



ESTADÍSTICA PARA NEGOCIOS Y ECONOMÍA

12ª EDICIÓN

Anderson · Sweeney · Williams · Camm · Cochran

ESTADÍSTICA
PARA NEGOCIOS
Y ECONOMÍA
12a. ed.

ESTADÍSTICA PARA NEGOCIOS Y ECONOMÍA 12a. ed.

David R. Anderson

University of Cincinnati

Dennis J. Sweeney

University of Cincinnati

Thomas A. Williams

Rochester Institute of Technology

Jeffrey D. Camm

University of Cincinnati

James J. Cochran

Louisiana Tech University

Traducción

Lorena Peralta Rosales

Traductor profesional

Revisión técnica

María de Guadalupe Arroyo Satisteban

Academia de Matemáticas ECEE

Universidad Panamericana

Ignacio García Juárez

Academia de Matemáticas ECEE

Universidad Panamericana

José Cruz Ramos Báez

Academia de Matemáticas ECEE

Universidad Panamericana

Iren Castillo Saldaña

Academia de Matemáticas ECEE

Universidad Panamericana



Estadística para negocios y economía**Décimosegunda edición**

David R. Anderson, Dennis J. Sweeney,
Thomas A. Williams, Jeffrey D. Camm,
James J. Cochran

Presidente de Cengage Learning**Latinoamérica**

Fernando Valenzuela Migoya

Director Editorial para Latinoamérica

Ricardo H. Rodríguez

Gerente de Adquisiciones para**Latinoamérica**

Claudia C. Garay Castro

Gerente de Manufactura para**Latinoamérica**

Antonio Mateos Martínez

Gerente Editorial de Contenidos en Español

Pilar Hernández Santamarina

Gerente de Proyectos Especiales

Luciana Rabuffetti

Coordinador de Manufactura

Rafael Pérez González

Editor

Javier Reyes Martínez

Diseño de portada

Lilia Palomino Viveros
RED Studio

Imagen de portada

©Shutterstock
254655451
Zadorozhnyi Viktor

Composición tipográfica

Heriberto Gachúz Chávez

© D.R. 2016 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.

ExamView® es una marca registrada de elnstruction Corp. Windows es una marca registrada de Microsoft Corporation; se utiliza aquí bajo licencia. Macintosh y Power Macintosh son marcas comerciales registradas de Apple Computer, Inc. utilizadas en este documento bajo licencia.

© 2008 Cengage Learning. Todos Los Derechos Reservados. *Microsoft Excel*® es una marca comercial registrada de Microsoft Corporation. © 2014 Microsoft.

Traducido del libro:

Statistics for Business and Economics
Twelfth edition

David R. Anderson, Dennis J. Sweeney,
Thomas A. Williams, Jeffrey D. Camm,
James J. Cochran

Publicado en inglés por

South-Western, Cengage Learning © 2014
ISBN 978-1-133-27453-7

Datos para catalogación bibliográfica:
Anderson, David R., Sweeney, Dennis J.
Williams, Thomas A., Camm, Jeffrey D.
Cochran, James J.

Estadística para negocios y economía
Décimosegunda edición
ISBN-13: 978-607-522-515-9

Visite nuestro sitio en:

<http://latinoamerica.cengage.com>

Contenido

Prefacio xxv
Acerca de los autores xxix

Capítulo 1 Los datos y la estadística 1

Estadística en la práctica. Bloomberg Businessweek 2

1.1 Aplicaciones en negocios y economía 3

Contabilidad 3
Finanzas 4
Marketing 4
Producción 4
Economía 4
Sistemas de información 5

1.2 Datos 5

Elementos, variables y observaciones 5
Escala de medición 7
Datos categóricos y cuantitativos 8
Datos de corte transversal y de series de tiempo 8

1.3 Fuentes de datos 11

Fuentes existentes 11
Estudios estadísticos 12
Errores en la adquisición de datos 14

1.4 Estadística descriptiva 14

1.5 Inferencia estadística 16

1.6 Computadoras y análisis estadístico 18

1.7 Minería de datos 18

1.8 Lineamientos éticos para la práctica de la estadística 19

Resumen 21

Glosario 21

Ejercicios complementarios 22

Apéndice Introducción a StatTools 29

Capítulo 2 Estadística descriptiva: Presentaciones tabulares y gráficas 33

Estadística en la práctica. Colgate-Palmolive Company 34

2.1 Resumen de datos para una variable cualitativa o categórica 35

Distribución de frecuencia 35
Distribuciones de frecuencia relativa y frecuencia porcentual 36

Gráficas de barras y circulares	36
2.2 Resumen de datos para una variable cuantitativa	42
Distribución de frecuencia	42
Distribuciones de frecuencia relativa y de frecuencia porcentual	43
Diagrama de puntos	44
Histograma	44
Distribuciones acumuladas	46
Diagrama de tallo y hoja	47
2.3 Cómo resumir datos para dos variables por medio de tablas	55
Tabulación cruzada	55
La paradoja de Simpson	58
2.4 Cómo resumir datos para dos variables por medio de representaciones gráficas	64
Diagrama de dispersión y línea de tendencia	64
Gráficas de barras apiladas y de barras agrupadas	65
2.5 Visualización de datos: Mejores prácticas en la elaboración de representaciones gráficas efectivas	70
Cómo elaborar representaciones gráficas efectivas	71
Selección del tipo de representación gráfica	72
Data dashboards o tableros de datos	72
La visualización de datos en la práctica: El zoológico y jardín botánico de Cincinnati	74
Resumen	77
Glosario	78
Fórmulas clave	79
Ejercicios complementarios	79
Caso práctico 1 Pelican Stores	84
Caso práctico 2 Industria del cine	85
Apéndice 2.1 Cómo usar Minitab para presentaciones tabulares y gráficas	86
Apéndice 2.2 Cómo usar Excel para presentaciones tabulares y gráficas	88
Apéndice 2.3 Cómo usar StatTools para presentaciones tabulares y gráficas	98
Capítulo 3 Estadística descriptiva: Medidas numéricas	99
<i>Estadística en la práctica. Small Fry Design</i>	100
3.1 Medidas de posición o localización	101
Media	101
Media ponderada	103
Mediana	104
Media geométrica	106
Moda	107
Percentiles	108
Cuartiles	109

3.2	Medidas de variabilidad	116
	Rango	116
	Rango intercuartílico o intercuartil	117
	Varianza	117
	Desviación estándar	118
	Coefficiente de variación	119
3.3	Medidas de la forma de la distribución, posición relativa y detección de observaciones atípicas	123
	Forma de la distribución	123
	Valor z	123
	Teorema de Chebyshev	125
	Regla empírica	126
	Detección de observaciones atípicas	127
3.4	Resúmenes de cinco números y diagramas de caja	130
	Resumen de cinco números	131
	Diagrama de caja	131
3.5	Medidas de asociación entre dos variables	136
	Covarianza	136
	Interpretación de la covarianza	138
	Coefficiente de correlación	140
	Interpretación del coeficiente de correlación	141
3.6	Data dashboards o tableros de datos: Incorporación de medidas numéricas para mejorar su eficacia	145
	Resumen	149
	Glosario	149
	Fórmulas clave	150
	Ejercicios complementarios	152
	<i>Caso práctico 1</i> Pelican Stores	157
	<i>Caso práctico 2</i> Industria del cine	158
	<i>Caso práctico 3</i> Escuelas de negocios de Asia-Pacífico	159
	<i>Caso práctico 4</i> Transacciones del sitio web de Heavenly Chocolates	161
	<i>Caso práctico 5</i> Las poblaciones de elefantes africanos	162
	Apéndice 3.1 Estadística descriptiva usando Minitab	163
	Apéndice 3.2 Estadística descriptiva usando Excel	165
	Apéndice 3.3 Estadística descriptiva usando StatTools	167

Capítulo 4 Introducción a la probabilidad **169**

Estadística en la práctica. La probabilidad al rescate **170**

4.1	Experimentos, reglas de conteo y asignación de probabilidades	171
	Reglas de conteo, combinaciones y permutaciones	172
	Asignación de probabilidades	176
	Probabilidades para el proyecto de KP&L	178
4.2	Eventos y sus probabilidades	181

4.3 Algunas relaciones básicas de probabilidad 185

Complemento de un evento 185

Ley de la adición 186

4.4 Probabilidad condicional 192

Eventos independientes 195

Ley de la multiplicación 195

4.5 Teorema de Bayes 200

Método tabular 203

Resumen 206

Glosario 206

Fórmulas clave 207

Ejercicios complementarios 208

Caso práctico Jueces del condado de Hamilton 212**Capítulo 5 Distribuciones discretas de probabilidad 215***Estadística en la práctica. Citibank* 216**5.1 Variables aleatorias 217**

Variables aleatorias discretas 217

Variables aleatorias continuas 218

5.2 Desarrollo de distribuciones discretas de probabilidad 220**5.3 Valor esperado y varianza 225**

Valor esperado 225

Varianza 225

5.4 Distribuciones bivariantes, covarianza y portafolios financieros 230

Distribución discreta de probabilidad bivariante empírica 230

Aplicaciones financieras 233

Resumen 236

5.5 Distribución binomial 239

Un experimento binomial 240

El problema de Martin Clothing Store 241

Cómo usar tablas de probabilidades binomiales 245

Valor esperado y varianza de la distribución binomial 246

5.6 Distribución de Poisson 250

Un ejemplo con intervalos de tiempo 250

Un ejemplo con intervalos de longitud o distancia 252

5.7 Distribución hipergeométrica 253

Resumen 257

Glosario 258

Fórmulas clave 258

Ejercicios complementarios 260

Apéndice 5.1 Distribuciones discretas de probabilidad con Minitab 263

Apéndice 5.2 Distribuciones discretas de probabilidad con Excel 263

Capítulo 6 Distribuciones continuas de probabilidad 265

Estadística en la práctica. Procter & Gamble 266

6.1 Distribución uniforme 267

El área como medida de la probabilidad 268

6.2 Distribución normal 271

Curva normal 271

Distribución normal estándar 273

Cálculo de probabilidades para cualquier distribución normal 278

El problema de Gear Tire Company 279

6.3 Aproximación mediante la distribución normal de las probabilidades binomiales 283

6.4 Distribución exponencial 287

Cálculo de probabilidades para la distribución exponencial 287

Relación entre las distribuciones de Poisson y exponencial 288

Resumen 290

Glosario 291

Fórmulas clave 291

Ejercicios complementarios 291

Caso práctico Specialty Toys 294

Apéndice 6.1 Distribuciones continuas de probabilidad con Minitab 295

Apéndice 6.2 Distribuciones continuas de probabilidad con Excel 296

Capítulo 7 Muestreo y distribuciones muestrales o de muestreo 298

Estadística en la práctica. MeadWestvaco Corporation 299

7.1 El problema de muestreo de Electronics Associates 300

7.2 Cómo seleccionar una muestra 301

Muestreo de una población finita 301

Muestreo de una población infinita 303

7.3 Estimación puntual 306

Consejo práctico 308

7.4 Introducción a las distribuciones muestrales o de muestreo 310

7.5 Distribución de muestreo de \bar{x} 312

Valor esperado de \bar{x} 312

Desviación estándar de \bar{x} 313

Forma de la distribución de muestreo de \bar{x} 314

Distribución de muestreo de \bar{x} en el problema de EAI 316

Valor práctico de la distribución de muestreo de \bar{x} 317

Relación entre el tamaño de la muestra y la distribución de muestreo de \bar{x} 318

7.6 Distribución de muestreo de \bar{p} 322

Valor esperado de \bar{p} 323

Desviación estándar de \bar{p} 323

Forma de la distribución de muestreo de \bar{p}	324
Valor práctico de la distribución de muestreo de \bar{p}	324
7.7 Propiedades de los estimadores puntuales	328
Insensibilidad	328
Eficiencia	329
Consistencia	330
7.8 Otros métodos de muestreo	331
Muestreo aleatorio estratificado	331
Muestreo por conglomerados (o clusters)	331
Muestreo sistemático	332
Muestreo de conveniencia	332
Muestreo subjetivo	333
Resumen	333
Glosario	334
Fórmulas clave	335
Ejercicios complementarios	335
Apéndice 7.1 Valor esperado y desviación estándar de \bar{x}	337
Apéndice 7.2 Muestreo aleatorio con Minitab	339
Apéndice 7.3 Muestreo aleatorio con Excel	340
Apéndice 7.4 Muestreo aleatorio con StatTools	341
Capítulo 8 Estimación por intervalo	342
<i>Estadística en la práctica. Food Lion</i>	343
8.1 Media poblacional: σ conocida	344
Margen de error y estimación por intervalo	344
Consejo práctico	348
8.2 Media poblacional: σ desconocida	350
Margen de error y estimación por intervalo	351
Consejo práctico	354
Uso de una muestra pequeña	354
Resumen de los procedimientos de estimación por intervalo	356
8.3 Determinación del tamaño de la muestra	359
8.4 Proporción poblacional	362
Determinación del tamaño de la muestra	364
Resumen	367
Glosario	368
Fórmulas clave	369
Ejercicios complementarios	369
<i>Caso práctico 1 Revista Young Professional</i>	372
<i>Caso práctico 2 Gulf Real Estate Properties</i>	373
<i>Caso práctico 3 Metropolitan Research, Inc.</i>	375
Apéndice 8.1 Estimación por intervalo con Minitab	375

Apéndice 8.2 Estimación por intervalo usando Excel 377

Apéndice 8.3 Estimación por intervalo con StatTools 380

Capítulo 9 Pruebas de hipótesis 382

Estadística en la práctica. John Morrell & Company 383

9.1 Formulación de las hipótesis nula y alternativa 384

La hipótesis alternativa como hipótesis de investigación 384

La hipótesis nula como supuesto para desafiado 385

Resumen de las formas para las hipótesis nula y alternativa 386

9.2 Errores tipo I y tipo II 387

9.3 Media poblacional: σ conocida 390

Prueba de una cola 390

Prueba de dos colas 396

Resumen y consejo práctico 398

Relación entre estimación por intervalo y prueba de hipótesis 400

9.4 Media poblacional: σ desconocida 405

Prueba de una cola 405

Prueba de dos colas 406

Resumen y consejo práctico 408

9.5 Proporción poblacional 411

Resumen 413

9.6 Pruebas de hipótesis y toma de decisiones 416

9.7 Cálculo de la probabilidad de los errores tipo II 417

9.8 Determinación del tamaño de la muestra en una prueba de hipótesis para la media poblacional 422

Resumen 425

Glosario 426

Fórmulas clave 427

Ejercicios complementarios 427

Caso práctico 1 Quality Associates, Inc. 430

Caso práctico 2 Conducta ética de los estudiantes de negocios en Bayview University 432

Apéndice 9.1 Pruebas de hipótesis con Minitab 433

Apéndice 9.2 Pruebas de hipótesis con Excel 435

Apéndice 9.3 Pruebas de hipótesis con StatTools 439

Capítulo 10 Inferencia estadística acerca de medias y proporciones con dos poblaciones 441

Estadística en la práctica. U.S. Food and Drug Administration 442

10.1 Inferencias acerca de la diferencia entre dos medias poblacionales: σ_1 y σ_2 conocidas 443

Estimación por intervalo para $\mu_1 - \mu_2$ 443

Pruebas de hipótesis acerca de $\mu_1 - \mu_2$	445
Consejo práctico	447
10.2 Inferencias acerca de la diferencia entre dos medias poblacionales: σ_1 y σ_2 desconocidas	450
Estimación por intervalo para $\mu_1 - \mu_2$	450
Pruebas de hipótesis acerca de $\mu_1 - \mu_2$	452
Consejo práctico	454
10.3 Inferencias acerca de la diferencia entre dos medias poblacionales: muestras pareadas	458
10.4 Inferencias acerca de la diferencia entre dos proporciones poblacionales	464
Estimación por intervalo para $p_1 - p_2$	464
Prueba de hipótesis acerca de $p_1 - p_2$	466
Resumen	470
Glosario	471
Fórmulas clave	471
Ejercicios complementarios	472
<i>Caso práctico</i> Par, Inc.	475
Apéndice 10.1 Inferencias acerca de dos poblaciones usando Minitab	476
Apéndice 10.2 Inferencias acerca de dos poblaciones usando Excel	478
Apéndice 10.3 Inferencias acerca de dos poblaciones usando StatTools	479

Capítulo 11 Inferencias acerca de varianzas poblacionales 482

<i>Estadística en la práctica.</i> La U.S. Government Accountability Office	483
11.1 Inferencias acerca de una varianza poblacional	484
Estimación por intervalo	484
Pruebas de hipótesis	488
11.2 Inferencias acerca de dos varianzas poblacionales	494
Resumen	501
Fórmulas clave	501
Ejercicios complementarios	501
<i>Caso práctico</i> Programa de capacitación de la Fuerza Aérea	503
Apéndice 11.1 Varianzas poblacionales con Minitab	504
Apéndice 11.2 Varianzas poblacionales con Excel	505
Apéndice 11.3 Desviación estándar poblacional simple con StatTools	505

Capítulo 12 Comparaciones de múltiples proporciones, pruebas de bondad de ajuste e independencia 507

<i>Estadística en la práctica.</i> United Way	508
12.1 Pruebas de igualdad de proporciones poblacionales para tres o más poblaciones	509
Procedimiento de comparación múltiple	514

12.2 Prueba de independencia 519**12.3 Prueba de bondad de ajuste 527**

Distribución de probabilidad multinomial 527

Distribución de probabilidad normal 530

Resumen 536

Glosario 536

Fórmulas clave 537

Ejercicios complementarios 537

Caso práctico Una agenda bipartidista para el cambio 540

Apéndice 12.1 Pruebas de ji-cuadrada utilizando Minitab 541

Apéndice 12.2 Pruebas de ji-cuadrada con Excel 542

Apéndice 12.2 Pruebas de ji-cuadrada con StatTools 544

Capítulo 13 Diseño de experimentos y análisis de varianza 545*Estadística en la práctica. Burke Marketing Services, Inc.* 546**13.1 Introducción al diseño de experimentos y al análisis de varianza 547**

Recolección de datos 548

Supuestos del análisis de varianza 549

Análisis de varianza: una perspectiva conceptual 549

13.2 Análisis de varianza y diseño completamente aleatorizado 552

Estimación de la varianza poblacional entre tratamientos 553

Estimación de la varianza poblacional dentro de los tratamientos 554

Comparación de las estimaciones de las varianzas: la prueba F 555

Tabla de ANOVA (análisis de varianza) 557

Resultados de computadora para el análisis de varianza 558

Prueba para la igualdad de k medias poblacionales: un estudio observacional 559**13.3 Procedimientos de comparación múltiple 563**

LSD de Fisher 563

Tasas de error tipo I 566

13.4 Diseño de bloques aleatorizado 569

Prueba de estrés para controladores de tráfico aéreo 570

Procedimiento ANOVA 571

Cálculos y conclusiones 572

13.5 Experimento factorial 576

Procedimiento ANOVA 578

Cálculos y conclusiones 578

Resumen 583

Glosario 584

Fórmulas clave 584

Ejercicios complementarios 586

- Caso práctico 1* Wentworth Medical Center 591
Caso práctico 2 Compensación para vendedores profesionales 592
 Apéndice 13.1 Análisis de varianza con Minitab 592
 Apéndice 13.2 Análisis de varianza con Excel 594
 Apéndice 13.3 Análisis de un diseño completamente aleatorizado usando StatTools 597

Capítulo 14 Regresión lineal simple 598

- Estadística en la práctica. Alliance Data Systems* 599
- 14.1 Modelo de regresión lineal simple 600**
 Modelo de regresión y ecuación de regresión 600
 Ecuación de regresión estimada 601
- 14.2 Método de mínimos cuadrados 603**
- 14.3 Coeficiente de determinación 614**
 Coeficiente de correlación 618
- 14.4 Supuestos del modelo 622**
- 14.5 Prueba de significancia 623**
 Estimación de σ^2 623
 Prueba t 624
 Intervalo de confianza para β_1 626
 Prueba F 627
 Algunas advertencias acerca de la interpretación de las pruebas de significancia 629
- 14.6 Uso de la ecuación de regresión estimada para estimación y predicción 632**
 Estimación por intervalo 633
 Intervalo de confianza para el valor medio de y 634
 Intervalo de predicción para un solo valor de y 635
- 14.7 Solución por computadora 640**
- 14.8 Análisis de residuales: validación de los supuestos del modelo 644**
 Gráfica de residuales contra x 645
 Gráfica de residuales contra \hat{y} 646
 Residuales estandarizados 648
 Gráfica de probabilidad normal 650
- 14.9 Análisis de residuales: observaciones atípicas y observaciones influyentes 653**
 Detección de observaciones atípicas 653
 Detección de observaciones influyentes 656
- Resumen 661
 Glosario 661
 Fórmulas clave 662
 Ejercicios complementarios 664
- Caso práctico 1* Medición del riesgo en el mercado bursátil 671
Caso práctico 2 Departamento de Transporte de Estados Unidos 672
Caso práctico 3 Cómo seleccionar una cámara digital 673

- Caso práctico 4* Cómo encontrar el automóvil de mayor valor 674
- Apéndice 14.1** Obtención de la fórmula de mínimos cuadrados con base en el cálculo 675
- Apéndice 14.2** Prueba de significancia usando correlación 677
- Apéndice 14.3** Análisis de regresión con Minitab 678
- Apéndice 14.4** Análisis de regresión con Excel 678
- Apéndice 14.5** Análisis de regresión con StatTools 681

Capítulo 15 Regresión múltiple 682

Estadística en la práctica. dunnhumby 683

- 15.1** Modelo de regresión múltiple 684
- Modelo de regresión y ecuación de regresión 684
 - Ecuación de regresión múltiple estimada 684
- 15.2** Método de mínimos cuadrados 685
- Ejemplo: Butler Trucking Company 686
 - Nota sobre la interpretación de los coeficientes 688
- 15.3** Coeficiente de determinación múltiple 694
- 15.4** Supuestos del modelo 698
- 15.5** Prueba de significancia 699
- Prueba F 699
 - Prueba t 702
 - Multicolinealidad 703
- 15.6** Uso de la ecuación de regresión estimada para estimación y predicción 706
- 15.7** Variables independientes cualitativas 709
- Ejemplo: Johnson Filtration, Inc. 709
 - Interpretación de los parámetros 711
 - Variables cualitativas más complejas 713
- 15.8** Análisis residual 717
- Detección de observaciones atípicas 719
 - Residuales estudentizados eliminados y observaciones atípicas 719
 - Observaciones influyentes 720
 - Uso de la medida de la distancia de Cook para identificar observaciones influyentes 720
- 15.9** Regresión logística 724
- Ecuación de regresión logística 725
 - Estimación de la ecuación de regresión logística 726
 - Prueba de significancia 728
 - Aplicación en la administración 729
 - Interpretación de la ecuación de regresión logística 729
 - Transformación logit 732
- Resumen** 736
- Glosario** 736
- Fórmulas clave** 737

Ejercicios complementarios 739

Caso práctico 1 Consumer Research, Inc. 745

Caso práctico 2 Predicción de victorias para los conductores de NASCAR 746

Caso práctico 3 Cómo encontrar el automóvil de mayor valor 747

Apéndice 15.1 Regresión múltiple con Minitab 748

Apéndice 15.2 Regresión múltiple con Excel 748

Apéndice 15.3 Regresión logística con Minitab 750

Apéndice 15.4 Análisis de regresión múltiple con StatTools 750

Capítulo 16 Análisis de regresión: construcción de modelos 751

Estadística en la práctica. Monsanto Company 752

16.1 Modelo lineal general 753

Modelado de relaciones curvilíneas 753

Interacción 756

Transformaciones que involucran a la variable dependiente 760

Modelos no lineales que son intrínsecamente lineales 763

16.2 Determinación de cuándo agregar o eliminar variables 767

Caso general 769

Uso de los valores- p 770

16.3 Análisis de un problema mayor 773

16.4 Procedimientos de selección de variables 777

Regresión por pasos 777

Selección hacia adelante 778

Eliminación hacia atrás 779

Regresión de los mejores subconjuntos 779

La elección final 780

16.5 Método de regresión múltiple para el diseño de experimentos 783

16.6 Autocorrelación y la prueba de Durbin-Watson 788

Resumen 792

Glosario 792

Fórmulas clave 792

Ejercicios complementarios 793

Caso práctico 1 Análisis de las estadísticas de la PGA Tour 796

Caso práctico 2 Clasificación de vinos de la Región de Piamonte en Italia 797

Apéndice 16.1 Procedimientos de selección de variables con Minitab 798

Apéndice 16.2 Procedimientos de selección de variables con StatTools 799

Capítulo 17 Análisis de series de tiempo y elaboración de pronósticos 800

Estadística en la práctica. Nevada Occupational Health Center 801

17.1 Patrones de una serie de tiempo 802

Patrón horizontal 802

	Patrón de tendencia	804
	Patrón o componente estacional	804
	Patrones estacional y de tendencia	805
	Patrón cíclico	805
	Selección de un método de elaboración de pronósticos	807
17.2	Exactitud del pronóstico	808
17.3	Promedios móviles y suavizamiento exponencial	813
	Promedios móviles	813
	Promedios móviles ponderados	816
	Suavizamiento exponencial	816
17.4	Proyección de la tendencia	823
	Regresión de tendencia lineal	823
	Suavizamiento exponencial lineal de Holt	828
	Regresión de tendencia no lineal	830
17.5	Estacionalidad y tendencia	836
	Estacionalidad sin tendencia	836
	Estacionalidad y tendencia	838
	Modelos basados en datos mensuales	841
17.6	Descomposición de series de tiempo	845
	Cálculo de los índices estacionales	846
	Desestacionalización de una serie de tiempo	849
	Uso de una serie de tiempo desestacionalizada para identificar tendencias	851
	Ajustes estacionales	852
	Modelos basados en datos mensuales	852
	Patrón o componente cíclico	852
	Resumen	855
	Glosario	856
	Fórmulas clave	857
	Ejercicios complementarios	857
	<i>Caso práctico 1</i> Pronóstico de ventas de alimentos y bebidas	861
	<i>Caso práctico 2</i> Elaboración del pronóstico de pérdida de ventas	862
	Apéndice 17.1 Elaboración de pronósticos con Minitab	864
	Apéndice 17.2 Elaboración de pronósticos con Excel	866
	Apéndice 17.2 Elaboración de pronósticos con StatTools	867
	Capítulo 18 Métodos no paramétricos	870
	<i>Estadística en la práctica. West Shell Realtors</i>	871
18.1	La prueba de signos	872
	Prueba de hipótesis acerca de una mediana poblacional	872
	Prueba de hipótesis con muestras pareadas	877
18.2	Prueba de rangos con signo de Wilcoxon	880
18.3	Prueba de Mann-Whitney-Wilcoxon	885
18.4	Prueba de Kruskal-Wallis	895

18.5 Correlación de rangos 900

Resumen 905

Glosario 905

Fórmulas clave 906

Ejercicios complementarios 907

Apéndice 18.1 Métodos no paramétricos con Minitab 910

Apéndice 18.2 Métodos no paramétricos con Excel 912

Apéndice 18.3 Métodos no paramétricos con StatTools 914

Capítulo 19 Métodos estadísticos de control de la calidad 916*Estadística en la práctica. Dow Chemical Company* 917**19.1 Filosofías y marcos de referencia 918**

El Malcolm Baldrige National Quality Award 919

ISO 9000 919

Six Sigma 919

Calidad en el sector servicios 922

19.2 Control estadístico de procesos 922

Gráficas de control 923

Gráfica \bar{x} : media y desviación estándar conocidas del proceso 924Gráfica \bar{x} : media y desviación estándar desconocidas del proceso 926Gráfica R 929Gráfica p 931Gráfica np 933

Interpretación de las gráficas de control 933

19.3 Muestreo de aceptación 936

KALI, Inc.: un ejemplo de muestreo de aceptación 937

Cálculo de la probabilidad de aceptación de un lote 938

Selección de un plan de muestreo de aceptación 941

Planes de muestreo múltiple 943

Resumen 944

Glosario 944

Fórmulas clave 945

Ejercicios complementarios 946

Apéndice 19.1 Gráficas de control con Minitab 948

Apéndice 19.2 Gráficas de control utilizando StatTools 949

Capítulo 20 Números índice 951*Estadística en la práctica. Oficina de Estadísticas Laborales, Departamento del Trabajo de Estados Unidos* 952**20.1 Precios relativos 953****20.2 Índices de precios agregados 953****20.3 Cálculo del índice de precios agregados a partir de los precios relativos 957**

20.4 Algunos índices importantes de precios 959

Índice de precios al consumidor 959

Índice de precios al productor 959

Promedios Dow Jones 960

20.5 Deflactación de una serie mediante índices de precios 961**20.6 Índices de precios: otras consideraciones 964**

Selección de artículos 964

Selección de un periodo base 965

Variaciones en la calidad 965

20.5 Índices de cantidad 965

Resumen 967

Glosario 967

Fórmulas clave 968

Ejercicios complementarios 968

Capítulo 21 Análisis de decisiones [sitio web](#)*Estadística en la práctica. Ohio Edison Company 21-2***21.1 Formulación del problema 21-3**

Tablas de pagos 21-4

Árboles de decisión 21-4

21.2 Toma de decisiones con probabilidades 21-5

Método del valor esperado 21-5

Valor esperado de la información perfecta 21-7

21.3 Análisis de decisiones con información muestral 21-13

Árbol de decisión 21-14

Estrategia de decisión 21-15

Valor esperado de la información muestral 21-18

21.4 Cálculo de probabilidades mediante el teorema de Bayes 21-24

Resumen 21-28

Glosario 21-29

Fórmulas clave 21-30

Ejercicios complementarios 21-30

Caso práctico Estrategia de defensa en una demanda 21-33

Apéndice Introducción a PrecisionTree 21-34

Apéndice Soluciones a los ejercicios de autoevaluación y a los ejercicios pares 21-39

Capítulo 22 Encuesta por muestreo [sitio web](#)*Estadística en la práctica. Duke Energy 22-2***22.1 Terminología que se utiliza en las encuestas por muestreo 22-2****22.1 Tipos de encuestas y métodos de muestreo 22-3**

22.3 Errores en las encuestas	22-5
Error que no es de muestreo	22-5
Error de muestreo	22-5
22.4 Muestreo aleatorio simple	22-6
Media poblacional	22-6
Total poblacional	22-7
Proporción poblacional	22-8
Determinación del tamaño de la muestra	22-9
22.5 Muestreo aleatorio estratificado simple	21-12
Media poblacional	22-12
Total poblacional	22-14
Proporción poblacional	22-15
Determinación del tamaño de la muestra	22-16
22.6 Muestreo por conglomerados	22-21
Media poblacional	22-23
Total poblacional	22-24
Proporción poblacional	22-25
Determinación del tamaño de la muestra	22-26
22.7 Muestreo sistemático	22-29
Resumen	22-29
Glosario	22-30
Fórmulas clave	22-30
Ejercicios complementarios	22-34
Apéndice Soluciones de los ejercicios de autoevaluación y respuestas de los ejercicios pares	22-37
Apéndice A Referencias y bibliografía	971
Apéndice B Tablas	974
Apéndice C Notación de suma	1001
Apéndice D Soluciones de las autoevaluaciones y respuestas a los ejercicios con números pares	1003
Apéndice E Microsoft Excel 2010 y sus herramientas para el análisis estadístico	1064
Apéndice F Cálculo de los valores-<i>p</i> utilizando Minitab y Excel	1076
Índice analítico	1080

CAPÍTULO 1

Los datos y la estadística

CONTENIDO

ESTADÍSTICA EN LA PRÁCTICA:
BLOOMBERG BUSINESSWEEK

- 1.1** APLICACIONES EN NEGOCIOS Y ECONOMÍA
 - Contabilidad
 - Finanzas
 - Marketing
 - Producción
 - Economía
 - Sistemas de información
- 1.2** DATOS
 - Elementos, variables y observaciones
 - Escalas de medición
 - Datos categóricos y cuantitativos
 - Datos de corte transversal y de series de tiempo
- 1.3** FUENTES DE DATOS
 - Fuentes existentes
 - Estudios estadísticos
 - Errores en la adquisición de datos
- 1.4** ESTADÍSTICA DESCRIPTIVA
- 1.5** INFERENCIA ESTADÍSTICA
- 1.6** COMPUTADORAS Y ANÁLISIS ESTADÍSTICO
- 1.7** MINERÍA DE DATOS
- 1.8** LINEAMIENTOS ÉTICOS PARA LA PRÁCTICA DE LA ESTADÍSTICA

ESTADÍSTICA *en* LA PRÁCTICA

*BLOOMBERG BUSINESSWEEK** NEW YORK, NEW YORK

Con una circulación global de más de un millón de ejemplares, *Bloomberg Businessweek* es la revista de negocios más leída en el mundo. Más de 1 700 reporteros en 145 oficinas de todo el mundo producen una variedad de artículos de interés para la comunidad de los negocios y la economía. Además de reportajes especiales sobre temas de actualidad, la revista contiene secciones regulares sobre administración internacional, análisis económico, procesamiento de información y ciencia y tecnología. La información contenida en los reportajes especiales y las secciones regulares ayuda a los lectores a mantenerse al día en los desarrollos actuales y evalúa su impacto en los negocios y la economía bajo las condiciones actuales.

La mayoría de los números de *Bloomberg Businessweek* anteriormente conocida sólo como *BusinessWeek* contiene un artículo de fondo sobre un tema de interés actual. Dichos artículos a menudo contienen hechos y resúmenes estadísticos que ayudan al lector a comprender la información de negocios y economía. Por ejemplo, el número del 3 de marzo de 2011 analizó el impacto en los negocios de movilizar su trabajo más importante hacia el cómputo en la nube; el número del 30 de mayo de 2011 incluyó un reportaje sobre la crisis del U.S. Postal Service, y el número del 1 de agosto de 2011 contenía un reportaje sobre las razones por las que la crisis por endeudamiento era peor de lo que se pensaba. Además, *Bloomberg Businessweek* proporciona estadísticas sobre el estado de la economía, que incluyen índices de producción, precios de las acciones, fondos de inversión y tasas de interés.

Bloomberg Businessweek también utiliza información estadística en la administración de su propia empresa. Por ejemplo, una encuesta anual aplicada a los suscriptores permite a la empresa obtener sus datos demográficos, hábitos de lectura, compras probables, su estilo de vida, etc. Los di-

* Los autores agradecen a Charlene Trentham, gerente de investigación de *BusinessWeek*, por proporcionar este artículo para la sección Estadística en la práctica.



Bloomberg Businessweek utiliza hechos estadísticos y resúmenes en muchos de sus artículos. © Kyodo/Photoshot.

rectivos de *Bloomberg Businessweek* utilizan resúmenes estadísticos de la consulta para brindar un mejor servicio a sus suscriptores y anunciantes. Una encuesta reciente entre los estadounidenses reveló que 90% de los suscriptores de *Bloomberg Businessweek* utiliza una computadora personal en su hogar, y que 64% realizó compras por computadora en el trabajo. Estas estadísticas alertaron a los directivos de la revista sobre el interés de los suscriptores en los nuevos avances en computación. Los resultados de la encuesta también se pusieron a disposición de los posibles anunciantes. El alto porcentaje de suscriptores que usan computadoras personales en el hogar y de los que realizan compras por internet en su trabajo son un incentivo para que un fabricante de estos equipos considere anunciarse en *Bloomberg Businessweek*.

En este capítulo se estudian los tipos de datos de que se dispone para el análisis estadístico y se describe cómo se obtienen los mismos. La estadística descriptiva y la inferencia estadística se presentan como medios para convertir los datos en información fácil de interpretar.

Es frecuente ver en los periódicos y las revistas las frases siguientes:

- El Departamento del Trabajo de Estados Unidos reportó una caída en la tasa de desempleo a 8.2%, su menor nivel en los últimos tres años (*The Washington Post*, 6 de abril de 2012).
- Cada estadounidense consume al año un promedio 23.2 cuartos de galón de helado, leche helada, sorbetes, hielos y otros lácteos congelados producidos comercialmente (sitio web makeicecream.com 2 de abril de 2012).
- La mediana del precio de venta de una casa vacacional es de 121 300 dólares (@CNN Money, 29 de marzo de 2012).

- La montaña rusa The Wild Eagle de Dollywood en Pigeon Forge, Tennessee, alcanza una velocidad máxima de 61 millas por hora (sitio web de *USA Today*, 5 de abril de 2012).
- El número de usuarios registrados en Pinterest, el servicio del sitio web para compartir fotografías, creció 85% entre mediados de enero y mediados de febrero (CNBC, 29 de marzo de 2012).
- El Pew Research Center reportó que la mediana de la edad a la que las novias contraen matrimonio por primera vez en Estados Unidos alcanzó un máximo histórico de 26.5 años (*Significance*, febrero de 2012).
- Los canadienses registraron un promedio de 45 horas conectados en línea durante el cuarto trimestre del 2011 (CBC News, 2 de marzo de 2012).
- La Reserva Federal de Estados Unidos reportó que la deuda promedio en tarjetas de crédito es de 5 204 dólares por persona (sitio web PRWeb, 5 de abril de 2012).

Los datos numéricos en las frases anteriores (8.2%, 23.2, \$121 300, 61, 85%, 26.5, 45, \$5 204) se llaman *estadísticas*. En este sentido, el término *estadística* se refiere a datos numéricos como promedios, medias, porcentajes e índices que nos ayudan a entender una variedad de situaciones de los negocios y la economía. Sin embargo, como verá más adelante, el campo, o materia, de la estadística abarca mucho más que los datos numéricos. En un sentido más amplio, la **estadística** se define como el arte y la ciencia de recolectar, analizar e interpretar datos. En particular en los negocios y la economía, la información que se obtiene a partir de la recolección, el análisis, la presentación y la interpretación de los datos permite a los administradores o gerentes y a quienes toman decisiones comprender mejor los entornos económico y de negocios, y por lo tanto tomar decisiones mejores y más informadas. En este libro se enfatiza el uso de la estadística para la toma de decisiones en ambos ámbitos.

El capítulo 1 comienza con algunos ejemplos de aplicaciones de la estadística en los negocios y la economía. En la sección 1.2 se define el término *dato* y se introduce el concepto de banco de datos. Esta sección también presenta términos clave como *variables* y *observaciones*; estudia la diferencia entre datos cuantitativos y categóricos, e ilustra los usos de los datos de corte transversal y de series de tiempo. En la sección 1.3 se analiza cómo se obtienen los datos de fuentes existentes o por medio de estudios experimentales diseñados para obtener datos nuevos. También se destaca el papel importante que juega hoy internet en la obtención de datos. Los usos de los datos en el desarrollo de la estadística descriptiva y la elaboración de inferencias estadísticas se describen en las secciones 1.4 y 1.5. Las últimas tres secciones tratan sobre la función de la computadora en el análisis estadístico, y presentan una introducción al campo relativamente nuevo de la minería de datos y un análisis de los lineamientos éticos para la práctica estadística. Al final del capítulo se incluye un apéndice con una introducción al complemento StatTools que se usa para ampliar las opciones estadísticas a los usuarios de Microsoft Excel.

1.1

Aplicaciones en negocios y economía

En el actual entorno global de los negocios y la economía, cualquier persona tiene acceso a una vasta cantidad de información estadística. Los gerentes y líderes de decisiones más exitosos entienden la información y saben cómo usarla de manera eficiente. En esta sección se proporcionan ejemplos que ilustran algunas aplicaciones de la estadística en los negocios y la economía.

Contabilidad

Las firmas de contabilidad públicas utilizan procedimientos de muestreo estadístico cuando realizan auditorías para sus clientes. Por ejemplo, suponga que una firma contable quiere determinar si el estado de cuenta de un cliente representa de manera precisa el monto real de las cuentas por cobrar. La gran cantidad de cuentas por cobrar hace que la revisión y la validación de cada cuenta consuman demasiado tiempo y dinero. Como práctica común en este tipo de situaciones, el personal de auditoría selecciona un subconjunto de las cuentas llamado *muestra*. Después de revisar la precisión de la selección que tomó como muestra, los auditores llegan a

una conclusión con respecto a si el monto de las cuentas por cobrar que aparece en el estado de cuenta del cliente es aceptable.

Finanzas

Los analistas financieros utilizan una variedad de información estadística como guía para sus recomendaciones de inversión. En el caso de las acciones, revisan diversos datos financieros que incluyen las razones precio/utilidad y el rendimiento por concepto de dividendos. Al comparar la información para una acción con datos sobre los promedios del mercado de valores, un analista financiero puede formular una conclusión acerca de si una acción es una buena inversión. Por ejemplo, el número del 19 de marzo de 2012 de *The Wall Street Journal* reportó que el rendimiento promedio por concepto de dividendos de las 500 grandes empresas S&P 500 fue de 2.2%. Microsoft, en específico, obtuvo un rendimiento por concepto de dividendos de 2.42%. En este caso, la información estadística sobre el rendimiento por concepto de dividendos indica una cifra superior de los dividendos de Microsoft sobre el promedio de dividendos de las empresas S&P 500. Esta información sobre Microsoft, junto con otra, ayuda a los analistas a realizar recomendaciones para conservar, vender o comprar acciones de Microsoft.

Marketing

Los escáneres electrónicos en las cajas de cobro de las tiendas minoristas recolectan datos para diversas aplicaciones de investigación de mercados. Por ejemplo, proveedores de datos como ACNielsen e Information Resources, Inc. compran datos de los escáneres en puntos de venta como las tiendas de abarrotes, los procesan y luego venden resúmenes estadísticos a los fabricantes. Estos últimos gastan cientos de miles de dólares por categoría de producto para obtener este tipo de datos. Los fabricantes también compran datos y resúmenes estadísticos sobre actividades promocionales, como la asignación de precios especiales y el uso de exhibidores dentro de las tiendas. Los gerentes de marca pueden revisar las estadísticas de los escáneres y de la actividad promocional para comprender mejor la relación entre las actividades de promoción y las ventas. Estos análisis a menudo son útiles para establecer estrategias de futuras marketing para diversos productos.

Producción

El énfasis actual en la calidad hace que su control sea una aplicación importante de la estadística en la producción. Diversas gráficas estadísticas de control de calidad se usan para monitorear el resultado de un proceso de producción. En particular, una gráfica x barra sirve para monitorear el resultado promedio. Suponga, por ejemplo, que una máquina llena envases con 12 onzas de una bebida refrescante. En forma periódica, un empleado de producción selecciona una muestra de envases y calcula el número promedio de onzas en la muestra. Este promedio, o valor x barra, se traza en una gráfica x barra. Un valor trazado sobre el límite superior de control de la gráfica indica que hay un exceso en el llenado, y un valor trazado por debajo del límite inferior de control indica que el llenado es deficiente. El proceso se considera “bajo control” y permite continuar siempre que los valores x barra trazados se encuentren dentro de los límites de control superior e inferior de la gráfica. Si se interpreta de manera adecuada, una gráfica x barra ayuda a determinar cuándo es necesario realizar ajustes para corregir un proceso de producción.

Economía

Los economistas a menudo proporcionan pronósticos sobre la economía o algún otro tema relacionado. Utilizan diferente información estadística para elaborarlos. Por ejemplo, para pronosticar las tasas de inflación recurren a información estadística sobre indicadores como el índice de precios al consumidor, la tasa de desempleo y el uso de la capacidad de manufactura. Estos indicadores se introducen con frecuencia en modelos computarizados de pronósticos que predicen las tasas de inflación.

Las aplicaciones de la estadística, como las descritas en esta sección, son parte integral de este libro. Estos ejemplos proporcionan una descripción general de gran diversidad de apli-

caciones. Para complementar estos ejemplos, profesionales en el campo de los negocios y la economía aportaron artículos para la sección *Estadística en la práctica* al inicio de cada capítulo, donde se presenta el material que cubre su contenido. Dichas aplicaciones muestran la importancia de la estadística en una amplia variedad de situaciones de negocios y economía.

Sistemas de información

Los administradores de sistemas de información son responsables de la operación cotidiana de las redes de cómputo de una organización. Una gran variedad de información estadística les ayuda a evaluar el desempeño de dichas redes, como las redes de área local (LAN), redes de área amplia (WAN), segmentos de redes, intranets y otros sistemas de comunicación de datos. Las estadísticas, como el número medio de usuarios del sistema, la cantidad de tiempo que un componente del sistema no funciona y la proporción de ancho de banda que se utiliza en varios momentos del día son ejemplos de información estadística que ayuda a los administradores de los sistemas a comprender y manejar mejor las redes de cómputo.

1.2

Datos

Los **datos** son los hechos y las cifras recabados, analizados y resumidos para su presentación e interpretación. Todos los datos recabados en un estudio en particular se conocen como **banco o conjunto de datos** del estudio. La tabla 1.1 presenta un banco de datos que contiene información sobre 60 países que participan en la Organización Mundial de Comercio. Esta organización fomenta el libre flujo del comercio internacional y proporciona un foro para resolver disputas comerciales.

Elementos, variables y observaciones

Los **elementos** son las entidades a partir de las cuales se recaban los datos. Cada uno de los países listados en la tabla 1.1 es un elemento con el nombre del país o del elemento en la primera columna. Puesto que se presentan 60 países, el banco de datos contiene 60 elementos.

Una **variable** es una característica de interés para los elementos. El banco de datos de la tabla 1.1 incluye las cinco variables siguientes:

- Estatus en la OMC (WTO Status): El estatus de membresía que tiene el país ante la Organización Mundial de Comercio; puede ser como miembro o como observador
- PIB per cápita [Per Capita GDP (\$)]: El producto total del país dividido entre su número de habitantes. Por lo general, esta variable se emplea para comparar la productividad económica de los países.
- Déficit comercial en miles de dólares (Trade Deficit \$1000s): La diferencia en dólares entre el valor de las importaciones y las exportaciones totales del país.
- Fitch Rating: La calificación de crédito soberano del país provista por el grupo Fitch¹. El rango de calificaciones para los créditos va desde AAA en su punto más alto hasta F en su punto más bajo, y pueden ser modificadas con atributos + o -.
- Fitch Outlook: Es un indicador de que la dirección de la calificación del crédito tiene posibilidades de modificarse durante los próximos dos años. El panorama puede ser negativo, estable o positivo.

En un estudio, las mediciones recabadas para cada elemento en cada variable proporcionan los datos. El conjunto de mediciones obtenido para un elemento en particular se llama **observación**. “Al analizar de nuevo la tabla 1.1, se observa que el conjunto de mediciones para la primera observación (Armenia) es Miembro, 5 400, 2 673 359, BB- y Estable. El conjunto de

¹ The Fitch Group es una de tres organizaciones de estadísticas, reconocidas a nivel nacional, designadas por la U.S. Securities and Exchange Commission. Las otras dos son Standard and Poor's y Moody's.

TABLA 1.1 Banco de datos de 60 países que participan en la Organización Mundial de Comercio

Nation	WTO Status	Per Capita GDP (\$)	Trade Deficit (\$1000s)	Fitch Rating	Fitch Outlook
Armenia	Member	5 400	2 673 359	BB-	Stable
Australia	Member	40 800	-33 304 157	AAA	Stable
Austria	Member	41 700	12 796 558	AAA	Stable
Azerbaijan	Observer	5 400	-16 747 320	BBB-	Positive
Bahrain	Member	27 300	3 102 665	BBB	Stable
Belgium	Member	37 600	-14 930 833	AA+	Negative
Brazil	Member	11 600	-29 796 166	BBB	Stable
Bulgaria	Member	13 500	4 049 237	BBB-	Positive
Canada	Member	40 300	-1 611 380	AAA	Stable
Cape Verde	Member	4 000	874 459	B+	Stable
Chile	Member	16 100	-14 558 218	A+	Stable
China	Member	8 400	-156 705 311	A+	Stable
Colombia	Member	10 100	-1 561 199	BBB-	Stable
Costa Rica	Member	11 500	5 807 509	BB+	Stable
Croatia	Member	18 300	8 108 103	BBB-	Negative
Cyprus	Member	29 100	6 623 337	BBB	Negative
Czech Republic	Member	25 900	-10 749 467	A+	Positive
Denmark	Member	40 200	-15 057 343	AAA	Stable
Ecuador	Member	8 300	1 993 819	B-	Stable
Egypt	Member	6 500	28 486 933	BB	Negative
El Salvador	Member	7 600	5 019 363	BB	Stable
Estonia	Member	20 200	802 234	A+	Stable
France	Member	35 000	118 841 542	AAA	Stable
Georgia	Member	5 400	4 398 153	B+	Positive
Germany	Member	37 900	-213 367 685	AAA	Stable
Hungary	Member	19 600	-9 421 301	BBB-	Negative
Iceland	Member	38 000	-504 939	BB+	Stable
Ireland	Member	39 500	-59 093 323	BBB+	Negative
Israel	Member	31 000	6 722 291	A	Stable
Italy	Member	30 100	33 568 668	A+	Negative
Japan	Member	34 300	31 675 424	AA	Negative
Kazakhstan	Observer	13 000	-33 220 437	BBB	Positive
Kenya	Member	1 700	9 174 198	B+	Stable
Latvia	Member	15 400	2 448 053	BBB-	Positive
Lebanon	Observer	15 600	13 715 550	B	Stable
Lithuania	Member	18 700	3 359 641	BBB	Positive
Malaysia	Member	15 600	-39 420 064	A-	Stable
Mexico	Member	15 100	1 288 112	BBB	Stable
Peru	Member	10 000	-7 888 993	BBB	Stable
Philippines	Member	4 100	15 667 209	BB+	Stable
Poland	Member	20 100	19 552 976	A-	Stable
Portugal	Member	23 200	21 060 508	BBB-	Negative
South Korea	Member	31 700	-37 509 141	A+	Stable
Romania	Member	12 300	13 323 709	BBB-	Stable
Russia	Observer	16 700	-151 400 000	BBB	Positive
Rwanda	Member	1 300	939 222	B	Stable
Serbia	Observer	10 700	8 275 693	BB-	Stable
Seychelles	Observer	24 700	666 026	B	Stable
Singapore	Member	59 900	-27 110 421	AAA	Stable
Slovakia	Member	23 400	-2 110 626	A+	Stable
Slovenia	Member	29 100	2 310 617	AA-	Negative
South Africa	Member	11 000	3 321 801	BBB+	Stable

WEB file
Nations

Los bancos de datos, como el de Nations, están disponibles en inglés en el sitio web de este libro.

Sweden	Member	40 600	-10 903 251	AAA	Stable
Switzerland	Member	43 400	-27 197 873	AAA	Stable
Thailand	Member	9 700	2 049 669	BBB	Stable
Turkey	Member	14 600	71 612 947	BB+	Positive
UK	Member	35 900	162 316 831	AAA	Negative
Uruguay	Member	15 400	2 662 628	BB	Positive
USA	Member	48 100	784 438 559	AAA	Stable
Zambia	Member	1 600	-1 805 198	B+	Stable

mediciones para la segunda observación (Austria) es Miembro, 40 800, -33 304 157, AAA y Estable. Un banco de datos de 60 elementos contiene 60 observaciones

Escalas de medición

La recolección de datos requiere una de las escalas de medición siguientes: nominal, ordinal, de intervalo o de razón. La escala de medición determina la cantidad de información contenida en los datos e indica la manera más apropiada de resumirlos y analizarlos estadísticamente.

Cuando los datos de una variable se componen de etiquetas o nombres utilizados para identificar un atributo del elemento, la escala de medición se considera una **escala nominal**. Por ejemplo, al observar los datos de la tabla 1.1 se observa que la escala de medición para la variable estatus en la OMC es nominal, porque “miembro (member)” y “observador (observer)” son etiquetas que se usan para identificar la categoría del estatus del país. En tales casos se puede utilizar un código numérico o etiquetas no numéricas. Por ejemplo, para facilitar la recolección y preparación de los datos con la finalidad de introducirlos en una base de datos computarizada, podría utilizarse un código numérico para la variable estatus en la OMC (WTO Status) que establezca que 1 denota un país miembro de la Organización Mundial de Comercio y 2 un país observador. La escala de medición es nominal a pesar de que los datos aparecen como valores numéricos.

La escala de medición de una variable se llama **escala ordinal** si los datos exhiben las propiedades de los datos nominales y su orden o clasificación es significativo. Por ejemplo, en relación con la tabla 1.1, la escala de medición para Ritch Rating es ordinal, porque las etiquetas de rating o calificación que van desde AAA hasta F pueden ordenarse desde la mejor calificación crediticia AAA hasta la peor, que es F. Las letras que indican la calificación proporcionan etiquetas similares a los datos nominales, pero además pueden clasificarse u ordenarse con base en la calificación crediticia, lo que hace que la escala sea ordinal. Los datos ordinales también pueden proporcionarse por medio de un código numérico, por ejemplo, su número de lista en clase.

En una **escala de intervalo** para una variable, los datos presentan todas las propiedades de los datos ordinales, y el intervalo entre los valores se expresa en términos de una unidad de medida fija. Los datos de intervalo son siempre numéricos. Las calificaciones de la prueba de aptitudes Scholastic Aptitude Test (SAT) son un ejemplo de datos escala de intervalo. Por ejemplo, tres estudiantes que obtuvieron las calificaciones 620, 550 y 470 en una prueba o examen de matemáticas llamada SAT pueden clasificarse u ordenarse en función del mejor al peor desempeño. Además, las diferencias entre las puntuaciones son significativas. Por ejemplo, el estudiante 1 obtuvo $620 - 550 = 70$ puntos más que el estudiante 2, mientras que éste obtuvo $550 - 470 = 80$ puntos más que el estudiante 3.

En una **escala de razón** para una variable los datos tienen todas las propiedades de los datos de intervalo, y la razón de los dos valores es significativa. Para la medición de variables como la distancia, la estatura, el peso y el tiempo se usa la escala de razón. Ésta requiere que se incluya un valor cero para indicar que en este punto no existe un valor para la variable. Por ejemplo, considere el costo de un automóvil. Un valor cero para el costo indicaría que el vehículo no tiene costo, es gratis. Además, si se compara el costo de un automóvil de \$30 000 con el

costo de un segundo automóvil de \$15000, la propiedad de la razón muestra que el primero cuesta $\$30\,000/\$15\,000 = 2$ veces, o el doble, que el segundo.

Datos categóricos y cuantitativos

Los datos se clasifican como categóricos o cuantitativos. Los que se agrupan por categorías específicas se conocen como **datos categóricos**. Este tipo de datos utiliza una escala de medición que puede ser nominal u ordinal. Los que utilizan valores numéricos para indicar cuánto o cuántos se conocen como **datos cuantitativos**; éstos se obtienen usando la escala de medición ya sea de intervalo o de razón.

El método estadístico apropiado para resumir los datos depende de que los datos sean categóricos o cuantitativos.

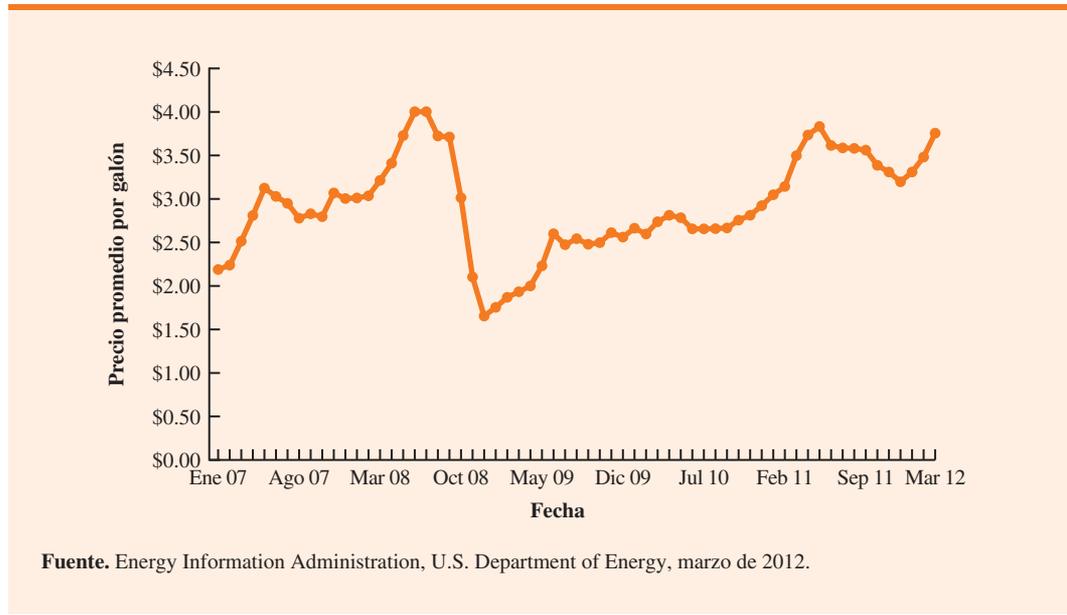
Una **variable categórica** incluye datos categóricos y una **variable cuantitativa** comprende datos cuantitativos. El análisis estadístico apropiado para una variable en particular depende de que ésta sea categórica o cuantitativa. Si la variable es categórica, el análisis estadístico es muy limitado. Los datos categóricos se resumen mediante el conteo del número de observaciones en cada categoría o por medio del cálculo de la proporción de las observaciones en cada categoría. Sin embargo, aun cuando estos datos se identifican por medio de un código aritmético, operaciones como la suma, la resta, la multiplicación y la división no producen resultados que tengan sentido. En la sección 2.1 se estudian algunas maneras de resumir los datos categóricos.

Las operaciones aritméticas sí proporcionan resultados con sentido para las variables cuantitativas. Por ejemplo, los datos cuantitativos pueden sumarse y luego dividirse entre el número de observaciones para calcular el valor promedio, el cual tiene significado y se interpreta con facilidad. En general, se tienen más alternativas para el análisis estadístico cuando los datos son cuantitativos. La sección 2.2 y el capítulo 3 proporcionan maneras de resumir este tipo de datos.

Datos de corte transversal y de series de tiempo

Para efectos del análisis estadístico es importante distinguir entre datos de corte transversal y datos de series de tiempo. Los **datos de corte transversal** son recabados en el mismo momento, o aproximadamente al mismo tiempo. Los de la tabla 1.1 son de corte transversal, debido a que describen las cinco variables para los 60 países de la Organización Mundial de Comercio en el mismo punto de tiempo. Los **datos de series de tiempo** o series temporales son recabados a lo largo de varios periodos. Por ejemplo, la serie de tiempo de la figura 1.1 muestra el precio promedio por galón de gasolina regular convencional en Estados Unidos entre 2007 y 2012. Observe que los precios más altos del hidrocarburo han tendido a presentarse en los meses del verano de 2008 y después descendieron abruptamente en otoño de 2008. Desde este año, el precio promedio por galón ha continuado su incremento de forma estable, alcanzando su máximo histórico nuevamente en 2012.

Las gráficas de los datos de series de tiempo a menudo se encuentran en publicaciones de negocios y economía; ayudan a los analistas a comprender lo que ocurrió en el pasado, identificar cualquier tendencia en el tiempo y proyectar niveles futuros para las series de tiempo. Las gráficas de este tipo pueden adoptar una variedad de formas, como lo muestra la figura 1.2. Con un poco de estudio, suelen ser fáciles de comprender e interpretar. Por ejemplo, la gráfica (A) de la figura 1.2 muestra el índice promedio industrial Dow Jones de 2002 a 2012. En abril de 2002, el índice del mercado de valores se encontraba cercano a 10 000 puntos. Durante los siguientes cinco años alcanzó su máximo histórico ligeramente por encima de los 14 000 puntos en octubre de 2007. Sin embargo, observe que existe un declive en las series de tiempo después de su punto más alto en 2007. Para marzo de 2009, las condiciones económicas deficientes provocaron que el promedio industrial Dow Jones regresara a niveles alrededor de los 7 000 puntos. Este fue un periodo de temor y desaliento para los inversionistas. Sin embargo, para finales de 2009, el índice mostró una recuperación al alcanzar los 10 000 puntos. Este índice ha aumentado de forma constante y se ubicó por encima de los 13 000 puntos a principios de 2012.

FIGURA 1.1 Precio promedio por galón de la gasolina regular en Estados Unidos

La gráfica (B) muestra la utilidad neta de McDonald's Inc. desde 2005 hasta 2011. Las condiciones económicas en declive de 2008 y 2009 fueron realmente benéficas para la empresa, ya que dicha utilidad alcanzó un máximo histórico. Este crecimiento en la utilidad neta demostró que la empresa estaba prosperando durante la crisis económica, cuando la gente empezó a restringir sus gastos y prefería las alternativas más económicas ofrecidas por McDonald's en lugar de los restaurantes tradicionales más caros. La utilidad neta de McDonald's continuó aumentando hasta alcanzar un nuevo máximo histórico en 2010 y 2011.

La gráfica (C) muestra la serie de tiempo para la tasa de ocupación de los hoteles en el sur de Florida durante un periodo de un año. Las tasas más altas, 95 y 98%, ocurren durante los meses de febrero y marzo, cuando el clima de la región es atractivo para los turistas. De hecho, la temporada de ocupación más alta para los hoteles del sur de Florida es de enero a abril de cada año. Por otra parte, observe las menores tasas de ocupación durante los meses de agosto a octubre, periodo en cual se encuentra el indicador más bajo de 50% durante septiembre. Las altas temperaturas y la temporada de huracanes son las razones principales de la caída en la ocupación de los hoteles durante este periodo.

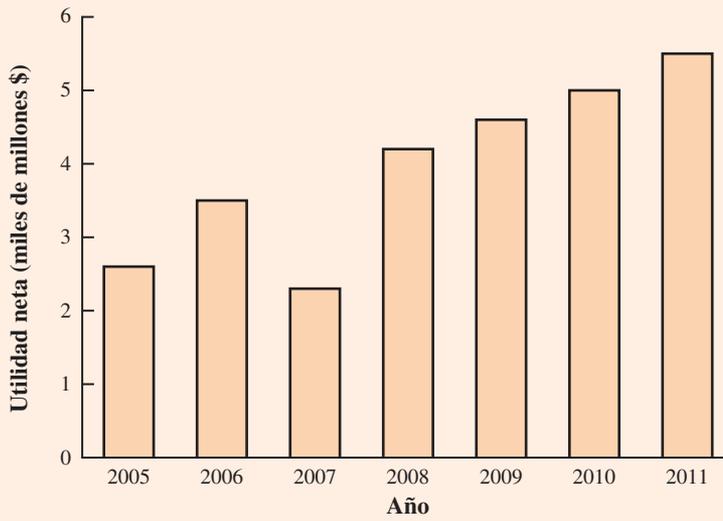
NOTAS Y COMENTARIOS

1. Una observación es el conjunto de mediciones obtenido para cada elemento de un banco de datos. Por consiguiente, el número de observaciones es siempre igual al número de elementos. El número de mediciones obtenidas para cada elemento es igual al número de variables. Por ende, el número total de elementos de datos se determina multiplicando el número de observaciones por el número de variables.
2. Los datos cuantitativos pueden ser discretos o continuos. Los datos cuantitativos que miden cuántos (por ejemplo, el número de llamadas recibidas en 5 minutos), son discretos. Los datos cuantitativos que miden cuánto (por ejemplo, el peso o el tiempo), son continuos debido a que no hay una separación entre los valores de datos posibles.

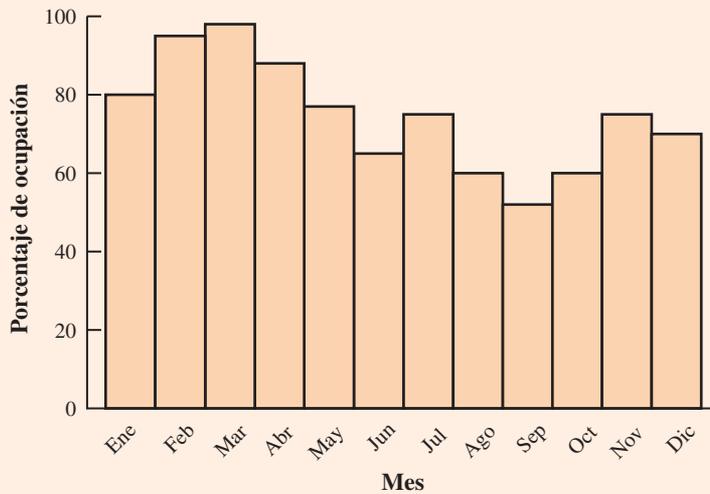
FIGURA 1.2 Varias gráficas de series de tiempo



(A) Promedio industrial Dow Jones



(B) Utilidad neta de McDonald's Inc.



(C) Tasa de ocupación de los hoteles del sur de Florida

1.3

Fuentes de datos

Los datos se obtienen de fuentes existentes o de encuestas y estudios experimentales diseñados para recabar datos nuevos.

Fuentes existentes

En algunos casos, los datos necesarios para una aplicación en particular ya existen. Las empresas mantienen diversas bases de datos de sus empleados, clientes y operaciones de negocios. Los datos sobre los sueldos, la edad y los años de experiencia de los empleados se obtienen por lo general de los registros internos del personal. Otros registros internos contienen datos sobre ventas, gastos de publicidad, costos de distribución, niveles de inventario y cantidades de producción. La mayoría de las empresas mantiene también datos detallados sobre sus clientes. La tabla 1.2 muestra algunos de los datos de que se dispone por lo general a partir de los registros internos de una empresa.

Las organizaciones que se especializan en la recolección y el mantenimiento de datos proveen cantidades significativas de información económica y de negocios. Las empresas tienen acceso a estas fuentes de datos externos por medio de acuerdos o al comprarlos. Dun & Bradstreet, Bloomberg y Dow Jones & Company son tres empresas que ofrecen extensos servicios de bases de datos a sus clientes. ACNielsen e Information Resources, Inc. ha logrado el éxito en su negocio de recolección y procesamiento de datos que vende a anunciantes y fabricantes de productos.

También se obtienen datos de diversas asociaciones de la industria y de organizaciones de interés especial. Travel Industry Association of America mantiene información relacionada con viajes, como el número de turistas y los gastos de viaje por estado. Estos datos son de interés para las empresas y personas de la industria del ramo. El Graduate Management Admission Council cuenta con datos sobre calificaciones de exámenes, características de los estudiantes y programas sobre administración de educación universitaria. La mayoría de los datos provenientes de estos tipos de fuentes se proveen a usuarios calificados por un costo moderado.

Internet es una fuente importante de datos e información estadística. Casi todas las empresas tienen una página web que proporciona información general acerca de la organización, así como datos sobre ventas, número de empleados y de productos, el precio de los productos y sus especificaciones. Además, varias empresas se especializan en proveer información a través de Internet, gracias a lo cual se puede tener acceso a cotizaciones de acciones, precios de los platillos en los restaurantes, datos sobre sueldos y una variedad casi infinita de información.

Las agencias gubernamentales son otra fuente importante de datos existentes. Por ejemplo, el Departamento del Trabajo de Estados Unidos maneja una gran cantidad de datos sobre las tasas de empleo, las tasas salariales, el porcentaje de la población activa y la afiliación a

TABLA 1.2 Ejemplos de datos disponibles de los registros o expedientes internos de una empresa

Fuente	Algunos datos comúnmente disponibles
Registros de empleados	Nombre, domicilio, número de Seguro Social, sueldo, número de días de vacaciones, número de días de incapacidad y bonos
Registros de producción	Número de parte o de producto, cantidad producida, costo de la mano de obra directa y costo de los materiales
Registros de inventarios	Número de parte o de producto, cantidad disponible de unidades, punto de reorden, lote económico o cantidad económica del pedido y programa de descuentos
Registros de ventas	Número de producto, volumen de ventas, volumen de ventas por región y volumen de ventas por tipo de cliente
Registros de crédito	Nombre del cliente, domicilio, número telefónico, límite de crédito y saldo de las cuentas por cobrar
Perfiles de clientes	Edad, género, nivel de ingresos, número de miembros en la familia, domicilio y preferencias

TABLA 1.3 Ejemplos de datos disponibles de algunas agencias gubernamentales

Agencia gubernamental	Algunos datos disponibles
Oficina del Censo	Datos poblacionales, número de familias e ingresos por familia
Consejo de la Reserva Federal	Datos sobre la masa monetaria, crédito a plazo, tipos de cambio y tasas de descuento
Oficina de Administración y Presupuesto	Datos sobre ingresos, gastos y deudas del gobierno federal
Departamento de Comercio	Datos sobre la actividad comercial, valor de las remesas por industria, nivel de utilidades por industria e industrias en crecimiento y en declive
Oficina de Estadísticas Laborales	Gasto de los consumidores, ingresos por hora, tasa de desempleo, registros de seguridad y estadísticas internacionales

sindicatos. La tabla 1.3 lista algunas agencias gubernamentales seleccionadas y algunos de los datos que proporcionan. La mayoría de las dependencias que recaba y procesa datos también los pone a disposición de los usuarios por medio de un sitio web. La figura 1.3 muestra la página principal del sitio web de la Oficina de Estadísticas Laborales de Estados Unidos.

Estudios estadísticos

Se cree que el estudio estadístico experimental más grande jamás realizado es el experimento para la vacuna Salk contra la polio del Servicio de Salud Pública efectuado en 1954. Se seleccionaron casi 2 millones de niños de 1o., 2o. y 3er. grados de enseñanza elemental de todo Estados Unidos.

Algunas veces los datos necesarios para alguna aplicación no están disponibles a través de las fuentes existentes. En estos casos suelen obtenerse mediante estudios estadísticos, los cuales se clasifican en *experimentales* u *observacionales*.

En un estudio experimental se identifica primero la variable de interés. Luego se toma una o más variables y se controlan para obtener datos de cómo influyen en la variable de interés. Por ejemplo, una compañía farmacéutica podría interesarse en realizar un experimento para enterarse de cómo afecta un nuevo medicamento la presión sanguínea. Ésta es la variable de interés en el estudio. El nivel de dosis del medicamento nuevo es otra variable que se espera que tenga un efecto causal en la presión sanguínea. Para obtener datos sobre el efecto del nuevo fármaco, los investigadores seleccionan una muestra de individuos. El nivel de dosis del medicamento está controlado, ya que a los distintos grupos de individuos se les suministran dosis diferentes.

FIGURA 1.3 Página principal de la Oficina de Estadísticas Laborales de Estados Unidos

The screenshot shows the homepage of the Bureau of Labor Statistics. The header includes the logo and name of the Bureau, along with navigation links like 'A to Z Index', 'FAQs', 'About BLS', and 'Contact Us'. A search bar is present in the top right corner. The main content area features several news items with dates and titles, such as 'JUN 27 May jobless rates down over the year in 331 of 372 metro areas; payroll jobs up in 266'. A sidebar on the right promotes BLS on Twitter. At the bottom, there is a 'read more »' link and a pagination control for archives.

CAPÍTULO 3

Estadística descriptiva: Medidas numéricas

CONTENIDO

ESTADÍSTICA EN LA PRÁCTICA:
SMALL FRY DESIGN

3.1 MEDIDAS DE POSICIÓN O LOCALIZACIÓN

Media
Media ponderada
Mediana
Media geométrica
Moda
Percentiles
Cuartiles

3.2 MEDIDAS DE VARIABILIDAD

Rango
Rango intercuartílico o intercuartil
Varianza
Desviación estándar
Coeficiente de variación

3.3 MEDIDAS DE LA FORMA DE LA DISTRIBUCIÓN, POSICIÓN RELATIVA Y DETECCIÓN DE OBSERVACIONES ATÍPICAS

Forma de la distribución
Valor z

Teorema de Chebyshev

Regla empírica

Detección de observaciones
atípicas

3.4 RESÚMENES DE CINCO NÚMEROS Y DIAGRAMAS DE CAJA

Resumen de cinco números
Diagrama de caja

3.5 MEDIDAS DE ASOCIACIÓN ENTRE DOS VARIABLES

Covarianza
Interpretación de la covarianza
Coeficiente de correlación
Interpretación del coeficiente
de correlación

3.6 DATA DASHBOARDS O TABLEROS DE DATOS: INCORPORACIÓN DE MEDIDAS NUMÉRICAS PARA MEJORAR SU EFICACIA

ESTADÍSTICA *en* LA PRÁCTICA

SMALL FRY DESIGN*

SANTA ANA, CALIFORNIA

Small Fry Design, fundada en 1997, es una empresa de juguetes y accesorios que diseña e importa productos para niños. Su línea de artículos incluye osos de peluche, móviles, juguetes musicales, sonajas y cobertores de seguridad, y presenta diseños de juguetes de alta calidad para bebé con un énfasis en los colores, las texturas y los sonidos. Los productos se diseñan en Estados Unidos y se fabrican en China.

Small Fry Design emplea representantes independientes para la venta de sus productos a minoristas de muebles infantiles, tiendas de accesorios y ropa para niños, negocios de regalos, tiendas departamentales exclusivas e importantes compañías de ventas por catálogo. En la actualidad, los productos de Small Fry Design se distribuyen en más de 1 000 puntos minoristas de venta en todo Estados Unidos.

La administración del flujo de efectivo es una de las actividades más importantes para la operación diaria de esta empresa. Garantizar que dicho flujo entrante sea suficiente para cumplir con las obligaciones de deudas tanto corrientes como a corto plazo puede significar la diferencia entre el éxito y el fracaso. Un factor crítico en la administración del flujo de efectivo es el análisis y control de las cuentas por cobrar. Al medir el tiempo promedio de cobro y el valor monetario de las facturas pendientes, la gerencia puede predecir la disponibilidad de efectivo y monitorear los cambios en el estado de las cuentas por cobrar. La empresa estableció las metas siguientes: la antigüedad promedio de las facturas pendientes no debe exceder los 45 días y el valor de las facturas con una antigüedad mayor de 60 días no debe exceder 5% del valor de todas las cuentas por cobrar.

En un resumen reciente del estado de las cuentas por cobrar se proporcionó los siguientes estadísticos descriptivos para la antigüedad de las facturas pendientes.

Media	40 días
Mediana	35 días
Moda	31 días

* Los autores agradecen a John A. McCarthy, presidente de Small Fry Design, por proporcionar este artículo para *Estadística en la práctica*.



Small Fry Design aplica la estadística descriptiva para monitorear sus cuentas por cobrar y su flujo de efectivo entrante. © Robert Dant/Alamy Limited.

La interpretación de estos datos estadísticos muestra que el tiempo promedio de cobro de una factura es de 40 días. La mediana señala que la mitad de estos documentos permanece pendiente 35 días o más. La moda de 31 días, el tiempo de cobro de una factura más frecuente, indica que el lapso más común en que ésta permanece pendiente es de 31 días. El resumen estadístico indica también que sólo 3% del valor de todas las cuentas por cobrar tiene un tiempo de cobro de más de 60 días. Con base en la información estadística, la gerencia quedó satisfecha, dado que las cuentas por cobrar y el flujo de efectivo entrante estaban bajo control.

En este capítulo aprenderá a calcular e interpretar algunas de las medidas estadísticas que utiliza Small Fry Design. Además de la media, la mediana y la moda, aprenderá otros estadísticos descriptivos, como el rango, la varianza, la desviación estándar, los percentiles y la correlación. Estas medidas numéricas ayudan a la comprensión e interpretación de los datos.

En el capítulo 2 se estudiaron las presentaciones tabulares y gráficas utilizadas para resumir los datos. En este capítulo se presentan varias medidas numéricas que proporcionan otras opciones para la misma tarea.

Primero se verá el desarrollo de medidas numéricas para conjuntos de datos que constan de una sola variable. Cuando un conjunto de datos contiene más de una variable, las mismas medidas numéricas se calculan por separado para cada variable. Sin embargo, en el caso de dos variables, se desarrollarán también medidas de la relación entre éstas.

Se presentan las medidas numéricas de posición, dispersión, forma y asociación. Si las medidas se calculan para los datos de una muestra, se les llama **estadístico muestral**. Si se calculan para los datos de una población, se les llama **parámetros poblacionales**. En la inferencia estadística, un estadístico muestral se conoce como **estimador puntual** del parámetro poblacional correspondiente. En el capítulo 7 se verá con más detalle el proceso de la estimación puntual.

En los tres apéndices del capítulo se explica cómo se usan Minitab, Excel y StatTools para calcular las medidas numéricas descritas en el capítulo.

3.1

Medidas de posición o localización

Media

La media también se conoce como media aritmética.

La **media**, o valor medio, es quizá la medida de posición más importante para una variable, pues proporciona una medida de la ubicación central de los datos. Si los datos son para una muestra, la media se denota \bar{x} ; si son para una población, se denota con la letra griega μ .

En las fórmulas estadísticas se acostumbra denotar el valor de la primera observación de la variable x mediante x_1 , el valor de la segunda observación de la variable x por medio de x_2 , y así sucesivamente. En general, el valor de la i -ésima observación de la variable x se representa por medio de x_i . Si se tiene una muestra con n observaciones, la fórmula para la media muestral es la siguiente:

La media muestral \bar{x} es un estadístico muestral.

MEDIA MUESTRAL

$$\bar{x} = \frac{\sum x_i}{n} \quad (3.1)$$

En la fórmula anterior, el numerador es la suma de los valores de las n observaciones. Es decir,

$$\sum x_i = x_1 + x_2 + \cdots + x_n$$

La letra griega Σ es el signo de sumatoria o sumatorio.

Para ilustrar el cálculo de una media muestral, considere los datos siguientes sobre el tamaño del grupo para una muestra de cinco grupos de estudiantes universitarios.

46 54 42 46 32

La notación x_1, x_2, x_3, x_4, x_5 se utiliza para representar el número de estudiantes en cada uno de los cinco grupos.

$$x_1 = 46 \quad x_2 = 54 \quad x_3 = 42 \quad x_4 = 46 \quad x_5 = 32$$

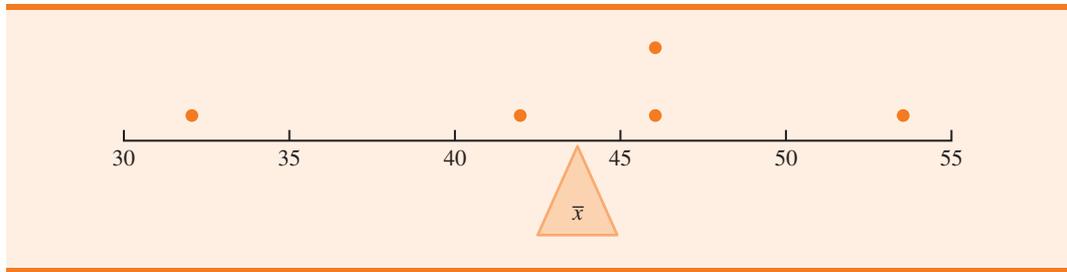
Por consiguiente, para calcular la media muestral se escribe

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{46 + 54 + 42 + 46 + 32}{5} = 44$$

La media muestral del tamaño del grupo es 44 estudiantes.

Para proporcionar una perspectiva visual de la media y mostrar cómo puede ser influida por valores extremos, observe el diagrama de puntos para el tamaño de un grupo que se muestra en la figura 3.1. Considere el eje horizontal del diagrama de puntos como una larga tabla estrecha

FIGURA 3.1 La media como centro de balance para el diagrama de puntos del tamaño de un aula de clase



en la que cada punto tiene el mismo peso fijo y la media es el punto en el que colocaríamos un soporte o punto de giro bajo la tabla para mantener en equilibrio el diagrama de puntos. Este es el mismo principio que usa un sube y baja en un parque, la única diferencia radica en que este juego tiene su soporte en el centro, para que cuando un extremo suba el otro baje. Ahora considere lo que ocurriría en la balanza si el número más grande fuera 114 en vez de 54. Para restablecer el equilibrio, tendríamos que mover hacia una dirección positiva el soporte debajo de un nuevo punto. Para determinar qué tanto debe moverse el soporte, se debe calcular la media muestral para el tamaño modificado del grupo.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{46 + 114 + 42 + 46 + 32}{5} = \frac{280}{5} = 56$$

De esta manera, la media para la nueva muestra de tamaño del grupo es 56, con un incremento de 12 estudiantes. En otras palabras, deberíamos mover 12 unidades el soporte de nuestra balanza bajo el nuevo diagrama de puntos.

Otro ejemplo del cálculo de una media muestral se da en la situación siguiente. Suponga que una agencia de colocación de empleos a nivel universitario envió un cuestionario a una muestra de licenciados en administración de empresas recién egresados (Graduate) solicitando información sobre los sueldos mensuales iniciales (Monthly Starting Salary). La tabla 3.1 muestra los datos recabados. El sueldo mensual inicial medio para la muestra de 12 licenciados en administración de empresas se calcula como sigue:

$$\begin{aligned} \bar{x} &= \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \cdots + x_{12}}{12} \\ &= \frac{3850 + 3950 + \cdots + 3880}{12} \\ &= \frac{47280}{12} = 3940 \end{aligned}$$

TABLA 3.1 Sueldos mensuales iniciales de una muestra de 12 licenciados en administración de empresas recién egresados

Graduate	Monthly Starting Salary (\$)	Graduate	Monthly Starting Salary (\$)
1	3850	7	3890
2	3950	8	4130
3	4050	9	3940
4	3880	10	4325
5	3755	11	3920
6	3710	12	3880

La ecuación (3.1) ilustra cómo se calcula la media para una muestra con n observaciones. La fórmula para determinar la media de una población es la misma, pero se emplea una notación diferente para indicar que se está trabajando con toda la población. El número de observaciones en una población se denota con N y el símbolo para la media poblacional es μ .

La media muestral \bar{x} es un estimador puntual de la media poblacional μ .

MEDIA POBLACIONAL

$$\mu = \frac{\sum x_i}{N} \quad (3.2)$$

Media ponderada

En la utilización de las fórmulas para calcular la media muestral y la media poblacional, se concede la misma importancia o peso a cada x_i . Por ejemplo, la fórmula de la media muestral podría escribirse de la siguiente manera:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1}{n} \left(\sum x_i \right) = \frac{1}{n} (x_1 + x_2 + \cdots + x_n) = \frac{1}{n} (x_1) + \frac{1}{n} (x_2) + \cdots + \frac{1}{n} (x_n)$$

Esto indica que cada uno de los elementos de la muestra tiene un peso de $1/n$. Aunque esta es la práctica más común, en algunos casos la media se calcula asignando a cada elemento un peso que refleja su importancia relativa. Cuando se calcula así la media, se conoce como **media ponderada**, que se calcula de la siguiente manera:

MEDIA PONDERADA

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} \quad (3.3)$$

donde

w_i = peso a observar de i

Cuando los datos provienen de una muestra, la ecuación (3.3) proporcionará como resultado la media muestral ponderada. Si los datos provienen de una población, μ , reemplaza a \bar{x} y la ecuación (3.3) proporcionará como resultado una media poblacional ponderada.

Para ejemplificar la necesidad de una media ponderada, considere la siguiente muestra de cinco compras de materia prima durante los últimos tres meses:

Compra	Costo por libra (\$)	Número de libras
1	3.00	1200
2	3.40	500
3	2.80	2750
4	2.90	1000
5	3.25	800

Observe que el costo por libra varía entre \$2.80 y \$3.40, y la cantidad adquirida varía entre 500 y 2750 libras. Suponga que un gerente desea conocer el costo medio por libra de materia prima. Dado que las cantidades solicitadas varían, debemos utilizar la fórmula de la media

ponderada. Los cinco valores de costo por libra son $x_1 = 3.00$, $x_2 = 3.40$, $x_3 = 2.80$, $x_4 = 2.90$ y $x_5 = 3.25$. Para determinar la media ponderada del costo por libra se debe asignar un peso a cada costo, con base en su cantidad correspondiente. Para este ejemplo, los pesos son $w_1 = 1200$, $w_2 = 500$, $w_3 = 2750$, $w_4 = 1000$ y $w_5 = 800$. Con base en la ecuación (3.3), la media ponderada se calcula así:

$$\begin{aligned}\bar{x} &= \frac{1200(3.00) + 500(3.40) + 2750(2.80) + 1000(2.90) + 800(3.25)}{1200 + 500 + 2750 + 1000 + 800} \\ &= \frac{18500}{6250} = 2.96\end{aligned}$$

De este modo, el cálculo de la media ponderada indica que el costo medio por libra de materia prima es de \$2.96. Observe que si utilizáramos la ecuación (3.1) en vez de la ecuación para la media ponderada (3.3), ello conduciría a un resultado erróneo. En este caso, la media muestral de los cinco valores de costo por libra es $(3.00 + 3.40 + 2.80 + 2.90 + 3.25)/5 = 15.35/5 = \3.07 , lo que exagera el costo medio real por libra adquirida.

La elección de las ponderaciones para un cálculo específico de media ponderada depende de la aplicación. Un ejemplo bien conocido por los estudiantes universitarios es el cálculo de una calificación promedio (grade point average, GPA). En este cálculo, los valores que se usan por lo general para los datos son 4 para una calificación de A, 3 para B, 2 para C, 1 para D y 0 para F. Estos pesos significan el número de horas acreditadas para cada calificación. El ejercicio 16 al final de esta sección proporciona un ejemplo para este cálculo de media ponderada. En otros cálculos de media ponderada, las cantidades como libras, dinero o volumen se emplean por lo general como pesos. En cualquiera de los casos, cuando los elementos varían en importancia, el analista debe elegir el peso que refleje mejor la importancia de cada elemento para determinar de la media.

Mediana

La **mediana** es otra medida de posición central; es el valor de en medio cuando los datos están acomodados en orden ascendente (del valor menor al mayor). Con un número impar de observaciones, la mediana es el valor de en medio. Con un número par, no hay valor de en medio. En este caso se sigue la convención y la mediana se define como el promedio de los valores de las dos observaciones de en medio. Por conveniencia, la definición de la mediana se replantea como sigue.

MEDIANA

Ordene los datos de forma ascendente (del valor menor al mayor).

- a) Para un número impar de observaciones, la mediana es el valor de en medio.
- b) Para un número par de observaciones, la mediana es el promedio de los dos valores de en medio.

Esta definición se aplica para calcular la mediana de los tamaños de grupo para la muestra de cinco grupos de estudiantes universitarios. Al ordenar los datos de forma ascendente se obtiene la lista siguiente:

32 42 46 46 54

Dado que $n = 5$ es impar, la mediana es el valor de en medio. Así, la mediana del tamaño del grupo es 46 estudiantes. Aun cuando este conjunto de datos contiene dos observaciones con valores de 46, cada una se trata de forma separada cuando los datos se acomodan en orden ascendente.

Suponga además que se calcula la mediana de los sueldos iniciales para los 12 licenciados en administración de empresas de la tabla 3.1. Primero se acomodan los datos en orden ascendente.

3710 3755 3850 3880 3880 3890 3920 3940 3950 4050 4130 4325
 Los dos valores de en medio

Como $n = 12$ es par, se identifican los dos valores de en medio: la mediana es el promedio de estos dos valores.

$$\text{Mediana} = \frac{3890 + 3920}{2} = 3905$$

El procedimiento que se utiliza para calcular la mediana depende de si el número de elementos es par o impar. A continuación se describe un enfoque más conceptual y visual utilizando el sueldo inicial de los 12 graduados en administración de empresas. Al igual que se hizo antes, comenzaremos por ordenar los datos en forma ascendente.

3710 3755 3850 3880 3880 3890 3920 3940 3950 4050 4130 4325

Una vez que los datos se encuentran en orden ascendente, se eliminan por pares los valores extremos altos y bajos, hasta que no queden pares de valores sin eliminar por completo los datos. Por ejemplo, después de eliminar el elemento menor (3710) y el mayor (4325), se obtiene un nuevo conjunto de datos con 10 elementos.

~~3710~~ 3755 3850 3880 3880 3890 3920 3940 3950 4050 4130 ~~4325~~

Al eliminar el valor menor que sigue (3755) y el siguiente valor mayor (4130), obtendremos un nuevo conjunto de datos con ocho observaciones.

~~3710~~ ~~3755~~ 3850 3880 3880 3890 3920 3940 3950 4050 ~~4130~~ ~~4325~~

Al continuar con el procedimiento, se obtienen los siguientes resultados:

~~3710~~ ~~3755~~ ~~3850~~ 3880 3880 3890 3920 3940 3950 ~~4050~~ ~~4130~~ ~~4325~~
~~3710~~ ~~3755~~ ~~3850~~ ~~3880~~ 3880 3890 3920 3940 ~~3950~~ ~~4050~~ ~~4130~~ ~~4325~~
~~3710~~ ~~3755~~ ~~3850~~ ~~3880~~ ~~3880~~ 3890 3920 ~~3940~~ ~~3950~~ ~~4050~~ ~~4130~~ ~~4325~~

En este punto no es posible continuar sin eliminar todos los datos, por lo tanto, la mediana es el promedio de los dos valores restantes. Cuando existe un número par de elementos, el proceso de eliminación siempre concluirá con dos valores restantes, y el promedio de estos dos valores será la mediana. Cuando existe un número impar de elementos, el proceso de eliminación siempre terminará en un valor final, que será la mediana. Por lo anterior, este método funciona tanto para un número de elementos par como para uno impar.

Aunque la media es la medida de posición central de uso más común, en algunas situaciones se prefiere la mediana, ya que los valores de datos muy pequeños y muy grandes influyen en la media. Por ejemplo, suponga que uno de los licenciados recién graduados (tabla 3.1) tenía un sueldo inicial de \$10000 al mes (tal vez la empresa es propiedad de su familia). Si se cambia el sueldo mensual inicial más alto de la tabla 3.1 de \$4325 a \$10000 y se vuelve a calcular la media, la media muestral pasa de \$3940 a \$4413. Sin embargo, la mediana de \$3905 permanece igual, ya que \$3890 y \$3920 siguen siendo los dos valores de en medio. Si el sueldo inicial es sumamente alto, la mediana proporciona una mejor medida de posición central que la media. Al hacer una generalización, se afirma que siempre que un conjunto de datos contiene valores extremos, la mediana suele ser la medida preferida de posición central.

La mediana es la medida de posición que más se utiliza para los datos de los ingresos anuales y el valor de propiedad, debido a que algunos ingresos o valores de propiedad muy grandes pueden inflar la media. En tales casos, la mediana es la medida preferida de posición central.

El propósito de este libro es proporcionar una introducción a los conceptos clave de la estadística, para lo cual emplea una orientación fundamental hacia las aplicaciones del análisis de datos y la metodología. El análisis y el desarrollo de cada técnica se presentan desde esta perspectiva, cuyos resultados estadísticos permiten comprender las decisiones y soluciones de los problemas.

Entre los cambios relevantes en esta edición, se pueden mencionar los siguientes:

- **ESTADÍSTICA DESCRIPTIVA.** Se incorpora material nuevo sobre visualización de datos, mejores prácticas, gráficas de barras apiladas y de barras agrupadas. También se integra una nueva sección sobre dashboards o tableros de datos y cómo incorporar resúmenes estadísticos para mejorar su eficacia.
- **DISTRIBUCIONES DISCRETAS.** Se incluye una nueva sección sobre distribuciones bivariadas discretas y sus aplicaciones en las finanzas, por ejemplo, cómo elaborar y analizar portafolios financieros por medio de estas distribuciones.
- **COMPARACIÓN DE MÚLTIPLES PROPORCIONES, PRUEBAS DE INDEPENDENCIA Y DE BONDAD DE AJUSTE.** Se incorpora una nueva sección sobre pruebas de igualdad para tres o más proporciones poblacionales, así como un procedimiento para llevar a cabo pruebas de comparaciones múltiples entre todos los pares de proporciones poblacionales. También se incluyen apéndices revisados con instrucciones detalladas para Minitab, Excel y StatTools.
- **NUEVOS CASOS PRÁCTICOS.** Se incluyen nuevos casos prácticos, los cuales brindan la oportunidad de analizar bancos de datos más grandes y preparar los informes gerenciales con base en los resultados del análisis.
- **NUEVAS APLICACIONES REALES.** Cada capítulo comienza con un caso real. Por ejemplo, en el capítulo 4 se describe cómo un equipo de la NASA aplicó la probabilidad para ayudar en el conocido rescate de los 33 mineros chilenos atrapados en un derrumbe.
- **NUEVOS EJEMPLOS Y EJERCICIOS CON DATOS REALES.** Se desarrollan explicaciones y ejercicios para mostrar los múltiples usos de la estadística en los negocios a partir de estudios reales publicados en fuentes como *The Wall Street Journal*, *USA Today* y *Barron's*. En total, esta edición contiene más de 350 ejemplos y ejercicios de este tipo.