels. (52) 5592 4277 y 5592 5600,
fax. 5705 3738, e-mail: oxford@oupmex.com.mx

Argentina
Oxford University Press Argentina, S.A.
Reconquista 661, piso 1 (1003), Capital Federal
Buenos Aires, tel. (541) 4312 7300, fax. 4313 5700

Brasil
Oxford University Press
Al. Joaquim Eugênio de Lima, 732,
6° andar, Jardim Paulista, 01403-000,
São Paulo - SP, tel. (5511) 253 9335

Centroamérica
Oxford University Press de Centroamérica, S.A.
7a. avenida, 19-35, zona 11,
Col. Mariscal, 01011, Guatemala, Ciudad
tels. (502) 473 3274 y 473 2541, fax 442 3552

Chile
Oxford University Press Chile, S.A.
Félix de Amesti 181, Las Condes, Santiago de Chile,
tels. (562) 228 6958 y 207 4323, fax. 207 4381

Colombia
Oxford University Press Colombia, S.A.
Carrera 11, núm. 93-53, piso 5.
Apartado Aéreo 253410, Santafé de Bogotá, D.C.
Colombia: tels. (571) 625 1048 y 635 1070,
fax 635 1115

Ecuador
Oxford University Press Ecuador, S.A.
Mariana de Jesús 867 y Amazonas
Depto. 2, piso 2, edif. Báez,
Quito, Ecuador, tel. (5932) 231 789, fax 565 368

España
Oxford University Press España
Parque Empresarial San Fernando
Edificio Atenas, 1a. planta
San Fernando de Henares, 28830,
Madrid, tel. (3491) 660 2600, fax 660 2626

Perú
Oxford University Press
130 Porta, Oficina 806, Miraflores, Lima,
tels. (511) 446 0235 y 444 5039, fax. 446 0235

Venezuela
Oxford University Press de Venezuela, C.A.
Av. Francisco de Miranda, cruce con calle Capitolio
Edificio Torre B.B., sector la California, piso 3, oficina
303-304, Caracas, tels. (582) 239 2176 y 239 6489,
fax. 239 4475

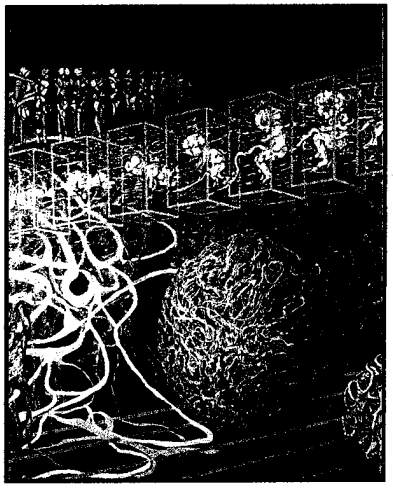
booksmedicos.org

Estadística para las ciencias sociales y del comportamiento

Es un texto actualizado, ya que los temas que contiene están incluidos en los nuevos programas de estudio en las carreras de psicología, sociología, trabajo social, administración (pública, de empresas turísticas, de instituciones de salud), ciencias de la educación (pedagogía), ciencias de la comunicación, informática, contaduría, diseño (ambiental, gráfico, industrial), mercadotecnia, ciencias de la salud, etc., que se imparten en las universidades e institutos tecnológicos de México y Latinoamérica, tanto en el nivel profesional como de posgrado.

Este texto estimula el aspecto activo, crítico, reflexivo y analítico del lector, por lo que la estadística puede ser incorporada a su propia estructura cognoscitiva, conjugando la estrategia académica de enseñanza-aprendizaje con los sistemas de educación tradicional, abierta y a distancia.

Haroldo Elorza Pérez-Tejada explica con claridad, generalidad y sencillez la estadística, que se ha convertido en un recurso fundamental para la mejor comprensión y prospectiva de los procesos sociales, de salud y del comportamiento, que interaccionan con el ser humano y su entorno ecológico, político, social y económico.



OXFORD
UNIVERSITY PRESS

ISBN 970-613-181-7

9 789706 131812