

PROBABILIDAD Y ESTADÍSTICA

para las ciencias sociales del comportamiento y la salud

Primera edición
en español



William Mendenhall III • Robert J. Beaver • Barbara M. Beaver

PROBABILIDAD Y ESTADÍSTICA

para las ciencias sociales del comportamiento y la salud

William Mendenhall III • Robert J. Beaver • Barbara M. Beaver

Probabilidad y estadística para las ciencias sociales del comportamiento y la salud



**EDICIÓN
EN ESPAÑOL**

William Mendenhall, III

University of Florida, Emérito

Robert J. Beaver

University of California, Riverside, Emérito

Barbara M. Beaver

University of California, Riverside, Emérito

Adaptación:

Haroldo Elorza Pérez-Tejada

Traducción:

Jorge Alberto Velázquez Arellano

Revisión técnica:

M.I. Ángel Leonardo Bañuelos Saucedo

Profesor de Carrera Titular

Facultad de Ingeniería

Universidad Nacional Autónoma de México (UNAM)



Probabilidad y estadística para las ciencias sociales del comportamiento y la salud
Primera edición en español

William Mendenhall, III; Robert J. Beaver y
Barbara M. Beaver

Director Editorial para Latinoamérica:
Ricardo H. Rodríguez

**Editora de Adquisiciones para
Latinoamérica:**
Claudia C. Garay Castro

**Gerente de Manufactura para
Latinoamérica:**
Antonio Mateos Martínez

**Gerente Editorial en Español para
Latinoamérica:**
Pilar Hernández Santamarina

Gerente de Proyectos Especiales:
Luciana Rabuffetti

Coordinador de Manufactura:
Rafael Pérez González

Editora:
Ivonne Arciniega Torres

Diseño de portada:
Gloria Ivonne Álvarez López

Imagen de portada:
© AMV_8o/Shutterstock.com

Composición tipográfica:
Tsuki Marketing S.A. de C.V.
Gerardo Larios García

© D.R. 2017 por Cengage Learning Editores, S.A. de C.V.,
una Compañía de Cengage Learning, Inc.
Corporativo Santa Fe
Av. Santa Fe núm. 505, piso 12
Col. Cruz Manca, Santa Fe
C.P. 05349, México, D.F.
Cengage Learning® es una marca registrada
usada bajo permiso.

DERECHOS RESERVADOS. Ninguna parte de
este trabajo amparado por la Ley Federal del
Derecho de Autor, podrá ser reproducida,
transmitida, almacenada o utilizada en
cualquier forma o por cualquier medio, ya sea
gráfico, electrónico o mecánico, incluyendo,
pero sin limitarse a lo siguiente: fotocopiado,
reproducción, escaneo, digitalización,
grabación en audio, distribución en Internet,
distribución en redes de información o
almacenamiento y recopilación en sistemas
de información a excepción de lo permitido
en el Capítulo III, Artículo 27 de la Ley Federal
del Derecho de Autor, sin el consentimiento
por escrito de la Editorial.

Traducido del libro
Introduction to Probability and Statistics,
Fourteenth Edition
William Mendenhall, III; Robert J. Beaver y Barbara M.
Beaver
Publicado en inglés por Brooks/Cole, una compañía de
Cengage Learning © 2013, 2009
ISBN: 978-1-133-10375-2

Adaptado del libro *Introducción a la probabilidad y
estadística*, 14a. ed.
William Mendenhall, III; Robert J. Beaver
y Barbara M. Beaver
Publicado por Cengage Learning © 2015
ISBN: 978-607-519-876-7

Datos para catalogación bibliográfica:
Mendenhall, William, III; Robert J. Beaver y Barbara M.
Beaver
*Probabilidad y estadística para las ciencias sociales del
comportamiento y la salud*, primera edición en español
ISBN: 978-607-526-309-0

Visite nuestro sitio en:
<http://latinoamerica.cengage.com>

Contenido

1

DESCRIPCIÓN DE DATOS CON MEDIDAS NUMÉRICAS 1

- 1.1 Aspectos fundamentales de la ciencia 2
 - Propósitos 2
 - Introducción 2
 - Relaciones entre estadística e investigación 7
 - Medición y estadística 9
 - Inferencia estadística y científica 11
 - Estadística e informe científico 13
- 1.2 Descripción de un conjunto de datos con medidas numéricas 14
- 1.3 Medidas de centro 14
 - Ejercicios 18
- 1.4 Medidas de variabilidad 20
 - Ejercicios 24
- 1.5 Sobre la significación práctica de la medición estándar 25
- 1.6 Una medición del cálculo de s 29
 - Ejercicios 31
- 1.7 Mediciones de posición relativa 34
- 1.8 El resumen de cinco números y la gráfica de caja 38
 - Ejercicios 41
 - Repaso del capítulo 44**
 - Tecnología actual 45**
 - Ejercicios suplementarios 48**

2

PROBABILIDAD 54

- 2.1 Introducción 55
- 2.2 El papel de la probabilidad en estadística 56
- 2.3 Eventos y el espacio muestral 56
- 2.4 Enfoques o escuelas de la probabilidad 59
- 2.5 Axiomas de probabilidad 60
- 2.6 Particiones 62
- 2.7 Cálculo de probabilidades con el uso de eventos sencillos 63
 - Ejercicios 66
- 2.8 Reglas útiles de conteo 68
 - Ejercicios 73

2.9	Relaciones de evento y reglas de probabilidad	74
	Cálculo de probabilidades para uniones y complementos	76
2.10	Independencia, probabilidad condicional y la regla de la multiplicación	79
2.11	Probabilidad condicional	81
	Ejercicios	84
2.12	Teorema de Bayes	87
	Ejercicios	90
2.13	¿Eventos mutuamente excluyentes o independientes?	91
2.14	Procesos estocásticos	93
	Cadenas de Markov	94
	Representación gráfica	94
	Representación matricial	96
	Resumen	107
	Ejercicios	107
	Ejercicios suplementarios	110

3

ALGUNAS DISTRIBUCIONES DE PROBABILIDAD IMPORTANTES 116

3.1	Introducción	117
3.2	Variables aleatorias discretas y sus distribuciones de probabilidad	117
	Variables aleatorias	117
	Distribuciones de probabilidad	117
	La media y desviación estándar para una variable aleatoria discreta	119
	Ejercicios	123
3.3	La distribución binomial de probabilidad	125
	Ejercicios	133
3.4	Distribuciones de probabilidad para variables aleatorias continuas	136
3.5	La distribución de probabilidad de Poisson	139
	Ejercicios	144
3.6	La distribución hipergeométrica de probabilidad	145
	Ejercicios	147
3.7	La distribución normal de probabilidad	148
3.8	Áreas tabuladas de la distribución normal de probabilidad	149
	La variable aleatoria normal estándar	149
	Cálculo de probabilidades para una variable aleatoria normal general	153
	Ejercicios	156
3.9	La aproximación de la distribución de probabilidad binomial a la normal	158
	Ejercicios	163
	Repaso del capítulo	164
	Tecnología actual	166

Ejercicios suplementarios 177

CASO PRÁCTICO: Un misterio: casos de cáncer cerca de un reactor 185

CASO PRÁCTICO: “¿Va a calificar por curva?” 186

4 DISTRIBUCIONES MUESTRALES 187

4.1 Introducción 188

4.2 Planes muestrales y diseños experimentales 188

Ejercicios 191

4.3 Estadísticas y distribuciones muestrales 192

4.4 El teorema central del límite 195

4.5 La distribución muestral de la media muestral 198

Error estándar 199

Ejercicios 202

4.6 La distribución muestral de la proporción muestral 204

Ejercicios 208

4.7 Una aplicación muestral: control estadístico de procesos (opcional) 209

Una gráfica de control para la media del proceso: la gráfica \bar{x} 209

Una gráfica de control para la proporción de piezas defectuosas: la gráfica p 211

Ejercicios 213

Repaso del capítulo 215

Tecnología actual 216

Ejercicios suplementarios 219

CASO PRÁCTICO: Muestreo de la Ruleta de Monte Carlo 222

5 ESTIMACIÓN DE MUESTRAS GRANDES 224

5.1 Dónde hemos estado 225

5.2 A dónde vamos: inferencia estadística 225

5.3 Tipos de estimadores 226

5.4 Estimación puntual 227

Ejercicios 232

5.5 Estimación de intervalo 233

Construcción de un intervalo de confianza 234

Intervalo de confianza de muestra grande para una media poblacional μ 236

Interpretación del intervalo de confianza 237

Intervalo de confianza de muestra grande para una proporción poblacional p 239

Ejercicios 241

5.6 Estimación de la diferencia entre dos medias poblacionales 242

Ejercicios 245

- 5.7 Estimación de la diferencia entre dos proporciones binomiales 248
Ejercicios 250
- 5.8 Límites de confianza a una cola 252
- 5.9 Selección del tamaño muestral 253
Ejercicios 257
- Repaso del capítulo 258**
- Ejercicios suplementarios 259**

6

INFERENCIA A PARTIR DE MUESTRAS PEQUEÑAS 262

- 6.1 Introducción 263
- 6.2 Distribución t de Student 263
Suposiciones detrás de la distribución t de Student 266
- 6.3 Inferencias de muestra pequeña respecto a una media poblacional 267
Ejercicios 271
- 6.4 Inferencias de muestra pequeña para la diferencia entre dos medias poblacionales: muestras aleatorias independientes 274
Ejercicios 280
- 6.5 Inferencias de muestra pequeña para la diferencia entre dos medias: una prueba de diferencia en pares 283
Ejercicios 288
- 6.6 Inferencias respecto a la varianza poblacional 290
Ejercicios 296
- 6.7 Comparación de dos varianzas poblacionales 297
Ejercicios 303
- 6.8 Repaso de suposiciones de muestra pequeña 305
Repaso del capítulo 306
Tecnología actual 307
Ejercicios suplementarios 313

7

ANÁLISIS DE VARIANZA 320

- 7.1 El diseño de un experimento 321
- 7.2 ¿Qué es un análisis de varianza? 322
- 7.3 Las suposiciones para un análisis de varianza 322
- 7.4 El diseño completamente aleatorizado: una clasificación en una dirección 323
- 7.5 El análisis de varianza para un diseño completamente aleatorizado 324
División de la variación total en un experimento 324
Prueba de igualdad de las medias de tratamiento 327
Estimación de diferencias en las medias de tratamiento 329
Ejercicios 332

7.6	Clasificación de medias poblacionales	335
	Ejercicios	338
7.7	Diseño de bloque aleatorizado: una clasificación en dos direcciones	339
7.8	El análisis de varianza para un diseño de bloque aleatorizado	340
	División de la variación total en el experimento	340
	Prueba de igualdad de las medias de tratamiento y de bloque	343
	Identificación de diferencias en las medias de tratamiento y de bloque	345
	Algunos comentarios de precaución en bloque	346
	Ejercicios	347
7.9	El experimento factorial $a \times b$: una clasificación en dos vías	351
7.10	El análisis de varianza para un experimento factorial $a \times b$	353
	Ejercicios	357
7.11	Repaso de las suposiciones del análisis de varianza	361
	Gráficas residuales	361
7.12	Un breve repaso	363
	Repaso del capítulo	364
	Tecnología actual	365
	Ejercicios suplementarios	370
	CASO PRÁCTICO: ¡Cómo ahorrar dinero en comestibles!	376

REGRESIÓN LINEAL Y CORRELACIÓN 377

8.1	Introducción	378
8.2	Modelo probabilístico lineal simple	378
8.3	El método de mínimos cuadrados	381
8.4	Un análisis de varianza para regresión lineal	383
	Ejercicios	386
8.5	Prueba de la utilidad del modelo de regresión lineal	389
	Inferencias respecto a β , la pendiente de la recta de medias	390
	El análisis de varianza de la prueba F	393
	Medir la fuerza de la relación: el coeficiente de determinación	393
	Interpretación de los resultados de una regresión significativa	394
	Ejercicios	395
8.6	Herramientas de diagnóstico para verificar suposiciones de la regresión	398
	Términos de error dependientes	398
	Gráficas residuales	398
	Ejercicios	399
8.7	Estimación y predicción usando la recta ajustada	402
	Ejercicios	406
8.8	Análisis de correlación	408
	Ejercicios	412

- 8.9 Covarianza 414
 - Repaso del capítulo 417**
 - Tecnología actual 418**
 - Ejercicios suplementarios 421**
 - CASO PRÁCTICO: ¿Su automóvil está “Hecho en EUA”? 426**

9

ANÁLISIS DE REGRESIÓN MÚLTIPLE 428

- 9.1 Introducción 429
- 9.2 El modelo de regresión múltiple 429
- 9.3 Un análisis de regresión múltiple 430
 - El método de mínimos cuadrados 431
 - El análisis de varianza para regresión múltiple 432
 - Prueba de la utilidad del modelo de regresión 433
 - Interpretación de los resultados de una regresión significativa 434
 - Comprobación de suposiciones de regresión 436
 - Uso del modelo de regresión para estimación y predicción 436
- 9.4 Un modelo de regresión polinomial 437
 - Ejercicios 440
- 9.5 Uso de variables predictoras cuantitativas y cualitativas en un modelo de regresión 444
 - Ejercicios 450
- 9.6 Prueba de conjuntos de coeficientes de regresión 453
- 9.7 Interpretación de gráficas residuales 456
- 9.8 Análisis de regresión por pasos 457
- 9.9 Interpretación errónea de un análisis de regresión 458
 - Causalidad 458
 - Multicolinealidad 458
- 9.10 Pasos a seguir al construir un modelo de regresión múltiple 460
 - Repaso del capítulo 460**
 - Tecnología actual 461**
 - Ejercicios suplementarios 463**

10

ANÁLISIS DE DATOS CATEGÓRICOS 471

- 10.1 Introducción 472
 - Estudios evaluativos, un enfoque actual 472
 - Los diferentes objetos de la evaluación 473
 - Estudios evaluativos: procedimientos generales 474
 - Áreas de interés del estudio evaluativo 477
 - Programas susceptibles de evaluación 477
 - Interpretación de los resultados 477

- 10.2 Una descripción del experimento 479
- 10.3 Estadístico ji cuadrada de Pearson 480
- 10.4 Prueba de probabilidades de celda especificada: la prueba de bondad de ajuste 481
 - Ejercicios 483
- 10.5 Tablas de contingencia: una clasificación de dos vías 485
 - La prueba de independencia ji cuadrada 485
 - Ejercicios 490
- 10.6 Procedimiento *post hoc* 491
 - Coeficiente $f(\Phi)$ 493
 - Coeficiente de contingencia (C) 496
 - Prueba de significancia 498
 - Coeficiente V de Kramer 499
- 10.7 Prueba exacta de Fisher 501
- 10.8 Prueba de McNemar 502
- 10.9 Comparación de varias poblaciones multinomiales: una clasificación de dos vías con totales de fila o columna fijos 504
 - Ejercicios 507
- 10.10 La equivalencia de pruebas estadísticas 509
- 10.11 Otras aplicaciones del estadístico de prueba ji cuadrada 509
 - Repaso del capítulo 511**
 - Tecnología actual 511**
 - Ejercicios suplementarios 515**

ESTADÍSTICA NO PARAMÉTRICA 521

- 11.1 Introducción 522
- 11.2 La prueba de suma de rango de Wilcoxon: muestras aleatorias independientes 522
 - Aproximación normal para la prueba de suma de rango de Wilcoxon 526
 - Ejercicios 529
- 11.3 La prueba del signo para un experimento de dos poblaciones 531
 - Aproximación normal para la prueba del signo 532
 - Ejercicios 534
- 11.4 Una comparación de pruebas estadísticas 535
- 11.5 La prueba de rango con signo de Wilcoxon para un experimento de dos poblaciones 536
 - Aproximación normal para la prueba de rango con signo de Wilcoxon 539
 - Ejercicios 540
- 11.6 La prueba H de Kruskal-Wallis para diseños completamente aleatorizados 542
 - Ejercicios 546
- 11.7 La prueba F_r de Friedman para diseños de bloque aleatorizados 548

- 11.8 Prueba de Nemenyi 551
 - Ejercicios 552
- 11.9 Prueba de la mediana 554
- 11.10 Coeficiente de correlación de rango 556
- 11.11 Prueba de significancia de r_s 560
- 11.12 Coeficiente tau (τ) de Kendall 561
- 11.13 Coeficiente de concordancia (ω) de Kendall 564
 - Prueba de significancia de ω 566
- 11.4 Coeficiente de correlación (r_{bp}) biserial de punto 566
 - Prueba de significancia de r_{bp} 569
- 11.15 Prueba de Kappa 570
 - Ejercicios 572
- 11.16 Resumen 575
 - Repaso del capítulo 576**
 - Tecnología actual 577**
 - Ejercicios suplementarios 580**
 - CASO PRÁCTICO: ¿Cómo está su nivel de colesterol? 585**

12

TEORÍA DE LA RESPUESTA AL ÍTEM 587

- 12.1 Introducción 588
- 12.2 Teoría clásica de los tests en la psicometría 588
 - Supuestos básicos de la teoría de la puntuación verdadera 589
 - Confiabilidad de un test 589
 - Condiciones de paralelismo 590
 - Características de los ítems en la TCT 590
 - Principales limitaciones de la teoría clásica de los tests 591
- 12.3 ¿Qué ofrece la teoría de la respuesta al ítem? 591
 - Curva característica del ítem (CCÍ) 592
 - Modelo ideal de Guttman y parámetros de un ítem 593
 - Índice de dificultad 593
 - Discriminación de un ítem 594
 - Parámetro de la pseudoadivinación 594
 - Modelo de ojiva normal 594
 - Reparametrización del modelo de ojiva normal 595
 - Modelo logístico de un parámetro o modelo de Rasch 597
 - Modelo logístico de dos parámetros 599
 - Modelo logístico con tres parámetros 599
- 12.4 Principales supuestos de la TRÍ 600
 - Unidimensionalidad del test 600
 - Indeterminación de la escala de rasgo latente 600

12.5	Estimación de parámetros del examinado y los ítems	601
	Método de estimación de máxima verosimilitud	601
	Estimación de los parámetros: a y b	602
12.6	Función de información	602
	Usos de la función de información	603
	Función de información del test	603
12.7	Evaluación de bondad de ajuste del modelo	604
	Interpretación del índice de bondad de ajuste	604
12.8	Modelos politómicos de la teoría de la respuesta al ítem	605
	Modelos politómicos para categorías ordenadas	606
	Modelo de respuesta graduada	607
	Ventajas de los modelos politómicos	609
12.9	Resumen	610

APÉNDICE A LOS ESCRITOS CIENTÍFICOS 611

APÉNDICE B MATRICES 631

ANEXO 643

FUENTES DE DATOS 684

RESPUESTAS A EJERCICIOS SELECCIONADOS 691

GLOSARIO 701

Descripción de datos con medidas numéricas

OBJETIVOS GENERALES

Las gráficas son sumamente útiles para la descripción visual de un conjunto de datos, pero no siempre son la mejor herramienta cuando se desea hacer inferencias acerca de una población a partir de la información contenida en una muestra. Para este propósito, es mejor usar medidas numéricas para construir una imagen mental de los datos.

ÍNDICE DEL CAPÍTULO

- Aspectos fundamentales de la ciencia (1.1)
- Descripción de un conjunto de datos con medidas numéricas (1.2)
- Medidas de centro (1.3)
- Medidas de variabilidad (1.4)
- Sobre la significación práctica de la medición estándar (1.5)
- Una medición del cálculo de s (1.6)
- Mediciones de posición relativa (1.7)
- El resumen de cinco números y la gráfica de caja (1.8)



NECESITO SABER...

Cómo calcular cuartiles muestrales

Los muchachos de verano

¿Los campeones de béisbol de hoy son mejores que los de “ayer”? ¿Los jugadores de la Liga Nacional batean mejor que los de la Liga Americana? El estudio práctico del final de este capítulo contiene los promedios de bateo de campeones de las ligas mayores. Se pueden usar medidas numéricas descriptivas para contestar éstas y otras preguntas similares.

ASPECTOS FUNDAMENTALES DE LA CIENCIA

Propósitos

El objetivo central del presente capítulo es que el lector sea capaz de ubicar a la estadística dentro del contexto de la ciencia y la investigación.

De igual forma, al término del mismo el lector podrá:

- Reconocer que la ciencia ha facilitado el desarrollo de las teorías que exponen la realidad.
- Comprender la conceptualización del empirismo y del positivismo en el proceso de acumulación del conocimiento.
- Explicar la forma en la que las teorías constituyen simplemente la organización lógica de las leyes empíricas.
- Enunciar la forma que tiene el empirismo de entender la ciencia.
- Reconocer que toda teoría, todo modelo y toda ley científica son una conjetura acerca de cómo es la realidad.
- Relacionar la opinión de Popper de que toda ley, principio, teoría o modelo es una conjetura o suposición.
- Diferenciar el punto de vista de los positivistas y los justificacionistas con relación a la ciencia.
- Considerar el punto de vista de Popper acerca de que: “Lo que caracteriza al hombre de ciencia no es la posesión del conocimiento o de verdades irrefutables, sino la investigación desinteresada e incesante de la verdad”.

Introducción

La ciencia es una de las empresas más humanas y productivas que haya desarrollado el hombre. Si lo que caracteriza al ser humano 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, entonces la ciencia es 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 basada en un proceso analítico y crítico produce el conocimiento que ha permitido una mejor comprensión de la realidad circundante. Asimismo, ha facultado al hombre para penetrar en los secretos más profundos del mundo, incluido el ser del hombre mismo. La ciencia ha facilitado el desarrollo de teorías que exponen la realidad, con base en un examen de la relación entre los intentos de explicación teórica, evidencia empírica y congruencia lógica, tanto interna a la explicación como en lo relativo a otras teorías con las que tienen vínculos conceptuales. Esto ha implicado que el científico pruebe sus teorías confrontándolas con la evidencia existente que, con el objeto de evaluar la teoría de que se trata, se acumula con procedimientos rigurosos. Asimismo, el científico está a la caza de inconsistencias internas en la lógica de las explicaciones, así como de las contradicciones entre las diversas teorías vinculadas.

La conceptualización del empirismo y del positivismo acerca de la naturaleza del proceso de acumulación de conocimiento se ha sustentado siempre en el proceso de inducción. Este principio señala, tal como lo plantea Hume, que si observa una cierta regularidad en los procesos naturales (incluida la naturaleza humana), entonces es posible generalizar a partir del establecimiento de una ley. De acuerdo con esta visión, el problema de la ciencia es observar cuidadosamente la naturaleza, evitando caer en errores debidos a la posible confusión de causas. El mejor modo de evitar el error es realizar una cuidadosa observación y medición del fenómeno y utilizar el método experimental para no confundir la verdadera causa de los fenómenos con otras que en apariencia los producen. 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. Ésta es la forma que tiene el empirismo de entender la ciencia y, con ciertas modificaciones, el positivismo.

Hume ya había planteado la naturaleza de las limitaciones lógicas del conocimiento inductivo: 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 solución planteada por Karl Popper (1972) a este dilema se hizo en términos de postular que nunca se puede partir de ninguna certidumbre acerca de nada de lo que se cree. 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 la 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, dice este autor, basan su desarrollo en la confrontación crítica con los hechos y con la lógica. En sus palabras, “... ningún conjunto de enunciados contrastadores verdaderos podrá justificar la pretensión de que una teoría explicativa universal es verdadera”.[†] Sin embargo, afirma que: “suponiendo que los enunciados contrastadores sean verdaderos, con base en ellos a veces podemos justificar la pretensión de que una teoría explicativa universal es falsa”.^{††} Esto desplaza el énfasis de la investigación al sentido contrario de como lo plantea el punto de vista tradicional científico: no es posible probar que las teorías sean verdaderas, sólo es factible eliminar las falsas. Por ello, Popper dice: “el método de la ciencia es el método de las conjeturas audaces e ingeniosas seguidas por intentos rigurosos de refutarlas”.^{†††}

Esto hace de la ciencia una aventura fascinante, donde las teorías se tienen que construir; hay que inventarlas sobre la base de lo que ya se comprende del fenómeno en cuestión. No obstante, lo que hace a la ciencia más emocionante aún, es la posibilidad de someter las teorías a rigurosas pruebas de evidencia. Por un lado, esto otorga un grado mucho mayor de libertad, pero también un enorme sentido de responsabilidad.

De acuerdo con la opinión de Popper, toda ley, todo principio, toda teoría o todo modelo es una conjetura, una suposición. Las teorías no surgen, como supondrían los positivistas, 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. Los datos, cuando se generan a *posteriori*, sirven también para poner a prueba la elaboración de una ley o teoría, y si ésta resulta rechazada, es precisamente la naturaleza de las fallas la que podría servir de inspiración para el posterior planteamiento.

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 permiten internarse en lo desconocido haciendo pronósticos no triviales y novedosos, sobre cuya base se les somete a pruebas rigurosas. El carácter riguroso de la

[†] K. R. Popper, *Conocimiento objetivo*, Tecnos, Madrid, 1974, p. 20.

^{††} *Ibid.*, p. 20.

^{†††} *Ibid.*, p. 83.

contrastación hace que las teorías cuantitativas sean mejores, permiten mayor precisión en la elaboración del pronóstico y, por tanto, en la prueba de ellas, ya que permite señalar con toda exactitud el grado de error de pronóstico y decidir si éste sólo se debe a un error de medición, o si se debe a una falla de la teoría.

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 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 relaciones con otros entes.

El propósito de la explicación es profundizar en la comprensión de los fenómenos. Por ejemplo, en el área de la química, Robert Boyle había desarrollado la distinción taxonómica entre elementos y compuestos. A partir de esa base, Proust elaboró la ley empírica de las proporciones constantes, la cual sostiene que los elementos tienen que combinarse en una determinada proporción de peso, para producir una reacción que genere un compuesto específico, sin que ninguno de los elementos que participaron en la reacción sobre, de modo que se requiere que estos elementos guarden una relación que se pueda expresar por medio de números enteros. Cuando esta proporción no se cumplía, la reacción no era completa y sobraban los elementos que tenían una proporción mayor a la estipulada. Esta ley empírica era suficiente para manejar coherentemente los fenómenos de la química que influían en las reacciones entre sustancias. Sin embargo, Dalton, un inglés, modesto profesor de primaria, introdujo una de las especulaciones más fructíferas 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. De acuerdo con Dalton, los átomos de cada elemento se unen en combinaciones determinadas para formar moléculas de compuestos, las cuales son apiñamientos de átomos en estructuras determinadas. Es entonces el número de átomos de cada clase, que existe en cada molécula de un compuesto específico, lo que define la proporción de los elementos que deben entrar en la reacción para que no sobren átomos de un tipo u otro.

No ha existido una propuesta más fértil que ésta. 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 al movimiento de los átomos y las moléculas.

Como puede observar ahora, las teorías son instrumentos intelectuales muy poderosos que 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, el que supone que es así en verdad. En este sentido, la ciencia es el instrumento intelectual más importante logrado por la humanidad, después de la invención de la escritura. La ciencia permite al hombre entender y anticipar el mundo que lo rodea, gracias al desarrollo de teorías que se asemejan cada vez más a la realidad, ya que, como lo señala Popper, las teorías van siendo, por selección natural, cada vez mejores mapas conceptuales de la realidad y cada vez más exactos y precisos. Las teorías se transforman en las mejores guías para la praxis humana, permitiendo el desarrollo de las poderosas tecnologías que caracterizan a la época moderna y haciendo factible el enorme éxito de la especie, por el que la humanidad ha logrado la población con la que actualmente cuenta.

Naturaleza de la investigación

La investigación se considera no sólo la parte creativa de la ciencia con la que se busca expandir el conocimiento y comprensión de la realidad, sino también la base que permitirá construir un mapa de ésta capaz de guiar al hombre en su búsqueda. Los mapas que proporciona la

ciencia no son únicamente esquemas descriptivos sino conceptuales-causales del mundo circundante; es decir, son guías en relación con las clases de objetos y eventos y sus conexiones causales recíprocas. Así, en función de esta situación el hombre avanza en su dominio cognoscitivo de la realidad.

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

La ciencia, como tal, surge en forma sistemática y organizada entre los griegos. La ciencia se desarrolló en 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; desde la filosofía de la ciencia de Aristóteles, Platón, Demócrito, etc., hasta las contribuciones empíricas y teóricas concretas de Anaxágoras, Aristarco, Arquímedes, entre otros.

Sin embargo, no fue sino hasta que inició el Renacimiento cuando surgió 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 y bajo el cobijo de la *filosofía empirista*. Cuatro astrónomos preeminentes en la creación de la ciencia: Copérnico, Kepler, Galileo y Newton, físicos además los dos últimos, impulsaron el 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.

Junto a quienes practicaban la ciencia como método empírico para abordar el conocimiento, surgían los filósofos empiristas, que fundamentaban el nuevo método de obtener conocimiento. Bacon, Hobbes, Locke, Berkeley y Hume instituyen el empirismo como el único camino al conocimiento, al establecer la experiencia empírica como la única posibilidad para conocer la verdad y la inducción como el método lógico que hacía posible esto al usar la inferencia como medio para el logro de conocimientos generales a partir de experiencias particulares. Ellos establecieron el conocimiento científico como un camino seguro a la verdad. Intentaban desarrollar un sistema de inferencia racional que hiciera 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 de manera perenne en el acervo de conocimiento verdadero, siendo la ciencia un proceso de acumulación de hechos. En pocas palabras, con ellos, la concepción de la ciencia se desarrolla como la búsqueda en la experiencia empírica de un camino para una seguridad absoluta que justifique 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.^{††} El positivismo considera a la experiencia como fuente de conocimiento, y los hechos generales o leyes son la única fuente de *certidumbre*. Encontramos a pensadores como Mach, Avenarius, Poincaré y Pearson, entre otros, como estructuradores de una filosofía que establecía a la ciencia sobre una base empírica que se pro-

[†] 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.

^{††} Comte fue, además, padre de la sociología, que desarrolla dentro del marco filosófico de su *método positivista* de hacer ciencia.

ponía como guía pragmática para enfrentar la vida. El Universo, incluyendo al hombre, estaría constituido por 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 que se le pruebe, y la prueba consiste en ponerlo bajo la hegemonía de la autoridad epistemológica pertinente, en este caso, la experiencia empírica.

En la actualidad, el trabajo de filósofos con enfoques diferentes, aunque con un núcleo central de acuerdo fundamental, culmina el 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 Feigl, 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 el siglo pasado, pero en la actualidad ha surgido con gran vigor la perspectiva de la ciencia, ya mencionada, llamada no justificacionista, que analiza el proceso de conocimiento científico sin recurrir al de la *justificación empírica* como base para el establecimiento de éste. Autores como Popper, Kuhn, Lakatos, Feyerabend y Weimer han jugado 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 llama *justificacionismo*[†] 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 *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 señala que es precisamente esa búsqueda de una base firme e incontrovertible la fuente de los problemas. Hace un análisis sobre la reflexión de Hume 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 es posible de modo alguno comprobar teorías, sí es factible refutarlas. Su solución para el funcionamiento de la ciencia puede resumirse 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 lógica. En esta perspectiva no *justificacionista*, la teoría no surge directamente de los datos a partir de

[†] 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.

un proceso de inducción, ya que cualquier proposición teórica, desde una simple ley empírica hasta un modelo teórico o una teoría, proviene de *una conjetura*. El origen puede ser, como se señaló anteriormente, cualquier posible fuente: la observación de una o varias regularidades, una especulación teórica, una analogía o algún otro proceso. Lo importante es, como ya se ha dicho, que las conjeturas científicas se ponen a prueba por medio de la crítica lógica y empírica (a diferencia de las conjeturas puramente especulativas en otros ámbitos). No obstante, si los hechos apoyan la teoría, no cabe pensar que la justifican, sólo que hasta ahora no la han refutado.

Justificación frente a confrontación

De acuerdo con Lakatos, 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; el investigador no se percata de qué variables debe controlar hasta que otra explicación sugiere los aspectos que debe considerar con más cuidado para decidir cuál explicación es la que mejor da cuenta de los hechos. Lakatos asevera que no es tan importante el choque entre teoría y datos como la competencia entre las teorías rivales. La actitud rigurosa no implica la supresión instantánea 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 para un investigador es permanecer comprometido con una posición para poder articularla plenamente y explorar sus consecuencias”.

¿De dónde surgen las teorías?

Como se ha visto, las teorías científicas son intentos de explicación de la realidad, confrontadas con los hechos de manera rigurosa, que compiten entre sí para tratar de encontrar la mejor manera de dar cuenta de los hechos. Son sistemas de creencias acerca del mundo, más explícitos, claros y precisos que otros conjuntos de creencias (la religión, el sentido común, las pseudociencias, etc.), y que son sometidos a una rigurosa prueba sistemática. Las teorías pueden tener una génesis muy diversa. Por una parte, se encuentra el conocimiento tácito de muchos aspectos de la realidad, donde el sentido común y el conocimiento personal son una fuente muy importante de hipótesis científicas. En la vida cotidiana observa casualmente muchos hechos que después lleva al laboratorio y examina con más cuidado. Con frecuencia, esas mismas observaciones inspiran los 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 llevan a encontrar lo que no se busca y se le ha denominado *serendipia*. 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, entendiendo que lo observado da claves para la construcción de la explicación. Otro origen frecuente de 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 alterna y resuelva el conflicto.

Relaciones entre estadística e investigación

El tema de este capítulo es examinar el papel que tiene la estadística en la investigación científica. La estadística es una rama de las matemáticas que se dedica a entender los fenómenos que tienen un cierto grado de azar. En la ciencia se enfrenta el problema de que los fenómenos son multicausales y existe una diversidad de aspectos de los que sólo se tiene un grado de control relativo. Frente a esta problemática, resulta útil emplear un método que permita lidiar con datos con una cierta dosis de incertidumbre. En realidad la estadística es un instrumento muy valioso para organizar la información científica y para tomar decisiones acerca de ella; sería imposible concebir la investigación científica moderna sin dicha 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 un radar) para controlar el error de medida, y luego usar la estadística para decidir si su nueva medición es igual o diferente a la que tuvo 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 estos grupos de sujetos.

Error de medida y error 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 usar la estadística en la investigación.

El error de medida es el que se comete al medir cualquier cosa a pesar del cuidado que se tenga. Por una variedad de razones es posible 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 aleatorio, que se refiere a inexactitudes de un instrumento al medir con él. El primero produce distorsiones de nuestros datos, que a la vez implican un error en nuestras conclusiones.

Los errores sistemáticos pueden radicar en fallas de calibración de los instrumentos. Los instrumentos de medición deben ser comparados con un estándar, el cual determina que el instrumento efectivamente arroja los valores adecuados a la escala que está usándose. Por ejemplo, el metro tiene como estándar de calibración una varilla de vanadio-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 (*tests*) se estandarizan (una forma de calibración) aplicándolos a una gran muestra de la población donde van a usarse (por ejemplo, la ciudad de La Plata o México), y luego se establecen las calificaciones estándar. Es decir, si se usa una prueba de inteligencia en México y se emplean estándares ingleses o argentinos, se estaría produciendo un error sistemático de medida.

Los errores sistemáticos también pueden ser causados por 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 con una aguja de bobina, como lo pudiera ser un sonómetro, o un efecto de una variable no adecuadamente controlada como el sexo o la clase social del encuestador en una prueba de personalidad.

Los errores aleatorios (de azar) son aquellos que se cometen por aspectos accidentales, tales como limitaciones perceptuales o inexactitud al momento de tomar una medida, como pudiera ser el caso de un error al leer una escala de manera distraída. Asimismo, los errores aleatorios también se deben a la influencia accidental, de carácter temporal, de otras variables, como el estado de ánimo de un sujeto al someterse a una prueba, 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 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 a través de la estadística. Es posible comparar medidas con error y estimar el valor casi exacto de cierta medida gracias a la estadística.

Medición y estadística

La estadística se aplica sobre medidas obtenidas de los diversos objetos de estudio en diferentes condiciones. Por ejemplo, si desea verificar si un curso de capacitación para soluciones de problemas mejora la inteligencia de quienes lo cursaron, puede tener un grupo al cual le mide la inteligencia antes y después de llevar el citado curso; es decir, aplica la estadística sobre medidas tomadas de los casos, antes y después de la intervención.

Medir, según Torgerson, es asignar números a una propiedad de acuerdo con una regla.[†] Es decir, medir es una forma particular de observación en la cual se asignan números a las propiedades observadas. Es de notar que esta asignación no es del todo arbitraria, ya que usa una regla de asignación de números a los valores de la propiedad. Algo que es necesario comprender es que debe abstraer la dimensión, lo cual es más difícil si se trata de aspectos no observables directamente, como el nivel del metabolismo basal, el peso de los átomos o la inteligencia.

En la vida cotidiana, sin duda, aparecen numerosas formas de medir, como usar una báscula para pesar. El peso se refiere a estándares, como el gramo, que es el peso de un centímetro cúbico de agua a nivel del mar. La regla para pesar consiste en comparar el peso del objeto de interés con el de un estándar. El número (valor) es asignado de acuerdo con la regla de que el peso del objeto sea igual o un múltiplo del peso del estándar.

Las balanzas son, tal vez, las que permiten ver esto de modo más directo; porque una varilla suspendida horizontalmente por el centro de un postecillo indica que se encuentra equilibrada y, si cuelga en los extremos unos platillos de igual peso, el equilibrio no se altera. En esta balanza pone el objeto que quiere pesar (harina) y se asegura de que tiene un kilogramo colocando en uno de los platillos el estándar de un kilogramo y en el otro la harina. Si el equilibrio se mantiene, entonces tiene el peso deseado. Si no fuese así, tendría que agregar o quitar harina hasta lograr el equilibrio, o puede 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, pregunta o algo similar, depende de diversas variables. Si selecciona una de esas variables para medirla, también puede escoger varios reactivos que supuestamente la midan, constituyendo una prueba o test con ellos. Usando estadísticas como la correlación y el análisis factorial, es posible ver qué tan efectivamente funciona cada reactivo (pregunta) en relación con la prueba e ir mejorándola de modo que obtenga una medida precisa, y que en efecto mida dicho atributo.

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

Escalas de medición

Como ya se mencionó, medir es asignar números a propiedades de un objeto de acuerdo con reglas, pero las reglas que es posible usar son de muy diferentes tipos. Al asignar números aproveche las propiedades de los sistemas numéricos. Stevens 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 nombres para los objetos. Éste sería el caso de usar el 0 para sexo femenino y el 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.

[†] Medición numérica.
Medición categórica: nominal y ordinal.

El segundo tipo, denominado *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 puede asignar números: 1, 2, 3, ..., etc., a los valores de una actitud. Es decir: “Indique el aprecio que tiene por el presidente de la República: 1. ninguno; 2. poco; 3. regular; 4. mucho”. En esta escala no es fácil 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.

En el tercer tipo, la *escala de intervalo*, no sólo se usa el ordenamiento, sino que establece que las distancias que hay entre número y número son iguales. Por ejemplo, las temperaturas medidas por los termómetros 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 °C o cuando se pasa de 10 a 15 °C.

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 anterior de la temperatura visto, las escalas hacen referencia a un cero que es arbitrario y no refleja la ausencia de la propiedad que se mide. El cero, en la escala Celsius, es el punto en que el hielo se derrite (o el agua se congela). En la escala Fahrenheit, la referencia es el alcohol en vez del agua. Ambos son ceros arbitrarios y por eso las escalas generan números negativos, es decir, hay temperaturas bajo cero. Por lo contrario, la escala Kelvin, sí hace referencia a un cero absoluto que implica la ausencia total de movimiento molecular y, por tanto, de temperatura.

Así, los diferentes tipos de escalas usan ciertas 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 cosas o personas. En este caso, los números sólo sirven como nombres y es indistinto el orden que se use. Aquí sólo se utiliza la propiedad de identidad de los números.

Las escalas ordinales usan la propiedad ordinal, es decir, el hecho de que se siga una secuencia. De este modo, sabe 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 puede 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 al cero.

Limitaciones de las estadísticas por nivel de medida

El uso de la estadística se ve limitado por el tipo de medidas que usa. Por ejemplo, las medidas de razón y de intervalo utilizan los procedimientos más poderosos, llamados *paramétricos*. Existen otros procedimientos que se aplican a los casos de las medidas ordinales y nominales y se les denomina *no paramétricos*. Algunos de ellos utilizan las propiedades de orden como Kolmogorov-Smirnov o la U de Mann-Whitney y otras como la ji cuadrada, que se utilizan para analizar términos de la probabilidad de clases de eventos. Estos procedimientos se verán más adelante con todo detalle; lo importante ahora es percatarse que el tipo de medidas determina el tipo de estadística.

Inferencia estadística y científica

La estadística funciona para hacer inferencias de las distribuciones de las medidas de los fenómenos; partiendo de la suposición de que varias muestras pertenecen a la misma población; y cuando la población a la que pertenecen difiere de ellas, esto se refleja en las muestras.

Para entender mejor esto es preciso decir qué se entiende por población y por muestra. La población es la totalidad de sujetos de una condición que se está observando; es difícil de abarcar y a veces incluye sujetos inaccesibles, como los muertos. Pero incluso los vivos son difíciles de incluir en su totalidad, por ejemplo todos los seres humanos (mayores de 18 años) en el planeta Tierra. Ni siquiera todos los niños menores de 12 años con síndrome de Down. Lo más frecuente es no referirse a sujetos u objetos en sí, sino a alguna dimensión o variable de éstos, como puede ser la estatura, la inteligencia, etcétera.

Una muestra es un subconjunto de la población seleccionado al azar (esto es lo ideal), donde todos los miembros de la población tienen probabilidad de formar parte de ella. Esto en la práctica es muy difícil y costoso, cuando no imposible.

La estadística usa la distribución de probabilidad de los estadísticos muestrales (media, desviación estándar, varianza, etc.). Por ejemplo, la media aritmética, que se verá en el capítulo 2, es una medida global que identifica a un grupo de medidas. Es el valor en el punto central o de equilibrio y que, por tanto, representa al grupo. Las medias aritméticas de muestras aun del mismo tamaño varían entre sí, no siendo exactamente iguales. La frecuencia de estas medias se distribuye de acuerdo con una forma (función de probabilidad) por ejemplo la *t* de Student). Esta distribución es más alta donde se encuentra la verdadera media aritmética de la población y disminuye a medida que se aleja. Esto implica que cuando toma una muestra aleatoria, la media de ésta tiene una mayor probabilidad de ser igual a la de la población, pero hay una probabilidad pequeña de que difieran.

La inferencia estadística se basa en llegar a una conclusión a partir de una probabilidad de que las medias de dos grupos pertenezcan a la misma población. Si la probabilidad es lo bastante baja se concluye que las muestras no pertenecen a dicha población y que por tanto la razón por la cual los grupos difieren (por ejemplo, una manipulación experimental o la procedencia de grupos con características distintas) genera diferentes poblaciones en esa medida.

Por ejemplo, si supone que el alcohol afecta la comprensión de un texto puede usar una medida del grado de comprensión que tiene un lector. Esta medida puede ser una serie de preguntas acerca del texto (que deberán ser tratadas psicométricamente). Ahora, suponga que forma tres grupos de estudiantes de psicología: al primer grupo le da una bebida sin alcohol, al segundo le da una copa de tequila y al tercero dos copas a cada uno de ellos. Les sugiere leer el texto (cada uno tiene una copia del mismo) a continuación les aplica un cuestionario que mide comprensión de lectura. Si los tres grupos provienen de la misma población (de comprensión de dicho texto) por probabilidad las medias aritméticas serían todas parecidas; pero si el consumo de alcohol tuvo un efecto en la comprensión de la lectura, estas medias diferirán. La diferencia (obtenida mediante un análisis de varianza) determina la probabilidad de que éstos pertenezcan a una población homogénea; y cuando la probabilidad es lo suficientemente baja implica que la hipótesis alterna, esto es que los grupos difieren entre sí, no se rechaza.

Este tipo de inferencia, al igual que la inferencia no estadística que se mencionó anteriormente, se debe tomar con la reserva debida. Por experiencia profesional, tal vez, surgió la hipótesis de que el consumo de alcohol afecta la comprensión de textos. Esta hipótesis se pone a prueba en dicha investigación y deberá seguirse contrastando con diferentes muestras, condiciones, sujetos y lecturas.

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 interesa ver su efecto sobre otra u otras variables (variables dependientes) para

Algunas distribuciones de probabilidad importantes

OBJETIVOS GENERALES

Las variables aleatorias discretas se emplean en numerosas aplicaciones prácticas. En este capítulo presentamos tres variables aleatorias discretas importantes: la binomial, la de Poisson y la hipergeométrica. Es frecuente que estas variables aleatorias se usen para describir el número de sucesos de un evento especificado en un número fijo de intentos o una unidad fija de tiempo o espacio. Usted aprendió acerca de variables aleatorias discretas y sus distribuciones de probabilidad. En este capítulo estudiaremos las variables aleatorias continuas y sus distribuciones de probabilidad, así como una variable aleatoria continua muy importante, la normal. Usted aprenderá cómo calcular probabilidades normales y, en ciertas condiciones, cómo usar la distribución normal de probabilidad para aproximar la distribución binomial de probabilidad. En los capítulos siguientes verá la forma en que la distribución normal de probabilidad desempeña un papel central en la inferencia estadística.

ÍNDICE DEL CAPÍTULO

- Introducción (3.1)
- Variables aleatorias discretas y sus distribuciones de probabilidad (3.2)
- La distribución binomial de probabilidad (3.3)
- Distribuciones de probabilidad para variables aleatorias continuas (3.4)
- La distribución de probabilidad de Poisson (3.5)
- La distribución hipergeométrica de probabilidad (3.6)
- La distribución normal de probabilidad (3.7)
- Áreas tabuladas de la distribución normal de probabilidad (3.8)
- La aproximación de la distribución de probabilidad binomial a la normal (3.9)



NECESITO SABER...

Cómo utilizar la tabla 1 para calcular probabilidades binomiales
 Cómo usar la tabla 2 para calcular probabilidades de Poisson
 Cómo utilizar la tabla 3 para calcular probabilidades bajo la curva normal estándar
 Cómo calcular probabilidades binomiales usando la aproximación normal

Un misterio: casos de cáncer cerca de un reactor

¿El reactor nuclear Pilgrim I es responsable del aumento en los casos de cáncer en el área circundante? Surgió una controversia política cuando el Departamento de Salud Pública de Massachusetts encontró un número anormalmente grande de casos en una franja costera de cuatro millas de ancho un poco al norte del reactor nuclear de Plymouth, Massachusetts. El caso práctico, que aparece al final de este capítulo, examina cómo esta pregunta puede contestarse usando una de las distribuciones discretas de probabilidad presentadas aquí.

“¿Va a calificar por curva?”

“Calificar por curva” no necesariamente significa que recibirá una calificación más alta en un examen, ¡aunque a muchos estudiantes les gustaría pensar que sí! Calificar por curva en realidad se refiere a un método para asignar las calificaciones con las letras A, B, C, D o F usando proporciones fijas de las calificaciones correspondientes a cada una de las calificaciones con letra. Una de dichas técnicas para calificar por curva supone que la distribución de las calificaciones es aproximadamente normal y usa estas proporciones.

Calificación por letra	A	B	C	D	F
Proporción de calificaciones	10%	20%	40%	20%	10%

En el caso práctico al final de este capítulo se examinará ésta y otras proporciones asignadas para calificar por curva.

3.1

INTRODUCCIÓN

Es posible encontrar ejemplos de *variables aleatorias discretas* en numerosas situaciones cotidianas y en casi todas las disciplinas académicas. No obstante, hay tres distribuciones discretas de probabilidad que sirven de *modelos* para un gran número de estas aplicaciones. En este capítulo estudiamos las distribuciones de probabilidad binomial, de Poisson e hipergeométrica y discutimos su utilidad en diferentes situaciones físicas.

3.2

VARIABLES ALEATORIAS DISCRETAS Y SUS DISTRIBUCIONES DE PROBABILIDAD

En el capítulo 1 las *variables* se definieron como características que cambian o varían con el tiempo y para diferentes personas u objetos en consideración. Las *variables cuantitativas* generan datos numéricos, en tanto que las *variables cualitativas* generan datos categóricos. No obstante, incluso las variables cualitativas generan datos numéricos si las categorías se codifican numéricamente para formar una escala. Por ejemplo, si se lanza al aire una sola moneda, el resultado cualitativo podría registrarse como “0” si es cara o como “1” si es cruz.

Variables aleatorias

Una variable x valuada numéricamente varía o cambia, dependiendo del resultado particular del experimento que se mida. Por ejemplo, suponga que se tira un dado y se mide x , el número observado en la cara superior. La variable x puede tomar cualquiera de seis valores: 1, 2, 3, 4, 5, 6, dependiendo del resultado *aleatorio* del experimento. Por esta razón, la variable x se conoce como **variable aleatoria**.

Definición Una variable x es **variable aleatoria** si el valor que toma, correspondiente al resultado de un experimento, es una probabilidad o evento aleatorio.

Se consideran numerosos ejemplos de variables aleatorias:

- x = Número de defectos en una pieza de mueble seleccionada al azar
- x = Calificación de examen de aptitud escolar (SAT) para un aspirante universitario seleccionado al azar
- x = Número de llamadas telefónicas recibidas por una línea directa de intervención en crisis durante un periodo seleccionado al azar

Las variables aleatorias cuantitativas se clasifican ya sea como *discretas* o como *continuas*, de acuerdo con los valores que x tome. Es importante distinguir entre variables aleatorias discretas y continuas, porque se usan técnicas diferentes para describir sus distribuciones. Nos concentramos en variables aleatorias discretas en el resto de este capítulo; las variables aleatorias continuas son el tema del capítulo 6.

Distribuciones de probabilidad

En los capítulos 1 y 2 usted aprendió a construir la *distribución de frecuencia relativa* para un conjunto de mediciones numéricas en una variable x . La distribución dio esta información acerca de x :

- ¿Qué valores de x se presentaron?
- ¿Con qué frecuencia se presentó cada valor de x ?

Usted también aprendió a usar la media y desviación estándar para medir el centro y variabilidad de este conjunto de datos.

En este capítulo, definimos la *probabilidad* como el valor limitando de la frecuencia relativa cuando el experimento se repite una y otra vez. Ahora definimos la **distribución de probabilidad** para una variable aleatoria x como la *distribución de frecuencia relativa* construida para toda la población de mediciones.

Definición La **distribución de probabilidad** para una variable aleatoria discreta es una fórmula, tabla o gráfica que da los posibles valores de x , y la probabilidad $p(x)$ asociada con cada valor de x .

Los valores de x representan eventos numéricos mutuamente excluyentes. Sumar $p(x)$ sobre todos los valores de x es equivalente a sumar las probabilidades de todos los eventos simples y por tanto es igual a 1.

REQUISITOS PARA UNA DISTRIBUCIÓN DE PROBABILIDAD DISCRETA

- $0 \leq p(x) \leq 1$
- $\sum p(x) = 1$

EJEMPLO

3.1

Lance al aire dos monedas imparciales y sea x igual al número observado de caras. Encuentre la distribución de probabilidad para x .



Solución Los eventos simples para este experimento con sus respectivas probabilidades se muestran en la tabla 3.1. Como $E_1 = HH$ resulta en dos caras, este evento simple resulta en el valor $x = 2$. Del mismo modo, el valor $x = 1$ se asigna a E_2 , y así sucesivamente.

TABLA 3.1

Eventos simples y probabilidades al lanzar al aire dos monedas

Evento simple	Moneda 1	Moneda 2	$P(E_i)$	x
E_1	H	H	1/4	2
E_2	H	T	1/4	1
E_3	T	H	1/4	1
E_4	T	T	1/4	0

Para cada valor de x , se calcula $p(x)$ al sumar las probabilidades de los eventos simples en ese evento. Por ejemplo, cuando $x = 0$, ocurre el evento simple E_4 , de modo que

$$p(0) = P(E_4) = \frac{1}{4}$$

y cuando $x = 1$,

$$p(1) = P(E_2) + P(E_3) = \frac{1}{2}$$

Los valores de x y sus probabilidades respectivas, $p(x)$, aparecen en la tabla 3.2. Observe que las probabilidades totalizan 1.

TABLA 3.2

Distribución de probabilidad para x (x = número de caras)

x	Eventos simples en x	$p(x)$
0	E_4	1/4
1	E_2, E_3	1/2
2	E_1	1/4
		$\Sigma p(x) = 1$



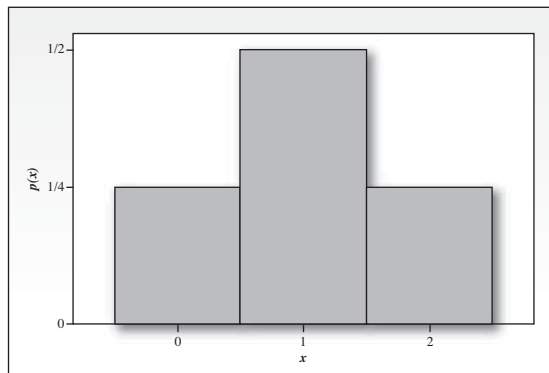
APPLET EN LÍNEA

Lanzar monedas

La distribución de probabilidad de la tabla 3.2 se grafica para formar el **histograma de probabilidad** en la figura 3.1.[†] Los tres valores de la variable aleatoria x se encuentran en el eje horizontal, y las probabilidades $p(x)$ están en el eje vertical (sustituyendo a las frecuencias relativas. Como el ancho de cada barra es 1, el área bajo la barra es la probabilidad de observar el valor particular de x y el área total es igual a 1.

FIGURA 3.1

Histograma de probabilidad para el ejemplo 3.1



La media y desviación estándar para una variable aleatoria discreta

La distribución de probabilidad para una variable aleatoria discreta luce muy semejante a la distribución de frecuencia relativa. La diferencia es que la distribución de frecuencia relativa describe una *muestra* de n mediciones, en tanto que la distribución de probabilidad se construye como un modelo para *toda la población* de mediciones. Así como la media \bar{x} y la desviación estándar s midieron el centro y dispersión de los datos muestrales, usted calculará medidas similares para describir el centro y dispersión de la población.

La media poblacional, que mide el valor promedio de x en la población, también se denomina **valor esperado** de la variable aleatoria x , y se escribe como $E(x)$. Es el valor que *se esperaría* observar en *promedio* si el experimento se repite una y otra vez. La fórmula para calcular la media poblacional es más fácil de entender por ejemplo. Lance otra vez al aire esas dos monedas imparciales, y sea x el número de caras observado. Construimos esta distribución de probabilidad para x :

x	0	1	2
$p(x)$	1/4	1/2	1/4

[†]La distribución de probabilidad de la tabla 3.8 también se puede presentar usando una fórmula, que se da en la sección 3.2.

Suponga que el experimento se repite un gran número de veces, por ejemplo $n = 4\,000\,000$ de veces. Intuitivamente, se esperaría observar alrededor de un millón de ceros, dos millones de números 1 y un millón de números dos. Entonces el valor promedio de x sería igual a

$$\begin{aligned}\frac{\text{Suma de las medidas}}{n} &= \frac{1\,000\,000(0) + 2\,000\,000(1) + 1\,000\,000(2)}{4,000,000} \\ &= \left(\frac{1}{4}\right)(0) + \left(\frac{1}{2}\right)(1) + \left(\frac{1}{4}\right)(2)\end{aligned}$$

Observe que el primer término de esta suma es $(0)p(0)$, el segundo es igual a $(1)p(1)$ y el tercero es $(2)p(2)$. El valor promedio de x , entonces, es

$$\sum xp(x) = 0 + \frac{1}{2} + \frac{2}{4} = 1$$

Este resultado da alguna justificación intuitiva para la definición del valor esperado de una variable aleatoria x discreta.

Definición Sea x una variable aleatoria discreta con distribución de probabilidad $p(x)$. La media o **valor esperado de x** está dada como

$$\mu = E(x) = \sum xp(x)$$

donde los elementos se suman sobre todos los valores de la variable aleatoria x .

Podríamos usar un argumento similar para justificar las fórmulas para la **varianza poblacional** σ^2 y la **desviación estándar de la población** σ . Estas medidas numéricas describen la dispersión o variabilidad de la variable aleatoria usando el “promedio” o “valor esperado” de $(x - \mu)^2$, el cuadrado de las desviaciones de los valores x desde su media μ .

Definición Sea x una variable aleatoria discreta con distribución de probabilidad $p(x)$ y media. La **varianza de x** es

$$\sigma^2 = E[(x - \mu)^2] = \sum (x - \mu)^2 p(x)$$

donde la sumatoria es sobre todos los valores de la variable aleatoria x .[†]

Definición La **desviación estándar σ de una variable aleatoria x** es igual a la raíz cuadrada positiva de su varianza.

EJEMPLO

3.2

Una tienda de electrónica vende un modelo particular de computadora portátil. Hay sólo cuatro computadoras en existencia y la gerente se pregunta cuál será la demanda de hoy para este modelo particular. Ella se entera en el departamento de marketing de que la distribución de probabilidad para x , la demanda diaria para la laptop, es como se muestra en la tabla. Encuentre la media, varianza y desviación estándar de x . ¿Es probable que cinco o más clientes deseen comprar una laptop hoy?

x	0	1	2	3	4	5
$p(x)$.10	.40	.20	.15	.10	.05

[†]Se puede demostrar (prueba omitida) que $\sigma^2 = \sum (x - \mu)^2 p(x) = \sum x^2 p(x) - \mu^2$. Este resultado es análogo a la fórmula de computación para la suma de cuadrados de las desviaciones dadas en el capítulo 1.

Solución La tabla 3.3 muestra los valores de x y $p(x)$, junto con los términos individuales empleados en las fórmulas para μ y σ^2 . La suma de los valores en la tercera columna es

$$\mu = \sum xp(x) = (0)(.10) + (1)(.40) + \dots + (5)(.05) = 1.90$$

en tanto que la suma de los valores en la quinta columna es

$$\begin{aligned} \sigma^2 &= \sum (x - \mu)^2 p(x) \\ &= (0 - 1.9)^2(.10) + (1 - 1.9)^2(.40) + \dots + (5 - 1.9)^2(.05) = 1.79 \end{aligned}$$

y

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.79} = 1.34$$

TABLA 3.3

Cálculos para el ejemplo 3.2

x	$p(x)$	$xp(x)$	$(x - \mu)^2$	$(x - \mu)^2 p(x)$
0	.10	.00	3.61	.361
1	.40	.40	.81	.324
2	.20	.40	.01	.002
3	.15	.45	1.21	.1815
4	.10	.40	4.41	.441
5	.05	.25	9.61	.4805
Totales	1.00	$\mu = 1.90$		$\sigma^2 = 1.79$

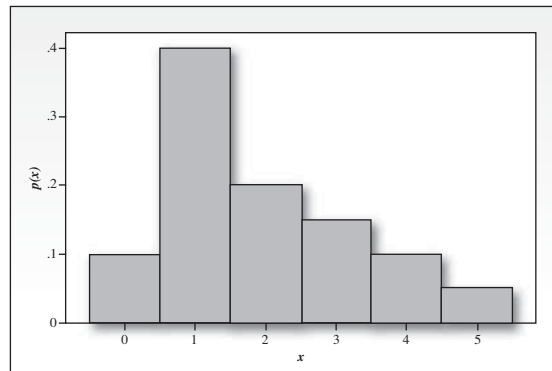
La gráfica de la distribución de probabilidad se muestra en la figura 3.2. Como la distribución tiene más o menos la forma de montículo, aproximadamente 95% de todas las mediciones deben estar a no más de *dos* desviaciones estándar de la media, es decir,

$$\mu \pm 2\sigma \Rightarrow 1.90 \pm 2(1.34) \quad \text{o} \quad -.78 \text{ a } 4.58$$

Como $x = 5$ está fuera de este intervalo, se dice que es improbable que cinco o más clientes deseen comprar una laptop hoy. De hecho, $P(x \geq 5)$ es exactamente .05, o sea 1 vez en 20.

FIGURA 3.2

Distribución de probabilidad para el ejemplo 3.2



EJEMPLO

3.3

En una lotería que se realiza a beneficio de una institución local de caridad, se deben vender 8000 boletos a \$10 cada uno. El premio es un automóvil de \$24000. Si usted compra dos boletos, ¿cuál es su ganancia esperada?

Solución Su ganancia x toma uno de dos valores. O bien perderá \$20 (es decir, su “ganancia” será -\$20) o ganará \$23 980, con probabilidades de 7998/8000 y 2/8000, respectivamente. La distribución de probabilidad para la ganancia x se muestra en la tabla:

x	$p(x)$
-\$20	7998/8000
\$23980	2/8000

La ganancia esperada será

$$\begin{aligned}\mu &= \sum xp(x) \\ &= (-\$20)\left(\frac{7\,998}{8\,000}\right) + (\$23\,980)\left(\frac{2}{8\,000}\right) = -\$14\end{aligned}$$

Recuerde que el valor esperado de x es el promedio de la población teórica que resultaría si en la lotería se repitiera un número infinitamente grande de veces. Si se hiciera esto, su ganancia promedio o esperada por boleto de lotería sería una pérdida de \$14.

EJEMPLO

3.4

Determine la prima anual para una póliza de seguro de \$10 000 que cubre un evento que, en un largo tiempo, ha ocurrido a razón de 2 veces en 100. Sea x igual a la ganancia financiera anual para la compañía de seguros, que resulte de la venta de la póliza, y sea C igual a la prima anual desconocida. Calcule el valor de C tal que la ganancia esperada $E(x)$ iguale a cero. Entonces C es la prima requerida para que haya punto de equilibrio. Para esto, la compañía sumaría los costos administrativos y la utilidad.

Solución El primer paso en la solución es determinar los valores que la ganancia x toma y luego determinar $p(x)$. Si el evento no ocurre durante el año, la compañía de seguros ganará la prima de $x = C$ dólares. Si el evento ocurre, la ganancia será negativa; esto es, la compañía perderá \$10 000 menos la prima de C dólares ya recolectada. Entonces $x = -(10\,000 - C)$ dólares. Las probabilidades asociadas con estos dos valores de x son $98/100$ y $2/100$, respectivamente. La distribución de probabilidad para la ganancia se muestra en la tabla:

$x =$ Ganancia	$p(x)$
C	$98/100$
$-(10\,000 - C)$	$2/100$

Como la compañía desea una prima de seguro C tal que, a largo plazo (para muchas pólizas similares), la ganancia media sea igual a cero, se puede establecer el valor esperado de x igual a cero y despejar C . Entonces

$$\begin{aligned}\mu &= E(x) = \sum xp(x) \\ &= C\left(\frac{98}{100}\right) + [-10\,000 + C]\left(\frac{2}{100}\right) = 0\end{aligned}$$

o

$$\frac{98}{100}C + \frac{2}{100}C - 200 = 0$$

Despejando C de esta ecuación, se obtiene $C = \$200$. Por tanto, si la compañía de seguros cobró una prima anual de \$200, el promedio de ganancia calculada para un gran número de pólizas similares sería igual a cero. La prima real sería igual a \$200 más los costos administrativos y la utilidad.

El método para calcular el valor esperado de x para una variable aleatoria continua es similar a lo que acabamos de hacer, pero en la práctica requiere el uso de cálculo. No obstante, los resultados básicos respecto a expectativas son los mismos para variables aleatorias continuas y discretas. Por ejemplo, sin considerar si x es continua o discreta, $\mu = E(x)$ y $\sigma^2 = E[(x - \mu)^2]$.

3.2 EJERCICIOS

TÉCNICAS BÁSICAS

3.1 ¿Discretas o continuas? Identifique las siguientes variables aleatorias como discretas o continuas:

- El número total de puntos anotados en un juego de fútbol.
- La duración en estante de un medicamento particular.
- La altura de la marea del océano en un lugar determinado.
- Longitud de una perca americana de 2 años de edad.
- El número de choques que casi ocurren de aviones en el aire en un año.

3.2 ¿Discretas o continuas? II Identifique las siguientes variables aleatorias como discretas o continuas:

- Aumento en tiempo de vida alcanzado por un paciente de cáncer como resultado de una cirugía.
- Resistencia a la ruptura (en libras por pulgada cuadrada) de un cable de acero de una pulgada de diámetro.
- Número de venados muertos por año en una reservación estatal de fauna silvestre.
- Número de cuentas vencidas en una tienda de departamentos en un tiempo particular.
- Su presión sanguínea.

3.3 Distribución de probabilidad I Una variable aleatoria x tiene esta distribución de probabilidad:

x	0	1	2	3	4	5
$p(x)$.1	.3	.4	.1	?	.05

- Encuentre $p(4)$.
- Construya un histograma de probabilidad para describir $p(x)$.
- Encuentre μ , σ^2 , y σ .
- Localice el intervalo $\mu \pm 2\sigma$ en el eje x del histograma. ¿Cuál es la probabilidad de que x caiga en este intervalo?
- Si seleccionáramos un número muy grande de valores de x de la población, ¿la mayoría caería en el intervalo $\mu \pm 2\sigma$? Explique.

3.4 Distribución de probabilidad II Una variable aleatoria x puede tomar cinco valores: 0, 1, 2, 3, 4. Una parte de la distribución de probabilidad se muestra aquí:

x	0	1	2	3	4
$p(x)$.1	.3	.3	?	.1

- Encuentre $p(3)$.
- Construya un histograma de probabilidad para $p(x)$.
- Calcule la media poblacional, varianza y desviación estándar.

d. ¿Cuál es la probabilidad de que x sea mayor que 2?

e. ¿Cuál es la probabilidad de que x sea 3 o menor?

3.5 Visitas de tienda Represente con x el número de veces que un cliente acude a una tienda en un periodo de una semana. Suponga que ésta es la distribución de probabilidad de x :

x	0	1	2	3
$p(x)$.1	.4	.4	.1

Encuentre el valor esperado de x , el número promedio de veces que un cliente acude a la tienda.

3.6 Si lanza un par de dados, la suma T de los números que aparecen en las caras superiores de los dados puede asumir el valor de un entero en el intervalo $2 \leq T \leq 12$.

- Encuentre la distribución de probabilidad para T . Presente esta distribución de probabilidad en una tabla.
- Construya un histograma de probabilidad para $P(T)$. ¿Cómo describiría la forma de esta distribución?

APLICACIONES

3.7 ¿Mensajes de texto mientras se conduce? La proporción de adultos (18 años o más) que admiten enviar mensajes de texto mientras conducen es 47%.¹ Suponga que selecciona al azar tres conductores adultos y les pregunta si envían mensajes de texto mientras conducen.

- Encuentre la distribución de probabilidad para x , el número de conductores en la muestra que admiten enviar mensajes de texto mientras conducen.
- Construya un histograma de probabilidad para $p(x)$.
- ¿Cuál es la probabilidad de que exactamente uno de los tres conductores envíe mensajes de texto mientras conduce?
- ¿Cuáles son la media y la desviación estándar poblacionales para la variable aleatoria x ?

3.8 ¿Sesgo de género? Una compañía tiene cinco solicitantes para dos puestos de trabajo: dos mujeres y tres hombres. Suponga que los cinco solicitantes son igualmente calificados y que no hay preferencia para elegir cualquier género. Sea x igual al número de mujeres seleccionadas para ocupar los dos puestos de trabajo.

- Encuentre $p(x)$.
- Construya un histograma de probabilidad para x .

3.9 Perforación de pozos petroleros La experiencia ha demostrado que, en promedio, sólo uno de cada 10 pozos produce petróleo. Sea x el número de perforaciones hasta el primer éxito (se encuentra petróleo). Suponga que las perforaciones representan eventos independientes.

- a. Encuentre $p(1)$, $p(2)$ y $p(3)$.
 b. Dé una fórmula para $p(x)$.
 c. Grafique $p(x)$.

3.10 ¿Alguien juega tenis? Dos jugadores profesionales de tenis, A y B , están programados para jugar un partido: el ganador del partido es el primero en ganar tres sets de un total de cinco. El evento en que A gane algún set es independiente del evento de que gane cualquier otro y la probabilidad de que gane cualquier set es igual a $.6$. Sea x igual al número total de sets del partido; esto es, $x = 3, 4$ o 5 . Encuentre $p(x)$.

3.11 Tenis, otra vez En el ejercicio 3.10 encontró la distribución de probabilidad para x , el número de sets requeridos para jugar un partido como el mejor de cinco sets, dado que la probabilidad de que A gane cualquier set —llamemos a esto $P(A)$ — es $.6$.

- a. Encuentre el número esperado de sets necesario para completar el partido para $P(A) = .6$.
 b. Encuentre el número esperado de sets necesario para completar el partido cuando los jugadores sean de igual capacidad, es decir, $P(A) = .5$.
 c. Encuentre el número esperado de sets necesario para completar el partido cuando los jugadores difieran en mucho en capacidad, es decir, por ejemplo, $P(A) = .9$.
 d. ¿Cuál es la relación entre $P(A)$ y $E(x)$, el número esperado de sets requeridos para completar el partido?

3.12 Prueba de la FDA La duración máxima de patente para un nuevo medicamento es 17 años. Restando el tiempo requerido por la FDA para probar y aprobar el medicamento proporciona la vida real de patente del medicamento, es decir, el tiempo que una compañía tiene para recuperar costos de investigación y desarrollo y obtener una utilidad. Suponga que la distribución de tiempos de vida de patente para nuevos medicamentos es como se muestra a continuación:

Años, x	3	4	5	6	7	8
$p(x)$.03	.05	.07	.10	.14	.20
Años, x	9	10	11	12	13	
$p(x)$.18	.12	.07	.03	.01	

- a. Encuentre el número esperado de años de vigencia de patente para un nuevo medicamento.
 b. Encuentre la desviación estándar de x .
 c. Encuentre la probabilidad de que x caiga en el intervalo $\mu \pm 2\sigma$.

3.13 Descanso para tomar café La mayoría de las personas que bebe café se da un poco de tiempo para hacerlo y muchas toman más de un descanso al día. La tabla siguiente, adaptada de un Snapshot de *USA Today*, muestra la distribución de probabilidad para x , el número de descansos por día que se dan quienes beben café.²

x	0	1	2	3	4	5
$p(x)$.28	.37	.17	.12	.05	.01

- a. ¿Cuál es la probabilidad de que una persona que toma café, seleccionada al azar, no se dé un descanso para tomar café durante el día?
 b. ¿Cuál es la probabilidad de que una persona que toma café, seleccionada al azar, se dé más de dos descansos para tomar café durante el día?
 c. Calcule la media y desviación estándar para la variable aleatoria x .
 d. Encuentre la probabilidad de que x caiga en el intervalo $\mu \pm 2\sigma$.

3.14 Actuarios El director de una empresa está considerando tomar una póliza de seguro para cubrir posibles pérdidas en que incurriría al vender un nuevo producto. Si el producto es un completo fracaso, el director incurrirá en una pérdida de \$800 000; si es sólo un éxito moderado, incurrirá en una pérdida de \$250 000. Los actuarios de seguros han determinado que las probabilidades de que ese producto sea un fracaso o sólo tenga un éxito moderado son $.01$ y $.05$, respectivamente. Suponiendo que el director de la empresa esté dispuesto a ignorar todas las otras posibles pérdidas, ¿qué prima debería cobrar la compañía de seguros por una póliza para no tener pérdida ni ganancia?

LA DISTRIBUCIÓN BINOMIAL DE PROBABILIDAD

3.3

El experimento de lanzar al aire una moneda es un ejemplo sencillo de una importante variable aleatoria discreta llamada **variable aleatoria binomial**. Muchos experimentos prácticos resultan en datos similares a que salgan cara o cruz al tirar la moneda. Por ejemplo, considere las encuestas políticas que se emplean para predecir las preferencias de los votantes en elecciones. Cada votante entrevistado se puede comparar a una moneda porque es probable que el votante esté a favor de nuestro candidato (una “cara”) o no (una “cruz”). Casi siempre, la proporción de votantes que están a favor de nuestro candidato no es igual a $1/2$, es decir, la moneda no es imparcial. De hecho, la encuesta está diseñada exactamente para determinar la proporción de votantes que están a favor de nuestro candidato.

Veamos aquí algunas otras situaciones semejantes al experimento de lanzar al aire una moneda:

- Un sociólogo está interesado en la proporción de maestros de escuela primaria que sean varones.
- Una comerciante en bebidas gaseosas está interesada en la proporción de consumidores de refrescos de cola que prefieren su marca.
- Un genetista está interesado en la proporción de la población que posee un gen vinculado a la enfermedad de Alzheimer.

Cada persona muestreada es análoga a lanzar al aire una moneda, pero la probabilidad de una “cara” no es necesariamente igual a $1/2$. Aun cuando estas situaciones tienen diferentes objetivos prácticos, todas exhiben las características comunes del **experimento binomial**.

Definición Un **experimento binomial** tiene estas cinco características:

1. El experimento consiste en n intentos idénticos.
2. Cada intento produce uno de dos resultados. Por falta de un mejor nombre, el resultado uno se llama éxito, S, y el otro, fracaso, F.
3. La probabilidad de éxito en un solo intento es igual a p y es el mismo de un intento tras otro. La probabilidad de fracaso es igual a $(1 - p) = q$.
4. Los intentos son independientes.
5. Estamos interesados en x , el número de éxitos observado durante los n intentos, para $x = 0, 1, 2, \dots, n$.

EJEMPLO

3.5

Suponga que hay alrededor de un millón de adultos en un condado y una proporción desconocida p están a favor de limitar el periodo de función de políticos. Se elegirá una muestra de mil adultos en forma tal que cada uno, del millón de adultos, tenga igual probabilidad de ser seleccionado y a cada uno se le preguntará si él o ella está a favor de limitar el periodo. (El objetivo final de esta encuesta es estimar la proporción desconocida p , un problema que veremos en el capítulo 8.) ¿Este experimento es binomial?

Solución ¿El experimento tiene las cinco características binomiales?

1. Un “intento” es la selección de un solo adulto de entre el millón de electores de la ciudad. Esta muestra consta de $n = 1000$ intentos idénticos.
2. Como cada elector estará o no a favor de limitar el periodo, hay dos resultados que representan los “éxitos” y “fracasos” del experimento binomial.[†]

[†]Aun cuando es tradicional que los dos posibles resultados de un intento se denominen “éxito” y “fracaso”, el resultado llamado “éxito” no necesita ser visto como éxito en el uso ordinario de la palabra.

3. La probabilidad de éxito, p , es la probabilidad de que un adulto esté a favor del límite del periodo. ¿Esta probabilidad sigue igual para cada uno de los electores de la muestra? Para todos los fines prácticos, la respuesta es sí. Por ejemplo, si 500 000 electores de la población están a favor de limitar el periodo, entonces la probabilidad de un “éxito” cuando se elija al primer elector es $500\,000/1\,000\,000 = 1/2$. Cuando se elija al segundo elector, la probabilidad p cambia ligeramente, dependiendo de la primera selección. Esto es, habrá 499 999 o 500 000 éxitos que queden entre los 999 999 adultos. En cualquiera de estos casos, p es todavía más o menos igual a $1/2$.
4. La independencia de los intentos está garantizada debido al grupo grande de adultos del que se toma la muestra. La probabilidad de que un elector esté a favor de limitar el periodo no cambia, dependiendo de las respuestas de las personas previamente elegidas.
5. La variable aleatoria x es el número de electores de la muestra que estén a favor de limitar el periodo.

Debido a que el estudio satisface las cinco características razonablemente bien, para todos los fines prácticos se le considera un experimento binomial.

EJEMPLO

3.6

Un paciente llena una receta para un régimen de 10 días de dos píldoras diarias. Sin que lo sepa el farmacéutico ni el paciente, las 20 pastillas están formadas por 18 píldoras del medicamento prescrito y dos píldoras que son el equivalente genérico del medicamento prescrito. El paciente selecciona dos píldoras al azar para la dosis del primer día. Si verificamos la selección y registramos el número de píldoras que son genéricas, ¿es éste un experimento binomial?

Solución Verifique de nuevo el procedimiento de muestra para las características de un experimento binomial.

1. Un “intento” es la selección de una píldora de entre las 20 de la receta. Este experimento consta de $n = 2$ intentos.
2. Cada intento resulta en uno de dos resultados. O bien, la píldora es genérica (llame “éxito” a esto) o no lo es (un “fracaso”).
3. Como las píldoras de una botella de receta se consideran “mezcladas” al azar, la probabilidad incondicional de sacar una píldora genérica en un intento determinado sería $2/20$.
4. La condición de independencia entre intentos *no* está satisfecha, porque la probabilidad de sacar una píldora genérica en el segundo intento depende del primer intento. Por ejemplo, si la primera píldora sacada es genérica entonces hay sólo una píldora genérica en las restantes 19. Por tanto,

$$P(\text{genérica en intento 2} | \text{genérica en intento 1}) = 1/19$$

Si la primera selección *no* resulta en una píldora genérica, entonces hay todavía dos píldoras genéricas en las restantes 19, y la probabilidad de un “éxito” (una píldora genérica) cambia a

$$P(\text{genérica en el intento 2} | \text{no genérica en el intento 1}) = 2/19$$

Por tanto, los intentos son dependientes y el muestreo no representa un experimento binomial.

Considere la diferencia entre estos dos ejemplos. Cuando la muestra (los n intentos idénticos) vinieron de una población grande, la probabilidad de éxito p siguió siendo más o menos la misma de un intento a otro. Cuando el tamaño poblacional N era pequeño, la probabilidad de éxito p cambió en forma considerable de un intento a otro, y el experimento *no fue* binomial.

REGLA PRÁCTICA

Si el tamaño muestral es grande respecto al tamaño poblacional, en particular si $n/N \geq .05$, entonces el experimento resultante no es binomial.

Si lanzamos al aire dos monedas “honestas” y construimos la distribución de probabilidad para x , el número de caras, un experimento binomial con $n = 2$ y $p = .5$. La distribución binomial general de probabilidad se construye en la misma forma, pero el procedimiento se complica cuando n se hace grande. Afortunadamente, las probabilidades $p(x)$ siguen un modelo general. Esto nos permite usar una sola fórmula para hallar $p(x)$ para cualquier valor dado de x .

LA DISTRIBUCIÓN BINOMIAL DE PROBABILIDAD

Un experimento binomial consta de n intentos idénticos con probabilidad p de éxito en cada intento. La probabilidad de k éxitos en n intentos es

$$P(x = k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

para valores de $k = 0, 1, 2, \dots, n$. El símbolo C_k^n es igual a,

$$\frac{n!}{k!(n-k)!}$$

donde $n! = n(n-1)(n-2) \dots (2)(1)$ y $0! \equiv 1$.

Las fórmulas generales para μ , σ^2 y σ se usan para obtener las siguientes fórmulas más sencillas para la media y la desviación estándar binomiales.

MEDIA Y DESVIACIÓN ESTÁNDAR PARA LA VARIABLE ALEATORIA BINOMIAL

La variable aleatoria x , el número de éxitos en n intentos, tiene una distribución de probabilidad con este centro y dispersión:

$$\begin{array}{ll} \text{Media:} & \mu = np \\ \text{Varianza:} & \sigma^2 = npq \\ \text{Desviación estándar:} & \sigma = \sqrt{npq} \end{array}$$

EJEMPLO

3.7

Encuentre $P(x = 2)$ para una variable aleatoria binomial con $n = 10$ y $p = .1$.

Solución $P(x = 2)$ es la probabilidad de observar 2 éxitos y 8 fracasos en una secuencia de 10 intentos. Se podrían observar 2 éxitos primero, seguidos de 8 fracasos consecutivos:

E, E, F, F, F, F, F, F, F, F

Como p es la probabilidad de éxito y q es la probabilidad de fracaso, esta secuencia particular tiene probabilidad

$$ppqqqqqqqq = p^2q^8$$

Sin embargo, puede también resultar muchas otras secuencias en $x = 2$ éxitos. La fórmula binomial utiliza C_2^{10} para contar el número de secuencias y da la probabilidad exacta cuando se usa la fórmula binomial con $k = 2$:

TIP MI CONSEJO

$n! = n(n-1)(n-2) \dots (2)(1)$
 Por ejemplo,
 $5! = 5(4)(3)(2)(1) = 120$
 y $0! \equiv 1$

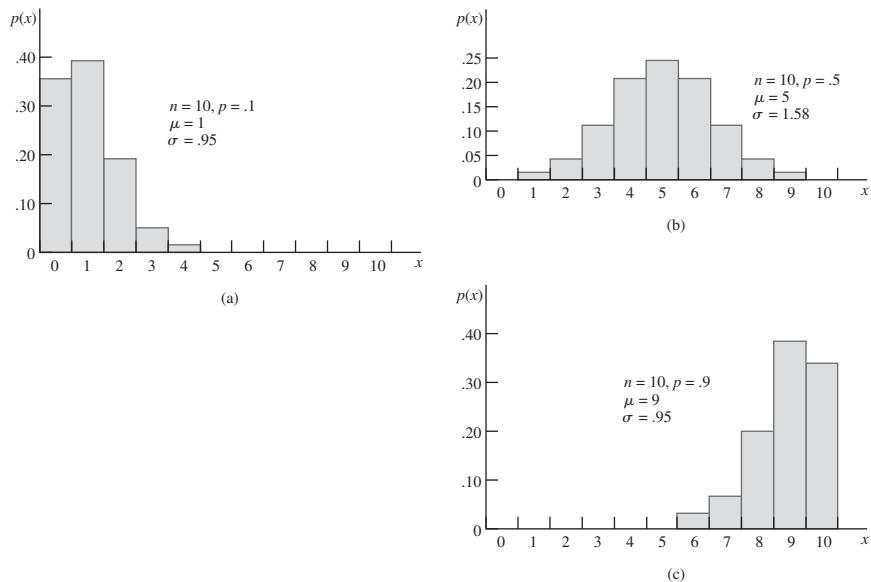
$$\begin{aligned}
 P(x = 2) &= C_2^{10}(.1)^2(.9)^{10-2} \\
 &= \frac{10!}{2!(10-2)!}(.1)^2(.9)^8 = \frac{10(9)}{2(1)}(.01)(.430467) = .1937
 \end{aligned}$$

Se podría repetir el procedimiento del ejemplo 3.7 para cada valor de x ($0, 1, 2, \dots, 10$) y encontrar todos los valores de $p(x)$ necesarios para construir un histograma de probabilidad para x . Éste sería un trabajo largo y tedioso, pero la gráfica resultante se vería como la figura 3.3a). Se puede verificar la altura de la barra para $x = 2$ y encontrar $p(2) = P(x = 2) = .1937$. La gráfica está sesgada a la derecha; esto es, casi todo el tiempo se observarían valores pequeños de x . La media o “punto de equilibrio” está alrededor de $x = 1$; de hecho, se utiliza la fórmula para hallar la media exacta:

$$\mu = np = 10(.1) = 1$$

Las figuras 3.3b) y 3.3c) muestran las otras dos distribuciones binomiales con $n = 10$ pero con diferentes valores de p . Vea las formas de estas distribuciones. Cuando $p = .5$, la distribución es exactamente simétrica alrededor de la media, $\mu = np = 10(.5) = 5$. Cuando $p = .9$, la distribución es la “imagen espejo” de la distribución para $p = .1$ y está sesgada a la izquierda.

FIGURA 3.3
Distribuciones de probabilidad binomial



EJEMPLO

3.8

Durante un largo periodo se ha observado que un jugador profesional de baloncesto puede hacer un tiro libre en un intento determinado con una probabilidad igual a $.8$. Suponga que él lanza cuatro tiros libres.

1. ¿Cuál es la probabilidad de que enceste exactamente dos tiros libres?
2. ¿Cuál es la probabilidad de que enceste al menos un tiro libre?

Solución Un “intento” es un solo tiro libre y se define un “éxito” como una canasta y un “fracaso” como una falla, de modo que $n = 4$ y $p = .8$. Si se supone que la probabilidad del jugador de encestar el tiro libre no cambia de un tiro a otro, entonces el número x de veces que enceste el tiro libre es una *variable aleatoria binomial*.

$$\begin{aligned}
 1. \quad P(x = 2) &= C_2^4(.8)^2(.2)^2 \\
 &= \frac{4!}{2!2!}(.64)(.04) = \frac{4(3)(2)(1)}{2(1)(2)(1)}(.64)(.04) = .1536
 \end{aligned}$$

La probabilidad es .1536 de que enceste exactamente dos tiros libres.

$$\begin{aligned}
 2. \quad P(\text{al menos uno}) &= P(x \geq 1) = p(1) + p(2) + p(3) + p(4) \\
 &= 1 - p(0) \\
 &= 1 - C_0^4(.8)^0(.2)^4 \\
 &= 1 - .0016 = .9984
 \end{aligned}$$

Aun cuando usted podría calcular $P(x = 1)$, $P(x = 2)$, $P(x = 3)$ y $P(x = 4)$ para hallar esta probabilidad, usar el complemento del evento hace más fácil su trabajo; es decir,

$$P(x \geq 1) = 1 - P(x < 1) = 1 - P(x = 0)$$

¿Considera alguna razón por la que su suposición de intentos independientes podría ser errónea? Si el jugador aprende de su intento previo (es decir, ajusta su tiro de acuerdo con su último intento), entonces su probabilidad p de encestar el tiro libre puede cambiar, posiblemente aumentar, de un tiro a otro. Los intentos *no* serían independientes y el experimento *no* sería binomial.

TIP MI CONSEJO

Use la tabla 1 del Anexo en lugar de la fórmula binomial siempre que sea posible. ¡Esta es una forma más fácil!

Calcular probabilidades binomiales puede ser tedioso incluso para valores relativamente pequeños de n . Cuando n se hace grande, es casi imposible sin ayuda de una calculadora o computadora. Por fortuna, tenemos estas dos herramientas. Las tablas de **probabilidades binomiales acumulativas** generadas por computadora se presentan en la tabla 1 del Anexo, para valores de n que van de 2 a 25 y para valores seleccionados de p . Estas probabilidades también pueden ser generadas si se usa el *MINITAB*, *MS Excel* o los applets Java en el sitio web CourseMate.

Las probabilidades binomiales *acumulativas* difieren de las probabilidades binomiales *individuales* que se calcularon con la fórmula binomial. Una vez que usted encuentre la columna de probabilidades para los valores correctos de n y p en la tabla 1, el renglón marcado como k proporciona la suma de todas las probabilidades binomiales de $x = 0$ a $x = k$. La tabla 3.5 muestra parte de la tabla 1 para $n = 5$ y $p = .6$. Si se observa el renglón marcado $k = 3$, se encuentra

$$P(x \leq 3) = p(0) + p(1) + p(2) + p(3) = .663$$

TABLA 3.4 Parte de la tabla 1 del Anexo para $n = 5$

k	p													k
	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	
0	—	—	—	—	—	—	—	.010	—	—	—	—	—	0
1	—	—	—	—	—	—	—	.087	—	—	—	—	—	1
2	—	—	—	—	—	—	—	.317	—	—	—	—	—	2
3	—	—	—	—	—	—	—	.663	—	—	—	—	—	3
4	—	—	—	—	—	—	—	.922	—	—	—	—	—	4
5	—	—	—	—	—	—	—	1.000	—	—	—	—	—	5

Si la probabilidad que necesita calcular no está en esta forma, deberá considerar una forma para reescribir su probabilidad y hacer uso de las tablas.

EJEMPLO

3.9

Use la tabla binomial acumulativa para $n = 5$ y $p = .6$ para hallar las probabilidades de estos eventos:

1. Exactamente tres éxitos
2. Tres o más éxitos

Solución

1. Si encuentra $k = 3$ en la tabla 3.5, el valor presentado es

$$P(x \leq 3) = p(0) + p(1) + p(2) + p(3)$$

Como usted desea sólo $P(x = 3) = p(3)$, debe restar la probabilidad no deseada:

$$P(x \leq 2) = p(0) + p(1) + p(2)$$

que se encuentra en la tabla 3.5 con $k = 2$. Entonces

$$\begin{aligned} P(x = 3) &= P(x \leq 3) - P(x \leq 2) \\ &= .663 - .317 = .346 \end{aligned}$$

2. Para hallar $P(\text{tres o más éxitos}) = P(x \geq 3)$ usando la tabla 3.1, se debe emplear el complemento del evento de interés. Escriba

$$P(x \geq 3) = 1 - P(x < 3) = 1 - P(x \leq 2)$$

Puede hallar $P(x \leq 2)$ en la tabla 3.5 con $k = 2$. Entonces

$$\begin{aligned} P(x \geq 3) &= 1 - P(x \leq 2) \\ &= 1 - .317 = .683 \end{aligned}$$

EJEMPLO

3.10

Consulte el ejemplo 3.5 y la variable aleatoria binomial x con $n = 5$ y $p = .6$. Use la tabla binomial acumulativa para encontrar las probabilidades binomiales restantes, $p(0)$, $p(1)$, $p(2)$, $p(4)$ y $p(5)$. Construya el histograma de probabilidad para la variable aleatoria x y describa su forma y ubicación.

Solución

1. Puede encontrar $P(x = 0)$ directamente de la tabla 3.5 con $k = 0$. Esto es, $p(0) = .010$.
2. Las otras probabilidades pueden encontrarse restando entradas sucesivas en la tabla 3.5. Entonces

$$P(x = 1) = P(x \leq 1) - P(x = 0) = .087 - .010 = .077$$

$$P(x = 2) = P(x \leq 2) - P(x \leq 1) = .317 - .087 = .230$$

$$P(x = 4) = P(x \leq 4) - P(x \leq 3) = .922 - .663 = .259$$

$$P(x = 5) = P(x \leq 5) - P(x \leq 4) = 1.000 - .922 = .078$$

El histograma de probabilidad se muestra en la figura 3.4. La distribución tiene una forma relativamente de montículo, con un centro alrededor de 3.



PROBABILIDAD Y ESTADÍSTICA

para las ciencias sociales
del comportamiento y la salud

Probabilidad y estadística para las ciencias sociales del comportamiento y la salud, primera edición en español, es una versión adaptada del libro *Introducción a la probabilidad y estadística*, decimocuarta edición, de William Mendenhall III, Robert J. Beaver y Barbara M. Beaver. En ella se han incluido temas que son indispensables para los profesionales de las ciencias sociales del comportamiento y la salud que, ya sea en su trabajo cotidiano o en la realización de estudios e investigaciones, deben manejar, analizar e interpretar grandes cantidades de datos: estadística y probabilidad dentro del contexto de la investigación, medición estándar, distribuciones, variables aleatorias discretas (binomial, de Poisson e hipergeométrica), estimación estadística de muestras, métodos para estimar los parámetros poblacionales y para datos categóricos, entre otros.

La obra conserva de la versión original algunas características importantes, pero en esta ocasión enfocadas en el estudio de la estadística desde la perspectiva de las ciencias sociales.



Visite nuestro sitio en <http://latinoamerica.cengage.com>

ISBN-13: 978-607-526-309-0
ISBN-10: 607-526-309-8



9 786075 263090