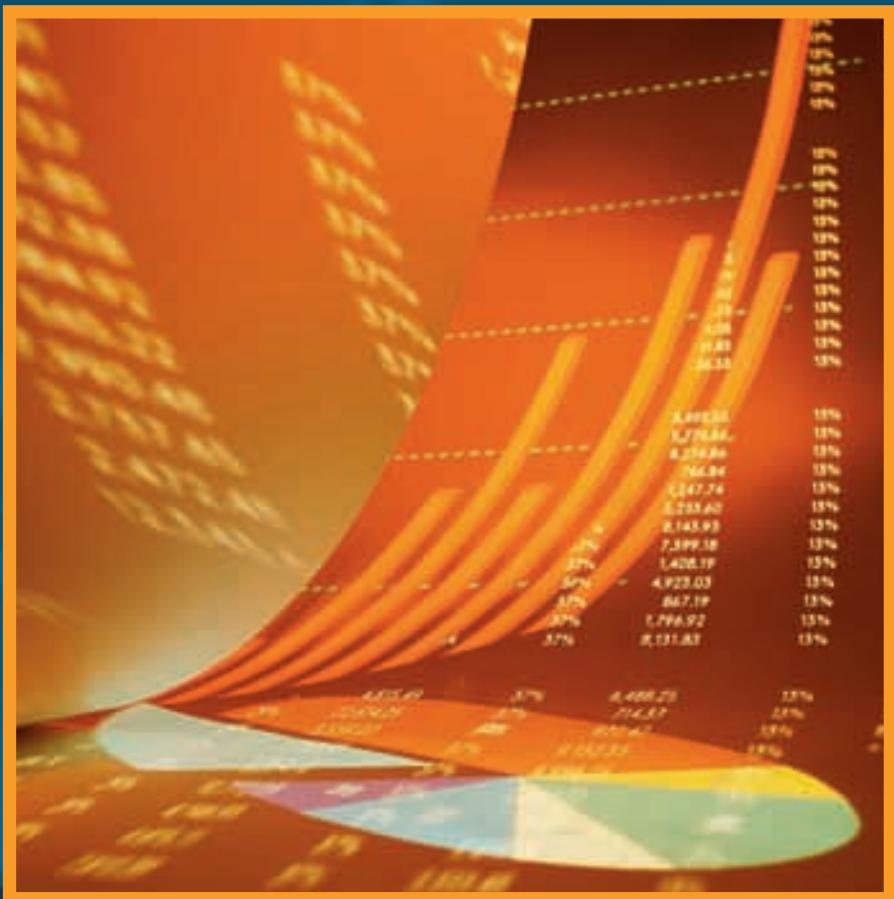


# PRONÓSTICOS EN LOS NEGOCIOS

OCTAVA EDICIÓN



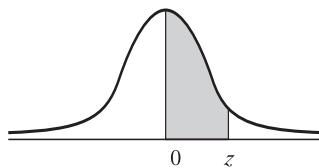
JOHN E. HANKE

DEAN W. WICHERN





## Tabla de áreas para la distribución de probabilidad normal estándar



Para  $z = 1.93$ , el área sombreada equivale a 0.4732 del área total de 1.

Octava  
edición

# PRONÓSTICOS EN LOS NEGOCIOS

John E. Hanke

*Eastern Washington University, Emeritus*

Dean W. Wichern

*Texas A&M University*

**Traducción:**

Astrid Mues Zepeda

Ekaterina Guerrero Ushakova  
*Traductor as profesionales*

**Revisión técnica:**

Andrés González Nucamendi

*Instituto Tecnológico de Estudios  
Superiores de Monterrey  
Campus Ciudad de México  
Profesor de planta  
Coordinador de pronósticos  
para la toma de decisiones*



México • Argentina • Brasil • Colombia • Costa Rica • Chile • Ecuador  
España • Guatemala • Panamá • Perú • Puerto Rico • Uruguay • Venezuela

Datos de catalogación bibliográfica	
<b>HANKE, JOHN E.</b>	
<b>Pronósticos en los negocios. Octava edición</b>	
PEARSON EDUCACIÓN, México, 2006	
ISBN: 970-26-0759-0	
Área: Administración	
Formato: 20 x 25.5 cm	Páginas: 552

Authorized translation from the English language edition, entitled *Business forecasting* 8th ed., by John E. Hanke and Dean W. Wichern, published by Pearson Education, Inc., publishing as PRENTICE HALL, INC., Copyright © 2005. All rights reserved.

ISBN 0-13-141290-6

Traducción autorizada de la edición en idioma inglés, titulada *Pronósticos en los negocios* 8/e de John E. Hanke and Dean W. Wichern publicada por Pearson Education, Inc., publicada como PRENTICE HALL INC., Copyright © 2005. Todos los derechos reservados.

Esta edición en español es la única autorizada.

#### **Edición en español**

Editor: Marisa de Anta

e-mail: marisa.anta@pearsoned.com

Supervisor de desarrollo: Astrid Mues Zepeda

Supervisor de producción: José D. Hernández Garduño

#### **Edición en inglés**

Executive Editor: Tom Tucker

Editor-in-Chief: P. J. Boardman

Editorial Assistant: Dawn Stapleton

Senior Media Project Manager: Nancy Welcher

Executive Marketing Manager: Debbie Clare

Marketing Assistant: Amanda Fisher

Managing Editor (Production): John Roberts

Production Editor: Kelly Warsak

Production Assistant: Joe DeProspero

Manufacturing Buyer: Michelle Klein

Cover Design: Bruce Kenselaar

Cover Illustration: Getty Images

Composition: Laserwords Private Ltd, India

Full-Service Project Management: Tempe Goodhue/nSight, Inc.

Printer/Binder: Hamilton Printing Company

Cover Printer: Phoenix Color Corp.

#### OCTAVA EDICIÓN, 2006

D.R. © 2006 por Pearson Educación de México, S.A. de C.V.

Atlacomulco núm. 500 – 5º piso

Col. Industrial Atoto

53519 Naucalpan de Juárez, Edo. de México

Cámara Nacional de la Industria Editorial Mexicana. Reg. Núm. 1031

Reservados todos los derechos. Ni la totalidad ni parte de esta publicación pueden reproducirse, registrarse o transmitirse, por un sistema de recuperación de información, en ninguna forma ni por ningún medio, sea electrónico, mecánico, fotoquímico, magnético o electroóptico, por fotocopia, grabación o cualquier otro, sin permiso previo por escrito del editor.

El préstamo, alquiler o cualquier otra forma de cesión de uso de este ejemplar requerirá también la autorización del editor o de sus representantes.

ISBN 970-26-0759-0

Impreso en México. *Printed in Mexico.*

1 2 3 4 5 6 7 8 9 0 - 09 08 07 06



Dedicado al recuerdo de Harry  
(quien en realidad no quería leerlo);

Geri  
(quien no necesita hacerlo);

Donna y Dorothy  
(quienes probablemente no lo hagan);

Katrina, Michael y Andrew  
(quienes quizá sí lo hagan);

Kevin  
(quien dice haberlo hecho);

y todos nuestros alumnos  
(a quienes más les vale hacerlo).



# ||| Contenido

## Prefacio xiii

## CAPÍTULO 1 Introducción a los pronósticos 1

Historia de los pronósticos	1
La necesidad de los pronósticos	1
Tipos de pronósticos	3
Consideraciones de pronósticos macroeconómicos	4
Elección de un método de pronósticos	4
Pasos de los pronósticos	5
Administración del proceso de pronósticos	6
Software para pronósticos	7
Información en línea	8
Ejemplo de un pronóstico	8
Resumen	9
Caso 1-1 Mr. Tux	10
Caso 1-2 Consumer Credit Counseling	10
Aplicaciones en Minitab	11
Aplicaciones de Excel	12
Referencias	12

## CAPÍTULO 2 Repaso de conceptos estadísticos básicos 15

Estadística descriptiva	15
Presentación de la información numérica	19
Distribuciones de probabilidad	22
Distribuciones muestrales	26
Inferencia a partir de una muestra	28
<i>Estimación</i>	28
Prueba de hipótesis	29
<i>Valor p</i>	31
Análisis de correlación	32
<i>Diagramas de dispersión</i>	32
Coeficiente de correlación	35

Ajuste de una línea recta	37
Evaluación de la normalidad	39
Aplicación a la administración	42
Glosario	43
Fórmulas clave	43
Problemas	45
Caso 2-1 Alcom electronics	50
Caso 2-2 Mr. Tux	51
Caso 2-3 Alomega food stores	52
Aplicaciones en Minitab	53
Aplicaciones en Excel	55
Referencias	56

### **CAPÍTULO 3 Exploración de patrones de datos y elección de una técnica de pronóstico 57**

Exploración de patrones de datos de series de tiempo	58
Exploración de patrones de datos con análisis de autocorrelación	60
<i>¿Son aleatorios los datos?</i>	65
<i>¿Tienen tendencia los datos?</i>	67
<i>¿Son estacionales los datos?</i>	69
Elección de una técnica de pronósticos	74
<i>Técnicas de pronóstico para datos estacionarios</i>	75
<i>Técnicas de pronóstico para datos con tendencia</i>	75
<i>Técnicas de pronóstico para datos estacionales</i>	76
<i>Técnicas de pronóstico para series cíclicas</i>	76
<i>Otros factores a considerar cuando se elige una técnica de pronósticos</i>	77
<i>Evaluación empírica de los métodos de pronóstico</i>	77
Medición del error de pronóstico	78
<i>Notación básica de pronósticos</i>	79
Determinación de la idoneidad de una técnica de pronóstico	81
Aplicación a la administración	83
Glosario	84
Fórmulas clave	84
Problemas	85
Caso 3-1A Murphy Brothers Furniture	90
Caso 3-1B Murphy Brothers Furniture	92
Caso 3-2 Mr. Tux	92
Caso 3-3 Consumer Credit Counseling	94
Caso 3-4 Alomega Food Stores	94
Aplicaciones en Minitab	95

Aplicaciones en Excel	98
Referencias	100

## CAPÍTULO 4 Promedios móviles y métodos de suavizamiento 101

Modelos informales	102
Métodos de pronósticos basados en promedios	105
<i>Promedios simples</i>	105
<i>Promedios móviles</i>	107
<i>Promedios móviles dobles</i>	110
Métodos de suavizamiento exponencial	114
<i>Suavizamiento exponencial ajustado a la tendencia: método de Holt</i>	121
<i>Suavizamiento exponencial ajustado para variaciones de tendencia y estacionales: método de Winters</i>	126
Aplicación a la administración	130
Glosario	131
Fórmulas clave	131
Problemas	133
Caso 4-1 La Solar Alternative Company 4	139
Caso 4-2 Mr. Tux	140
Caso 4-3 Consumer Credit Counseling	141
Caso 4-4 Murphy Brothers Furniture	141
Caso 4-5 Proyección de utilidades para cinco años de Downtown Radiology	142
Aplicaciones en Minitab	148
Aplicaciones en Excel	150
Aplicaciones en Excel: CB Predictor	151
Referencias	155

## CAPÍTULO 5 Series de tiempo y sus componentes 157

Descomposición	158
Tendencia	160
<i>Curvas de tendencia no lineales</i>	164
<i>Pronóstico de la tendencia</i>	166
<i>Estacionalidad</i>	167
Datos ajustados a la estacionalidad	171
<i>Variaciones cíclicas e irregulares</i>	172
Pronóstico de una serie de tiempo estacional	177
Método de descomposición de censo II	179
Aplicación a la administración	181
Apéndice: índice de precios	182
Glosario	184

Fórmulas clave	184
Problemas	185
Caso 5-1 Small Engine Doctor	191
Caso 5-2 Mr. Tux	192
Caso 5-3 Consumer Credit Counseling	196
Caso 5-4 Murphy Brothers Furniture	197
Caso 5-5 AAA Washington	200
Caso 5-6 Alomega Food Stores	202
Aplicaciones en Minitab	203
Aplicaciones en Excel	206
Referencias	209

## **CAPÍTULO 6 Regresión lineal simple 211**

Línea de regresión	212
Error estándar de la estimación	216
Pronóstico de $Y$	217
Descomposición de la varianza	220
Coeficiente de determinación	224
Prueba de hipótesis	226
Análisis de residuales	229
Resultados en computadora al usar minitab	231
Transformación de variables	233
Curvas de crecimiento	237
Aplicación a la administración	242
Glosario	243
Fórmulas clave	243
Problemas	245
Caso 6-1 Tiger Transport	254
Caso 6-2 Butcher Products, Inc.	256
Caso 6-3 Ace Manufacturing	257
Caso 6-4 Mr. Tux	258
Caso 6-5 Consumer Credit Counseling	259
Caso 6-6 AAA Washington	260
Aplicaciones en Minitab	262
Aplicaciones en Excel	265
Referencias	267

## **CAPÍTULO 7 Análisis de regresión múltiple 269**

Algunas variables explicativas	269
Matriz de correlación	270

Modelo de regresión múltiple	271
<i>Modelo estadístico para regresión múltiple</i>	271
Interpretación de los coeficientes de regresión	273
Inferencias en modelos de regresión múltiple	274
<i>Error estándar de la estimación</i>	275
<i>Significancia de la regresión</i>	276
<i>Variables explicativas individuales</i>	278
<i>Pronóstico de una respuesta futura</i>	279
Resultado en computadora	280
Variables ficticias	281
Multicolinealidad	285
Selección de la “mejor” ecuación de regresión	288
<i>Todas las regresiones posibles</i>	290
<i>Regresión por pasos</i>	292
<i>Notas finales sobre la regresión por pasos</i>	294
Diagnósticos de regresión y análisis residual	295
Advertencias en el pronóstico	297
<i>Sobreajuste</i>	297
<i>Regresiones útiles, valores grandes calculados de F</i>	298
Aplicaciones a la administración	298
Glosario	300
Fórmulas clave	300
Problemas	301
Caso 7-1 El Mercado de Bonos	310
Caso 7-2 AAA Washington	313
Caso 7-3 Beisbol de Fantasía (A)	315
Caso 7-4 Beisbol de Fantasía (B)	320
Aplicaciones en Minitab	324
Aplicaciones en Excel	326
Referencias	326

## **CAPÍTULO 8 Regresión con datos de series de tiempo 327**

Datos de series de tiempo y el problema de la autocorrelación	327
Prueba de Durbin-Watson para correlación serial	331
Soluciones para los problemas de autocorrelación	334
<i>Error de especificación del modelo (omisión de una variable)</i>	335
<i>Regresión con diferencias</i>	337
<i>Errores autocorrelacionados y diferencias generalizadas</i>	342
<i>Modelos autorregresivos</i>	345

Datos de las series de tiempo y el problema de la heteroscedasticidad	346
Uso de la regresión para pronosticar datos estacionales	349
Pronóstico econométrico	352
Aplicación a la administración	353
Glosario	353
Fórmulas clave	353
Problemas	355
Caso 8-1 Compañía de su elección	362
Caso 8-2 Índice de actividad empresarial en el condado de Spokane	363
Caso 8-3 Ventas de un restaurante	367
Caso 8-4 Mr. Tux	369
Caso 8-5 Consumer Credit Counseling	371
Caso 8-6 AAA Washington	372
Caso 8-7 Alomega Food Stores	375
Aplicaciones en Minitab	376
Aplicaciones en Excel	377
Referencias	379
<b>CAPÍTULO 9 La metodología Box-Jenkins (ARIMA) 381</b>	
La metodología Box-Jenkins	381
<i>Modelos autorregresivos</i>	386
<i>Modelos de promedio móvil</i>	387
<i>Modelos de promedios móviles autorregresivos</i>	388
<i>Resumen</i>	389
Aplicación de una estrategia para la construcción de un modelo	389
<i>Paso 1: identificación del modelo</i>	389
<i>Paso 2: estimación de modelos</i>	391
<i>Paso 3: evaluación del modelo</i>	392
<i>Paso 4: realización de pronósticos con el modelo</i>	392
<i>Comentarios finales</i>	411
<i>Criterio para la selección de un modelo</i>	412
<i>Modelos para datos estacionales</i>	414
<i>Suavizamiento exponencial simple y un modelo ARIMA</i>	424
<i>Ventajas y desventajas de los modelos ARIMA</i>	426
Aplicación a la administración	428
Glosario	429
Fórmulas clave	429
Problemas	430
Caso 9-1 Ventas del restaurante	440

Caso 9-2	Mr. Tux	442
Caso 9-3	Consumer Credit Counseling	444
Caso 9-4	Lydia E. Pinkham Medicine Company	444
Caso 9-5	La Ciudad de College Station	447
Caso 9-6	Division de finanzas de Ups Air	450
Caso 9-7	AAA Washington	453
Aplicaciones en Minitab		455
Aplicaciones en Excel: CB predictor		457
Referencias		460

## **CAPÍTULO 10 Pronóstico por apreciación y ajuste de pronóstico 463**

El método Delphi	464	
Formulación de escenarios	467	
Combinación de pronósticos	468	
Redes neurales y pronósticos	470	
Resumen del pronóstico por apreciación subjetiva	472	
Otras herramientas útiles para juzgar el futuro	473	
Fórmulas clave	477	
Problemas	478	
Caso 10-1	Restaurante Golden Gardens	478
Caso 10-2	Alomega Food Stores	479
Caso 10-3	Otra visita a la Lydia E. Pinkham Medicine Company	480
Referencias		482

## **CAPÍTULO 11 Administración del proceso de pronósticos 485**

El proceso de pronósticos	485	
Monitoreo de los pronósticos	486	
Revisión de las etapas del pronóstico	491	
Responsabilidad del proceso de pronóstico	492	
Costos del pronóstico	493	
El pronóstico y el sistema MIS	493	
Pronósticos y administración de las ventas	494	
El pronóstico y su futuro	494	
Problemas	495	
Caso 11-1	Boundary Electronics	495
Caso 11-2	Busby Associates	496
Caso 11-3	Consumer Credit Counseling	499
Caso 11-4	Mr. Tux	500
Caso 11-5	Alomega Food Stores	501
Referencias		501

**APÉNDICE A Deducción de fórmulas**

- Deducción de fórmulas de correlación 503
- Deducción de los estimadores de mínimos cuadrados 503
- Derivadas parciales 503

**APÉNDICE B Datos para el caso 7.1 505**

**APÉNDICE C Tablas 507**

- TABLA C-1 Términos individuales de la distribución binomial 507
- TABLA C-2 Tabla de las áreas para la distribución de probabilidad normal estándar 509
- TABLA C-3 Valores críticos de la estadística *t* de *Student* 510
- TABLA C-4 Valores críticos de la distribución chi cuadrada 511
- TABLA C-5 Tabla de la distribución *F* 513
- TABLA C-6 Límites en la prueba Durbin-Watson 514

**APÉNDICE D Series de datos y bases de datos 517**

**Índice analítico 531**

# Prefacio

El objetivo de la octava edición de *Pronósticos en los negocios* es el mismo que en las ediciones anteriores: presentar las técnicas estadísticas fundamentales que resultan útiles para preparar pronósticos individuales de negocios y planes de largo plazo. El libro está escrito en un estilo simple y directo que hace amplio uso de ejemplos prácticos de negocios. Al final de los capítulos aparecen 53 casos que proporcionan al estudiante el vínculo necesario entre los conceptos teóricos y sus aplicaciones prácticas. Se enfatiza la aplicación de las técnicas de la administración para la toma de decisiones. Asimismo, se supone que los lectores han tomado un curso de introducción a la estadística y están familiarizados con el uso de la computadora, lo que les facilita usar paquetes de software tales como procesadores de palabras y hojas de cálculo.

## ORGANIZACIÓN

---

Todos los capítulos se han revisado para mejorar la claridad del texto y aumentar la efectividad de la enseñanza y el aprendizaje. El contenido está organizado en seis secciones.

La primera (capítulos 1 y 2) presenta material de antecedentes. La naturaleza de los pronósticos y una revisión rápida de conceptos estadísticos básicos establecen las bases para la cobertura de técnicas, la cual comienza en la segunda sección.

La segunda sección (capítulo 3) enfatiza la exploración de patrones de datos y la elección de una técnica de pronósticos. La tercera (capítulos 4 y 5) cubre las técnicas de promedio y suavizamiento, así como una introducción a la descomposición de series de tiempo en términos de sus componentes subyacentes. La cuarta (capítulos 6 y 7) enfatiza las técnicas causales de pronósticos tales como análisis de correlación, regresión lineal simple y regresión múltiple.

La quinta sección (capítulos 8 y 9) estudia las técnicas utilizadas para pronosticar datos de series de tiempo. El libro concluye con una sección final (capítulos 10 y 11) sobre los pronósticos por apreciación y ajustes de pronósticos, así como un análisis sobre la administración y supervisión del proceso de pronósticos.

## CAMBIOS EN LA OCTAVA EDICIÓN

---

La octava edición se ha escrito nuevamente en su totalidad. Aunque el “sabor” de las ediciones anteriores se ha mantenido, se ha puesto mayor énfasis en los desarrollos teóricos y los hallazgos empíricos más recientes. El material obsoleto se ha eliminado y el libro se reorganizó por completo con la incorporación de nuevos problemas, ejemplos, conjuntos de datos y casos. Las siguientes características son nuevas o mejoradas en esta edición:

- Se han añadido siete casos nuevos.
- Se muestra un ejemplo de cómo utilizar Minitab (versión 14) en una situación de pronósticos al final de casi todos los capítulos.

- Se incluye un manual de soluciones en el CD para el alumno que consiste en las respuestas a los problemas nenes de los capítulos.
- Se encuentra disponible un manual actualizado para el instructor.
- Al final de la mayoría de los capítulos se muestra un ejemplo del uso de Excel 2000 para resolver un problema de pronósticos.
- Se incluyen ejemplos que requieren el CB Predictor de Crystal Ball (un complemento para Excel) en el capítulo 4, “Promedios móviles y métodos de suavización”, y en el capítulo 9, “La metodología Box-Jenkins (ARIMA)”.
- Las series de datos para Minitab, Excel y otros programas están disponibles en el CD incluido en este texto. Dichas series de datos también están en Internet.
- Asimismo, en el CD se incluye la versión para el alumno de CB Predictor.
- Se han añadido varios problemas nuevos a esta edición.
- A su vez, las series de datos están actualizadas.
- El material sobre curvas de crecimiento se ha movido del capítulo 10 al capítulo 6.
- En el capítulo 10 se incluye un análisis ampliado del método Delphi.

## LA FUNCIÓN DE LA COMPUTADORA

---

En las primeras siete ediciones, se reconoció a la computadora como una herramienta poderosa en los pronósticos. En la actualidad, la computadora es aún más importante y los gerentes aprovechan la facilidad y disponibilidad de pronósticos sofisticados que se pueden obtener mediante computadoras de escritorio, laptop y portátiles y las capacidades de conectarse en red.

Los autores han pasado varias noches de insomnio decidiendo qué hacer en torno a la computadora. Un estudio nacional de todas las instituciones que pertenecen a la AACSB realizado por los autores a fin de determinar lo que hace el cuerpo docente sobre el uso de computadoras para enseñar los pronósticos, mostró que (1) la mayoría del profesorado (94.2%) intenta dar a los estudiantes experiencia práctica en el uso de la computadora, y (2) varios paquetes de estadística y paquetes personales específicos se mencionaron en la encuesta. Los paquetes citados con más frecuencia fueron Minitab, SAS, Eviews y hojas de cálculo.

Los autores han probado diversos enfoques para ayudar a los docentes y alumnos a utilizar la computadora, a fin de que hagan pronósticos. Esta edición incluye lo siguiente:

1. Se presentan instrucciones para Minitab al final de la mayoría de los capítulos.
2. De la misma manera, se proporcionan instrucciones para Excel.
3. Las instrucciones para CB Predictor están al final de los capítulos 4 y 9.
4. En el CD incluido en este libro, así como en la Internet, se encuentran disponibles tres colecciones de datos (Minitab, Excel y otros programas). Cada colección contiene datos de los ejemplos y problemas del texto. Para entrar a los conjuntos de datos en Internet, vaya a la página web de Prentice Hall en [www.pearsoneducacion.net/hanke](http://www.pearsoneducacion.net/hanke).
5. A lo largo del texto hay ejemplos de diferentes resultados por computadora.

## RECONOCIMIENTOS

---

Los autores están en deuda con los muchos instructores alrededor del mundo que han utilizado las primeras siete ediciones y que han proporcionado sugerencias invaluables para mejorar el libro. Un agradecimiento especial al profesor Frank Forest de Marquette University, al profesor William Darrow de Townsend State University, a Susan Winters de Northwestern State University, al profesor Shik Chun Young de Eastern Washington University; a Mark Craze, Judy Johnson, Steve Brandon y Dorothy Mercer por proporcionar los casos y a Jennifer Dahl por la elaboración del índice.

Partes de este texto, en particular varios conjuntos de datos, se adaptaron de los que aparecieron en la segunda edición de *Understanding Business Statistics* de Hanke y Reitsch (1999), publicado por Richard D. Irwin, a quien le reconocemos aquí por esta reutilización.

También agradecemos a los revisores John Liechty, de la University of Michigan; John Tamura, de University of Washington, y a Ted Taukahara, de St. Mary's University, por sus comentarios tan constructivos en la revisión de este libro. Otros revisores que merecen nuestro agradecimiento son Perry Sadorsky, de York University; Shady Kholdy, de California State Polytechnic University–Pomona; Michael Niemira, de New York University; Fred Zufryden, de University of Southern California, y Haizheg Li, de Georgia Institute of Technology. Si tuviéramos el suficiente talento para lograr todo lo que sugirieron los revisores, el libro mejoraría 100 por ciento.

Por último, les agradecemos a nuestras computadoras y nos preguntamos cómo pudimos escribir un libro sin ellas alguna vez. Nosotros, no las computadoras, somos los responsables de cualquier error.



## CAPÍTULO

# 1

# INTRODUCCIÓN A LOS PRONÓSTICOS

Este libro se centra en los métodos utilizados para predecir la naturaleza incierta de las tendencias en los negocios, en un esfuerzo por ayudar a los administradores a tomar decisiones y a mejorar sus planes. A menudo, tales denuedos implican el estudio y manipulación de datos históricos para buscar los patrones que puedan extrapolarse de manera eficaz para producir pronósticos.

En este texto se les recuerda con regularidad a los lectores que se debe utilizar un juicio razonable, en combinación con resultados numéricos para obtener un buen pronóstico. El ejemplo en este capítulo y los casos prácticos que están al final de cada capítulo enfatizan este punto. Al respecto, se presentan análisis en éste y el capítulo final.

## HISTORIA DE LOS PRONÓSTICOS

---

Muchas de las técnicas de pronóstico utilizadas actualmente y presentadas en este libro se desarrollaron en el siglo XIX; un ejemplo de ellas son los procedimientos de análisis de regresión. En contraste, algunos de los temas en este libro recientemente han captado atención. Los procedimientos de descomposición, suavizamiento y Box-Jenkins entran en esta categoría.

Con el desarrollo de técnicas de pronóstico más sofisticadas, más el advenimiento de las computadoras —en especial la proliferación de la computadora personal (PC) y su *software* asociado— la realización de pronósticos ha recibido cada vez más atención. Ahora, todo administrador tiene la capacidad de utilizar técnicas muy sofisticadas de análisis de datos para fines de pronóstico; en consecuencia, es esencial entenderlas. Por esta razón, los agentes que necesitan de los de pronósticos (por ejemplo, los administradores) deben estar alerta ante el uso inadecuado de las técnicas de pronósticos, ya que cuando éstos son imprecisos pueden llevar a malas decisiones.

A medida que crece la preocupación por el proceso del pronóstico, el desarrollo de nuevas técnicas para pronósticos continúa. Un foco de atención particular al respecto se encuentra en los errores que son inherentes a cualquier procedimiento de pronóstico. Las predicciones de los resultados rara vez son precisas, y quienes pronostican sólo pueden intentar que los errores que se cometen de manera inevitable se minimicen tanto como sea posible.

## LA NECESIDAD DE LOS PRONÓSTICOS

---

En vista de que las imprecisiones no se pueden separar del proceso, ¿por qué son necesarios los pronósticos? La respuesta es que todas las organizaciones operan en una atmósfera de incertidumbre; pero hoy deben tomarse decisiones que afectarán su futuro. Para los administradores de una empresa, las conjeturas sobre el futuro, con base en cierta información, son más valiosas que las presunciones sin bases. Este libro presenta varias maneras de pronosticar que se basan en métodos lógicos de manipulación de datos, los cuales han sido generados por acontecimientos históricos.

Esto no quiere decir que los pronósticos intuitivos sean malos. Al contrario, es frecuente que “las coronadas” de quienes dirigen las organizaciones proporcionen los únicos pronósticos disponibles. Este texto presenta las técnicas de pronóstico que pueden utilizarse para complementar el sentido común y la habilidad directiva de quienes toman las decisiones; los elementos de juicio personal en el proceso de pronósticos se presentan en el capítulo 10. Nuestro punto de vista es que quienes toman las decisiones se encuentran en una mejor posición si entienden las técnicas de pronósticos —tanto cuantitativas como cualitativas— y las utilizan con prudencia; en cambio, obtendrán resultados inferiores si se ven forzados a planear sin tener el beneficio de buenos pronósticos.

La función de los pronósticos por juicio personal ha cambiado en años recientes. Antes de la llegada de las técnicas modernas de pronóstico y el poder de la computadora, el juicio del administrador era la única herramienta de pronóstico posible. No obstante, de acuerdo con Makridakis (1986), los pronósticos generados al utilizar sólo el juicio personal no son tan precisos como aquellos que combinan los juicios personales con las técnicas cuantitativas:

Los humanos poseen conocimientos únicos e información interna que no se encuentran en los métodos cuantitativos. Sin embargo, los estudios empíricos y los experimentos de laboratorio han demostrado sorpresivamente que esos pronósticos no tienen mayor precisión que aquellos generados por métodos cuantitativos. Los humanos tienden a ser optimistas y subestiman la incertidumbre futura. Además, el costo de los pronósticos con métodos de juicio personal comúnmente es más alto que cuando se utilizan métodos cuantitativos.<sup>1</sup>

Se cree que es más eficiente el pronosticador capaz de formular una mezcla habiliosa de técnicas cuantitativas de pronósticos con un buen juicio personal y que evita los extremos donde se depende completamente de cualquiera de las dos. En un extremo se encuentra el ejecutivo que, debido a su ignorancia y miedo a las técnicas cuantitativas y a las computadoras, depende únicamente de su intuición y sentimientos. En el otro extremo está un pronosticador experto en las últimas técnicas de manipulación de datos, pero que es incapaz o no está dispuesto a relacionar el proceso de pronósticos con las necesidades de la organización y de quienes toman las decisiones dentro de ella. El punto de vista de este libro es que las técnicas cuantitativas de pronósticos, presentadas en la mayor parte de su contenido, sólo son el punto de partida en el pronóstico eficaz de resultados importantes para la organización. Las técnicas son importantes, sin embargo, para generar los resultados finales, además del análisis se deben incluir el juicio personal, el sentido común y la experiencia en los negocios.

Debido a que el mundo en el que operan las organizaciones cambia constantemente, siempre han sido necesarios los pronósticos. Sin embargo, en años recientes se ha incrementado la dependencia de los métodos que implican técnicas sofisticadas de manipulación de datos. De la noche a la mañana han surgido nuevas tecnologías y disciplinas, se ha intensificado la actividad gubernamental en todos sus niveles, la competencia en muchas áreas se ha vuelto más reñida, ha aumentado el comercio internacional en casi todas las industrias, se han creado y han crecido agencias de ayuda social y de servicios, e Internet se ha vuelto una fuente importante de datos e información para la toma de decisiones. Estos factores se combinan para crear un clima organizacional más complejo, rápido y competitivo que en el pasado. Las organizaciones que no reaccionen de inmediato ante las condiciones cambiantes y que no puedan prever el futuro con un cierto grado de precisión, están condenadas a extinguirse.

Las computadoras, junto con las técnicas cuantitativas que éstas hacen posible, han dejado de ser sólo una comodidad y se han transformado en una necesidad de las organizaciones modernas. Las dificultades ya señaladas generan una enorme cantidad de datos y la necesidad abrumadora de extraer información útil. Las herramientas modernas de pronósticos, junto con las capacidades de la computadora, se han vuelto indispensables para las organizaciones que operan en el mundo moderno.

<sup>1</sup> Makridakis (1986), p. 17.

¿Quién necesita pronósticos? Casi todas las organizaciones—grandes y pequeñas, privadas y públicas—utilizan los pronósticos de manera explícita o implícita, puesto que deben planear para satisfacer las condiciones del futuro sobre las que tienen un conocimiento imperfecto. Además, la necesidad de tener pronósticos está en todas las líneas funcionales, así como en todo tipo de organizaciones. Se requieren pronósticos en las áreas de finanzas, marketing, personal y producción; en organizaciones de gobierno y empresas que buscan utilidades, en clubes sociales pequeños y partidos políticos nacionales. Considere las siguientes preguntas que sugieren la necesidad de ciertos procedimientos de pronóstico:

- Si se aumenta nuestro presupuesto publicitario en 10%, ¿cómo se verán afectadas las ventas?
- ¿Qué ingresos puede esperar el gobierno estatal en los dos próximos años?
- ¿Cuántas unidades se deben vender para recuperar la inversión fija en el equipo de producción?
- ¿Qué factores ayudan a explicar la variabilidad en las ventas unitarias mensuales?
- ¿Cuál es la predicción del monto de créditos otorgados por algún banco en cada uno de los próximos diez años?
- ¿Habrá una recesión? Si es así, ¿cuándo comenzará, qué tan severa será y cuándo terminará?

Un pasaje de Bernstein (1996) resume de manera eficaz la función de los pronósticos en las organizaciones.

No se planea enviar bienes a través del océano o ensamblar mercancía para su venta ni se pide dinero prestado, sin primero intentar determinar lo que el futuro podría traer. Asegurarse de que los materiales que se ordenan se entregarán a tiempo, verificar que los artículos que se planea vender se produzcan dentro del programa y tener las instalaciones de ventas en su lugar, son cuestiones que deben plantearse antes del momento en que los clientes lleguen y pongan su dinero sobre el mostrador. El ejecutivo de negocios exitoso es, en primer lugar, un pronosticador; las compras, el marketing, la fijación de precios y la organización vendrán después.<sup>2</sup>

## TIPOS DE PRONÓSTICOS

---

Cuando los administradores se enfrentan a la necesidad de tomar decisiones en una atmósfera de incertidumbre, ¿qué tipos de pronósticos están disponibles para ellos? Primero, los procedimientos de pronóstico podrían clasificarse como de largo o corto plazos. Los pronósticos de largo plazo son necesarios para establecer el curso general de una organización y son el enfoque exclusivo de la alta dirección. Los pronósticos de corto plazo se usan para diseñar estrategias inmediatas, y los mandos medios y las gerencias de primera línea los usan para cubrir las necesidades del futuro inmediato.

También se podrían clasificar los pronósticos en términos de su posición dentro de un continuo micro y macro; es decir, en la medida que involucren pequeños detalles en comparación con una gran escala. Por ejemplo, a un gerente de planta le podría interesar el pronóstico del número de trabajadores necesarios durante los próximos meses (un micropronóstico), mientras que el gobierno federal pronostica la cantidad total de personas empleadas en el país entero (un macropronóstico). Una vez más, los diferentes niveles administrativos dentro de una organización tienden a centrarse en diferentes niveles del continuo micro y macro. Por ejemplo, a la alta dirección le interesaría pronosticar las ventas para toda la compañía, mientras que a los vendedores individuales les sería de más interés pronosticar sus volúmenes individuales de ventas.

---

<sup>2</sup> Bernstein (1996) pp. 21-22.

Los procedimientos de pronósticos también pueden clasificarse según sean más cuantitativos o cualitativos. En un extremo, una técnica totalmente cualitativa no requiere manipulación abierta de datos. Solamente se utiliza el juicio de quien pronostica. Incluso aquí, en realidad, el juicio de esta persona es resultado de la manipulación mental de datos históricos. En el otro extremo, las técnicas puramente cuantitativas no necesitan elementos de juicio; son procedimientos mecánicos que producen resultados cuantitativos. Por supuesto, algunos procedimientos cuantitativos requieren una manipulación mucho más sofisticada de los datos que otros. Este texto enfatiza las técnicas cuantitativas de pronóstico debido a que se requiere un entendimiento más amplio de estos procesos, los cuales son muy útiles en la dirección eficaz de las organizaciones modernas. Sin embargo, se pone de relieve, una vez más, que deben utilizarse el sentido común y la apreciación junto con los procedimientos mecánicos y de manipulación de datos. Sólo de esta manera puede obtenerse un pronóstico inteligente.

## CONSIDERACIONES DE PRONÓSTICOS MACROECONÓMICOS

---

Con frecuencia se piensa en los pronósticos en términos de predecir variables importantes para una compañía específica o, quizás, para algún departamento de una empresa. Las ventas mensuales de la compañía, las ventas unitarias de alguna de sus tiendas y las horas de ausencia de cada empleado por mes dentro de una fábrica son ejemplos de esto.

En contraste, hay un creciente interés en el pronóstico de variables importantes para la economía entera de un país o para la economía mundial. Se ha realizado mucho trabajo en métodos de evaluación para hacer este tipo de pronósticos económicos generales, llamados macroeconómicos. Algunos ejemplos de interés para el gobierno de Estados Unidos son las tasas de desempleo, el producto interno bruto y la tasa de interés preferencial. La política económica se basa, en parte, en las proyecciones de indicadores económicos importantes, como los anteriores. Por esta razón, existe gran interés en mejorar los métodos de pronóstico para los principales indicadores generales del desempeño económico de un país.

Un cambio inesperado e importante en un factor económico fundamental es uno de los principales obstáculos para desarrollar pronósticos precisos para la actividad económica general. Las grandes variaciones en los precios del petróleo, los picos de inflación y las amplias transformaciones políticas de un gobierno son ejemplos de cambios en un factor fundamental que puede afectar a la economía global.

La posibilidad de cambios importantes en la escena económica ha provocado una pregunta clave en los pronósticos macroeconómicos: ¿deberían modificarse los pronósticos generados por su modelo mediante el uso del juicio de quien pronostica? A menudo, el trabajo actual sobre metodología de pronósticos implica esta cuestión.

Se sigue haciendo mucho trabajo, tanto teórico como práctico, en el tema de los pronósticos macroeconómicos. Al considerar la importancia de los pronósticos económicos precisos para la formulación de la política económica en diversos países, se puede esperar una permanente atención en tales pronósticos. Una buena referencia de introducción a los pronósticos macroeconómicos se encuentra en Pindyck y Rubinfeld (1998).

## ELECCIÓN DE UN MÉTODO DE PRONÓSTICOS

---

El análisis anterior sugiere varios factores que deben considerarse al elegir un método de pronóstico. Debe tomarse en cuenta el nivel de detalle. ¿Son necesarios los pronósticos de detalles específicos (un micropronóstico)? ¿O se necesita la predicción del futuro de un factor general o de gran escala (un macropronóstico)? ¿Es imperioso el pronóstico para una fecha próxima (un pronóstico de corto plazo) para una fecha intermedia o para un futuro lejano (un pronóstico de largo plazo)? ¿En qué medida son apropiados los métodos cualitativos (de juicio) y cuantitativos (de manipulación de datos)?

La consideración primordial en la elección de un método de pronóstico es que los resultados deben facilitar el proceso de la toma de decisiones de los administradores de la organización. Rara vez un solo método funciona para todos los casos. Para seleccionar un método de pronóstico deben considerarse el tipo de productos (por ejemplo, nuevos frente a establecidos), los objetivos (como la simple predicción en contraposición con la necesidad de controlar un impulsor de negocios importante para los valores futuros) y limitantes (como costo, experiencia requerida o la prisa). Con la disponibilidad de software moderno para pronósticos, es mejor pensar en los métodos de pronóstico como herramientas genéricas que pueden aplicarse de manera simultánea. Asimismo, pueden probarse varios métodos en cualquier situación determinada. La metodología que produzca los pronósticos más precisos en un caso, quizás no sea la mejor en otra situación. Sin embargo, el(los) método(s) elegido(s) debe(n) producir un pronóstico adecuado, oportuno y entendible para los administradores, de tal forma que pueda ayudar a tomar mejores decisiones.

La elección de la técnica de pronóstico se analiza en el capítulo 3 y se resume en la tabla 3.6. Las metodologías de pronóstico consideradas en este libro se resumen en la tabla 11.1 (capítulo 11). Estas tablas, junto con el análisis adicional disponible en Chase (1997), pueden ayudar a elegir un grupo inicial de procedimientos de pronósticos.

## PASOS DE LOS PRONÓSTICOS

---

Todos los procedimientos formales de pronóstico implican el entendimiento de las experiencias del pasado para proyectarlas hacia el futuro. Por lo tanto, suponen que las condiciones futuras serán las mismas que generaron los datos en el pasado, excepto aquellas variables reconocidas de forma explícita por el modelo de pronóstico.

Un departamento de recursos humanos contrata empleados, en parte, con base en el resultado de un examen de admisión a la compañía, debido a que en el pasado el resultado del examen parecía ser una herramienta importante para predecir la calificación de productividad laboral. A medida que esta relación continúa siendo cierta, los pronósticos de productividad futura en el trabajo —y, por lo tanto, las decisiones de contratación— se pueden mejorar mediante el uso de resultados en los exámenes. Si por alguna razón la asociación entre el resultado del examen y la productividad en el trabajo cambia, entonces el pronóstico de la calificación de productividad a partir de resultados de exámenes, mediante el modelo histórico, dará lugar a pronósticos imprecisos y decisiones de contratación potencialmente malas. Esto hace que el acto de pronosticar sea difícil. El futuro no siempre se parece al pasado. En la medida en que lo haga, los métodos cuantitativos de pronóstico funcionarán bien. Si no lo hace, pueden resultar pronósticos imprecisos. Sin embargo, en general es mejor tener un pronóstico construido de manera razonable que no tener ninguno.

El reconocimiento de que las técnicas de pronósticos operan con los datos generados por sucesos históricos lleva a la identificación de los siguientes cinco pasos en el proceso de pronóstico:

1. Formulación del problema y recolección de datos
2. Manipulación y limpieza de datos
3. Construcción y evaluación del modelo
4. Aplicación del modelo (el pronóstico real)
5. Evaluación del pronóstico

En el paso 1, la formulación del problema y la recolección de datos se tratan como un paso único, debido a que están íntimamente relacionados. El problema determina los datos adecuados. Si se considera una metodología cuantitativa de pronóstico, los datos relevantes deben estar disponibles y ser correctos. A menudo, la tarea de obtener y recopilar los datos adecuados es desafiante y consume mucho tiempo. Si no se dispone de los datos adecuados, el problema podría tener que redefinirse o se tendría que emplear una metodología

no cuantitativa de pronóstico. En general, surgen problemas de recolección y control de calidad siempre que es necesario obtener datos pertinentes para un esfuerzo por hacer pronósticos en los negocios.

El paso 2, manipulación y limpieza de datos, con frecuencia es necesario. En el proceso de pronóstico es posible tener demasiados datos o muy pocos. Algunos datos pueden ser irrelevantes al problema. Otros podrían tener valores omitidos que deban estimarse. Algunos datos podrían tener que expresarse en unidades diferentes de las originales. Otros quizás deban procesarse previamente (por ejemplo, los que se acumulan de diversas fuentes y se suman). Otros datos podrían ser adecuados, aunque solamente durante ciertos períodos históricos (por ejemplo, al pronosticar las ventas de automóviles pequeños, acaso se desee utilizar sólo las ventas de autos después del embargo de petróleo de la década de 1970 en lugar de los datos a lo largo de los últimos 50 años). Por lo general, se requiere cierto esfuerzo para obtener los datos de la forma requerida a fin de utilizar determinados procedimientos de pronóstico.

El paso 3, construcción y evaluación del modelo, implica adecuar los datos recolectados en un modelo de pronósticos que sea adecuado en términos de minimización del error de pronóstico. Cuanto más simple sea el modelo, resulta mejor para obtener la aceptación del proceso de pronóstico por parte de los administradores que toman las decisiones en la empresa. Con frecuencia debe llegarse a un equilibrio entre un método de pronóstico sofisticado, que es un poco más preciso, y uno simple que se entienda con facilidad y obtenga el apoyo de —y lo utilicen activamente— quienes toman las decisiones en la compañía. Obviamente, el juicio intuitivo está implicado en este proceso de selección. Puesto que este libro presenta modelos de pronóstico de aplicabilidad diversa, deseamos que, después de estudiar el material, aumente la destreza del lector para ejercer buen juicio en la elección y uso de los modelos de pronósticos más adecuados.

El paso 4, aplicación del modelo, consiste en los pronósticos reales del modelo que se generan una vez que se han recolectado y quizás reducido a sólo los datos adecuados, tan pronto se ha elegido un modelo adecuado de pronósticos. Frecuentemente, los pronósticos de períodos recientes en los que se conocen los valores históricos reales se utilizan para verificar la precisión del proceso. Los errores de pronóstico que se observan y se resumen de cierta forma aparecen en el paso 5.

El paso 5, evaluación del pronóstico, implica comparar los valores del pronóstico con los valores históricos reales. En este proceso, algunos de los más recientes valores de datos se retienen del grupo de datos que se analiza. Una vez que se ha completado el modelo de pronósticos, se realizan los pronósticos para los diversos períodos y se comparan con los valores históricos conocidos. Algunos procedimientos de pronóstico reportan la suma de los valores absolutos de los errores o la dividen entre el número de intentos de pronóstico para reportar el error absoluto promedio del pronóstico. Otros procedimientos toman en cuenta la suma de los cuadrados del error, lo cual se compara con cifras similares de métodos alternos de pronóstico. Algunos procedimientos también localizan y reportan la magnitud de los términos de los errores a lo largo del período de pronóstico. Frecuentemente, el examen de los patrones de errores lleva al analista a modificar el procedimiento de pronósticos. Cerca del final del capítulo 3 se analizan métodos específicos para medir los errores de pronósticos.

## **ADMINISTRACIÓN DEL PROCESO DE PRONÓSTICOS**

---

El análisis en el presente capítulo sirve para subrayar nuestra creencia en que la habilidad directiva y el sentido común deben estar implicados en el proceso de pronóstico. El pronosticador debería ser percibido como un asesor del administrador, más que como el supervisor de un aparato automático de toma de decisiones. Desafortunadamente, esto último a veces es el caso en la práctica, sobre todo por el aura de la computadora. Una vez más, las técnicas cuantitativas en el proceso de pronósticos deben percibirse como lo que realmente son;

es decir, herramientas que utiliza el administrador para lograr mejores decisiones. De acuerdo con Makridakis (1986):

La utilidad y funcionalidad de los pronósticos pueden aumentarse si la administración adopta una actitud más realista. Los pronósticos no deben verse como un sustituto de las profecías, sino más bien como la mejor forma de identificar y extrapolar los patrones establecidos o las relaciones, a fin de pronosticar. Si se acepta tal actitud, los errores en los pronósticos deben considerarse como inevitables y habrán de investigarse las circunstancias que los causan.<sup>3</sup>

Dado lo anterior, deben surgir varias preguntas fundamentales si se quiere una administración adecuada del proceso.

- ¿Por qué es necesario un pronóstico?
- ¿Quién utilizará el pronóstico y cuáles son sus requisitos específicos?
- ¿Qué nivel de detalle o de agregación se requiere y cuál es el horizonte de tiempo adecuado?
- ¿Qué datos están disponibles y cuáles serán suficientes para generar el pronóstico necesario?
- ¿Cuánto costará el pronóstico?
- ¿Qué tan preciso se espera que sea el pronóstico?
- ¿Se realizará el pronóstico a tiempo para ayudar al proceso de toma de decisiones?
- ¿El pronosticador entiende claramente cómo se utilizará el pronóstico en la organización?
- ¿Está disponible un proceso de retroalimentación para evaluar el pronóstico después de que se haya hecho y para ajustar el proceso de pronósticos de manera conveniente?

## SOFTWARE PARA PRONÓSTICOS

---

El desarrollo de mayor trascendencia en los pronósticos durante los últimos años es el software diseñado para manejar directamente varios métodos de pronóstico. Hay dos tipos de paquetes de cómputo que interesan a los pronosticadores: (1) paquetes estadísticos que incluyen análisis de regresión, exploración de series de tiempo y otras técnicas usadas con frecuencia por los pronosticadores, y (2) paquetes de pronósticos diseñados específicamente para aplicaciones de pronóstico.

Los ejemplos de software independiente con herramientas de pronóstico incluyen Minitab, SAS, EViews y SPSS. Además, hoy existen muchos complementos o programas supplementarios que proporcionan herramientas de pronóstico en un entorno de hoja de cálculo. Por ejemplo, el complemento de Analysis ToolPak para Microsoft Excel proporciona algunas capacidades de análisis de regresión y suavización. Algunos complementos más detallados, como Crystal Ball (CB) Predictor y Forecast X, proporcionan una gama (casi) completa de capacidades de pronóstico.

Los administradores que tienen computadoras personales en sus escritorios y conocimientos sobre las técnicas, ya no dependen del personal para hacer sus pronósticos y pueden aprovechar la existencia y facilidad de métodos sofisticados de pronóstico. El software para este fin se actualiza de manera consistente con mejoras en la facilidad de usarlo. Ahora está de moda, en particular en un ambiente de hoja de cálculo, tener disponible el pronóstico “automático”. Es decir, el software selecciona el mejor modelo o procedimiento para efectuar el pronóstico y lo genera de inmediato. Advertimos, sin embargo, que esta funcionalidad tiene un precio. Los procesos automáticos producen números, pero rara vez dan una perspectiva real de la naturaleza y calidad de los pronósticos a quien los realiza. Se piensa que la generación de pronósticos significativos requiere de la intervención humana,

---

<sup>3</sup> Makridakis (1986), p. 33.

un dar y recibir entre el conocimiento del problema y los procedimientos de pronóstico (software).

Muchas de las técnicas en este libro se ilustrarán con Minitab 14 y Excel 2000 con el complemento de CB Predictor. Usted encontrará descripciones y pantallas de estos programas a lo largo del texto. Ocasionalmente se mostrarán resultados seleccionados de EViews.<sup>4</sup> Otros paquetes de pronóstico tendrán la misma funcionalidad básica, aunque las ventanas de entrada y resultados serán un tanto diferentes. Los programas que se usan en este libro se eligieron, principalmente, por su facilidad de uso y su amplia disponibilidad.

Un resumen reciente de software de pronósticos se encuentra en la encuesta de software de análisis estadístico que aparece en la publicación INFORMS de *ORMS Today*, volumen 30, número 1, de febrero de 2003.

## INFORMACIÓN EN LÍNEA

---

En Internet hay información de interés para los pronosticadores. Quizás la mejor forma de aprender acerca de lo que está disponible en el ciberespacio es pasar cierto tiempo buscando lo que le interese mediante un navegador, como Netscape o Internet Explorer de Microsoft.

Es probable que cualquier lista de páginas web para pronosticadores esté anticuada en el momento que esta edición aparezca en el mercado; sin embargo, hay dos sitios en la web que tienen probabilidades de estar disponibles por cierto tiempo. B&E DataLinks, que se encuentra en [www.econ-datalinks.org/](http://www.econ-datalinks.org/) es una página web que mantiene la Business and Economic Section de la American Statistical Association. Esta página web contiene extensos vínculos con fuentes de datos económicos y financieros de interés para los pronosticadores. El segundo sitio, Resources for Economists, patrocinado por la American Statistical Association en [rfe.wustl.edu/EconFAQ.html](http://rfe.wustl.edu/EconFAQ.html) contiene un extenso grupo de vínculos con fuentes de datos, publicaciones, organizaciones profesionales y demás.

## EJEMPLO DE UN PRONÓSTICO

---

Los análisis en este capítulo enfatizan que los pronósticos requieren de una gran capacidad intuitiva además de la manipulación matemática de los datos recopilados. El siguiente ejemplo muestra el tipo de pensamiento que frecuentemente antecede al esfuerzo de pronosticar en una empresa real. Observe que los valores de datos que generarán pronósticos útiles, cuando existen, no son evidentes al inicio del proceso y podrían haber pasado inadvertidos. En otras palabras, los denuedos iniciales podrían resultar inútiles para la administración. No se muestran aquí los resultados de cómputo del intento de pronosticar cuando se utilizan las variables identificadas, ya que implican temas que se describen a lo largo del texto. Sin embargo, busque las técnicas descritas en los capítulos subsecuentes para aplicarlas a estos datos. El [ejemplo 1.1](#) se resolverá en el [capítulo 11](#). Por el momento, se espera que este ejemplo ilustre el alcance del esfuerzo de pronósticos al cual se enfrentan los administradores en la práctica.

### Ejemplo 1.1

Alomega Food Stores es un proveedor de alimentos al menudeo con 27 tiendas en un estado del oeste medio. La compañía participa en varios tipos de publicidad y, hasta hace poco, nunca había estudiado el efecto de su presupuesto publicitario en las ventas, aunque se han recopilado y guardado algunos datos durante tres años.

Los ejecutivos en Alomega decidieron comenzar a rastrear sus esfuerzos publicitarios junto con los volúmenes de ventas para cada mes. Su esperanza era que después de varios meses,

---

<sup>4</sup> EViews, un software popular para pronósticos, se puede encontrar en Quantitative Micro Systems LLC, Irvine, CA.

los datos recolectados pudieran examinarse para revelar posibles relaciones que ayudarían a determinar los gastos futuros en publicidad.

El departamento de contabilidad comenzó a ampliar sus registros históricos al anotar los volúmenes de venta de cada mes junto con el presupuesto para anuncios comerciales, tanto en periódicos como en televisión. También registraron las ventas y los valores publicitarios de uno y dos meses antes. Esto se hizo debido a que algunas personas en el comité ejecutivo pensaban que las ventas podrían depender de los gastos publicitarios de los meses anteriores, y no de aquellos del mes en que aparecían los anuncios.

También se creía que las ventas experimentaban un efecto estacional. Por esta razón, se utilizó una variable ficticia o categórica para indicar cada mes. La administración también se preguntaba si habría alguna tendencia en el volumen de ventas.

Por último, se pensaba que los fondos publicitarios de Alomega podrían tener un efecto en los presupuestos publicitarios de sus principales competidores en el mes próximo. Se decidió que para cada mes siguiente se podría clasificar la publicidad de la competencia en cantidades: (1) pequeña, (2) moderada o (3) grande.

Después de algunos meses de recolección de datos y análisis de registros pasados, el departamento de contabilidad terminó un arreglo de datos de 48 meses mediante el uso de las siguientes variables:

- Presupuesto publicitario
- Presupuesto publicitario en periódicos
- Presupuesto publicitario en televisión
- Código del mes, donde: enero = 1, febrero = 2, hasta diciembre = 12
- Una serie de 11 variables ficticias para indicar el mes
- Publicidad en periódico de un mes antes
- Publicidad en periódico de dos meses antes
- Publicidad en televisión de un mes antes
- Publicidad en televisión de dos meses antes
- Número del mes desde 1 hasta 48
- Código 1, 2 o 3 para indicar los esfuerzos publicitarios de la competencia al mes siguiente

La dirección de Alomega, en especial Julie Ruth, presidenta de la compañía, ahora quiere aprender lo que pueda a partir de los datos que ha recopilado. Además de aprender sobre los efectos de la publicidad en los volúmenes de ventas y la publicidad de la competencia, Julie se pregunta si hay alguna tendencia y efectos estacionales en las ventas. Sin embargo, el gerente de producción de la empresa, Jackson Tilson no comparte su entusiasmo. Al final de la junta de planeación de pronósticos, hizo la siguiente declaración: “He tratado de mantener la boca cerrada durante esta junta, pero realmente es demasiado. Creo que propiciamos que mucha gente pierda el tiempo con toda esta recolección de datos con las computadoras. Todo lo que hay que hacer es hablar con nuestro personal en piso y con los gerentes de las tiendas de abarrotes para comprender lo que está sucediendo. Ya vi esto anteriormente, y aquí vamos de nuevo. Algunos de ustedes necesitan apagar sus computadoras, salir de sus oficinas elegantes y hablar con unas cuantas personas de verdad”.

## Resumen

El propósito de un pronóstico es reducir el rango de incertidumbre dentro del cual deben hacerse las estimaciones relacionadas con la administración. Para esto se sugieren dos reglas principales a las que debe adherirse el proceso de pronóstico:

1. El pronóstico debe ser técnicamente correcto y producir pronósticos precisos que sean suficientes para satisfacer las necesidades de la empresa.
2. El procedimiento de pronóstico y sus resultados deben presentarse de forma eficaz a la administración, a fin de que los pronósticos se utilicen durante el proceso de toma de decisiones en beneficio de la empresa; los resultados también deben justificarse sobre una base de costo-beneficio.

Es común que los pronosticadores pongan atención especial en la primera regla y se esmeren menos en la segunda. Pero si los pronósticos bien preparados y eficientes por su costo van a beneficiar a la empresa, deberán utilizarse por quienes tengan la autoridad para tomar decisiones. De aquí surge la cuestión de lo que podría llamarse la “política” de los

pronósticos. Los gastos y asignaciones de recursos considerables, y a veces importantes, dentro de la empresa dependen con frecuencia del punto de vista de la administración sobre el curso de acontecimientos futuros. Puesto que el movimiento de recursos y el poder dentro de una organización a menudo se basa en la percepción que la dirección de la empresa tenga sobre el futuro (pronósticos), no es sorprendente encontrar una cierta cantidad de intriga política alrededor del proceso de pronóstico. La necesidad de ser capaz de vender con eficacia los pronósticos a la dirección es, al menos, tan importante como el desarrollo de las predicciones.

El resto de este libro analiza varios modelos y procedimientos de pronóstico. Primero se revisarán los conceptos estadísticos fundamentales y se proporcionará una introducción al análisis de correlación y regresión. Después se dedica un capítulo a los métodos de recolección de datos y la exploración de grupos de datos en busca de patrones subyacentes. Muchos métodos específicos de pronóstico se detallan en las secciones que siguen, y los últimos dos capítulos del texto están dedicados a aspectos de pronósticos por juicios personales y a la administración del proceso de pronóstico.

## CASOS

---

### CASO 1-1 MR. TUX

John Mosby es propietario de varias tiendas de alquiler de trajes de etiqueta Mr. Tux, la mayoría de ellas en el área de Spokane, en Washington.<sup>5</sup> Su tienda de Spokane también fabrica camisas para esmoquin, las cuales distribuye en tiendas de alquiler de todo el país. Puesto que la actividad del alquiler varía de una temporada a otra debido a las graduaciones, reuniones y otras actividades, John sabe que su negocio depende de esos períodos. Le gustaría medir este efecto estacional como un apoyo para administrar su negocio y utilizarlo en las negociaciones de pago de créditos con su banco.

Es de interés aún mayor para él, encontrar una manera de pronosticar sus ventas mensuales. Su negocio continúa creciendo, lo cual requiere mayor capital

y deudas de largo plazo. Tiene fuentes para financiar ambos tipos de crecimiento, pero los inversionistas y banqueros se interesan en una forma concreta de pronosticar las ventas en el futuro. Aunque confían en John, su expresión de que el futuro de su negocio “se ve maravilloso” los deja inquietos.

Como primer paso en la construcción de un modelo de pronósticos, John instruye a uno de sus empleados, McKennah Lane, que recopile datos de ventas mensuales de los últimos años. En los capítulos siguientes se utilizan varias técnicas para pronosticar tales datos de ventas para Mr. Tux. En el capítulo 11 se resumen estos esfuerzos y John Mosby intentará elegir la técnica de pronósticos que cubra mejor sus necesidades. ■

### CASO 1-2 CONSUMER CREDIT COUNSELING

Consumer Credit Counseling (CCC), una corporación privada no lucrativa, se fundó en 1982.<sup>6</sup> El propósito de CCC fue brindar ayuda a los consumidores para planear y apegarse a presupuestos; también, ayudarles a lograr acuerdos con sus acreedores a fin de pagar obligaciones con mora, así como instruirlos en la administración del dinero.

Se daba asesoría financiera privada sin costo a las familias e individuos que estuvieran en dificultades financieras y que quisieran mejorar su habilidad para administrar el dinero. Se impartieron programas educativos de administración monetaria en escuelas, grupos comunitarios y negocios. Asimismo, se ofreció un programa de administración de deudas como una alternativa a la bancarrota.

<sup>5</sup> Estamos en deuda con John Mosby, propietario de las tiendas de alquiler Mr. Tux por su ayuda para preparar este caso.

<sup>6</sup> Estamos en deuda con Marv Harnishfeger, directora ejecutiva de Consumer Credit Counseling, de Spokane, así como con Dorothy Mercer, presidenta del consejo de administración, por su ayuda para preparar el caso. Dorothy es una antigua estudiante de MBA de JH, quien en forma consistente lo ha mantenido en contacto con el uso de los métodos cuantitativos en el mundo práctico de los negocios.

A través de este programa, CCC negoció con los acreedores en nombre del cliente para lograr arreglos de pago especiales. El cliente hacía un pago único a CCC que enseguida se entregaba a los acreedores.

CCC tenía una mezcla de personal voluntario y a sueldo; de hecho, la cantidad de voluntarios superaba tres a uno el número de empleados a sueldo. Siete de estos últimos se encargaban de la administración, el trabajo de oficina y cubrían alrededor de la mitad de las necesidades de asesoría de CCC. La otra mitad de los requerimientos de asesoramiento del servicio las cubrían 21 consejeros voluntarios. CCC dependía principalmente de fondos corporativos para mantener las operaciones y los servicios. El Fair Share Funding Program permitía a los acreedores que recibieran pagos de programas de administración de deuda de clientes, donar al servi-

cio una porción de los fondos que les eran devueltos a través de dichos programas.

Una porción importante del apoyo corporativo provenía de una empresa de servicio público local que proporcionaba fondos para mantener una posición de consejero de tiempo completo, así como espacio para asesoría en todas las oficinas.

Además, las cuotas de los clientes constituyan una fuente de fondos. Los clientes que participaban en la administración de deudas pagaban una cuota mensual de 15 dólares para ayudar a cubrir el costo administrativo del programa. (En el caso de los clientes que no podían pagar, estas cuotas se reducían o se cancelaban).

Estos antecedentes se utilizarán en los capítulos siguientes, a medida que CCC se enfrenta a problemas difíciles relacionados con los pronósticos de variables importantes. ■

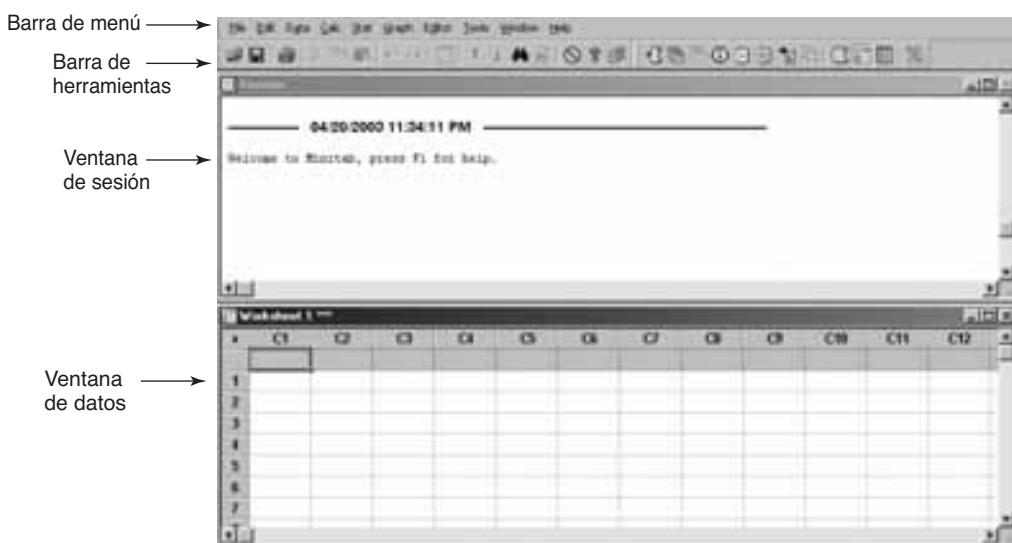
## Aplicaciones en Minitab

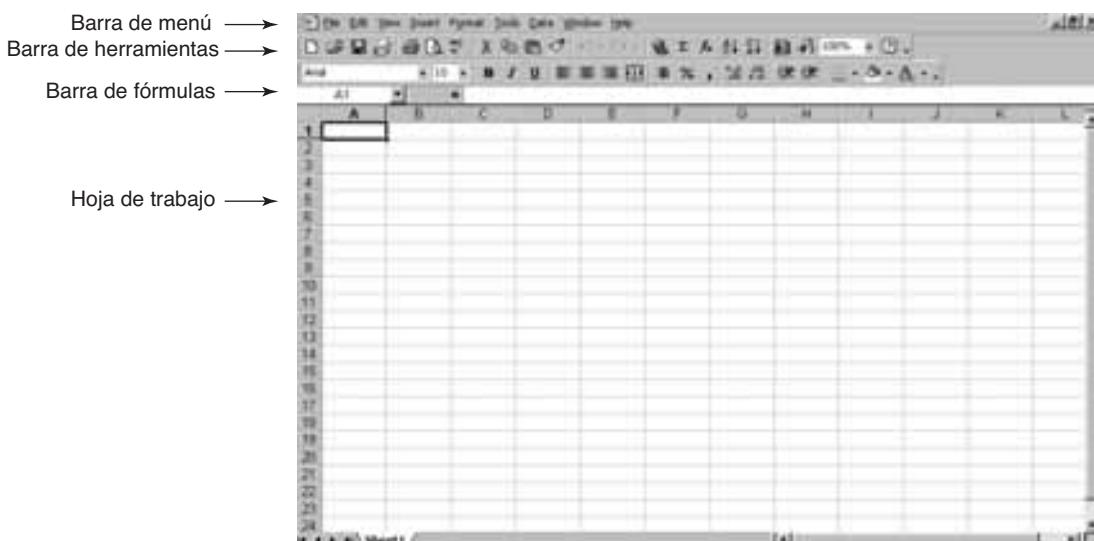
Minitab es un programa estadístico sofisticado que mejora con cada nueva versión. Aquí se describe la número 14.

La figura 1.1 muestra cuatro aspectos importantes de Minitab. La barra de menú es donde se eligen los comandos. Por ejemplo, al hacer clic en Stat, aparece un menú desplegable que contiene todas las técnicas estadísticas disponibles. La barra de herramientas muestra botones para funciones que se utilizan comúnmente. Observe que estos botones cambian según la ventana de Minitab que esté abierta. Hay dos ventanas separadas en la pantalla de Minitab: la ventana de datos donde se integran, editan y ven las columnas de datos en cada hoja de cálculo, y la ventana de sesión que muestra los resultados de texto, como tablas de estadísticas.

En los siguientes capítulos se dan instrucciones específicas para que pueda guardar datos en la hoja de cálculo de Minitab y activar los procesos de pronósticos, a fin de realizar los pronósticos necesarios.

**FIGURA 1.1 Pantalla básica de Minitab**





**FIGURA 1.2 Pantalla básica de Excel**

## Aplicaciones de Excel

Excel es un programa de hojas de cálculo que se usa con frecuencia para hacer pronósticos. La figura 1.2 muestra la pantalla de apertura de la versión 2000. Los datos se colocan en las filas y columnas de la hoja de cálculo (observe que en la figura 1.2 se enfatiza la fila 1, columna A) y a continuación se dan comandos para realizar varias operaciones con los datos incorporados.

Por ejemplo, los salarios anuales para un número de empleados podría agregarse a la columna 1, y podría calcularse el promedio de estos valores mediante Excel. Otro ejemplo podría ser colocar las edades de los empleados en la columna 2 y examinar la relación entre la edad y el salario. Los capítulos siguientes muestran cómo utilizar Excel para resolver éste y otros problemas de pronóstico.

Hay varias funciones estadísticas disponibles en Excel que podrían faltar en los menús desplegables de su pantalla. Para activar estas funciones haga clic en:

Tools > Add-Ins

Aparecerá el cuadro de diálogo de *Add-Ins* (complementos). Seleccione *Analysis ToolPak* (herramientas de análisis) y haga clic en OK. Las funciones disponibles en ToolPak se requerirán en capítulos posteriores.

Se aconseja mucho el uso de un complemento de Excel para ayudar con la multitud de cálculos estadísticos requeridos por las técnicas de pronósticos presentadas en este libro.

También se recomienda el complemento de Excel Crystal Ball (CB) Predictor para ampliar las capacidades de pronóstico de Excel. El uso de CB Predictor se ilustra en los capítulos 4 y 9.

## Referencias

Bernstein, P. L., *Against the Gods: The Remarkable Story of Risk*. Nueva York: John Wiley & Sons, 1996.

Carlberg, C., "Use Excel's Forecasting to Get Terrific Projections", *Denver Business Journal* 47 (18) (1996): 2B.

- Chase Jr., C. W., "Selecting the Appropriate Forecasting Method", *Journal of Business Forecasting* (otoño de 1997): 2.
- Diebold, F. X., *Elements of Forecasting*, tercera edición, Cincinnati, Ohio: South-Western, 2004.
- Georgoff, D. M. y R. G. Mardick, "Manager's Guide to Forecasting", *Harvard Business Review* 1 (1986): 110-120.
- Hogarth, R. M. y S. Makridakis, "Forecasting and Planning: An Evaluation", *Management Science* 27 (2) (1981): 115-138.
- Makridakis, S., "The Art and Science of Forecasting", *International Journal of Forecasting* 2 (1986): 15-39.
- Newbold, P. y T. Bos, *Introductory Business and Economic Forecasting*, segunda edición, Cincinnati, Ohio: South-Western, 1994.
- Ord, K. y S. Lowe, "Automatic Forecasting", *American Statistician*, 50 (1996): 88-94.
- Perry, S., "Applied Business Forecasting", *Management Accounting* 72 (3) (1994): 40.
- Pindyck, R. S. y D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, cuarta edición, Nueva York: McGraw-Hill, 1998.
- Wright, G. y P. Ayton, eds., *Judgmental Forecasting*, Nueva York: John Wiley & Sons, 1987.



## CAPÍTULO

# 2

## REPASO DE CONCEPTOS ESTADÍSTICOS BÁSICOS

La mayoría de las técnicas de pronóstico se basan en los conceptos fundamentales que son el tema de los libros de texto de estadísticas para los negocios y los cursos de introducción a la estadística. Este capítulo revisa algunos de los conceptos que son la base de gran parte del material en el resto del libro.

Casi todos los procedimientos estadísticos hacen inferencias acerca de los asuntos de interés de la llamada *población*, después de seleccionar y medir un subgrupo de estos elementos que se conoce como *muestra*. La selección cuidadosa y uso de una muestra representativa que sea lo suficientemente grande, son componentes importantes del proceso estadístico de inferencia con el fin de tener un grado de riesgo aceptablemente bajo.

Generalmente, en los pronósticos se trabaja con datos históricos en un intento por predecir o pronosticar el futuro incierto. Por esta razón, nos concentraremos en examinar datos de muestras, manipular estos datos de alguna manera y utilizar los resultados para elaborar pronósticos.

### ESTADÍSTICA DESCRIPTIVA

El propósito de los procedimientos de la estadística descriptiva es explicar de forma breve una amplia colección de medidas, con unos cuantos valores fundamentales. El resumen más común se obtiene al promediar los valores. En estadística, el proceso de promediar usualmente se logra al calcular la *media*, lo cual implica sumar todos los valores y dividirlos entre el número de valores.

La media de la muestra (con barra) se calcula utilizando la ecuación 2.1.

$$\bar{X} = \frac{\Sigma X}{n} \quad (2.1)$$

donde

$\bar{X}$  = media de la muestra

$\Sigma \bar{X}$  = suma de todos los valores de la muestra

$n$  = tamaño de la muestra

Para simplificar los cálculos en este texto, se utilizan algunas notaciones abreviadas. En la notación simplificada para sumar todos los valores de  $X$  (vea la ecuación 2.1), se entiende que las sumatorias se extienden desde uno hasta  $n$ . Una notación más formal para este procedimiento es

$$\sum_{i=1}^n X_i$$

donde el subíndice  $i$  variará su valor inicial de 1 hasta  $n$  en incrementos de uno. Dado que casi todas las sumas van desde 1 hasta  $n$ , los índices del principio y el final ( $n$ ) se omitirán y se utilizará la notación más simple, excepto donde sea necesaria la notación más completa por cuestiones de claridad.

Además de medir la tendencia central de un grupo de valores mediante el cálculo de la media, la medida en la cual los valores se dispersan alrededor de la media es, a menudo, de interés. La *desviación estándar* se puede concebir como una manera de medir la dispersión de los datos en torno a la media. La ecuación 2.2 es la fórmula para la desviación estándar:

$$S = \sqrt{\frac{\sum(X - \bar{X})^2}{n-1}} = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} \quad (2.2)$$

donde el numerador representa la suma de las diferencias al cuadrado entre los valores medidos y su media.

Muchos procedimientos estadísticos utilizan la *varianza* muestral, la cual se define como el cuadrado de la desviación estándar de un conjunto de medidas. Así que la varianza de la muestra ( $S^2$ ) se calcula como

$$S^2 = \frac{\sum(X - \bar{X})^2}{n-1} = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1} \quad (2.3)$$

### Ejemplo 2.1

Considere la siguiente colección de edades de personas:

$$23, 38, 42, 25, 60, 55, 50, 42, 32, 35$$

Para esta muestra,  $n = 10$  y

$$\sum_{i=1}^n X_i = X_1 + X_2 + \cdots + X_{10} = 23 + 38 + \cdots + 35 = 402$$

$$\bar{X} = \frac{\sum X}{n} = \frac{402}{10} = 40.2$$

$$S^2 = \frac{\sum(X - \bar{X})^2}{n-1} = \frac{(23 - 40.2)^2 + \cdots + (35 - 40.2)^2}{10-1} = \frac{1339.6}{9} = 148.84$$

$$S = \sqrt{S^2} = \sqrt{148.84} = 12.2$$

Los cálculos se muestran en la tabla 2.1. La media de la muestra es 40.2 años, la varianza de la muestra es 148.84 y la desviación estándar de la muestra es de 12.2 años.

El término *grados de libertad* se utiliza para indicar el número de datos que son independientes unos de otros, en el sentido de que no pueden calcularse a partir de los demás y, por lo tanto, llevan piezas únicas de información. Por ejemplo, suponga que se hacen las siguientes tres declaraciones:

*Estoy pensando en el número 5. Estoy pensando en el número 7. La suma de los dos números en mi mente es 12.*

A primera vista, se presentan tres piezas de información. Sin embargo, si se conocen dos de estas afirmaciones, se puede determinar la otra. Podría decirse que sólo hay dos piezas únicas de información en las tres declaraciones o, para usar el término estadístico, nada

TABLA 2.1 Cálculo de $S^2 (\bar{X} = 40.2)$		
$X$	$X - \bar{X}$	$(X - \bar{X})^2$
23	-17.2	295.84
38	-2.2	4.84
42	1.8	3.24
25	-15.2	231.04
60	19.8	392.04
55	14.8	219.04
50	9.8	96.04
42	1.8	3.24
32	-8.2	67.24
35	-5.2	27.04
		$\Sigma(X - \bar{X})^2 = 1,339.60$
		$S^2 = \frac{1,339.60}{10 - 1} = 148.84$

más hay dos grados de libertad puesto que sólo dos de los valores tienen libertad para variar, el tercero no.

Los *grados de libertad* se refieren al número de datos que son independientes unos de otros y que son piezas únicas de información.

En el ejemplo que se presenta en la tabla 2.1, las edades de 10 personas constituyen una muestra con 10 grados de libertad. La edad de cualquiera de ellas pudo haberse incluido en la muestra y, por lo tanto, cada una de las edades tiene libertad para variar. Cuando la media se calcula, las diez edades se utilizan para explicar una edad media total de 40.2 años.

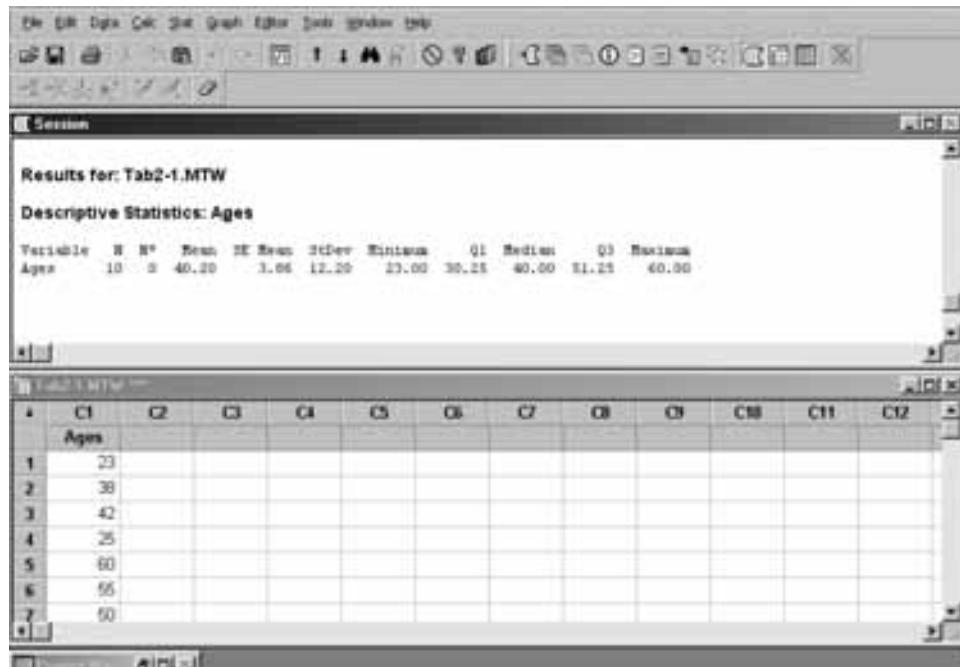
El cálculo de la desviación estándar de la muestra es diferente. Cuando se calcula la desviación estándar de la muestra, se utiliza un valor estimado de la media de la población (la media de la muestra  $\bar{X}$ ). Al utilizar la media de la muestra como valor estimado de la media de la población en los cálculos, por lo general se obtendrá una desviación estándar más pequeña que la desviación estándar de la población. Sin embargo, este problema puede de corregirse si se divide el valor  $\Sigma(X - \bar{X})^2$  entre los grados de libertad adecuados. Una vez que se ha calculado la media de la muestra en este ejemplo, sólo se requieren nueve de las desviaciones  $X - \bar{X}$  para calcular la desviación estándar de la muestra. Dadas nueve de las desviaciones, la última desviación, la número 10, es fija debido a que  $\Sigma(X - \bar{X})$  debe ser igual a cero. En consecuencia, se dice que la desviación estándar de la muestra (o varianza de la muestra) tiene nueve grados de libertad. En general, siempre que se utilice un estadístico de muestra como estimado de un parámetro de la población en un cálculo, se pierde un grado de libertad.

Las estadísticas descriptivas pueden definirse para las poblaciones. A fin de distinguir las estadísticas de población de las estadísticas de la muestra, se usa una notación diferente. La tabla 2.2 muestra los símbolos utilizados tanto para estadísticas de población como de muestra.

La media y la desviación estándar son las formas más comunes de describir, de forma breve y significativa, los datos de muestras. Sin embargo, en ocasiones también se utilizan otras estadísticas descriptivas. En ocasiones, la *mediana* se utiliza para indicar un valor central en una colección de datos. La mediana es aquel valor en el que la mitad de los valores en la colección es mayor que ella y la otra mitad es menor.

**TABLA 2.2 Notación para estadísticas de población y de muestra**

Estadística	Símbolo de la población	Símbolo de la muestra
Media	$\mu$	$\bar{X}$
Varianza	$\sigma^2$	$S^2$
Desviación estandar	$\sigma$	$S$



**FIGURA 2.1 Estadísticas descriptivas en un reporte de Minitab**

A veces, el *rango* de los datos se presenta como una medida aproximada de dispersión. El rango simplemente indica la diferencia entre los valores mínimo y máximo. El rango de edades en la tabla 2.1, por ejemplo, era 37 (60 – 23).

Los *cuartiles* dividen la colección de datos en cuatro partes iguales, después que se han arreglado los valores numéricos desde el menor hasta el mayor. La mediana divide el arreglo en dos partes iguales y, en ocasiones, es denominado *segundo cuartil* y se denota mediante  $Q_2$ . El *primer cuartil* ( $Q_1$ ) divide la mitad inferior en dos partes iguales y el *tercer cuartil* ( $Q_3$ ) divide la mitad superior en dos partes iguales. La colección de edades en la tabla 2.1, por ejemplo, tiene un primer cuartil de  $Q_1 = 30.25$ , una mediana (o segundo cuartil) de  $Q_2 = 40$ , y un tercer cuartil de  $Q_3 = 51.25$ .

Por último, el *rango intercuartil* proporciona una indicación de la variabilidad de un conjunto de datos. Es tan sólo la diferencia entre el tercero y el primer cuartil ( $Q_3 - Q_1$ ) o del rango para el 50% intermedio de los valores de los datos. Para los datos de edades, el rango intercuartil es de 21 (51.25 – 30.25).

Para calcular la mayoría de las estadísticas descriptivas que se han presentado hasta ahora, se pueden utilizar Minitab y Excel. La figura 2.1 muestra los resultados de Minitab para los datos de edades presentados en la tabla 2.1. Las instrucciones para calcular estadísticas descriptivas mediante Minitab y Excel se presentan en los módulos de aplicaciones Minitab y Excel que están al final de este capítulo.

## PRESENTACIÓN DE LA INFORMACIÓN NUMÉRICA

---

En los pronósticos nos preocupan dos tipos de datos: los datos de *corte transversal* en que todas las observaciones se hacen en el mismo marco temporal, y los datos de *series de tiempo* que consisten en una secuencia de observaciones a lo largo del tiempo. La tabla 2.3 muestra un ejemplo de datos de corte transversal: el ingreso neto como un porcentaje del capital de 209 compañías de la encuesta Fortune 500. Otros ejemplos son los ingresos anuales de los ejecutivos de corporaciones y los precios de venta de casas en una ciudad específica. Algunos ejemplos de datos de series de tiempo son el número de visitantes por mes del Coulee Dam Visitor Center, las ventas mensuales de Sears Corporation, el precio diario de las acciones de IBM y la producción anual de trigo de Estados Unidos.

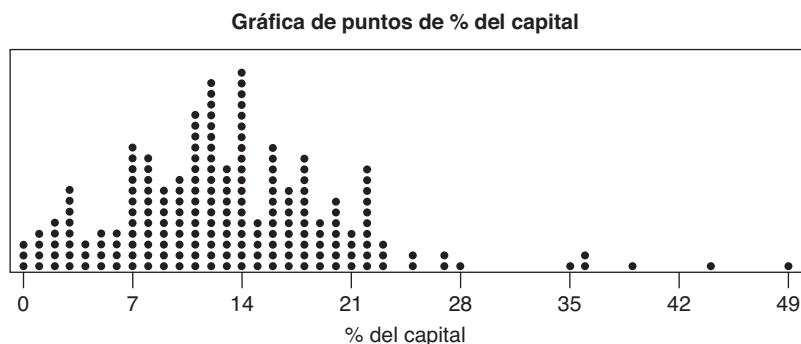
Una de las cosas más importantes por hacer cuando se comienza a explorar una variable es visualizar los datos mediante tablas y gráficas. Las características básicas de los datos, que incluyen observaciones inusuales y patrones únicos, son más fáciles de detectar de manera visual. A veces, las gráficas pueden sugerir explicaciones posibles para algunas de las variaciones en los datos.

Una de las formas más sencillas de visualizar los datos de manera descriptiva es una *gráfica de puntos*. En la figura 2.2, el eje horizontal muestra el rango de datos de valores de ingresos netos como un porcentaje del capital de la tabla 2.3. Cada observación se representa mediante un punto colocado por arriba del eje. Las gráficas de puntos muestran los detalles de los datos y también permiten al pronosticador que compare dos o más grupos de datos.

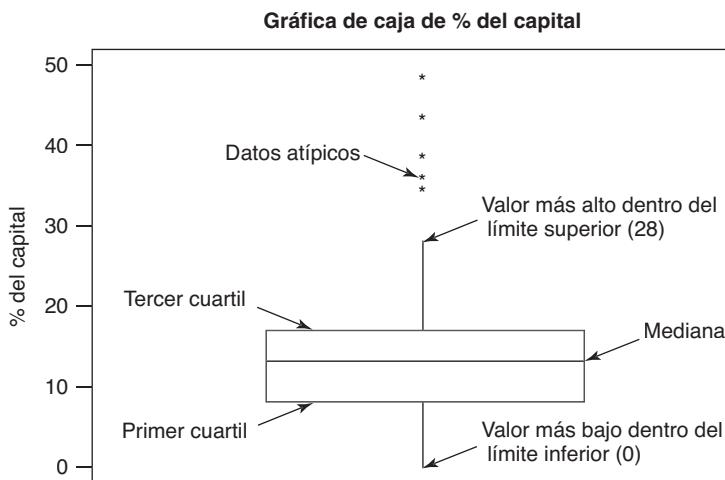
Una *gráfica de caja*, también llamada una gráfica de caja y filamento, es útil para mostrar las características de distribución de los datos. La figura 2.3 muestra una gráfica de caja para el ingreso neto como un porcentaje de los datos de capital. Se dibuja una línea a través de la caja donde está la mediana. Esta línea divide los datos en dos secciones iguales. El borde inferior de la caja es el primer cuartil ( $Q_1$ ) y el borde superior es el tercer cuartil ( $Q_3$ ). Se pueden construir límites adicionales al utilizar el rango intercuartil ( $Q_3 - Q_1$ ). El límite inferior se ubica en  $Q_1 - 1.5(Q_3 - Q_1)$  y el límite superior en  $Q_3 + 1.5(Q_3 - Q_1)$ . Los *datos atípicos* son puntos fuera de los límites superior e inferior y se marcan con aste-

**TABLA 2.3** Ingreso neto como porcentaje del capital para una muestra de 209 compañías de la encuesta de Fortune 500

17	23	22	18	8	7	12	2	49	14
14	36	16	7	3	8	10	11	20	17
15	25	18	12	20	7	5	11	0	22
14	10	14	19	8	12	13	21	3	22
11	18	2	18	14	11	36	16	7	14
12	14	10	8	20	13	8	23	6	21
9	23	7	14	25	12	12	8	11	5
18	13	14	9	16	2	19	21	18	9
14	2	20	17	11	16	13	12	22	16
7	6	14	10	1	21	35	20	18	28
17	15	9	12	5	10	14	1	17	14
14	14	6	22	16	13	14	8	12	6
15	10	22	19	16	4	20	18	2	3
20	7	15	39	4	3	10	7	15	16
12	13	12	11	18	10	13	7	13	12
14	8	11	17	11	22	16	11	12	11
9	11	13	0	12	3	9	9	13	27
1	16	18	12	11	0	10	9	12	22
18	44	4	3	17	12	8	16	7	16
27	11	19	12	22	3	14	14	7	8
11	1	3	17	8	7	5	19	22	



**FIGURA 2.2 Gráfica de puntos del ingreso neto como porcentaje del capital**

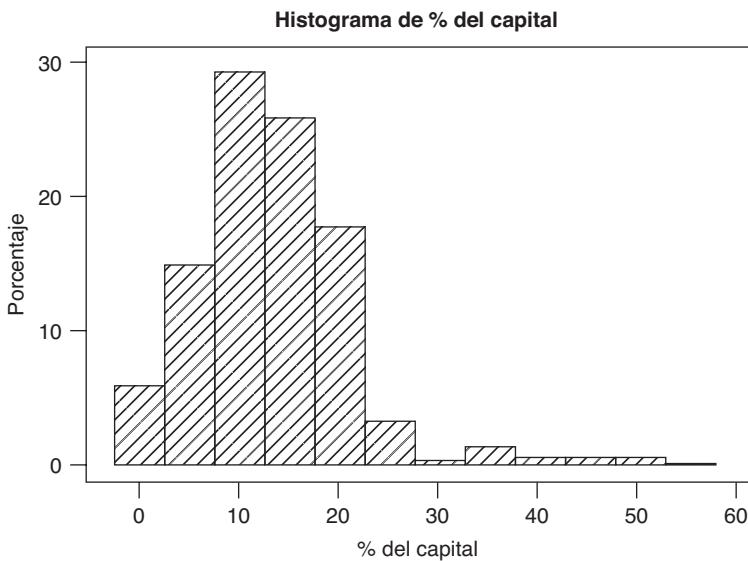


**FIGURA 2.3 Gráfica de caja del ingreso neto como porcentaje del capital**

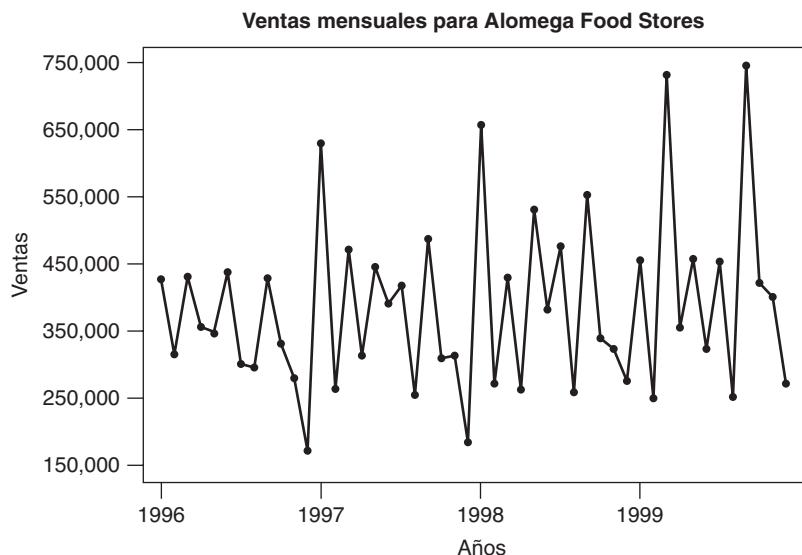
riscos. En la figura 2.3 el primer cuartil es 8, la mediana es 13, el tercer cuartil es 17 y el rango intercuartiles 9 ( $17 - 8$ ). El límite inferior es  $-5.5$  ( $8 - 1.5 \times 9$ ) y el límite superior es  $30.5$  ( $17 + 1.5 \times 9$ ). Observe que el valor más bajo dentro del límite inferior,  $-5.5$ , es cero y el valor más alto dentro del límite superior,  $30.5$ , es 28. Seis valores (35, 36, 36, 39, 44 y 49) exceden el límite superior de  $30.5$  y son datos atípicos.

El *histograma* condensa los datos al agrupar valores similares en clases. Un histograma puede construirse al colocar la variable de interés en el eje horizontal y la frecuencia, la frecuencia relativa, o el porcentaje de frecuencia en el eje vertical. Al observar un histograma como el de la figura 2.4, se puede reconocer la proporción del área total sobre un intervalo del eje horizontal. Un total de 61 de las 209 compañías, o 29.2%, tuvo un ingreso neto como porcentaje del capital en un intervalo entre 7.5% y 12.5%. En consecuencia, el tercer rectángulo de la figura 2.4 contiene 0.292 del área total representada por las barras verticales sombreadas.

A pesar de que los histogramas proporcionan buenas descripciones visuales de los conjuntos de datos, en especial cuando éstos son muy grandes, no permiten identificar observaciones individuales. En contraste, en una gráfica de puntos cada uno de los valores originales queda visible.



**FIGURA 2.4 Histograma del ingreso neto como porcentaje del capital**



**FIGURA 2.5 Ventas mensuales para Alomega Food Stores del ejemplo 1.1**

Los *diagramas de dispersión* se utilizan para visualizar la relación existente entre dos variables y se presentarán más adelante en este capítulo en la sección de “Análisis de correlación”.

Para datos en orden cronológico, la presentación utilizada con mayor frecuencia es una *gráfica de series de tiempo*, en la cual los datos se grafican a lo largo del tiempo. La figura 2.5 muestra una gráfica de series de tiempo de las ventas mensuales analizadas en el ejemplo 1.1 para Alomega Food Stores. Una gráfica de series de tiempo revela la variabilidad

de los datos y el tiempo en que ocurren los picos y los valles. También muestra el tamaño relativo de los picos y los valles en comparación con el resto de la serie.

Un paso importante al elegir una técnica de pronósticos adecuada es la identificación de los patrones de datos que existen dentro de una serie de tiempo. Una vez que se han identificado los patrones de datos, pueden utilizarse los métodos de pronósticos más adecuados a tales patrones. Se pueden identificar cuatro tipos de patrones de datos de series de tiempo: horizontal, tendencia, cíclico y estacional. En el capítulo 3 se presentará con mayor detalle cada uno de ellos.

El correlograma, o función de autocorrelación, es una herramienta que se utiliza con frecuencia para ayudar a identificar patrones de datos de series de tiempo. Es un instrumento gráfico para mostrar las correlaciones entre diversos retrasos de una serie de tiempo. Los correlogramas también se presentarán en el capítulo 3.

## DISTRIBUCIONES DE PROBABILIDAD

---



---

Se le llama *variable aleatoria* a una cantidad que es capaz de tomar diferentes valores de una prueba a otra en un experimento, donde el resultado exacto es un suceso fortuito o aleatorio. Si sólo son posibles ciertos valores especificados, a la variable aleatoria se le denomina *variable discreta*. Ejemplos de ella incluyen el número de habitaciones en una casa, la cantidad de gente que llega a la caja de un supermercado en una hora y el número de unidades defectuosas en un lote de refacciones electrónicas. Si cualquier valor de la variable aleatoria es posible dentro de un rango, se le nombra *variable continua*. Los ejemplos de este tipo de variable son los pesos de las personas, la longitud de una pieza manufacturada y el tiempo entre las llegadas de los automóviles a una caseta de peaje.

Una *variable aleatoria discreta* puede tomar sólo valores de un grupo previamente determinado. Comúnmente, estos resultados se presentan mediante números enteros. Una *variable aleatoria continua* puede tomar cualquier valor dentro de un rango especificado. Estos resultados se presentan de forma numérica mediante un intervalo de valores.

La *distribución de probabilidad* de una variable aleatoria discreta enlista todos los valores posibles que puede tomar dicha variable, junto con la probabilidad de cada uno. El *valor esperado* de una variable aleatoria es el valor medio que toma la variable durante muchas pruebas. Se puede encontrar el valor esperado,  $E(X)$ , para una distribución de probabilidad discreta si se multiplica cada valor posible de  $X$  por su probabilidad  $P(X)$  y después se suman estos productos. La ecuación 2.4 muestra este cálculo:

$$E(X) = \sum [X \times P(X)] \quad (2.4)$$

El *valor esperado* de una variable aleatoria es el valor medio de la variable a lo largo de muchas pruebas u observaciones.

### Ejemplo 2.2

El número de días sin ventas de un vendedor durante un mes se describe mediante la distribución de probabilidad que se muestra en la tabla 2.4. Estos valores se basan en la experiencia del vendedor y se utilizarán para predecir la actividad futura de ventas. La columna  $X$  hace una lista de todos los valores posibles (días sin ventas), mientras que la columna  $P(X)$  tiene una lista de las probabilidades correspondientes. Observe que, puesto que todos los valores posibles de  $X$  están en la lista, las probabilidades suman 1.00 o 100%. Esto es cierto para todas las distribuciones de probabilidad, si se descarta el redondeo de los errores.

**TABLA 2.4 Distribución de probabilidad discreta**

<b>X</b>	<b>P(X)</b>
1	.10
2	.20
3	.25
4	.15
5	.30

Para la distribución de probabilidad dada en la tabla 2.4, el valor esperado es:

$$E(X) = 1(.1) + 2(.2) + 3(.25) + 4(.15) + 5(.30) = 3.35$$

Así, si se observara al vendedor durante muchos meses y se registrara el número de días sin ventas, la media sería 3.35. Esto ocurrirá si la actividad futura se predice de manera correcta a partir de los datos históricos en los que se basa la distribución de probabilidad.

Observe que la media cae cerca de la mitad de los valores de  $X$ . Se ve atraída hacia el extremo superior de la distribución de probabilidad debido a la probabilidad relativamente grande que se asocia con  $X = 5$ .

Para una distribución continua, la probabilidad de obtener un valor específico se approxima a cero. Por ejemplo, la probabilidad de que alguien pese 150 libras podría considerarse cero debido a que esto significaría que tal peso es exactamente 150.0000, sin importar cuán precisa sea la báscula que se utilice. En el caso de la distribución continua, se asignan probabilidades a los intervalos o rangos de valores. Por ejemplo, podría calcularse la probabilidad de que el peso de una persona caiga en el intervalo de 145 a 155 libras.

Algunas distribuciones teóricas ocurren una y otra vez en aplicaciones estadísticas prácticas y, por esta razón, es importante examinar sus propiedades. Una de ellas es la *distribución binomial*, que a menudo se utiliza para representar una variable aleatoria discreta. Los requisitos para un experimento binomial son:

1. Hay  $n$  pruebas idénticas, cada una de las cuales produce uno de dos resultados posibles, digamos éxito o fracaso.
2. La probabilidad de éxito de cada resultado permanece fija de una prueba a otra.
3. Las pruebas son independientes.

El interés es encontrar la probabilidad de  $X$  sucesos exitosos en  $n$  pruebas, donde un éxito se define arbitrariamente como uno de los dos resultados posibles. Los diversos valores de  $X$ , junto con sus probabilidades, forman la distribución binomial. Estas probabilidades pueden encontrarse a partir de la siguiente fórmula binomial:

$$P(X) = \binom{n}{X} p^X (1-p)^{n-X} \quad \text{para } X = 0, 1, \dots, n \quad (2.5)$$

donde

$\binom{n}{X}$  = número de combinaciones de  $n$  cosas de donde se toman  $X$  cada vez

$p$  = probabilidad de éxito en cada prueba

$X$  = número específico de éxitos

$n$  = número de pruebas

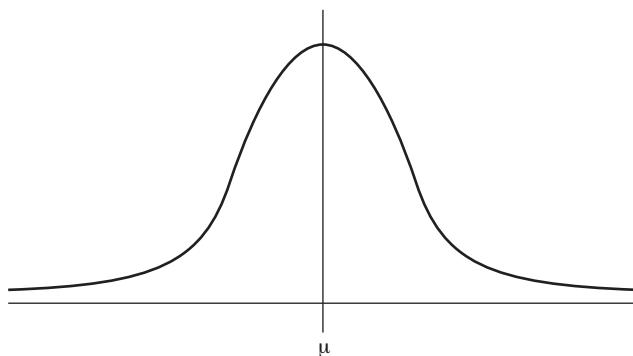


FIGURA 2.6 Distribución normal

La *distribución binomial* es una distribución de probabilidad discreta que describe la posibilidad de  $X$  éxitos en  $n$  pruebas de un experimento binomial.

Una forma más fácil de encontrar las probabilidades binomiales, en lugar de emplear la ecuación 2.5, consiste en referirse a una tabla de distribución binomial como la que se encuentra en el Apéndice C, tabla C-1. Los bloques que representan  $n$  agrupan las probabilidades. Cada bloque tiene una columna con un encabezado  $p$  y una fila indicada mediante  $X$ .<sup>1</sup>

### Ejemplo 2.3

Suponga que se sacan ocho artículos al azar de una línea de producción de la cual se sabe que hace partes defectuosas 5% del tiempo. ¿Cuál es la probabilidad de obtener exactamente cero defectos? La respuesta, a partir de la tabla binomial del Apéndice C-1, es 0.6634. (En este caso,  $n = 8$ ,  $p = 0.05$  y  $X = 0$ .)

Una distribución continua importante, debido a que muchas poblaciones de números se pueden aproximar mediante ella, es la *distribución normal*. Es necesario el conocimiento de la media y la desviación estándar para identificar una distribución normal específica. Una curva normal es simétrica y tiene forma de campana como se muestra en la figura 2.6. Esta distribución representa muchas variables prácticas que se miden en una escala continua.

Las probabilidades de los valores sacados de una distribución normal que caen en varios intervalos se encuentran al convertir, primero, todos los intervalos en unidades de desviación estándar llamados resultados  $Z$  como número de varianzas.<sup>2</sup> El resultado de  $Z$  como número de varianzas de cualquier valor de  $X$  es el número de desviaciones estándar desde el valor central de la curva ( $\mu$ ) hasta ese valor. Así, la fórmula es

$$Z = \frac{X - \mu}{\sigma} \quad (2.6)$$

donde

$X$  = valor de interés

$\mu$  = media

$\sigma$  = desviación estándar

<sup>1</sup> Excel y Minitab también calculan probabilidades binomiales.

<sup>2</sup> Se puede demostrar que si la variable aleatoria  $X$  tiene una distribución normal con media  $\mu$  y distribución estándar  $\sigma$ , entonces la variable aleatoria  $Z = (X - \mu)/\sigma$  tiene una distribución normal con una media 0 y una desviación estándar 1. Esta distribución normal específica se conoce como *distribución normal estándar*.

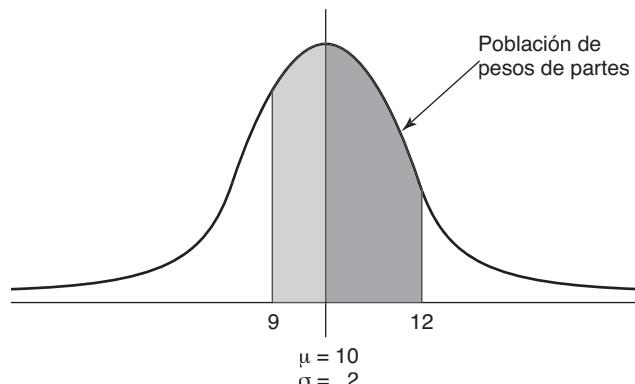


FIGURA 2.7 Áreas de curva normal para el ejemplo 2.4

Después de que se ha calculado  $Z$  como número de varianzas, se puede consultar la tabla de curva normal a fin de encontrar el área bajo la curva entre el centro de la curva original ( $\mu$ ) y el valor de interés  $X$ .<sup>3</sup>

La *distribución normal* tiene forma de campana y se determina completamente por su media y desviación estándar.

#### Ejemplo 2.4

Los pesos de una población de partes fabricadas por cierta máquina se distribuyen normalmente con una media de 10 libras y una desviación estándar de 2 libras. ¿Cuál es la probabilidad de que una parte que se tome al azar de la máquina pese entre 9 y 12 libras? La curva normal con el área sombreada apropiada se muestra en la figura 2.7.

Puesto que las tablas de la curva normal con frecuencia se diseñan para proporcionar áreas desde el centro de la curva hasta cierto punto, deben encontrarse dos áreas separadas: una a cada lado de la media. Luego, estas áreas se suman. El proceso generará la probabilidad de un valor que caiga en este intervalo. Los dos valores de  $Z$  como número de varianzas son:

$$Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{9 - 10}{2} = -.50$$

$$Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{12 - 10}{2} = 1.00$$

En consecuencia, el área bajo la curva normal entre 9 y 12 en la figura 2.7 es la misma que el área bajo la curva normal de  $Z$  (véase la nota 2 al pie) entre  $-0.5$  y  $1$ . Para la tabla normal en este libro, el signo negativo en el primer valor de  $Z$  como número de varianzas no se considera puesto que el área bajo la curva normal para  $Z$  entre  $-0.5$  y  $0$  es la misma que el área entre  $0$  y  $0.5$ . Estos dos valores de  $Z$  como número de varianzas se refieren a la tabla de la curva normal en el Apéndice C, tabla C-2, para producir las dos áreas, las cuales se suman juntas:

$$Z_1 = -.50 \rightarrow .1915$$

$$Z_2 = 1.00 \rightarrow \underline{.3413}$$

$$.5328$$

Se concluye que hay una probabilidad cercana a 53% de que un segmento tomado al azar de esta población de partes pese entre 9 y 12 libras.

<sup>3</sup> Las probabilidades normales se pueden calcular mediante Excel o Minitab.

## DISTRIBUCIONES MUESTRALES

En la mayoría de las aplicaciones estadísticas se toma una muestra al azar de la población que se investiga, se calcula una estadística a partir de los datos de la muestra y se sacan conclusiones sobre la población con base en esa muestra. Una *distribución muestral* es la distribución de todos los valores posibles de una estadística muestral, la cual puede obtenerse a partir de la población para un determinado tamaño de muestra. Por ejemplo, podría tomarse una muestra aleatoria de 100 personas de una población y pesarlas para calcular su peso medio. Esta media de la muestra ( $\bar{X}$ ) puede concebirse como aquella que se sacó de la distribución de todas las medias posibles de la muestra —de 100 elementos— que pudo haberse tomado de la población. Más en general, cada estadística de muestra que pueda calcularse a partir de datos de muestra puede considerarse como sacada de una distribución muestral.

Una *distribución muestral* es el arreglo de todos los valores posibles de una estadística de muestra que pueden sacarse de entre una población para un determinado tamaño de muestra.

El *teorema del límite central* establece que a medida que aumenta el tamaño de la muestra, la distribución muestral de las medias de la muestra tiende hacia la distribución normal, y que la media de esta distribución normal es  $\mu$ , la media de la población; asimismo, la desviación estándar es  $\sigma/\sqrt{n}$ , la desviación estándar de la población dividida entre la raíz cuadrada del tamaño de la muestra. Esta cantidad,  $\sigma/\sqrt{n}$ , se conoce como el *error estándar* de la media muestral. Así, la distribución muestral de la media de la muestra tenderá hacia la distribución normal sin importar la forma de la distribución de la población, de la cual se sacaron las muestras. La figura 2.8 comprueba cómo podría aparecer tal distribución muestral.

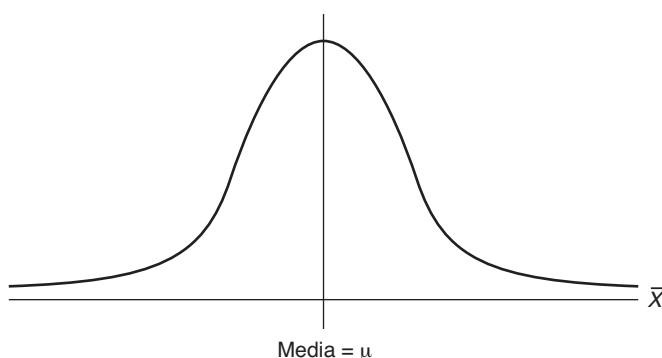
El teorema del límite central es de particular importancia en las estadísticas, ya que permite al analista calcular la probabilidad de varios resultados de muestra mediante el conocimiento de las probabilidades de curva normal.

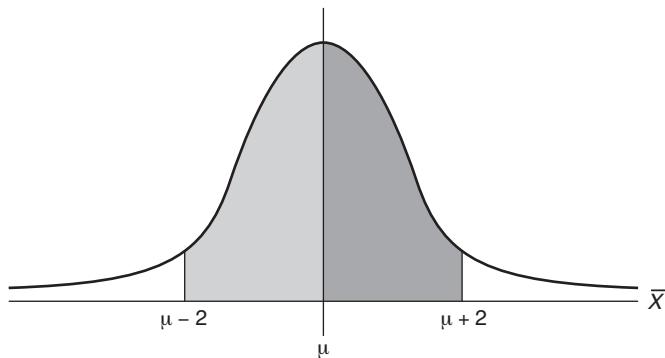
### Ejemplo 2.5

¿Cuál es la probabilidad de que la media de una muestra aleatoria de 100 pesos obtenidos de una población se encuentre dentro de dos libras de la media del peso verdadero de la población, si se estima que la desviación estándar de la población es de 15 libras? La figura 2.9 ilustra la distribución muestral adecuada.

El error estándar es  $\sigma/\sqrt{n} = 15/\sqrt{100} = 1.5$ , lo cual arroja un valor  $Z$  como número de variaciones de  $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n}) = 2/1.5 = 1.33$ , y un área a partir de la tabla de curva normal de 0.4082. Si se duplica dicha área para representar ambos lados de la media se obtiene como resultado un área total de 0.8164. La posibilidad es cercana a 82% de que la media de la muestra

**FIGURA 2.8 Distribución muestral de  $\bar{X}$**





**FIGURA 2.9 Áreas de distribución muestral para el ejemplo 2.5**

se encuentre dentro de dos libras de la media verdadera, considerando el tamaño de la muestra de 100 y una variabilidad estimada de la población,  $\sigma = 15$ . Como se verá, esta habilidad para calcular probabilidades de resultados de muestras le permitirá al analista hacer inferencias sobre la población con base en dicha muestra, lo cual tendrá aplicación directa en los pronósticos.

Para muestras pequeñas, podría ser que las medias de la muestra no se encuentren distribuidas de manera normal. Sin embargo, si la población de la cual se elige la muestra es normal en esencia, y si la desviación estándar de la población se estima mediante la desviación estándar de la muestra  $S$ , se conoce la distribución muestral de la relación:

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (2.7)$$

También llamada *distribución t de Student* (o simplemente la distribución *t*) con  $n - 1$  grados de libertad (df en forma abreviada, por sus siglas en inglés). La distribución *t* se centra en 0 y tiene forma de campana, pero con los extremos más amplios que la distribución normal. Sin embargo, al aumentar el número de grados de libertad (y de forma equivalente, el tamaño  $n$  de la muestra) la distribución *t* se aproxima cada vez más a una distribución normal con una media de 0 y una desviación estándar de 1.

El Apéndice C, tabla C-3, muestra valores tomados de la distribución *t*. Observe que se necesita especificar un valor antes de referirse a la tabla, a saber, los grados de libertad. Una vez que se conoce el valor de estos últimos, pueden encontrarse los valores *t* que excluyen los porcentajes deseados del área bajo la curva. Por ejemplo, si la distribución *t* de interés tiene 12 grados de libertad, entonces un valor *t* de 2.179 a cada lado de cero incluirá 95% del área bajo la curva y excluirá 5% de ella; es decir, 2.5% en cada extremo.

### Ejemplo 2.6

El gerente de una compañía de tarjetas de crédito declara que el tiempo medio para resolver reclamaciones es de 30 días. Para probar esta afirmación, se recolectaron datos de una muestra de 15 quejas. Éstas tenían un tiempo medio para resolverse de 35.9 días y una desviación estándar de 10.2 días.

Si se consideran los datos como una muestra aleatoria de una población normal, ¿arrojan alguna sospecha sobre la declaración del gerente?

Si lo aseverado por el gerente es correcto y la media de la población es  $\mu = 30$ , la variable aleatoria

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución *t* de Student con  $n - 1 = 15 - 1 = 14$  grados de libertad. La declaración del gerente sería sospechosa si el valor de esta variable fuera demasiado grande o demasiado

pequeño debido a que esto sucede cuando el valor de  $\bar{X}$  se encuentra lejos de  $\mu$ . Al consultar la tabla C-3 para 14 grados de libertad, se observa que  $t_{0.025} = 2.145$ . Antes de obtener esta muestra, la probabilidad de que  $t$  exceda a 2.145 es de 0.025. De manera similar, la probabilidad de que  $t$  sea menor que  $-2.145$  es de 0.025. Al sumar estas dos probabilidades iguales, la probabilidad es  $0.025 + 0.025 = 0.05$  de que  $t$  será mayor que 2.145 o menor que  $-2.145$ . Ahora estamos listos para determinar si el valor  $\bar{X} = 35.9$ , se encuentra alejado de  $\mu = 30$ . El valor observado de  $t$  es

$$t = \frac{35.9 - 30}{10.2/\sqrt{15}} = 2.240$$

el cual excede  $t_{0.025} = 2.145$ . En vez de aceptar la explicación de que ha ocurrido un evento con una probabilidad pequeña, se cuestiona la afirmación de que la media sea de 30 días. Un valor mayor para la media hipotética sería más coherente con los datos.

## INFERENCIA A PARTIR DE UNA MUESTRA

---

### Estimación

Existen dos razones principales para examinar la evidencia muestral cuando medir la población entera de interés no es viable o no es posible. La primera de ellas se conoce como *estimación*, en la cual los resultados de la muestra se utilizan para evaluar características desconocidas de la población. Aunque estimación es el término estadístico común para esta tarea, también podría llamársele pronóstico en muchas situaciones de negocios debido a que los datos consisten en una colección de observaciones históricas, y el valor para el cual se desea la estimación —o pronóstico— se encuentra en el futuro desconocido. El segundo propósito fundamental del examen de evidencia de muestras, la prueba de hipótesis, se analizará en la siguiente sección.

Una *estimación puntual* de un parámetro de población (pronóstico) es un valor único que se calcula a partir de los datos de la muestra y que estima el valor desconocido de la población. Una estimación puntual es la “mejor conjectura” de un parámetro de población calculado a partir de la muestra. A menudo, la mejor conjectura de un parámetro de población la proporciona la correspondiente cantidad obtenida a partir de la muestra. Por ejemplo, la mejor conjectura del valor de la media de la población está dada por el valor de la media de la muestra. La tabla 2.2 contiene tres parámetros de población y las estadísticas de la muestra que proporcionan estimaciones puntuales de ellos.

Un *intervalo estimado* es aquel donde probablemente esté el parámetro de la población de interés. Se encuentra al construir un intervalo alrededor de la estimación puntual, de la siguiente forma:

$$\text{Estimación puntual} \pm \text{Múltiplo} \times (\text{Desviación estándar estimada de la estimación puntual}) \quad (2.8)$$

donde el “múltiplo” frecuentemente es un punto porcentual de distribución normal o un punto porcentual de distribución  $t$ .

De manera ideal, nos gustaría tener una muestra y entonces determinar un intervalo que, por ejemplo, contuviera definitivamente la media de la población  $\mu$ . Sin embargo, debido a la variabilidad de una muestra a otra, este objetivo no puede lograrse. En vez de ello, se especifica una probabilidad elevada —por ejemplo, 0.95 o 0.99— de que el intervalo propuesto cubrirá el valor real del parámetro de la población de interés. Ya que esta probabilidad pertenece al intervalo antes de que se observe la muestra, se conoce como *nivel de confianza*. Un intervalo obtenido mediante el procedimiento que satisface el requisito de probabilidad se conoce como *intervalo de confianza*. Comúnmente, la cantidad a la derecha del signo  $\pm$  en la ecuación 2.8 se conoce como el *error permitido* y su tamaño dependerá directamente del nivel de confianza (mediante el múltiplo) y de la información en la muestra (mediante la desviación estándar estimada de la estimación puntual).

Una *estimación puntual* es una estimación de valor único de un parámetro de la población.

Un *intervalo estimado* es aquel dentro del cual es probable que se encuentre el parámetro de la población.

### Ejemplo 2.7

Se elige una muestra aleatoria de 500 consumidores en el centro comercial Northgate para determinar la distancia promedio que recorrieron para llegar ahí. Un análisis de los resultados de la muestra revela que  $\bar{X} = 23.5$  millas y  $S = 10.4$  millas.

La estimación puntual para la distancia media desconocida de la población de todos los clientes de un centro comercial es de 23.5 millas que es la media de la muestra. El intervalo estimado para  $\mu$  al utilizar un nivel de confianza de 95% y mediante la ecuación 2.9 que está a continuación, es:

$$\bar{X} \pm Z \frac{S}{\sqrt{n}} \quad (2.9)$$

o

$$23.5 \pm 1.96 \frac{10.4}{\sqrt{500}}$$

$$23.5 \pm 1.96(.465)$$

$$23.5 \pm .91 \rightarrow (22.6, 24.4)$$

Observe que el múltiplo en la ecuación 2.8, en este caso, es el punto 0.025 (2.5%) superior de una distribución normal estándar (vea la nota 2 al pie), y  $S/\sqrt{n} = 10.4/\sqrt{500}$  es la desviación estándar estimada de  $\bar{X}$  para un tamaño de muestra  $n = 500$ .

Se puede establecer, con una confianza de 95%, que el número medio de millas recorridas hasta el centro comercial por la población de compradores se encuentra entre 22.6 y 24.4 millas. La confianza de 95% se refiere al hecho de que, por ejemplo, si se seleccionaran 100 muestras de tamaño 500 y se calcularan las medias de la muestra y las desviaciones estándar, y se construyeran los intervalos estimados, 95 de cada 100 intervalos contendrían, de hecho, la media real de la población. En consecuencia, es altamente probable que el intervalo específico que se construyó anteriormente contenga la media de millas recorridas por la población hasta el centro comercial.

## PRUEBA DE HIPÓTESIS

En muchas situaciones estadísticas, que incluyen los pronósticos, el interés está en probar cierta afirmación sobre la población más que en estimar o predecir alguno de sus parámetros. A este proceso se le llama prueba de hipótesis y es el segundo propósito básico que se puede perseguir al examinar la evidencia de las muestras. Probar las hipótesis implica los siguientes pasos:

**Paso 1.** Formular la hipótesis que se ha de probar (llamada hipótesis nula, cuyo símbolo es  $H_0$ ) y establecer la hipótesis alternativa (la cual concluye si se rechaza  $H_0$ , su símbolo es  $H_1$ ).

**Paso 2.** Recopilar una muestra aleatoria de artículos de la población, medirlos y calcular la estadística adecuada de la prueba de la muestra.

**Paso 3.** Suponer que la hipótesis nula es verdadera y determinar la distribución muestral de la estadística de la prueba.

**Paso 4.** Calcular la probabilidad de que el valor de la estadística de la muestra sea, por lo menos, tan grande como el que se observó que podría haberse sacado de esta distribución muestral.

**Paso 5.** Si esta probabilidad es alta, no se rechaza la hipótesis nula; si esta probabilidad es baja, la hipótesis nula se desacredita y puede rechazarse con una pequeña posibilidad de error.

Cuando se siguen estos pasos, pueden ocurrir dos tipos de errores, como se muestra en la tabla 2.5. Se espera llegar a la decisión correcta respecto a la hipótesis nula después de que

TABLA 2.5 Resultados de una prueba de hipótesis

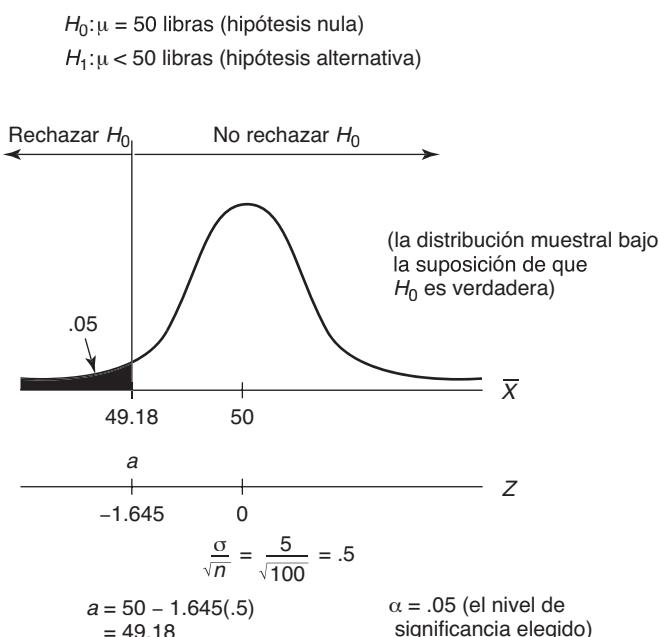
Estado de la naturaleza	$H_0$ verdadera	Acción	
		No rechazar $H_0$	Rechazar $H_0$
	$H_0$ verdadera	Decisión correcta	Error tipo II: probabilidad $\alpha$
	$H_0$ falsa	Error tipo I: probabilidad $\beta$	Decisión correcta

se ha examinado la evidencia de la muestra, pero siempre existe la posibilidad de que se rechace una  $H_0$  verdadera y de no rechazar una  $H_0$  falsa. Las probabilidades de estos eventos se conocen como alfa ( $\alpha$ ) y beta ( $\beta$ ), respectivamente. Alfa también es conocida como *nivel de significancia* de la prueba.

### Ejemplo 2.8

El supervisor de un taller sospecha que el peso medio de las partes producidas por cierta máquina ha disminuido. Se desea probar la hipótesis de que el peso medio de las piezas fabricadas todavía es de 50 libras, que es peso medio de las partes en años pasados. Se toma una muestra aleatoria de 100 piezas y se supone que la desviación estándar de los pesos de las partes es de cinco libras sin importar el peso medio, debido a que este valor se ha mantenido constante en estudios anteriores sobre las partes. Se utilizará la media de la muestra como estadística de prueba y se tomará como hipótesis nula el *status quo*; es decir, que el peso medio  $\mu$  de cada parte es de 50 libras. En consecuencia, si la hipótesis nula  $H_0$  es verdadera, el teorema del límite central dice que la distribución muestral de  $\bar{X}$  es normal en este caso, con una media  $\mu = 50$  y una desviación estándar  $\sigma/\sqrt{n} = 5/\sqrt{100} = 0.5$ . Un valor de la media de la muestra,  $\bar{X}$ , cercano o mayor que 50 apoya a  $H_0$ , y un valor de  $\bar{X}$  mucho menor a 50 es evidencia en contra de  $H_0$ , puesto que tal evento es poco probable si  $H_0$  es verdadera. Una regla sensata para decidir es, entonces, la que rechaza  $H_0$  por un valor de mucho menor a 50. Sin embargo, con esta regla es posible que se cometa un error tipo I, o sea, que se rechace una  $H_0$  verdadera. Se puede controlar la posibilidad de rechazar una  $H_0$  verdadera al seleccionar alfa, el nivel de significancia. Suponga que se disminuye el valor alfa, digamos 0.05 (5%). Entonces se rechazaría  $H_0$  si el valor observado de la estadística de prueba,  $\bar{X}$ , cae más allá del punto inferior 0.05 (5%) de la distribución muestral por debajo de  $H_0$ . La prueba se muestra en la figura 2.10.

FIGURA 2.10 Prueba de hipótesis para el ejemplo 2.8



La regla de decisión para esta prueba es:

$$\text{Si } \bar{X} < 49.18, \text{ rechazar la hipótesis nula } H_0: \mu = 50$$

Si la media de la muestra resulta ser 49.6 libras, no se podría rechazar  $H_0$  al nivel de 5%. Por otro lado, si el peso de la media de la muestra resultara ser de 48.6 libras, se rechazaría  $H_0$ . Observe que para  $\bar{X} = 48.6$ , la probabilidad de obtener un valor así de extremo (mucho menor a 50) si  $H_0$  es verdadera es  $P(\bar{X} < 48.6) = P(Z < \frac{48.6 - 50}{\frac{5}{\sqrt{15}}}) = P(Z < -2.8) = P(Z > 2.8) = (0.5000 - 0.4974) = 0.0026$ , un evento poco probable.

En la prueba de hipótesis siempre hay algo de ambigüedad asociada con la elección de las hipótesis nula y alternativa. En general, la hipótesis nula es el *statu quo* o hipótesis “sin cambio”. La hipótesis alternativa es la hipótesis de investigación o de cambio. Para descubrir si lo que se ha realizado ha generado alguna diferencia o ha derivado en un cambio del *statu quo* o de los procedimientos actuales, se formula la diferencia o cambio como la hipótesis alternativa. El objetivo es ver si se puede rechazar la hipótesis nula, con una posibilidad pequeña de un error tipo I en favor de la alternativa.

### Valor $p$

En el ejemplo 2.8 se calculó la probabilidad de obtener un valor de la media de la muestra tan extremo como  $\bar{X} = 48.6$  si la hipótesis nula  $H_0: \mu = 50$  fuera verdadera. La probabilidad resultó ser 0.0026. Esta probabilidad se conoce como *valor-prob*, o *valor p*, de la prueba. En lugar de seleccionar un nivel de significancia, hoy es común que se calcule y reporte el valor  $p$  de la prueba; de hecho, los paquetes de software estadístico reportan de manera rutinaria valores  $p$  asociados con estadísticas de pruebas. Los valores  $p$  y su uso en las pruebas de hipótesis se analizan en los pasos 4 y 5 de la prueba de hipótesis presentada al principio de esta sección. Un valor  $p$  pequeño sugiere un rechazo fuerte de la hipótesis nula. El valor  $p$  puede verse como evidencia en favor (valor  $p$  grande) o en contra (valor  $p$  pequeño) de la hipótesis nula. Los puntos de corte típicos para valores  $p$  pequeños son 0.05 o 0.01.

En el ejemplo 2.8, si se hubiera llevado a cabo una prueba de dos colas  $H_0: \mu = 50$  contra  $H_1: \mu \neq 50$ , y  $\bar{X} = 48.6$ , el valor  $p$  es  $P(\bar{X} < 48.6 \text{ o } \bar{X} > 51.4) = 2P(Z > \frac{51.4 - 50}{\frac{5}{\sqrt{15}}})$ , o dos veces el área bajo la curva normal estándar a la derecha de  $Z = 2.8$ . En este ejemplo, el valor  $p$  sería igual a  $2(0.0026) = 0.0052$ .

El *valor p*, o *probabilidad de la significancia*, es la probabilidad de obtener, cuando menos, un resultado de la muestra tan extremo como el que se observaría en realidad si  $H_0$  fuera verdadera.

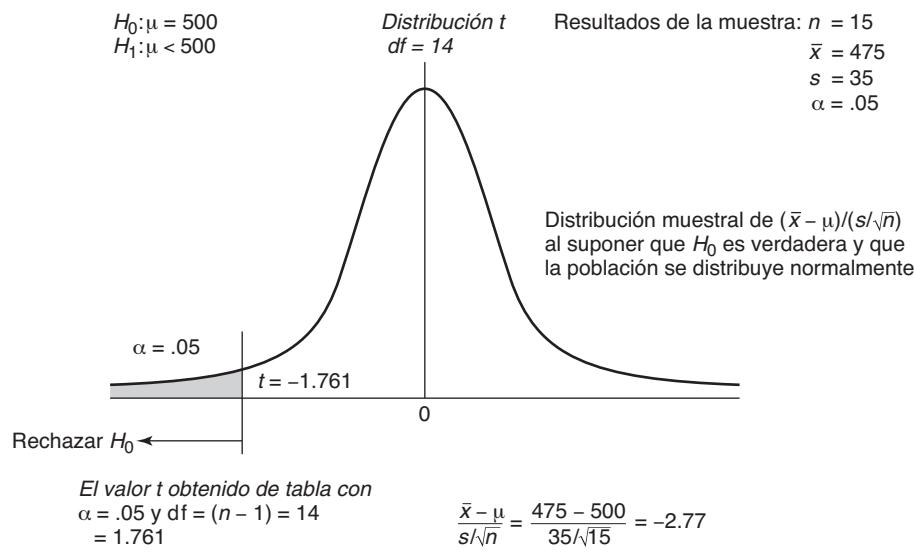
### Ejemplo 2.9

Suponga que se desea probar la hipótesis de que el puntaje promedio de los estudiantes en un examen nacional es de 500, contra la hipótesis alternativa de que es menor a 500. Se toma una muestra aleatoria de 15 estudiantes de entre la población y produce un puntaje promedio de  $\bar{X} = 475$  en tal muestra. La desviación estándar de la población se estima mediante la desviación estándar de la muestra,  $S = 35$ . Suponga que la población de los puntajes en los exámenes se distribuye normalmente.

Se quiere saber si los puntajes promedio de la prueba son menores que su valor histórico de 500. Se tomará una decisión con base en la evidencia muestral y, en particular, con base en el valor de la media de la muestra. Las hipótesis en competencia son

$$\begin{aligned} H_0: \mu &= 500 \\ H_1: \mu &< 500 \end{aligned}$$

Si  $H_0$  es verdadera, se esperaría ver un valor promedio de la muestra cercano o mayor a 500. Si  $H_0$  es falsa ( $H_1$  es verdadera), se esperaría ver un valor del promedio de la muestra algo menor que 500. Debido a que la muestra es pequeña, la población es normal y la desviación estándar de la población se estima mediante la desviación estándar de la muestra. Una estadística adecua-

**FIGURA 2.11 Prueba de hipótesis para el ejemplo 2.9**

da es la estadística  $t$  que se da en la ecuación 2.7, donde  $n - 1 = 15 - 1 = 14$  df. Un elevado valor negativo de la estadística  $t$  es evidencia en contra de  $H_0$  (el promedio de la muestra es mucho menor que el promedio hipotético de 500 de la población). Primero, construyamos la prueba para un nivel de significancia de  $\alpha = 0.05$ . La prueba (regla de decisión) se muestra en la figura 2.11. Puesto que

la regla de decisión es: si  $t$  observado  $< -1.76$ , rechazar  $H_0: \mu = 500$

y la estadística observada  $t$  es

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{475 - 500}{35/\sqrt{15}} = -2.77$$

se rechazaría la hipótesis nula, y se concluiría que el resultado medio de las pruebas de los estudiantes en el examen nacional es menor a 500. De manera alterna, el valor  $p$  en este caso es: valor  $p = P(t < -2.77)$ . Si se usa la tabla C-3 con 14 df y la simetría de la distribución  $t$ , esta probabilidad está entre 0.010 y 0.005. Nuestra estadística observada de la prueba es muy poco probable si  $H_0$  es verdadera. El valor  $p$  nos dice que se debe rechazar la hipótesis nula.

## ANÁLISIS DE CORRELACIÓN

Al construir modelos estadísticos para pronósticos, a menudo resulta útil examinar la relación entre dos variables. Aquí se revisan dos técnicas, los análisis de correlación y de regresión. Además, en capítulos posteriores se consideran casos especiales de correlación y regresión. Este énfasis en la correlación y la regresión se justifica en vista del amplio uso de estas técnicas en todo tipo de aplicaciones de pronósticos.

### Diagramas de dispersión

El estudio de la relación entre variables comienza con el caso más sencillo, el de la relación que existe entre dos variables. Suponga que se han tomado medidas en cada uno de dos objetos diversos. Un analista desea determinar si una de estas variables —que se pueden medir— llamada  $Y$ , tiende a aumentar o a disminuir a medida que cambia la otra variable, llamada  $X$ . Por ejemplo, suponga que se han medido tanto la edad como los ingresos de varios

individuos, como se muestra en la tabla 2.6. ¿Qué puede decirse sobre la relación entre  $X$  y  $Y$ ?

De la tabla 2.6 parece que  $Y$  y  $X$  tienen una relación definitiva. A medida que aumenta  $X$  también se acrecienta  $Y$ . Parece que en esta muestra de cinco personas cuanto mayor es el sujeto más dinero gana. Por supuesto, es peligroso llegar a conclusiones con base en una muestra pequeña, tema del que se hablará posteriormente. Aun así, con estas observaciones, parece existir una relación definitiva entre  $Y$  y  $X$ .

Estos cinco puntos de datos se pueden trazar en una escala bidimensional, con valores de  $X$  sobre el eje horizontal y los valores de  $Y$  sobre el vertical. Tal gráfica se conoce como *diagrama de dispersión* y aparece en la figura 2.12.

Un *diagrama de dispersión* es una gráfica de puntos de datos  $X$ - $Y$  en sistema de coordenadas bidimensional.

El diagrama de dispersión ayuda a ilustrar lo que sugiere la intuición cuando se observan a primera vista los datos sin procesar, en especial, la aparición de una relación lineal entre  $Y$  y  $X$ . A esta relación se le llama *positiva* debido a que al aumentar  $X$  también lo hace  $Y$ .

En otras situaciones donde participan dos variables, podrían surgir diferentes patrones de diagramas de dispersión. Considere las gráficas en la figura 2.13.

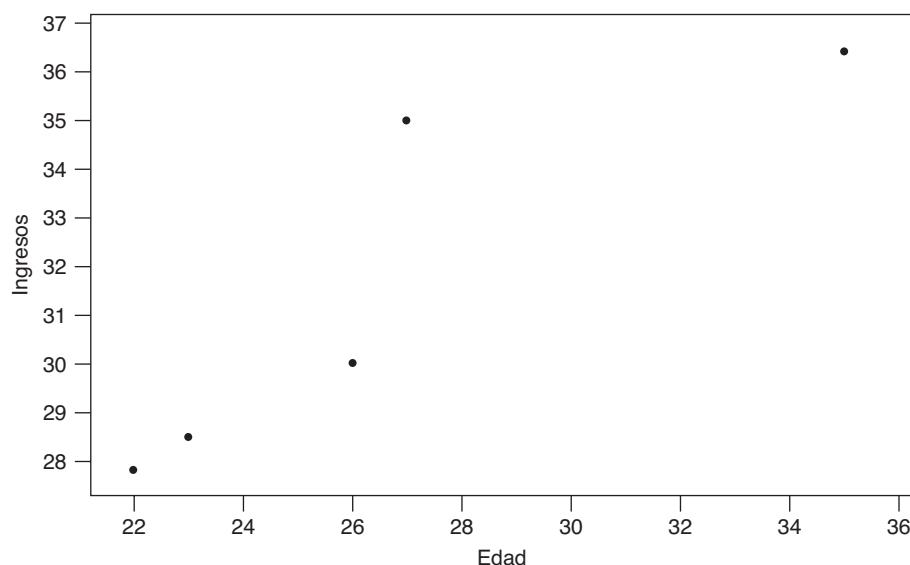
El diagrama (a) de la figura 2.13 sugiere lo que se llama una *relación lineal positiva perfecta*. Al incrementarse  $X$ , también aumenta  $Y$ , y lo hace de una forma perfectamente

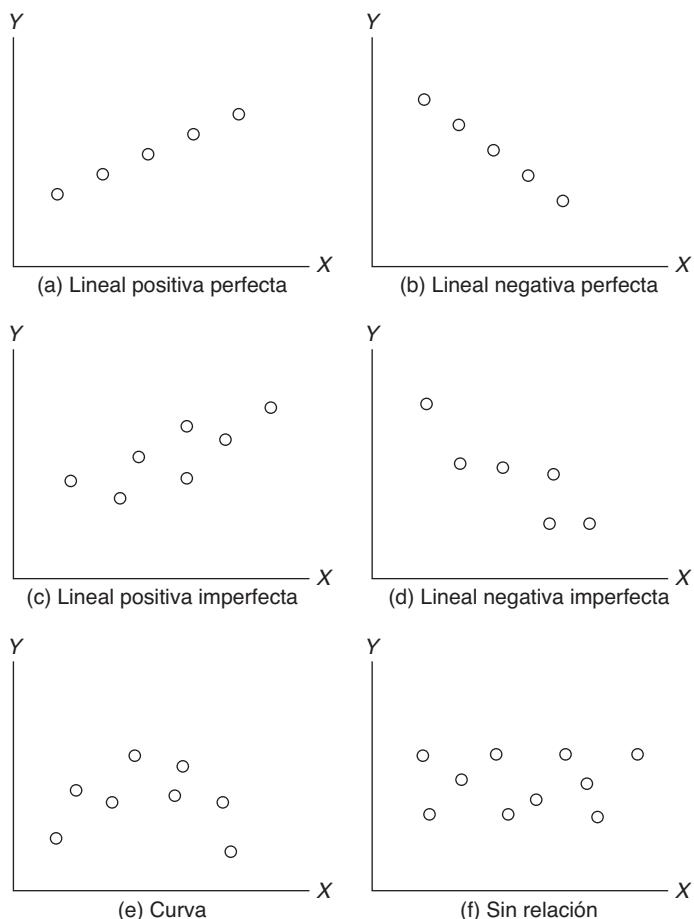
**TABLA 2.6 Medidas de ingresos y edad**

Persona	$Y$ , ingresos (miles de \$)	$X$ , edad (años)
1	27.8	22
2	28.5	23
3	30.0	26
4	35.0	27
5	36.4	35

**FIGURA 2.12 Diagrama de dispersión para edad e ingresos**

**Gráfica de dispersión de edad contra ingresos**





**FIGURA 2.13** Diagrama de dispersión para edad e ingresos

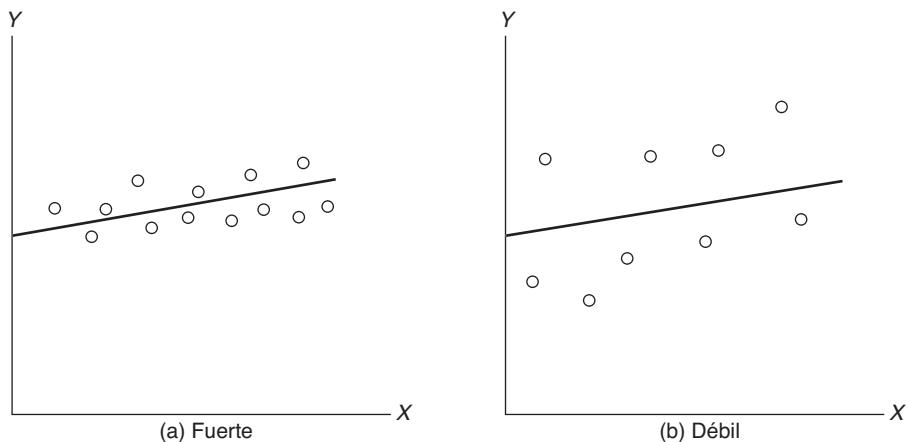
predecible. Es decir, los puntos de los datos de  $X$  y  $Y$  parecen caer sobre una línea recta. El diagrama (b) sugiere una *relación lineal negativa perfecta*. A medida que  $X$  aumenta,  $Y$  disminuye de forma predecible.

La figura 2.13 (c) y (d) ilustra, respectivamente, una *relación lineal positiva imperfecta* y una *negativa imperfecta*. Al aumentar  $X$  en estos diagramas de dispersión,  $Y$  incrementa (c) o reduce (d), pero no de una forma perfectamente predecible. Así,  $Y$  podría ser un poco más alta o más baja de lo “esperado”. Es decir, los puntos  $X$ - $Y$  no caen sobre una línea recta.

Los diagramas de dispersión en la figura 2.13, desde (a) hasta (d), ilustran *relaciones lineales*. La relación  $X$ - $Y$ , ya sea perfecta o imperfecta, se puede resumir mediante una línea recta. En comparación, una *relación curva* aparece en el diagrama (e).

Por último, el diagrama (f) de la figura 2.13 sugiere que no existe relación alguna entre las variables  $X$  y  $Y$ . A medida que  $X$  aumenta, parece que  $Y$  ni aumenta ni disminuye de forma predecible. Con base en la evidencia de la muestra que aparece en el diagrama (f), se podría concluir que en el conjunto universal que contiene todos los puntos de datos  $X$ - $Y$  no existe relación lineal alguna —ni cualquiera otra— entre las variables  $X$  y  $Y$ .

Ahora considere los dos diagramas de dispersión en la figura 2.14. Ambos sugieren relaciones lineales positivas imperfectas entre  $Y$  y  $X$ . La figura 2.14 (a) muestra una relación fuerte, puesto que todos los puntos de datos están bastante cerca de la línea recta que pasa entre ellos. En la figura 2.14 (b) los puntos de datos se encuentran más alejados de la



**FIGURA 2.14 Asociación lineal fuerte y débil para gráficas de datos X-Y**

línea recta que pasa entre ellos, lo cual sugiere una relación lineal más débil. Más adelante en este capítulo, se demostrará cómo medir la fortaleza de la relación que existe entre dos variables.

Como sugieren los dos diagramas de dispersión en la figura 2.14, con frecuencia es deseable resumir la relación entre dos variables al ajustar una línea recta entre los puntos de datos. Pronto aprenderá cómo se realiza esto, pero por el momento se puede decir que es posible ajustar una línea recta a los puntos de un diagrama de dispersión para que resulte un “buen” ajuste. Una pregunta interesante para los pronósticos es: ¿qué tan rápido sube o baja esta línea recta?

Responder esta pregunta requiere del cálculo de la pendiente de la línea. La pendiente de cualquier línea recta se define como el cambio en  $Y$  asociado al aumento de una unidad en  $X$ .

Para resumir, cuando se investiga la relación entre dos variables primero debe saberse si es lineal (una línea recta) o no lineal. Si es lineal, es necesario saber si tal relación es positiva o negativa y qué tan abruptamente sube o cae la recta que se ajusta a los puntos de datos. Por último, se requiere la fortaleza de la relación, es decir, qué tan cerca se encuentran los puntos de datos de la recta que se ajusta mejor a ellos.

## COEFICIENTE DE CORRELACIÓN

La fortaleza de la relación lineal que existe entre dos variables se mide mediante la *correlación* que existe entre ellas. El *coeficiente de correlación* mide la fuerza de la siguiente forma: dos variables con una relación negativa perfecta tienen un coeficiente de correlación igual a  $-1$ . En el otro extremo, dos variables con una relación positiva perfecta tienen un coeficiente de correlación igual a  $+1$ . Así, el coeficiente de correlación varía entre  $-1$  y  $+1$ , incluyéndolos, dependiendo de la cantidad de asociación entre las dos variables que se miden.

El *coeficiente de correlación* mide qué tanto se relacionan linealmente dos variables entre sí.

El diagrama de dispersión (a) de la figura 2.13 ilustra una situación que produciría un coeficiente de correlación de  $+1$ . El diagrama de dispersión (b) tiene un coeficiente de correlación de  $-1$ . Los diagramas (e) y (f) muestran dos variables que no se relacionan lineal-

mente. Los coeficientes de correlación para estas relaciones son iguales a 0; es decir, no hay relación lineal alguna.

A los pronosticadores les preocupa tanto la población como los datos de la muestra. En la *población* que contiene todos los puntos de datos  $X-Y$  de interés, existe un coeficiente de correlación cuyo símbolo es  $\rho$ , la letra griega rho. Si se dibuja una muestra aleatoria de estos puntos de datos  $X-Y$ , el coeficiente de correlación de éstos se denota mediante  $r$ .

Con frecuencia,  $X$  y  $Y$  se miden en unidades distintas, como libras y dólares, unidades de venta e importe de venta, o tasa de desempleo o el importe del producto interno bruto (PIB). A pesar de estas unidades de medida diferentes para  $X$  y  $Y$ , aún es importante medir cuánto se relacionan  $X$  y  $Y$ . Esta medida se realiza al convertir primero las variables  $X$  y  $Y$  en unidades estandarizadas, o  $Z$  como número de varianza.

Una vez que las medidas  $X-Y$  se han convertido en  $Z$  como número de varianza, se multiplican estos valores  $Z$  para cada medida  $X-Y$ , lo que da como resultado productos cruzados para cada caso. Estos productos cruzados son interesantes puesto que la media de los valores es el coeficiente de correlación. El cálculo del coeficiente de correlación como el producto cruzado medio de  $Z$  como número de varianza producirá en esencia el valor correcto, pero en la mayoría de los casos el coeficiente de correlación se calcula directamente desde los valores  $X-Y$ . La ecuación 2.10 muestra cómo calcular el coeficiente de correlación de la muestra  $r$  a partir de  $Z$  como número de varianza y de las medidas  $X-Y$ . Aquí,  $Z_X = (X - \bar{X}) / S_X$  y  $Z_Y = (Y - \bar{Y}) / S_Y$ .

$$\begin{aligned} r &= \frac{1}{n-1} \sum Z_X Z_Y = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}} \\ &= \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \end{aligned} \quad (2.10)$$

Una tabla tal como la 2.7 facilita el cálculo de  $r$ .

### Ejemplo 2.10

Si se estudia la relación entre la edad y los ingresos (vea la tabla 2.6), podría ser interesante conocer el valor de  $r$  para estos datos. Los cálculos requeridos aparecen en la tabla 2.7. Los totales de la tabla 2.7 se sustituyen en la ecuación 2.10, de donde resulta:

$$\begin{aligned} r &= \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \\ &= \frac{5(4,266.1) - (133)(157.7)}{\sqrt{5(3,643) - (133)^2} \sqrt{5(5,035.05) - (157.7)^2}} \\ &= \frac{21,330.5 - 20,974.1}{\sqrt{(18,215 - 17,689) \sqrt{(25,175.25 - 24,869.29)}}} = \frac{356.4}{401.17} = .89 \end{aligned}$$

Se puede observar que el coeficiente de correlación de la muestra confirma lo observado en la figura 2.12. El valor para  $r$  es positivo, lo cual sugiere una relación lineal positiva entre edad e ingresos. Además, en una escala de 0 a 1, el valor de  $r$  es bastante alto (0.89). Este resultado sugiere una relación lineal fuerte, más que una débil. La cuestión que queda es si la combinación del tamaño de la muestra y el coeficiente de correlación es tan fuerte como para hacer afirmaciones significativas sobre la población de la que se obtuvieron los datos.

Ahora, deben enfatizarse dos puntos importantes en el análisis de las correlaciones. Primero, siempre debe considerarse que se está midiendo la *correlación*, no la *causalidad*.

**TABLA 2.7 Cálculo para la correlación de edad e ingresos del ejemplo 2.10**

Personas	Y	X	$Y^2$	$X^2$	$XY$
1	27.8	22	772.84	484	611.6
2	28.5	23	812.25	529	655.5
3	30.0	26	900.00	676	780.0
4	35.0	27	1,225.00	729	945.0
5	36.4	35	1,324.96	1,225	1,274.0
Totales	157.7	133	5,035.05	3,643	4,266.1

Podría ser perfectamente válido decir que dos variables se relacionan con base en un alto coeficiente de correlación. Podría ser válido, o no, decir que una variable está causando el movimiento de otra; ésa es una cuestión para el juicio del analista. Por ejemplo, podría ser cierto que el volumen de ventas de una tienda provincial en un área con poca población tiene una alta correlación con el precio promedio del mercado de valores de Nueva York. Se podría concluir, después de examinar una muestra grande de estas dos variables, que esa elevada correlación sí existe. Probablemente sea falso que una de estas variables cause movimiento en la otra. De hecho, los movimientos de ambas variables quizás estén causados por un tercer factor: el estado general de la economía. Con frecuencia, políticos, publicistas y otros cometan el error de suponer causalidad con base en la correlación.

Segundo, observe que el coeficiente de correlación mide una relación *lineal* entre dos variables. En caso de que el coeficiente de correlación sea bajo, se puede concluir que las dos variables no están muy relacionadas linealmente. Podría ser que estén relacionadas estrechamente de forma no lineal o curva. Así, un coeficiente de correlación bajo no significa que las dos variables no se relacionen, sólo que aparentemente no existe una relación lineal o de líneas rectas.

## AJUSTE DE UNA LÍNEA RECTA

Como se mencionó anteriormente, a menudo resulta interesante ajustar una línea recta entre una colección de puntos de datos  $X-Y$  de forma que resulte un “buen ajuste”. Como se mostrará en capítulos posteriores, tal recta bien ajustada se puede utilizar para predecir valores desconocidos de  $Y$ , la variable de interés.

En la práctica, se utiliza un procedimiento matemático bien definido para calcular tanto la intersección con el eje  $Y$  como la pendiente de la línea recta que mejor se ajusta. El método más común para determinar dicha línea mejor ajustada se conoce como *método de mínimos cuadrados*. En este procedimiento, se formula una línea que minimiza la suma de las distancias al cuadrado entre los puntos de los datos y la línea, medidos en la dirección vertical ( $Y$ ).

El método de mínimos cuadrados elige los valores para la pendiente y para la intersección con el eje  $Y$ , a fin de minimizar la suma de los cuadrados del error –distancias– (SSE, siglas de *sum of squared errors*) entre los valores  $Y$  y la recta:

$$SSE = \Sigma(Y - \hat{Y})^2 = \Sigma(Y - b_0 - b_1 X)^2 \quad (2.11)$$

donde  $\hat{Y} = b_0 + b_1 X$  es el valor estimado de  $Y$  en la recta ajustada.

En los cálculos (véase el Apéndice A) se pueden derivar expresiones algebraicas específicas para los valores de mínimos cuadrados. De manera particular,

$$b_1 = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} \quad (2.12)$$

$$b_0 = \bar{Y} - b_1 \bar{X} = \frac{\Sigma Y}{n} - \frac{b_1 \Sigma X}{n} \quad (2.13)$$

donde

$b_1$  = pendiente de la recta

$b_0$  = intersección con el eje  $Y$

El *método de mínimos cuadrados* se utiliza para calcular la ecuación de una línea recta que minimiza la suma de las distancias al cuadrado entre los puntos de datos  $X-Y$  y la línea, medida en la dirección del eje vertical ( $Y$ ).

### Ejemplo 2.11

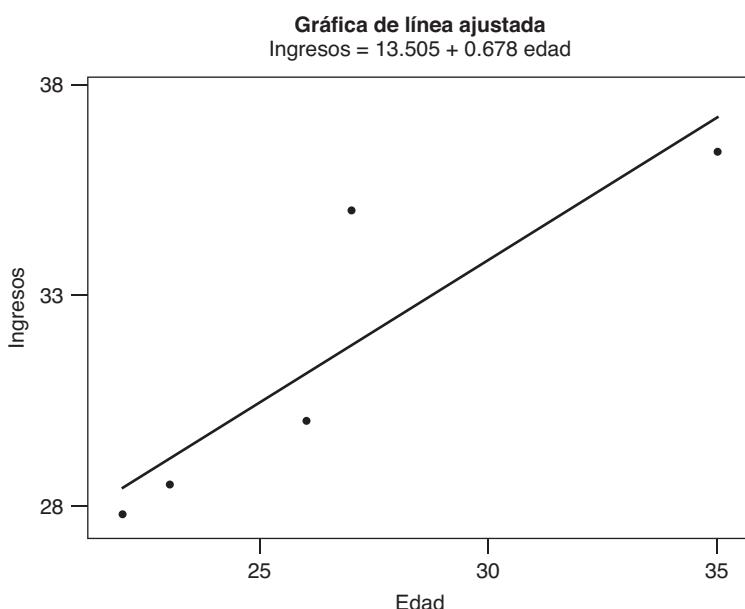
El ejemplo 2.10 sugirió una fuerte relación lineal positiva entre la edad y los ingresos. Al sustituir los totales de la tabla 2.7 en las ecuaciones 2.12 y 2.13, se calcula la ecuación de una línea recta que se ajuste mejor a los puntos:

$$b_1 = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} = \frac{5(4,266.1) - (133)(157.7)}{5(3,643) - (133)^2} = \frac{356.4}{526} = .678$$

$$b_0 = \frac{\Sigma Y}{n} - \frac{b_1 \Sigma X}{n} = \frac{157.7}{5} - \frac{(.678)(133)}{5} = 31.540 - 18.035 = 13.505$$

La línea que mejor se ajusta a los datos,  $\hat{Y} = 13.505 + .678X$  se muestra en la figura 2.15.

**FIGURA 2.15 Línea recta ajustada para el ejemplo 2.11**



Los administradores pueden emplear de forma rentable la ecuación calculada en el ejemplo 2.11, junto con otros valores que pueden calcularse a partir de los datos de la muestra, para pronosticar valores futuros de alguna variable importante y para evaluar por anticipado qué tan precisos podrían ser tales pronósticos. En un capítulo posterior se verá cómo extraer bastante información de los datos de muestras y cómo utilizarla para elaborar pronósticos mediante el análisis de regresión.

El coeficiente de la pendiente de mínimos cuadrados se relaciona con el coeficiente de correlación de la muestra. De forma específica,

$$b_1 = \left( \frac{\sqrt{\sum(Y - \bar{Y})^2}}{\sqrt{\sum(X - \bar{X})^2}} \right) r \quad (2.14)$$

Así,  $b_1$  y  $r$  son proporcionales una a otra y tienen el mismo signo.

#### Ejemplo 2.12

Al emplear los resultados de la tabla 2.7 y de los ejemplos 2.10 y 2.11, se puede verificar la ecuación 2.14 de forma numérica. Al usar la ecuación 2.14 se obtiene

$$b_1 = \left( \frac{\sqrt{\sum(Y - \bar{Y})^2}}{\sqrt{\sum(X - \bar{X})^2}} \right) r = \frac{\sqrt{61.192}}{\sqrt{105.2}} \cdot .89 = (.763)(.89) = .679$$

Este valor, dentro del error redondeado, concuerda con el valor del coeficiente de la pendiente que se ha calculado de manera directa en el ejemplo 2.11 mediante la ecuación 2.12.

#### Ejemplo 2.13

Suponga que el CEO (del inglés *chief executive officer*, presidente) de una gran empresa constructora sospecha que los gastos estimados de los proyectos de construcción de su compañía no coinciden del todo con los gastos reales al terminar los proyectos. Los datos que se muestran en la tabla 2.8 se recopilaron durante los últimos proyectos a fin de analizar la relación entre los costos estimados y los reales. Puesto que se dispone de Minitab, los datos se analizan con este programa.

La figura 2.16 muestra que la correlación entre los costos de construcción estimados y reales, con base en los datos de la muestra, es  $r = 0.912$ . El CEO de la compañía se sorprende al saber lo alto que es.

La figura 2.17 muestra los datos graficados como un diagrama de dispersión, así como la línea que mejor se ajusta a estos datos,  $\hat{Y} = 0.683 + 0.922X$ . Ahora, el CEO puede hacer un pronóstico de un costo de construcción real ( $Y$ ) después de que se prepara el monto estimado ( $X$ ) para el proyecto.

## EVALUACIÓN DE LA NORMALIDAD

---

Muchas técnicas estadísticas, entre ellas algunas de las que se utilizan en los pronósticos, requieren el supuesto de que un conjunto de datos sigue una distribución normal. Por

**FIGURA 2.15 Resultados de regresión con Minitab para el ejemplo 2.13**

Pearson correlation of Actual and Estimate = 0.912

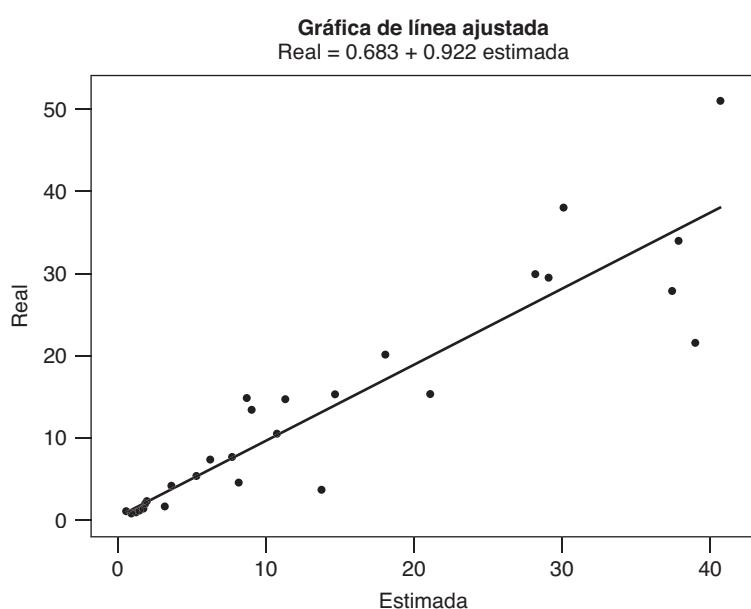
#### Regression Analysis: Actual versus Estimate

The regression equation is  
Actual = 0.683 + 0.922 Estimate

**TABLA 2.8** Costos estimados y reales de construcción de proyectos

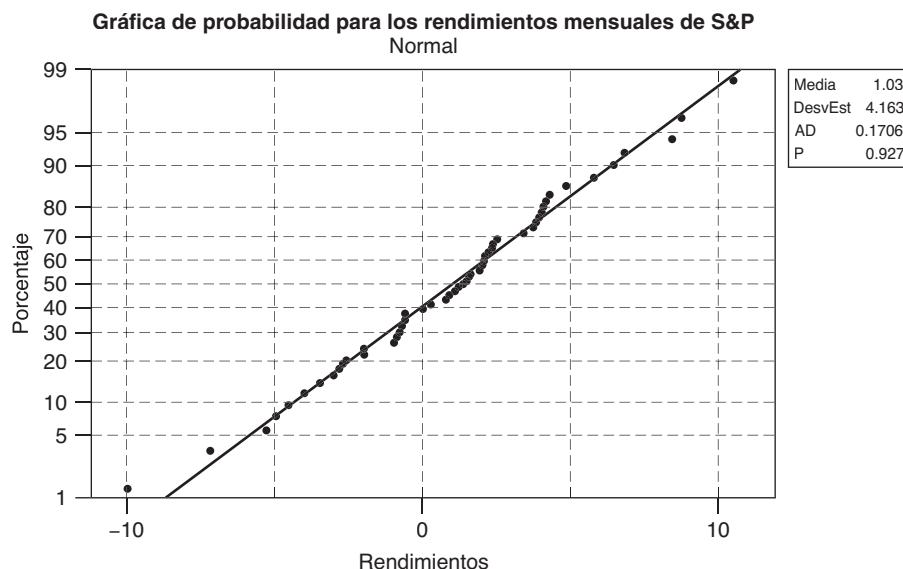
Fila	Real	Estimado
1	0.918	0.575
2	7.214	6.127
3	14.577	11.215
4	30.028	28.195
5	38.173	30.100
6	15.320	21.091
7	14.837	8.659
8	51.284	40.630
9	34.100	37.800
10	2.003	1.803
11	20.099	18.048
12	4.324	8.102
13	10.523	10.730
14	13.371	8.947
15	1.553	3.157
16	4.069	3.540
17	27.973	37.400
18	7.642	7.650
19	3.692	13.700
20	29.522	29.003
21	15.317	14.639
22	5.292	5.292
23	0.707	0.960
24	1.246	1.240
25	1.143	1.419
26	21.571	38.936

**FIGURA 2.17** Gráfica de línea ajustada para el ejemplo 2.13



**TABLA 2.9 Tasas mensuales de rendimiento (%) sobre el índice Standard & Poor 500**

	<i>Año</i>			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
Ene.	—	6.87	-7.13	4.07
Feb.	4.10	-2.94	0.85	6.51
Mar.	-3.39	2.06	2.40	2.20
Abr.	0.94	4.89	-2.73	0.03
May.	0.32	3.45	8.80	3.79
Jun.	4.23	-0.80	-0.89	-4.91
Jul.	-0.54	8.47	-0.52	4.39
Ago.	-3.94	1.54	-9.91	1.95
Sep.	3.90	-0.66	-5.25	-1.93
Oct.	2.56	-2.55	-0.67	1.18
Nov.	-1.91	1.64	5.82	-4.49
Dic.	1.46	2.12	2.45	10.58



**FIGURA 2.18 Gráfica normal de probabilidades para las tasas mensuales de rendimiento**

esta razón, se han desarrollado técnicas estadísticas que prueban las hipótesis de que una colección de datos de muestras se obtuvo de una población distribuida normalmente.

Considere las tasas de rendimiento mensuales del mercado de valores dados por el índice de S&P 500 que se muestran en la tabla 2.9. ¿Se puede suponer que estos datos siguen una distribución normal? La respuesta, producida por Minitab, se encuentra en la figura 2.18.

La línea recta que se muestra en la figura 2.18 muestra cómo se verían los puntos de una curva normal perfecta si se graficaran en esta escala especial.<sup>4</sup> Como se muestra, los datos de la tabla 2.9 se aproximan mucho a esta línea, lo cual sugiere un buen ajuste entre los datos de S&P y una distribución normal.

La prueba de normalidad predeterminada de Minitab es la prueba Anderson-Darling, cuyos resultados se muestran en la figura 2.18. Los detalles de esta prueba no nos conciernen en este momento, pero observe el valor  $p$  (etiquetado como  $P$  en el recuadro superior derecho de la figura) de 0.927. En este caso, el valor  $p$  tan alto sugiere que la muestra de datos de S&P es coherente con la hipótesis nula (lo cual ciertamente no es raro, si la hipótesis nula es verdadera). Rechazar la hipótesis nula de normalidad provocaría un error casi seguro. Por lo tanto, no debe rechazarse la hipótesis nula y puede hacerse con seguridad el supuesto de que los datos de S&P siguen una distribución normal.

## APLICACIÓN A LA ADMINISTRACIÓN

---

Muchos de los conceptos en este capítulo de repaso podrían considerarse material precursor, necesario para comprender las técnicas de pronósticos más avanzadas que se encuentran en el resto del libro. Sin embargo, los conceptos de este capítulo también tienen valor propio en muchas aplicaciones estadísticas. Aunque algunas de esas aplicaciones no cayeran lógicamente bajo el encabezado de “pronósticos”, de cualquier modo implican el uso de datos recopilados para responder a incertidumbres en la operación de negocios, especialmente en los resultados inciertos del futuro.

Los procedimientos de la estadística descriptiva que se mencionaron al principio del capítulo se utilizan ampliamente dondequiera que deban describirse, de manera convincente, grandes cantidades de datos a fin de poder utilizarlos en la toma de decisiones. Sería casi imposible pensar en una sola área que implicara medidas numéricas en las que las colecciones de datos no se resumieran de forma rutinaria el uso de estadísticas descriptivas. Este hecho se aplica en particular a la media, comúnmente señalada como “promedio”, y —en menor medida— a la desviación estándar. Los promedios son cosa corriente y se han utilizado durante muchos años para proporcionar medidas centrales de arreglos de datos. El énfasis reciente en la calidad requiere entender la variación; en consecuencia, las medidas de dispersión, como la desviación estándar, aparecen con mayor frecuencia en la práctica de los negocios.

Las distribuciones binomial y normal son buenos ejemplos de distribuciones teóricas que sirven como modelos de muchas situaciones de la vida real. Como tales, su uso es común en muchas aplicaciones, como en los pronósticos.

Las estimaciones y las pruebas de hipótesis son los dos pilares de la inferencia estadística básica. Los pronósticos o estimaciones de valores interesantes de una población, a partir de una muestra aleatoria, se emplean de manera rutinaria siempre que las restricciones de tiempo y costo no permitan un examen de todos los artículos bajo estudio. El muestreo está muy difundido en las auditorías. Las pruebas de hipótesis se usan ampliamente para comparar valores de población de un periodo anterior con valores presentes, a fin de contrastar los valores de parámetros asociados con, por ejemplo, dos sucursales o ubicaciones diferentes de un negocio, y para detectar cambios en parámetros fundamentales de procesos de producción con el propósito de mejorar el producto.

La correlación se utiliza ampliamente para examinar las relaciones entre pares de variables numéricas. Como se verá en capítulos posteriores, estas relaciones son de gran importancia para pronosticar, ya que frecuentemente son necesarios los valores de variables relevantes relacionadas para pronosticar una variable. Tanto el análisis de regresión simple como el múltiple son técnicas de pronósticos que entran en esta categoría.

<sup>4</sup> Existen otras gráficas de probabilidad normal. Una de ellas, llamada *gráfica de resultados normales*, se utiliza con frecuencia. En todas estas gráficas, se indica la normalidad si los datos graficados caen cerca de una línea recta.

## Glosario

**Coeficiente de correlación.** Mide la magnitud de la relación lineal, de dos variables entre sí.

**Diagrama de dispersión.** Trazo de puntos de datos  $X$ - $Y$  en una gráfica bidimensional.

**Distribución binomial.** Tipo de distribución de probabilidad discreta que describe la posibilidad de  $X$  éxitos en  $n$  pruebas de un experimento binomial.

**Distribución muestral.** El arreglo de todos los valores posibles de una estadística de muestra que se puede obtener de una población para un tamaño determinado de muestra.

**Distribución normal.** Tiene forma de campana y está determinada completamente por su media y desviación estándar.

**Estimación puntual.** Estimación de un parámetro de la población mediante un valor único.

**Grados de libertad.** Se refieren al número de datos independientes entre sí que llevan piezas únicas de información.

**Intervalo estimado.** Intervalo numérico dentro del cual es probable que se encuentre un parámetro de la población.

**Método de mínimos cuadrados.** Se utiliza para calcular la ecuación de una línea recta de forma que se minimice la suma de las distancias al cuadrado entre los puntos de datos  $X$ - $Y$  y la línea, medidos en la dirección vertical ( $Y$ ).

**Valor esperado.** El valor medio de la variable a lo largo de muchas pruebas u observaciones es valor esperado de una variable aleatoria.

**Valor  $p$ .** También llamado probabilidad de significancia, es la probabilidad de obtener un resultado de muestra por lo menos tan extremo como el que se observa realmente si  $H_0$  es verdadera.

**Variable aleatoria continua.** Aquella que puede tomar cualquier valor dentro de un intervalo especificado. Estos resultados se representan numéricamente mediante un intervalo de valores.

**Variable aleatoria discreta.** Tipo de variable que sólo puede tomar ciertos valores de un grupo predeterminado. A menudo, estos resultados se representan con números enteros.

## Fórmulas clave

### Media de la muestra

$$\bar{X} = \frac{\Sigma X}{n} \quad (2.1)$$

### Desviación estándar de la muestra

$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}} \quad (2.2)$$

### Varianza de la muestra

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1} \quad (2.3)$$

### Valor esperado

$$E(X) = \Sigma[X \times P(X)] \quad (2.4)$$

**Distribución de probabilidad binomial**

$$P(X) = \binom{n}{X} p^X (1-p)^{n-X} \quad \text{para } X = 0, 1, \dots, n \quad (2.5)$$

**Z como número de varianzas**

$$Z = \frac{X - \mu}{\sigma} \quad (2.6)$$

**Estadística de prueba *t* (distribución *t* de Student)**

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (2.7)$$

**Intervalo de confianza para la media de la población (muestra grande)**

$$\bar{X} \pm Z \frac{S}{\sqrt{n}} \quad (2.9)$$

**Coeficiente de correlación de la muestra**

$$\begin{aligned} r &= \frac{1}{n-1} \sum Z_X Z_Y = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2} \sqrt{\sum (Y - \bar{Y})^2}} \\ &= \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \end{aligned} \quad (2.10)$$

El método de mínimos cuadrados determina  $b_0$  y  $b_1$  que minimizan esta expresión para

$$SSE = \sum (Y - \hat{Y})^2 = \sum (Y - b_0 - b_1 X)^2 \quad (2.11)$$

**Ecuación de la pendiente de la recta de regresión ajustada**

$$b_1 = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} \quad (2.12)$$

**Ecuación de la intersección con el eje *Y* de la recta de la regresión ajustada**

$$b_0 = \bar{Y} - b_1 \bar{X} = \frac{\sum Y}{n} - \frac{b_1 \sum X}{n} \quad (2.13)$$

**Ecuación alternativa de la pendiente de la recta de regresión ajustada**

$$b_1 = \left( \frac{\sqrt{\sum (Y - \bar{Y})^2}}{\sqrt{\sum (X - \bar{X})^2}} \right) r \quad (2.14)$$

## Problemas

1. Dick Hoover, propietario de Modern Office Equipment, está preocupado por los costos de transporte y oficina en los que se incurre con los pedidos pequeños. En un esfuerzo por reducir los gastos en esta área, ha decidido introducir una política de descuentos que recompense los pedidos de más de 40 dólares, con la esperanza de que esto provoque que los clientes consoliden varias órdenes pequeñas en otras más grandes. Los siguientes datos muestran las cantidades por transacción para una muestra de 28 clientes:

10, 15, 20, 25, 17, 41, 50, 5, 9,  
 12, 14, 35, 18, 19, 17, 28, 29, 11, 11,  
 43, 54, 7, 8, 16, 13, 37, 18

- a) Calcule la media de la muestra.
  - b) Calcule la desviación estándar de la muestra.
  - c) Calcule la varianza de la muestra.
  - d) Si tiene éxito la política, ¿aumentará, disminuirá o permanecerá igual la media de la distribución?
  - e) Si la política es exitosa, ¿aumentará, disminuirá o permanecerá igual la desviación estándar de la distribución?
  - f) Con los datos anteriores, pronostique la cantidad para el pedido siguiente de un cliente.
2. Sandy James cree que los precios de las viviendas se han estabilizado en los últimos meses. Para convencer a su jefe, planea comparar los precios actuales con los del año pasado. Entonces, recopila 12 precios de viviendas de los anuncios y enseguida calcula la media y la desviación estándar de los precios que ha encontrado. ¿Cuáles son estos dos valores?

125,900 253,000 207,500 146,950 121,450 135,450  
 175,000 200,000 210,950 166,700 185,000 191,950

3. Una gran compañía constructora intentaba establecer una manera conveniente de ver las utilidades típicas de trabajos obtenidos mediante licitación competitiva. Ya que los trabajos varían sustancialmente tanto en el tamaño como en la cantidad total de la subasta exitosa, la compañía decidió expresar las utilidades como ganancias porcentuales:

$$\text{Ganancias porcentuales} = 100 \times \frac{\text{Ganancias}}{\text{Costos reales de construcción}}$$

Cuando se pierde dinero en un proyecto, las ganancias son negativas, de la misma manera que la utilidad neta resultante. Una muestra de 30 trabajos arrojó las ganancias porcentuales:

15.9	21.3	-1.8	6.6	.4	53.6	19.7	-.5	6.7	-2.3
11.9	-.3	19.0	12.8	-9.6	26.8	21.0	32.0	-.4	10.9
6.9	-8.5	3.5	3.5	-1.9	4.0	13.0	15.1	9.7	33.9

- a) ¿Cuál es un valor factible para las ganancias porcentuales medias de la población de trabajos al incluir todos los trabajos potenciales?
- b) Construya un intervalo de confianza de 95% para las ganancias porcentuales medias de la población de trabajos mediante un argumento de muestra grande.
- c) Levante un intervalo de confianza de 95% para las ganancias porcentuales medias de la población de trabajos suponiendo que 30 es una muestra pequeña. ¿Qué supuesto adicional se necesita hacer en este caso?
- d) Compare los dos intervalos en los incisos b) y c). Explique por qué a menudo una muestra de tamaño 30 se considera el límite entre las muestras grandes y pequeñas.

4. A partir de los datos en una muestra grande de transacciones de ventas, el propietario de un pequeño negocio reporta que un intervalo de confianza de 95% para la utilidad media por transacción,  $\mu$ , es de (23.41, 102.59). Utilice estos datos para determinar:
  - a) Una estimación puntual (mejor conjetura) de la media,  $\mu$ , y su margen de error de 95%.
  - b) Un intervalo de confianza de 90% para la media,  $\mu$ .
5. Se desea predecir si el número anual medio de días de ausencias ha aumentado en un conjunto grande de trabajadores. Hace un año se sabía que la media era 12.1. Una muestra reciente de 100 empleados reveló una media de la muestra de 13.5 con la desviación estándar de la muestra de 1.7 días. Utilice el nivel de significancia 0.05 para determinar si la media de la población ha aumentado, o si la diferencia entre 13.5 y 12.1 solamente representa un error muestral.
6. New Horizons Airlines quiere predecir el número medio de asientos vacíos por vuelo a Alemania en el próximo año. Para desarrollar este pronóstico, de los archivos del año pasado se eligen de manera aleatoria los registros de 49 vuelos, y se anota el número de asientos libres en cada vuelo. La media de la muestra es de 8.1 asientos y la desviación estándar, de 5.7 asientos. Desarrolle una estimación puntual y un intervalo de confianza de 95% para el número medio de asientos libres por vuelo durante el año pasado. Para el próximo año pronostique el número medio de asientos libres por vuelo a Alemania. Analice la precisión de este pronóstico.
7. A lo largo de varios años, una crema dental ha recibido una calificación media de 5.9 en una escala de 7 puntos en lo que se refiere a la satisfacción general del cliente con el producto. Debido a un pequeño cambio en el dentífrico que no se ha anunciado, existe cierta preocupación de que la satisfacción del cliente haya cambiado. Suponga que las calificaciones de satisfacción de una muestra de 60 clientes tienen una media de 5.60 y una desviación estándar de 0.87. ¿Indican estos datos que la calificación media de satisfacción es diferente a 5.9? Pruebe con  $\alpha = 0.05$ . ¿Cuál es el valor  $p$  para la prueba?
8. La gerente de una tienda de yogur congelado alega que una ración mediana contiene un promedio de más de cuatro onzas del producto. A partir de una muestra aleatoria de 14 raciones, se obtiene una media de 4.31 onzas y una desviación estándar de 0.52 onzas. Pruebe con  $\alpha = 0.05$ , lo que la gerente dice que es cierto. Suponga que la distribución de peso por ración es normal.
9. Con base en experiencias pasadas, la California Power Company predice que el uso residencial medio de electricidad por hogar será de 700 kwh en enero próximo. En ese mes la compañía elige una muestra aleatoria simple de 50 hogares y calcula una media y una desviación estándar de 715 y 50, respectivamente. Utilice el nivel de significancia 0.05 para determinar si el pronóstico de California Power es razonable. Calcule e interprete el valor  $p$  para la prueba.
10. Los expertos en población indican que el tamaño de las familias ha disminuido en los últimos años. Hace diez años, el promedio de una familia era de 2.9 elementos. Considere la población de 200 tamaños de familias que ofrece la tabla P.10. Elija al azar una muestra de 30 de esos datos y pruebe la hipótesis de que el tamaño de la familia promedio no ha cambiado en los últimos diez años.
11. James Dobbins, supervisor de mantenimiento de Atlanta Transit Authority, desea determinar si existe una relación positiva entre el costo anual de mantenimiento de un autobús y su edad. Si existe tal relación, James cree que puede predecir mejor el presupuesto anual de mantenimiento de los autobuses. El supervisor recolecta los datos que se muestran en la tabla P.11.

**TABLA P.10**

(1) 3	(35) 1	(69) 2	(102) 1	(135) 5	(168) 6
(2) 2	(36) 2	(70) 4	(103) 2	(136) 2	(169) 3
(3) 7	(37) 4	(71) 3	(104) 5	(137) 1	(170) 2
(4) 3	(38) 1	(72) 7	(105) 3	(138) 4	(171) 3
(5) 4	(39) 4	(73) 2	(106) 2	(139) 2	(172) 4
(6) 2	(40) 2	(74) 6	(107) 1	(140) 4	(173) 2
(7) 3	(41) 1	(75) 2	(108) 2	(141) 1	(174) 2
(8) 1	(42) 3	(76) 7	(109) 2	(142) 2	(175) 1
(9) 5	(43) 5	(77) 3	(110) 1	(143) 4	(176) 5
(10) 3	(44) 2	(78) 6	(111) 4	(144) 1	(177) 3
(11) 2	(45) 1	(79) 4	(112) 1	(145) 2	(178) 2
(12) 3	(46) 4	(80) 2	(113) 1	(146) 2	(179) 4
(13) 4	(47) 3	(81) 3	(114) 2	(147) 5	(180) 3
(14) 1	(48) 5	(82) 5	(115) 2	(148) 3	(181) 5
(15) 2	(49) 2	(83) 2	(116) 1	(149) 1	(182) 3
(16) 2	(50) 4	(84) 1	(117) 4	(150) 2	(183) 1
(17) 4	(51) 1	(85) 3	(118) 2	(151) 6	(184) 2
(18) 4	(52) 6	(86) 3	(119) 1	(152) 2	(185) 4
(19) 3	(53) 2	(87) 2	(120) 3	(153) 5	(186) 3
(20) 2	(54) 5	(88) 4	(121) 5	(154) 1	(187) 2
(21) 1	(55) 4	(89) 1	(122) 1	(155) 2	(188) 5
(22) 5	(56) 1	(90) 2	(123) 2	(156) 1	(189) 3
(23) 2	(57) 2	(91) 3	(124) 3	(157) 4	(190) 4
(24) 1	(58) 1	(92) 3	(125) 4	(158) 2	(191) 3
(25) 4	(59) 5	(93) 2	(126) 3	(159) 2	(192) 2
(26) 3	(60) 2	(94) 4	(127) 2	(160) 7	(193) 3
(27) 2	(61) 7	(95) 1	(128) 1	(161) 4	(194) 2
(28) 3	(62) 1	(96) 2	(129) 6	(162) 2	(195) 5
(29) 6	(63) 2	(97) 4	(130) 1	(163) 1	(196) 3
(30) 1	(64) 6	(98) 3	(131) 2	(164) 7	(197) 3
(31) 2	(65) 4	(99) 2	(132) 5	(165) 2	(198) 2
(32) 4	(66) 1	(100) 6	(133) 2	(166) 7	(199) 5
(33) 3	(67) 2	(101) 4	(134) 1	(167) 4	(200) 1
(34) 2	(68) 1				

**TABLA P.11**

<i>Autobús</i>	<i>Costo de man-</i> <i>tenimiento (\$)</i>	<i>Edad</i> <i>(años)</i>
	<i>Y</i>	<i>X</i>
1	859	8
2	682	5
3	471	3
4	708	9
5	1,094	11
6	224	2
7	320	1
8	651	8
9	1,049	12

**TABLA P.12**

<i>Semana</i>	<i>Libros vendidos</i> <i>Y</i>	<i>Espacio de anaquel</i> <i>X</i>
1	275	6.8
2	142	3.3
3	168	4.1
4	197	4.2
5	215	4.8
6	188	3.9
7	241	4.9
8	295	7.7
9	125	3.1
10	266	5.9
11	200	5.0

- a) Grafique un diagrama de dispersión.  
 b) ¿Qué tipo de relación existe entre estas dos variables?  
 c) Calcule el coeficiente de correlación.
12. Anna Sheehan es gerente de la cadena de supermercados Spendwise. Desea poder predecir las ventas de libros de pasta rústica (ejemplares por semana) con base en la cantidad proporcionada de espacio de exhibición en los anaqueles (pies). Anna recopila los datos para una muestra de 11 semanas, como se muestra en la tabla P.12.
- a) Grafique un diagrama de dispersión.  
 b) ¿Qué tipo de relación existe entre estas dos variables?  
 c) Calcule el coeficiente de correlación. Utilice el método de los mínimos cuadrados para determinar la pendiente y la intersección con el eje Y de la ecuación de la recta. Use esta ecuación para predecir el número de libros vendidos si se usan 5.2 pies de espacio de anaquel (es decir,  $X = 5.2$ ).
13. Considere una población que consiste de las 200 observaciones semanales que se presentan en la tabla P.13. La variable independiente  $X$  es la temperatura promedio semanal de Spokane, Washington. La variable dependiente  $Y$  es el número de acciones de la compañía Sunshine Mining con las que se hicieron operaciones en la bolsa de valores de Spokane en una semana. Elija datos de forma aleatoria para 16 semanas y calcule el coeficiente de correlación. (Sugerencia: asegúrese de que su muestra se tome de forma aleatoria de entre la población.) Luego, determine la recta de mínimos cuadrados y elabore un pronóstico de  $Y$  para una temperatura promedio semanal de 63.
14. Un inversionista en bienes raíces recopila los siguientes datos de una muestra aleatoria de apartamentos en el lado oeste de College Station, Texas.

Renta (dólares por mes)	Tamaño (pies <sup>2</sup> )	Renta (dólares por mes)	Tamaño (pies <sup>2</sup> )
720	1,000	650	800
595	900	748	960
915	1,200	685	650
760	810	755	970
1,000	1,210	815	1,000
790	860	745	1,000
880	1,135	715	1,000
845	960	885	1,180

TABLA P.13

<i>OBS.</i>	<i>Y</i>	<i>X</i>									
(1)	50	37	(51)	54	86	(101)	22	43	(151)	79	85
(2)	90	77	(52)	76	48	(102)	32	5	(152)	79	27
(3)	46	55	(53)	55	48	(103)	24	13	(153)	48	61
(4)	47	27	(54)	12	15	(104)	63	3	(154)	5	7
(5)	12	49	(55)	5	70	(105)	16	58	(155)	24	79
(6)	23	23	(56)	2	9	(106)	4	13	(156)	47	49
(7)	65	18	(57)	77	52	(107)	79	18	(157)	65	71
(8)	37	1	(58)	6	71	(108)	5	5	(158)	56	27
(9)	87	41	(59)	67	38	(109)	59	26	(159)	52	15
(10)	83	73	(60)	30	69	(110)	99	9	(160)	17	88
(11)	87	61	(61)	3	13	(111)	76	96	(161)	45	38
(12)	39	85	(62)	6	63	(112)	15	94	(162)	45	31
(13)	28	16	(63)	70	65	(113)	10	30	(163)	90	35
(14)	97	46	(64)	33	87	(114)	20	41	(164)	69	78
(15)	69	88	(65)	13	18	(115)	37	1	(165)	62	93
(16)	87	87	(66)	10	4	(116)	56	27	(166)	0	51
(17)	52	82	(67)	21	29	(117)	6	73	(167)	8	68
(18)	52	56	(68)	56	21	(118)	86	19	(168)	47	30
(19)	15	22	(69)	74	9	(119)	27	94	(169)	7	81
(20)	85	49	(70)	47	8	(120)	67	5	(170)	48	30
(21)	41	44	(71)	34	18	(121)	22	31	(171)	59	46
(22)	82	33	(72)	38	84	(122)	32	13	(172)	76	99
(23)	98	77	(73)	75	64	(123)	90	11	(173)	54	98
(24)	99	87	(74)	0	81	(124)	88	50	(174)	95	11
(25)	23	54	(75)	51	98	(125)	35	40	(175)	7	6
(26)	77	8	(76)	47	55	(126)	57	80	(176)	24	83
(27)	42	64	(77)	63	40	(127)	73	44	(177)	55	49
(28)	60	24	(78)	7	14	(128)	13	63	(178)	41	39
(29)	22	29	(79)	6	11	(129)	18	74	(179)	14	16
(30)	91	40	(80)	68	42	(130)	70	40	(180)	24	13
(31)	68	35	(81)	72	43	(131)	9	53	(181)	36	31
(32)	36	37	(82)	95	73	(132)	93	79	(182)	62	44
(33)	22	28	(83)	82	45	(133)	41	9	(183)	77	11
(34)	92	56	(84)	91	16	(134)	17	52	(184)	32	60
(35)	34	33	(85)	83	21	(135)	10	82	(185)	12	82
(36)	34	82	(86)	27	85	(136)	69	37	(186)	85	7
(37)	63	89	(87)	13	37	(137)	5	57	(187)	90	68
(38)	30	78	(88)	6	89	(138)	18	62	(188)	78	10
(39)	31	24	(89)	76	76	(139)	88	21	(189)	60	27
(40)	84	53	(90)	55	71	(140)	99	94	(190)	96	90
(41)	56	61	(91)	13	53	(141)	86	99	(191)	51	6
(42)	48	18	(92)	50	13	(142)	95	45	(192)	9	62
(43)	0	45	(93)	60	12	(143)	78	19	(193)	93	78
(44)	58	4	(94)	61	30	(144)	3	76	(194)	61	22
(45)	27	23	(95)	73	57	(145)	38	81	(195)	5	99
(46)	78	68	(96)	20	66	(146)	57	95	(196)	88	51
(47)	78	79	(97)	36	27	(147)	77	30	(197)	45	44
(48)	72	66	(98)	85	41	(148)	25	59	(198)	34	86
(49)	21	80	(99)	49	20	(149)	99	93	(199)	28	47
(50)	73	99	(100)	83	66	(150)	9	28	(200)	44	49

- a) Grafique los datos como un diagrama de dispersión en donde  $Y$  = renta y  $X$  = tamaño.
- b) Determine la ecuación de la recta de mínimos cuadrados que relacione la renta con el tamaño.
- c) ¿Cuál es el aumento estimado en la renta por pie cuadrado adicional de espacio?
- d) Pronostique la renta mensual para un apartamento de 750 pies cuadrados.
15. La firma Abbot & Sons necesita predecir la edad media  $\mu$  de sus trabajadores contratados por hora. Se toma una muestra aleatoria de los archivos de personal que arroja los resultados que se encuentran a continuación. Prepare tanto una estimación puntual como un intervalo de confianza de 98% para la edad media de los trabajadores. Pruebe la hipótesis  $H_0: \mu = 44$  contra  $H_1: \mu \neq 44$  en el nivel de 2%. ¿Son coherentes los resultados de la prueba de la hipótesis con el intervalo de confianza para  $\mu$ ? ¿Esperaría que así fueran?

$$\bar{X} = 45.2 \quad S = 10.3 \quad n = 175$$

## CASOS

---

### CASO 2-1 ALCOM ELECTRONICS

Recientemente, Jerrick Tilby recibió un título en administración de negocios de una pequeña universidad y comenzó a trabajar en Alcom Electronics, un fabricante de diversos componentes electrónicos para la industria. Después de algunas semanas de trabajar allí, fue llamado a la oficina de la dueña y gerente general de Alcom, McKennah Labrum, quien le pidió que investigara un asunto concerniente a cierto transistor que la compañía fabricaba debido a que una empresa grande de televisión estaba interesada en realizar una compra importante.

McKennah quería pronosticar la vida promedio de los transistores, asunto que era una gran preocupación para la compañía de televisión. Las unidades que actualmente están en existencia podrían ser representativas de las que se fabricarían durante la vigencia del nuevo contrato, si es que éste fuese aceptado.

Jerrick decidió tomar una muestra aleatoria de los transistores en cuestión y formuló un plan para lograr esta labor. Marcó numéricamente los contenedores de almacenamiento de los transistores, les puso números aleatorios y muestreó todos los transistores en cada uno de los contenedores seleccionados para la muestra. Debido a que cada contenedor tenía cerca de 20 transistores, Jerrick seleccionó 10 números aleatorios que le proporcionaron una muestra final de 205 transistores. Debido a que había elegido 10 de los 55 contenedores, pensó que tenía una muestra representativa y que podría utilizar los resultados de esta muestra para generalizar la población completa de los transistores en el inventario, así como las unidades que aún habrían de ser producidas mediante el mismo proceso.

Entonces, Jerrick consideró la cuestión de la vida promedio de las unidades. Puesto que la vida de los tran-

sistores puede extenderse por varios años, se dio cuenta de que ninguna de las unidades muestreadas podría probarse si se deseaba obtener una respuesta oportuna. Por lo tanto, decidió contactar a varios usuarios de este componente para determinar si había registros concernientes al tiempo de duración del transistor. Afortunadamente, encontró tres compañías que ya habían utilizado el transistor y que habían hecho registros acerca de la vida de los componentes. En total, recibió datos de 38 transistores cuyos tiempos de falla eran conocidos. Ya que estos transistores fueron fabricados mediante el mismo proceso que se utiliza actualmente, Jerrick pensó que los resultados de esta muestra podrían utilizarse para inferir las condiciones de las unidades que estaban en inventario y las que aún habrían de producirse.

A continuación se presentan los resultados de los cálculos que Jerrick llevó a cabo con los datos del tiempo de vida de esta muestra:

$$n = 38$$

vida promedio  $\bar{X} = 4\,805$  horas  
desviación estándar de los tiempos de vida  $S = 675$  horas

Después de observar que la vida promedio de la muestra era sólo de 4 805 horas, Jerrick se preocupó porque sabía que los otros proveedores de los componentes estaban garantizando un promedio de vida de 5000 horas. A pesar de que el promedio de la muestra estaba por debajo de las 5000 horas, se dio cuenta de que el tamaño de la muestra era únicamente de 38 y que esto no constituía una prueba contundente de que la calidad de Alcom fuera inferior a la de los otros proveedores.

Decidió probar la hipótesis de que la vida promedio de todos los transistores de Alcom era de 5000 horas,

en comparación con la alternativa de que fuera inferior. A continuación se muestran los cálculos que realizó Jerrick utilizando  $\alpha = 0.01$ :

$$\begin{aligned} H_0: \mu &= 5000 \\ H_1: \mu &< 5000 \end{aligned}$$

Si  $S = 675$ , entonces

**Punto de la regla de decisión:**  $5,000 - 2.33 \frac{675}{\sqrt{38}} = 4,744.9$

**Regla de decisión:** Si  $\bar{X} < 4,744.9$  rechace  $H_0$

Debido a que la media de la muestra (4 805) no estuvo por debajo del punto de la regla de decisión (4 744.9), Jerrick no rechazó la hipótesis de que la du-

ración media de todos los componentes era igual a 5000 horas. Pensó que éstas serían buenas noticias para McKennah Labrum y anexó un resumen de lo que descubrió a su informe final. Pocos días después de que entregó su informe escrito y verbal, McKennah le llamó a su oficina para felicitarle por su buen trabajo y para compartirle su preocupación relacionada con los descubrimientos. Ella le dijo: “Estoy preocupada por el nivel tan bajo de significancia de tu prueba de hipótesis. Tan sólo tomaste una posibilidad de 1% para rechazar la hipótesis nula si ésta fuera verdadera. Esto me parece una posición bastante conservadora. Me preocupa el hecho de que firmemos un contrato y después veamos que nuestro nivel de calidad no satisface las especificaciones requeridas, de 5000 horas.” ■

#### PREGUNTA

¿Qué respondería al comentario de McKennah Labrum?

## CASO 2-2 MR. TUX

John Mosby, propietario de varias tiendas de alquiler Mr. Tux, está interesado en predecir su volumen mensual de ventas (véase el caso de Mr. Tux que aparece al final del capítulo 1). Como etapa inicial, John recopila datos de ventas del periodo que va de 1989 a 1996, como se muestra en la tabla 2.10.

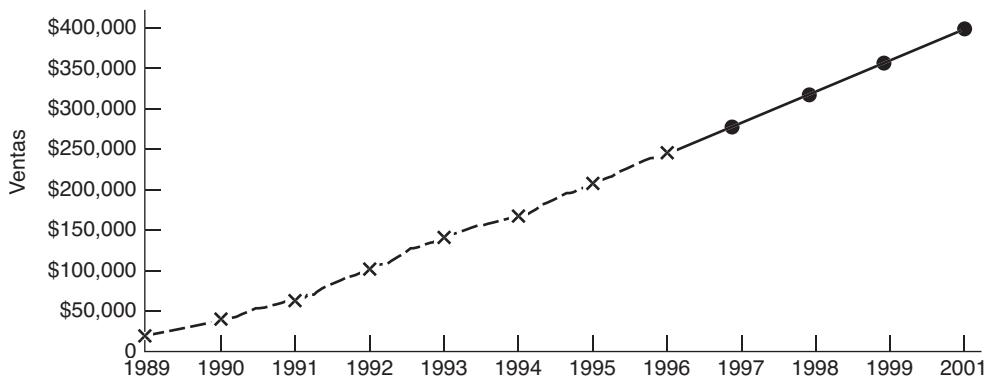
A continuación, John calcula el valor promedio de las ventas mensuales de cada año (por ejemplo, suma

los 12 valores de 1989 y los divide entre 12). También calcula la desviación estándar de los 12 valores mensuales para cada año. Los resultados se muestran en la tabla 2.11. Asimismo, John decide hacer una gráfica de series de tiempo, que se muestra en la figura 2.19. Mosby grafica la media de los valores mensuales de venta en el eje Y y el tiempo en el eje X. ■

**TABLA 2.10** Datos de las ventas mensuales de Mr. Tux, en dólares

	1989	1990	1991	1992	1993	1994	1995	1996
Ene.	6,028	16,850	15,395	27,773	31,416	51,604	58,843	71,043
Feb.	5,927	12,753	30,826	36,653	48,341	80,366	82,386	152,930
Mar.	10,515	26,901	25,589	51,157	85,651	208,938	224,803	250,559
Abr.	32,276	61,494	103,184	217,509	242,673	263,830	354,301	409,567
May.	51,920	147,862	197,608	206,229	289,554	252,216	328,263	394,747
Jun.	31,294	57,990	68,600	110,081	164,373	219,566	313,647	272,874
Jul.	23,573	51,318	39,909	102,893	160,608	149,082	214,561	230,303
Ago.	36,465	53,599	91,368	128,857	176,096	213,888	337,192	375,402
Sep.	18,959	23,038	58,781	104,776	142,363	178,947	183,482	195,409
Oct.	13,918	41,396	59,679	111,036	114,907	133,650	144,618	173,518
Nov.	17,987	19,330	33,443	63,701	113,552	116,946	139,750	181,702
Dic.	15,294	22,707	53,719	82,657	127,042	164,154	184,546	258,713

<b>TABLA 2.11 Valores promedio de las ventas mensuales de Mr. Tux</b>		
<i>Año</i>	<i>Promedio (media)</i> \$	<i>Desviación estándar</i> \$
1989	22,013	13,165
1990	44,603	35,290
1991	64,841	47,217
1992	103,610	57,197
1993	141,381	70,625
1994	169,432	63,376
1995	213,866	96,387
1996	247,231	99,153

**FIGURA 2.19 Media de las ventas mensuales de Mr. Tux****PREGUNTAS**

1. ¿Qué ideas de pronóstico le vienen a la mente cuando estudia los valores medios de las ventas mensuales que John ha obtenido para los años de sus datos?
2. Suponga que John decide trazar una línea recta en su diagrama de dispersión para que “ajuste bien” y después la extiende hacia el futuro, para lo cual usa puntos a lo largo de la línea a manera de pronósticos mensuales. ¿Qué tan precisos cree que sean estos pronósticos? Utilice los valores de la

desviación estándar que John calculó al responder esta pregunta. Con base en su análisis, ¿alentaría a John a que continuara buscando un método de pronóstico más exacto? John cuenta con la versión más reciente de Minitab. ¿Piensa usted que debería utilizar la función de análisis de regresión de Minitab para calcular una línea de mínimos cuadrados? Si así fuera, ¿cuál variable  $X$  se emplearía para pronosticar las ventas mensuales ( $Y$ )?

## CASO 2-3 ALOMEGA FOOD STORES

En el ejemplo 1.1, la presidenta de Alomega Food Stores, Julie Ruth, había recopilado datos del funcionamiento de su compañía. Obtuvo datos de varios meses de ventas junto con diversas variables explicativas que eran posibles, como el importe gastado en publicidad

(repase la situación del ejemplo 1.1). Mientras que su equipo de análisis trabajaba con los datos en un intento por pronosticar las ventas mensuales, ella se impacientó y se preguntó cuál de las posibles variables explicativas sería la mejor para este fin.

Ya que contaba con un programa estadístico en su computadora, decidió observar los datos por sí misma. En primer lugar, encontró los coeficientes de correlación entre las ventas mensuales variables y diversas variables explicativas potenciales. En lo particular, estaba interesada en la correlación existente entre las ventas y el importe de la publicidad en periódicos, el importe mensual de la publicidad de televisión, el de la publicidad en periódico de uno y dos meses antes, la publicidad en televisión de uno a dos meses antes y las calificaciones de publicidad de la competencia. Los valores de  $r$  (coeficientes de correlación) fueron:

Coeficiente de correlación ( $r$ ) entre las ventas variables y	
Importe mensual de la publicidad en periódico	0.45
Importe mensual de la publicidad en televisión	0.60
Importe mensual de la publicidad en periódico de un mes antes	-0.32
Importe mensual de la publicidad en periódico de dos meses antes	0.21
Importe mensual de la publicidad en televisión de un mes antes	-0.06
Importe mensual de la publicidad en televisión de dos meses antes	0.03
Calificación de la publicidad de la competencia	-0.18

Julie no se sorprendió al encontrar que la correlación más alta estaba entre las ventas mensuales y los importes de la publicidad por televisión ( $r = 0.60$ ); pero esperaba que la correlación fuese más fuerte. Decidió emplear una característica de regresión para calcular la ecuación de la recta de los mínimos cuadrados, utilizando las ventas como variable dependiente y la publicidad

mensual por televisión como variable explicativa. Los resultados de esta corrida fueron:

$$\begin{aligned} \text{Ventas} &= 341.663 + 0.336 \\ (\text{publicidad mensual por televisión}) \\ r\text{-cuadrada} &= 0.36 (36\%) \quad \text{valor } p = 0.000 \end{aligned}$$

Julie tuvo que buscar en su libro de estadística de la universidad para interpretar los valores que imprimió de  $r$ -cuadrada y  $p$ . Después de leer, recordó que  $r$  cuadrada (que es el cuadrado del coeficiente de correlación,  $r$ ) mide el porcentaje de variabilidad en las ventas que puede explicarse mediante la variabilidad en la publicidad televisiva (esto se explicará en el capítulo 6). También el valor  $p$  indica que el coeficiente de la pendiente (0.336) es importante; es decir, puede rechazarse la hipótesis de que es cero en una población de la que se obtuvo la muestra, casi sin que exista posibilidad de error.

La presidenta de Alomega concluyó que la ecuación de la regresión que encontró era importante y que podía emplearse para pronosticar las ventas mensuales, si se conocía el presupuesto asignado a la publicidad por televisión. Debido a que los gastos en publicidad televisiva están bajo el control de la compañía, creyó que tenía una buena forma de pronosticar las ventas mensuales. En una breve conversación con el líder del departamento de administración de datos, Roger Jackson, Julie le comentó lo que había descubierto. Él respondió: "Sí, nosotros también lo descubrimos. Sin embargo, ten en cuenta que los anuncios por televisión sólo explican aproximadamente la tercera parte de la variabilidad de las ventas. Bueno, 36%. No pensamos que sea lo suficientemente elevado y estamos intentando usar diversas variables para tratar de obtener un valor más elevado de  $r$ -cuadrada. Además, pensamos que estamos en un método que de cualquier forma hará un mejor trabajo que el de un análisis de regresión." ■

## PREGUNTAS

1. ¿Qué opina del análisis de Julie Ruth?
2. Defina los residuales (errores) como diferencias entre los valores de las ventas reales y los valores

predichos por la línea recta. ¿De qué manera examinaría los residuales para decidir si la representación de la línea recta que hizo Julie es adecuada o no?

## Aplicaciones en Minitab

**El problema.** En el ejemplo 2.1 se llevó a cabo la recopilación de edades utilizando estadística descriptiva.

### Solución con Minitab

1. Introduzca el nombre de la variable como *Ages* (edades) debajo de C1.
2. Introduzca los datos de la columna C1.
3. Dé clic en los menús que se muestran a continuación:

Stat>Basic Statistics>Graphical Summary

4. Aparece el cuadro de diálogo de la sección *Graphical Summary* (resumen gráfico).
  - a) Dé clic en *Variables* y seleccione *Ages* (C1) como se muestra en la figura 2.20.
  - b) Dé clic en *OK* en el cuadro de diálogo de *Graphical Summary* y aparecerá el resumen que se muestra en la figura 2.21.
  - c) Para imprimir esta gráfica dé clic en los siguientes menús:

File>Print Graph

5. Dé clic en el menú que se presenta a continuación:

Stat>Basic Statistics>Display Descriptive Statistics

6. Aparece el cuadro de diálogo de *Display Descriptive Statistics* (mostrar estadísticas descriptivas)
  - a) Dé clic en *Variables* y seleccione *Ages* (C1)
  - b) Dé clic en *OK* y aparecerá el resumen de la figura 2.1 en la ventana *Session* (sesión). (Para obtener estadísticas descriptivas adicionales, dé clic en *Statistics* y selecciones medidas adicionales descriptivas antes de dar clic en *OK*.)
  - c) Para imprimir la ventana de la sesión que contenga un resumen de la estadística descriptiva, dé clic en:

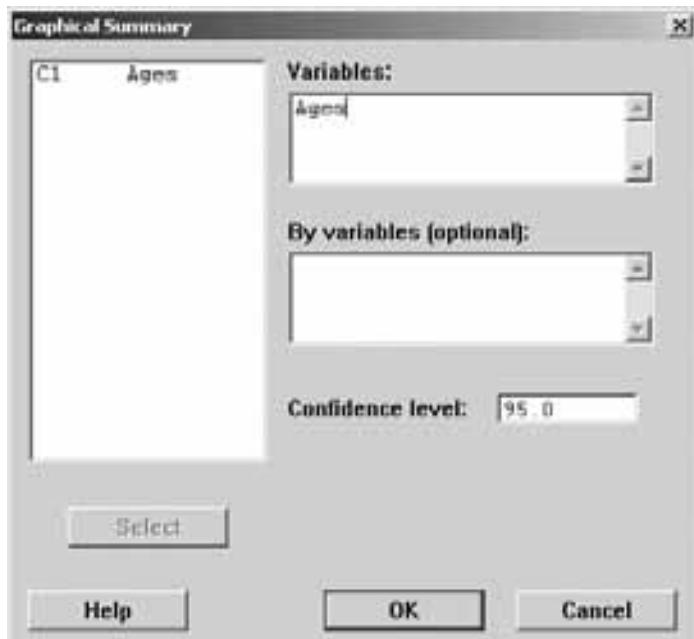
File>Print Session Window

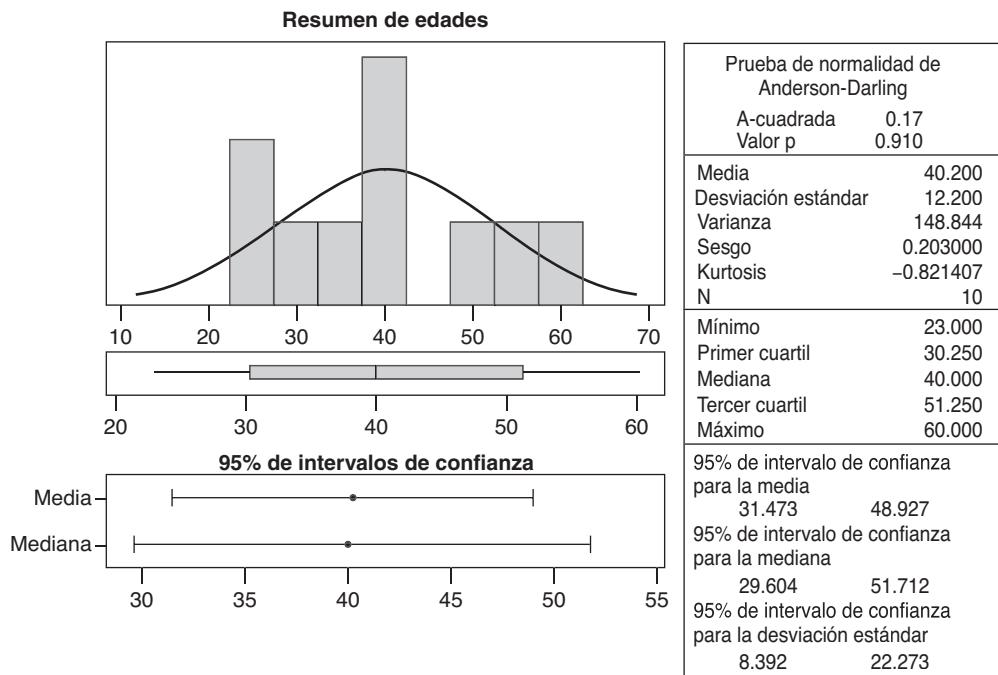
Los resultados se muestran en la figura 2.1 (vea la página 18).

7. Los siguientes comandos le permitirán desarrollar un histograma, gráfica de puntos y gráfica de caja.

Graph>Histogram  
 Graph>Dotplot  
 Graph>Boxplot

**FIGURA 2.20 Pantalla del cuadro de diálogo de Graphical Summary en Minitab**





**FIGURA 2.21 Pantalla de Graphical Summary en Minitab**

Las instrucciones acerca de cómo utilizar Minitab para correr los análisis de correlación y regresión se presentan al final del capítulo 6.

## Aplicaciones en Excel

**El problema.** En el problema 1, Dick Hoover, propietario de Modern Office Equipment, está preocupado por los costos de flete y de administración en los que se incurre cuando hay pedidos pequeños.

### Solución de Excel.

1. Se abre el programa Excel y aparece la hoja de cálculo que se muestra en la figura 2.1, al final del capítulo 1. Mueva el cursor de manera que la celda resaltada se encuentre en la esquina superior izquierda en la hoja de cálculo, la celda A1.
2. Introduzca el primer valor, 10, seguido por la tecla *return*, continúe con los datos que siguen.
3. Después de que se hayan introducido los 28 valores en la columna A, el cursor se coloca en la celda donde se desea tener los resultados del primer cálculo, A30.
4. El promedio de los datos de las celdas, desde A1 hasta A28, se calcula al colocar la fórmula en la celda A30. Para introducir la fórmula, es necesario precederla del signo =. La fórmula es =Average(A1:A28)<sup>1</sup>. Observe que A30 aparece a la izquierda de la barra de fórmulas y =Average(A1:A28) a la derecha, por arriba de la hoja de cálculo.
5. El mismo método se emplea para calcular la desviación estándar. La fórmula =Stdev(A1:A28)<sup>2</sup> se introduce en la celda A31. Los resultados se muestran en la figura 2.22.

El promedio y la desviación estándar podrían haberse calculado también al utilizar la función *Insert* (insertar) la herramienta *Data Analysis*. Estos métodos se presentarán en capítulos posteriores.

Las instrucciones sobre cómo utilizar Excel para ejecutar el análisis de correlación y de regresión se presentan al final del capítulo 6.

<sup>1</sup> PROMEDIO

<sup>2</sup> DESVEST

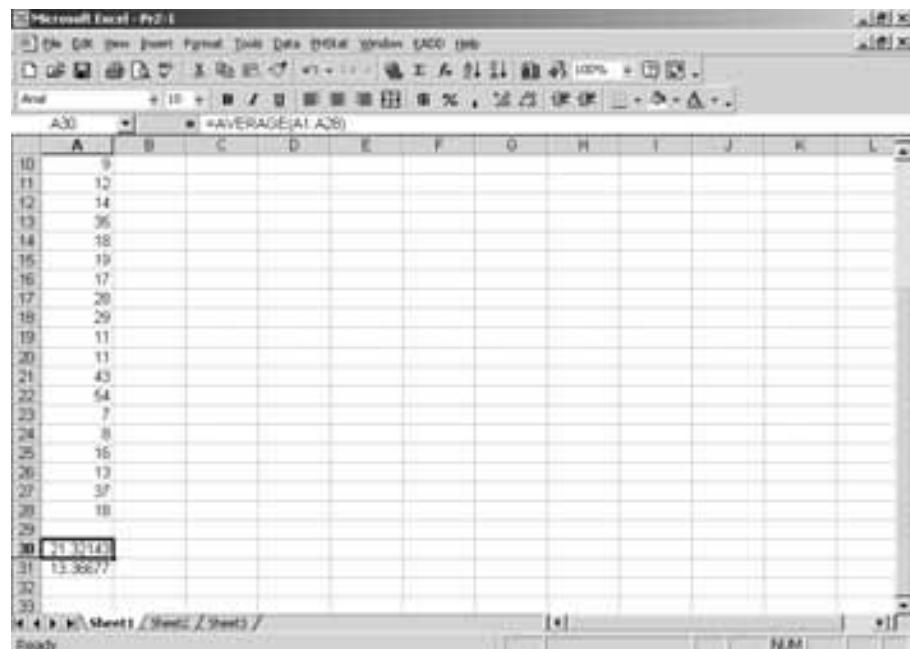


FIGURA 2.22 Hoja de cálculo de Excel para el problema 1

Si usted está empleando el programa Excel o Minitab, se le sugiere que utilice diversos grupos de datos y rutinas estadísticas para familiarizarse con estos programas tan poderosos. La habilidad que obtenga le será de gran utilidad a medida que aprenda los procedimientos de pronóstico que se presentan en este texto.

## Referencias

- Anderson, D. R., D. J. Sweeney y T. A. Williams, *Modern Business Statistics with Microsoft Excel*, Cincinnati, OH, South-Western, 2003.
- Cryer, J. D. y R. B. Miller, *Statistics for Business: Data Analysis and Modeling*, segunda edición, Belmont, CA, Duxbury Press, 1994.
- Hanke, J. y A. Reitsch, *Understanding Business Statistics*, segunda edición, Homewood, IL, Richard D. Irwin, 1994.
- Johnson, R. A. y D. W. Wichern, *Business Statistics: Decision Making with Data*, Nueva York, John Wiley & Sons, 1997.
- Keller, G. y B. Warrack, *Statistics for Management and Economics*, sexta edición, Pacific Grove, CA, Brooks/Cole, 2003.

## EXPLORACIÓN DE PATRONES DE DATOS Y ELECCIÓN DE UNA TÉCNICA DE PRONÓSTICO

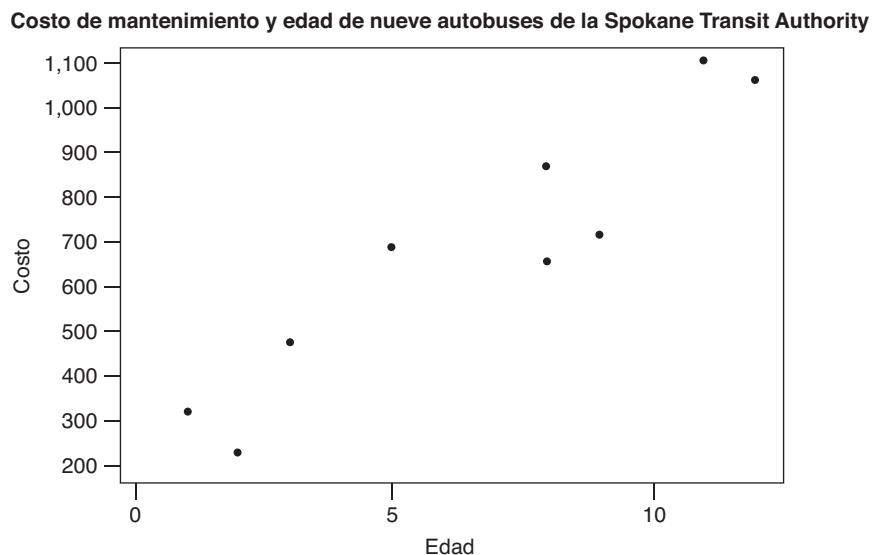
Una de las partes más difíciles de los pronósticos y que toma más tiempo es la recolección de datos válidos y confiables. Al personal de procesamiento de datos le gusta utilizar la expresión “entra basura, sale basura” (GIGO, del inglés *garbage in, garbage out*). Esta frase también se aplica a los pronósticos. Un pronóstico no puede ser más preciso que los datos en los que se basa. El modelo de pronóstico más sofisticado fallará si se aplica a datos que no sean confiables.

El advenimiento de la computadora ha propiciado una acumulación increíble de datos sobre casi todos los temas. La difícil tarea a la que se enfrenta la mayoría de los pronosticadores consiste en cómo encontrar datos relevantes que les ayuden a resolver los problemas propios de la toma de decisiones.

Pueden aplicarse cuatro criterios para determinar si los datos serán o no útiles:

1. Los datos deben ser fidedignos y precisos. Se debe tener cuidado que los datos sean recopilados de una fuente confiable y que se le dé la atención adecuada a la precisión.
2. Los datos deben ser relevantes. Deben ser representativos de las circunstancias para las que se utilizan.
3. Los datos deben ser coherentes. Cuando cambian las definiciones en lo que concierne a cómo se recopilan los datos, es necesario hacer ajustes para mantener la coherencia en los patrones históricos. Esto puede representar un problema, por ejemplo, cuando las agencias gubernamentales cambian la miscelánea o “canasta básica” que se utiliza para determinar el índice del costo de vida. Hace 30 años las computadoras personales no formaban parte de la combinación de productos que compraban los consumidores, ahora sí lo son.
4. Los datos deben ser oportunos. Los datos que se recopilan, se resumen y se publican oportunamente serán los más valiosos para quien haga los pronósticos. Puede haber muy pocos datos (historia insuficiente para basar en ella los resultados futuros) o demasiados (datos de períodos históricos irrelevantes del pasado remoto).

En general, hay dos tipos de datos que son de interés para quien hace pronósticos. El primero consiste en los datos recolectados en un solo punto del tiempo, ya sea una hora, un día, una semana, un mes o un trimestre. El segundo tipo radica en las observaciones de datos que se hacen a lo largo del tiempo. Cuando todas las observaciones son del mismo periodo, se llaman datos de *corte transversal*. El objetivo es examinar tales datos y luego extrapolar o extender las relaciones descubiertas hacia la población en general. Un ejemplo es tomar una muestra aleatoria de archivos de personal con el fin de estudiar las circunstancias de los empleados de una empresa. La recopilación de datos sobre la edad y el costo actual de mantenimiento de nueve autobuses de la Spokane Transit Authority es otro. Un diagrama de dispersión, como el de la figura 3.1, ayuda a visualizar la relación y sugiere que la edad podría utilizarse como auxiliar para pronosticar el presupuesto anual de mantenimiento.



**FIGURA 3.1** Diagrama de dispersión de edad contra costo de mantenimiento de nueve autobuses de la Spokane Transit Authority

Los datos de *corte transversal* son observaciones recolectadas en un solo punto del tiempo.

Cualquier variable que consiste en datos que se recopilan, registran u observan a lo largo de incrementos sucesivos de tiempo se conoce como una *serie de tiempo*. La producción mensual de cerveza en Estados Unidos es un ejemplo de una serie de tiempo.

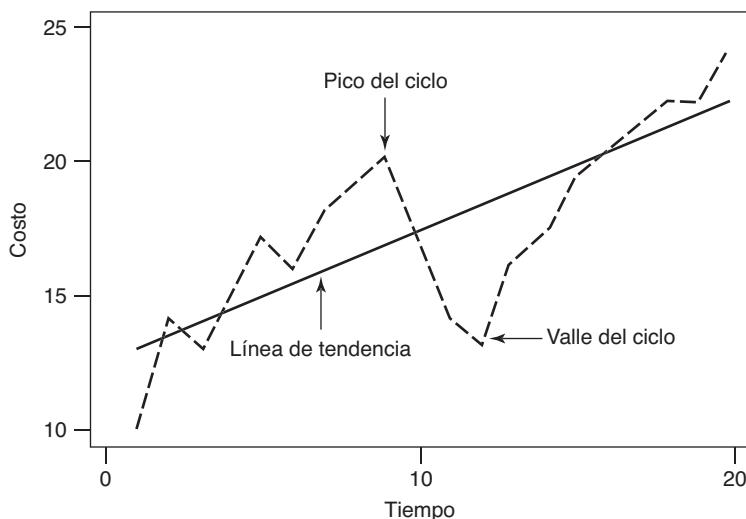
Una *serie de tiempo* consiste en datos que se recopilan, registran u observan a lo largo de incrementos sucesivos de tiempo.

## EXPLORACIÓN DE PATRONES DE DATOS DE SERIES DE TIEMPO

Al seleccionar un método de pronósticos adecuado para los datos de series de tiempo, considerar las distintas clases de patrones de datos es uno de los aspectos más importantes. Existen cuatro tipos generales: horizontales, tendencias, estacionales y cílicos.

Cuando las observaciones de datos fluctúan alrededor de un nivel constante o medio, existe un patrón *horizontal*. Este tipo de serie se conoce como *estacionaria* en su media. Se consideraría que las ventas mensuales de un producto que no aumentan ni disminuyen de manera consistente a lo largo del tiempo tienen un patrón horizontal.

Cuando las observaciones de datos crecen o disminuyen en un periodo largo, existe un patrón de *tendencia*. La figura 3.2 muestra el crecimiento de largo plazo (tendencia) de una serie de tiempo (costos de vivienda) con datos anuales. Para ilustrar este crecimiento, se ha trazado una línea de tendencia lineal. Aunque los costos variables de vivienda no han aumentado cada año, en general, el movimiento de la variable ha ido hacia arriba entre los períodos 1 y 20. Algunos ejemplos de las fuerzas básicas que afectan y ayudan a explicar las tendencias de una serie son el crecimiento de la población, la inflación de los precios, los cambios tecnológicos, las preferencias de los consumidores y los aumentos en la productividad.



**FIGURA 3.2 Componentes de tendencia y ciclo de una serie de tiempo anual, ejemplo de costos de vivienda**

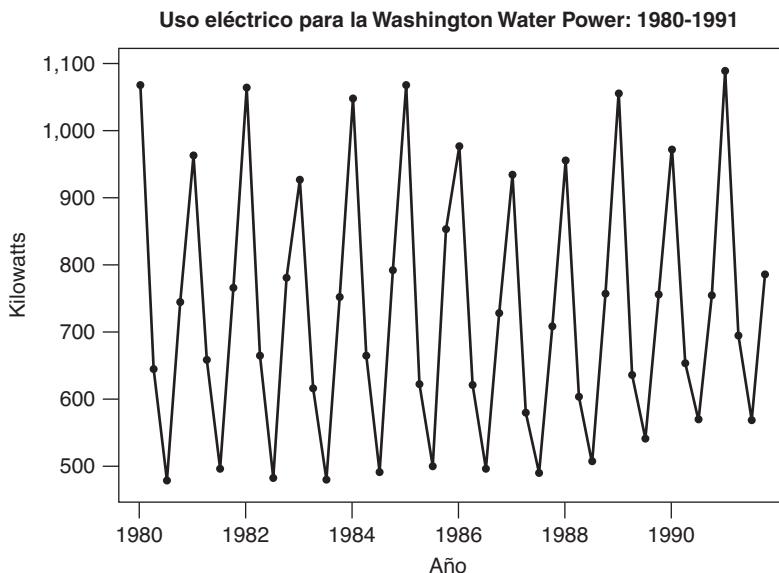
Muchas variables macroeconómicas, como el Producto Interno Bruto de Estados Unidos (PIB), el empleo y la producción industrial, exhiben comportamientos parecidos a las tendencias. La figura 3.10 (vea la página 71) contiene otro ejemplo de una serie de tiempo con una tendencia preponderante. Esta figura muestra el crecimiento de los ingresos operativos de Sears desde 1955 hasta 2000.

*La **tendencia** es el componente de largo plazo que representa el crecimiento o decrecimiento en la series de tiempo a lo largo de un periodo extenso.*

Cuando las observaciones exhiben aumentos y caídas que no se refieren a un periodo fijo, existe un patrón *cíclico*. El componente cíclico es la oscilación alrededor de la tendencia que por lo común es afectada por las condiciones económicas generales. A menudo, las fluctuaciones cíclicas se ven influidas por cambios en las expansiones y contracciones económicas: por lo general, se les conoce como ciclo del negocio. La figura 3.2 muestra una serie de tiempo con un componente cíclico. El pico del ciclo en el período 9 ilustra una expansión económica y el valle del ciclo en el período 12, una contracción económica.

*El **componente cíclico** es la oscilación alrededor de la tendencia.*

Cuando las observaciones se ven influidas por factores temporales, existe un patrón *estacional*. El componente estacional se refiere a un patrón de cambio que se repite año con año. En una serie mensual, el componente estacional mide la variabilidad de la serie cada enero, cada febrero y así sucesivamente. En una serie trimestral, hay cuatro elementos estacionales, uno para cada trimestre. La figura 3.3 muestra que el gasto eléctrico de los clientes residenciales de la Washington Water Power es más alto en el primer trimestre (meses de invierno) de cada año. La figura 3.14 (vea la página 73) muestra que las ventas trimestrales de Outboard Marine son típicamente bajas durante el primer trimestre de cada año. La variación estacional podría reflejar las condiciones climáticas, días feriados o la longitud del calendario.



**FIGURA 3.3 Gasto eléctrico para la Washington Water Power Company, 1980-1991**

El *componente estacional* es un patrón de cambios que se repite año tras año.

## EXPLORACIÓN DE PATRONES DE DATOS CON ANÁLISIS DE AUTOCORRELACIÓN

Cuando se mide una variable a lo largo del tiempo, las observaciones en diferentes períodos con frecuencia están relacionadas o correlacionadas. Esta correlación se mide mediante el uso del coeficiente de autocorrelación.

*Autocorrelación* es la correlación que existe entre una variable retrasada uno o más períodos consigo misma.

Los patrones de datos, que incluyen componentes como la tendencia y la estacionalidad, se pueden estudiar si se usan autocorrelaciones. Los coeficientes de autocorrelación para diferentes retrasos de una variable se usan para identificar patrones de datos de series de tiempo.

El concepto de autocorrelación se ilustra con los datos del ejemplo 3.1 que se presentan en la tabla 3.1. Observe que las variables  $Y_{t-1}$  y  $Y_{t-2}$  son en realidad los valores  $Y$  que se han retrasado uno y dos períodos, respectivamente. Los valores de marzo, que se muestran en la fila para el período 3, son las ventas de marzo,  $Y_t = 125$ ; las ventas de febrero,  $Y_{t-1} = 130$  y las de enero,  $Y_{t-2} = 123$ .

La ecuación 3.1 contiene la fórmula para calcular el coeficiente de autocorrelación ( $r_k$ ) entre las observaciones  $Y_t$  y  $Y_{t-k}$ , las cuales se encuentran a  $k$  períodos de distancia.

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad k = 0, 1, 2, \dots \quad (3.1)$$

**TABLA 3.1 Datos de grabadoras de vídeo para el ejemplo 3.1**

<i>Tiempo</i> <i>t</i>	<i>Mes</i>	<i>Datos originales</i> <i>Y<sub>t</sub></i>	<i>Y retrasada un periodo</i> <i>Y<sub>t-1</sub></i>	<i>Y retrasada dos periodos</i> <i>Y<sub>t-2</sub></i>
1	Enero	123		
2	Febrero	130	123	
3	Marzo	125	130	123
4	Abril	138	125	130
5	Mayo	145	138	125
6	Junio	142	145	138
7	Julio	141	142	145
8	Agosto	146	141	142
9	Septiembre	147	146	141
10	Octubre	157	147	146
11	Noviembre	150	157	147
12	Diciembre	160	150	157

donde

$r_k$  = coeficiente de autocorrelación para un retraso de  $k$  periodos

$\bar{Y}$  = media de los valores de la serie

$Y_t$  = observación en el periodo  $t$

$Y_{t-k}$  = observación de  $k$  periodos anteriores o durante un periodo  $t - k$

### Ejemplo 3.1

Harry Vernon ha recopilado datos sobre el número de grabadoras de vídeo (VCR) vendidas en Vernon's Music Store. Los datos se presentan en la tabla 3.1. La tabla 3.2 muestra las operaciones que llevaron al cálculo del coeficiente de autocorrelación para el retraso 1. La figura 3.4 contiene un diagrama de dispersión de los pares de observaciones  $(Y_{t-1}, Y_t)$ . Queda claro a partir del diagrama de dispersión que la correlación del retraso 1 será positiva.

El coeficiente de autocorrelación del retraso 1 ( $r_1$ ) o la autocorrelación entre  $Y_t$  y  $Y_{t-1}$ , se calcula al utilizar los totales de la tabla 3.2 y la ecuación 3.1. Así,

$$r_1 = \frac{\sum_{t=1+1}^n (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{843}{1,474} = .572$$

Como lo sugiere la gráfica de la figura 3.4, en esta serie de tiempo existe una autocorrelación positiva para el retraso 1. La correlación entre  $Y_t$  y  $Y_{t-1}$ , o la autocorrelación para el retraso 1, es de 0.572. Esto significa que las ventas mensuales sucesivas de videograbadoras están, en cierto modo, correlacionadas unas con otras. Esta información podría darle a Harry perspectivas valiosas sobre su serie de tiempo, le ayudaría a prepararse para utilizar un método avanzado de pronósticos y le sugeriría usar el análisis de regresión. Todas estas ideas se presentarán en capítulos posteriores.

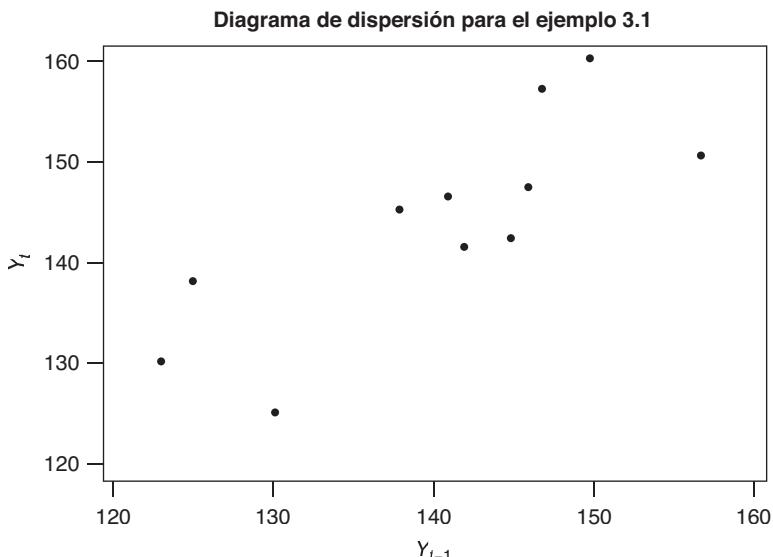
El coeficiente de autocorrelación de segundo orden ( $r_2$ ) o la correlación entre  $Y_t$  y  $Y_{t-2}$  para los datos de Harry también se calcula utilizando la ecuación 3.1.

$$r_2 = \frac{\sum_{t=2+1}^n (Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} = \frac{682}{1,474} = .463$$

TABLA 3.2 Cálculo del coeficiente de autocorrelación para el retraso 1 de los datos en la tabla 3.1							
Tiempo, $t$	$Y_t$	$Y_{t-1}$	$(Y_t - \bar{Y})$	$(Y_{t-1} - \bar{Y})$	$(Y_t - \bar{Y})^2$	$(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})$	
1	123	—	-19	—	361	—	
2	130	123	-12	-19	144	228	
3	125	130	-17	-12	289	204	
4	138	125	-4	-17	16	68	
5	145	138	3	-4	9	-12	
6	142	145	0	3	0	0	
7	141	142	-1	0	1	0	
8	146	141	4	-1	16	-4	
9	147	146	5	4	25	20	
10	157	147	15	5	225	75	
11	150	157	8	15	64	120	
12	160	150	18	8	324	144	
Total	1,704		0		1,474	843	

$\bar{Y} = \frac{1,704}{12} = 142$

$r_1 = \frac{843}{1,474} = .572$



**FIGURA 3.4 Diagrama de dispersión para los datos de Vernon's Music Store del ejemplo 3.1**

Parece que existe una autocorrelación moderada en esta serie de tiempo para dos períodos de retraso. La correlación entre  $Y_t$  y  $Y_{t-2}$ , o correlación para el retraso 2, es de 0.463. Observe que el coeficiente de autocorrelación en el retraso 2 (0.463) es menor que el coeficiente de autocorrelación en el retraso 1 (0.572). En general, conforme aumenta el número de retrasos de tiempo,  $k$ , disminuyen las magnitudes de los coeficientes de autocorrelación.

La figura 3.5 muestra una gráfica de las autocorrelaciones contra los retrasos de tiempo para los datos de Harry Vernon utilizados en el ejemplo 3.1. La escala horizontal en la parte de abajo de la gráfica muestra cada uno de los retrasos de tiempo que son de interés,

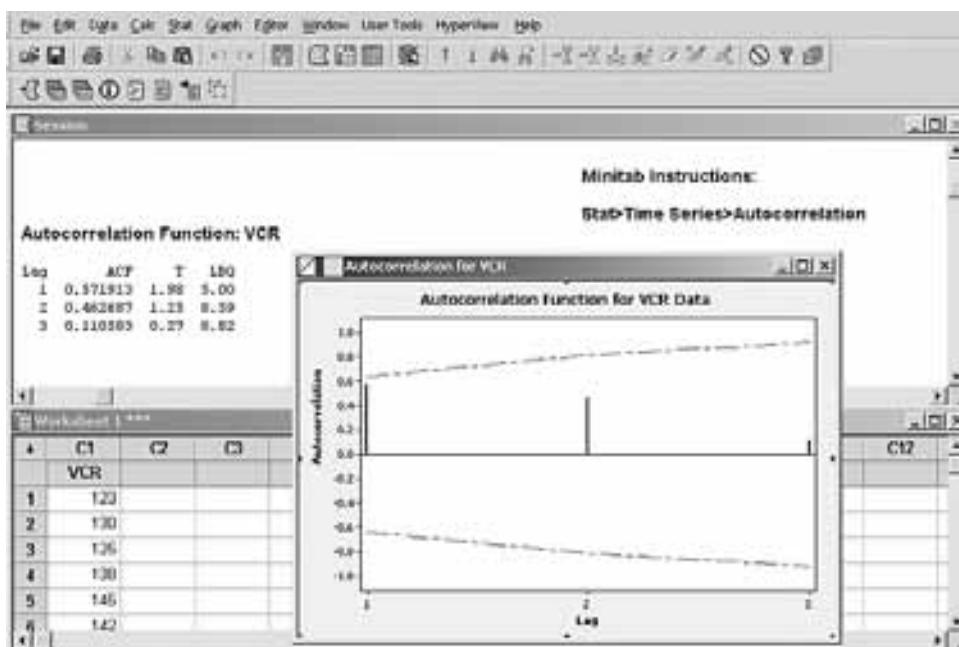


FIGURA 3.5 Función de autocorrelación para los datos utilizados en el ejemplo 3.1

1, 2, 3 y así sucesivamente. La escala vertical del lado izquierdo muestra el posible rango de un coeficiente de autocorrelación, de  $-1$  a  $+1$ . La línea horizontal a la mitad de la gráfica representa autocorrelaciones de cero. La línea vertical que se extiende hacia arriba por encima del retraso de tiempo 1 muestra un coeficiente de autocorrelación de  $0.57$ ,  $r_1 = 0.57$ . La línea vertical que se extiende hacia arriba por encima del retraso de tiempo 2 muestra un coeficiente de autocorrelación de  $0.46$ ,  $r_2 = 0.46$ . La línea punteada y las estadísticas T y LBQ que se muestran en la ventana Session se analizarán en los ejemplos 3.2 y 3.3. Los patrones de un correlograma se utilizan para analizar las características fundamentales de los datos, un concepto que se demuestra en la siguiente sección. El software Minitab (vea la sección “Aplicaciones en Minitab” al final del capítulo para obtener instrucciones específicas) se puede utilizar para calcular autocorrelaciones y desarrollar correlogramas.

El correlograma o función de autocorrelación es una gráfica de las autocorrelaciones para varios retrasos de una serie de tiempo.

Con una pantalla como la de la figura 3.5 pueden estudiarse los patrones de datos, incluso, la tendencia y la estacionalidad. Asimismo, pueden utilizarse coeficientes de autocorrelación para diferentes retrasos de tiempo para una variable, a fin de contestar las siguientes preguntas sobre una serie de tiempo:

1. ¿Los datos son aleatorios?
2. ¿Tienen tendencia los datos (no estacionarios)?
3. ¿Son estacionarios los datos?
4. ¿Los datos son estacionales?

Si una serie es aleatoria, las correlaciones entre  $Y_t$  y  $Y_{t-k}$  para cualquier retraso  $k$  son cercanas a cero. Los valores sucesivos de una serie de tiempo no se relacionan unos con otros.

Si una serie tiene tendencia, las observaciones sucesivas están muy correlacionadas y es típico que los coeficientes de correlación sean bastante diferentes de cero para los

primeros retrasos de tiempo y entonces, de forma gradual, caen hacia cero conforme aumenta el número de retrasos. Frecuentemente, el coeficiente de autocorrelación para un retraso de tiempo 1 es muy grande (cercano a 1). También el coeficiente de autocorrelación para el retraso de tiempo 2 será grande. Sin embargo, no será tan grande como para el retraso de tiempo 1.

Si una serie tiene un patrón estacional, se dará un importante coeficiente de autocorrelación en el retraso de tiempo estacional o en múltiplos del retraso estacional. El retraso estacional se considera de 4 para datos trimestrales y de 12 para datos mensuales.

¿Cómo determina un analista si un coeficiente de autocorrelación es significativamente diferente de cero para los datos de la tabla 3.1? Quenouille (1949) y otros han demostrado que los coeficientes de autocorrelación de datos aleatorios tienen una distribución muestral que puede aproximarse mediante una curva normal con media cero y una desviación estándar aproximada de  $1/\sqrt{n}$ . Al saber esto, el analista puede comparar los coeficientes de autocorrelación con esta distribución muestral teórica y determinar si, para retrasos de tiempo dados, provienen de una población cuya media es de cero.

De hecho, algunos paquetes de software utilizan una fórmula un poco diferente para calcular las desviaciones estándar (o errores estándar) de los coeficientes de autocorrelación, como se muestra en la ecuación 3.2. Esta fórmula supone que cualquier autocorrelación anterior al retraso  $k$  es diferente de cero y que cualquier autocorrelación en retrasos mayores o iguales a  $k$  es cero. Para una autocorrelación en el retraso 1 se utiliza el error estándar  $1/\sqrt{n}$ .

$$SE(r_k) = \sqrt{\frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{n}} \quad (3.2)$$

donde

$SE(r_k)$  = error estándar de la autocorrelación en el retraso  $k$

$r_i$  = autocorrelación en el retraso  $i$

$k$  = el retraso del tiempo

$n$  = número de observaciones en la serie de tiempo

Este cálculo se demostrará en el ejemplo 3.2. Si la serie es realmente aleatoria, casi todos los coeficientes de autocorrelación de la muestra deberían encontrarse dentro de un rango especificado por cero, más o menos cierto número de errores estándar. En un nivel de confianza determinado, una serie se puede considerar aleatoria si los coeficientes de autocorrelación calculados se encuentran dentro del intervalo alrededor de cero dado por:

$$0 \pm t \times SE(r_k)$$

donde el multiplicador  $t$  es un punto porcentual adecuado de una distribución  $t$ .

Aunque resulta útil la prueba de cada una de las  $r_k$  para saber si en lo individual son significativamente diferentes de 0, también es una buena práctica examinar una secuencia de  $r_k$  como un grupo. Se puede utilizar una prueba de conjunto para ver si, por ejemplo, de los primeros 10 valores de  $r_k$ , es significativamente distinto de un grupo en el cual todos los 10 valores son cero.

Una prueba de baúl común es la estadística modificada  $Q$  de Box-Pierce (ecuación 3.3) desarrollada por Ljung y Box. Por lo general, dicha prueba se aplica a los residuales de un modelo de pronóstico. Si las autocorrelaciones se calculan a partir de un proceso aleatorio (o de ruido blanco), la estadística  $Q$  tiene una distribución chi cuadrada con  $m$  grados de libertad (el número de retrasos de tiempo por probarse). Sin embargo, para los residuales de un modelo de pronóstico, la estadística  $Q$  tiene una distribución chi cuadrada con los grados de libertad iguales a  $m$ , menos el número de parámetros estimados en el

modelo. El valor de la estadística  $Q$  puede compararse con la tabla de chi cuadrada (tabla C-4), a fin de determinar si es mayor de lo que se esperaría bajo la hipótesis nula de que todas las autocorrelaciones del conjunto son cero. De manera alterna, se puede calcular e interpretar el valor  $p$  generado mediante la estadística  $Q$  de prueba. La estadística  $Q$  está dada en la ecuación 3.3. Se demostrará en el ejemplo 3.3.

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2}{n - k} \quad (3.3)$$

donde

$n$  = número de observaciones en la serie de tiempo

$k$  = el retraso del tiempo

$m$  = el número de retrasos de tiempo a ser probados

$r_k$  = función de autocorrelación de la muestra de los residuales retrasados  $k$  periodos

### ¿Son aleatorios los datos?

La ecuación 3.4 es un modelo aleatorio simple llamado *modelo de ruido blanco*. La observación  $Y_t$  está compuesta de dos partes:  $c$ , el nivel general, y  $\varepsilon_t$ , el componente de error aleatorio. Es importante notar que se supone que el componente  $\varepsilon_t$  no está correlacionado de un periodo a otro.

$$Y_t = c + \varepsilon_t \quad (3.4)$$

¿Los datos de la tabla 3.1 son congruentes con este modelo? Esta cuestión se explorará en los ejemplos 3.2 y 3.3.

#### Ejemplo 3.2

Se desarrolla una prueba de hipótesis para determinar si un coeficiente de autocorrelación específico es diferente de cero, en una forma significativa, para el correlograma mostrado en la figura 3.5. Las hipótesis nula y alternativa para probar la significancia del coeficiente de autocorrelación para la población del retraso 1 son:

$$\begin{aligned} H_0: \rho_1 &= 0 \\ H_1: \rho_1 &\neq 0 \end{aligned}$$

Si la hipótesis nula es verdadera, la estadística de prueba

$$t = \frac{r_1 - \rho_1}{SE(r_1)} = \frac{r_1 - 0}{SE(r_1)} = \frac{r_1}{SE(r_1)} \quad (3.5)$$

tiene una distribución  $t$  con  $df = n - 1$ . Aquí,  $n - 1 = 12 - 1 = 11$ , así que para un nivel de significancia de 5% la regla de decisión es:

**Regla de decisión:** si  $t < -2.2$  o  $t > 2.2$ , se rechaza  $H_0$  y se concluye que la autocorrelación para el retraso 1 es significativamente diferente de cero.

Los valores críticos  $\pm 2.2$  son los puntos 0.025 superiores e inferiores de una distribución  $t$  con 11 grados de libertad. El error estándar de  $r_1$  es  $SE(r_1) = \sqrt{1/12} = \sqrt{0.083} = 0.289$  y el valor de la estadística de prueba se convierte en

$$t = \frac{r_1}{SE(r_1)} = \frac{.572}{.289} = 1.98$$

y, si se usa la regla de decisión anterior, no puede rechazarse  $H_0: \rho_1 = 0$ , debido a que  $-2.2 < 1.98 < 2.2$ . Observe que el valor de nuestra estadística de prueba,  $t = 1.98$ , es el mismo que la cantidad en la fila del retraso 1 por debajo del encabezado T en los resultados de Minitab de la figura 3.5. Los valores T en los resultados de Minitab tan sólo son los valores de la estadística de prueba al probar una autocorrelación cero en los diferentes retrasos.

Para probar una autocorrelación de cero en un retraso de tiempo 2, se considera

$$\begin{aligned} H_0: \rho_2 &= 0 \\ H_1: \rho_2 &\neq 0 \end{aligned}$$

y la estadística de prueba

$$t = \frac{r_2 - \rho_2}{SE(r_2)} = \frac{r_2 - 0}{SE(r_2)} = \frac{r_2}{SE(r_2)}$$

Al usar la ecuación (3.2),

$$SE(r_2) = \sqrt{\frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{n}} = \sqrt{\frac{1 + 2 \sum_{i=1}^{2-1} r_i^2}{n}} = \sqrt{\frac{1 + 2(0.572)^2}{12}} = \sqrt{\frac{1.6544}{12}} = \sqrt{.138} = .371$$

y

$$t = \frac{.463}{.371} = 1.25$$

Este resultado concuerda con el valor T para el retraso 2 en los resultados de Minitab de la figura 3.5.

Al usar la regla de decisión anterior,  $H_0: \rho_2 = 0$  no puede rechazarse en el nivel 0.05 debido a que  $-2.2 < 1.25 < 2.2$ . Una forma alterna de verificar una autocorrelación significativa es construir, digamos, límites de confianza de 95% centrados en cero. Estos límites para los retrasos 1 y 2 están dados por

$$\begin{aligned} \text{retraso 1: } 0 &\pm t_{0.025} \times SE(r_1) \quad \text{o} \quad 0 \pm 2.2 (0.289) \rightarrow (-0.636, 0.636) \\ \text{retraso 2: } 0 &\pm t_{0.025} \times SE(r_2) \quad \text{o} \quad 0 \pm 2.2 (0.371) \rightarrow (-0.816, 0.816) \end{aligned}$$

Una autocorrelación significativamente distinta de cero se indica siempre y cuando el valor de  $r_k$  caiga fuera de los límites de confianza correspondientes. Los límites de confianza de 95% se muestran en la figura 3.5 mediante líneas punteadas en la expresión gráfica de la función de autocorrelación.

### Ejemplo 3.3

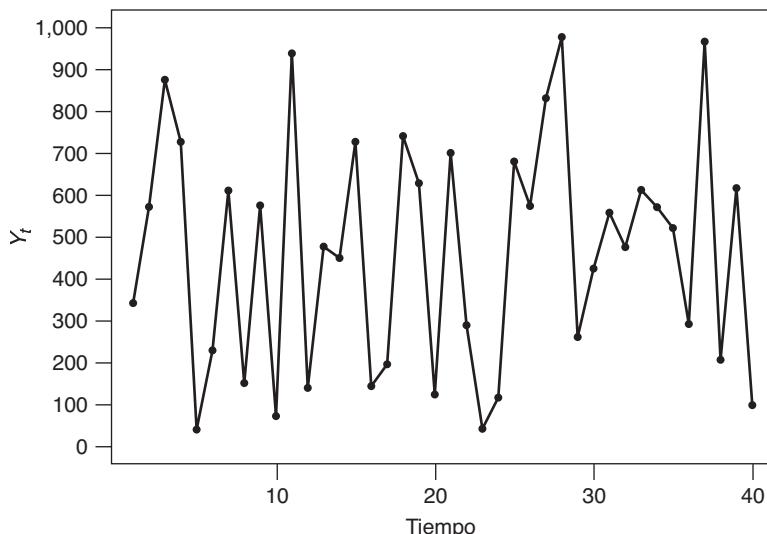
Se utilizó Minitab para generar la serie de tiempo de 40 números seudoaleatorios de tres dígitos que se muestran en la tabla 3.3. La figura 3.6 muestra una gráfica de serie de tiempo para estos datos. Puesto que los números son aleatorios (independientes unos de otros y todos de entre la misma población), en teoría, las autocorrelaciones para todos los retrasos de tiempo deberían ser iguales a cero. Por supuesto, los 40 valores de la tabla 3.3 tan sólo son un grupo de un gran número posible de muestras de tamaño 40. Cada muestra producirá diferentes autocorrelaciones. La mayoría de ellas generan coeficientes de autocorrelación de la muestra cercanos a cero. Sin embargo, es posible que una muestra origine un coeficiente de autocorrelación bastante diferente de cero tan sólo por casualidad.

A continuación se construye la función de autocorrelación que se muestra en la figura 3.7 mediante el uso de Minitab (vea la página 68). Observe que las dos líneas punteadas muestran los límites de confianza de 95%. Se examinan diez retrasos de tiempo y dentro de esos límites están todos los coeficientes de autocorrelación individuales. No hay razón para dudar de que cada una de las autocorrelaciones para los siguientes diez retrasos sea cero. Sin embargo, aunque las autocorrelaciones individuales de la muestra no son significativamente diferentes de cero, ¿las magnitudes de los primeros  $10r_k$ , consideradas como un grupo, son mayores de lo que se esperaría bajo la hipótesis de que no hay autocorrelación en retraso alguno? Esta pregunta se contesta mediante la estadística  $Q$  de Ljung-Box (LBQ en Minitab).

Si no existe autocorrelación en cualquiera de los retrasos, la estadística  $Q$  tiene una distribución de chi cuadrada, en este caso, con un  $df = 10$ . En consecuencia, un valor alto para  $Q$  (en un extremo de la distribución de chi cuadrada) es evidencia en contra de la hipótesis nula. A partir de la figura 3.7, el valor de  $Q$  (LBQ) para diez retrasos de tiempo es de 7.75. De la tabla C-4, el punto 0.05 superior de una distribución de chi cuadrada con 10 grados de libertad es de 18.31. Puesto que  $7.75 < 18.31$ , no puede rechazarse la hipótesis nula en el nivel de significancia de 5%. Estos datos no están correlacionados en cualquier retraso de tiempo, lo cual es un resultado coherente con el modelo de la ecuación 3.4.

**TABLA 3.3 Serie de tiempo de 40 números aleatorios para el ejemplo 3.3**

$t$	$Y_t$	$t$	$Y_t$	$t$	$Y_t$	$t$	$Y_t$
1	343	11	946	21	704	31	555
2	574	12	142	22	291	32	476
3	879	13	477	23	43	33	612
4	728	14	452	24	118	34	574
5	37	15	727	25	682	35	518
6	227	16	147	26	577	36	296
7	613	17	199	27	834	37	970
8	157	18	744	28	981	38	204
9	571	19	627	29	263	39	616
10	72	20	122	30	424	40	97



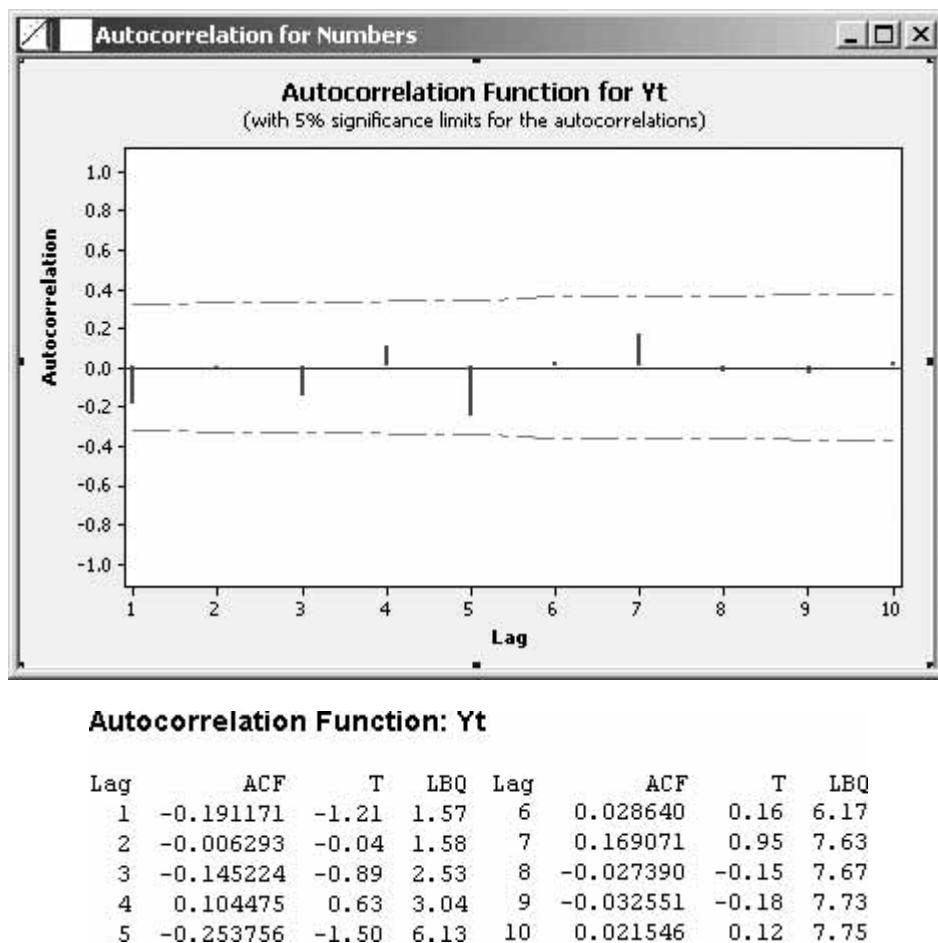
**FIGURA 3.6 Gráfica de la serie de tiempo de números aleatorios utilizados en el ejemplo 3.3**

### ¿Tienen tendencia los datos?

Si una serie tiene tendencia, entonces hay una relación significativa entre valores sucesivos de series de tiempo. Es típico que los coeficientes de autocorrelación sean grandes en los primeros retrasos de tiempo y después, gradualmente, caen hacia cero conforme aumenta el número de retrasos.

Una serie de tiempo *estacionaria* es aquella cuyas propiedades estadísticas básicas, tales como la media y la varianza, permanecen constantes a lo largo del tiempo. En consecuencia, una serie que varía alrededor de un nivel fijo a lo largo del tiempo (sin crecimiento ni decrecimiento) se dice que es estacionaria. Una serie que contiene una tendencia es *no estacionaria*. Los coeficientes de autocorrelación para una serie estacionaria caen hacia cero bastante rápido, en general, después del segundo o tercer retrasos de tiempo. Por otro lado, las autocorrelaciones de muestras permanecen bastante grandes en una serie no estacionaria durante varios períodos. A menudo, para analizar las series no estacionarias, se elimina la tendencia antes de realizar modelos adicionales. Los procedimientos que se presentan en el capítulo 9 utilizan este enfoque.

Un método llamado *construcción de diferencias* puede usarse habitualmente para eliminar la tendencia de una serie no estacionaria. Los datos de las videocámaras presentados



**FIGURA 3.7** Función de autocorrelación para los datos usados en el ejemplo 3.3

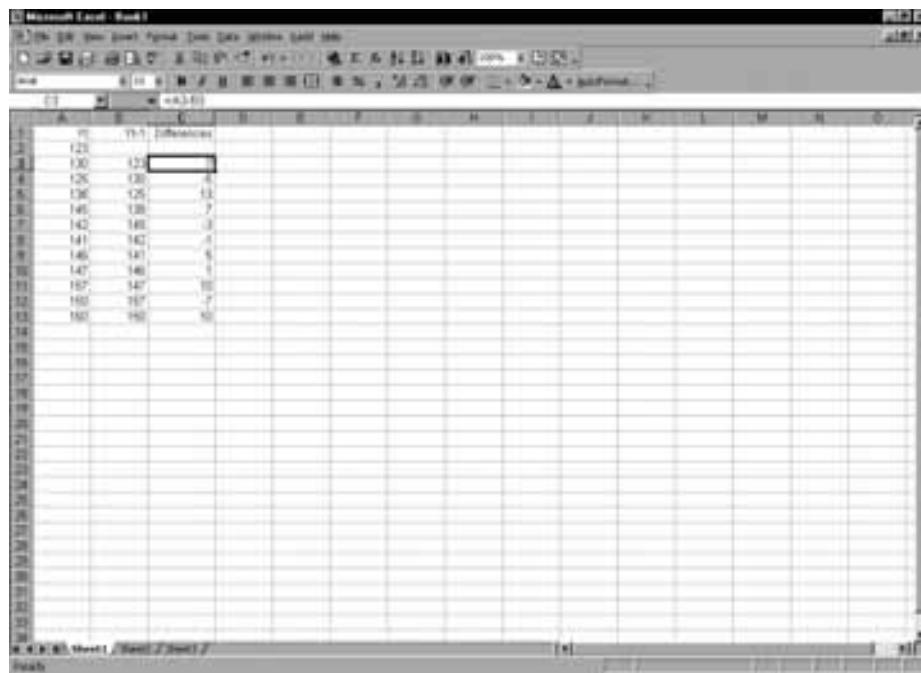
originalmente en la tabla 3.1 se muestran otra vez en la columna A de la figura 3.8. El valor  $Y_t$  retrasado un periodo,  $Y_{t-1}$ , se muestra en la columna B. Las diferencias,  $Y_t - Y_{t-1}$  (columna A - columna B) se muestran en la columna C. Por ejemplo, el primer valor de las diferencias es  $Y_2 - Y_1 = 130 - 123 = 7$ . Observe el crecimiento o tendencia de los datos para las videogramadoras que se muestra en la gráfica A de la figura 3.9. Luego, observe el patrón estacionario de los datos diferenciados en la gráfica B. La construcción de diferencias de los datos ha quitado la tendencia.

#### Ejemplo 3.4

Maggie Trymane, analista de Sears, tiene la tarea de pronosticar los ingresos operativos para el 2001. Recopila los datos por año desde 1955 hasta 2000, como se muestra en la tabla 3.4. Los datos se grafican como una serie de tiempo en la figura 3.10 (vea la página 71). Primero, calcula un intervalo de confianza de 95% para los coeficientes de autocorrelación en el retraso de tiempo 1 utilizando  $0 \pm Z_{0.025} (1/\sqrt{n})$  donde, para muestras grandes, la distribución normal estándar correspondiente a 0.025 ha reemplazado a la distribución *t* correspondiente:

$$0 \pm 1.96 (\sqrt{1/46}) \\ 0 \pm 0.289$$

A continuación, Maggie procesa los datos en Minitab y produce la función de autocorrelación que se muestra en la figura 3.11. Al examinarla, nota que las autocorrelaciones de los primeros tres retrasos de tiempo son significativamente diferentes de cero (0.96, 0.92 y 0.87) y que



**FIGURA 3.8 Resultados de Excel para la diferenciación de los datos de videograbadoras del ejemplo 3.1**

entonces los valores caen de forma gradual hasta cero. Como verificación final, Maggie observa la estadística  $Q$  para los 10 retrasos de tiempo. El LBQ es de 274.97, lo cual es mayor que el valor chi cuadrada de 18.3 (correspondiente al nivel de significancia 0.05 de una distribución de chi cuadrada con 10 grados de libertad). Entonces decide que los datos tienen una gran autocorrelación y que exhiben un comportamiento con tendencia.

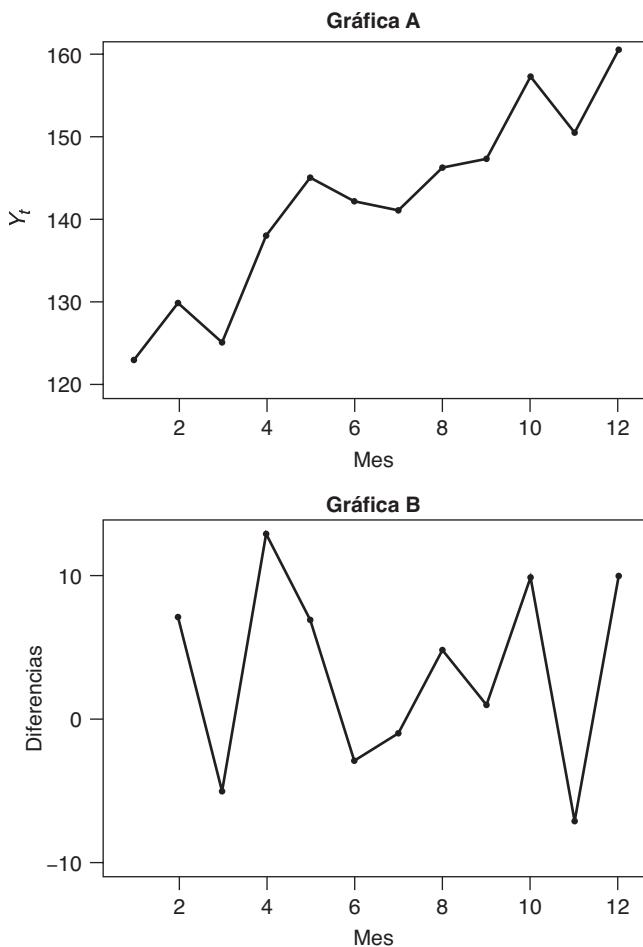
Maggie sospecha de que la serie se puede diferenciar para quitar la tendencia y para crear una serie estacionaria. Comparó los datos (vea “Aplicaciones en Minitab” en la página 95) con los resultados que se muestran en la figura 3.12. La serie diferenciada no muestra evidencia alguna de tendencia y la función de autocorrelación que se muestra en la figura 3.13 parece apoyar esta conclusión. Al examinar la figura 3.13, Maggie nota que el coeficiente de autocorrelación para el retraso de tiempo 3, de 0.32, es significativamente diferente de cero (en el nivel de significancia 0.05). Las autocorrelaciones en retrasos diferentes al retraso 3 son pequeñas, y ella se pregunta si hay algún patrón en estos datos del que pueda surgir un modelo mediante una de las técnicas de pronósticos más avanzadas, como las que se presentan en el capítulo 9.

### ¿Son estacionales los datos?

Si una serie es estacional, se repite un mismo patrón año con año en fechas o períodos específicos del calendario. Las observaciones en las mismas fechas o períodos estacionales tienden a estar relacionadas. Si se analizan los datos trimestrales con un patrón estacional, los primeros trimestres tienden a parecerse, igual que los segundos y así sucesivamente; entonces aparecerá un coeficiente de autocorrelación significativo en el retraso de 4 períodos. Si se analizan los datos mensuales, un coeficiente de autocorrelación significativo aparecerá en el retraso de tiempo 12. Es decir, enero se correlacionará con otros eneros, febrero lo hará con otros febreros y así sucesivamente. El ejemplo 3.5 analiza una serie estacional.

#### Ejemplo 3.5

Perkin Kendell es un analista de Outboard Marine Corporation. Siempre ha creído que las ventas son estacionales. Perkin recopila los datos que se muestran en la tabla 3.5 (vea la página 73) para las ventas trimestrales de Outboard Marine Corporation desde 1984 hasta 1996 y elabora



**FIGURA 3.9** Gráficas de series de tiempo de los datos de videogramadoras y datos diferenciados de videogramadoras

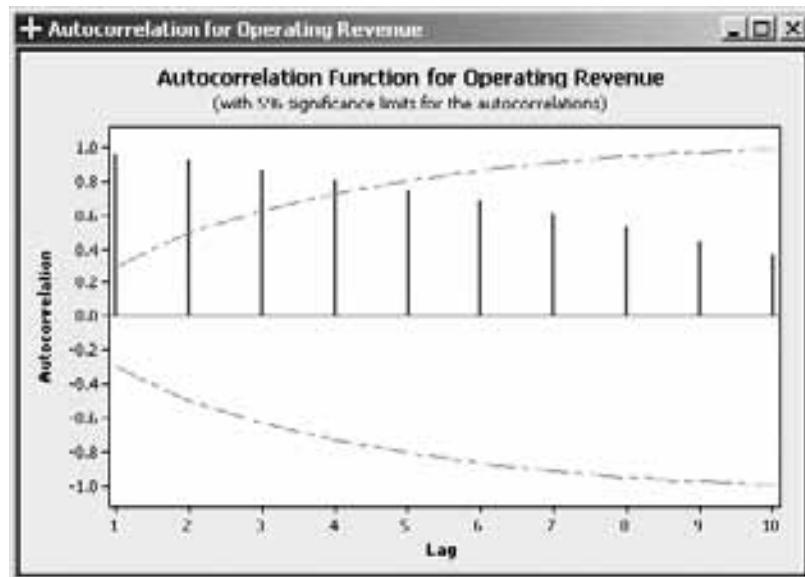
**TABLA 3.4** Ingresos operativos anuales de Sears Roebuck & Co., 1955-2000 del ejemplo 3.4

Año	$Y_t$	Año	$Y_t$	Año	$Y_t$	Año	$Y_t$
1955	3,307	1967	7,296	1979	17,514	1991	57,242
1956	3,556	1968	8,178	1980	25,195	1992	52,345
1957	3,601	1969	8,844	1981	27,357	1993	50,838
1958	3,721	1970	9,251	1982	30,020	1994	54,559
1959	4,036	1971	10,006	1983	35,883	1995	34,925
1960	4,134	1972	10,991	1984	38,828	1996	38,236
1961	4,268	1973	12,306	1985	40,715	1997	41,296
1962	4,578	1974	13,101	1986	44,282	1998	41,322
1963	5,093	1975	13,639	1987	48,440	1999	41,071
1964	5,716	1976	14,950	1988	50,251	2000	40,937
1965	6,357	1977	17,224	1989	53,794		
1966	6,769	1978	17,946	1990	55,972		

Fuente: *Industry Surveys*, varios años.



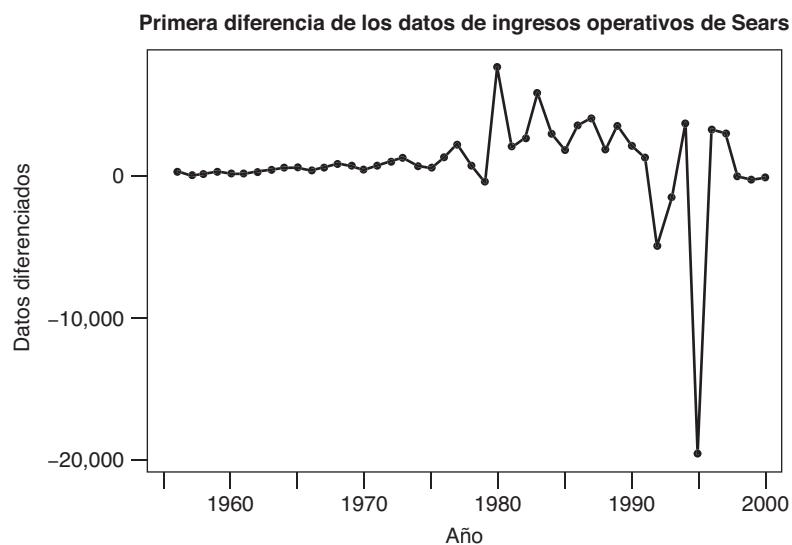
**FIGURA 3.10** Gráfica de series de tiempo de ingresos operativos de Sears



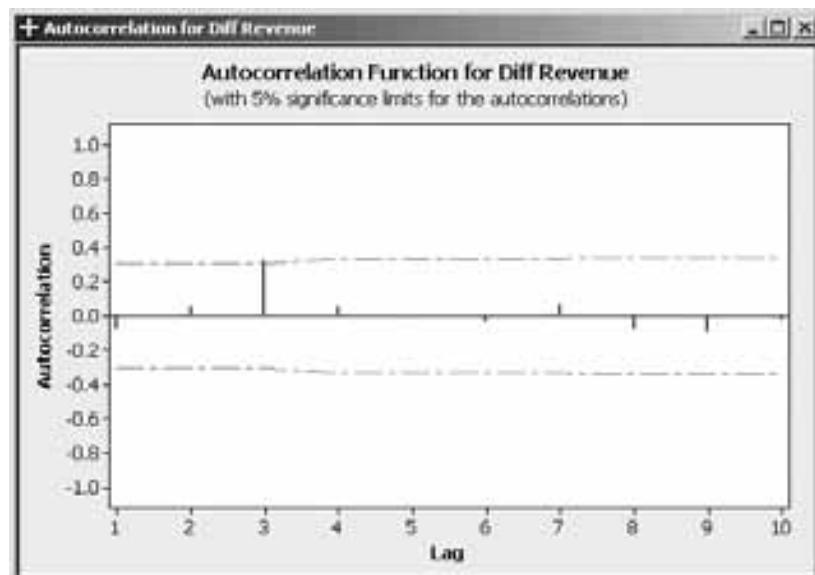
**Autocorrelation Function: Revenue**

Lag	ACF	T	LBI	Lag	ACF	T	LBI
1	0.957322	6.49	44.97	6	0.692167	1.61	217.26
2	0.915920	3.69	87.07	7	0.611867	1.35	238.45
3	0.870327	2.78	125.96	8	0.530836	1.13	254.82
4	0.810476	2.24	160.49	9	0.449109	0.93	266.86
5	0.750323	1.88	190.81	10	0.363433	0.74	274.97

**FIGURA 3.11** Función de autocorrelación para los ingresos operativos de Sears



**FIGURA 3.12** Gráfica de series de tiempo de las primeras diferencias de los ingresos operativos de Sears



**Autocorrelation Function: Diff Revenue**

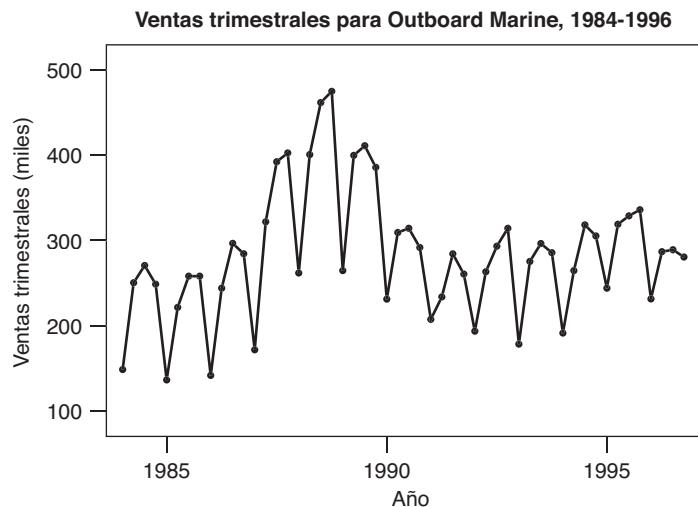
Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	-0.002142	-0.55	0.32	6	-0.032326	-0.20	5.96
2	0.050143	0.33	0.45	7	0.060489	0.36	6.17
3	0.324999	2.16	5.77	8	-0.080592	-0.48	6.54
4	0.051342	0.31	5.90	9	-0.092040	-0.55	7.03
5	-0.006329	-0.04	5.91	10	-0.027916	-0.17	7.08

**FIGURA 3.13** Función de autocorrelación para las primeras diferencias de los ingresos operativos de Sears

**TABLA 3.5 Ventas trimestrales para Outboard Marine, 1984-1996, del ejemplo 3.5**

<i>Fin del año fiscal</i>	<i>31 de diciembre</i>	<i>31 de marzo</i>	<i>30 de junio</i>	<i>30 de septiembre</i>
1984	147.6	251.8	273.1	249.1
1985	139.3	221.2	260.2	259.5
1986	140.5	245.5	298.8	287.0
1987	168.8	322.6	393.5	404.3
1988	259.7	401.1	464.6	479.7
1989	264.4	402.6	411.3	385.9
1990	232.7	309.2	310.7	293.0
1991	205.1	234.4	285.4	258.7
1992	193.2	263.7	292.5	315.2
1993	178.3	274.5	295.4	286.4
1994	190.8	263.5	318.8	305.5
1995	242.6	318.8	329.6	338.2
1996	232.1	285.6	291.0	281.4

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1988, 1992, 1996), p. 1773.



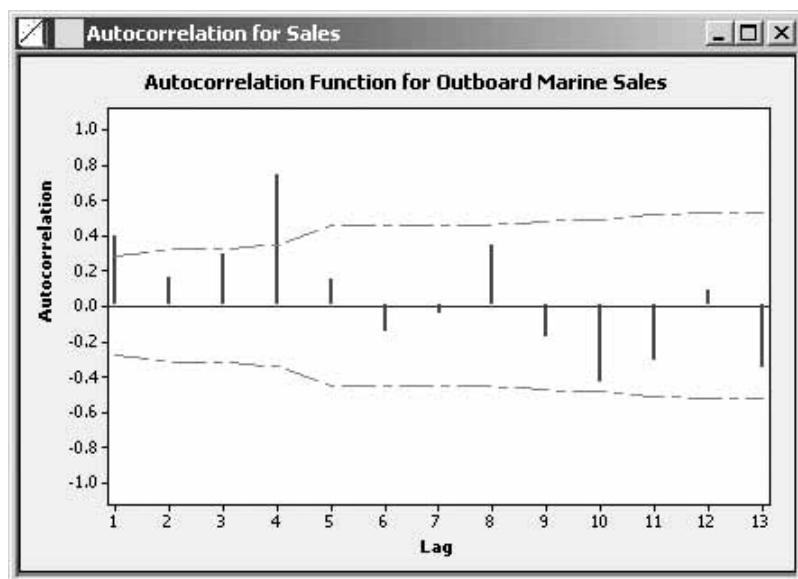
**FIGURA 3.14 Gráfica de series de tiempo para las ventas trimestrales de Outboard Marine**

la gráfica de serie de tiempo que se muestra en la figura 3.14. A continuación, calcula un intervalo de confianza de 95% para una muestra grande para el coeficiente de autocorrelación en el retraso de tiempo 1:

$$0 \pm 1.96\sqrt{1/52}$$

$$0 \pm 0.272$$

A continuación, Perkin calcula los coeficientes de autocorrelación que se muestran en la figura 3.15. Observa que los coeficientes de autocorrelación en los retrasos de tiempo 1 y 4 son significativamente diferentes de cero ( $r_1 = 0.39 > 0.272$  y  $r_4 = 0.74 > 0.333$ ). Concluye que las ventas de Outboard Marine son estacionales de forma trimestral.



Autocorrelation Function: Sales

Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0.392861	2.83	8.50	8	0.346975	1.51	57.72
2	0.153945	0.97	9.83	9	-0.182601	-0.76	59.90
3	0.293782	1.82	14.77	10	-0.434729	-1.80	72.53
4	0.743520	4.34	47.11	11	-0.315031	-1.23	79.33
5	0.150655	0.67	48.47	12	0.091203	0.35	79.91
6	-0.153008	-0.67	49.90	13	-0.353274	-1.34	88.90
7	-0.046978	-0.21	50.04				

FIGURA 3.15 Función de autocorrelación para las ventas trimestrales de Outboard Marine

## ELECCIÓN DE UNA TÉCNICA DE PRONÓSTICOS

Este texto está dedicado en su mayor parte a explicar las diversas técnicas de pronósticos y a demostrar su utilidad. Primero, se aborda la importante labor de elegir la mejor técnica de pronósticos.

Algunas de las preguntas que deben plantearse antes de decidir la técnica de pronósticos más apropiada para un problema específico son las siguientes:

- ¿Por qué se necesita un pronóstico?
- ¿Quién utilizará el pronóstico?
- ¿Cuáles son las características de los datos disponibles?
- ¿Qué periodo debe pronosticarse?
- ¿Cuáles son los requisitos mínimos de datos?
- ¿Qué tanta precisión se desea?
- ¿Cuánto costará el pronóstico?

A fin de seleccionar adecuadamente la técnica conveniente de pronósticos, el pronosticador debe ser capaz de:

- Definir la naturaleza del problema de pronóstico.
- Explicar la naturaleza de los datos que se investigan.

- Describir las capacidades y limitaciones de técnicas de pronósticos potencialmente útiles.
- Desarrollar algunos criterios predeterminados sobre los que se pueda tomar la decisión de selección.

Un factor importante que influye en la selección de una técnica de pronóstico es identificar y entender los patrones históricos de los datos. Si se pueden reconocer patrones de tendencia, cílicos o estacionales, pueden seleccionarse técnicas capaces de extrapolarlos de manera eficaz.

### Técnicas de pronóstico para datos estacionarios

Ya se definió una serie estacionaria como aquella cuyo valor medio no cambia a lo largo del tiempo. Tales situaciones surgen cuando los patrones de demanda que influyen en la series son relativamente estables. En su forma más simple, el pronóstico de una serie estacionaria implica utilizar la historia disponible de la serie para estimar su valor medio, lo que a su vez se convierte en el pronóstico de períodos futuros. Las técnicas más sofisticadas implican actualizar el estimado conforme se disponga de información nueva. Estas técnicas son útiles cuando los estimados iniciales no son confiables o cuando la estabilidad del promedio está en duda. Además, las técnicas de actualización proporcionan cierto grado de receptividad a los cambios en la estructura subyacente de la serie.

Las técnicas estacionarias de pronóstico se usan siempre que:

- *Las fuerzas que generan una serie se han estabilizado y el entorno donde existe la serie permanece relativamente sin cambios.* Ejemplos de ello son el número de descomposturas por semana en una línea de ensamble que tiene una tasa uniforme de producción, las ventas unitarias de un producto o servicio en la etapa de maduración de su ciclo de vida y el número de ventas resultantes de un nivel constante de esfuerzo.
- *Se requiere un modelo muy simple debido a una falta de datos para la explicación o aplicación.* Un ejemplo es cuando un negocio u organización es nuevo y hay muy pocos datos históricos disponibles.
- *Se puede lograr estabilidad al hacer correcciones simples para factores como el crecimiento de la población o la inflación.* Los ejemplos son el cambio de ingreso a ingreso per cápita o importe de las ventas a valor constante en relación a una fecha determinada.
- *La serie puede convertirse en una serie estable.* Por ejemplo, transformar una serie al tomar logaritmos, raíces cuadradas o diferencias.
- *La serie es un grupo de errores de pronóstico de una técnica de pronóstico que se considera adecuada.* (Vea el ejemplo 3.7 en la página 82.)

Las técnicas que deben considerarse cuando se pronostican series estacionarias incluyen los métodos informales, los de promedio simple, los de promedio móvil y los modelos autorregresivos de promedio móvil (ARMA, del inglés *autoregressive moving average*) y los modelos Box-Jenkins.

### Técnicas de pronóstico para datos con tendencia

Se ha definido a una serie con *tendencia* como una serie de tiempo que contiene un componente de largo plazo que representa el crecimiento o el decremento. En otras palabras, se dice que una serie de tiempo tiene tendencia si su valor promedio cambia a lo largo del tiempo, de forma que se espera que aumente o disminuya durante el periodo para el que se desea hacer el pronóstico. Es común que las series de tiempo económicas contengan tendencia.

Las técnicas de pronóstico para datos con tendencia se usan siempre que:

- *La productividad aumentada y la tecnología nueva lleven a cambios en el estilo de vida.* Los ejemplos pueden incluir la demanda de componentes electrónicos que aumentaron con la llegada de la computadora y el uso del ferrocarril, que disminuyó con el advenimiento del aeroplano.

- *La población creciente genere aumentos en la demanda de bienes y servicios.* Ejemplos de ello son los ingresos por ventas de bienes de consumo, la demanda por consumo de energía y el uso de materias primas.
- *El poder de compra de la moneda afecte las variables económicas, debido a la inflación.* Los ejemplos son los salarios, costos de producción y precios.
- *La aceptación del mercado de un producto aumente.* Un ejemplo es el periodo de crecimiento en el ciclo de vida de un producto nuevo.

Las técnicas que deben considerarse al pronosticar series con tendencia incluyen los modelos de promedios móviles, de suavizamiento exponencial lineal de Holt, de regresión simple, curvas de crecimiento, exponenciales y los modelos autorregresivos integrados de promedio móvil (ARIMA, del inglés *autoregressive integrated moving average*) también conocidos como métodos Box-Jenkins.

### Técnicas de pronóstico para datos estacionales

Ya se dijo anteriormente que una serie estacional es una serie de tiempo con un patrón de cambio que se repite año tras año. Una forma de desarrollar pronósticos implica la selección de un método de descomposición multiplicativo o aditivo y, enseguida, la estimación de los índices estacionales a partir de la historia de la serie. Después, estos índices se utilizan para incluir la estacionalidad en los pronósticos o para eliminar sus efectos en los valores observados. Este último proceso se conoce como ajuste estacional de los datos y se presenta en el [capítulo 5](#).

Las técnicas de pronóstico para datos estacionales se usan siempre que:

- *El clima influya en la variable de interés.* Los ejemplos son el consumo de energía eléctrica, las actividades de verano e invierno (por ejemplo, deportes como el esquí), la ropa y las temporadas de crecimiento agrícola.
- *El calendario anual influya en la variable de interés.* Los ejemplos incluyen las ventas al menudeo influídas por los días festivos, los fines de semana de tres días y los calendarios escolares.

Las técnicas que deben considerarse al pronosticar series estacionales incluyen los modelos de descomposición clásica, censo X-12, de suavización exponencial de Winter, de regresión múltiple y ARIMA (métodos Box-Jenkins).

### Técnicas de pronóstico para series cíclicas

El efecto cíclico ya fue definido como la oscilación alrededor de la tendencia. Es difícil hacer modelos de los patrones cíclicos puesto que es típico que sus patrones no sean estables. Las fluctuaciones ascendentes y descendentes como olas alrededor de las tendencias rara vez se repiten a intervalos fijos, y la magnitud de las fluctuaciones también tiende a variar. Los métodos de descomposición pueden extenderse para el análisis de datos cíclicos ([capítulo 5](#)). Sin embargo, dado el comportamiento irregular de los ciclos, el análisis de un componente cíclico de una serie a menudo requiere que se encuentren indicadores económicos principales o fortuitos.

Las técnicas de pronóstico para datos cíclicos se utilizan siempre que:

- *El ciclo del negocio influya en la variable de interés.* Los ejemplos pueden ser los factores económicos, de mercado y competitivos.
- *Haya cambios en los gustos populares.* Ejemplos de ello son las modas, la música y la comida.
- *Se den cambios en la población.* Los ejemplos son: guerras, hambrunas, epidemias y desastres naturales.
- *Ocurran cambios en el ciclo de vida de un producto.* Los ejemplos son la introducción, crecimiento, maduración y saturación del mercado, y su declive.

Las técnicas que deben considerarse cuando se pronostican series cíclicas incluyen los modelos de descomposición clásica, de indicadores económicos, econométricos, de regresión múltiple y ARIMA (métodos Box-Jenkins).

## Otros factores a considerar cuando se elige una técnica de pronósticos

El horizonte de tiempo para un pronóstico tiene una relación directa con la selección de la técnica para llevarlo a cabo. Para pronósticos de corto y mediano plazos puede aplicarse una variedad de técnicas cuantitativas. Sin embargo, conforme aumenta el horizonte de pronóstico, algunas de estas técnicas se vuelven menos aplicables. Por ejemplo, los promedios móviles, el suavizamiento exponencial y los modelos ARIMA no son buenos para predecir momentos decisivos de la economía, mientras que los modelos econométricos pueden resultar más útiles. Los modelos de regresión son adecuados para los plazos corto, mediano y largo. Las medias, promedios móviles, descomposición clásica y proyecciones de tendencias son técnicas cuantitativas adecuadas para los horizontes de tiempo cortos e intermedios. Las técnicas más complejas de Box-Jenkins y las econométricas también son adecuadas para pronósticos en los términos corto y mediano. Con frecuencia se emplean los métodos cualitativos para horizontes de tiempo más largos (vea el capítulo 10).

Generalmente, el pronosticador basa en su experiencia la aplicabilidad de las técnicas de pronóstico. Asimismo, los administradores requieren pronósticos en un tiempo relativamente corto. El suavizamiento exponencial, la proyección de tendencias, los modelos de regresión y los métodos de descomposición clásica tienen una ventaja en esta situación (vea la tabla 3.6).

Los costos de las computadoras ya son parte fundamental de la selección de la técnica. Las computadoras de escritorio (que utilizan microprocesadores) y el software de pronósticos se están volviendo comunes en muchas organizaciones. Debido a este desarrollo, a menudo otros criterios opacan las consideraciones de costo de la computadora.

Por último, se presentará el pronóstico a la dirección para su autorización y su uso en la toma de decisiones. Por lo tanto, la facilidad de entendimiento e interpretación de los resultados es una consideración importante. Los modelos de regresión, las proyecciones de tendencia, la descomposición clásica y las técnicas de suavizamiento exponencial tienen un buen desempeño en este criterio.

## Evaluación empírica de los métodos de pronóstico

La investigación empírica ha revelado que la precisión de los pronósticos de métodos simples es, por lo menos, tan buena como la de las técnicas estadísticas complejas y sofisticadas.<sup>1</sup> Los resultados de la M3-IJF Competition, en la cual diferentes expertos utilizaron su metodología de pronósticos favorita para generar pronósticos de tres mil tres diferentes series de tiempo, tendieron a apoyar este hallazgo.<sup>2</sup> Aun cuando parecería que cuanto más compleja es una técnica estadística, mejor debería predecir los patrones de series de tiempo. Desafortunadamente, los patrones establecidos de series de tiempo pueden cambiar y en el futuro lo hacen. Así que tener un modelo que represente de la mejor forma los datos históricos (aquellos que los métodos complejos hacen bien) no necesariamente garantiza mayor precisión en las predicciones futuras. Por supuesto, la habilidad del pronosticador también desempeña una función importante en el desarrollo de un buen pronóstico.

La M3-IJF Competition se llevó a cabo en 1997. Los pronósticos producidos por las diversas técnicas de pronósticos se compararon en toda la muestra de tres mil tres series de tiempo y su precisión fue evaluada utilizando diversas medidas de error. El objetivo del estudio de 1997 era verificar las cuatro conclusiones principales de la original *M-Competition* utilizando un grupo de datos más grande.<sup>3</sup> Makridakis y Hibon (2000) resumieron la última competencia de la siguiente forma:

1. Al igual que en el estudio previo, los métodos estadísticamente sofisticados o complejos no necesariamente producen pronósticos más precisos que los métodos más simples.
2. Las diversas medidas de precisión (MAD, MSE, MAPE) producen resultados congruentes cuando se utilizan para evaluar diferentes métodos de pronóstico.

<sup>1</sup> Véase Fildes *et al.* (1997), Makridakis *et al.* (1993) y Makridakis e Hibon (2000).

<sup>2</sup> Véase Makridakis e Hibon (2000).

<sup>3</sup> Véase Makridakis *et al.* (1982).

**TABLA 3.6 Elección de una técnica de pronóstico**

<i>Método</i>	<i>Patrón de datos</i>	<i>Horizonte de tiempo</i>	<i>Tipo de modelo</i>	<i>Requisitos mínimos de datos</i>	
				<i>No estacionales</i>	<i>Estacionales</i>
Informal	ST, T, S	S	TS	1	
Promedios simples	ST	S	TS	30	
Promedios móviles	ST	S	TS	4–20	
Suavizamiento exponencial	ST	S	TS	2	
Suavizamiento exponencial lineal	T	S	TS	3	
Suavizamiento exponencial cuadrático	T	S	TS	4	
Suavizamiento exponencial estacional	S	S	TS		2 × s
Filtración adaptativa	S	S	TS		5 × s
Regresión simple	T	I	C	10	
Regresión múltiple	C, S	I	C	10 × V	
Descomposición clásica	S	S	TS		5 × s
Modelos de tendencia exponencial	T	I, L	TS	10	
Ajuste de curva S	T	I, L	TS	10	
Modelos Gompertz	T	I, L	TS	10	
Curvas de crecimiento	T	I, L	TS	10	
Censo X-12	S	S	TS		6 × s
Box-Jenkins	ST, T, C, S	S	TS	24	3 × s
Indicadores principales	C	S	C	24	
Modelos econométricos	C	S	C	30	
Regresión múltiple con series de tiempo	T, S	I, L	C		6 × s

*Patrones de datos:* ST, estacionarios; T, tendencia; S, estacional; C, cíclico.

*Horizonte de tiempo:* S, corto plazo (menos de tres meses); I, intermedio; L, largo plazo.

*Tipo de modelo:* TS, serie de tiempo; C, causal.

*Estacional:* s, longitud de la estacionalidad.

*Variable:* V, número de variables.

3. La combinación de tres métodos de suavizamiento exponencial supera, en promedio, a los métodos individuales que se combinan y funciona bien en comparación con otros.
4. El desempeño de varios métodos de pronóstico depende de la longitud del horizonte de pronóstico y del tipo de datos analizados (anuales, trimestrales, mensuales). Algunos métodos se desempeñan de manera más precisa para los horizontes cortos, mientras otros son más adecuados para horizontes más largos. Algunos métodos funcionan mejor con datos anuales y otros son más adecuados para datos trimestrales y mensuales.

Como parte de la selección final, cada técnica debe ser evaluada en términos de su confiabilidad y aplicabilidad al problema específico, por su efectividad y precisión comparadas con las técnicas que le compiten y por la aceptación de la administración. La tabla 3.6 resume las técnicas de pronóstico apropiadas para patrones de datos particulares. Esta tabla representa un punto de partida; es decir, muestra métodos a considerar para datos con ciertas características. Finalmente, cualquier método que se elija debe supervisarse constantemente para asegurar que su operación es adecuada para lo que fue planeado.

## MEDICIÓN DEL ERROR DE PRONÓSTICO

Puesto que las técnicas cuantitativas de pronósticos con frecuencia implican datos de series de tiempo, se desarrolló una notación matemática para referirse a cada periodo específico.

La letra  $Y$  se utiliza para indicar una variable de serie de tiempo, a menos que haya más de una variable implicada. El periodo asociado con una observación se muestra como un subíndice. Así,  $Y_t$  se refiere al valor de una serie de tiempo en el periodo  $t$ . Los datos trimestrales de Outboard Marine Corporation que se presentaron en el ejemplo 3.5 (vea la página 73) se indicarían como  $Y_1 = 147.6$ ,  $Y_2 = 251.8$ ,  $Y_3 = 273.1$ , ...,  $Y_{52} = 281.4$ .

La notación matemática también debe ser desarrollada para distinguir entre un valor real de una serie de tiempo y un valor pronosticado. Un  $\hat{\phantom{Y}}$  (sombro) se colocará encima de un valor para indicar que se está pronosticando. El valor de pronóstico para  $Y_t$  es  $\hat{Y}_t$ . La precisión de una técnica de pronóstico con frecuencia se juzgará mediante la comparación de la serie original  $Y_1, Y_2, \dots$  con la serie de valores pronosticados  $\hat{Y}_1, \hat{Y}_2, \dots$

### Notación básica de pronósticos

La notación básica para pronósticos se resume de la siguiente forma:

$Y_t$  = valor de una serie de tiempo en el periodo  $t$

$\hat{Y}_t$  = valor pronosticado de  $Y_t$

$e_t = Y_t - \hat{Y}_t$  = residual o error residual de pronóstico

Se han creado diversos métodos para resumir los errores generados por una técnica de pronósticos específica. La mayoría de estas medidas son el promedio de alguna función de la diferencia entre los valores real y de pronóstico. Estas diferencias se conocen como *residuales*.

Un *residual* es la diferencia entre el valor real y su valor de pronóstico.

Para calcular el error de pronóstico o residual de cada periodo pronosticado se utiliza la ecuación 3.6.

$$e_t = Y_t - \hat{Y}_t \quad (3.6)$$

donde

$e_t$  = error de pronóstico en el periodo  $t$

$Y_t$  = valor real en el periodo  $t$

$\hat{Y}_t$  = valor de pronóstico para el periodo  $t$

Un método para evaluar las técnicas de pronósticos utiliza la suma de los errores absolutos. La desviación absoluta media (*MAD*, del inglés *mean absolute deviation*) mide la precisión del pronóstico al promediar las magnitudes de los errores de pronóstico (valores absolutos de cada error). *MAD* es más útil cuando el analista quiere medir el error de pronóstico en las mismas unidades que la serie original. La ecuación 3.7 muestra cómo se calcula la *MAD*.

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (3.7)$$

El error cuadrático medio (*MSE*, del inglés *mean squared error*) es otro método para evaluar una técnica de pronóstico. Cada error de pronóstico o residual se eleva al cuadrado; luego, se suman y se dividen entre el número de observaciones. Este método penaliza los errores grandes de pronóstico debido a que los errores se elevan al cuadrado, lo cual es importante; una técnica que produce errores moderados podría ser preferible a una que,

por lo general, tiene errores pequeños, pero que en ocasiones produce errores muy grandes. El *MSE* está dado por la ecuación 3.8.

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (3.8)$$

A veces, es más útil calcular los errores de pronóstico en términos de porcentajes en lugar de cantidades. El error porcentual absoluto medio (*MAPE*, del inglés *mean absolute percentage error*) se calcula al encontrar el error absoluto en cada periodo, dividiéndolo entre el valor real observado para ese periodo y luego promediando los errores porcentuales absolutos. Este método es útil cuando el tamaño o magnitud de la variable del pronóstico es importante para evaluar la precisión del pronóstico. El *MAPE* proporciona una indicación de cuán grandes son los errores de pronóstico en comparación con los valores reales de la serie. Esta técnica es especialmente útil cuando los valores de  $Y_t$  son grandes. También puede utilizarse el *MAPE* para comparar la precisión de las mismas o diferentes técnicas en dos series totalmente distintas. La ecuación 3.9 muestra cómo se calcula el *MAPE*.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (3.9)$$

A veces es necesario determinar si un método de pronóstico tiene sesgo (produce pronósticos más altos o más bajos de manera sistemática). En estos casos se usa el error porcentual medio (*MPE*, del inglés *mean percentage error*). Se calcula al encontrar el error en cada periodo y al dividir el resultado entre el valor real para dicho periodo; a continuación, se promedian estos errores porcentuales. Si el método de pronóstico no tiene sesgo, el *MPE* producirá un número cercano a cero. Si el resultado es un alto porcentaje negativo, el método sobreestima de forma consistente, y si el resultado es un porcentaje alto positivo, el método subestima consistentemente. El *MPE* está dado por

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} \quad (3.10)$$

Parte de la decisión de utilizar una técnica específica de pronóstico implica la determinación de si esta técnica producirá errores de pronóstico que se juzguen ser suficientemente pequeños. Se puede esperar que una buena técnica de pronóstico produzca errores de pronóstico relativamente pequeños de forma persistente.

Las cuatro medidas de precisión de pronóstico que recién se han descrito se utilizan para:

- Comparar la precisión de dos (o más) técnicas diferentes.
- Medir la utilidad o confiabilidad de una técnica específica.
- Ayudar a buscar una técnica óptima.

El ejemplo 3.6 ilustra cómo se calcula cada una de estas medidas de error.

#### Ejemplo 3.6

La tabla 3.7 muestra los datos para el número diario de clientes  $Y_t$ , que requieren trabajos de reparación, así como un pronóstico  $\hat{Y}_t$  de tales datos, para Gary's Chevron Station. La técnica de pronóstico utilizó el número de clientes atendidos en el periodo anterior como pronóstico del

**TABLA 3.7 Cálculo de métodos de evaluación de pronóstico del ejemplo 3.6**

Tiempo	Clientes	Pronóstico	Error				
<i>t</i>	<i>Y<sub>t</sub></i>	<i>Ŷ<sub>t</sub></i>	<i>e<sub>t</sub></i>	<i>e<sub>t</sub></i>	<i>e<sub>t</sub><sup>2</sup></i>	<i>e<sub>t</sub></i>  / <i>Y<sub>t</sub></i>	<i>e<sub>t</sub></i> / <i>Y<sub>t</sub></i>
1	58	—	—	—	—	—	—
2	54	58	-4	4	16	.074	-.074
3	60	54	6	6	36	.100	.100
4	55	60	-5	5	25	.091	-.091
5	62	55	7	7	49	.113	.113
6	62	62	0	0	0	.000	.000
7	65	62	3	3	9	.046	.046
8	63	65	-2	2	4	.032	-.032
9	70	63	7	7	49	.100	.100
Totales		—	12	34	188	.556	.162

periodo actual. Esta técnica sencilla se analizará en el capítulo 4. Por medio de *MAD*, *MSE*, *MAPE* y *MPE* se emplearon los siguientes cálculos para evaluar el modelo.

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| = \frac{34}{8} = 4.3$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \frac{188}{8} = 23.5$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} = \frac{.556}{8} = .0695 (6.95\%)$$

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} = \frac{.162}{8} = .0203 (2.03\%)$$

*MAD* indica que cada pronóstico se desvió en un promedio de 4.3 clientes. El *MSE* de 23.5 y el *MAPE* de 6.95% se compararían con los *MSE* y *MAPE* de cualquier otro método que se utilizara para pronosticar estos datos. Por último, el pequeño *MPE* de 2.03% indica que la técnica no tiene sesgo: puesto que el valor es cercano a cero, la técnica ni sobreestima ni subestima consistentemente el número de clientes atendidos diariamente.

## DETERMINACIÓN DE LA IDONEIDAD DE UNA TÉCNICA DE PRONÓSTICO

Antes de pronosticar con una técnica seleccionada, se debe evaluar la idoneidad de la elección. El pronosticador deberá contestar las siguientes preguntas:

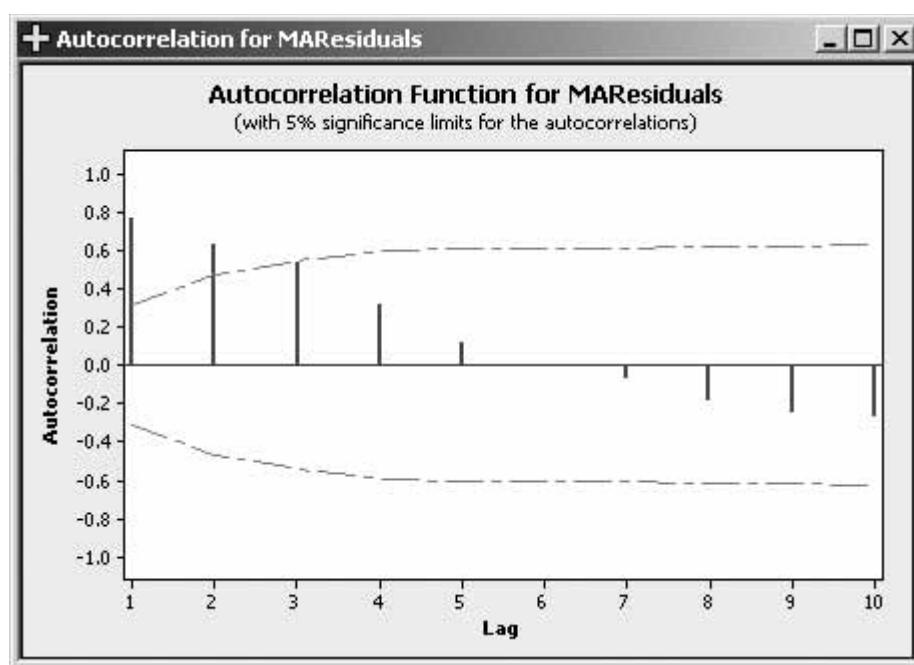
- ¿Los coeficientes de autocorrelación de los residuales indican que existe una serie aleatoria? Esta pregunta puede contestarse al examinar la función de autocorrelación para los residuales, tal como la que se demuestra en el [ejemplo 3.7](#).
- ¿Se distribuyen casi normalmente los residuales? Esta pregunta puede contestarse al analizar un histograma de los residuales o una gráfica de probabilidad normal.
- ¿Son significativas las estadísticas *t* para todos los valores estimados de los parámetros? La distribución *t* se revisa en el [capítulo 2](#) y las aplicaciones de las estadísticas *t* se demuestran en el [ejemplo 3.2](#) (vea las páginas 65 y 66) y en los [capítulos 6 al 9](#).
- ¿Quienes toman las decisiones entienden la técnica y la emplean con facilidad?

El requisito básico de que el patrón residual sea aleatorio se verifica al examinar los coeficientes de autocorrelación de los residuales. No debería haber coeficientes de autocorrelación importantes. El ejemplo 3.2 ilustró cómo puede utilizarse una función de autocorrelación para determinar si una serie es aleatoria. La estadística  $Q$  de Ljung-Box también se utiliza para probar que las autocorrelaciones para todos los retrasos, hasta el retraso  $m$ , son iguales a cero. El ejemplo 3.7 ilustra este procedimiento con los residuales de dos modelos ajustados.

### Ejemplo 3.7

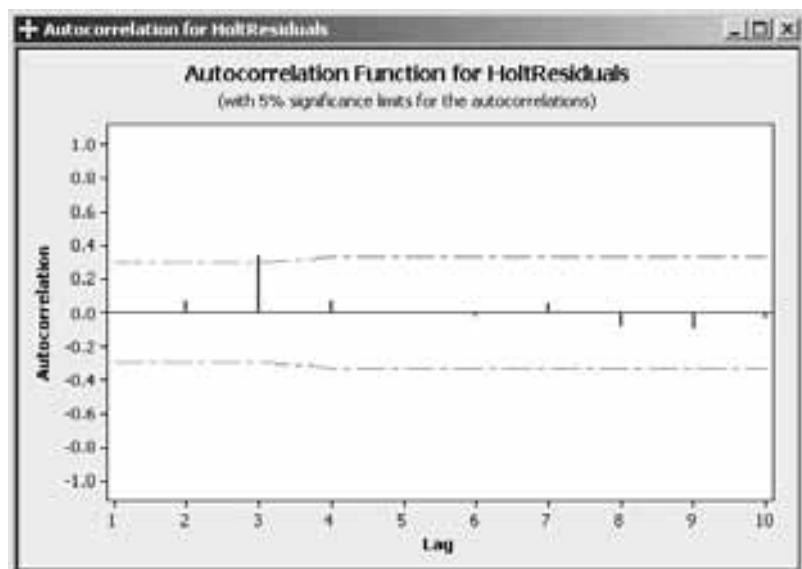
A Maggie Trymane, analista de Sears, le solicitaron pronosticar las ventas para el 2001. Los datos del periodo 1955-2000 se muestran en la tabla 3.4. Primero, ella trata de pronosticar los datos mediante un promedio móvil de cinco períodos (vea el capítulo 4 para una descripción de esta técnica). Los residuales, diferencia entre los valores reales y los pronósticos de ajustados se calculan y se almacenan. Los coeficientes de autocorrelación para estos residuales se muestran en la figura 3.16. Un examen de tales coeficientes de autocorrelación indica que dos de ellos son significativamente diferentes de cero,  $r_1 = 0.77$  y  $r_2 = 0.63$ . Los coeficientes de autocorrelación importantes indican cierta asociación o patrón en los residuales. Además, la función de autocorrelación tiene, en sí misma, un patrón de coeficientes que decrecen suavemente. Al examinar las primeras diez autocorrelaciones como grupo, la estadística  $Q$  de 10 retrasos es de 72.26, mucho mayor que el valor 18.3 de una variable chi cuadrada con 10 grados de libertad, para una significancia de 0.05. La hipótesis de que las primeras diez autocorrelaciones son congruentes con las de una serie aleatoria se rechaza claramente en el nivel de 5%. Maggie considera que la técnica de promedio móvil de cinco períodos es inadecuada, debido a que uno de los requisitos básicos para una técnica de pronósticos es que proporcione una serie de residuales o errores esencialmente aleatorios.

**FIGURA 3.16 Función de autocorrelación para residuales con un patrón**



**Autocorrelation Function: MAResiduals**

Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0.768208	4.92	26.01	6	-0.000650	-0.00	62.36
2	0.628865	2.73	43.89	7	-0.076490	-0.25	62.66
3	0.531352	1.97	56.99	8	-0.184331	-0.61	64.48
4	0.313974	1.07	61.68	9	-0.248870	-0.81	67.89
5	0.117308	0.39	62.36	10	-0.277090	-0.89	72.26

**Autocorrelation Function: HoltResiduals**

Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	-0.010293	-0.07	0.01	6	-0.028641	-0.17	6.41
2	0.072582	0.49	0.27	7	0.055955	0.34	6.59
3	0.336058	2.27	6.07	8	-0.081304	-0.49	6.97
4	0.075102	0.46	6.37	9	-0.098068	-0.59	7.55
5	-0.001224	-0.01	6.37	10	-0.036649	-0.22	7.63

**FIGURA 3.17 Función de autocorrelación para residuales que son esencialmente aleatorios**

Ahora, Maggie prueba con el suavizamiento exponencial lineal de Holt (vea el capítulo 4 para una descripción de esta técnica). La función de autocorrelación para la serie residual generada por esta técnica se muestra en la figura 3.17. Un examen de estos coeficientes de autocorrelación indica que solamente el coeficiente de autocorrelación residual del retraso 3 —que es de 0.34— es significativamente diferente de cero (en el nivel de 5%). También se examina la estadística  $Q$  para 10 retrasos de tiempo. El valor de LBQ de 7.63 en los resultados de Minitab es menor que el valor 15.5 de una variable chi cuadrada con ocho grados de libertad, para un nivel de significancia de 0.05. (En este caso, los grados de libertad son iguales al número de retrasos que deben probarse menos el número de parámetros en el modelo de suavizamiento exponencial lineal que se han ajustado a los datos.) Aunque la autocorrelación residual del retraso 3 de alguna manera es grande, como grupo, las primeras diez autocorrelaciones residuales no son distintas de aquellas para una serie totalmente aleatoria. Maggie decide considerar la técnica de suavizamiento exponencial lineal de Holt como un modelo posible para pronosticar los ingresos operativos de Sears para el 2001.

## APLICACIÓN A LA ADMINISTRACIÓN

Los conceptos en este capítulo proporcionan la base para seleccionar una técnica de pronósticos conveniente en una situación determinada. En los capítulos siguientes, muchas de las técnicas de pronósticos más importantes se analizan y se aplican en situaciones de pronóstico. Es importante observar que, en muchas situaciones prácticas, más de un método o modelo podría producir pronósticos aceptables y casi imposibles de distinguir. De hecho,

es una buena práctica probar con varias técnicas de pronóstico razonables. A menudo debe utilizarse la apreciación, con base en la facilidad de uso, costos, condiciones ambientales externas y demás, a fin de seleccionar un grupo específico de pronósticos de entre, por ejemplo, dos grupos de valores casi imposibles de distinguir.

Los siguientes son algunos ejemplos de situaciones que surgen constantemente en los negocios, para las cuales una técnica sensata de pronóstico ayudaría en la toma de decisiones.

- Una compañía de bebidas refrescantes quiere proyectar la demanda mensual de su producto principal durante los dos próximos años.
- Una importante empresa de telecomunicaciones desea pronosticar los pagos trimestrales de dividendos de su más grande competidor en los próximos tres años.
- Una universidad necesita hacer un pronóstico trimestral de las horas-crédito por estudiante para los siguientes cuatro años, a fin de desarrollar proyecciones de presupuestos para la legislatura estatal.
- Una empresa de contabilidad pública necesita pronósticos mensuales de los importes de facturación a fin de poder planear nuevos puestos de contabilidad y comenzar a reclutar.
- El gerente de control de calidad de una fábrica que produce lingotes de aluminio necesita un pronóstico semanal de los defectos de producción para la dirección de la empresa.
- Un banquero quiere ver proyectados los ingresos mensuales de un pequeño fabricante de bicicletas que busca obtener un gran crédito para triplicar su capacidad de producción.
- Con el propósito de hacer recomendaciones de reglamentación, una agencia del gobierno federal necesita las proyecciones anuales de las millas promedio por galón de los automóviles que se fabricarán en Estados Unidos durante los próximos 10 años.
- El gerente de personal de cierta compañía necesita el pronóstico mensual de inasistencia de la fuerza laboral para planear los gastos por horas extras.
- En un intento por evitar la bancarrota, una empresa de ahorro y crédito necesita un pronóstico de los créditos atrasados para los próximos dos años.
- A fin de planear su presupuesto de investigación y desarrollo, una compañía que fabrica chips de computadora necesita un pronóstico del número de computadoras personales que se venderán en los próximos cinco años.

## Glosario

**Autocorrelación.** Correlación que existe entre una variable retrasada uno o más períodos consigo misma.

**Componente cíclico.** Fluctuación parecida a olas alrededor de la tendencia.

**Componente estacional.** Patrón de cambios que se repite año tras año.

**Correlograma o función de autocorrelación.** Gráfica de las autocorrelaciones para varios retrasos de una serie de tiempo.

**Corte transversal.** Observaciones recolectadas en un único periodo.

**Residual.** Diferencia entre el valor real y su valor de pronóstico.

**Serie de tiempo.** Consiste en datos que se recopilan, registran u observan a lo largo de incrementos sucesivos de tiempo.

**Serie estacionaria.** Aquella cuyas particulares estadísticas fundamentales, como la media y la varianza, permanecen constantes a lo largo del tiempo.

**Tendencia.** Componente de largo plazo que representa el crecimiento o decremento en las series de tiempo durante un amplio periodo.

## Fórmulas clave

### Coeficiente de autocorrelación del $k$ -ésimo orden

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (3.1)$$

**Error estándar del coeficiente de autocorrelación**

$$SE(r_k) = \sqrt{\frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{n}} \quad (3.2)$$

**Estadística  $Q$  Ljung-Box**

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2}{n - k} \quad (3.3)$$

**Modelo aleatorio**

$$Y_t = c + \varepsilon_t \quad (3.4)$$

**Estadística para probar la significancia de la autocorrelación del retraso 1**

$$t = \frac{r_1}{SE(r_1)} \quad (3.5)$$

**Error de pronóstico o residual**

$$e_t = Y_t - \hat{Y}_t \quad (3.6)$$

**Desviación media absoluta**

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (3.7)$$

**Error cuadrático medio**

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (3.8)$$

**Error porcentual absoluto medio**

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \quad (3.9)$$

**Error porcentual medio**

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} \quad (3.10)$$

**Problemas**

1. Explique las diferencias entre las técnicas cualitativas y cuantitativas de pronóstico.
2. ¿Qué es una serie de tiempo?
3. Describa cada uno de los componentes en una serie de tiempo.
4. ¿Qué es una autocorrelación?
5. ¿Qué mide un coeficiente de autocorrelación?

6. Describa cómo se usan los correlogramas para analizar autocorrelaciones de varios retrasos de una serie de tiempo.
7. Cada una de las siguientes oraciones describe una serie estacionaria o no estacionaria. Indique cuál es en cada caso.
  - a) Una serie que tiene tendencia.
  - b) Una serie cuya media y varianza permanecen constantes a lo largo del tiempo.
  - c) Una serie cuyo valor medio cambia a lo largo del tiempo.
  - d) Una serie que no contiene ni crecimiento ni decremento.
8. Se dieron descripciones para varios tipos de series: aleatorias, estacionarías, con tendencia y estacionales. Identifique cada tipo de serie.
  - a) Una serie que tiene propiedades estadísticas básicas, tales como media y varianza, que permanecen constantes a lo largo del tiempo
  - b) Los valores sucesivos de una serie de tiempo no se relacionan unos con otros.
  - c) Existe una alta relación entre cada valor sucesivo de una serie.
  - d) En el retraso de tiempo 4, para el caso de datos trimestrales, aparece un coeficiente de autocorrelación importante.
  - e) Una serie que no contiene ni crecimiento ni decremento.
  - f) Es típico que los coeficientes de autocorrelación sean significativamente distintos de cero para los primeros retrasos de tiempo y después disminuyan hacia cero, de manera gradual, conforme aumenta el número de retrasos.
9. Haga una lista de algunas técnicas de pronósticos que deben considerarse cuando se hacen pronósticos de series estacionarías. Dé ejemplos de situaciones en las cuales estas técnicas serían aplicables.
10. Elabore una lista con algunas de las técnicas de pronóstico que deben considerarse cuando se pronostica una serie con tendencia. Dé ejemplos de situaciones en las cuales estas técnicas serían aplicables.
11. Haga una lista de algunas de las técnicas de pronóstico que deberían considerarse al hacer pronósticos de series estacionales. Dé ejemplos de situaciones en las cuales estas técnicas serían aplicables.
12. Elabore una lista con algunas de las técnicas de pronóstico que deben considerarse cuando se trata de una serie cíclica. Dé ejemplos de situaciones en las que estas técnicas sean aplicables.
13. El número de matrimonios en Estados Unidos se encuentra en la tabla P.13. Calcule las primeras diferencias para estos datos. Haga una gráfica de los datos originales y de los datos diferenciados como una serie de tiempo. ¿Hay tendencia en alguna de estas series? Haga un análisis.

**TABLA P.13**

<i>Año</i>	<i>Matrimonios</i>	<i>Año</i>	<i>Matrimonios</i>
1985	2413	1992	2362
1986	2407	1993	2334
1987	2403	1994	2362
1988	2396	1995	2336
1989	2403	1996	2344
1990	2443	1997	2384
1991	2371	1998	2244

Fuente: *Statistical Abstract of the United States*, 2002.

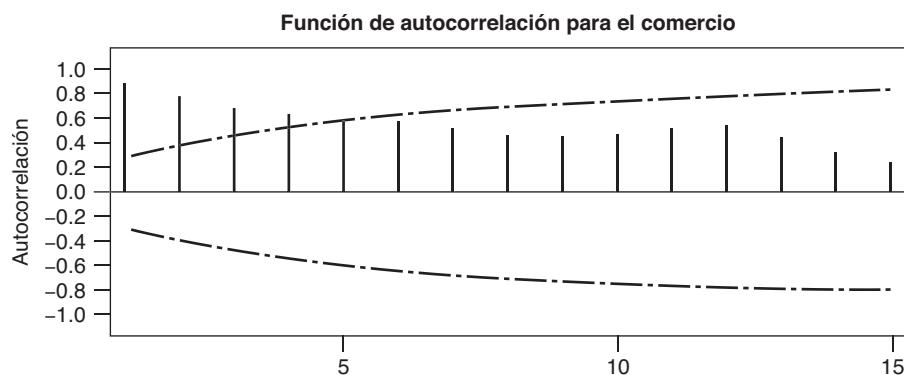
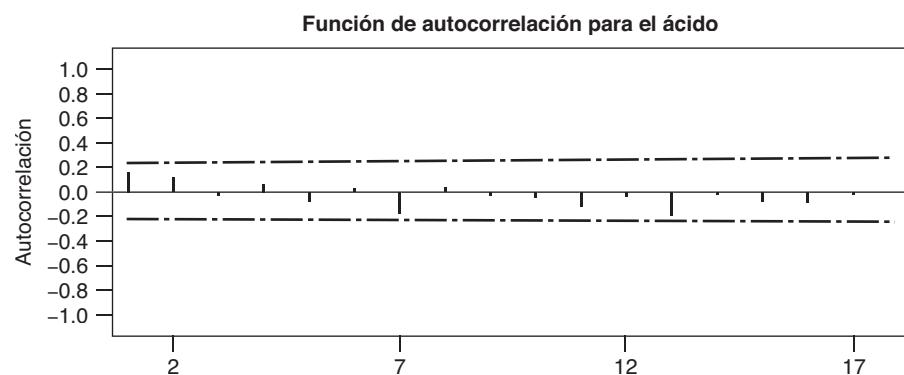
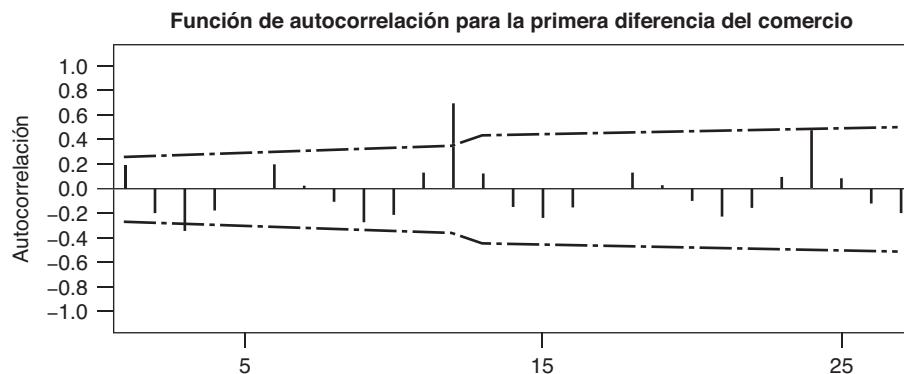
14. Calcule el intervalo de confianza de 95% para el coeficiente de autocorrelación para el retraso de tiempo 1 de una serie que contiene 80 períodos.
15. ¿Qué medida de precisión de pronóstico debe utilizarse en cada una de las siguientes situaciones?
  - a) El analista necesita determinar si el método de pronóstico tiene sesgo.
  - b) El analista cree que el tamaño o la magnitud de la variable de pronóstico es importante para evaluar la precisión del pronóstico.
  - c) El analista necesita penalizar los errores grandes de pronóstico.
16. ¿Cuál de las siguientes declaraciones es cierta en lo relativo a las medidas de precisión utilizadas para evaluar pronósticos?
  - a) El *MAPE* toma en consideración la magnitud de los valores que se pronostican.
  - b) El *MSE* penaliza los errores grandes.
  - c) El *MPE* se usa para determinar si un modelo predice demasiado alto o demasiado bajo de forma sistemática.
  - d) La ventaja del método *MAD* es que relaciona el tamaño del error con la observación real.
17. Allie White, funcionario en jefe de créditos de Dominion Bank, desea analizar la cartera de créditos del banco para el periodo 1998-2003. Los datos se muestran en la tabla P.17.
  - a) Calcule las autocorrelaciones para los retrasos de tiempo 1 y 2. Realice la prueba para determinar si estos coeficientes de autocorrelación son significativamente diferentes de cero en el nivel de significancia 0.05.
  - b) Utilice un software para hacer una gráfica de los datos y calcule las autocorrelaciones de los primeros seis retrasos de tiempo. ¿Es estacionaria esta serie de tiempo?
18. Esta pregunta se refiere al problema 17. Calcule las primeras diferencias de los datos trimestrales de los créditos de Dominion Bank.
  - a) Calcule el coeficiente de autocorrelación para el retraso de tiempo 1; emplee los datos diferenciados.
  - b) Utilice un software para hacer una gráfica de los datos diferenciados y para calcular las autocorrelaciones de los datos diferenciados para los primeros seis retrasos de tiempo. ¿Es estacionaria esta serie?
19. Analice los coeficientes de autocorrelación mostrados en las figuras 3.18 hasta la 3.21. Describa brevemente cada serie.
20. A un analista le gustaría determinar si existe un patrón en las ganancias por acción para la Price Company, la cual operaba un negocio de mayoreo-menudeo en varios estados bajo el nombre de Price Club. Los datos se muestran en la tabla P.20.  
Describa los patrones que existen en estos datos.

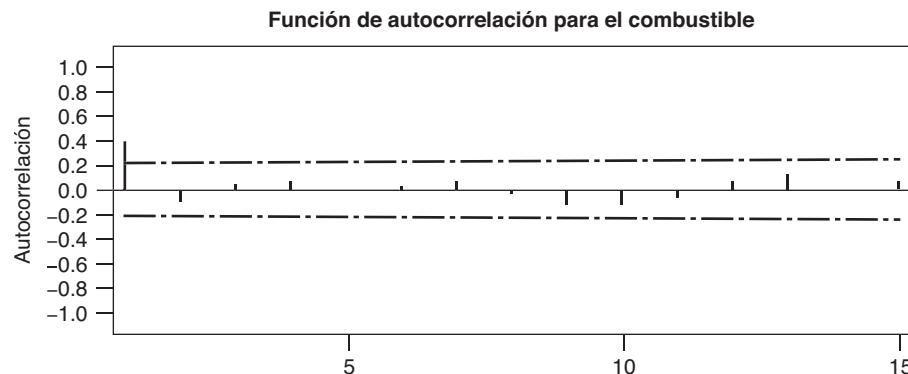
**TABLA P.17 Créditos trimestrales otorgados por Dominion Bank 1998-2003 (millones de \$)**

Calendario	31 mar.	30 jun.	30 sep.	31 dic.
1998	2313	2495	2609	2792
1999	2860	3099	3202	3161
2000	3399	3471	3545	3851
2001	4458	4850	5093	5318
2002	5756	6013	6158	6289
2003	6369	6568	6646	6861

Fuente: registros de Dominion Bank.

- a) Encuentre el valor de pronóstico para las ganancias trimestrales por acción de Price Club para cada trimestre mediante un método informal (por ejemplo, el pronóstico del primer trimestre de 1987 es el valor de las ganancias del cuarto trimestre de 1986, es decir 0.32).
- b) Evalúe el pronóstico informal utilizando  $MAD$ .
- c) Evalúe el pronóstico informal por medio de  $MSE$ .
- d) Evalúe el pronóstico informal utilizando  $MAPE$ .
- e) Evalúe el pronóstico informal; use  $MPE$ .
- f) Describa sus hallazgos en un resumen.

**FIGURA 3.18 Función de autocorrelación para el problema 19****FIGURA 3.19 Función de autocorrelación para el problema 19****FIGURA 3.20 Función de autocorrelación para el problema 19**

**FIGURA 3.21 Función de autocorrelación para el problema 19**

**TABLA P.20 Ganancias trimestrales por acción para Price Company: 1986-1993 Trimestres**

<i>Calendario</i>	<i>1o.</i>	<i>2o.</i>	<i>3o.</i>	<i>4o.</i>
1986	.40	.29	.24	.32
1987	.47	.34	.30	.39
1988	.63	.43	.38	.49
1989	.76	.51	.42	.61
1990	.86	.51	.47	.63
1991	.94	.56	.50	.65
1992	.95	.42	.57	.60
1993	.93	.38	.37	.57

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1994), p. 1646.

**TABLA P.21 Ventas semanales de un artículo de comida**

(Leer transversalmente)

2649.9	2898.7	2897.8	3054.3	3888.1	3963.6	3258.9	3199.6
3504.3	2445.9	1833.9	2185.4	3367.4	1374.1	497.5	1699.0
1425.4	1946.2	1809.9	2339.9	1717.9	2420.3	1396.5	1612.1
1367.9	2176.8	2725.0	3723.7	2016.0	862.2	1234.6	1166.5
1759.5	1039.4	2404.8	2047.8	4072.6	4600.5	2370.1	3542.3
2273.0	3596.6	2615.8	2253.3	1779.4	3917.9	3329.3	1864.4
3318.9	3342.6	2131.9	3003.2				

21. La tabla P.21 contiene las ventas mensuales, durante 52 semanas consecutivas, de un producto alimentario.
- Utilice una computadora para hacer una gráfica de los datos de ventas como una serie de tiempo.
  - ¿Cree que esta serie es estacionaria o no estacionaria?
  - Utilice Minitab o un programa similar y calcule las autocorrelaciones de las series de ventas para los primeros 10 retrasos de tiempo. ¿Es coherente el comportamiento de la autocorrelación con su elección en el inciso b)? Explique su respuesta.

22. Esta pregunta se refiere al problema 21.
- Ajuste el modelo aleatorio dado por la ecuación 3.4 a los datos de la tabla P.21 al estimar  $c$  con una media de la muestra  $\bar{Y}$  de manera que  $\hat{Y}_t = \bar{Y}$ . Calcule los residuales.

$$e_t = Y_t - \hat{Y}_t = Y_t - \bar{Y}.$$

- Utilice Minitab o un programa semejante y calcule las autocorrelaciones de los residuales para los primeros 10 retrasos de tiempo. ¿Es adecuado el modelo aleatorio para los datos de ventas? Explique su respuesta.

## CASOS

---

### CASO 3-1A

### MURPHY BROTHERS FURNITURE

En 1958, los hermanos Murphy fundaron una tienda de muebles en el centro de Dallas, Texas. Con el paso de los años tuvieron mucho éxito y ampliaron su negocio a lo largo del oeste y medio-oeste de Estados Unidos. Para 1996, su cadena de mueblerías estaba bien establecida en 36 estados.

Julie Murphy, hija de uno de los fundadores, recientemente se incorporó a la compañía. Su padre y su tío eran diestros en muchas cosas, pero no en el área de habilidades cuantitativas. En lo particular, ambos sentían que no podían pronosticar con precisión las ventas futuras de la empresa por medio de técnicas computarizadas modernas. Por esta razón, acudieron a Julie para que les ayudara con esto como parte de su nuevo empleo.

Al principio, Julie consideró utilizar los importes de las ventas de las mueblerías como la variable, pero descubrió que faltaban varios años de la historia de ventas. Le preguntó a su padre, Glen, acerca de esto pero él dijo que en ese momento “no pensó que fuera importante”. Ella le explicó la importancia de los datos anteriores y él indicó que conservaría los datos futuros.

Julie concluyó que las ventas de Murphy probablemente eran parecidas a las cifras nacionales de ventas y comenzó a buscar una variable apropiada en alguna de las muchas publicaciones federales. Después de bus-

car en un número reciente de *Survey of Current Business*, encontró la historia de las ventas mensuales de todas las tiendas al menudeo en Estados Unidos y decidió utilizar esta variable como un sustituto de la variable que le interesaba; es decir, el importe de las ventas de Murphy Stores. Ella pensó que si podía establecer pronósticos precisos para las ventas nacionales, también podría relacionarlos con las ventas de Murphy y de esta manera obtener los pronósticos deseados.

La tabla 3.8 indica los datos que Julie recopiló y la figura 3.22 muestra el diagrama de datos proporcionados por el programa de su computadora. Con esto, comenzó su análisis mediante la computadora para desarrollar una gráfica de los coeficientes de autocorrelación.

Después de examinar la función de autocorrelación producida en la figura 3.23, le resultó obvio que sus datos reflejaban una tendencia. Los coeficientes tempranos de autocorrelación eran bastante grandes y tenían la tendencia de disminuir gradualmente hasta cero. Para que la serie fuera estacionaria, de manera que varios métodos de pronóstico pudieran considerarse, Julie decidió primero diferenciar sus datos para ver si podía eliminarse la tendencia. La función de autocorrelación para los primeros datos diferenciados se muestra en la figura 3.24.■

---

## PREGUNTAS

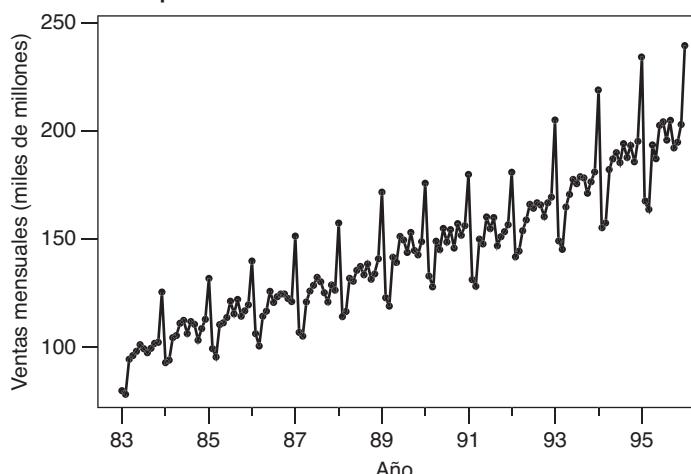
- ¿Qué debería concluir Julie acerca de las series de ventas al menudeo?
- ¿Ha progresado en cuanto a encontrar una técnica de pronóstico?
- ¿Qué técnicas de pronóstico debería poner a prueba?
- ¿Cómo sabrá qué técnica funciona mejor?

**TABLA 3.8 Ventas mensuales (miles de millones de \$) para todas las ventas al menudeo, 1983-1995, para el caso 3.1 A**

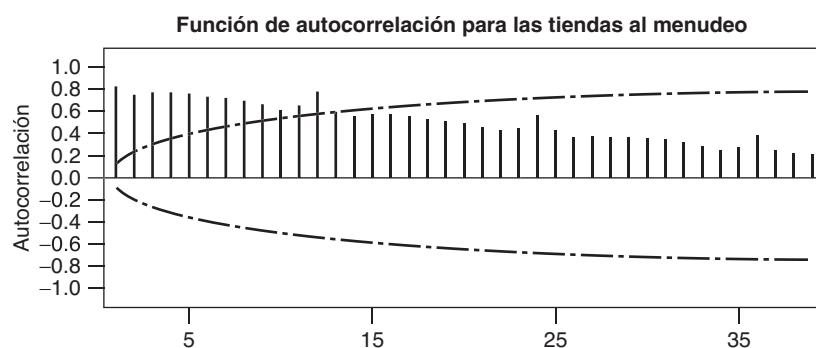
	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
Ene.	81.3	93.1	98.8	105.6	106.4	113.6	122.5	132.6	130.9	142.1	148.4	154.6	167.0
Feb.	78.9	93.7	95.6	99.7	105.8	115.0	118.9	127.3	128.6	143.1	145.0	155.8	164.0
Mar.	93.8	103.3	110.2	114.2	120.4	131.6	141.3	148.3	149.3	154.7	164.6	184.2	192.1
Abr.	93.8	103.9	113.1	115.7	125.4	130.9	139.8	145.0	148.5	159.1	170.3	181.8	187.5
May.	97.8	111.8	120.3	125.4	129.1	136.0	150.3	154.1	159.8	165.8	176.1	187.2	201.4
Jun.	100.6	112.3	115.0	120.4	129.0	137.5	149.0	153.5	153.9	164.6	175.7	190.1	202.6
Jul.	99.4	106.9	115.5	120.7	129.3	134.1	144.6	148.9	154.6	166.0	177.7	185.8	194.9
Ago.	100.1	111.2	121.1	124.1	131.5	138.7	153.0	157.4	159.9	166.3	177.1	193.8	204.2
Sep.	97.9	104.0	113.8	124.4	124.5	131.9	144.1	145.6	146.7	160.6	171.1	185.9	192.8
Oct.	100.7	109.6	115.8	123.8	128.3	133.8	142.3	151.5	152.1	168.7	176.4	189.7	194.0
Nov.	103.9	113.5	118.1	121.4	126.9	140.2	148.8	156.1	155.6	167.2	180.9	194.7	202.4
Dic.	125.8	132.3	138.6	152.1	157.2	171.0	176.5	179.7	181.0	204.1	218.3	233.3	238.0

Fuente: Survey of Current Business, varios años.

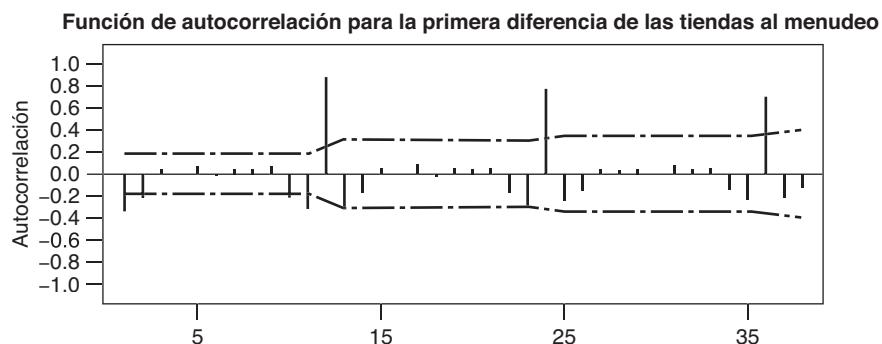
**Ventas mensuales para todas las tiendas al menudeo en Estados Unidos: 1983-1995**



**FIGURA 3.22 Gráfica de las series de tiempo en las ventas mensuales de las tiendas al menudeo en Estados Unidos**



**FIGURA 3.23 Función de autocorrelación para las ventas mensuales en todas las tiendas al menudeo de Estados Unidos**



**FIGURA 3.24 Función de autocorrelación para la primera diferencia de todas las tiendas al menudeo de Estados Unidos**

**TABLA 3.9 Ventas mensuales de Murphy Brothers Furniture, 1992–1995, para el caso 3.1 B**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
1992	4906	5068	4710	4792	4638	4670	4574	4477	4571	4370	4293	3911
1993	5389	5507	4727	5030	4926	4847	4814	4744	4844	4769	4483	4120
1994	5270	5835	5147	5354	5290	5271	5328	5164	5372	5313	4924	4552
1995	6283	6051	5298	5659	5343	5461	5568	5287	5555	5501	5201	4826

Fuente: Registros de ventas mensuales de Murphy Brothers.

## CASO 3-1B MURPHY BROTHERS FURNITURE

Glen Murphy no estaba muy contento por haber sido regañado por su hija, así que emprende una búsqueda exhaustiva en los registros de la compañía. Después de iniciar la investigación, está feliz de haber descubierto datos sobre las ventas para el periodo que va de 1992 a 1995, como se muestra en la tabla 3.9. Pero se sorprende

al descubrir que Julie no comparte su entusiasmo. Ella sabe que haber adquirido datos reales sobre las ventas de los últimos cuatro años es un hecho positivo. Su problema es que no está segura sobre lo que debe hacer con los datos recién adquiridos.■

### PREGUNTAS

1. ¿Qué debe concluir Julie acerca de estos datos de las ventas de Murphy Brothers?
2. ¿Cómo se compara el patrón en las ventas reales con el patrón de los datos en las ventas al menudeo que se presentaron en el caso 3.1 A?
3. ¿Qué datos deberá utilizar para desarrollar un modelo de pronóstico?

## CASO 3-2 MR. TUX

John Mosby, propietario de varias tiendas de alquiler Mr. Tux, está comenzando a pronosticar su variable más importante del negocio, el importe de las ventas mensua-

les (vea los casos de Mr. Tux que aparecen al final de los capítulos 1 y 2). Una de sus empleadas, Virginia Perot, recopiló los datos de ventas que se muestran en

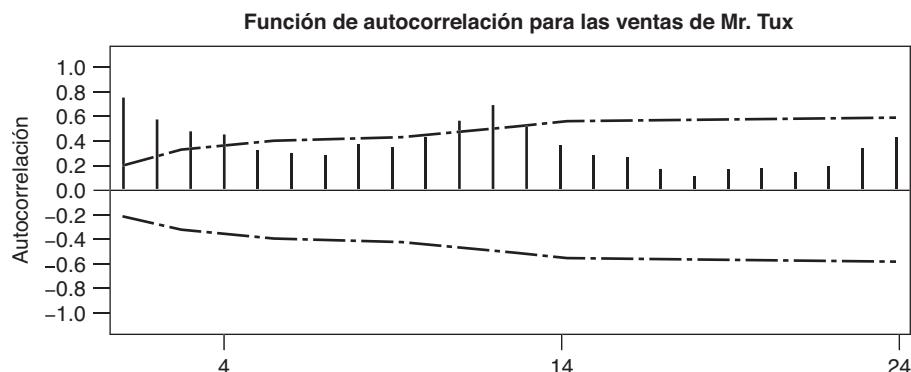


FIGURA 3.25 Función de autocorrelación para los datos de Mr. Tux

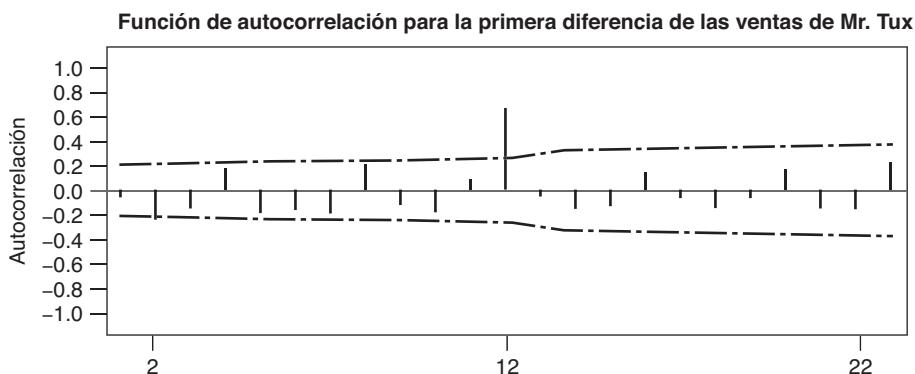


FIGURA 3.26 Función de autocorrelación para la primera diferencia de datos de Mr. Tux

el caso del capítulo 2. John decide utilizar los datos que se han recopilado de 96 meses. Corre los datos en Minitab y obtiene el patrón de coeficiente de autocorrelación que se muestra en la figura 3.25. Debido a que todos los coeficientes de autocorrelación son positivos y se desenvuelven muy lentamente, John concluye que sus datos tienen una tendencia.

A continuación, John pide al programa que calcule las primeras diferencias de los datos. La figura 3.26 muestra los coeficientes de autocorrelación para los datos diferenciados. El coeficiente de autocorrelación para el retraso de tiempo 12,  $r_{12} = 0.68$  es significativamente distinto de cero.

Por último, John utiliza otro software para calcular el porcentaje de la varianza en los datos originales, explicados por la tendencia y los componentes estacionales y aleatorios. El programa calcula el porcentaje de la varianza en los datos originales, explicados por los factores en el análisis:

FACTOR	% EXPLICADO
Datos	100
Tendencia	6
Estacional	45
Aleatorio	49

#### PREGUNTAS

- Resuma los resultados del análisis de John en un párrafo, de manera que los entienda un administrador y no un pronosticador.
- Describa la tendencia y los efectos estacionales que parecen estar presentes en los datos de las ventas de Mr. Tux.

**TABLA 3.10 Número de clientes de CCC desde enero de 1985 a marzo de 1993**

	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>May.</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sep.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
1985	182	136	99	77	75	63	87	73	83	82	74	75
1986	102	121	128	128	112	122	104	108	97	141	97	87
1987	145	103	113	150	100	131	96	92	88	118	102	98
1988	101	153	138	107	100	115	78	106	94	93	103	104
1989	150	102	151	100	100	98	97	120	98	135	141	67
1990	127	146	175	110	153	117	121	121	131	147	121	110
1991	171	185	172	168	142	152	151	141	128	151	121	126
1992	166	138	175	108	112	147	168	149	145	149	169	138
1993	152	151	199									

3. ¿Cómo explicaría el renglón “49% aleatorio”?  
 4. Suponga que las autocorrelaciones  $r_{24}$  y  $r_{36}$  son significativas, ¿concluiría usted que la estacionalidad

tiene una tendencia? Si es así, explique lo que se quiere decir por tendencia en la estacionalidad.

## CASO 3-3 CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se describió en el capítulo 1 (caso 1.2).

Marv Harnishfeger, director ejecutivo, estaba preocupado por las dimensiones y la programación del personal para el resto de 1993. Explicó el problema a Dorothy Mercer, recientemente elegida presidenta del comité ejecutivo. Dorothy meditó el problema y concluyó que CCC necesitaba analizar el número de nuevos clientes que adquiría cada mes. Dorothy estaba familiarizada con diversas técnicas de exploración de datos y estuvo de acuerdo

en analizar el problema, pues antes había trabajado para una empresa de servicios públicos. Entonces, le pidió a Marv que le proporcionara datos mensuales acerca del número de nuevos clientes vistos. Marv le proporcionó los datos mensuales que se muestran en la tabla 3.10 correspondientes al número de clientes nuevos de CCC para el periodo comprendido entre enero de 1985 y marzo de 1993.

Dorothy analizó estos datos mediante una gráfica de series de tiempo y un análisis de autocorrelación.■

### PREGUNTAS

- Explique la manera en que Dorothy utilizó el análisis de autocorrelación para explorar el patrón de los datos para el número de nuevos clientes vistos por CCC.
- ¿Qué concluyó ella al terminar su análisis?
- ¿Qué tipo de técnica de pronóstico recomendó Dorothy para este conjunto de datos?

## CASO 3-4 ALOMEGA FOOD STORES

En el ejemplo 1.1 la presidenta de Alomega, Julie Ruth, había recopilado datos acerca de las operaciones de su compañía. Encontró datos de ventas de algunos meses

con diversas variables explicativas posibles (revise esta situación en el ejemplo 1.1). Mientras su equipo de análisis trabajaba con los datos en un intento por pro-

**TABLA 3.11 Ventas mensuales para 27 tiendas Alomega Food Stores, 1999-2002, para el caso 3.4**

Mes	1999	2000	2001	2002
Ene.	425,075	629,404	655,748	455,136
Feb.	315,305	263,467	270,483	247,570
Mar.	432,101	468,612	429,480	732,005
Abr.	357,191	313,221	260,458	357,107
May.	347,874	444,404	528,210	453,156
Jun.	435,529	386,986	379,856	320,103
Jul.	299,403	414,314	472,058	451,779
Ago.	296,505	253,493	254,516	249,482
Sep.	426,701	484,365	551,354	744,583
Oct.	329,722	305,989	335,826	421,186
Nov.	281,783	315,407	320,408	397,367
Dic.	166,391	182,784	276,901	269,096

nosticar las ventas mensuales, ella se impacientó y se preguntó qué variables explicativas eran las mejores para este propósito.

En el caso 2.3, Julie investigó la relación entre las ventas y las posibles variables explicativas. Ahora se da

cuenta de que este paso fue prematuro, puesto que no conoce siquiera el patrón de datos de sus ventas (vea la tabla 3.11).■

## PREGUNTAS

1. ¿Cuál fue la conclusión de Julie acerca del patrón de datos de las ventas de Alomega?

## Aplicaciones en Minitab

**El problema.** En el ejemplo 3.4, Maggie Trymane, analista de Sears, desea hacer un pronóstico de las ventas para 2001. Es necesario determinar el patrón para los datos de ventas del periodo 1995-2000.

### Solución con Minitab

1. Introduzca los datos de Sears que se muestran en la tabla 3.4 en la columna C1. Puesto que los datos ya están almacenados en un archivo llamado Tab3-4.MTW, dé clic en

File>Open Worksheet

y obtendrá el cuadro de diálogo que se muestra en la figura 3.27. Localice el archivo que desea abrir e introduzca el nombre, en este caso, Tab3-4.MTW, a la derecha del nombre File (archivo) y dé clic en Open (abrir). Aparecerán los datos de Sears en C1.

2. Para construir una función de autocorrelación, dé clic en los siguientes menús, como se muestra en la figura 3.28:

Stat>Time Series>Autocorrelation

3. Aparece el cuadro de diálogo de la función de autocorrelación que se muestra en la figura 3.29.

a) Dé doble clic en la variable *Revenue* (ingresos) y ésta aparecerá a la derecha de *Series*.

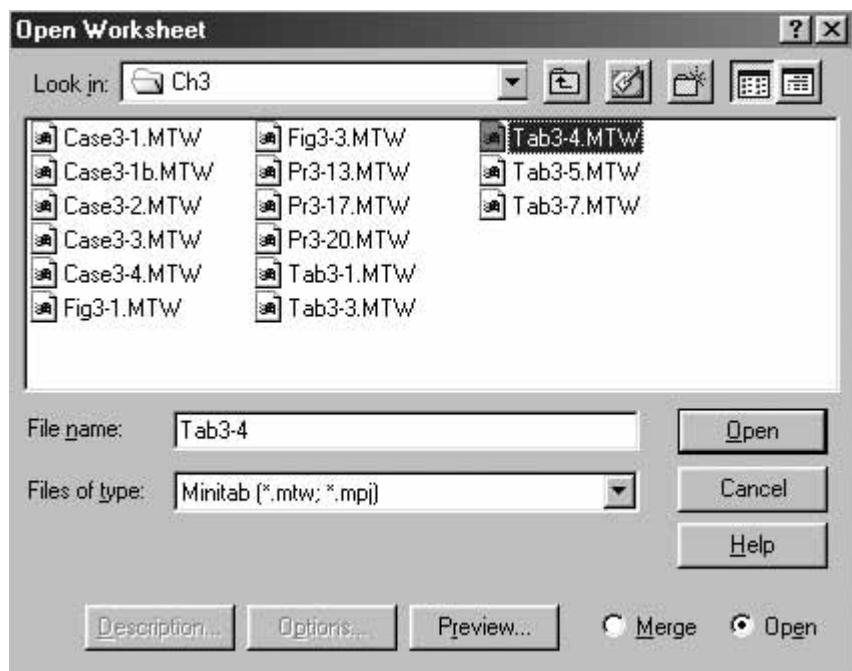


FIGURA 3.27 Menú de apertura de la hoja de cálculo

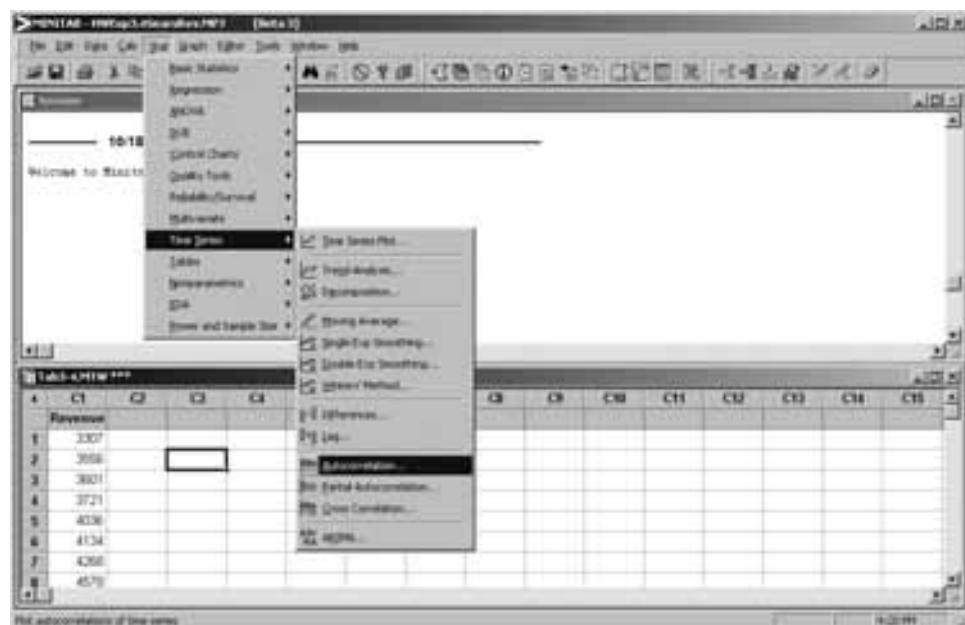
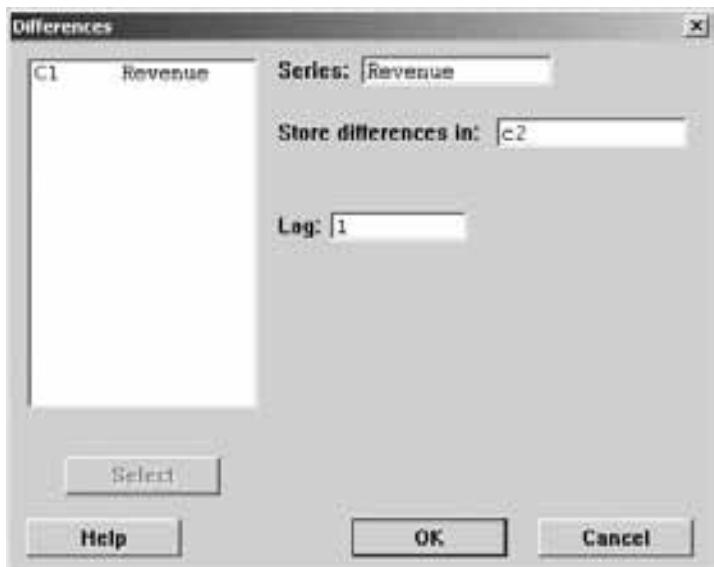


FIGURA 3.28 Menú de autocorrelación de Minitab



**FIGURA 3.29 Cuadro de diálogo de la función de autocorrelación en Minitab**



**FIGURA 3.30 Cuadro de diálogo de diferencias en Minitab**

- b) Dé clic en *Number of lags* (número de retrasos) e introduzca 10 en el cuadro de la derecha.  
 c) Introduzca un *Title* (título) en el espacio apropiado y dé clic en OK. La función de autocorrelación resultante se muestra en la figura 3.11 de la página 71.
4. Con el fin de diferenciar los datos, dé clic en los siguientes menús:
- ```
Stat>Time Series>Differences
```
- La opción de *Differences* (diferencias) se encuentra por encima de la opción de *Auto-correlation* (autocorrelación) que se ilustra en la figura 3.28.
5. Aparece el cuadro de diálogo de *Differences* que se muestra en la figura 3.30.
- Dé doble clic en la variable *Revenue* y ésta aparecerá en el lado derecho de *Series*.
  - Utilice el tabulador para introducir las diferencias en el campo *Store differences in: C2*. Los datos diferenciados aparecerán en la columna C2 de la hoja de cálculo.

## Aplicaciones en Excel

**El problema.** Harry Vernon desea utilizar Excel para calcular los coeficientes de correlación y establecer un correlograma para los datos que se presentan en la tabla 3.1 (vea la página 61).

### Solución de Excel

La mejor manera de resolver este problema sería utilizar un complemento de análisis estadístico para Excel, como CB Predictor. El método que usa el complemento se presentará al final del capítulo 9. Los siguientes comandos resuelven el problema sin hacer uso de CB Predictor.

1. Cree un nuevo archivo haciendo clic en los siguientes menús:

File>New

2. Coloque el cursor en A1. Observe que siempre que se coloque el cursor en una celda, ésta se verá resaltada. Escriba el título VERNON'S MUSIC STORE. Coloque el cursor en A2 y escriba el título NUMBER OF VCRS SOLD (número de videograbadoras vendidas).
3. Coloque el cursor en A4 y escriba Month (mes). Oprima <enter> y la celda A5 se resalta. Ahora introduzca los meses comenzando por enero y terminando con diciembre en A16.
4. Coloque el cursor en B4 y escriba Y. Introduzca los datos de la tabla 3.1 comenzando por la celda B5. Coloque el cursor en C4 y escriba Z.
5. Resalte las celdas B4:C16 y dé clic en los siguientes menús:

Insert>Name>Create

En el cuadro de diálogo de *Create Names* (crear nombres), dé clic en el recuadro de *Top row* (fila superior) y dé clic en OK. Este paso crea el nombre Y para el rango B5:B16 y el nombre Z para el rango C5:C16.

6. Resalte C5 e introduzca la fórmula

= (B5-AVERAGE(Y)) / STDEV(Y)

Copie C5 al resto de la columna al resaltarla, después dé clic en la barra que se encuentra en la barra *Edit* (edición) o bien con el cursor situado en C5 lleve el ratón a la esquina inferior derecha de C5 y arrástrelo hacia abajo, hasta la celda C16. Aún con las celdas C5:C16 resaltadas dé clic en el botón *Decrease Decimal* (disminuir decimales, el cual se muestra en la figura 3.31 hacia la esquina superior derecha) hasta que se desplieguen tres decimales. El botón *Decrease Decimal* está en la barra de herramientas de *Format* (formato). Esta barra de herramientas puede desplegarse al dar clic con el botón derecho del ratón en *File* (archivo) y después con el botón izquierdo del ratón en *Formatting* (formato).

7. Introduzca las etiquetas LAG y ACF en las celdas E4 y F4. Para poder examinar los primeros seis retrasos de tiempo, introduzca los dígitos del 1 al 6 en las celdas E5:E10.
8. Resalte F5 e introduzca la fórmula

=SUMPRODUCT(OFFSET(Z,E5,0,12-E5),OFFSET(Z,0,0,12-E5))/11<sup>1</sup>

Resalte F5, dé clic en la barra *Fill* (llenar) ubicada en la esquina inferior izquierda y deslícela hacia la celda F10. Con las celdas F5:F10 aún resaltadas, dé clic en el botón *Decrease Decimal* hasta que se observen tres espacios decimales. Los resultados se muestran en la figura 3.31.

9. Para desarrollar la función de autocorrelación, resalte las celdas F5:F10. Dé clic en la herramienta *Chart Wizard* (asistente de gráficos, que se muestra en la figura 3.32 hacia la parte superior derecha).
10. Aparece el cuadro de diálogo de los pasos 1 al 4 para la herramienta *Chart Wizard*. En el paso 1 seleccione el tipo de gráfica al dar clic en *Column* (columnas), dé clic en *Next* (siguiente). En el siguiente cuadro de diálogo para el paso 2, dé clic en la pestaña

<sup>1</sup> El comando OFFSET equivale a DESREF en la versión en español de Excel.

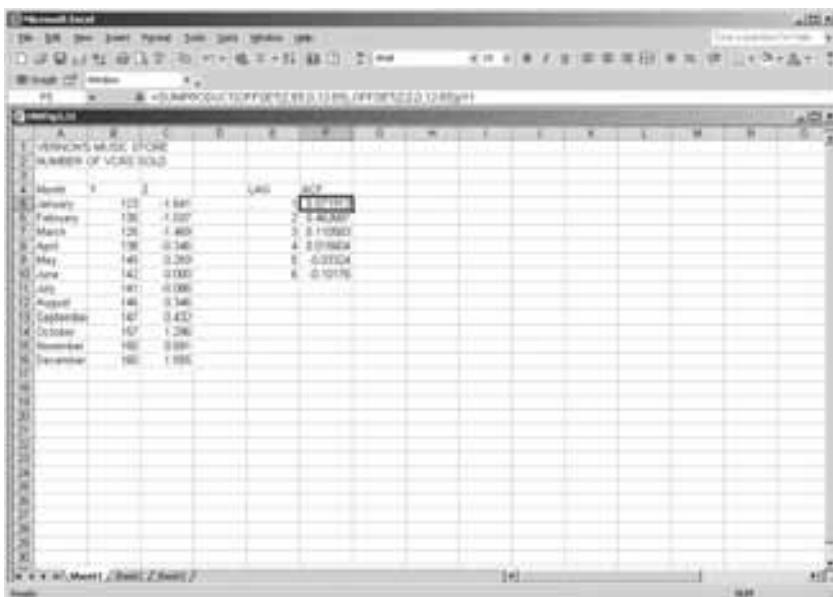


FIGURA 3.31 Hoja de cálculo de Excel para los datos de VCR

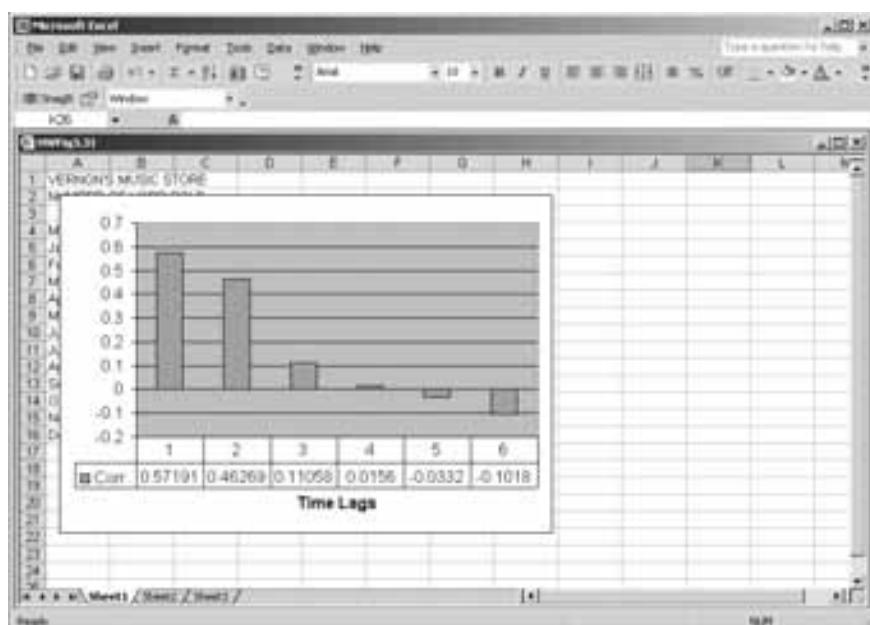


FIGURA 3.32 Resultados de Excel para la función de autocorrelación

de Series. En el espacio en blanco junto a *Name* (nombre) escriba *Corr*. Dé clic en *Next* y aparecerá el cuadro de diálogo del paso 3. Debajo del título de la gráfica borre *Corr*. Bajo el eje de *Category* (categoría) *X* escriba *Time Lags* (retrasos). Ahora dé clic en la pestaña de *Data Table* (tabla de datos) y verifique la casilla que se encuentra junto a la tabla de *Show data* (mostrar datos). Para poder acceder al cuadro de diálogo del paso 4 dé clic en *Finish* (finalizar) para producir la función de correlación que se muestra en la figura 3.31. Dé clic en una de las esquinas del gráfico y muévalo hacia fuera para agrandar la función de autocorrelación.

11. Para poder guardar los datos y usarlos en el capítulo 9, dé clic en

File>Save As

En el cuadro de diálogo de *Save As* (guardar como) escriba Tab3-1 en el espacio que se encuentra a la derecha del nombre del archivo (*File*): dé clic en *Save* (guardar) y el archivo se guardará con el nombre de Tab3-1.xls.

## Referencias

- Armstrong, J. S., Ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, MA: Kluwer Academic Publishers, 2001.
- Diebold, F. X., *Elements of Forecasting*, tercera edición. Cincinnati, OH: South-Western, 2004.
- Ermer, C. M., “Cost of Error Affects the Forecasting Model Selection.” *Journal of Business Forecasting* (primavera de 1991): 10-12.
- Fildes, R., M., Hibon, S. Makridakis y N. Meade, “The Accuracy of Extrapolative Forecasting Methods: Additional Empirical Evidence”, *International Journal of Forecasting* (1997): 13.
- Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen y R. Winkler, “The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition”, *Journal of Forecasting* 1 (1982): 111-153.
- Makridakis, S., C. Chatfield, M. Hibon, M. Lawrence, T. Mills, K. Ord, y L. F. Simmons, “The M2-Competition: A Real Time Judgmentally Based Forecasting Study.” *International Journal of Forecasting* 9 (1993): 5-30.
- Makridakis, S., y M. Hibon, “The M3-Competition: Results, Conclusions and Implications.” *International Journal of Forecasting* 16 (2000): 451-476.
- Newbold, P., y T. Bos, *Introductory Business and Economic Forecasting*, segunda edición. Cincinnati, OH: South-Western, 1994.
- Quenouille, M. H., “The Joint Distribution of Serial Correlation Coefficients”, *Annals of Mathematical Statistics* 20 (1949): 561-571.
- Wilkinson, G. F., “How a Forecasting Model Is Chosen”, *Journal of Business Forecasting* (verano 1989): 7-8.

## PROMEDIOS MÓVILES Y MÉTODOS DE SUAVIZAMIENTO

Este capítulo describe tres enfoques simples para pronosticar una serie de tiempo: el informal, el de promedio y el de suavizamiento. Los métodos *informales* se usan para desarrollar modelos sencillos, suponen que los datos más recientes proporcionan los mejores elementos para predecir el futuro. Los métodos de *promedio* generan pronósticos con base en el promedio de observaciones pasadas. Los métodos de *suavizamiento* generan pronósticos con base en promedios ponderados, en donde las ponderaciones decrecen exponencialmente para los valores precedentes.

La figura 4.1 muestra un esbozo del proceso de pronóstico para los métodos presentados en este capítulo. Mírese usted en una escala de tiempo: se encuentra en el punto en la figura 4.1 y puede ver hacia atrás sobre las observaciones pasadas de la variable de interés ( $Y$ ) o hacia adelante en el futuro. Una vez que se ha elegido la técnica de pronóstico, ésta se ajusta a los datos conocidos y se obtienen los valores de pronóstico ( $\hat{Y}_t$ ). Una vez disponibles los valores de pronóstico, se pueden comparar con las observaciones conocidas y así puede calcularse el error de pronóstico ( $e_t$ ).

Una buena estrategia para evaluar los métodos de pronóstico implica los siguientes pasos:

1. Se elige un método de pronóstico con base en el análisis y la intuición del pronosticador sobre la naturaleza de los datos.
2. El conjunto de datos se divide en dos secciones: una de ajustes y otra de validación de la técnica pronóstico
3. La técnica de pronóstico elegida se utiliza para obtener pronósticos de prueba o valores de ajuste con la primera porción de datos.
4. La técnica se utiliza para pronosticar la segunda sección, los pronósticos obtenidos se comparan con los datos a fin de evaluar el error de pronóstico (vea el capítulo 3 para un repaso sobre las medidas de precisión de pronósticos).
5. Se toma una decisión. Se podría utilizar la técnica en su forma actual, modificarla o desarrollar un pronóstico mediante otra técnica y luego comparar los resultados.

**FIGURA 4.1 Esquema de un pronóstico**

|       | Datos pasados                | Usted está aquí                                         | Periodos a pronosticar                               |
|-------|------------------------------|---------------------------------------------------------|------------------------------------------------------|
|       | $Y_{t-3}, Y_{t-2}, Y_{t-1},$ | $t$                                                     | $\hat{Y}_{t+1}, \hat{Y}_{t+2}, \hat{Y}_{t+3}, \dots$ |
| donde | $Y_t$                        | es la observación más reciente de una variable          |                                                      |
|       | $\hat{Y}_{t+1}$              | es el pronóstico para el siguiente periodo en el futuro |                                                      |

## MODELOS INFORMALES

Con frecuencia, los negocios nuevos enfrentan el dilema de ensayar pronósticos con grupos de datos muy pequeños. Esta situación induce un problema real debido a que muchas técnicas de pronóstico requieren grandes cantidades de datos. Los modelos informales son una solución potencial, ya que se basan únicamente en la información más reciente que está disponible.

Los pronósticos informales suponen que los períodos recientes son los mejores para predecir el futuro. El modelo más simple es

$$\hat{Y}_{t+1} = Y_t \quad (4.1)$$

donde  $\hat{Y}_{t+1}$  es el pronóstico realizado en el tiempo  $t$  (el origen del pronóstico) para el tiempo  $t + 1$ . El pronóstico informal para cada período es la observación inmediatamente anterior. Se le da 100% del peso al valor actual de la serie. En ocasiones, el pronóstico informal es llamado pronóstico “de no cambio”.

Dado que el pronóstico informal (ecuación 4.1) descarta todas las demás observaciones, este esquema detecta los cambios rápidamente. El problema con este enfoque es que las fluctuaciones aleatorias se reproducen con la misma fidelidad que los cambios en los datos fundamentales.

### Ejemplo 4.1

La figura 4.2 muestra las ventas trimestrales de serruchos en la Acme Tool Company. La técnica informal predice que las ventas del siguiente trimestre serán las mismas que en el trimestre anterior. La tabla 4.1 muestra los datos desde 1996 hasta 2002. Si se utilizan los datos de 1996 a 2001 como la parte inicial (de ajustes) y el 2002 como la parte de prueba (validación), el pronóstico para el primer trimestre de 2002 será

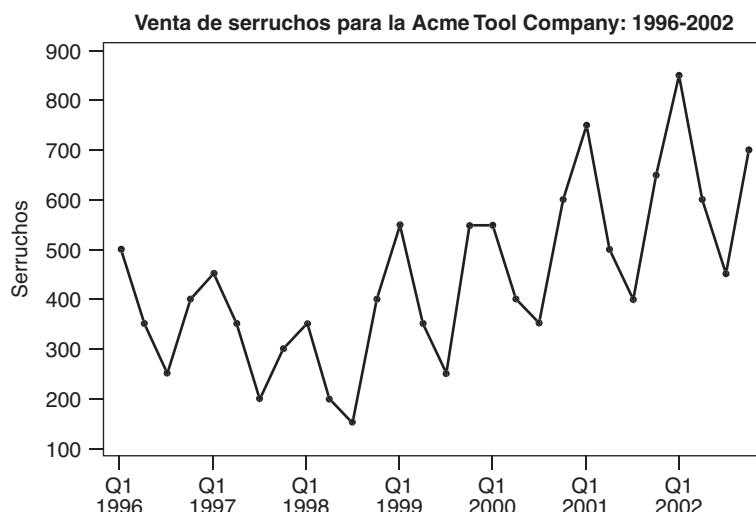
$$\begin{aligned}\hat{Y}_{24+1} &= Y_{24} \\ \hat{Y}_{25} &= 650\end{aligned}$$

El error de pronóstico se determina mediante la ecuación 3.6. El error del período 25 es:

$$e_{25} = Y_{25} - \hat{Y}_{25} = 850 - 650 = 200$$

De manera similar, el pronóstico para el período 26 es de 850 con un error de -250. La figura 4.2 muestra que estos datos tienen una tendencia ascendente y parece que hay un patrón estacional (los trimestres primero y cuarto son relativamente altos), por lo cual se toma una decisión de modificar el modelo informal.

**FIGURA 4.2 Gráfica de serie de tiempo para las ventas de serruchos de la Acme Tool Company**



**TABLA 4.1 Ventas de serruchos de la  
Acme Tool Company, 1996-2002**

| Año  | Trimestre | t  | Ventas |
|------|-----------|----|--------|
| 1996 | 1         | 1  | 500    |
|      | 2         | 2  | 350    |
|      | 3         | 3  | 250    |
|      | 4         | 4  | 400    |
| 1997 | 1         | 5  | 450    |
|      | 2         | 6  | 350    |
|      | 3         | 7  | 200    |
|      | 4         | 8  | 300    |
| 1998 | 1         | 9  | 350    |
|      | 2         | 10 | 200    |
|      | 3         | 11 | 150    |
|      | 4         | 12 | 400    |
| 1999 | 1         | 13 | 550    |
|      | 2         | 14 | 350    |
|      | 3         | 15 | 250    |
|      | 4         | 16 | 550    |
| 2000 | 1         | 17 | 550    |
|      | 2         | 18 | 400    |
|      | 3         | 19 | 350    |
|      | 4         | 20 | 600    |
| 2001 | 1         | 21 | 750    |
|      | 2         | 22 | 500    |
|      | 3         | 23 | 400    |
|      | 4         | 24 | 650    |
| 2002 | 1         | 25 | 850    |
|      | 2         | 26 | 600    |
|      | 3         | 27 | 450    |
|      | 4         | 28 | 700    |

El examen de los datos en el [ejemplo 4.1](#) conduce a concluir que los valores aumentan a lo largo del tiempo. Cuando esto sucede, se dice que los valores son *no estacionarios* en su nivel o que tienen una *tendencia*. Si se usa la ecuación 4.1, las proyecciones son consistentemente bajas. Sin embargo, la técnica puede ajustarse para considerar las tendencias mediante la suma de la diferencia entre este periodo y el anterior. La ecuación de pronóstico es

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \quad (4.2)$$

La ecuación 4.2 considera el monto del cambio que ocurrió entre los trimestres.

#### Ejemplo 4.1 (continúa)

Si se usa la ecuación 4.2, la ecuación para el primer trimestre de 2002 los cálculos serían

$$\hat{Y}_{24+1} = Y_{24} + (Y_{24} - Y_{24-1})$$

$$\hat{Y}_{25} = Y_{24} + (Y_{24} - Y_{23})$$

$$\hat{Y}_{25} = 650 + (650 - 400)$$

$$\hat{Y}_{25} = 650 + 250 = 900$$

El error de pronóstico de este modelo es

$$e_{25} = Y_{25} - \hat{Y}_{25} = 850 - 900 = -50$$

Para ciertos fines, podría ser más adecuada la tasa de cambio que la cantidad absoluta de cambio. Si es así, es razonable generar pronósticos de acuerdo con

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} \quad (4.3)$$

La inspección visual de los datos en la tabla 4.1 indica que, al parecer, existe una variación estacional. Las ventas en los trimestres primero y cuarto son más grandes que en los demás trimestres. Si el patrón estacional es fuerte, una ecuación adecuada de pronósticos para los datos trimestrales podría ser

$$\hat{Y}_{t+1} = Y_{t-3} \quad (4.4)$$

La ecuación 4.4 dice que para siguiente trimestre la variable se pronostica con el valor que tomó en el trimestre correspondiente del año anterior.

La principal debilidad de este enfoque es que ignora todo lo ocurrido desde el año pasado, así como cualquier tendencia. Hay varias formas de introducir información más reciente. Por ejemplo, el analista puede combinar estimados estacionales y de tendencia y pronosticar el siguiente trimestre si utiliza

$$\hat{Y}_{t+1} = Y_{t-3} + \frac{(Y_t - Y_{t-1}) + \dots + (Y_{t-3} - Y_{t-4})}{4} = Y_{t-3} + \frac{Y_t - Y_{t-4}}{4} \quad (4.5)$$

donde el término  $Y_{t-3}$  pronostica el patrón estacional y el otro término es el promedio de las diferencias entre los últimos cuatro trimestres y proporciona un estimado de la tendencia.

Los modelos informales de pronóstico de las ecuaciones 4.4 y 4.5 se dan para datos trimestrales. Pueden hacerse ajustes para datos recopilados durante diferentes períodos. Por ejemplo, en el caso de datos mensuales el período es de 12 y no de 4; y el pronóstico para el siguiente período (mes) dado por la ecuación 4.4 es  $\hat{Y}_{t+1} = Y_{t-11}$ .

Es evidente que el número y la complejidad de los posibles modelos informales sólo está limitado por el ingenio del analista, pero el uso de estas técnicas debe guiarse por el buen juicio.

Los métodos informales también se usan como base para hacer comparaciones contra las cuales se juzga el desempeño de métodos más sofisticados.

#### Ejemplo 4.1 (continúa)

Los pronósticos del primer trimestre del 2002 al usar las ecuaciones 4.3, 4.4 y 4.5 son

$$\begin{aligned} \hat{Y}_{24+1} &= Y_{24} \frac{Y_{24}}{Y_{24-1}} = Y_{24} \frac{Y_{24}}{Y_{23}} \\ \hat{Y}_{25} &= 650 \frac{650}{400} = 1,056 \end{aligned} \quad (\text{Ecuación 4.3})$$

$$\begin{aligned} \hat{Y}_{24+1} &= Y_{24-3} = Y_{21} \\ \hat{Y}_{25} &= Y_{21} = 750 \end{aligned} \quad (\text{Ecuación 4.4})$$

$$\hat{Y}_{24+1} = Y_{24-3} + \frac{(Y_{24} - Y_{24-1}) + \dots + (Y_{24-3} - Y_{24-4})}{4} \quad (\text{Ecuación 4.5})$$

$$\hat{Y}_{25} = Y_{24-3} + \frac{Y_{24} - Y_{24-4}}{4}$$

$$\hat{Y}_{25} = Y_{21} + \frac{Y_{24} - Y_{20}}{4} = 750 + \frac{650 - 600}{4}$$

$$\hat{Y}_{25} = 750 + 12.5 = 762.5$$

## MÉTODOS DE PRONÓSTICOS BASADOS EN PROMEDIOS

---

Con frecuencia, la administración se encuentra con una situación en la que debe actualizar —por día, semana o mes— los pronósticos de inventarios que contienen cientos o miles de artículos. Otras veces no es posible desarrollar técnicas sofisticadas de pronóstico para cada uno de dichos artículos. En vez de ello, se necesitan algunas herramientas rápidas, de bajo costo, muy sencillas y de corto plazo para lograr esta tarea.

Un administrador que se enfrente a tal situación probablemente use una técnica de promedio o de suavizamiento. Este tipo de técnicas emplean una forma de promedio ponderado de observaciones pasadas a fin de suavizar las fluctuaciones del corto plazo. La suposición sobre la cual se basan estas técnicas es que las fluctuaciones en los valores previos representan salidas aleatorias de cierta estructura subyacente. Una vez identificada tal estructura, se puede proyectar hacia el futuro para producir un pronóstico.

### Promedios simples

Los datos históricos pueden suavizarse de muchas formas. El objetivo es utilizar datos anteriores para desarrollar un modelo de pronósticos para períodos futuros. En esta sección se considera el método de promedios simples. De la misma forma que con los métodos informales, se toma una decisión para utilizar los primeros puntos de datos como la parte de inicialización y los datos restantes como la parte de prueba. A continuación, se usa la ecuación 4.6 para promediar (o calcular la media) de la parte de inicialización de los datos y para pronosticar el siguiente periodo.

$$\hat{Y}_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i \quad (4.6)$$

Cuando está disponible una nueva observación, el pronóstico para el periodo siguiente,  $\hat{Y}_{t+2}$  es el promedio —o media— que se calcula utilizando la ecuación 4.6 y esta nueva observación.

Cuando se pronostica un gran número de series de manera simultánea (por ejemplo, en la administración de inventarios), el almacenamiento de datos puede ser un problema. La ecuación 4.7 resuelve ese problema potencial. Sólo el pronóstico y la observación más recientes deben almacenarse al pasar el tiempo.

$$\hat{Y}_{t+2} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t+1} \quad (4.7)$$

El método de *promedios simples* es una técnica adecuada cuando se han estabilizado las fuerzas que generan las series a pronosticar y cuando, por lo general, no cambia el entorno donde existe la serie. Los ejemplos de este tipo de series son el volumen de ventas que resultan del esfuerzo constante de los vendedores, la cantidad de ventas de un produc-

to en la etapa madura de su ciclo de vida y el número de citas por semana solicitadas a un dentista, médico o abogado cuya base de pacientes o clientes es constante.

Un *promedio simple* usa la media de todas las observaciones históricas relevantes como pronóstico del siguiente periodo.

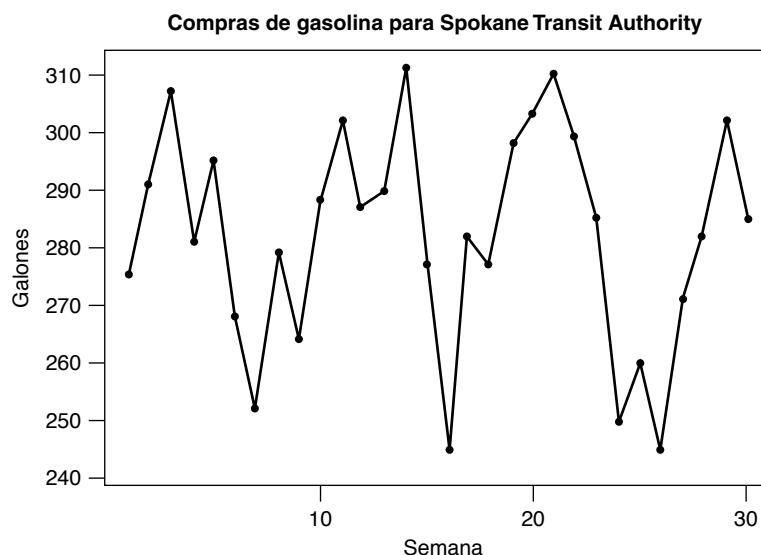
#### Ejemplo 4.2

La Spokane Transit Authority (STA) opera una flotilla de camionetas que se usan para transportar tanto a los discapacitados como a las personas de edad avanzada. En la tabla 4.2 se muestra un registro de la gasolina que se compró para la flotilla. La cantidad real de gasolina consumida por uno de los vehículos se determina mediante la naturaleza aleatoria de las llamadas y los destinos. Al examinar las compras de gasolina graficadas en la figura 4.3 se aprecia que los datos son muy estables. Dado que aparentemente son datos estacionarios, se utiliza el método de

**TABLA 4.2 Compras de gasolina para Spokane Transit Authority**

| Semana   | Galones              | Semana   | Galones              | Semana   | Galones              |
|----------|----------------------|----------|----------------------|----------|----------------------|
| <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> |
| 1        | 275                  | 11       | 302                  | 21       | 310                  |
| 2        | 291                  | 12       | 287                  | 22       | 299                  |
| 3        | 307                  | 13       | 290                  | 23       | 285                  |
| 4        | 281                  | 14       | 311                  | 24       | 250                  |
| 5        | 295                  | 15       | 277                  | 25       | 260                  |
| 6        | 268                  | 16       | 245                  | 26       | 245                  |
| 7        | 252                  | 17       | 282                  | 27       | 271                  |
| 8        | 279                  | 18       | 277                  | 28       | 282                  |
| 9        | 264                  | 19       | 298                  | 29       | 302                  |
| 10       | 288                  | 20       | 303                  | 30       | 285                  |

**FIGURA 4.3 Gráfica de serie de tiempo para las compras de gasolina de Spokane Transit Authority**



promedios simples para las semanas que van de la 1 a la 28 para pronosticar las compras de gasolina durante las semanas 29 y 30. El pronóstico para la 29 es

$$\hat{Y}_{28+1} = \frac{1}{28} \sum_{i=1}^{28} Y_i$$

$$\hat{Y}_{29} = \frac{7,874}{28} = 281.2$$

El error de pronóstico es

$$e_{29} = Y_{29} - \hat{Y}_{29} = 302 - 281.2 = 20.8$$

El pronóstico para la semana 30 incluye un punto más de datos (302) que se añadió al periodo de inicialización. Si se usa la ecuación 4.7, el pronóstico es:

$$\hat{Y}_{28+2} = \frac{28\hat{Y}_{28+1} + Y_{28+1}}{28 + 1} = \frac{28\hat{Y}_{29} + Y_{29}}{29}$$

$$\hat{Y}_{30} = \frac{28(281.2) + 302}{29} = 281.9$$

El error de pronóstico es

$$e_{30} = Y_{30} - \hat{Y}_{30} = 285 - 281.9 = 3.1$$

Al usar el método de los promedios simples, el pronóstico de galones de gasolina comprados en la semana 31 es

$$\hat{Y}_{30+1} = \frac{1}{30} \sum_{i=1}^{30} Y_i = \frac{8,461}{30} = 282$$

## Promedios móviles

El método de promedios simples utiliza la media de todos los datos para hacer el pronóstico. ¿Qué pasa si el analista se preocupa más por las observaciones recientes? Se puede especificar un número constante de puntos de datos al inicio y se puede calcular una media para las observaciones más recientes. El término *promedio móvil* se usa para describir este enfoque. Conforme se tienen nuevas observaciones, se calcula una nueva media al sumar el valor más reciente y al eliminar el más antiguo. Este promedio móvil se usa para pronosticar el siguiente periodo. La ecuación 4.8 da el pronóstico como el promedio móvil simple. Un promedio móvil de orden  $k$ ,  $MA(k)$ , se calcula mediante

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-k+1}}{k} \quad (4.8)$$

donde

$$\begin{aligned}\hat{Y}_{t+1} &= \text{valor pronosticado para el siguiente periodo} \\ Y_t &= \text{valor real en el periodo } t \\ k &= \text{número de términos en el promedio móvil}\end{aligned}$$

El promedio móvil para el periodo  $t$  es la media aritmética de las observaciones más recientes  $k$ . En un promedio móvil se asignan pesos iguales a cada observación. Cada nuevo punto de datos se incluye en el promedio al estar disponible y se elimina el punto de datos más antiguo. La tasa de respuesta a los cambios en el patrón subyacente de datos depende del número de periodos,  $k$ , incluidos en el promedio móvil.

Observe que la técnica de los promedios móviles sólo maneja los últimos períodos  $k$  de datos conocidos; el número de puntos de datos en cada promedio no cambia al pasar el tiempo. El modelo del promedio móvil no maneja muy bien la tendencia o la estacionalidad, aunque lo hace mejor que el método de promedio simple.

El analista debe elegir el número de períodos,  $k$ , en un promedio móvil. Un promedio móvil de orden 1, MA(1), usaría la observación actual,  $Y_t$ , para pronosticar  $Y$  para el siguiente periodo. Éste tan sólo es el enfoque informal de la ecuación 4.1.

Un *promedio móvil* de orden  $k$  es la media aritmética de  $k$  observaciones consecutivas. El valor más reciente del promedio móvil proporciona un pronóstico para el siguiente periodo.

**TABLA 4.3 Compras de gasolina para Spokane Transit Authority**

| $t$ | Galones | $\hat{Y}_t$ | $e_t$ |
|-----|---------|-------------|-------|
| 1   | 275     | —           | —     |
| 2   | 291     | —           | —     |
| 3   | 307     | —           | —     |
| 4   | 281     | —           | —     |
| 5   | 295     | —           | —     |
| 6   | 268     | 289.8       | 21.8  |
| 7   | 252     | 288.4       | 36.4  |
| 8   | 279     | 280.6       | 1.6   |
| 9   | 264     | 275.0       | 11.0  |
| 10  | 288     | 271.6       | 16.4  |
| 11  | 302     | 270.2       | 31.8  |
| 12  | 287     | 277.0       | 10.0  |
| 13  | 290     | 284.0       | 6.0   |
| 14  | 311     | 286.2       | 24.8  |
| 15  | 277     | 295.6       | 18.6  |
| 16  | 245     | 293.4       | 48.4  |
| 17  | 282     | 282.0       | 0.0   |
| 18  | 277     | 281.0       | 4.0   |
| 19  | 298     | 278.4       | 19.6  |
| 20  | 303     | 275.8       | 27.2  |
| 21  | 310     | 281.0       | 29.0  |
| 22  | 299     | 294.0       | 5.0   |
| 23  | 285     | 297.4       | 12.4  |
| 24  | 250     | 299.0       | 49.0  |
| 25  | 260     | 289.4       | 29.4  |
| 26  | 245     | 280.8       | 35.8  |
| 27  | 271     | 267.8       | 3.2   |
| 28  | 282     | 262.2       | 19.8  |
| 29  | 302     | 261.6       | 40.4  |
| 30  | 285     | 272.0       | 13.0  |

**Ejemplo 4.3**

La tabla 4.3 demuestra la técnica de pronósticos mediante promedio móvil con los datos de Spokane Transit Authority, donde se utiliza un promedio móvil de cinco semanas. El pronóstico de promedio móvil para la semana 29 es

$$\hat{Y}_{28+1} = \frac{Y_{28} + Y_{28-1} + \cdots + Y_{28-5+1}}{5}$$

$$\hat{Y}_{29} = \frac{Y_{28} + Y_{27} + Y_{26} + Y_{25} + Y_{24}}{5}$$

$$\hat{Y}_{29} = \frac{282 + 271 + 245 + 260 + 250}{5} = \frac{1,308}{5} = 261.6$$

Cuando se conoce el valor real de la semana 29, se calcula el error de pronóstico:

$$e_{29} = Y_{29} - \hat{Y}_{29} = 302 - 261.6 = 40.4$$

El pronóstico de la semana 31 es:

$$\hat{Y}_{28+1} = \frac{Y_{28} + Y_{28-1} + \cdots + Y_{28-5+1}}{5}$$

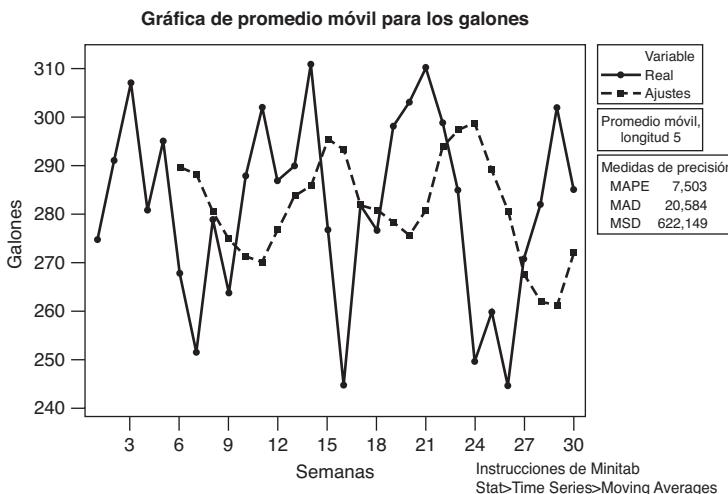
$$\hat{Y}_{29} = \frac{Y_{28} + Y_{27} + Y_{26} + Y_{25} + Y_{24}}{5}$$

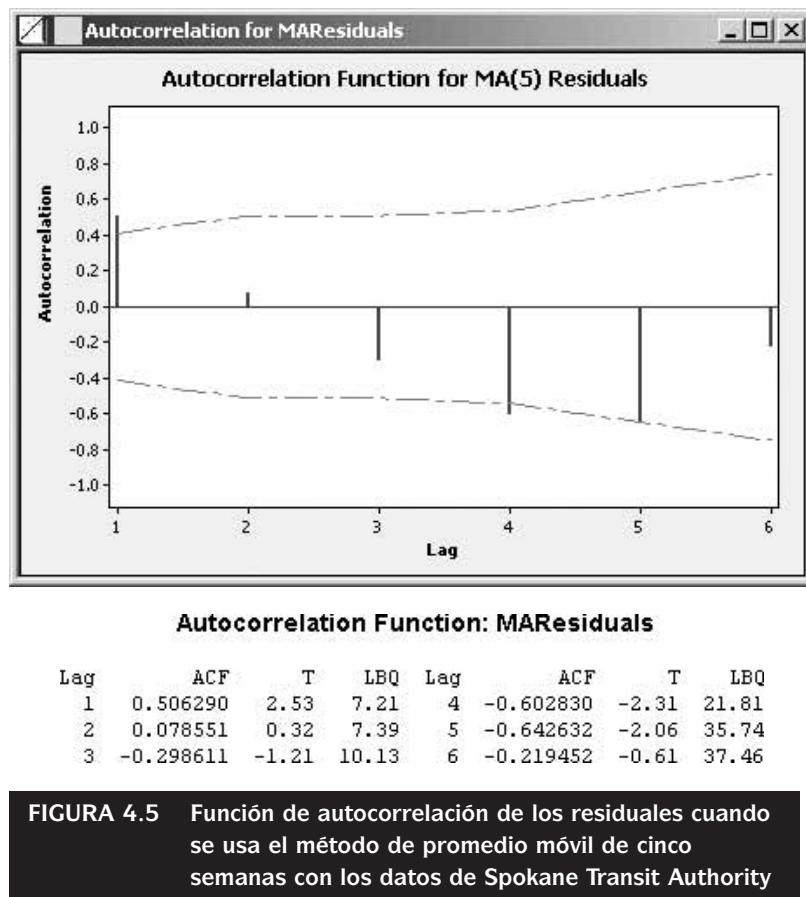
$$\hat{Y}_{29} = \frac{282 + 271 + 245 + 260 + 250}{5} = \frac{1,308}{5} = 261.6$$

Minitab puede usarse para calcular el promedio móvil de cinco semanas (vea “Aplicaciones en Minitab” al final del capítulo para leer las instrucciones). La figura 4.4 muestra el promedio móvil de cinco semanas graficado contra los datos reales, MAPE, MAD, MSD y las instrucciones básicas de Minitab. Observe que Minitab le llama *MSD* (desviación cuadrática media) al error cuadrático medio.

La figura 4.3 muestra la función de autocorrelación de los residuales del método de promedio móvil de cinco semanas. Los límites de error para las autocorrelaciones individuales centradas en 0 y la estadística *Q* de Ljung-Box indican que existe una autocorrelación residual significativa; es decir, que los residuales no son aleatorios. La asociación contenida en los residuales en ciertos retrasos puede usarse para mejorar el modelo de pronóstico.

**FIGURA 4.4 Promedio móvil de cinco semanas aplicado a las compras semanales de gasolina de la Spokane Transit Authority**





**FIGURA 4.5 Función de autocorrelación de los residuales cuando se usa el método de promedio móvil de cinco semanas con los datos de Spokane Transit Authority**

El analista debe apelar a su juicio cuando determine sobre cuántos días, semanas, meses o trimestres debe basarse el promedio móvil. Cuanto menor sea el número, mayor peso se les dará a los períodos recientes. Por el contrario, cuanto mayor sea el número, se les dará menos peso a los períodos más recientes. Un número pequeño es más deseable cuando existen cambios repentinos en el nivel de la serie y pone un mayor peso en la historia reciente, lo cual permite que los pronósticos alcancen más rápido el nivel actual. Un número alto es deseable cuando hay fluctuaciones amplias y poco frecuentes en la serie.

Los promedios móviles se usan con frecuencia con los datos trimestrales o mensuales para ayudar a suavizar los componentes dentro de una serie de tiempo, como se muestra en el capítulo 5. Para los datos trimestrales, un promedio móvil de cuatro trimestres, MA(4), arroja el promedio de los cuatro trimestres y en el caso de datos mensuales, un promedio móvil de 12 meses, MA(12), elimina o promedia los efectos estacionales. Cuanto mayor sea el orden del promedio móvil, mayor será el efecto de suavización.

En el ejemplo 4.3 se utilizó la técnica de promedios móviles con datos estacionarios. En el ejemplo 4.4 se muestra lo que sucede cuando se usa el método de promedios móviles con datos con tendencia. A continuación, se presenta la técnica de promedio móvil doble que se diseñó para manejar datos con tendencia.

### Promedios móviles dobles

Una forma de pronosticar datos de series de tiempo que tienen una tendencia lineal es usar los promedios móviles dobles. Este método hace lo que está implícito en su nombre: se calcula un grupo de promedios móviles y, entonces, se calcula un segundo grupo como un promedio móvil del primer conjunto.

**TABLA 4.4 Rentas mensuales para Movie Video Store del ejemplo 4.4**

| <i>t</i> | <i>Y<sub>t</sub></i> | Total | $\hat{Y}_{t+1}$ | <i>e<sub>t</sub></i> |
|----------|----------------------|-------|-----------------|----------------------|
| 1        | 654                  | —     | —               | —                    |
| 2        | 658                  | —     | —               | —                    |
| 3        | 665                  | 1,977 | —               | —                    |
| 4        | 672                  | 1,995 | 659             | 13                   |
| 5        | 673                  | 2,010 | 665             | 8                    |
| 6        | 671                  | 2,016 | 670             | 1                    |
| 7        | 693                  | 2,037 | 672             | 21                   |
| 8        | 694                  | 2,058 | 679             | 15                   |
| 9        | 701                  | 2,088 | 686             | 15                   |
| 10       | 703                  | 2,098 | 696             | 7                    |
| 11       | 702                  | 2,106 | 699             | 3                    |
| 12       | 710                  | 2,115 | 702             | 8                    |
| 13       | 712                  | 2,124 | 705             | 7                    |
| 14       | 711                  | 2,133 | 708             | 3                    |
| 15       | 728                  | 2,151 | 711             | 17                   |
| 16       | —                    | —     | 717             | —                    |

MSE = 133.

La tabla 4.4 muestra los datos para las rentas semanales de la Movie Video Store junto con los resultados de haber utilizado un promedio móvil de tres semanas, a fin de pronosticar las ventas futuras. Al examinar la columna de error en la tabla 4.4, se puede ver que cada entrada es positiva, lo cual significa que los pronósticos no alcanzaron la tendencia. El promedio móvil de tres semanas y el promedio móvil doble de estos datos se muestran en la figura 4.6. Observe cómo los pronósticos mediante los promedios móviles de tres semanas se atrasan con respecto a los valores reales en períodos comparables: esto ilustra lo que sucede cuando se usa a la técnica de promedios móviles con datos con tendencia. Observe también que los pronósticos realizados mediante promedios móviles dobles se atrasan respecto al primer conjunto de pronósticos casi tanto como éste se atrasa respecto de los valores reales. La diferencia entre los dos grupos de promedios móviles se suma al promedio móvil de tres semanas, a fin de pronosticar los valores reales.

Las ecuaciones de la 4.9 a la 4.12 resumen la construcción de promedios móviles dobles. Primero se utiliza la ecuación 4.8 para calcular el promedio móvil del orden *k*.

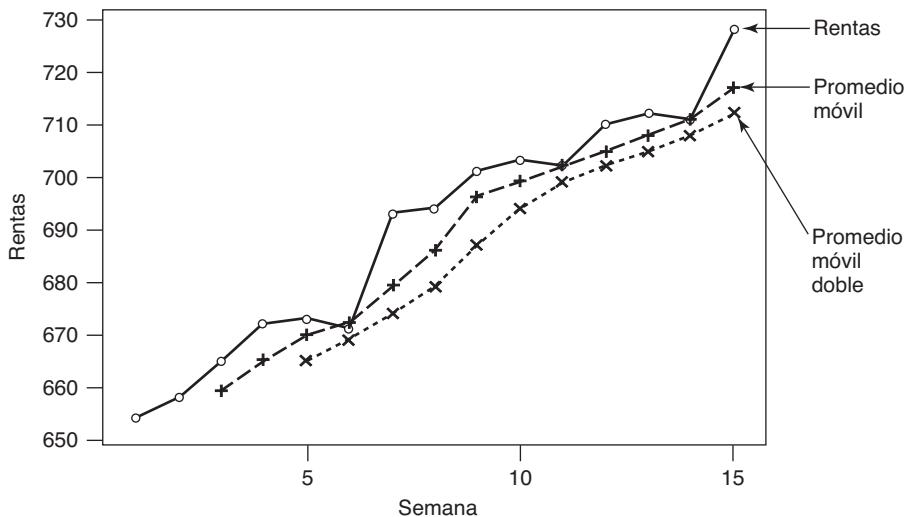
$$M_t = \hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \cdots + Y_{t-k+1}}{k}$$

Entonces, se usa la ecuación 4.9 para calcular el segundo promedio móvil.

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \cdots + M_{t-k+1}}{k} \quad (4.9)$$

La ecuación 4.10 se usa para desarrollar un pronóstico que consiste en sumar la diferencia entre los dos promedios móviles al primero.

$$a_t = M_t + (M_t - M'_t) = 2M_t - M'_t \quad (4.10)$$



**FIGURA 4.6** Promedios móviles sencillo y doble de tres semanas para los datos de Movie Video Store

La ecuación 4.11 es un factor de ajuste adicional, que es como la medida del cambio en la pendiente a lo largo de la serie.

$$b_t = \frac{2}{k-1} (M_t - M'_t) \quad (4.11)$$

Por último, se usa la ecuación 4.12 para hacer el pronóstico de  $p$  períodos en el futuro.

$$\hat{Y}_{t+p} = a_t + b_t p \quad (4.12)$$

donde

$k$  = número de períodos en el promedio móvil

$p$  = número de períodos futuros por pronosticar

#### Ejemplo 4.4

La Movie Video Store tiene en operación varios establecimientos de renta en Denver, Colorado. La compañía está creciendo y necesita expandir su inventario a fin de dar cabida a la demanda creciente por sus servicios. El presidente de la compañía comisiona a Jill Ottenbreit para que pronostique las rentas del mes siguiente. Los datos de rentas de las últimas 15 semanas están disponibles y se presentan en la tabla 4.5. Al principio, Jill trata de desarrollar un pronóstico con el uso de un promedio móvil de tres semanas. El  $MSE$  para este modelo es de 133. Dado que los datos tienen una tendencia obvia, se encuentra con que sus pronósticos subestiman las rentas reales de manera consistente. Por esta razón, decide intentar con un promedio móvil doble. Los resultados se presentan en la tabla 4.5. A continuación, se presentan los cálculos con el propósito de comprender el pronóstico para la semana 16. Se emplea la ecuación 4.8 para calcular el promedio móvil de tres semanas (columna tres).

$$M_{15} = \hat{Y}_{15+1} = \frac{Y_{15} + Y_{15-1} + Y_{15-3+1}}{3}$$

$$M_{15} = \hat{Y}_{16} = \frac{728 + 711 + 712}{3} = 717$$

**TABLA 4.5 Pronóstico de promedio móvil doble para Movie Video Store del ejemplo 4.4**

| (1)                     | (2)                       | (3)                                                            | (4)                                                   | (5)               | (6)               | (7)                                                   | (8)                  |
|-------------------------|---------------------------|----------------------------------------------------------------|-------------------------------------------------------|-------------------|-------------------|-------------------------------------------------------|----------------------|
| <i>Tiempo semanales</i> | <i>Ventas</i><br><i>t</i> | <i>Promedio móvil de tres semanas,</i><br><i>M<sub>t</sub></i> | <i>Promedio móvil doble,</i><br><i>M'<sub>t</sub></i> | <i>Valor de a</i> | <i>Valor de b</i> | <i>Pronóstico</i><br><i>a + bp</i><br>( <i>p</i> = 1) | <i>e<sub>t</sub></i> |
| 1                       | 654                       | —                                                              | —                                                     | —                 | —                 | —                                                     | —                    |
| 2                       | 658                       | —                                                              | —                                                     | —                 | —                 | —                                                     | —                    |
| 3                       | 665                       | 659                                                            | —                                                     | —                 | —                 | —                                                     | —                    |
| 4                       | 672                       | 665                                                            | —                                                     | —                 | —                 | —                                                     | —                    |
| 5                       | 673                       | 670                                                            | 665                                                   | 675               | 5                 | —                                                     | —                    |
| 6                       | 671                       | 672                                                            | 669                                                   | 675               | 3                 | 680                                                   | -9                   |
| 7                       | 693                       | 679                                                            | 674                                                   | 684               | 5                 | 678                                                   | 15                   |
| 8                       | 694                       | 686                                                            | 679                                                   | 693               | 7                 | 689                                                   | 5                    |
| 9                       | 701                       | 696                                                            | 687                                                   | 705               | 9                 | 700                                                   | 1                    |
| 10                      | 703                       | 699                                                            | 694                                                   | 704               | 5                 | 714                                                   | -11                  |
| 11                      | 702                       | 702                                                            | 699                                                   | 705               | 3                 | 709                                                   | -7                   |
| 12                      | 710                       | 705                                                            | 702                                                   | 708               | 3                 | 708                                                   | 2                    |
| 13                      | 712                       | 708                                                            | 705                                                   | 711               | 3                 | 711                                                   | 1                    |
| 14                      | 711                       | 711                                                            | 708                                                   | 714               | 3                 | 714                                                   | -3                   |
| 15                      | 728                       | 717                                                            | 712                                                   | 722               | 5                 | 717                                                   | 11                   |
| 16                      | —                         | —                                                              | —                                                     | —                 | —                 | 727                                                   | —                    |

$$MSE = 63.7.$$

A continuación se usa la ecuación 4.9 para calcular el promedio móvil doble (columna 4).

$$M'_{15} = \frac{M_{15} + M_{15-1} + M_{15-3+1}}{3}$$

$$M'_{15} = \frac{717 + 711 + 708}{3} = 712$$

La ecuación 4.10 se utiliza para calcular la diferencia entre los dos promedios móviles (columna 5).

$$a_{15} = 2M_{15} - M'_{15} = 2(717) - 712 = 722$$

La ecuación 4.11 ajusta la pendiente (columna 6).

$$b_{15} = \frac{2}{3-1}(M_{15} - M'_{15}) = \frac{2}{2}(717 - 712) = 5$$

Se utiliza la ecuación 4.12 para hacer el pronóstico de un periodo en el futuro (columna 7).

$$\hat{Y}_{15+1} = a_{15} + b_{15}p = 722 + 5(1) = 727$$

El pronóstico de cuatro semanas en el futuro es

$$\hat{Y}_{15+4} = a_{15} + b_{15}p = 722 + 5(4) = 742$$

Observe que el *MSE* se ha reducido de 133 a 63.7.

Parece razonable que las observaciones más recientes quizás contengan más información importante. En la siguiente sección se introduce un procedimiento que pone mayor énfasis en las observaciones más recientes.

## MÉTODOS DE SUAVIZAMIENTO EXPONENCIAL

---

Mientras que el método de promedios móviles sólo toma en cuenta las observaciones más recientes, el suavizamiento exponencial simple proporciona un promedio móvil con un peso exponencial de todos los valores observados con anterioridad. Con frecuencia, el modelo es apropiado para los datos que no tienen una tendencia predecible hacia arriba o hacia abajo. El objetivo es estimar el nivel real. Luego, esta estimación del nivel se utiliza como el pronóstico de los valores futuros.

El suavizamiento exponencial revisa continuamente el valor estimado a la luz de experiencias más recientes. Este método se basa en promediar (suavizar) los valores pasados de una serie en una forma exponencialmente decreciente. La observación más próxima recibe el peso más grande  $\alpha$  (donde  $0 < \alpha < 1$ ), la siguiente observación más cercana recibe menos peso  $\alpha(1 - \alpha)$ , la observación de dos períodos anteriores en el pasado recibe incluso menos peso  $\alpha(1 - \alpha)^2$ , y así sucesivamente.

En una representación de suavizamiento exponencial, el pronóstico nuevo (para el tiempo  $t + 1$ ) puede concebirse como la suma ponderada de la nueva observación (en el tiempo  $t$ ) y el pronóstico viejo (para el tiempo  $t$ ). El peso  $\alpha$  (donde  $0 < \alpha < 1$ ) se da al nuevo valor observado y el peso  $(1 - \alpha)$  se da al último pronóstico. Así,

$$\text{Nuevo pronóstico} = [\alpha \times (\text{observación nueva})] + [(1 - \alpha) \times (\text{último pronóstico})]$$

De manera más formal, la ecuación del suavizamiento exponencial es:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (4.13)$$

donde

$\hat{Y}_{t+1}$  = nuevo valor suavizado o el valor del pronóstico para el siguiente periodo

$\alpha$  = constante de suavizamiento ( $0 < \alpha < 1$ )

$Y_t$  = nueva observación o valor real de una serie en el periodo  $t$

$\hat{Y}_t$  = antiguo valor suavizado o pronóstico para el periodo  $t$

La ecuación 14.3 puede expresarse como

$$\begin{aligned}\hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha) \hat{Y}_t = \alpha Y_t + \hat{Y}_t - \alpha \hat{Y}_t \\ \hat{Y}_{t+1} &= \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)\end{aligned}$$

De esta manera, el nuevo pronóstico ( $\hat{Y}_{t+1}$ ) es el viejo ( $\hat{Y}_t$ ) ajustado en  $\alpha$  veces el error ( $Y_t - \hat{Y}_t$ ) en el pronóstico antiguo.

En la ecuación 4.13, la constante de suavizamiento  $\alpha$  sirve como factor de peso. El valor de  $\alpha$  determina la medida en que la observación actual influye en el pronóstico de la siguiente observación. Cuando  $\alpha$  se acerca a uno, básicamente el nuevo pronóstico será la última observación real. (De forma equivalente, el nuevo pronóstico será el pronóstico antiguo más un ajuste sustancial por el error que haya ocurrido en el pronóstico anterior.) A la inversa, cuando  $\alpha$  se acerca a cero, el pronóstico nuevo será muy similar al pronóstico anterior y la última observación real tendrá poca importancia.

El *suavizamiento exponencial* es un procedimiento para revisar de forma continua un pronóstico a la luz de experiencias más recientes.

Por último, la ecuación 4.13 supone que, para el tiempo  $t$ ,  $\hat{Y}_t = \hat{Y}_{t-1} + (1 - \alpha)$   $\hat{Y}_{t-1}$  y al sustituir  $\hat{Y}_t$  en la ecuación 4.13 se obtiene:

$$\begin{aligned}\hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha)\hat{Y}_t = \alpha Y_t + (1 - \alpha)[\alpha Y_{t-1} + (1 - \alpha)\hat{Y}_{t-1}] \\ \hat{Y}_{t+1} &= \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + (1 - \alpha)^2\hat{Y}_{t-1}\end{aligned}$$

Al continuar con la sustitución (a continuación sustituya por  $\hat{Y}_{t-1}$  y así sucesivamente) se muestra que  $\hat{Y}_{t+1}$  puede expresarse como la suma ponderada de la  $Y$  actual y las anteriores con pesos que descienden exponencialmente, o

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2Y_{t-2} + \alpha(1 - \alpha)^3Y_{t-3} + \dots \quad (4.14)$$

Es decir,  $\hat{Y}_{t+1}$  es un valor suavizado exponencialmente. La velocidad a la que pierden su impacto las observaciones pasadas depende del valor de  $\alpha$ , como se demuestra en la tabla 4.6.

Las ecuaciones 4.13 y 4.14 son equivalentes, aunque la primera es la que normalmente se utiliza para calcular el pronóstico  $\hat{Y}_{t+1}$  porque requiere menos almacenamiento de datos y se aplica con mayor facilidad.

El valor asignado a  $\alpha$  es fundamental para el análisis. Si se desea que las predicciones sean estables y que se suavicen las variaciones aleatorias, se requiere un  $\alpha$  de valor pequeño. Si se desea una respuesta rápida a un cambio real en el patrón de observaciones, es apropiado un  $\alpha$  de valor mayor. Un método para estimar  $\alpha$  es mediante un procedimiento iterativo para minimizar el error cuadrático medio (*MSE*) dado por la ecuación 3.8. Se calculan pronósticos utilizando valores 0.1, 0.2, ..., 0.9, como una aproximación a un decimal para  $\alpha$ , a continuación se calcula la suma de los errores cuadráticos pronosticados para cada valor. Se elige el valor de  $\alpha$  que produzca el error más pequeño para usarlo en la generación futura de pronósticos.

Para aplicar el algoritmo representado por la ecuación 4.13, se debe fijar un pronóstico inicial. Una manera de fijar el primer pronóstico es tomarlo como la primera observación. El ejemplo 4.5 lo ilustrará. Otro método consiste en utilizar el promedio de las primeras cinco o seis observaciones del valor suavizado inicial.

#### Ejemplo 4.5

La técnica de suavizamiento exponencial para Acme Tool Company —desde 1996 hasta el 2002— se demuestra en la tabla 4.7 y en la figura 4.7, donde se usan constantes de suavizamiento de 0.1 y 0.6. Los datos para el primer trimestre de 2002 se usarán como datos de prueba para ayudar a determinar el mejor valor de  $\alpha$  (entre los dos que se consideren). La serie suavizada de forma exponencial se calcula al fijar el pronóstico inicial  $\alpha_1$  igual a 500 que es la primera observación. Si hay datos anteriores, sería posible usarlos para desarrollar una serie suavizada hasta 1996 y emplear esta experiencia como el valor inicial de la serie suavizada. A continuación, se muestran los cálculos que llevan al pronóstico de los períodos 3 y 4.

**TABLA 4.6 Comparación de los efectos de diferentes constantes de suavizamiento**

| <i>Periodo</i>  | $\alpha = .1$          |             | $\alpha = .6$          |             |
|-----------------|------------------------|-------------|------------------------|-------------|
|                 | <i>Cálculos</i>        | <i>Peso</i> | <i>Cálculos</i>        | <i>Peso</i> |
| $t$             |                        | .100        |                        | .600        |
| $t - 1$         | .9 × .1                | .090        | .4 × .6                | .240        |
| $t - 2$         | .9 × .9 × .1           | .081        | .4 × .4 × .6           | .096        |
| $t - 3$         | .9 × .9 × .9 × .1      | .073        | .4 × .4 × .4 × .6      | .038        |
| $t - 4$         | .9 × .9 × .9 × .9 × .1 | .066        | .4 × .4 × .4 × .4 × .6 | .015        |
| Todas las demás |                        | .590        |                        | .011        |
|                 |                        |             |                        |             |
|                 | Totales                | 1.000       |                        | 1.000       |

**TABLA 4.7 Pronósticos de ajuste por suavizamiento exponencial para las ventas de Acme Tool Company del ejemplo 4.5**

| <i>Año</i> | <i>Trimestres</i> | <i>Tiempo</i> | <i>Valor real</i> | <i>Valor pronosticado por suavizamiento</i> | <i>Error de pronóstico</i> | <i>Valor pronosticado por suavizamiento</i> | <i>Error de pronóstico</i> |
|------------|-------------------|---------------|-------------------|---------------------------------------------|----------------------------|---------------------------------------------|----------------------------|
|            |                   |               |                   | $\hat{Y}_t (\alpha = .1)$                   | $e_t$                      | $\hat{Y}_t (\alpha = .6)$                   | $e_t$                      |
| 1996       | 1                 |               | 500               | 500.0                                       | 0.0                        | 500.0                                       | 0.0                        |
|            | 2                 |               | 350               | 500.0                                       | 150.0                      | 500.0                                       | 150.0                      |
|            | 3                 |               | 250               | 485.0 (1)                                   | 235.0 (2)                  | 410.0                                       | 160.0                      |
|            | 4                 |               | 400               | 461.5 (3)                                   | 61.5                       | 314.0                                       | 86.0                       |
| 1997       | 5                 |               | 450               | 455.4                                       | 5.4                        | 365.6                                       | 84.4                       |
|            | 6                 |               | 350               | 454.8                                       | 104.8                      | 416.2                                       | 66.2                       |
|            | 7                 |               | 200               | 444.3                                       | 244.3                      | 376.5                                       | 176.5                      |
|            | 8                 |               | 300               | 419.9                                       | 119.9                      | 270.6                                       | 29.4                       |
| 1998       | 9                 |               | 350               | 407.9                                       | 57.9                       | 288.2                                       | 61.8                       |
|            | 10                |               | 200               | 402.1                                       | 202.1                      | 325.3                                       | 125.3                      |
|            | 11                |               | 150               | 381.9                                       | 231.9                      | 250.1                                       | 100.1                      |
|            | 12                |               | 400               | 358.7                                       | 41.3                       | 190.0                                       | 210.0                      |
| 1999       | 13                |               | 550               | 362.8                                       | 187.2                      | 316.0                                       | 234.0                      |
|            | 14                |               | 350               | 381.6                                       | 31.5                       | 456.4                                       | 106.4                      |
|            | 15                |               | 250               | 378.4                                       | 128.4                      | 392.6                                       | 142.6                      |
|            | 16                |               | 550               | 365.6                                       | 184.4                      | 307.0                                       | 243.0                      |
| 2000       | 17                |               | 550               | 384.0                                       | 166.0                      | 452.8                                       | 97.2                       |
|            | 18                |               | 400               | 400.6                                       | 0.6                        | 511.1                                       | 111.1                      |
|            | 19                |               | 350               | 400.5                                       | 50.5                       | 444.5                                       | 94.5                       |
|            | 20                |               | 600               | 395.5                                       | 204.5                      | 387.8                                       | 212.2                      |
| 2001       | 21                |               | 750               | 415.9                                       | 334.1                      | 515.1                                       | 234.9                      |
|            | 22                |               | 500               | 449.3                                       | 50.7                       | 656.0                                       | 156.0                      |
|            | 23                |               | 400               | 454.4                                       | 54.4                       | 562.4                                       | 162.4                      |
|            | 24                |               | 650               | 449.0                                       | 201.0                      | 465.0                                       | 185.0                      |
| 2002       | 25                |               | 850               | 469.0                                       | 381.0                      | 576.0                                       | 274.0                      |

Los números entre paréntesis se refieren a las explicaciones dadas en el ejemplo 4.5 del texto.

1. Mediante la ecuación 4.13 en el periodo 2, el pronóstico para el periodo 3 con  $\alpha = 0.1$  es

$$\begin{aligned}\hat{Y}_{2+1} &= \alpha Y_2 + (1 - \alpha) \hat{Y}_2 \\ \hat{Y}_3 &= .1(350) + .9(500) = 485\end{aligned}$$

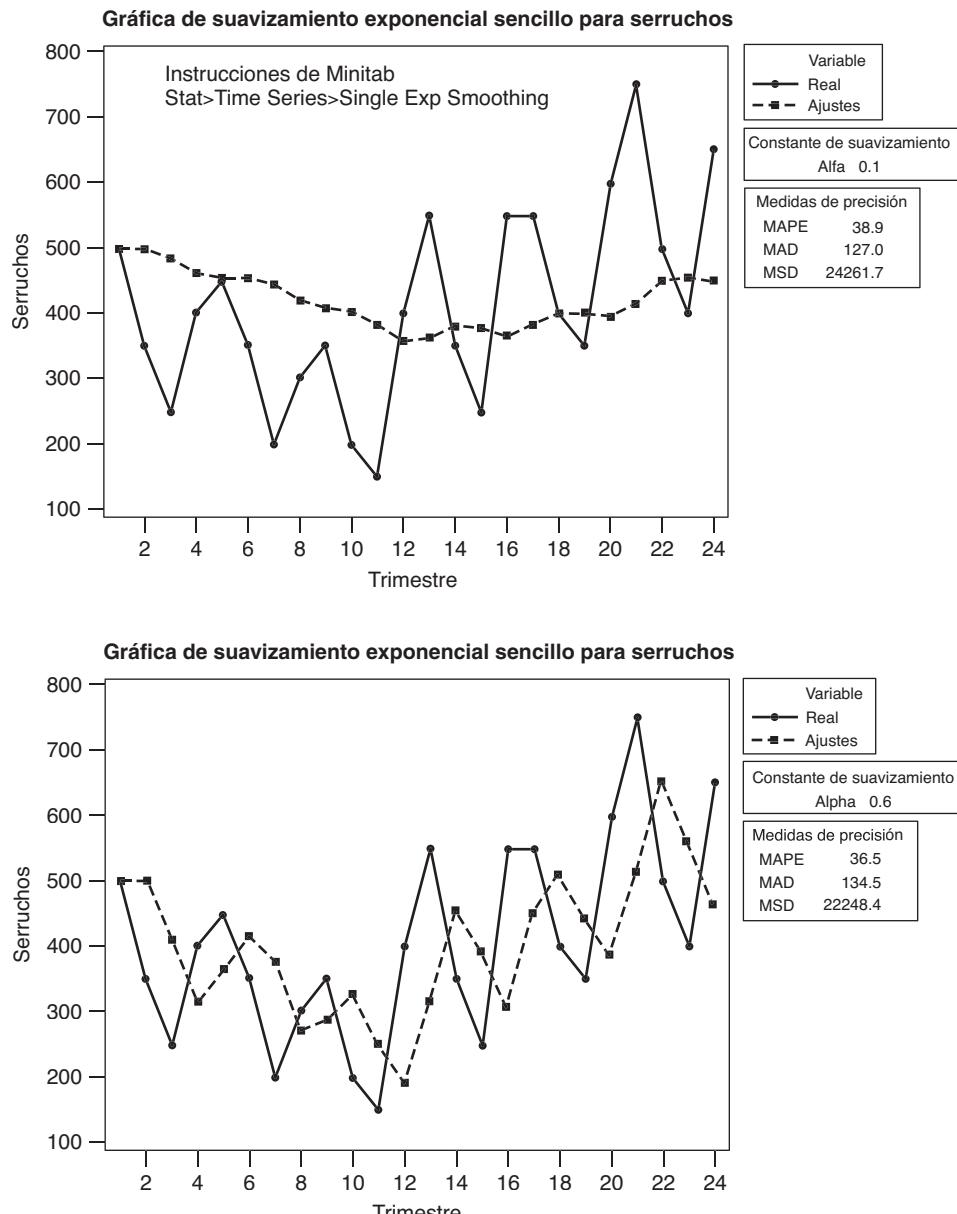
2. El error en este pronóstico es

$$e_3 = Y_3 - \hat{Y}_3 = 250 - 485 = -235$$

3. El pronóstico para el periodo 4 es

$$\begin{aligned}\hat{Y}_{3+1} &= \alpha Y_3 + (1 - \alpha) \hat{Y}_3 \\ \hat{Y}_4 &= .1(250) + .9(485) = 461.5\end{aligned}$$

De la tabla 4.7, cuando la constante de suavizamiento es 0.1, el pronóstico para el primer trimestre de 2002 es 469 con un error cuadrático de 145 161. Cuando la constante de suavizamiento es de 0.6, el pronóstico para el primer trimestre de 2002 es de 576 con un error cuadrático de 75 076. Con base en esta evidencia limitada, se desempeña mejor el suavizamiento exponencial con  $\alpha = 0.6$  que con  $\alpha = 0.1$ .



**FIGURA 4.7** Suavizamiento exponencial para los datos de Acme Tool Company:  
 $\alpha = 0.1$  (superior) y  $\alpha = 0.6$  (inferior)

Observe en la figura 4.7 que los valores suavizados son muy estables para la constante de suavizamiento 0.1. Con base en la minimización del error cuadrático medio *MSE* (al *MSE* se le llama *MSD* en los resultados de Minitab) durante los primeros 24 trimestres, la constante de suavizamiento de 0.6 es mejor. Si se comparan los errores porcentuales absolutos medios (*MAPE*), la constante de suavizamiento de 0.6 también es mejor. Para resumir:

$$\begin{array}{lll} \alpha = .1 & MSE = 24,262 & MAPE = 38.9\% \\ \alpha = .6 & MSE = 22,248 & MAPE = 36.5\% \end{array}$$

Sin embargo, tanto *MSE* como *MAPE* son grandes y, con base en este resumen estadístico, es evidente que el suavizado exponencial no es un método adecuado de pronóstico para estos datos. Como se verá, un método de suavizado que tome en cuenta la estacionalidad funciona mejor para predecir la venta de serruchos de Acme Tool Company.

Además del valor de  $\alpha$ , los valores de los pronósticos se ven afectados por la elección del valor inicial,  $\hat{Y}_1$ , para la serie suavizada. En la tabla 4.7 (vea el ejemplo 4.5), se usó  $\hat{Y}_1 = Y_1$  como valor inicial suavizado. Esta elección tiende a darle demasiado peso a  $Y_1$  en pronósticos subsecuentes. Por fortuna, la influencia del pronóstico inicial disminuye en gran medida al aumentar  $t$ .

Otro método para iniciar el procedimiento de suavizado consiste en promediar las primeras  $k$  observaciones. Entonces, el suavizado comienza como

$$\hat{Y}_1 = \frac{1}{k} \sum_{t=1}^k Y_t$$

A menudo se elige como  $k$  un número relativamente pequeño. Por ejemplo, el enfoque predeterminado en Minitab está fijo en  $k = 6$ .

#### Ejemplo 4.6

El cálculo del valor inicial como promedio para los datos de Acme Tool Company, presentado en el ejemplo 4.5, se muestra a continuación. Si  $k$  se elige igual a 6, el valor inicial es

$$\hat{Y}_1 = \frac{1}{6} \sum_{t=1}^6 Y_t = \frac{1}{6}(500 + 350 + 250 + 400 + 450 + 350) = 383.3$$

El *MSE* y el *MAPE* de cada alfa, cuando se usa un pronóstico inicial de 383.3, se muestran a continuación.

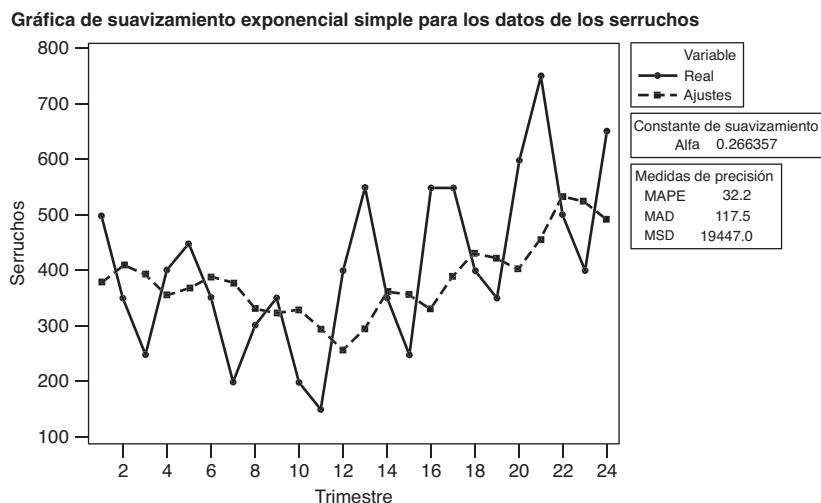
|               |                |                 |
|---------------|----------------|-----------------|
| $\alpha = .1$ | $MSE = 21,091$ | $MAPE = 32.1\%$ |
| $\alpha = .6$ | $MSE = 22,152$ | $MAPE = 36.7\%$ |

El valor inicial  $\hat{Y}_1 = 383.3$  llevó a una disminución en *MSE* y *MAPE* para  $\alpha = 0.1$ , pero no tuvo mayor efecto cuando  $\alpha = 0.6$ . Ahora el mejor modelo, basado en las medidas de resumen de *MSE* y *MAPE* parece ser el que use  $\alpha = 0.1$  en lugar de 0.6.

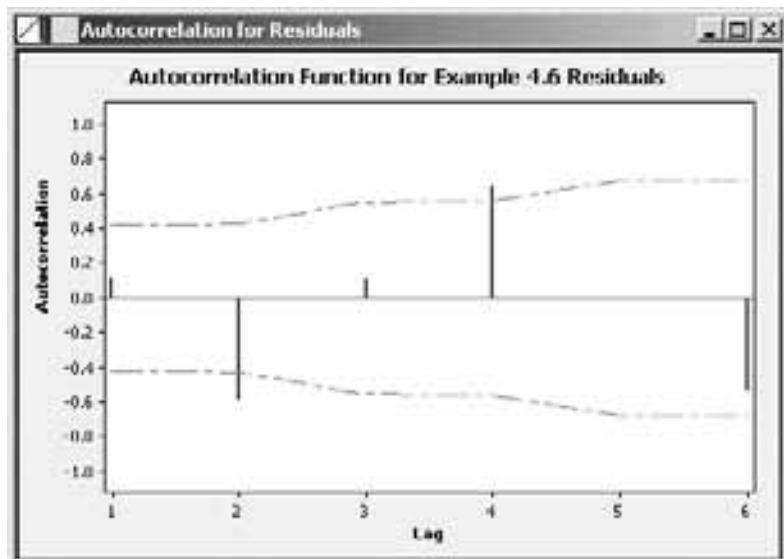
La figura 4.8 muestra los resultados del ejemplo 4.5 cuando los datos se corren en Minitab (vea la sección “Aplicaciones en Minitab” al final del capítulo para leer las instrucciones). La constante de  $\alpha = 0.266$  se seleccionó de forma automática al minimizar el *MSE*. Éste se reduce a 19 477, el *MAPE* es igual a 32.2% y, aunque no se muestra, el *MPE* es igual a -6.4%. El pronóstico para el primer trimestre de 2002 es de 534.

La figura 4.9 muestra la función de autocorrelación para los residuales del método del suavizado exponencial si se usa un alfa de 0.266. Al llevar a cabo la prueba Ljung-Box con seis retrasos de tiempo, el gran valor de LBQ (33.86) muestra que el conjunto de las primeras seis autocorrelaciones residuales es mayor de lo que se esperaría si los residuales fueran aleatorios. Específicamente, las autocorrelaciones de residuales tan significativamente grandes en los retrasos 2 y 4 indican que la variación estacional de los datos no se explica mediante un suavizado exponencial simple.

A menudo, el suavizado exponencial es un buen procedimiento de pronóstico cuando una serie de tiempo no aleatoria exhibe un comportamiento con tendencia. Resulta útil desarrollar una medida que se pueda aplicar para determinar cuándo ha cambiado el patrón básico de una serie de tiempo. Una *señal de control* es una forma de supervisar los cambios. Ésta implica calcular una medida de errores de pronóstico a lo largo del tiempo y, también, establecer límites para que cuando los errores salgan de dichas fronteras, se alerte al pronosticador.



**FIGURA 4.8** Suavizado exponencial con  $\alpha = 0.266$  para los datos de Acme Tool Company



Autocorrelation Function: Residuals

| Lag | ACF       | T     | LBQ   | Lag | ACF       | T     | LBQ   |
|-----|-----------|-------|-------|-----|-----------|-------|-------|
| 1   | 0.121421  | 0.59  | 0.40  | 4   | 0.646967  | 2.40  | 23.65 |
| 2   | -0.586941 | -2.84 | 10.24 | 5   | -0.006908 | -0.02 | 23.65 |
| 3   | 0.109122  | 0.41  | 10.59 | 6   | -0.542625 | -1.65 | 33.86 |

**FIGURA 4.9** Función de autocorrelación para los residuales obtenidos al usar el suavizado exponencial con  $\alpha = 0.266$  con los datos de Acme Tool Company

Una *señal de control* implica calcular una medida de errores de pronóstico a lo largo del tiempo y establecer límites a fin de que, cuando el error acumulado salga de esas fronteras, se alerte al pronosticador.

Por ejemplo, podría usarse una señal de control para determinar cuándo debería cambiarse el tamaño de la constante de suavizado alfa ( $\alpha$ ). Debido a que, en general, se pronostica un gran número de artículos, es una práctica común continuar con el mismo valor de  $\alpha$  para varios períodos antes de intentar establecer si es necesaria una revisión. Desafortunadamente, la simplicidad de emplear un modelo de suavizado exponencial establecido es un gran motivador para no llevar a cabo el cambio. Pero en algún momento puede ser necesario actualizar o abandonar por completo el suavizado exponencial. Cuando el modelo genera pronósticos que contienen muchos errores, es adecuado hacer un cambio.

Un sistema de control es un método para supervisar la necesidad de cambio. Éste contiene un rango de desviaciones previsibles del pronóstico con respecto a los valores reales. Siempre que los pronósticos generados mediante suavizado exponencial caigan dentro de este rango, es innecesario hacer cambios en  $\alpha$ . No obstante, si un pronóstico se encuentra fuera del rango, el sistema señala que es necesario actualizar  $\alpha$ .

Por ejemplo, si las cosas van bien, la técnica de pronóstico debería sobreestimar y subestimar con igual frecuencia. Puede desarrollarse una señal de control que se base en esta lógica.

Sea  $U$  igual al número de subestimaciones de los últimos  $k$  pronósticos. En otras palabras,  $U$  es el número de errores entre los últimos valores  $k$  que son positivos. Si el proceso está controlado, el valor esperado de  $U$  es  $k/2$ , aunque está implicada la variabilidad muestral, así que los valores cercanos a  $k/2$  no serían raros. Por otro lado, los valores que no se acercan a  $k/2$  indicarían que la técnica está produciendo pronósticos con sesgo.

#### Ejemplo 4.7

Suponga que Acme Tool Company ha decidido utilizar la técnica de suavizado exponencial con  $\alpha$  igual a 0.1, como se muestra en el ejemplo 4.5 (vea la página 115). Si el proceso está controlado y el analista decide supervisar los últimos 10 valores del error,  $U$  tiene un valor esperado de 5. En realidad, los valores para  $U$  de 2, 3, 4, 6, 7 u 8 no serían demasiado alarmantes; sin embargo, los valores de 0, 1, 9 o 10 serían preocupantes dado que la probabilidad de obtener tales valores sólo por casualidad sería de 0.022 (basado en la distribución binomial). Con esta información, se puede desarrollar un sistema de control basado en las siguientes reglas.

Si  $2 \leq U \leq 8$  ser entonces el proceso está controlado.

Si  $U < 2$  o  $U > 8$  entonces el proceso está fuera de control.

Suponga que, de los siguientes 10 pronósticos que usaron esta técnica, sólo uno tiene un error positivo. Puesto que la probabilidad de obtener sólo un error positivo de entre 10 es bastante baja —de 0.011—, se considera que el proceso está fuera de control (sobreestima) y el valor de  $\alpha$  debe cambiarse.

Otra forma de controlar una técnica de pronóstico es determinar el rango en que deben estar los errores de pronóstico, lo cual puede lograrse al utilizar el *MSE* que se estableció cuando se determinó  $\alpha$  del tamaño óptimo. Si la técnica de suavizado exponencial es razonablemente precisa, el error de pronóstico deberá distribuirse, casi normalmente, alrededor de una media de cero. En esta condición, existe una posibilidad cercana a 95% de que la observación real caiga dentro de aproximadamente dos desviaciones estándar del pronóstico. Si se usa  $\sqrt{MSE}$  como estimado de la desviación estándar del error de pronóstico, se pueden determinar límites de error aproximados a 95%. Los errores de pronóstico que caigan dentro de estos límites indican que no hay razón para alarmarse. Los errores (en particular, una secuencia de errores) fuera de los límites sugieren un cambio. El ejemplo 4.8 ilustra este enfoque.

**Ejemplo 4.8**

En el ejemplo de Acme Tool Company, se determinó que el valor alfa óptimo sería  $\alpha = 0.266$  con  $MSE = 19\,447$ . Un estimado de la desviación estándar de los errores de pronóstico es  $\sqrt{MSE} = \sqrt{19\,447} = 139.5$ . Si los errores de pronóstico se distribuyen de manera casi normal alrededor de una media de cero, hay una posibilidad cercana a 95% de que la observación real caiga dentro de dos desviaciones estándar del pronóstico o dentro de

$$\pm 2\sqrt{MSE} = \pm 2\sqrt{19,447} = \pm 2(139.5) = \pm 279$$

Para este ejemplo, el error absoluto permisible es de 279. Si en cualquier pronóstico futuro el valor absoluto del error es mayor que 279, habrá razón para creer que la constante de suavizamiento óptimo alfa debe actualizarse o debe considerarse otro método de pronóstico.

El análisis anterior sobre las señales de control también tiene aplicación en los métodos de suavizamiento que se presentarán en el resto del capítulo.

El suavizamiento exponencial simple funciona bien cuando los datos varían alrededor de un nivel que cambia con poca frecuencia. Siempre que exista una tendencia sostenida, el suavizamiento exponencial se retrasará con respecto a los valores reales a lo largo del tiempo. La técnica de suavizamiento exponencial lineal de Holt, que se diseñó para manejar datos con una tendencia bien definida, trata este problema y se presenta a continuación.

### Suavizamiento exponencial ajustado a la tendencia: método de Holt

El nivel de la serie de tiempo puede cambiar ocasionalmente y cuando se usa el suavizamiento exponencial simple se requieren actualizaciones del nivel. En algunas situaciones, los datos observados tendrán una tendencia clara e información que permita anticipar movimientos futuros hacia arriba. Cuando éste sea el caso, será necesaria una función de pronóstico de tendencia lineal. Dado que las series económicas y de negocios rara vez exhiben una tendencia lineal fija, se considera la posibilidad de hacer un modelo de las tendencias lineales locales que evolucionan con el tiempo. Holt (1957) desarrolló su *método de suavizamiento exponencial lineal*,<sup>1</sup> el cual toma en cuenta las tendencias lineales locales en evolución dentro de una serie de tiempo y puede usarse para generar pronósticos.

Cuando se anticipa la tendencia en la serie de tiempo, se requiere un estimado de la pendiente y del nivel actual. La técnica de Holt suaviza el nivel y la pendiente de manera directa al usar diferentes constantes de suavizamiento para cada una. Éstas proporcionan estimados del nivel y de la pendiente que se adaptan a lo largo del tiempo conforme aparecen nuevas observaciones. Una de las ventajas de la técnica de Holt es que su flexibilidad al seleccionar los coeficientes que controlan el nivel y la tendencia.

Las tres ecuaciones que se usan en el método de Holt son:

1. La serie suavizada exponencialmente, o estimado del nivel actual:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.15)$$

2. El estimado de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.16)$$

3. Pronóstico del periodo  $p$  en el futuro:

$$\hat{Y}_{t+p} = L_t + pT_t \quad (4.17)$$

---

<sup>1</sup> El suavizamiento exponencial lineal de Holt a veces se llama *suavizamiento exponencial doble*.

donde

- $L_t$  = nuevo valor suavizado (estimado del nivel actual)
- $\alpha$  = constante de suavizado para el nivel ( $0 < \alpha < 1$ )
- $Y_t$  = observación nueva o valor real de la serie en el periodo  $t$
- $\beta$  = constante de suavizado para el estimado de tendencia ( $0 < \beta < 1$ )
- $T_t$  = estimado de tendencia
- $p$  = periodo a pronosticarse en el futuro
- $\hat{Y}_{t+p}$  = pronóstico para el periodo  $p$  en el futuro

La ecuación 4.15 es muy similar a la ecuación del suavizado exponencial simple —4.13—, excepto que se ha incorporado el término ( $T_{t-1}$ ) para actualizar adecuadamente el nivel cuando existe una tendencia. En otras palabras, el nivel actual ( $L_t$ ) se calcula al tomar un promedio ponderado de dos estimados del nivel —un estimado está dado por la observación actual ( $Y_t$ ) y el otro estimado se obtiene al sumar la tendencia anterior ( $T_{t-1}$ ) al nivel suavizado anteriormente ( $L_{t-1}$ ). Si no existe tendencia en los datos, no habrá necesidad de que exista el término  $T_{t-1}$  en la ecuación 4.15, de hecho, reduciéndola a la ecuación 4.13. Tampoco hay necesidad de la ecuación 4.16.

Se usa una segunda constante de suavizado,  $\beta$ , para crear el estimado de la tendencia. La ecuación 4.16 muestra que la tendencia actual ( $T_t$ ) es un promedio ponderado (con pesos  $\beta$  y  $1 - \beta$ ) de dos estimados de tendencia: uno está dado por el cambio en el nivel desde el tiempo  $t - 1$  hasta  $t$  ( $L_t - L_{t-1}$ ), y el otro es la tendencia suavizada anteriormente ( $T_{t-1}$ ). La ecuación 4.16 es similar a la ecuación 4.15, excepto que el suavizado se realiza para la tendencia y no para los datos reales.

La ecuación 4.17 muestra el pronóstico para el periodo  $p$  en el futuro. Para un pronóstico que se realiza en el tiempo  $t$ , el estimado actual de la tendencia ( $T_t$ ) se multiplica por el número del periodo a pronosticarse ( $p$ ), y el producto se suma al nivel actual ( $L_t$ ). Observe que los pronósticos para los períodos futuros se encuentran en una línea recta con una pendiente  $T_t$  e intersección con el eje  $Y$  de  $L_t$ .

Igual que con el suavizado exponencial simple, las constantes de suavizado  $\alpha$  y  $\beta$  pueden seleccionarse de manera subjetiva o mediante la minimización de una medida del error de pronóstico como el *MSE*. Los pesos grandes propician cambios más rápidos en el componente; los pesos pequeños dan como resultado cambios menos rápidos. Por lo tanto, cuanto mayores sean los pesos, los valores de suavizado seguirán más a los datos y cuanto menores sean los pesos, los valores de suavizado seguirán más a los valores suavizado previos.

Se podría desarrollar una retícula de valores de  $\alpha$  y  $\beta$  (por ejemplo, cada combinación de  $\alpha = 0.1, 0.2, \dots, 0.9$  y  $\beta = 0.1, 0.2, \dots, 0.9$ ) y a continuación seleccionar la combinación que proporcione el *MSE* más bajo. La mayoría del software de pronóstico utiliza un algoritmo de optimización para minimizar el *MSE*. Se podría decir que  $\alpha = \beta$ , y de esta manera proporcionar iguales cantidades de suavizado para el nivel y la tendencia. En el caso especial donde  $\alpha = \beta$ , el enfoque de Holt es el mismo que el suavizado exponencial doble de Brown.

Para comenzar, deben determinarse los valores iniciales para  $L$  y  $T$  en las ecuaciones 4.15 y 4.16. Un enfoque es fijar el primer estimado del nivel suavizado igual a la primera observación, y se estima que la tendencia inicial es cero. Un segundo enfoque consiste en utilizar el promedio de las primeras cinco o seis observaciones como el valor inicial suavizado  $L$ . A continuación, se estima la tendencia con el uso de la pendiente de una línea que se ajusta a estas cinco o seis observaciones. Minitab lo hace mediante una regresión en donde  $Y$  es la variable dependiente y  $X$  es la variable independiente (el tiempo). La constante de esta ecuación es el estimado inicial del componente nivel y la pendiente, o coeficiente de regresión, es el estimado inicial del componente tendencia.

**Ejemplo 4.9**

En el ejemplo 4.6 el suavizamiento exponencial simple no originó pronósticos exitosos de las ventas de serruchos de Acme Tool Company. Puesto que la figura 4.8 sugiere que podría existir tendencia en los datos, se usa el suavizamiento exponencial de Holt para desarrollar pronósticos. Para comenzar los cálculos mostrados en la tabla 4.8, se requieren dos valores estimados iniciales; específicamente, el nivel inicial y el valor inicial de la tendencia. El estimado del nivel se estableció como igual a la primera observación. Se estimó que la tendencia es igual a cero. La técnica se demuestra en la tabla 4.8 para valores de  $\alpha = 0.3$  y  $\beta = 0.1$ .

El valor de  $\alpha$  utilizado aquí es cercano al alfa óptimo ( $\alpha = 0.266$ ) para el suavizamiento exponencial simple del ejemplo 4.6. Alfa se usa para suavizar los datos a fin de eliminar lo aleatorio y estimar el nivel. La constante de suavizamiento  $\beta$  es como  $\alpha$ , a excepción de que se usa para suavizar la tendencia en los datos. Ambas constantes de suavizamiento se usan para promediar los valores anteriores y, de esa forma, eliminar lo aleatorio. A continuación, se muestran los cálculos que llevan al pronóstico para el periodo 3.

1. Actualizar el nivel la serie suavizada exponencialmente:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$L_2 = .3Y_2 + (1 - .3)(L_1 + T_1)$$

$$L_2 = .3(350) + .7(500 + 0) = 455$$

**TABLA 4.8** Valores de las ventas de Acme Tool Company suavizados exponencialmente mediante el método de Holt para el ejemplo 4.9

| (1)<br>Año | (2)<br><i>t</i> | (3)<br><i>Y<sub>t</sub></i> | (4)<br><i>L<sub>t</sub></i> | (5)<br><i>T<sub>t</sub></i> | (6)<br>$\hat{Y}_{t+p}$ | <i>e<sub>t</sub></i> |
|------------|-----------------|-----------------------------|-----------------------------|-----------------------------|------------------------|----------------------|
| 1996       | 1               | 500                         | 500.0                       | 0.0                         | 500.0                  | 0.0                  |
|            | 2               | 350                         | 455.0                       | -4.5                        | 500.0                  | -150.0               |
|            | 3               | 250                         | 390.4                       | -10.5                       | 450.5                  | -200.5               |
|            | 4               | 400                         | 385.9                       | -9.9                        | 379.8                  | 20.2                 |
| 1997       | 5               | 450                         | 398.2                       | -7.7                        | 376.0                  | 74.0                 |
|            | 6               | 350                         | 378.3                       | -8.9                        | 390.5                  | -40.5                |
|            | 7               | 200                         | 318.6                       | -14.0                       | 369.4                  | -169.4               |
|            | 8               | 300                         | 303.2                       | -14.1                       | 304.6                  | -4.6                 |
| 1998       | 9               | 350                         | 307.4                       | -12.3                       | 289.1                  | 60.9                 |
|            | 10              | 200                         | 266.6                       | -15.2                       | 295.1                  | -95.0                |
|            | 11              | 150                         | 221.0                       | -18.2                       | 251.4                  | -101.4               |
|            | 12              | 400                         | 262.0                       | -12.3                       | 202.8                  | 197.2                |
| 1999       | 13              | 550                         | 339.8                       | -3.3                        | 249.7                  | 300.3                |
|            | 14              | 350                         | 340.6                       | -2.9                        | 336.5                  | 13.5                 |
|            | 15              | 250                         | 311.4                       | 5.5                         | 337.7                  | -287.7               |
|            | 16              | 550                         | 379.1                       | 1.8                         | 305.9                  | 244.1                |
| 2000       | 17              | 550                         | 431.7                       | 6.9                         | 381.0                  | 169.0                |
|            | 18              | 400                         | 427.0                       | 5.7                         | 438.6                  | -38.6                |
|            | 19              | 350                         | 407.9                       | 3.3                         | 432.7                  | -82.7                |
|            | 20              | 600                         | 467.8                       | 8.9                         | 411.2                  | 188.8                |
| 2001       | 21              | 750                         | 558.7                       | 17.1                        | 476.8                  | 273.2                |
|            | 22              | 500                         | 553.1                       | 14.8                        | 575.9                  | -75.9                |
|            | 23              | 400                         | 517.6                       | 9.8                         | 567.9                  | -167.9               |
|            | 24              | 650                         | 564.2                       | 13.5                        | 527.4                  | 122.6                |
| 2002       | 25              | 850                         | —                           | —                           | 577.7                  | 273.3                |

MSE = 20,515.5.

## 2. Actualizar el estimado de tendencia

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$T_2 = .1(L_2 - L_{2-1}) + (1 - .1)T_{2-1}$$

$$T_2 = .1(455 - 500) + .9(0) = -4.5$$

## 3. Pronosticar el periodo uno en el futuro:

$$\hat{Y}_{t+p} = L_t + pT_t$$

$$\hat{Y}_{2+1} = L_2 + 1T_2 = 455 + 1(-4.5) = 450.5$$

## 4. Determinar el error de pronóstico:

$$e_3 = Y_3 - \hat{Y}_3 = 250 - 450.5 = -200.5$$

El pronóstico para el periodo 25 se calcula de la siguiente manera:

## 1. Actualizar la serie o nivel suavizado exponencialmente:

$$L_{24} = .3Y_{24} + (1 - .3)(L_{24-1} + T_{24-1})$$

$$L_{24} = .3(650) + .7(517.6 + 9.8) = 564.2$$

## 2. Actualizar el estimado de tendencia:

$$T_{24} = .1(L_{24} - L_{24-1}) + (1 - .1)T_{24-1}$$

$$T_{24} = .1(564.2 - 517.6) + .9(9.8) = 13.5$$

## 3. Pronosticar un periodo hacia el futuro:

$$\hat{Y}_{24+1} = L_{24} + 1T_{24}$$

$$\hat{Y}_{25} = 564.2 + 1(13.5) = 577.7$$

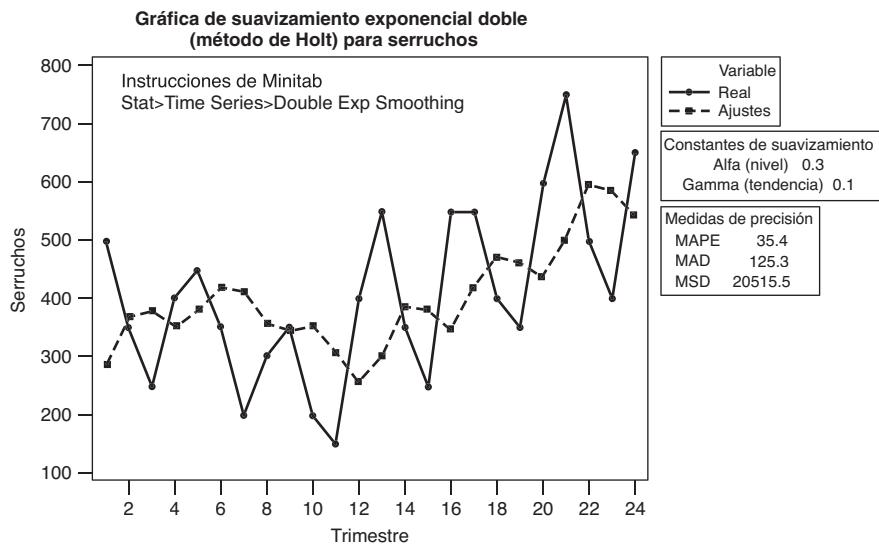
Mediante el criterio del *MSE* mínimo durante el periodo de 1996 a 2002, el suavizado lineal de Holt (con  $\alpha = 0.3$  y  $\beta = 0.1$ ) no reproduce los datos mejor que el suavizado exponencial simple que usó una constante de suavizado de 0.266. Si se comparan los *MAPE*, se puede ver que son casi iguales. Cuando los pronósticos de las ventas reales del primer trimestre de 2002 se comparan, otra vez el suavizado exponencial simple y el suavizado de Holt son similares. El resumen es:

$$\alpha = .266 \quad MSE = 19,447 \quad MAPE = 32.2\%$$

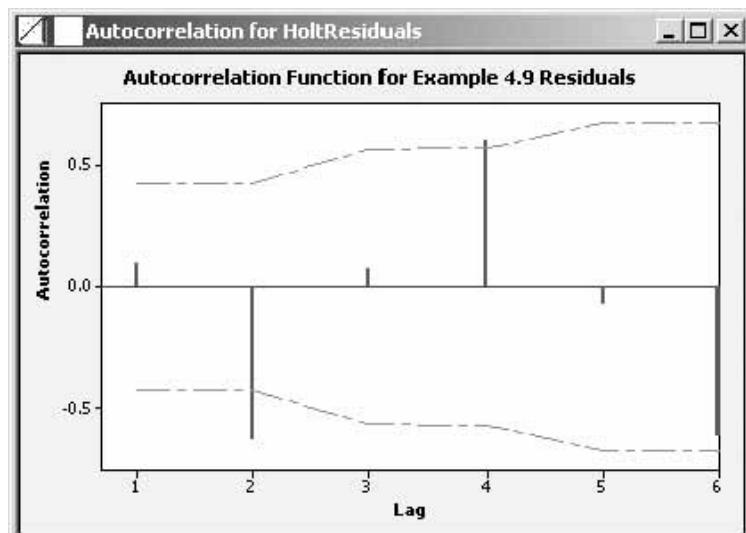
$$\alpha = .3, \beta = .1 \quad MSE = 20,516 \quad MAPE = 35.4\%$$

La figura 4.10 muestra los resultados cuando se utiliza el método de Holt con  $\alpha = 0.3$  y  $\beta = 0.1$  en Minitab.<sup>2</sup> La función de autocorrelación para los residuales del suavizado exponencial lineal de Holt se dan en la figura 4.11. Los coeficientes de autocorrelación de los retrasos de tiempo 2 y 4 parecen ser importantes. Además, cuando se calcula la estadística *Q* de Ljung-Box para los seis retrasos de tiempo, el alto valor de LBQ (36.33) muestra que los residuales contienen amplia autocorrelación: no son aleatorios. Las gran-

<sup>2</sup> En el programa Minitab el parámetro de tendencia gamma ( $\gamma$ ) corresponde al de nuestra beta ( $\beta$ ).



**FIGURA 4.10 Suavizado exponencial lineal de Holt: datos de Acme Tool Company**



**Autocorrelation Function: Residuals**

| Lag | ACF       | T     | LBQ   | Lag | ACF       | T     | LBQ   |
|-----|-----------|-------|-------|-----|-----------|-------|-------|
| 1   | 0.092547  | 0.45  | 0.23  | 4   | 0.601371  | 2.18  | 22.92 |
| 2   | -0.629408 | -3.06 | 11.47 | 5   | -0.070132 | -0.22 | 23.08 |
| 3   | 0.074355  | 0.27  | 11.63 | 6   | -0.618267 | -1.90 | 36.33 |

**FIGURA 4.11 Función de autocorrelación para los residuales del suavizado exponencial lineal de Holt: datos de Acme Tool Company**

des autocorrelaciones residuales en los retrasos 2 y 4 sugieren que podría haber un componente estacional en los datos de Acme Tool Company.

Los resultados en el ejemplo 4.6 y 4.9 (vea las figuras 4.8 y 4.10) no son muy diferentes debido a que la constante de suavizado  $\alpha$  es casi la misma en ambos casos, y la constante de suavizado  $\beta$  en el ejemplo 4.9 es pequeña. (Cuando  $\beta = 0$ , el suavizado lineal de Holt se reduce a un suavizado exponencial simple.)

### Suavizado exponencial ajustado para variaciones de tendencia y estacionales: método de Winters

El examen de los datos de Acme Tool Company en la tabla 4.8 indica que las ventas son más elevadas, de manera consistente, durante los trimestres primero y cuarto y más bajas durante el tercero. Tal parece que hay un patrón estacional. El método de Winters de tres parámetros de suavizado exponencial lineal y estacional —que es una extensión del método de Holt— podría describir mejor los datos y reducir el error de pronóstico. En el método de Winters se usa una ecuación adicional para estimar la estacionalidad. En la versión multiplicativa del método de Winters, el estimado de estacionalidad será como un índice estacional y se calcula mediante la ecuación 4.20. Dicha ecuación muestra que el componente estacional actual,  $S_t$ , se calcula  $\gamma$  veces, como un estimado del índice estacional dado por  $Y_t/L_t$ , sumado a  $(1 - \gamma)$  veces el componente estacional anterior  $S_{t-s}$ . Este procedimiento es equivalente a suavizar los valores actuales y previos de  $Y_t/L_t$ . La razón por la cual  $Y_t$  se divide entre el estimado del nivel actual  $L_t$  es para crear un índice (proporción) que pueda utilizarse de forma multiplicativa, con el fin de ajustar un pronóstico y explicar los picos y valles estacionales.

Las cuatro ecuaciones utilizadas en el suavizado (multiplicativo) de Winters son:

1. El nivel de la serie estimado por suavizado exponencial:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.18)$$

2. El estimado de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.19)$$

3. El estimado de estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (4.20)$$

4. Pronóstico del periodo  $p$  en el futuro:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (4.1)$$

donde

$L_t$  = nuevo valor suavizado o nivel actual estimado

$\alpha$  = constante de suavizado para el nivel

$Y_t$  = nueva observación o valor real en el periodo  $t$

$\beta$  = constante de suavizado para el estimado de tendencia

$T_t$  = estimado de tendencia

$\gamma$  = constante de suavizado para el estimado de estacionalidad

$S_t$  = estimado de estacionalidad

$p$  = número del periodo futuro a pronosticarse

$s$  = longitud de la estacionalidad

$\hat{Y}_{t+p}$  = pronóstico para el periodo  $p$  en el futuro

La ecuación 4.18 actualiza la serie suavizada. Hay una pequeña diferencia en esta ecuación que la distingue de la correspondiente en el procedimiento de Holt —ecuación 4.15—. En la ecuación 4.18,  $Y_t$  se divide entre  $S_{t-s}$ , lo cual ajusta a  $Y_t$  a la estacionalidad y, de esta manera, se eliminan los efectos estacionales que pudieran existir en los datos originales  $Y_t$ .

Después de que se han suavizado el estimado de tendencia y el estimado de estacionalidad en las ecuaciones 4.19 y 4.20, se obtiene un pronóstico mediante la ecuación 4.21. Es casi igual a la fórmula correspondiente —la ecuación 4.17— utilizada para obtener un pronóstico mediante el suavizamiento de Holt. La diferencia está en que este estimado de los períodos futuros,  $t + p$ , se multiplica por  $S_{t-s+p}$ . Este índice estacional es el último disponible y, por lo tanto, se usa para ajustar el pronóstico a la estacionalidad.

Igual que con el suavizamiento exponencial lineal de Holt, los pesos  $\alpha$ ,  $\beta$  y  $\gamma$  pueden elegirse de forma subjetiva o mediante un criterio de mínimo error de pronósticos, como el *MSE*. El enfoque más común para determinar estos valores es utilizar un algoritmo de optimización a fin de encontrar las constantes de suavizamiento óptimas.

Para comenzar el algoritmo de la ecuación 4.18, deben fijarse los valores iniciales de la serie suavizada  $L_t$ , la tendencia  $T_t$  y los índices de estacionalidad  $S_t$ . Un enfoque consistente en fijar el primer estimado de la serie suavizada (nivel) igual a la primera observación. A continuación, se estima la tendencia para que sea igual a cero y los índices estacionales se fijan en 1.0, cada uno. Existen otros métodos para iniciar los estimados de nivel, tendencia y estacionalidad. Por ejemplo, Minitab desarrolla una ecuación de regresión que usa el tiempo como la variable independiente  $X$  y la variable de interés como  $Y$ . La constante de esta ecuación es el estimado inicial de la serie suavizada —o componente nivel— y la pendiente —o coeficiente de regresión— es el estimado inicial del componente tendencia. Los valores iniciales de los componentes estacionales se obtienen a partir de una regresión de variable ficticia que usa datos sin tendencia (vea el capítulo 8).

#### Ejemplo 4.10

La técnica de Winters se muestra en la tabla 4.9 para  $\alpha = 0.4$ ,  $\beta = 0.1$  y  $\gamma = 0.3$  para los datos de la Acme Tool Company. El valor de alfa es similar al utilizado para el suavizamiento exponencial simple del ejemplo 4.6 y se usa para suavizar los datos y crear un estimado del nivel. La constante de suavizamiento beta ( $\beta$ ) se usa para crear un estimado de la tendencia. La constante de suavizamiento gamma ( $\gamma$ ) se usa para crear un estimado suavizado del componente estacional en los datos.

Minitab puede usarse para resolver este ejemplo (vea la sección de “Aplicaciones en Minitab” al final del capítulo para leer las instrucciones).<sup>3</sup> Los resultados se muestran en la tabla 4.9 y la figura 4.12. El pronóstico del primer trimestre del año 2002 es de 778.2. A continuación, se muestran los cálculos que llevaron a este valor pronosticado para el primer trimestre del 2002, o periodo 25.

1. La serie suavizada exponencialmente o estimado de nivel:

$$\begin{aligned} L_t &= \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ L_{24} &= .4 \frac{Y_{24}}{S_{24-4}} + (1 - .4)(L_{24-1} + T_{24-1}) \\ L_{24} &= .4 \frac{650}{1.39628} + (1 - .4)(501.286 + 9.148) \\ L_{24} &= .4(465.52) + .6(510.434) = 492.469 \end{aligned}$$

<sup>3</sup> En el programa Minitab, el parámetro de tendencia gamma ( $\gamma$ ) corresponde a nuestro beta ( $\beta$ ) y el parámetro estacional delta ( $\delta$ ) equivale a nuestra gamma ( $\gamma$ ) en las ecuaciones 4.19 y 4.20, respectivamente.

**TABLA 4.9** Valores suavizados exponencialmente mediante el método de Winters para las ventas de Acme Tool Company del ejemplo 4.10

| <i>Año</i> | (1)<br><i>t</i> | (2)<br><i>Y<sub>t</sub></i> | (3)<br><i>L<sub>t</sub></i> | (4)<br><i>T<sub>t</sub></i> | (5)<br><i>S<sub>t</sub></i> | (6)<br><i>Ŷ<sub>t+p</sub></i> | (7)<br><i>e<sub>t</sub></i> |
|------------|-----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-------------------------------|-----------------------------|
| 1996       | 1               | 500                         | 415.459                     | -41.9541                    | 1.26744                     | 563.257                       | -63.257                     |
|            | 2               | 350                         | 383.109                     | -40.9937                    | 0.89040                     | 328.859                       | 21.141                      |
|            | 3               | 250                         | 358.984                     | -39.3068                    | 0.66431                     | 222.565                       | 27.435                      |
|            | 4               | 400                         | 328.077                     | -38.4668                    | 1.18766                     | 375.344                       | 24.656                      |
| 1997       | 5               | 450                         | 315.785                     | -35.8494                    | 1.31471                     | 367.063                       | 82.937                      |
|            | 6               | 350                         | 325.194                     | -31.3235                    | 0.94617                     | 249.255                       | 100.745                     |
|            | 7               | 200                         | 296.748                     | -31.0358                    | 0.66721                     | 195.221                       | 4.779                       |
|            | 8               | 300                         | 260.466                     | -31.5604                    | 1.17690                     | 315.576                       | -15.576                     |
| 1998       | 9               | 350                         | 243.831                     | -30.0679                    | 1.35093                     | 300.945                       | 49.055                      |
|            | 10              | 200                         | 212.809                     | -30.1632                    | 0.94426                     | 202.255                       | -2.255                      |
|            | 11              | 150                         | 199.515                     | -28.4764                    | 0.69259                     | 121.863                       | 28.137                      |
|            | 12              | 400                         | 238.574                     | -21.7228                    | 1.32682                     | 201.294                       | 198.706                     |
| 1999       | 13              | 550                         | 292.962                     | -14.1117                    | 1.50886                     | 292.949                       | 257.051                     |
|            | 14              | 350                         | 315.575                     | -10.4393                    | 0.99371                     | 263.306                       | 86.694                      |
|            | 15              | 250                         | 327.466                     | -8.2062                     | 0.71385                     | 211.335                       | 38.665                      |
|            | 16              | 550                         | 357.366                     | -4.3956                     | 1.39048                     | 423.599                       | 126.401                     |
| 2000       | 17              | 550                         | 357.588                     | -3.9339                     | 1.51763                     | 532.584                       | 17.416                      |
|            | 18              | 400                         | 373.206                     | -1.9787                     | 1.01713                     | 351.428                       | 48.572                      |
|            | 19              | 350                         | 418.856                     | 2.7843                      | 0.75038                     | 264.999                       | 85.001                      |
|            | 20              | 600                         | 425.586                     | 3.1788                      | 1.39628                     | 586.284                       | 13.716                      |
| 2001       | 21              | 750                         | 454.936                     | 5.7959                      | 1.55691                     | 650.706                       | 99.294                      |
|            | 22              | 500                         | 473.070                     | 7.0297                      | 1.02907                     | 468.626                       | 31.374                      |
|            | 23              | 400                         | 501.286                     | 9.1484                      | 0.76465                     | 360.255                       | 39.745                      |
|            | 24              | 650                         | 492.469                     | 7.3518                      | 1.37336                     | 712.712                       | -62.712                     |
| 2002       | 25              | 850                         | —                           | —                           | —                           | 778.179                       | —                           |
|            | 26              | 600                         | —                           | —                           | —                           | 521.917                       | —                           |
|            | 27              | 450                         | —                           | —                           | —                           | 393.430                       | —                           |
|            | 28              | 700                         | —                           | —                           | —                           | 716.726                       | —                           |

*MSE* = 7,636.86.

2. El estimado de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$T_{24} = .1(L_{24} - L_{24-1}) + (1 - .1)T_{24-1}$$

$$T_{24} = .1(492.469 - 501.286) + .9(9.148)$$

$$T_{24} = .1(-8.817) + .9(9.148) = 7.352$$

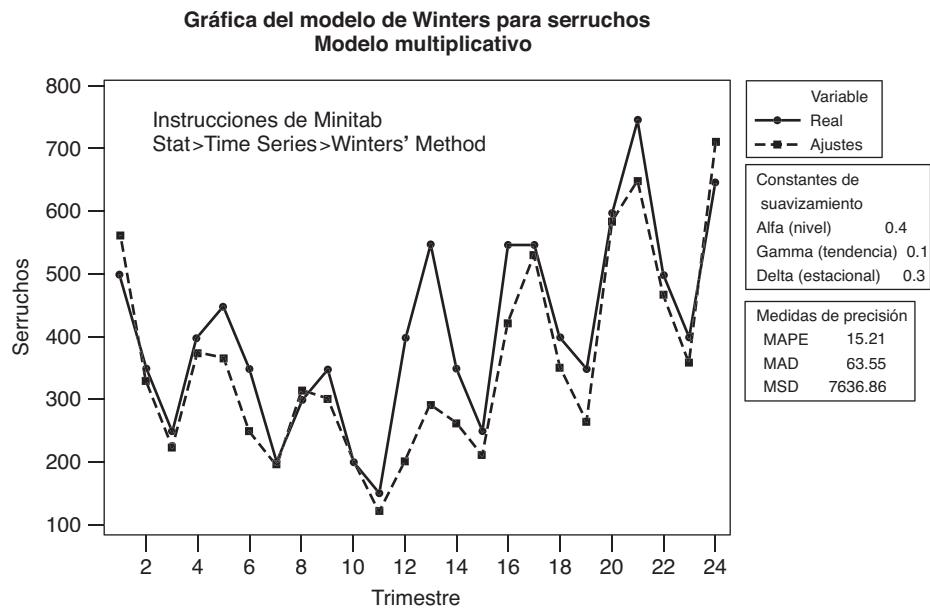
3. El estimado de estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s}$$

$$S_{24} = .3 \frac{Y_{24}}{L_{24}} + (1 - .3)S_{24-4}$$

$$S_{24} = .3 \frac{650}{492.469} + .7(1.39628)$$

$$S_{24} = .3(1.3199) + .9774 = 1.3734$$



**FIGURA 4.12 Suavizamiento exponencial de Winters: datos de Acme Tool Company**

4. Pronóstico para el primer periodo,  $p = 1$  en el futuro:

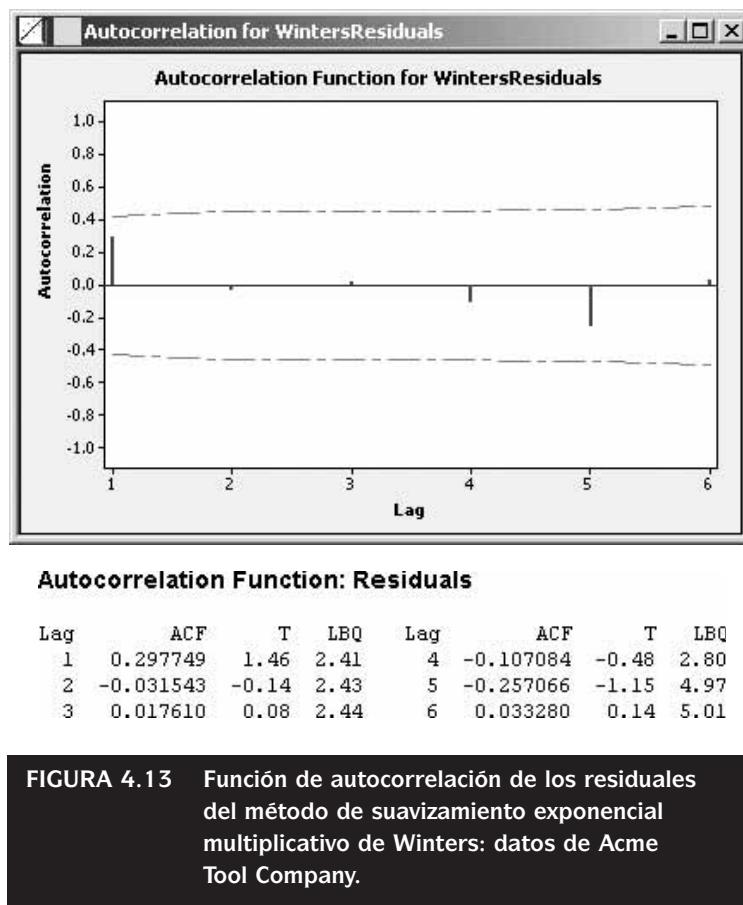
$$\begin{aligned}\hat{Y}_{24+1} &= (L_{24} + 1T_{24})S_{24-4+1} \\ \hat{Y}_{25} &= (492.469 + 1(7.352))1.5569 = 778.17\end{aligned}$$

Para los valores de parámetros que se han considerado, la técnica de Winters es mejor que los dos procedimientos de suavizamiento presentados antes en términos de minimizar el *MSE*. Cuando se comparan los pronósticos de las ventas reales del primer trimestre de 2002, la técnica de Winters también parece funcionar mejor. La figura 4.13 muestra la función de autocorrelación para los residuales de suavizamiento exponencial de Winters. Al parecer, ninguno de los coeficientes de autocorrelación residual es mucho mayor que cero. Cuando la estadística  $Q$  de Ljung-Box se calcula para seis retrasos de tiempo, el valor pequeño de LBQ (5.01) muestra que la serie residual es aleatoria. En apariencia, el método de suavización exponencial de Winters proporciona pronósticos adecuados para los datos de Acme Tool Company.

El método de Winters proporciona una forma fácil de explicar la estacionalidad cuando los datos tienen un patrón estacional. Un método alterno consiste en eliminar la estacionalidad o hacer un ajuste estacional de los datos. Éste es un proceso que elimina los efectos de la estacionalidad de los datos originales que se mostrará en el capítulo 5. El modelo de pronóstico se aplica a los datos sin estacionalidad y, si fuera necesario, se reinserta el componente estacional para proporcionar pronósticos precisos.

El suavizamiento exponencial es una técnica popular para los pronósticos de corto plazo. Sus mayores ventajas son bajo costo y simplicidad. Cuando se necesitan pronósticos para sistemas de inventario que contienen miles de artículos, los métodos de suavización suelen ser el único enfoque aceptable.

Los promedios móviles simples y el suavizamiento exponencial basan los pronósticos en promedios ponderados de mediciones pasadas. La lógica es que los valores pasados contienen información sobre lo que ha de ocurrir en el futuro. Dado que los valores pasados incluyen fluctuaciones aleatorias, así como información acerca del el patrón subyacente de la



variable, se hace un intento por suavizar los valores. Este enfoque supone que las fluctuaciones extremas representan lo aleatorio en una serie de observaciones históricas.

Por otro lado, los promedios móviles son medias aritméticas de un cierto número,  $k$ , de valores de una variable. El promedio más reciente es el pronóstico para el siguiente periodo. Este enfoque asigna un peso igual a cada valor pasado que entra al promedio. Sin embargo, se puede dar un argumento convincente a favor de utilizar todos los datos, aunque enfatizando los valores más recientes. Los métodos de suavizamiento exponencial son atractivos, ya que generan pronósticos al asignar pesos que disminuyen de forma exponencial conforme las observaciones se vuelven más antiguas.

## APLICACIÓN A LA ADMINISTRACIÓN

El pronóstico es uno de los apoyos más importantes que los administradores tienen en la toma de decisiones. Prácticamente toda decisión de trabajo importante depende, en cierta medida, de un pronóstico. El departamento de producción tiene que programar las necesidades de contratar personal y los pedidos de materia prima con uno o dos meses de anticipación; el departamento de finanzas debe determinar las mejores oportunidades de inversión; el de mercadeo debe pronosticar la demanda de un nuevo producto. La lista de aplicaciones de los pronósticos es bastante larga.

Los ejecutivos conocen muy bien la importancia de los pronósticos. De hecho, se pasan mucho tiempo estudiando las tendencias en los asuntos económicos y políticos y cómo

certos acontecimientos podrían afectar la demanda de productos o servicios. Un punto de interés es la importancia que los ejecutivos dan a los métodos cuantitativos de pronóstico en comparación con sus propias opiniones. Esta cuestión es especialmente sensible cuando se implican sucesos que tienen una repercusión significativa sobre la demanda. Un problema con los métodos cuantitativos de pronóstico es que dependen de los datos históricos. Por esta razón, son relativamente ineficaces para anticipar un cambio dramático que suele propiciar una demanda mucho más alta o más baja.

Los métodos de promedio y suavizamiento presentados en este capítulo son útiles por su relativa simplicidad. Asimismo, tienden a ser menos costosos, más fáciles de aplicar y de entender que los métodos complejos. Con frecuencia, el costo y las dificultades de los modelos más sofisticados superan cualquier ganancia en precisión. Por ello, los pequeños negocios descubren que los métodos simples son prácticos. Los negocios sin computadoras o personal capaz de manejar modelos estadísticos también recurren a los métodos sencillos. Es frecuente que los administradores de negocios se enfrenten a la necesidad de preparar pronósticos de corto plazo para una serie de artículos diferentes. Un ejemplo típico es el administrador que debe planear la producción con base en algún pronóstico de la demanda de varios cientos de diferentes productos de una línea específica. Por último, los negocios nuevos sin grandes bases de datos históricos encuentran útiles estos métodos.

Con una elección juiciosa del orden  $k$ , el método de promedio móvil se ajusta bien a los cambios en los niveles. Su actualización es económica y no requiere mucho almacenamiento de datos. Los métodos de promedios móviles se usan con mayor frecuencia cuando se requieren pronósticos repetidos.

El suavizamiento exponencial es una técnica popular cuya fortaleza se encuentra en la buena precisión de corto plazo, en combinación con las actualizaciones rápidas y de bajo costo. Comúnmente, la técnica se usa cuando se necesitan pronósticos mensuales o semanales de forma regular para un gran número, quizás miles, de artículos. El control de inventarios es un ejemplo en el que los métodos de suavizamiento exponencial se usan de manera rutinaria.

## Glosario

**Promedio móvil.** El valor medio de  $k$  observaciones consecutivas es un promedio móvil de orden  $k$ . El valor más reciente del promedio móvil ofrece un pronóstico para el siguiente periodo.

**Promedio simple.** Usa la media de todas las observaciones históricas relevantes como pronóstico del siguiente periodo.

**Señal de control.** Ésta implica calcular una medida de errores de pronóstico a lo largo del tiempo y establecer límites a fin de que, cuando el error acumulativo salga de dichos límites, se alerte al pronosticador.

**Suavizamiento exponencial.** Es un procedimiento para revisar continuamente un pronóstico a la luz de las experiencias más recientes.

## Fórmulas clave

### Modelo informal

$$\hat{Y}_{t+1} = Y_t \quad (4.1)$$

### Modelo informal con tendencia

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \quad (4.2)$$

### Modelo informal de tasa de cambio

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} \quad (4.3)$$

**Modelo informal estacional para datos trimestrales**

$$\hat{Y}_{t+1} = Y_{t-3} \quad (4.4)$$

**Modelo informal estacional y con tendencia para datos trimestrales**

$$\hat{Y}_{t+1} = Y_{t-3} + \frac{Y_t - Y_{t-4}}{4} \quad (4.5)$$

**Modelo de promedio simple**

$$\hat{Y}_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i \quad (4.6)$$

**Promedio simple actualizado, nuevo periodo**

$$\hat{Y}_{t+2} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t+1} \quad (4.7)$$

**Modelo de promedio móvil para  $k$  periodos**

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-k+1}}{k} \quad (4.8)$$

**Promedio móvil doble**

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \dots + M_{t-k+1}}{k} \quad (4.9)$$

$$a_t = 2M_t - M'_t \quad (4.10)$$

$$b_t = \frac{2}{k-1}(M_t - M'_t) \quad (4.11)$$

$$\hat{Y}_{t+p} = a_t + b_t p \quad (4.12)$$

**Suavizado exponencial simple**

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \quad (4.13)$$

Expresión alterna equivalente:

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2Y_{t-2} + \alpha(1-\alpha)^3Y_{t-3} + \dots \quad (4.14)$$

**Suavizado lineal de Holt**

Estimado del nivel actual de la serie suavizada exponencialmente:

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (4.15)$$

Estimado de tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (4.16)$$

Pronóstico del periodo  $p$  en el futuro:

$$\hat{Y}_{t+p} = L_t + pT_t \quad (4.17)$$

### Suavizamiento exponencial de Winters con estacionalidad multiplicativa

Nivel estimado de serie suavizada exponencialmente:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.18)$$

Estimado de tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.19)$$

Estimado de estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (4.20)$$

Pronóstico del periodo  $p$  en el futuro:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (4.21)$$

## Problemas

1. ¿Cuál técnica de pronóstico revisa continuamente un estimado a la luz de las experiencias más recientes?
2. ¿Cuál técnica de pronóstico utiliza el valor del periodo actual como un pronóstico para el periodo siguiente?
3. ¿Cuál técnica de pronóstico asigna pesos iguales a cada una de las observaciones?
4. ¿Cuál(es) técnica(s) de pronóstico debería(n) aplicarse si los datos muestran una tendencia?
5. ¿Cuáles técnicas de pronóstico deberían aplicarse si los datos son estacionales?
6. Appex Mutual Fund es una sociedad de inversión que invierte principalmente en valores de tecnología. Los precios de las acciones de la sociedad de inversión al final de cada mes para los 12 meses de 2003 se presentan en la tabla P.6:
  - a) Encuentre el valor de pronóstico de las acciones del fondo mutuo o sociedad de inversión para cada mes por medio de un modelo informal. El valor para diciembre de 2002 fue de 19.00.

**TABLA P.6**

| Mes        | Precio de las acciones de la sociedad de inversión |
|------------|----------------------------------------------------|
| Enero      | 19.39                                              |
| Febrero    | 18.96                                              |
| Marzo      | 18.20                                              |
| Abril      | 17.89                                              |
| Mayo       | 18.43                                              |
| Junio      | 19.98                                              |
| Julio      | 19.51                                              |
| Agosto     | 20.63                                              |
| Septiembre | 19.78                                              |
| Octubre    | 21.25                                              |
| Noviembre  | 21.18                                              |
| Diciembre  | 22.14                                              |

- b) Evalúe este método de pronóstico por medio de *MAD*.  
 c) Evalúe este método de pronóstico por medio de *MSE*.  
 d) Evalúe este método de pronóstico por medio de *MAPE*.  
 e) Evalúe este método de pronóstico por medio de *MPE*.  
 f) Por medio del modelo informal usado, pronostique el precio de las acciones de la sociedad de inversión para enero de 2004.  
 g) Redacte un resumen con sus descubrimientos.
7. Refiérase al problema 6 y utilice un promedio móvil de tres meses para pronosticar el precio de las acciones del fondo para enero de 2004. ¿Este pronóstico es mejor que el que se realizó por medio del modelo informal? Explíquelo.
8. Segundo la serie  $Y_t$  de la tabla P.8:
- ¿Cuál es el pronóstico para el periodo 9 cuando se utiliza un promedio móvil de cinco periodos?
  - Si se utiliza un suavizamiento exponencial simple con una constante de suavizamiento de 0.4, ¿cuál es el pronóstico para el periodo 4?
  - En el inciso b), ¿cuál es el error de pronóstico para el periodo 3?
9. El rendimiento en un bono de obligación general para la ciudad de Davenport fluctúa con el mercado. Las cotizaciones mensuales para 2002 se presentan en la tabla P.9.
- Encuentre el valor de pronóstico para el rendimiento de los bonos de obligación para cada mes, comenzando por abril, por medio de un promedio móvil de tres meses.
  - Encuentre el valor de pronóstico para el rendimiento de los bonos de obligación para cada mes, comenzando en junio, mediante un promedio móvil de cinco meses.
  - Evalúe estos métodos de pronóstico por medio de *MAD*.

**TABLA P.8**

| <i>Periodo</i> | $Y_t$ | $\hat{Y}_t$ | $e_t$ |
|----------------|-------|-------------|-------|
| 1              | 200   | 200         | —     |
| 2              | 210   | —           | —     |
| 3              | 215   | —           | —     |
| 4              | 216   | —           | —     |
| 5              | 219   | —           | —     |
| 6              | 220   | —           | —     |
| 7              | 225   | —           | —     |
| 8              | 226   | —           | —     |

**TABLA P.9**

| <i>Mes</i> | <i>Rendimiento</i> |
|------------|--------------------|
| Enero      | 9.29               |
| Febrero    | 9.99               |
| Marzo      | 10.16              |
| Abril      | 10.25              |
| Mayo       | 10.61              |
| Junio      | 11.07              |
| Julio      | 11.52              |
| Agosto     | 11.09              |
| Septiembre | 10.80              |
| Octubre    | 10.50              |
| Noviembre  | 10.86              |
| Diciembre  | 9.97               |

- d)* Evalúe estos métodos de pronóstico por medio de *MSE*.  
*e)* Evalúe estos métodos de pronóstico por medio de *MAPE*.  
*f)* Evalúe estos métodos de pronóstico por medio de *MPE*.  
*g)* Pronostique el rendimiento para enero de 2003 por medio de la mejor técnica.  
*h)* Redacte un resumen de sus descubrimientos.
10. Esta pregunta se refiere al problema 9. Utilice el suavizamiento exponencial con una constante de suavizamiento de 0.2 y un valor inicial de 9.29 para pronosticar el rendimiento de enero de 2003. ¿Este pronóstico es mejor que el realizado por el modelo de promedio móvil? Explique.
11. La Hughes Supply Company utiliza un método de administración de inventarios para determinar las demandas mensuales de diversos productos. Los valores en la demanda para los últimos 12 meses de cada producto han sido registrados y están disponibles para pronósticos futuros. Los valores de la demanda para los 12 meses de 2002 para un instrumento eléctrico se presentan en la tabla P.11.
- Utilice un suavizamiento exponencial con una constante de suavizamiento de 0.5 y un valor inicial de 205 para pronosticar la demanda para enero de 2003.
12. General American Investors Inc., una compañía de administración de inversiones, invierte principalmente en acciones de calidad media y elevada. Jim Campbell está estudiando el valor de los activos por acción para su compañía y desea pronosticar esta variable para los siguientes trimestres de 1996. Los datos se presentan en la tabla P.12.
- Evalúe la habilidad para pronosticar la variable del valor de los activos por acción; utilice los siguientes métodos de pronóstico: informal, de promedio móvil y de suavizamiento exponencial. Cuando compare las técnicas, considere que el valor real de los activos por acción para el segundo trimestre de 1996 fue de 26.47. Escriba un informe para Jim en donde indique cuál método debería utilizar y por qué.
13. Southdown Inc., uno de los productores más importantes de Estados Unidos, continúa operando un programa de quema de combustible de desecho. El costo total de la operación para Southdown será de \$37 millones. Para esta empresa, es extremadamente importante tener un pronóstico preciso de las utilidades para el primer trimestre de 2000. Los datos se presentan en la tabla P.13.

TABLA P.11

| <i>Mes</i> | <i>Demanda</i> |
|------------|----------------|
| Enero      | 205            |
| Febrero    | 251            |
| Marzo      | 304            |
| Abril      | 284            |
| Mayo       | 352            |
| Junio      | 300            |
| Julio      | 241            |
| Agosto     | 284            |
| Septiembre | 312            |
| Octubre    | 289            |
| Noviembre  | 385            |
| Diciembre  | 256            |

*Fuentes:* Registros de Hughes Supply Company.

**TABLA P.12**

| Año  | Trimestre |       |       |       |
|------|-----------|-------|-------|-------|
|      | 1         | 2     | 3     | 4     |
| 1985 | 16.98     | 18.47 | 17.63 | 20.65 |
| 1986 | 21.95     | 23.85 | 20.44 | 19.29 |
| 1987 | 22.75     | 23.94 | 24.84 | 16.70 |
| 1988 | 18.04     | 19.19 | 18.97 | 17.03 |
| 1989 | 18.23     | 19.80 | 22.89 | 21.41 |
| 1990 | 21.50     | 25.05 | 20.33 | 20.60 |
| 1991 | 25.33     | 26.06 | 28.89 | 30.60 |
| 1992 | 27.44     | 26.69 | 28.71 | 28.56 |
| 1993 | 25.87     | 24.96 | 27.61 | 24.75 |
| 1994 | 23.32     | 22.61 | 24.08 | 22.31 |
| 1995 | 22.67     | 23.52 | 25.41 | 23.94 |
| 1996 | 25.68     | —     | —     | —     |

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993, 1996), p. 2187.

**TABLA P.13 Utilidades de Southdown  
1986-1999**

| Año  | Trimestre |       |       |       |
|------|-----------|-------|-------|-------|
|      | 1         | 2     | 3     | 4     |
| 1986 | 77.4      | 88.8  | 92.1  | 79.8  |
| 1987 | 77.5      | 89.1  | 92.4  | 80.1  |
| 1988 | 74.4      | 185.2 | 162.4 | 178.1 |
| 1989 | 129.1     | 158.4 | 160.6 | 138.7 |
| 1990 | 127.2     | 149.8 | 151.7 | 132.9 |
| 1991 | 103.0     | 136.8 | 141.3 | 123.5 |
| 1992 | 107.3     | 136.1 | 138.6 | 123.7 |
| 1993 | 106.1     | 144.4 | 156.1 | 138.2 |
| 1994 | 111.8     | 149.8 | 158.5 | 141.8 |
| 1995 | 119.1     | 158.0 | 170.4 | 151.8 |
| 1996 | 127.4     | 178.2 | 189.3 | 169.5 |
| 1997 | 151.4     | 187.2 | 199.2 | 181.4 |
| 1998 | 224.9     | 317.7 | 341.4 | 300.7 |
| 1999 | 244.9     | 333.4 | 370.0 | 326.7 |

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993, 1996, 1999) p. 896.

- a) Utilice el suavizamiento exponencial con una constante de suavizamiento de 0.4 y un valor inicial de 77.4 para pronosticar las utilidades trimestrales para el primer trimestre de 2000.
- b) Ahora utilice una constante de suavizamiento de 0.6 y un valor inicial de 77.4 para pronosticar las utilidades trimestrales del primer trimestre de 2000.
- c) ¿Cuál constante de suavizamiento proporciona el mejor pronóstico?
- d) Remítase al inciso c). Examine las autocorrelaciones de los residuales. ¿Está satisfecho con el suavizamiento exponencial simple que se presenta en este ejemplo? Explique.

| <b>TABLA P.14</b> |                          | <b>Ventas por acción de Triton 1974 - 1999</b> |                          |
|-------------------|--------------------------|------------------------------------------------|--------------------------|
| <i>Año</i>        | <i>Ventas por acción</i> | <i>Año</i>                                     | <i>Ventas por acción</i> |
| 1974              | .93                      | 1987                                           | 5.33                     |
| 1975              | 1.35                     | 1988                                           | 8.12                     |
| 1976              | 1.48                     | 1989                                           | 10.65                    |
| 1977              | 2.36                     | 1990                                           | 12.06                    |
| 1978              | 2.45                     | 1991                                           | 11.63                    |
| 1979              | 2.52                     | 1992                                           | 6.58                     |
| 1980              | 2.81                     | 1993                                           | 2.96                     |
| 1981              | 3.82                     | 1994                                           | 1.58                     |
| 1982              | 5.54                     | 1995                                           | 2.99                     |
| 1983              | 7.16                     | 1996                                           | 3.69                     |
| 1984              | 1.93                     | 1997                                           | 3.98                     |
| 1985              | 5.17                     | 1998                                           | 4.39                     |
| 1986              | 7.72                     | 1999                                           | 6.85                     |

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993, 1996, 1999), p. 1872.

14. La Triton Energy Corporation explora y produce aceite y gas. El presidente de la compañía Gail Freeman desea que el analista de la compañía pronostique las ventas por cada acción de la compañía para el año 2000. Éste será un pronóstico importante porque los planes de reestructuración de la compañía enfrentan dificultades. Los datos se muestran en la tabla P.14.

Determine el mejor método de pronóstico y realice el pronóstico de las ventas por acción para el año 2000.

15. La Consolidated Edison Company vende electricidad (82% de utilidades), gas (13%) y vapor (5%) en la ciudad de Nueva York y el condado de Westchester. Bart Thomas, quien hace los pronósticos de la compañía, tiene la consigna de pronosticar las utilidades trimestrales de la empresa para el resto de 2002 y 2003. Él recopila los datos que se muestran en la tabla P.15.

Determine la mejor técnica de pronóstico y pronostique la utilidad trimestral para el resto de 2002 y 2003.

16. Un fabricante que se especializa en partes de reemplazo no cuenta con un sistema de pronóstico y programa la cantidad de manufactura de productos con base en las ventas del último mes. En la tabla P.16 aparecen los 24 meses de datos de ventas que están disponibles.

a) Grafique los datos de ventas como series de tiempo. ¿Los datos son estacionales?

*Pista:* Para los datos mensuales, el periodo estacional es  $s = 12$ . ¿Existe algún patrón, como ventas en el verano relativamente bajas o las de otoño relativamente altas, que tienda a repetirse cada 12 meses?

b) Utilice un modelo informal para generar los pronósticos de ventas mensuales (ejemplo: el pronóstico de febrero de 2001 se otorga por el valor de enero de 2001 y así sucesivamente). Calcule el *MAPE*.

c) Utilice un suavizamiento exponencial simple con una constante de suavizamiento de 0.5 y un valor inicial de 430 para generar pronósticos de ventas de cada mes. Calcule el *MAPE*.

d) ¿Cree usted que cualquiera de los modelos que se mencionan en los incisos b) y c) tiene la probabilidad de generar pronósticos precisos para futuras ventas mensuales? Explique.

**TABLA P.15 Utilidades trimestrales para Consolidated Edison (millones de dólares), 1986 - 1999**

| <i>Año</i> | <i>Mar. 31</i> | <i>Jun. 30</i> | <i>Sept. 30</i> | <i>Dic. 31</i> |
|------------|----------------|----------------|-----------------|----------------|
| 1985       | 1,441          | 1,209          | 1,526           | 1,321          |
| 1986       | 1,414          | 1,187          | 1,411           | 1,185          |
| 1987       | 1,284          | 1,125          | 1,493           | 1,192          |
| 1988       | 1,327          | 1,102          | 1,469           | 1,213          |
| 1989       | 1,387          | 1,218          | 1,575           | 1,371          |
| 1990       | 1,494          | 1,263          | 1,613           | 1,369          |
| 1991       | 1,479          | 1,330          | 1,720           | 1,344          |
| 1992       | 1,456          | 1,280          | 1,717           | 1,480          |
| 1993       | 1,586          | 1,396          | 1,800           | 1,483          |
| 1994       | 1,697          | 1,392          | 1,822           | 1,461          |
| 1995       | 1,669          | 1,460          | 1,880           | 1,528          |
| 1996       | 1,867          | 1,540          | 1,920           | 1,632          |
| 1997       | 1,886          | 1,504          | 2,011           | 1,720          |
| 1998       | 1,853          | 1,561          | 2,062           | 1,617          |
| 1999       | 1,777          | 1,479          | 2,346           | 1,889          |
| 2000       | 2,318          | 2,042          | 2,821           | 2,250          |
| 2001       | 2,886          | 2,112          | 2,693           | 1,943          |
| 2002       | 2,099          | 1,900          | —               | —              |

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993, 1996, 1999, 2001).

**TABLA P.16**

| <i>Mes</i> | <i>Ventas</i> | <i>Mes</i> | <i>Ventas</i> |
|------------|---------------|------------|---------------|
| Enero 2001 | 430           | Enero 2002 | 442           |
| Febrero    | 420           | Febrero    | 449           |
| Marzo      | 436           | Marzo      | 458           |
| Abril      | 452           | Abril      | 472           |
| Mayo       | 477           | Mayo       | 463           |
| Junio      | 420           | Junio      | 431           |
| Julio      | 398           | Julio      | 400           |
| Agosto     | 501           | Agosto     | 487           |
| Septiembre | 514           | Septiembre | 503           |
| Octubre    | 532           | Octubre    | 503           |
| Noviembre  | 512           | Noviembre  | 548           |
| Diciembre  | 410           | Diciembre  | 432           |

- e) Utilice Minitab o CB Predictor (vea “Aplicaciones en Excel: CB Predictor” que se encuentra al final de este capítulo); asimismo, utilice el método de suavizamiento multiplicativo de Winters con las constantes de suavizamiento  $\alpha = \beta = \gamma = 0.5$  para generar un pronóstico para enero de 2003. Guarde los residuales.
- f) Remítase al inciso e). Compare el *MAPE* para el método Winters de la versión impresa de la computadora con los *MAPE* en los incisos b) y c). ¿Cuál de los tres procedimientos prefiere usted?
- g) Remítase al inciso e). Calcule las autocorrelaciones (para seis retrasos) para los residuales del procedimiento multiplicativo de Winters. ¿Las autocorrelaciones de los residuales le sugieren que el procedimiento de Winters funciona bien para estos datos? Explique.

## CASOS

---

### CASO 4-1 LA SOLAR ALTERNATIVE COMPANY<sup>4</sup>

La Solar Alternative Company está a punto de comenzar su tercer año de operación. Bob y Mary Johnson, quienes enseñan ciencias en la preparatoria local, fundaron esta empresa con el propósito de complementar sus ingresos académicos. Con base en su investigación acerca de los sistemas de energía, pudieron ensamblar un sistema solar para calentar el agua de las casas. El sistema consiste en un tanque de almacenamiento de fibra de vidrio con capacidad de 100 galones, dos celdas solares de 36 pies cada una, controles electrónicos, un tubo de PVC y varios elementos de ajuste.

El periodo de recuperación de la inversión en el sistema es de 10 años. A pesar de que esta situación no ofrece una oportunidad de inversión atractiva desde el punto de vista financiero, hay suficiente interés en lo novedoso del concepto como para lograr un nivel moderado de ventas. Los Johnson obtienen aproximadamente 75 dólares por cada sistema instalado, que tiene un precio de dos mil dólares, después de considerar costos y gastos. Los costos de material y equipo representan aproximadamente 75% del costo de un sistema instalado. Una ventaja que ayuda a compensar el bajo margen de ganancia es que este producto no es suficientemente rentable como para generar cualquier competencia significativa que ofrezcan los contratistas que se dediquen a la calefacción. Los Johnson operan su negocio desde su casa. Hay una oficina en el sótano y su cochera la utilizan sólo para almacenar los componentes y materiales del sistema. Como resultado, los gastos se encuentran reducidos al mínimo. Los Johnson disfrutan de un ingreso complementario modesto a partir del funcionamiento de la compañía. El negocio también proporciona varias ventajas fiscales.

Bob y Mary están complacidos con el crecimiento de su negocio. A pesar de que las ventas varían de un mes a otro, en general, el segundo año fue mucho mejor que el primero. Muchos de los clientes del segundo año son vecinos de las personas que durante el primer año adquirieron el sistema. En apariencia, después de ver durante un año que el sistema funciona de manera exitosa, otros están dispuestos a probar el concepto de energía solar. Las ventas se dan a lo largo de todo el año. La demanda se incrementa a finales del verano y a principios del otoño, cuando los lugareños preparan sus casas para el invierno, temporada que requiere mayor calefacción.

Con el crecimiento anticipado del negocio, los Johnson pensaron que necesitaban un pronóstico de ventas para administrar eficazmente el año siguiente. Usualmente, se requieren de 60 a 90 días para recibir tanques de almacenamiento después de haber colocado el pedido. Las celdas solares están disponibles la mayor parte del año; sin embargo, a finales del verano y durante el otoño el tiempo de espera puede ser hasta de 90 o 100 días. A pesar de que hay una competencia limitada, la pérdida de ventas es una posibilidad real si los clientes potenciales deben esperar varios meses para que les instalen su sistema. Quizás aún más importante sea la necesidad de realizar proyecciones precisas de las ventas, a fin de aprovechar un descuento en compras por volumen. Estos factores, en combinación con el elevado costo de los componentes del sistema y el reducido espacio de la cochera como almacén, hacen necesario desarrollar un pronóstico confiable. El historial de las ventas para los primeros dos años de la compañía se presenta en la tabla 4.10.■

**TABLA 4.10**

| Mes     | 2001 | 2002 | Mes        | 2001 | 2002 |
|---------|------|------|------------|------|------|
| Enero   | 5    | 17   | Julio      | 23   | 44   |
| Febrero | 6    | 14   | Agosto     | 26   | 41   |
| Marzo   | 10   | 20   | Septiembre | 21   | 33   |
| Abril   | 13   | 23   | Octubre    | 15   | 23   |
| Mayo    | 18   | 30   | Noviembre  | 12   | 26   |
| Junio   | 15   | 38   | Diciembre  | 14   | 17   |

<sup>4</sup> Este caso fue aportado por William P. Darrow, de la Towson State University, Towson, Maryland.

### PREGUNTAS

1. Identifique el modelo que Bob y Mary deberían utilizar como base de sus planes de negocio en 2003 y explique por qué seleccionó este modelo.
2. Pronostique las ventas para 2003.

## CASO 4-2 MR. TUX

John Mosby, propietario de varias tiendas de alquiler Mr. Tux, comienza a pronosticar su variable más importante de negocio: el importe mensual de ventas (vea los casos de Mr. Tux en los capítulos anteriores). Una de sus empleadas, Virginia Perot, recopiló los datos de ventas que se muestran en el caso 2.2. Ahora, John desea pronosticar las ventas a partir de estos datos por medio de las técnicas de promedio móvil y suavizamiento exponencial.

John utilizó Minitab en el caso 3.2 para determinar que estos datos ofrecían una tendencia y eran estacionales. Aunque se le ha dicho que las técnicas simples de promedios móviles y de suavizamiento exponencial no funcionarán con estos datos, decide descubrirlo por sí mismo.

Mosby comienza por intentar un promedio móvil de tres meses. El programa calcula diversas medidas del error del pronóstico. Estos valores resumen los errores que se encuentran al predecir datos históricos reales por medio de un promedio móvil de tres meses. Él registra tres de estas mediciones de error:

$$MAD = 54\,373$$

$$MPE = -16.5\%$$

$$MAPE = 47.0\%$$

La *MAD* (desviación media absoluta) es el promedio de los valores absolutos de los errores que se obtienen en los pronósticos de ajuste para datos pasados. Cada pronóstico que utiliza un método de promedio móvil de tres meses es erróneo en un promedio de 54 373. El *MPE* (error porcentual medio) mide el sesgo; por tanto, el  $-16.5\%$  indica que la técnica de pronóstico sobreestima sistemáticamente los valores. Los pronósticos son mayores que los valores reales. *MAPE* (error porcentual absoluto medio) muestra el error como porcentaje del valor real que habrá de ser pronosticado. El error de promedio obtenido mediante una técnica de promedio móvil de tres meses es de 47%, o casi la mitad del valor que habrá de ser pronosticado.

A continuación, John prueba con una técnica de suavizamiento exponencial simple. El programa le pide que proporcione la constante de suavizamiento ( $\alpha$ ), o bien, que ingrese el valor óptimo que deberá calcularse. John hace esto último y el programa establece que el valor óptimo de  $\alpha$  es 0.867. De nuevo registra las mediciones de error apropiadas:

$$MAD = 46\,562$$

$$MPE = -11.8\%$$

$$MAPE = 44.0\%$$

John ordena al programa que aplique el suavizamiento exponencial de Holt a sus datos. El programa no sólo utiliza el método de suavizamiento exponencial, también puede tomar en cuenta que los datos tienen una tendencia. John aplica una constante de suavizamiento de 0.4, tanto para  $\alpha$  como para  $\beta$ . Las tres mediciones de error de resumen para el método de Holt son:

$$MAD = 63\,579$$

$$MPE = -15.3\%$$

$$MAPE = 59.0\%$$

John se sorprende al encontrar errores mayores de medición con esta técnica y decide que el aspecto estacional de los datos es el problema. El suavizamiento exponencial multiplicativo de Winters es el siguiente método que John pone a prueba. Este método puede tomar en cuenta los factores estacionales, así como la tendencia. John utiliza las constantes de suavizamiento de  $\alpha = 0.2$ ,  $\beta = 0.2$  y  $\gamma = 0.2$ . Las mediciones de error son:

$$MAD = 25\,825$$

$$MPE = -4.9\%$$

$$MAPE = 22.0\%$$

Cuando John se sienta a estudiar los resultados de su análisis, se decepciona, pues aunque el método de Winters ofrece una gran mejoría, el *MAPE* aún es de 22%. Mosby esperaba que uno de los métodos que utilizó produjera pronósticos precisos de los períodos anteriores; entonces podría utilizar este método para

pronosticar los niveles de ventas para los meses siguientes del año entrante. No obstante, los errores absolutos promedio (*MAD*) y los errores porcentuales (*MAPE*)

para los métodos lo llevan a buscar otra manera de pronosticar.■

#### PREGUNTAS

1. Determine el nivel de error del pronóstico para el mejor método que John haya encontrado por medio de Minitab.
2. ¿Cuál es el valor ideal de *MPE* cuando se utiliza cualquier método de suavizamiento exponencial?
3. ¿Cuál es la implicación de un signo negativo en el *MPE*?

4. ¿Está usted de acuerdo con John acerca de sentirse decepcionado por sus resultados?
5. ¿Qué debería hacer John para determinar hasta dónde se puede mejorar el pronóstico mediante la técnica de Winters?

## CASO 4-3 CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se ha descrito en el capítulo 1 (caso 1.2). El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de clientes que habrían en el resto de 1993. Marv proporcionó a Dorothy Mercer los datos mensuales

del número de clientes nuevos que CCC había visto en el periodo de enero de 1985 a marzo de 1993 (vea el caso 3.3 de la página 94). En ese caso, Dorothy empleó el análisis de correlación para explorar el patrón de datos. Utilice los resultados de esta investigación para realizar lo siguiente.■

#### PREGUNTAS

1. Desarrolle un modelo informal para pronosticar el número de nuevos clientes de la CCC para el resto de 1993.
2. Desarrolle un modelo de promedio móvil para pronosticar el número de nuevos clientes de la CCC para el resto de 1993.
3. Desarrolle un procedimiento de suavizamiento exponencial para pronosticar el número de nuevos clientes de la CCC para el resto de 1993.

4. Evalúe los métodos de pronósticos anteriores por medio de los diversos errores de pronóstico que se presentan en el capítulo 3.
5. Seleccione el mejor modelo y pronostique el número de nuevos clientes para el resto de 1993.
6. Determine la idoneidad del modelo de pronóstico que usted ha seleccionado.

## CASO 4-4 MURPHY BROTHERS FURNITURE

Julie Murphy sabe que las decisiones operativas más importantes dependen, hasta cierto grado, de un pronóstico. Para Murphy Brothers Furniture, los pronósticos de ventas ejercen influencia en cuanto a añadir nuevas líneas de muebles o sacar las antiguas, planear las compras, fijar las cuotas de ventas y tomar deci-

siones acerca de personal, publicidad y finanzas. De manera específica, Julie está consciente de varias necesidades de pronóstico actuales. Ella sabe que el departamento de producción tiene que planear la cantidad de empleados y determinar los pedidos de materia prima para uno o dos meses. También sabe que su padre,

Glen Murphy, necesita determinar las mejores decisiones de inversión y debe pronosticar la demanda para una línea futura.

En el caso 3.1-A Julie Murphy utilizó las ventas mensuales nacionales de todas las tiendas al menudeo de 1983 a 1995 (vea la tabla 3.8) para desarrollar un patrón de las ventas de Murphy Brothers Furniture. En el caso 3.1B Glen Murphy descubrió los datos reales

de las ventas correspondientes a los últimos cuatro años, de 1992 a 1995 (vea la tabla 3.9). Julie no estaba entusiasmada con el descubrimiento de su padre porque no estaba segura acerca de cuáles datos emplear para desarrollar un pronóstico para 1996. Luego determinó que las ventas de todas las tiendas minoristas tenían el mismo patrón que los datos reales de Murphy Brothers.■

## PREGUNTAS

1. ¿Alguno de los modelos de pronóstico estudiados en este capítulo funciona con los datos de las ventas nacionales?
2. ¿Alguno de los modelos de pronóstico estudiados en este capítulo funciona con los datos reales de las ventas de Murphy Brothers?
3. ¿Cuáles datos y modelos de pronóstico deberá emplear Julie a fin pronosticar las ventas para 1996?

## CASO 4-5 PROYECCIÓN DE UTILIDADES PARA CINCO AÑOS DE DOWNTOWN RADIOLOGY

Hace algunos años, Downtown Radiology desarrolló un centro de imágenes médicas más completo y avanzado tecnológicamente que cualquiera otro en el área oriental de Washington y la región septentrional de Idaho, conocida como Inland Empire. El equipo planeado para el centro radiológico equiparaba o superaba a las instalaciones de imágenes médicas de todos los centros médicos de la zona. Al principio, el centro contaba con un escáner para tomografías computadorizadas CT de la serie 9800 y un equipo de resonancia magnética nuclear (MRI, del inglés *magnetic resonance imaging*). También tenía aparatos para ultrasonido, medicina nuclear, angiografía de sustracción digital (DSA, del inglés *digital subtraction angiography*), mamografía, radiología convencional y fluoroscopia. La propiedad de este centro se puso a la venta mediante una oferta al público y Downtown Radiology empleó una evaluación independiente del mercado, que fue estimado por Professional Marketing Associates Inc. y completó una proyección de utilidades a cinco años.

### EXPOSICIÓN DEL PROBLEMA

El propósito de este estudio es pronosticar las utilidades de los siguientes cinco años para el centro de imágenes médicas propuesto, suponiendo que usted fuera empleado de Professional Marketing Associates, Inc., durante 1984.

### Objetivos

Los objetivos de este estudio son:

- Identificar las áreas de mercado para cada tipo de procedimiento médico que habrá de ofrecerse en las nuevas instalaciones.
- Recopilar y analizar los datos existentes acerca de las utilidades del área del mercado para cada uno de los servicios que se ofrecerán en las nuevas instalaciones.
- Identificar las tendencias en la industria de la salud que afectarán, positiva o negativamente, las utilidades de los servicios que se ofrecerán en las nuevas instalaciones.
- Identificar los factores en el negocio, el mercadeo y la planeación de las instalaciones del nuevo proyecto que pudieran afectar, positiva o negativamente, las perspectivas de ganancias.
- Analizar los servicios anteriores de Downtown Radiology como una base de datos para el modelo de pronóstico que habrá de desarrollarse.
- Utilizar un modelo adecuado de pronóstico cuantitativo que estime las proyecciones de utilidades para los siguientes cinco años en el centro planeado.

## METODOLOGÍA

### Servicios médicos

Se llevaron a cabo los siguientes pasos para completar la proyección de las utilidades a cinco años. Se hizo un análisis de los últimos servicios médicos. Se desarrolló el modelo adecuado de pronóstico y se utilizó para determinar el punto de partida para la proyección de cada procedimiento.

1. Se estableció el área de mercado para cada tipo de procedimiento y se obtuvieron los pronósticos del tamaño de población para 1986 y 1990.
2. Se estudiaron los patrones de recomendación de los médicos para determinar el porcentaje de los médicos que remitían a sus pacientes a Downtown Radiology y el número promedio de remisiones hechas por cada uno.
3. Las tarifas nacionales se obtuvieron del National Center for Health Statistics. Estas tarifas se compararon con los números reales obtenidos de la Hospital Commission.
4. La participación de mercado de Downtown Radiology se calculó con base en el número real de tomografías CT que hay en el área de mercado. (La participación de mercado para otros servicios se determinó con base en la participación de Downtown Radiology, en comparación con las tarifas proporcionadas por el National Center for Health Statistics.)

### Suposiciones

Para desarrollar un pronóstico cuantitativo fueron necesarias ciertas suposiciones. Se tomaron en cuenta las siguientes:

- El 1 de enero de 1985, el nuevo centro de imágenes médicas será operable con todo el equipo, a excepción del MRI.
- El equipo de resonancia magnética nuclear funcionará a partir de abril de 1985.
- El ofrecimiento participación en la propiedad bajo el esquema de sociedad de responsabilidad limitada se venderá con éxito a por lo menos 50 médicos que estén en el área de servicio.
- Los médicos que tengan interés financiero en el nuevo centro de imágenes médicas incrementarán sus remisiones al nosocomio.
- No habrá otros MRI en el área de mercado antes de 1987.
- El nuevo centro de imágenes médicas ofrecerá otros servicios a precios más bajos que los de la competencia.

- Se realizará un esfuerzo eficaz de marketing, especialmente concentrado en grandes empresarios, compañías de seguros y sindicatos.
- El MRI reemplazará, durante los siguientes seis meses de su puesta en marcha, aproximadamente 60% de las tomografías de la cabeza que actualmente se hacen con el escáner CT y 70% durante los siguientes 12 meses.
- El público continuará presionando a la industria de la salud para que baje los costos.
- El monto de las erogaciones en la industria de la salud se incrementó 13.2% anual desde 1971 hasta 1981. La Health Care Financial Administration calcula que la tasa anual promedio del incremento se verá reducida aproximadamente de 11 a 12% entre 1981 y 1990 (*Industry Surveys*, abril de 1984).
- Las compañías de seguros reembolsarán a los pacientes (en el peor de los casos) de cero a 100% del costo de imágenes por resonancia nuclear magnética (*Imaging News*, febrero de 1984).

### Modelos

Con base en la experiencia, tarifas de la industria y presunciones racionales, se desarrolló un pronóstico para cada uno de los servicios. Debido a que los modelos se desarrollaron sobre la base de las suposiciones anteriores, si éstas no resultan válidas, los modelos serán inexactos.

## ANÁLISIS DE LOS DATOS PRECEDENTES

### Oficina de rayos X

El número de aplicaciones de rayos X realizados se analizó desde julio de 1981 hasta mayo de 1984. Los datos incluyeron rayos X de diagnóstico, rayos X gástrico-intestinales, imágenes médicas del pecho, inyecciones y servicios especiales. El examen de estos datos indica que no hubo patrones de tendencia, ni estacional, ni cíclico. Por esta razón, se eligió un suavizado exponencial simple como método de pronóstico apropiado. Se analizaron diversas constantes de suavización, de las cuales se encontró una de 0.3 que proporciona el mejor modelo. Los resultados se presentan en la figura 4.14. El pronóstico para junio de 1984 es de 855 aplicaciones de rayos X.

### Oficina de ultrasonido

El número de prácticas de ultrasonido se analizó de julio de 1981 a mayo de 1984. La figura 4.15 muestra el patrón de datos. De nuevo, no hay patrones de tendencia, ni estacional, ni cíclico. Una suavización exponencial con una constante de suavizado de  $\alpha = 0.5$  se determinó como el mejor de los modelos. El pronóstico para junio de 1984 es de 127 aplicaciones de ultrasonido.

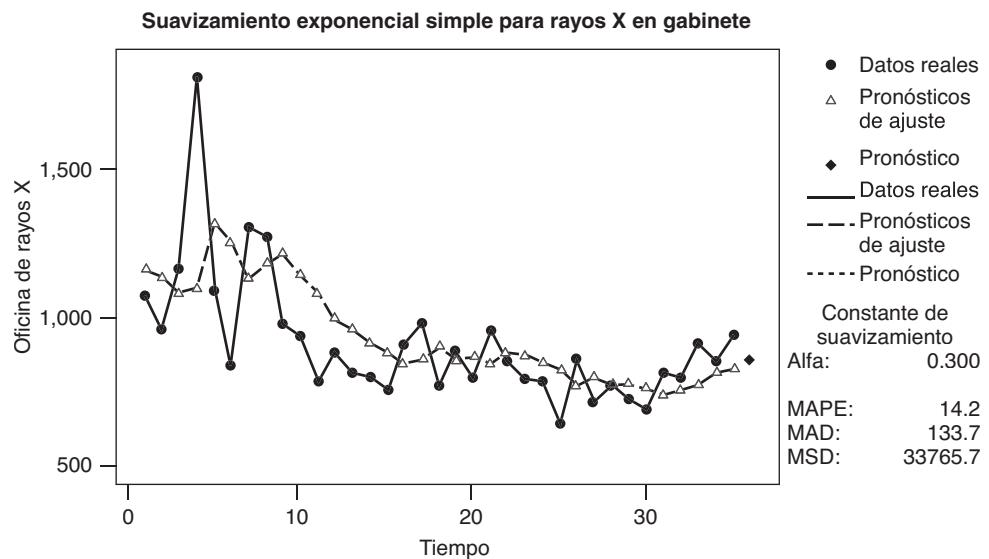


FIGURA 4.14 Suavizado exponencial simple: rayos X de Downtown Radiology

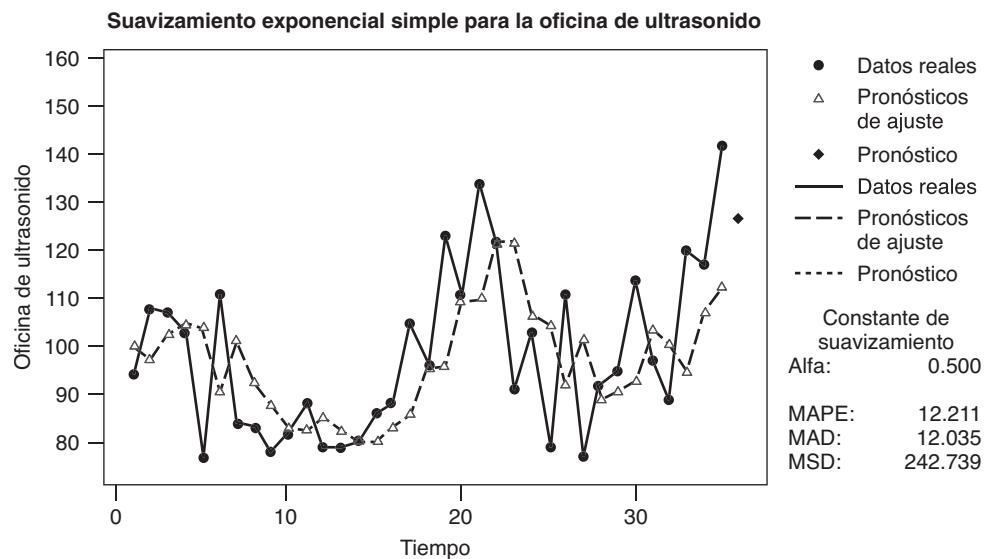
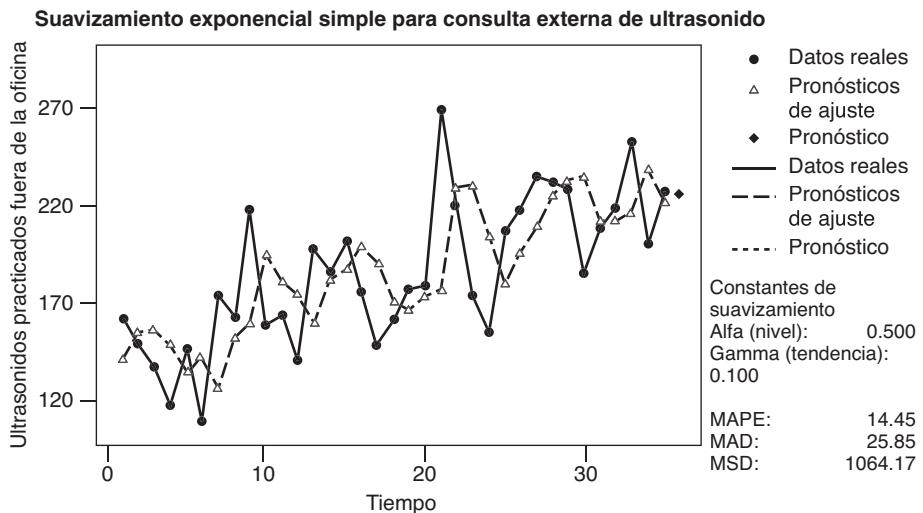


FIGURA 4.15 Suavizado exponencial simple: ultrasonido de Downtown Radiology

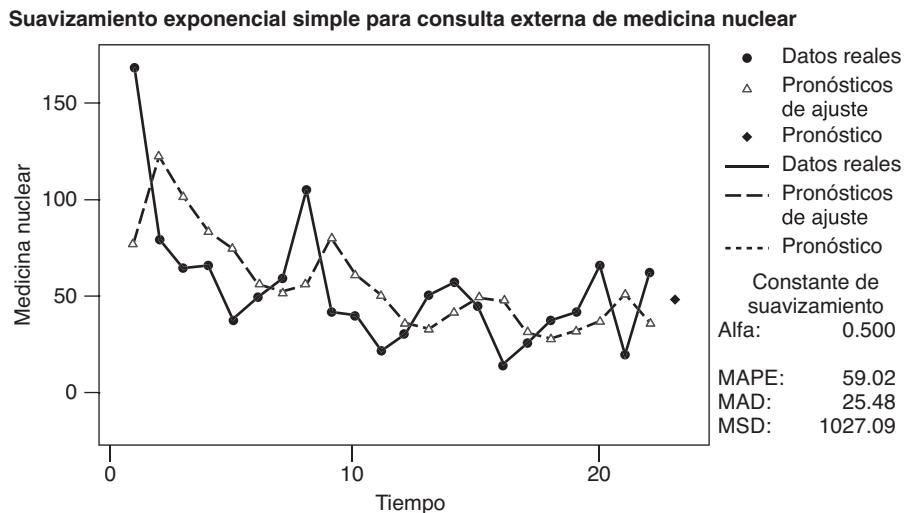
El número de ultrasonidos hechos por las dos unidades móviles que posee Downtown Radiology fue analizado de julio de 1981 a mayo de 1984. La figura 4.16 muestra el patrón de datos. Una tendencia ascendente es notoria y puede modelarse mediante un suavizado exponencial lineal de dos parámetros de Holt. Se usan las constantes de suavizado de  $\alpha = 0.5$  y  $\beta = 0.1$  y el pronóstico para julio de 1984 es de 227.

### Servicios de medicina nuclear

El número de prácticas de medicina nuclear que se llevan a cabo en las dos unidades móviles de Downtown Radiology, se analizó de agosto de 1982 a mayo de 1984. La figura 4.17 muestra el patrón de datos. Éstos no fueron estacionales y no tuvieron patrones de tendencia o cíclicos. Por esta razón, se eligió el suavizado exponencial simple como método de pronóstico apropiado. El factor de suavizado de  $\alpha = 0.5$



**FIGURA 4.16 Suavizado exponencial lineal de Holt: ultrasonido de Downtown Radiology**



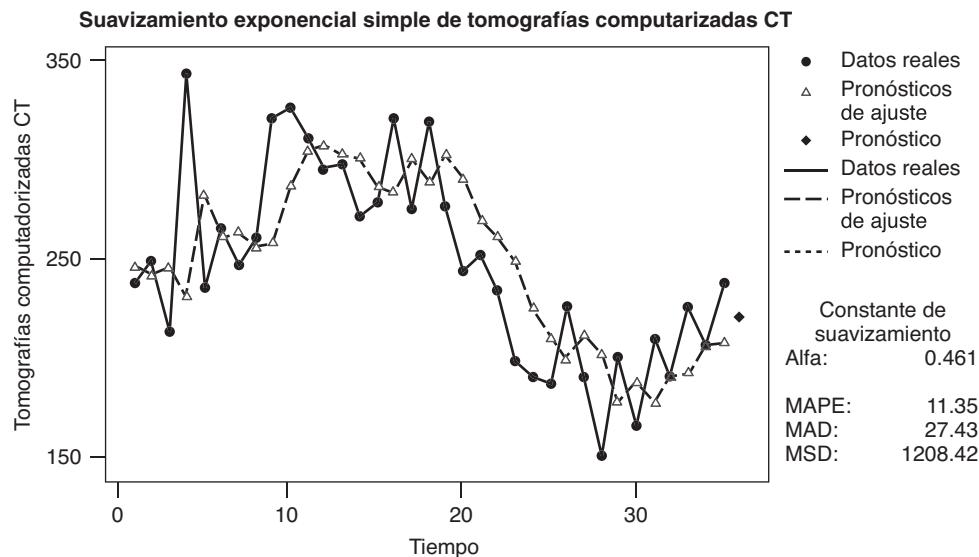
**FIGURA 4.17 Suavizado exponencial simple: medicina nuclear de Downtown Radiology**

proporcionó el mejor modelo. El pronóstico para junio de 1984 es de 48 servicios de medicina nuclear.

#### Oficina de tomografías CT

El número de tomografías CT que se llevaron a cabo fue analizado de julio de 1981 a mayo de 1984. No se encontró estacionalidad y el número de tomografías CT no mostraba tendencia alguna. Sin embargo, al parecer había un patrón cíclico. Sería importante

saber cuántas tomografías se efectuaron el mes pasado para el pronóstico de lo que sucederá este mes. Un modelo autorregresivo (vea los capítulos 8 y 9) fue examinado y comparado con un modelo de suavizado exponencial de  $\alpha = 0.461$ . La constante de suavizado más grande proporciona una mayor ponderación en las observaciones más recientes en el pronóstico. Se determinó que el modelo de suavizado exponencial es mejor que el modelo



**FIGURA 4.18 Suavizado exponencial simple: número de tomografías computarizadas CT de Downtown Radiology**

autorregresivo. La figura 4.18 muestra que la proyección del número de tomografías CT para junio de 1984 es de 221.

## ANÁLISIS DEL ÁREA DE MERCADO

Las áreas de mercado fueron determinadas para los servicios que actualmente se realizan en Downtown Radiology al examinar los registros de los pacientes y los patrones de remisión de los médicos. Se determinaron las áreas de mercado para los servicios que *no* se realizan en Downtown Radiology mediante la investigación de la competencia y el análisis de las áreas geográficas que atienden.

### Área de mercado del escáner para tomografías computarizadas CT

El área de mercado de las tomografías CT para el centro de imágenes médicas propuesto abarca las circunscripciones de Spokane, Whitman, Adams, Lincoln, Stevens y Pend Oreille, en el estado de Washington, y Bonner, Boundary, Kootenai, Benewah y Shoshone en Idaho. Con base en las proyecciones de porcentaje adecuadas, el área de mercado del CT tendrá un tamaño de población de 630 655 individuos en 1985 y de 696 018 en 1990.

### Estimados cuantitativos

Para proyectar las utilidades es necesario determinar ciertos estimados cuantitativos. El más importante tiene que ver con el número de doctores que participarán en la sociedad de responsabilidad limitada. El estimado utilizado en los cálculos futuros indica que participará, por lo menos, 8% de la población médica de la circunscripción de Spokane.

La siguiente incertidumbre que debe cuantificarse implica establecer cómo se afectará el patrón de pacientes referidos con la participación de 50 médicos en la sociedad de responsabilidad limitada. Se supone que 10 de los futuros socios, de un grupo de 30, no incrementarán el número de sus pacientes referidos y los otros 20 lo duplicarán. También se presume que 20 médicos que jamás han hecho traslados a Downtown se integrarán a la sociedad de responsabilidad limitada y comenzarán a enviar, por lo menos, la mitad de sus pacientes a Downtown Radiology.

La cuantificación de los pacientes referidos adicionales deberá ser aclarada mediante ciertas observaciones cualitativas. El cálculo de 50 médicos que se integrarían a la sociedad de responsabilidad limitada es conservador. Hay una gran posibilidad de que se unan médicos de las áreas vecinas de Spokane. De manera tradicional, los cambios en los patrones de referencia médica ocurren muy lentamente. No obstante, la súbita naturaleza competitiva del mercado probablemente tenga repercusiones en ellos. Si la sociedad de responsabilidad limitada se promueve entre especialistas con un elevado potencial de prescripciones radiológicas, el número de éstas podría aumentar más de lo proyectado. La variabilidad en el número de prescripciones para cada procedimiento es extremadamente elevada. Unos cuantos médicos prescribieron un muy alto porcentaje de los servicios realizados por Downtown Radiology. Si se reclutara a algunos médicos nuevos con un elevado número de pacientes referidos, se podría tener un efecto importante en el número total de servicios realizados de entre los servicios individuales proporcionados por Downtown Radiology.

Por último, debe estimarse el efecto que tendría un nuevo centro de imágenes médicas sobre la participación de mercado de Downtown Radiology, ya que contará con el mejor equipo y estará preparado para realizar el rango total de servicios a costos más bajos. El número de nuevos médicos que refieran sus pacientes a este centro deberá incrementarse con base en las recomendaciones verbales que hagan los nuevos médicos inversionistas. Si las compañías de seguros, grandes empresas y sindicatos participan en acuerdos con el nuevo centro de imágenes médicas, Downtown deberá tener la capacidad de incrementar su participación de mercado en por lo menos 4% en 1985; 2% en 1986, y 1% en 1987, además de conservar su participación de mercado en 1988 y 1989. Este incremento en la participación de mercado deberá señalarse como el *efecto de imagen total* en el resto del reporte de proyecciones de las utilidades.

### Proyecciones de las utilidades

Las proyecciones de las utilidades se completaron para cada servicio. En este caso únicamente se muestran las proyecciones para las tomografías computadorizadas CT.

### Proyecciones del número de tomografías computadorizadas CT

Con base en el modelo de suavizamiento exponencial que ya había tenido lugar en los primeros cinco meses de 1984, el pronóstico de este servicio para 1984 (de enero de 1984 a enero de 1985) es de 2 600.

El National Center for Health Statistics reporta una tasa mensual de 261 tomografías CT por cada 100,000 habitantes. Utilizando la población de 630 655 proyectada para el área mercantil del CT, el mercado deberá ser de 19 752 servicios para todo 1985. El número real de tomografías CT efectuadas en el área de mercado durante 1983 se calculó en 21 600. Este estimativo se basó en datos conocidos de servicios reales de Downtown Radiology (2260), Sacred Heart (4970), Deaconess (3850), Valley (2300) y Kootenai (1820), así como los estimados para Radiation Therapy (2400) y Northwest Imaging (4000). Si los cálculos son precisos, Downtown Radiology tenía una participación de mercado cercana a 10.5% en 1983. Los valores reales también se analizaron para 1982, y se proyectó que Downtown Radiology tendría aproximadamente 15.5% del mercado de tomografías CT durante ese año. Por lo tanto, se pronostica que Downtown Radiology tendrá aproximadamente 13% del mercado.

Con base en el incremento de pacientes referidos hecho por los médicos integrados a la sociedad de responsabilidad limitada y en un análisis del promedio de prescripciones de tomografías computadorizadas CT, se proyecta un aumento de 320 servicios de tomografías CT para 1985, a partir de esta fuente. Si se emplean los valores reales para 1983, la cuota del área de mercado de las tomografías CT para Inland Empire es de 3 568

(21 600/6 054) por cada 100 000 habitantes. Si este patrón continúa, el número de tomografías CT en el área de mercado aumentará a 22 514 ( $3568 \times 6.31$ ) en 1985. Por lo tanto, la participación de mercado de Downtown Radiology está proyectada para ser de 13% ( $2\ 920/22\ 514$ ). Cuando se suma el incremento de 4% en la participación de mercado basado en el total de imágenes tomadas, la participación de mercado de Downtown Radiology aumenta a 17.0% y su número proyectado de tomografías CT es de 3827 ( $22\ 514 \times 0.17$ ).

No obstante, la investigación parece indicar que las resonancias magnéticas reemplazarán a un gran número de tomografías CT de cráneo (*Applied Radiology*, mayo-junio de 1983, y *Diagnostic Imaging*, febrero de 1984). El National Center for Health Statistics indicó que 60% de todas las tomografías CT fueron craneales. Los registros de Downtown Radiology mostraron que 59% de sus servicios en 1982 fueron tomografías de la cabeza, y en 1983 equivalieron a 54%. Si 60% de las tomografías CT de Downtown Radiology son de cráneo y las resonancias magnéticas reemplazan aproximadamente 60% de ellas, son necesarias nuevas proyecciones de tomografías CT para 1985. Debido a que la resonancia magnética sólo funcionará la mitad del año, se pronostica una reducción de 689 ( $3827/2 \times 0.60 \times 0.60$ ) en el número de tomografías CT.

El número proyectado de tomografías CT para 1985 es de 3138. El costo promedio de este servicio es de 360 dólares por cada examen y la ganancia esperada es de \$1 129 680. La tabla 4.11 muestra el ingreso proyectado de las tomografías CT para los siguientes cinco años. El costo de los servicios está calculado para aumentar 11% anual, aproximadamente.

Sin considerar el efecto de la resonancia magnética, la proyección para las tomografías CT en 1986 se estima en 4363 ( $6.31 \times 1.02 \times 3568 \times 0.19$ ). No obstante, si 60% corresponde a tomografías CT craneales, y la resonancia magnética reemplaza a 70% de tales exámenes, el número proyectado de estos servicios deberá caer a 2531 [ $4\ 363 - (4\ 363 \times 0.60 \times 0.70)$ ].

La proyección del número de tomografías CT sin el efecto de la MRI para 1987 es de 4683 ( $6.31 \times 1.04 \times 3568 \times 0.20$ ). El pronóstico con el efecto de MRI es de 2716 [ $4683 - (4\ 683 \times 0.60 \times 0.70)$ ].

**TABLA 4.11 Utilidades proyectadas a cinco años para las tomografías CT**

| Año  | Número de servicios | Ingreso     |
|------|---------------------|-------------|
| 1985 | 3,138               | \$1,129,680 |
| 1986 | 2,531               | 1,012,400   |
| 1987 | 2,716               | 1,205,904   |
| 1988 | 2,482               | 1,223,626   |
| 1989 | 2,529               | 1,383,363   |

La proyección del número de tomografías CT sin el efecto de la MRI para 1988 es de  $4773 (6.31 \times 1.06 \times 3568 \times 0.20)$ . El pronóstico con el efecto MRI es de  $2482 [4773 - (4773 \times 0.60 \times 0.80)]$ .

La proyección del número de tomografías CT sin el efecto de la MRI para 1989 es de  $4863 (6.31 \times 1.08 \times 3568 \times 0.20)$ . El pronóstico con el efecto MRI es de  $2529 [4863 - (4863 \times 0.60 \times 0.80)]$ . ■

### PREGUNTA

- El contador de Downtown Radiology anunció que los ingresos serían considerablemente más altos. Dado que el interés económico por la propiedad se pondrá a disposición por medio de cierto tipo de oferta pública, la administración de Downtown

Radiology debe tomar una decisión concerniente a la puntualidad de las proyecciones elaboradas por Professional Marketing Associates. Se le solicita que analice el informe. ¿Qué recomendaciones haría usted?

### Aplicaciones en Minitab

**El problema.** En el ejemplo 4.3, los datos de Spokane Transit Authority necesitan pronosticarse a partir de un promedio móvil de cinco semanas.

#### Solución con Minitab

- Introduzca los datos de Spokane Transit Authority que se muestran en la tabla 4.2 (vea la página 106) en la columna C1 o abra la hoja de cálculo Tab4-2. Dé clic en los siguientes menús:

Stat>Time Series>Moving Average

- Aparece el cuadro de diálogo de *Moving Average* (promedio móvil).
  - Dé doble clic en la variable *Gallons* (galones) y aparecerá a la derecha de *Variable*.
  - Ya que queremos un promedio de movimiento a cinco meses, indique 5 en la longitud de *MA*.
  - No seleccione el cuadro de *Center moving average* (promedio móvil centrado), el cual se utilizará para suavizar los datos en el capítulo 5.
  - Dé clic en *Generate forecasts* (generar pronósticos) e indique 1 a la derecha de *Number of forecasts* (número de pronósticos).
  - Dé clic en *OK* y aparecerá la figura 4.4.

**El problema.** En el ejemplo 4.6, los datos de la Acme Tool Company necesitan pronosticarse mediante un suavizamiento exponencial simple.

#### Solución con Minitab

- Introduzca los datos de Acme Tool Company que se muestran en la tabla 4.1 (vea la página 103) para el periodo de 1996 a 2001 en la columna C1. Dé clic en los siguientes menús:

Stat>Time Series>Single Exponential Smoothing

- Aparece el cuadro de diálogo *Simple Exponential Smoothing* (suavizamiento exponencial simple).
  - Dé doble clic en la variable *Saws* (serruchos) y aparecerá a la derecha de *Variable*.
  - Bajo *Weights to Use in Smoothing* (pesos para usar en el suavizamiento) elija *Optimal ARIMA* (ARIMA óptimo), después dé clic en *OK*. El resultado se muestra en la figura 4.8 (vea la página 119).

**El problema.** En el ejemplo 4.10, los datos de la Acme Tool Company necesitan pronosticarse mediante un suavizamiento exponencial ajustado a la tendencia y a la estacionalidad.

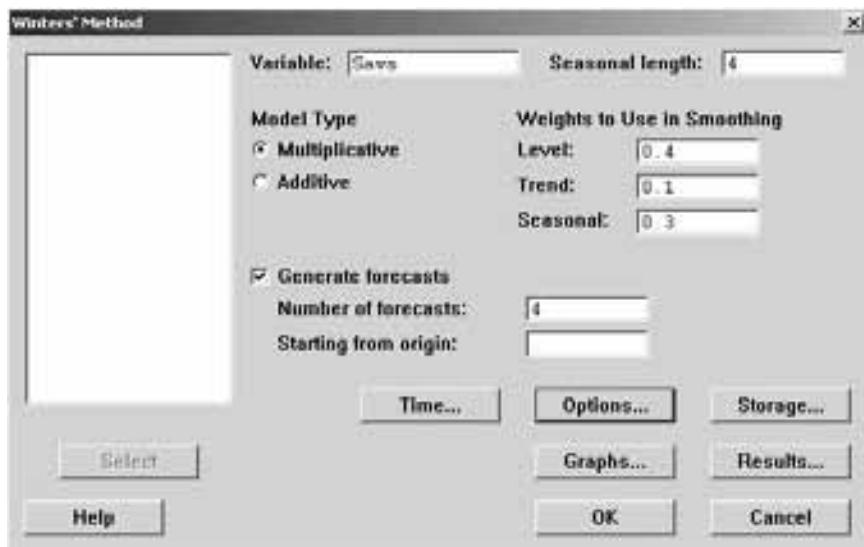


FIGURA 4.19 Cuadro de diálogo del método de Winters en Minitab

### Solución en Minitab

1. Introduzca los datos de Acme Tool Company que se muestran en la tabla 4.1 (vea la página 103) para los años de 1996 a 2001 en la columna C1. Dé clic en los siguientes menús:

Stat>Time Series>Winters' Method

2. Aparece el cuadro de diálogo de *Winters Method* (método de Winters) tal y como se muestra en la figura 4.19.
  - a) La variable de interés es *Saws*.
  - b) Debido a que los datos son trimestrales, indique 4 en *Seasonal lenght* (longitud estacional).
  - c) Los *Weights to Use in Smoothing* son *Level* (nivel): 0.4, *Trend* (tendencia): 0.1 y *Seasonal* (estacional): 0.3.
  - d) Dé clic en *Generate forecasts* (generar pronósticos) y para *Number of forecasts* (número de pronósticos) indique 4.
  - e) Dé clic en *Storage* (almacenamiento).
3. Aparece el cuadro de diálogo de *Winters method*.
  - a) Dé clic en *Level estimates*, *Trend estimates*, *Seasonal estimates*, *Fits* (pronóstico con un periodo de anticipación) y en *Residuals*.
  - b) Dé clic en OK tanto en el cuadro de diálogo de *Winters Method Storage* como en el de *Winters Method*. Los resultados se muestran en la tabla 4.9 y en la figura 4.12. El pronóstico para el primer trimestre de 2002 es de 778.2.
4. A fin de almacenar los datos para usos futuros dé clic en los siguientes menús:

File>Save Worksheet As

5. Aparece el cuadro de diálogo de *Save Worksheet* (guardar hoja de cálculo).
  - a) Escriba un nombre en el espacio de *File Name* (nombre de archivo), tal como *Saws* (serruchos).
  - b) El espacio *Save as Type* (guardar como archivo tipo) le permite seleccionar cómo desea guardar su archivo. La mayor parte de las ocasiones seleccionará Minitab. Sin

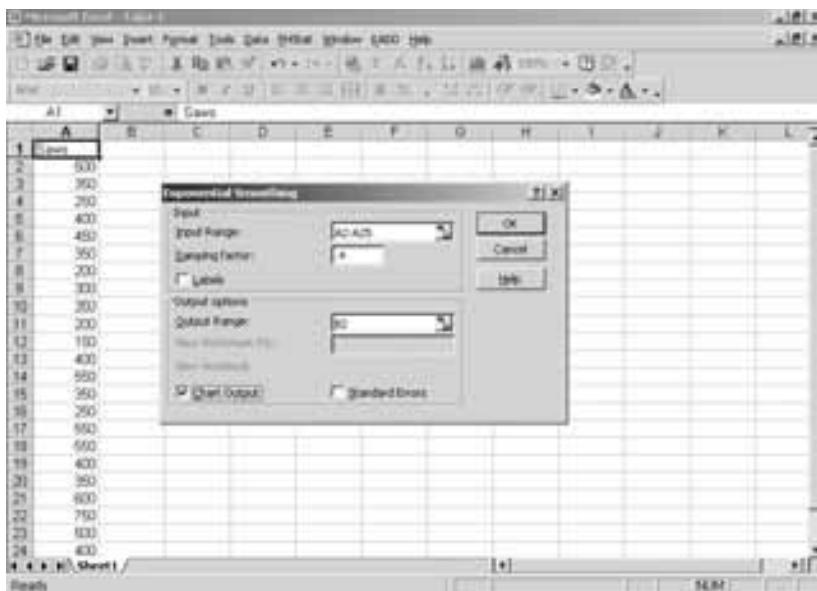


FIGURA 4.20 Cuadro de diálogo del suavizamiento exponencial en Excel

embargo, es posible guardar su archivo de manera que pueda ser utilizado por diversos programas. Por ejemplo, podría elegir guardarlo como un archivo de Excel. Éste se guarda como *Saws.xls* y se utilizará en la siguiente sección de Aplicaciones en Excel.

## Aplicaciones en Excel

**El problema.** En el ejemplo 4.5 los datos de la Acme Tool Company se pronosticaron mediante una suavización exponencial simple con una constante de suavizamiento igual a 0.6.

### Solución en Excel

1. Abra el archivo que contiene los datos presentados en la tabla 4.1 (vea la página 103) al dar clic en los siguientes menús:

File>Open

Busque el archivo llamado *Saws.xls*

2. Dé clic en los siguientes menús:

Tools>Data Analysis

Aparece el cuadro de diálogo de *Data Analysis* (análisis de datos). Bajo *Analysis Tools* (herramientas de análisis) elija *Exponential Smoothing* (suavización exponencial) y dé clic en OK. El cuadro de diálogo se muestra en la figura 4.20.

3. a) Introduzca A2:A25 en el cuadro de edición de *Input Range* (rango de entrada) de datos.  
b) Introduzca 0.4 en el cuadro de edición del *Damping Factor* (factor de suavización) ya que éste se define como complemento de la constante de suavizamiento.  
c) Seleccione el cuadro de verificación de *Labels* (rótulos).  
d) Introduzca B2 en la caja de edición de *Output Range* (rango de salida).  
e) Verifique el cuadro de *Chart Output* (crear gráfico).  
f) Ahora dé clic en OK.
4. Los resultados (columna B) y la gráfica se muestran en la figura 4.21. Observe que la herramienta de análisis del *Exponential Smoothing* coloca fórmulas en la hoja de cálculo. La celda B4 está resaltada y la fórmula = 0.6 \* A3 + 0.4 \* B3 se muestra en la barra de fórmulas.

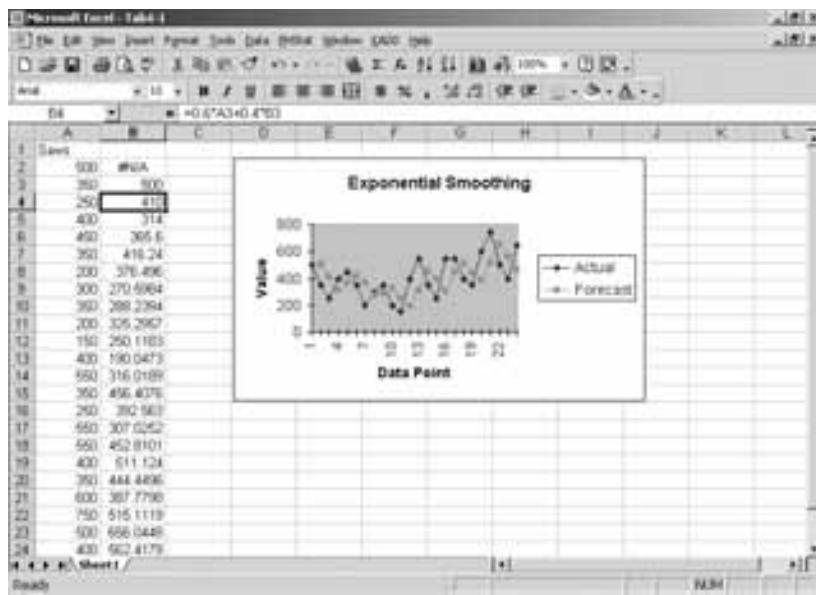


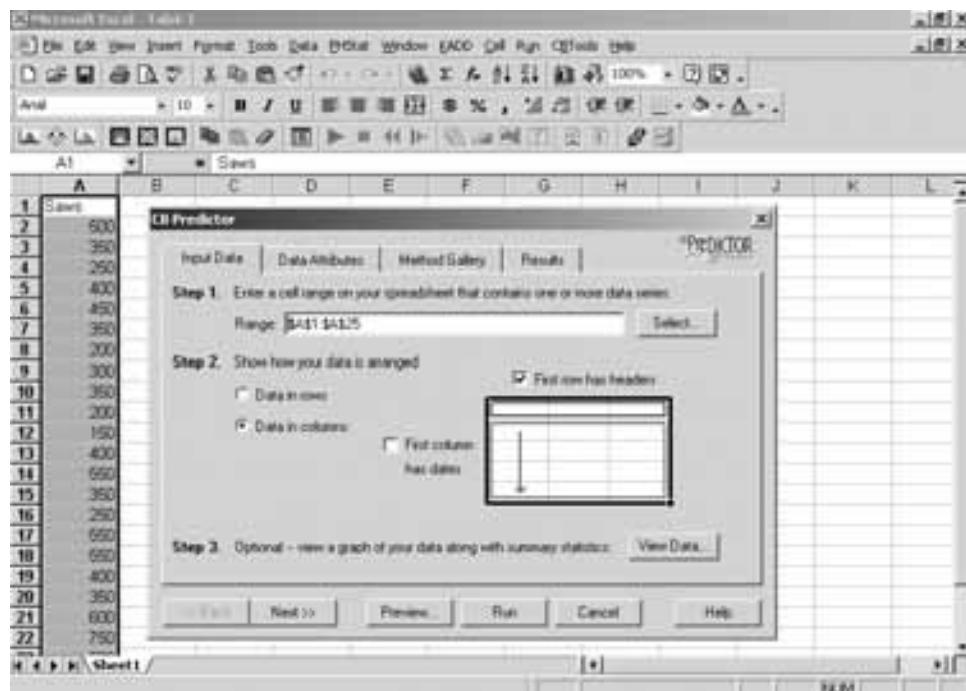
FIGURA 4.21 Resultados del ejemplo 4.5 del suavizado exponencial en Excel

## Aplicaciones en Excel: CB Predictor

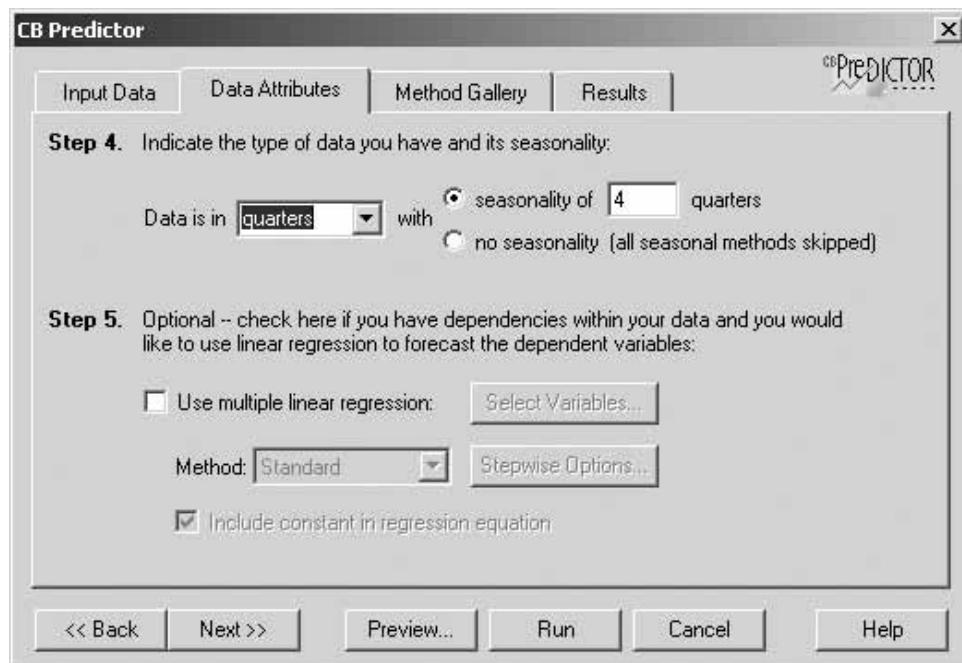
**El problema.** En el ejemplo 4.10 los datos de la Acme Tool Company necesitan pronosticarse mediante un suavizado exponencial ajustado a la tendencia y a la estacionalidad.

### Solución en CB Predictor

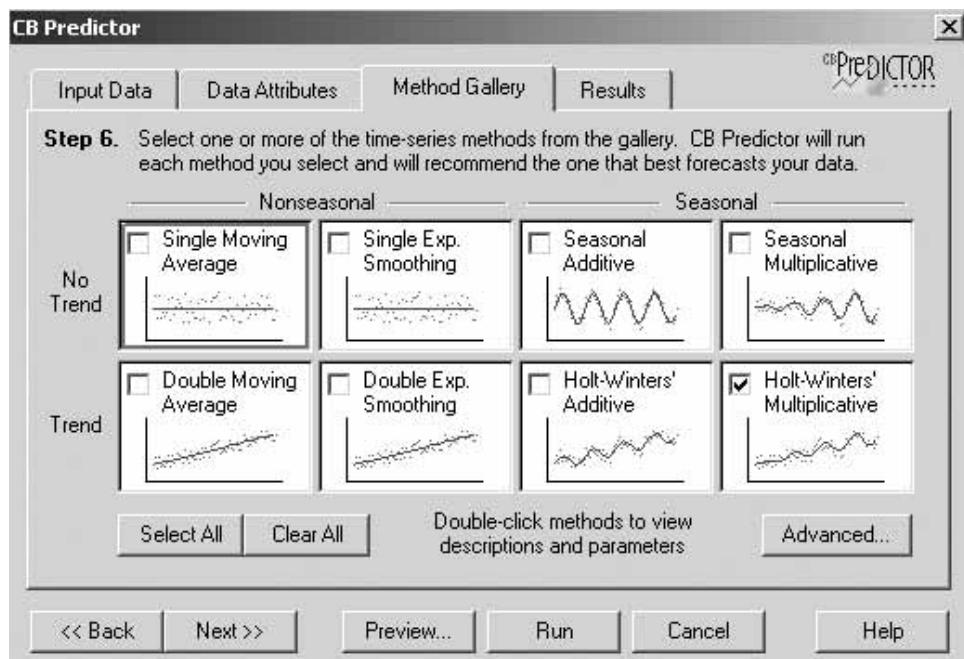
1. Abra el complemento de Excel, Crystal Ball.
2. Abra el archivo que contiene los datos que se proporcionan en la tabla 4.1 al dar clic en los menús:  
File>Open (Busque el archivo *Saws.xls*)
3. Dé clic en los siguientes menús y aparecerá el cuadro de diálogo que se observa en la figura 4.22:  
CBTools > CB Predictor
4. a) En el **Paso 1** introduzca A1:A5 en el cuadro de edición de *Range* (rango).  
b) En el **Paso 2** seleccione los encabezados de la primera línea en el recuadro de verificación  
c) En el **Paso 3** (opcional) dé clic en *View Data* (ver datos) para que se muestre una gráfica de serie de tiempo de las ventas de los serruchos.  
d) Seleccione *Next* (siguiente) o el tabulador de *Data Attributes* (atributos de datos) y aparecerá el cuadro de diálogo que se muestra en la figura 4.23.
5. a) En el **Paso 4** introduzca los trimestres en el cuadro de edición de *Data* (datos). Dé clic en *Seasonality* (estacionalidad) e introduzca 4 en el cuadro de edición.  
b) Omita el **Paso 5**.  
c) Seleccione *Next* o el tabulador de *Method Gallery* (galería de métodos) y aparecerá el cuadro de diálogo que se muestra en la figura 4.24.
6. a) En el **Paso 6** dé clic en *Holt-Winters Multiplicative* (multiplicativo de Holt-Winters).  
b) Coloque el cursor sobre las palabras “*Holt-Winters Multiplicative*” y dé clic. Aparece la ventana que se muestra en la figura 4.25 (vea la página 153).  
c) Seleccione *User defined* (definido por el usuario) e introduzca 0.4 en la caja Alfa, 0.1 en la caja Beta y 0.3 en la caja Gamma. Dé clic en OK.



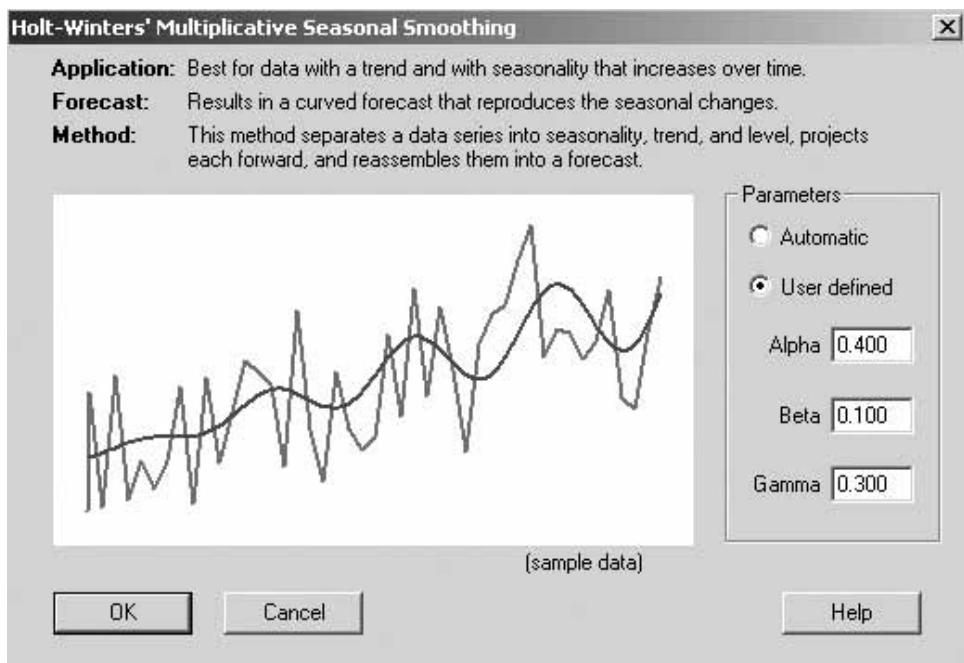
**FIGURA 4.22 Cuadro de diálogo de *Input Data* de CB Predictor: datos de la Acme Tool Company**



**FIGURA 4.23 Cuadro de diálogo de *Data Attributes* en CB Predictor: datos de la Acme Tool Company**



**FIGURA 4.24 Cuadro de diálogo de *Method Gallery*: datos de la Acme Tool Company**



**FIGURA 4.25 Ventana de CB Predictor del suavizamiento multiplicativo de Holt-Winters**

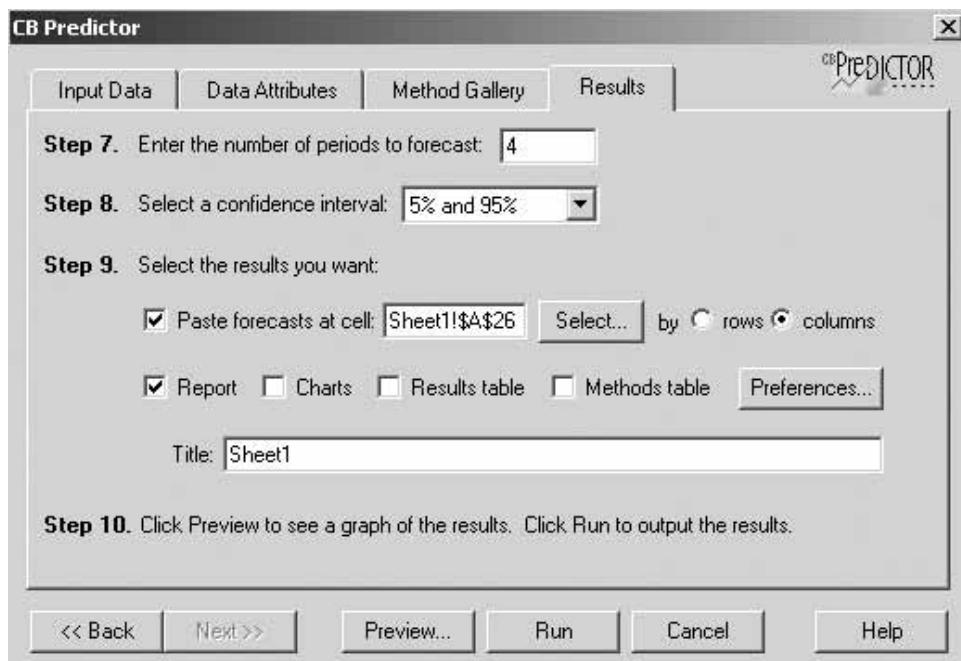


FIGURA 4.26 Cuadro de diálogo de resultados de CB Predictor: datos de la Acme Tool Company

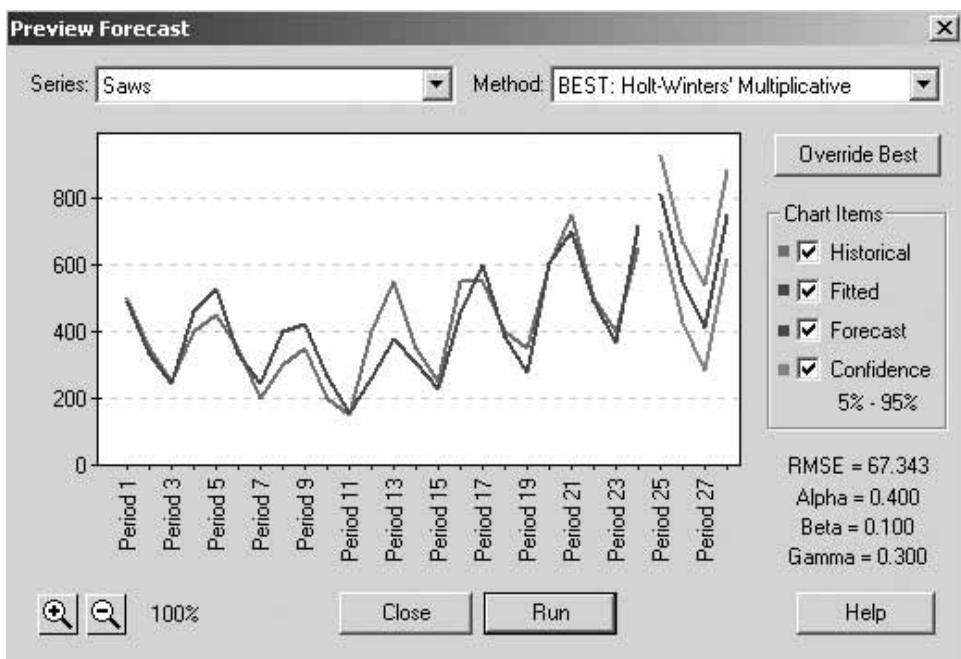


FIGURA 4.27 Pantalla de una vista previa del pronóstico de CB Predictor

- d) Seleccione *Next* la pestaña de *Results* (resultados) y aparecerá el cuadro de diálogo que se observa en la figura 4.26 (vea la página 154).
7. a) En el **Paso 7** introduzca 4 en el cuadro de edición de *Edit the number of periods to forecast* (edición del número de periodos por pronosticar).
- b) En el **Paso 8** utilice los valores establecidos de 5 y 95% en el cuadro de edición de *Select a confidence interval* (elegir un intervalo de confianza).
- c) En el **Paso 9** verifique la celda de *Paste forecasts* (pegar pronósticos) y utilice lo que ahí aparece predeterminado o introduzca, por decir, A26 para formar pronósticos a los valores de los datos. Seleccione columnas. Verifique *Report* (reporte) para obtener un resumen completo de los resultados, incluyendo una gráfica de datos y pronóstico. Si se desea, escriba un encabezado en el cuadro de edición de *Title* (título).
- d) En el **Paso 10** (opcional) dé clic en *Preview* (presentación preliminar) para ver una gráfica inicial de los datos, valores suavizados y pronósticos. Vea la figura 4.27. (Nota:  $RMSE = \sqrt{MSE}$ .)
- e) Dé clic en *Run* (ejecutar) y los resultados del modelo aparecerán en una hoja de reporte por separado. Los pronósticos para los siguientes cuatro trimestres aparecerán en las celdas de A26 a A29 en la hoja de datos.

## Referencias

- Aaker, D. A., y R. Jacobson, "The Sophistication of 'Naive' Modeling", *International Journal of Forecasting* 3, núm. 314 (1987): 449-452.
- Dalrymple, D. J., y B. E. King, "Selecting Parameters for Short-Term Forecasting Techniques", *Decision Sciences* 12 (1981): 661-669.
- Gardner, E. S. Jr., y D. G. Dannenbring, "Forecasting with Exponential Smoothing: Some Guidelines for Model Selection", *Decision Sciences* 11 (1980): 370-383.
- Holt, C. C., "Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages", Office of Naval Research, memorándum núm. 52, 1957.
- Holt, C. C., F. Modigliani, J. F. Muth y H. A. Simon, *Planning Production Inventories and Work Force*, Upper Saddle River, NJ, Prentice Hall, 1960.
- Ledolter, J., y B. Abraham, "Some Comments on the Initialization of Exponential Smoothing", *Journal of Forecasting* 3, núm. 1, (1984): 79-84.
- Makridakis, S., S. C. Wheelwright y R. Hyndman, *Forecasting Methods and Applications*, Nueva York, John Wiley & Sons, 1998.
- McKenzie, E., "An Analysis of General Exponential Smoothing", *Operations Research* 24, (1976): 131-140.
- Newbold, P., y T. Bos, *Introductory Business and Economic Forecasting*, segunda edición, Cincinnati, OH, South-Western, 1994.
- Winters, P. R., "Forecasting Sales by Exponentially Weighted Moving Averages", *Management Science* 6, (1960): 324-342.



## SERIES DE TIEMPO Y SUS COMPONENTES

Como se ha indicado en capítulos anteriores, las observaciones de una variable  $Y$  que se hacen disponibles a través del tiempo se conocen como *datos de una serie de tiempo* o, de manera más sencilla, una *serie de tiempo*. Con frecuencia, estas observaciones se registran en intervalos fijos. De esta manera, por ejemplo,  $Y$  puede representar las ventas y la serie asociada de tiempo podría ser una secuencia de las cifras de ventas anuales. Otros ejemplos de series de tiempo incluyen ganancias trimestrales, niveles de inventarios mensuales y tipos de cambio. En general, las series de tiempo no se comportan como una muestra aleatoria y requieren de métodos especiales para su análisis. De manera ordinaria, las observaciones de una serie de tiempo se relacionan entre ellas (autocorrelacionadas). Esta dependencia genera patrones de variabilidad que pueden utilizarse para pronosticar los valores futuros y asistir en la administración de las operaciones del negocio. Considere estas situaciones.

American Airlines (AA) compara las reservaciones actuales con los pronósticos a partir de patrones históricos. Según sea el caso, si las reservaciones actuales están retrasadas o exceden las proyecciones, AA ajusta de manera adecuada la proporción de los asientos que se venden con descuento. Las adaptaciones se hacen para cada segmento de vuelo en el sistema de AA.

Un importador canadiense de flores adquiere sus productos de diversos floricultores de Estados Unidos, México, Centroamérica y Sudamérica. Sin embargo, debido a que estos productores compran su material y químicos en Estados Unidos, todos los precios de venta se tasan en dólares estadounidenses al momento de la transacción. Una factura no se paga de forma inmediata y debido a las fluctuaciones entre las monedas canadiense y estadounidense, al momento de la compra el importador desconoce el costo en dólares canadienses. Si el tipo de cambio no varía antes de que se pague la factura, no hay riesgo de tipo de cambio para el importador. Si el índice se eleva, el importador pierde dinero en cada compra que realice en dólares estadounidenses. Si el índice cae, el importador gana. Éste emplea los pronósticos semanales del tipo de cambio entre ambas monedas para administrar el inventario de las flores.

A pesar de que con frecuencia las series de tiempo se generan de manera interna y son únicas para la organización, muchas series de tiempo de interés en los negocios pueden obtenerse de fuentes externas. Las publicaciones como *Statistical Abstract of the United States*, *Survey of Current Business*, *Monthly Labor Review* y *Federal Reserve Bulletin* contienen series de tiempo de toda índole. La literatura de este género proporciona datos de series de tiempo sobre precios, producción, ventas, empleo, desempleo, horas trabajadas, consumo de combustible, generación de energía, ganancias, etc., que se reportan cada mes, trimestre o año. Actualmente, existen amplias colecciones de series de tiempo que están disponibles en sitios de Internet que mantienen agencias del gobierno estadounidense, organizaciones estadísticas, universidades y particulares.

Es importante que los administradores comprendan el pasado y empleen su buen juicio y los datos históricos a fin de realizar una planeación inteligente que satisfaga las demandas del futuro. Los pronósticos de series de tiempo que se construyen adecuadamente ayudan a disminuir la incertidumbre asociada con el futuro y pueden ayudar a la administración a determinar estrategias optionales. Desde luego, la planeación anticipada no lo es todo. No obstante, en un ambiente empresarial dinámico, la falta de planeación puede ser

desastrosa. Por ejemplo, un fabricante de grandes computadoras que hace algunos años ignoraba la tendencia hacia las computadoras personales y las estaciones de trabajo, hubiera perdido gran parte de su participación de mercado rápidamente.

A pesar de que se enfocará la atención en un método basado en los modelos de análisis de las series de tiempo, que dependa principalmente de los datos, una revisión subjetiva del esfuerzo para pronosticar resulta muy importante. Cada vez que se examina el pasado para tener pistas acerca del futuro, es relevante sólo en la medida que las condiciones causales continúen vigentes en el siguiente periodo. La multitud de factores causales vigentes tienden a cambiar continuamente, de manera que la conexión entre el pasado, el presente y el futuro debe evaluarse de forma reiterada.

Las técnicas de series de tiempo proporcionan un método conceptual de pronóstico que ha probado ser bastante útil. Los pronósticos se realizan con la ayuda de una serie de procedimientos específicos y formales, y los juicios que le siguen se indican explícitamente.

## DESCOMPOSICIÓN

---

Un método para el análisis de los datos de series de tiempo incluye un intento por identificar los factores que influyen en cada valor de la serie. Este procedimiento de análisis se llama *descomposición*. Cada componente se estudia por separado. Las proyecciones de cada componente se pueden combinar para producir pronósticos de valores futuros de una serie de tiempo. Los métodos de descomposición se utilizan para los pronósticos de corto y largo plazos. También se emplean para mostrar de manera simple el crecimiento subyacente o la declinación de la serie, así como para ajustar las series al eliminar uno o más de los componentes.

El análisis de una serie de tiempo mediante la descomposición en sus partes tiene una larga historia. Sin embargo, últimamente los métodos de descomposición del pronóstico han perdido algo de su brillo. Con frecuencia, la proyección de los componentes individuales hacia el futuro y la recombinación de esas proyecciones para formar un pronóstico de las series subyacentes, no funcionan bien en la práctica. La dificultad radica en obtener pronósticos precisos de los componentes. El desarrollo de procedimientos más flexibles basados en modelos de pronóstico (algunos se analizan en capítulos posteriores) ha hecho de la descomposición una herramienta para comprender una serie de tiempo, en lugar de utilizarla como un método de pronóstico por derecho propio.

Para comprender la descomposición, se comenzará con los cuatro componentes de una serie de tiempo que se presentaron en el capítulo 3. Éstos son el de tendencia, el cíclico, el estacional y el irregular o aleatorio.

1. *Tendencia.* La tendencia es el componente que representa el crecimiento subyacente (o la declinación) en una serie de tiempo. Por ejemplo, la tendencia puede ser generada por un cambio constante en la población o por la inflación, las innovaciones tecnológicas y los incrementos de productividad. La tendencia se denota con  $T$ .
2. *Cíclico.* El componente cíclico es una serie de fluctuaciones (como un oleaje) o ciclos de más de un año de duración. Por lo general, las condiciones económicas cambiantes producen ciclos. La  $C$  denota al componente cíclico.

En la práctica resulta difícil identificar los ciclos y con frecuencia se le considera parte de la tendencia. En este caso, el componente del crecimiento general subyacente (o declinación) es conocido como *ciclo-tendencia* y se denota por la  $T$ . Se usa la notación para la tendencia porque frecuentemente el componente cíclico no puede ser separado de ésta.

3. *Estacional.* De manera típica, las fluctuaciones estacionales se encuentran en datos trimestrales, mensuales o semanales. La variación estacional se refiere a un patrón de cambio más o menos estable que aparece anualmente y se repite con esta periodicidad.

Los patrones estacionales se deben a la influencia del clima o a sucesos relativos al calendario, como las vacaciones escolares o los días feriados. La  $S$  denota al componente estacional.

4. *Irregular.* El componente irregular consiste en las fluctuaciones impredecibles o aleatorias. Dichas fluctuaciones son el resultado de muchos acontecimientos que de manera individual podrían carecer de importancia, pero cuyo efecto combinado podría ser grande. La  $I$  denota al componente irregular.

Para estudiar los componentes de una serie de tiempo, el analista debe considerar la manera en que se relacionan los componentes con las series originales. La tarea se lleva a cabo al especificar un *modelo* (relación matemática) que exprese la variable  $Y$  de la serie de tiempo en términos de los componentes  $T, C, S$  e  $I$ . Un modelo que trata a los valores de las series de tiempo como la suma de sus componentes se conoce como *modelo de componentes aditivos*. Un modelo que trata a los valores de la serie de tiempo como el producto de sus componentes se conoce como *modelo de componentes multiplicativos*. Algunas veces, a estos modelos se les conoce como *modelos de componentes no observados* porque en la práctica, aunque se observen los valores de las series de tiempo, no se observan los valores de los componentes. El método para el análisis de series de tiempo descrito en este capítulo se basa en un intento por estimar los valores de los componentes de las series. Estos estimados pueden utilizarse para pronosticar o exhibir las series que no se ven alteradas por las fluctuaciones estacionales. Este último proceso se conoce como *ajuste estacional*.

Es difícil manejar los componentes cílicos de las series de tiempo. El grado en que pueden determinarse los ciclos a partir de datos históricos, tanto sus longitudes (medidas en años) como sus magnitudes (diferencias entre altos y bajos), están lejos de ser constantes. Esta carencia de un patrón de oleaje consistente hace difícil que los ciclos se distingan de las tendencias que evolucionan con suavidad. En consecuencia, para mantener las cosas relativamente simples, se supondrá que cualquier ciclo en los datos es parte de la tendencia. Al principio, sólo se consideran los tres componentes  $T, S$  e  $I$ . Una breve presentación de cómo se manejan las fluctuaciones cílicas en el método de descomposición para el análisis de series de tiempo, se puede encontrar en la sección “Variaciones cílicas e irregulares” de este capítulo (vea la página 172).

Los dos modelos más simples que relacionan los valores observados ( $Y_t$ ) de una serie de tiempo con los componentes de tendencia ( $T_t$ ), estacional ( $S_t$ ) e irregular ( $I_t$ ) son el modelo de componentes aditivos

$$Y_t = T_t + S_t + I_t \quad (5.1)$$

y el modelo de componentes multiplicativos

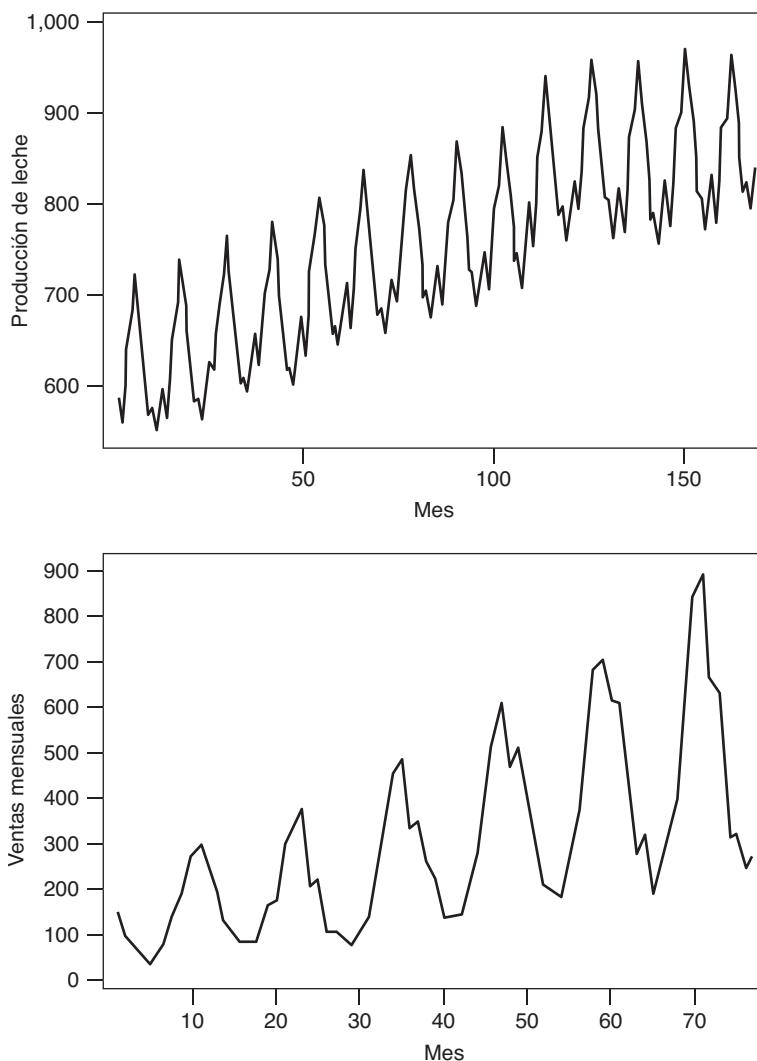
$$Y_t = T_t \times S_t \times I_t \quad (5.2)$$

El modelo de componentes aditivos funciona mejor cuando la serie de tiempo que se analiza tiene aproximadamente la misma variabilidad a lo largo de la serie. Es decir, todos los valores de la serie caen esencialmente dentro de una banda de anchura constante que se centra sobre la tendencia.

El modelo de componentes multiplicativos funciona mejor cuando la variabilidad de la serie de tiempo se incrementa al aumentar el nivel.<sup>1</sup> Es decir, los valores de la serie se extienden en la medida que se incrementa la tendencia, y el conjunto de observaciones tiene la apariencia de un megáfono o embudo. En la figura 5.1 se muestran una serie de tiempo con una variabilidad constante y una serie de tiempo con una variabilidad que crece al

---

<sup>1</sup> Es posible convertir una descomposición multiplicativa en una descomposición aditiva con los logaritmos de los datos. Al usar la ecuación 5.2 y las propiedades de los logaritmos se tiene que  $Y = \log(T \times S \times I) = \log T + \log S + \log I$ . La descomposición de los datos transformados con logaritmos se explora en el problema 5.15.



**FIGURA 5.1** Serie de tiempo con variabilidad constante (arriba)  
y una serie de tiempo con variabilidad creciente  
con nivel (abajo)

aumentar el nivel. Ambas series mensuales tienen una tendencia creciente y se definen con claridad en el patrón estacional.<sup>2</sup>

## TENDENCIA

Las tendencias son movimientos de largo plazo en una serie de tiempo que en ocasiones se pueden describir por una línea recta o una curva suavizada. Algunos ejemplos de las fuerzas

<sup>2</sup> Existen variantes de los modelos de descomposición (vea las ecuaciones 5.1 y 5.2) que contienen tanto términos aditivos como multiplicativos. Por ejemplo, algunos software llevan a cabo la descomposición “multiplicativa” mediante el modelo  $Y = T \times S + I$ .

básicas que producen o afectan la tendencia de una serie son los cambios en la población, en el precio y en la tecnología, además de los incrementos en la productividad y los ciclos de vida del producto.

Un incremento de la población podría causar que las ventas al menudeo se incrementaran cada año por un periodo prolongado. Asimismo, las ventas actuales en dólares pudieron haberse ido al alza durante el mismo periodo debido a los incrementos generales en los precios en los bienes de consumo, incluso cuando no hubiera cambiado el volumen general de ventas de los bienes de consumo.

El cambio tecnológico puede causar que una serie de tiempo se vaya hacia abajo o hacia arriba. El desarrollo de chips electrónicos de alta velocidad para computadoras, aparatos de memoria optimizados y pantallas mejoradas, además de los adelantos en la tecnología de las telecomunicaciones, ha resultado en un espectacular aumento en el uso de computadoras personales y teléfonos móviles. Desde luego, estos avances han llevado a una tendencia descendente en la producción de calculadoras mecánicas y teléfonos de disco.

Los incrementos en la productividad, que podrían deberse al cambio tecnológico, les dan una pendiente ascendente a muchas series de tiempo. Cualquier medida de la producción total, como las ventas de manufactura, se ve afectada por los cambios en la productividad.

Para las series de tiempo económicas y de negocios, lo mejor es ver a la tendencia (o ciclo de tendencia) como un cambio sin conflictos a través del tiempo. Es raro que se suponga con realismo que la tendencia puede representarse por alguna función simple, como una línea recta a lo largo todo el periodo, para la cual se observa la serie de tiempo. Sin embargo, es conveniente ajustar la curva de tendencia a una serie de tiempo, por dos razones: (1) proporciona cierta indicación de la dirección general de la serie observada, y (2) puede eliminarse de la serie original para obtener una imagen más clara de la estacionalidad.

Si la tendencia parece ser aproximadamente lineal, es decir, que aumenta o disminuye como una línea recta, se representa por la ecuación

$$\hat{T}_t = b_0 + b_1 t \quad (5.3)$$

Aquí,  $\hat{T}_t$  es el valor predicho para la tendencia en el tiempo  $t$ . El símbolo  $t$  utilizado para la variable independiente representa el tiempo y de forma ordinaria supone valores enteros de 1, 2, 3... correspondientes a los periodos consecutivos. El coeficiente de la pendiente  $b_1$  es el incremento promedio o decremento en  $T$  para cada incremento de un solo periodo en el tiempo.

Las ecuaciones de tendencias en el tiempo, entre ellas la tendencia en línea recta, pueden ajustarse a los datos por medio del *método de mínimos cuadrados*. Recuerde que este método selecciona los valores de los coeficientes en la ecuación de la tendencia ( $b_0$  y  $b_1$  en el caso de la línea recta) de forma que los valores de la tendencia estimada  $\hat{T}_t$  son los más cercanos a los valores reales de  $Y_t$  si la proximidad se mide mediante el criterio de la suma del cuadrado de los errores residuales.

$$SSE = \sum (Y_t - \hat{T}_t)^2 \quad (5.4)$$

### Ejemplo 5.1

Los datos en los registros anuales de nuevos automóviles de pasajeros en Estados Unidos, de 1960 a 1992, se muestran en la tabla 5.1 y están graficados en la figura 5.2. Los valores de 1960 a 1992 se utilizan para desarrollar una ecuación de tendencia. Los registros conforman la variable dependiente y la variable independiente es el tiempo  $t$  codificado como 1960 = 1, 1961 = 2, y así sucesivamente.

La tendencia ajustada por medio de una linea recta es

$$\hat{T}_t = 7.988 + .0687t$$

La pendiente de la ecuación de la tendencia indica que se estima que los registros se incrementarán en un promedio de 68 700 por año. La figura 5.3 muestra la tendencia en línea recta ajustada a los datos reales. La figura 5.3 también muestra pronósticos para registros de autos nuevos para los años 1993 y 1994 ( $t = 34$  y  $t = 35$ ) obtenidos por la extrapolación de la línea de tendencia. Más adelante se presentará más información acerca del pronóstico de una tendencia.

**TABLA 5.1 Registro de vehículos nuevos de pasajeros en Estados Unidos, 1960 – 1992, para el ejemplo 5.1**

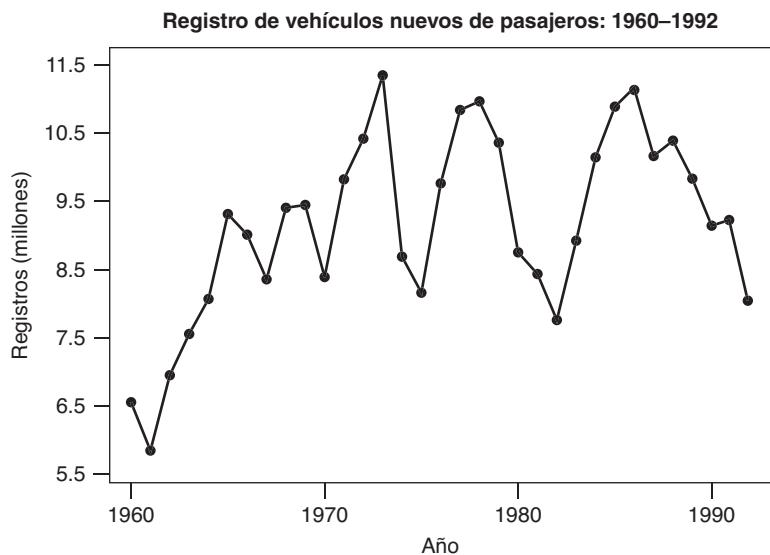
| <i>Año</i> | <i>Registros</i><br>(millones) | <i>Tiempo</i> | <i>Estimados de la tendencia</i><br>(millones) |           | <i>Error</i><br>(millones) |
|------------|--------------------------------|---------------|------------------------------------------------|-----------|----------------------------|
|            |                                |               | <i>Y</i>                                       | $\hat{T}$ |                            |
| 1960       | 6.577                          | 1             | 8.0568                                         | -1.4798   |                            |
| 1961       | 5.855                          | 2             | 8.1255                                         | -2.2705   |                            |
| 1962       | 6.939                          | 3             | 8.1942                                         | -1.2552   |                            |
| 1963       | 7.557                          | 4             | 8.2629                                         | -0.7059   |                            |
| 1964       | 8.065                          | 5             | 8.3316                                         | -0.2666   |                            |
| 1965       | 9.314                          | 6             | 8.4003                                         | 0.9138    |                            |
| 1966       | 9.009                          | 7             | 8.4690                                         | 0.5401    |                            |
| 1967       | 8.357                          | 8             | 8.5376                                         | -0.1807   |                            |
| 1968       | 9.404                          | 9             | 8.6063                                         | 0.7977    |                            |
| 1969       | 9.447                          | 10            | 8.6750                                         | 0.7720    |                            |
| 1970       | 8.388                          | 11            | 8.7437                                         | -0.3557   |                            |
| 1971       | 9.831                          | 12            | 8.8124                                         | 1.0186    |                            |
| 1972       | 10.409                         | 13            | 8.8811                                         | 1.5279    |                            |
| 1973       | 11.351                         | 14            | 8.9498                                         | 2.4012    |                            |
| 1974       | 8.701                          | 15            | 9.0185                                         | -0.3175   |                            |
| 1975       | 8.168                          | 16            | 9.0872                                         | -0.9192   |                            |
| 1976       | 9.752                          | 17            | 9.1559                                         | 0.5961    |                            |
| 1977       | 10.826                         | 18            | 9.2246                                         | 1.6014    |                            |
| 1978       | 10.946                         | 19            | 9.2933                                         | 1.6527    |                            |
| 1979       | 10.357                         | 20            | 9.3620                                         | 0.9950    |                            |
| 1980       | 8.761                          | 21            | 9.4307                                         | -0.6697   |                            |
| 1981       | 8.444                          | 22            | 9.4994                                         | -1.0554   |                            |
| 1982       | 7.754                          | 23            | 9.5681                                         | -1.8141   |                            |
| 1983       | 8.924                          | 24            | 9.6368                                         | -0.7128   |                            |
| 1984       | 10.118                         | 25            | 9.7055                                         | 0.4125    |                            |
| 1985       | 10.889                         | 26            | 9.7742                                         | 1.1148    |                            |
| 1986       | 11.140                         | 27            | 9.8429                                         | 1.2971    |                            |
| 1987       | 10.183                         | 28            | 9.9116                                         | 0.2714    |                            |
| 1988       | 10.398                         | 29            | 9.9803                                         | 0.4177    |                            |
| 1989       | 9.833                          | 30            | 10.0490                                        | -0.2160   |                            |
| 1990       | 9.160                          | 31            | 10.1177                                        | -0.9577   |                            |
| 1991       | 9.234                          | 32            | 10.1863                                        | -0.9524   |                            |
| 1992       | 8.054                          | 33            | 10.2550                                        | -2.2010   |                            |

Fuente: Datos del U.S. Department of Commerce, *Survey of Current Business* (varios años).

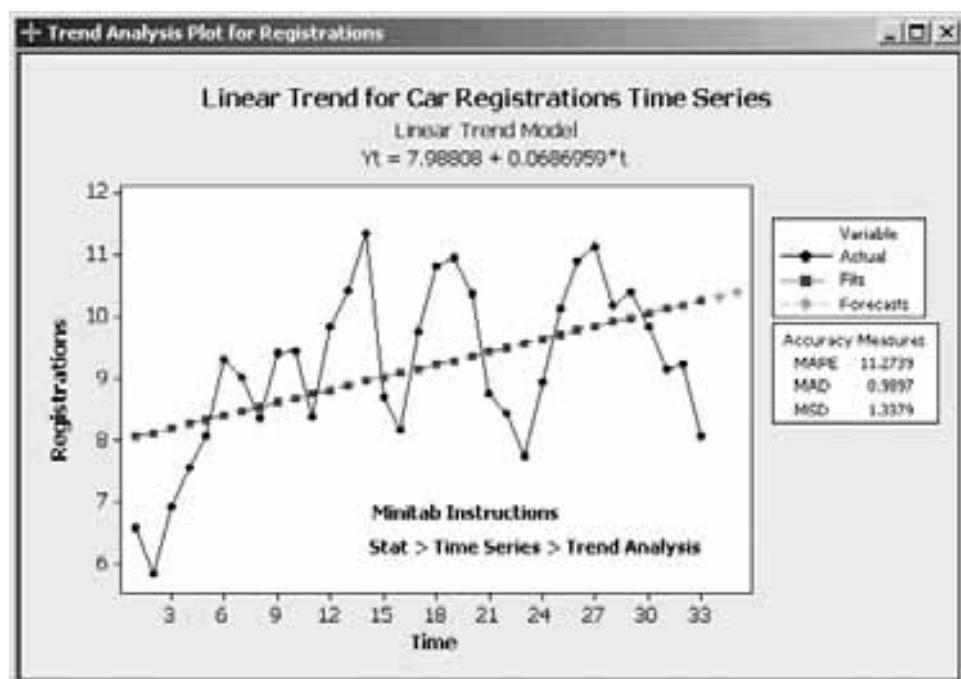
Los valores estimados de la tendencia para los registros de vehículos nuevos de pasajeros, de 1960 a 1992, se muestran en la tabla 5.1 bajo  $\hat{T}$ . Por ejemplo, la ecuación de la tendencia estima que los registros en 1992 ( $t = 33$ ) son

$$\hat{T}_{33} = 7.988 + .0687(33) = 10.225$$

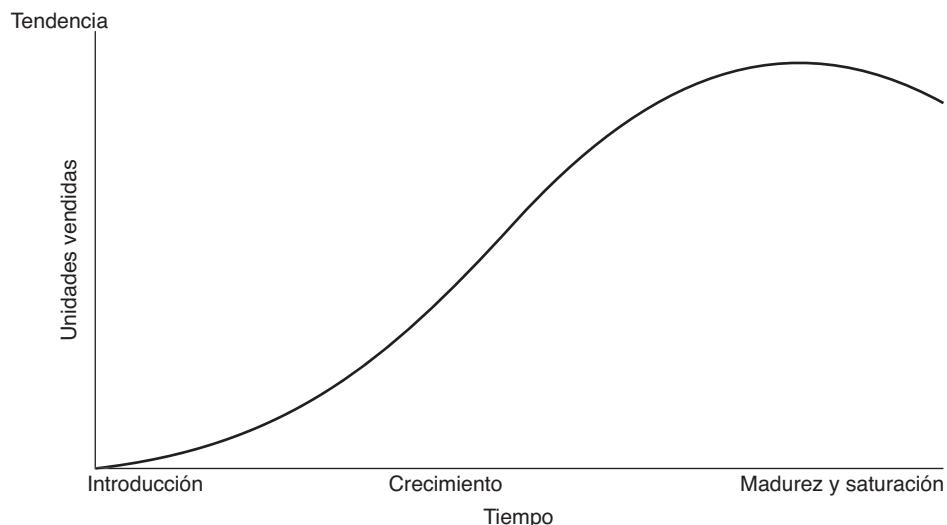
o 10 255 000 registros. Los registros para los vehículos nuevos de pasajeros fueron en realidad de 8 054 000 en 1992. Para 1992, la ecuación de tendencia sobreestima los registros por aproximadamente 2.2 millones. Éste y los demás errores de estimación definidos como  $Y - \hat{T}$  se listan en la tabla 5.1. Estos errores de estimación se utilizaron para calcular la precisión del ajuste mediante, *MAD*, *MSD* y *MAPE*, que se muestran en la figura 5.3. Éstas son las mismas medidas de precisión del pronóstico que fueron definidas por las ecuaciones (3.7), (3.8) y (3.9), respectivamente, en



**FIGURA 5.2 Serie de tiempo de registros de automóviles del ejemplo 5.1**



**FIGURA 5.3 Línea de tendencia para la serie de tiempo de registros de automóviles el ejemplo 5.1**



**FIGURA 5.4 Ciclo de vida típico de un nuevo producto.**

el capítulo 3. (Los comandos en Minitab empleados para producir los resultados en el ejemplo 5.1 se proporcionan al final de este capítulo.)

### Curvas de tendencia no lineales

El ciclo de vida de un nuevo producto tiene tres etapas: introducción, crecimiento, y madurez y saturación. Una curva que representa las ventas (en importe o unidades) sobre un nuevo ciclo de vida de un producto se muestra en la figura 5.4. El tiempo, que se muestra en el eje horizontal, puede variar de días a años según la naturaleza del mercado. Una tendencia en línea recta no funcionaría para estos datos. Los modelos lineales consideran que una variable es creciente (o decreciente) por una cantidad constante en cada periodo. Los incrementos por cada periodo en la curva del ciclo de vida del producto son distintos, según sea la fase del ciclo. Una curva, y no una línea recta, es necesaria para modelar la tendencia a lo largo del ciclo de vida de un nuevo producto.

Una función simple que permite la curvatura es la tendencia cuadrática

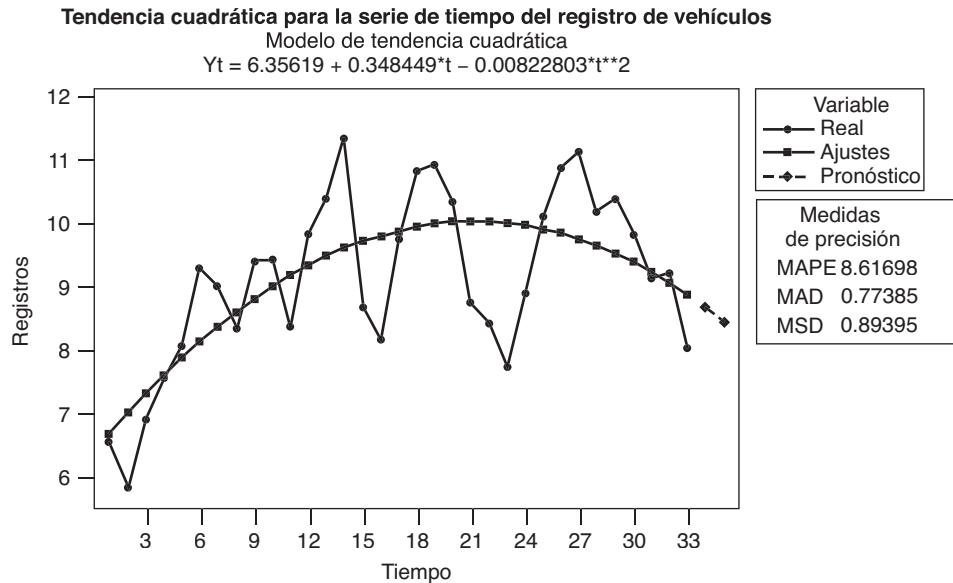
$$\hat{T}_t = b_0 + b_1 t + b_2 t^2 \quad (5.5)$$

La figura 5.5 muestra una curva de tendencia cuadrática ajustada a los datos de los registros de vehículos de pasajeros que se mencionan en el ejemplo 5.1 mediante el criterio SSE. La tendencia cuadrática puede proyectarse más allá de los datos, por ejemplo, para los años 1993 y 1994. En la siguiente sección, “Pronóstico de la tendencia”, se consideran las implicaciones de esta proyección.

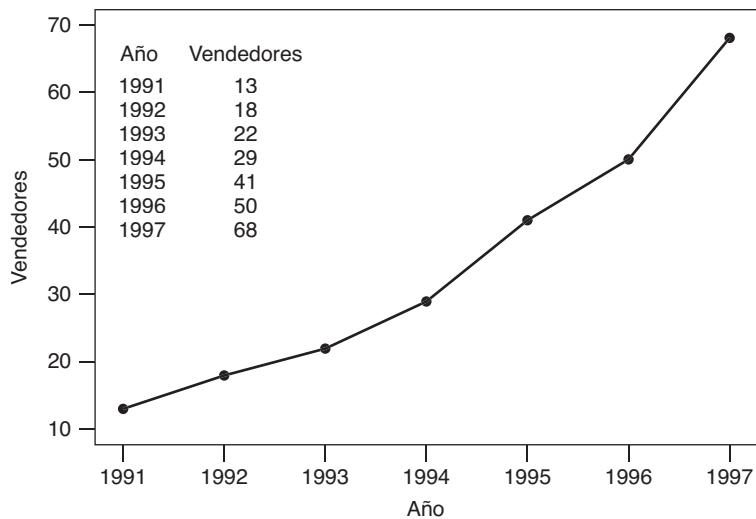
Con base en las medidas de precisión de MAPE, MAD y MSD, parece que la tendencia cuadrática representa mejor la tendencia de la serie de registros de vehículos que la tendencia lineal mostrada en la figura 5.3. ¿Cuál modelo es el apropiado? Antes de considerar esta cuestión, se presentan unas cuantas curvas de tendencia útiles.

Cuando una serie de tiempo comienza con lentitud y después aparenta incrementarse a un ritmo creciente (vea la figura 5.4) de manera que la diferencia porcentual que existe de una observación a otra es constante, puede ajustarse una tendencia exponencial. La tendencia exponencial está dada por

$$\hat{T}_t = b_0 b_1^t \quad (5.6)$$



**FIGURA 5.5 Curva de tendencia cuadrática para la serie de tiempo del registro de vehículos del ejemplo 5.1**



**FIGURA 5.6 Gráfica del número de vendedores de fondos mutuos**

El coeficiente  $b_1$  se relaciona con la tasa de crecimiento. Si la tendencia exponencial se ajusta a los datos anuales, se estima que la tasa de crecimiento anual será  $100(b_1 - 1)\%$ .

La figura 5.6 contiene el número de vendedores de fondos mutuos de una compañía en particular durante varios años consecutivos. El incremento del número de vendedores no es constante. Parece que un número cada vez mayor de personas se estuvieran sumando en los años recientes.

Una curva exponencial ajustada a los datos de los vendedores tiene la ecuación

$$\hat{T}_t = 10.016(1.313)^t$$

la cual implica una tasa de crecimiento anual de 31%, aproximadamente. En consecuencia, si el modelo estima 51 personas para 1996, el incremento para 1997 sería de  $16(51) \times 0.31$  para un estimado total de 67. Esto puede compararse con el valor real de 68 vendedores.

Una tendencia lineal ajustada a los datos del número de vendedores indicaría un incremento del promedio constante de nueve vendedores por año, aproximadamente. Esta tendencia sobreestima el incremento real en los primeros años y lo subestima en el último. No modela bien la tendencia como lo hace la curva exponencial.

Queda claro que extrapolar la curva exponencial con una tasa de crecimiento de 31% producirá rápidamente cifras muy elevadas. Éste es un problema potencial con un modelo de tendencias exponenciales. ¿Qué pasa cuando la economía entra en calma y los precios accionarios comienzan a retraerse? La demanda de vendedores de fondos mutuos disminuirá y el número de ellos incluso podría declinar. El pronóstico de la tendencia hecho por la curva exponencial será demasiado alto.

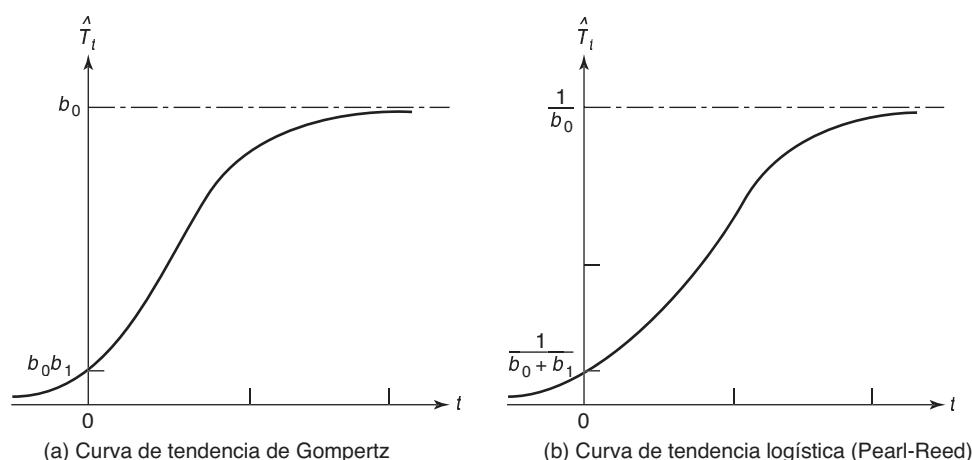
Las curvas de crecimiento de los tipos logísticos y de Gompertz representan la tendencia a crecer de muchas industrias o líneas de productos en una tasa decreciente en la medida que maduran. Si los datos graficados reflejan una situación en la que las ventas comienzan bajas, después aumentan conforme el producto comienza a darse a conocer y finalmente adquieren un nivel fijo cuando se alcanza la saturación, podrían ser adecuados la curva de Gompertz o el modelo de logística de Pearl-Reed. La figura 5.7 muestra una comparación de las formas generales de la curva de Gompertz (a) y el modelo de logística de Pearl-Reed (b). Observe que la curva de logística es muy similar a la de Gompertz, con una pendiente ligeramente más suave. La figura 5.7 muestra la manera en la que las intercepciones con el eje Y y los valores máximos para estas curvas están relacionados con algunos de los coeficientes en sus formas funcionales. Las fórmulas para estas curvas de tendencia son complejas y no están al alcance de este texto. Muchos paquetes de software de estadística, entre ellos Minitab, permiten que uno ajuste varios de los modelos de tendencia presentados en esta sección.

A pesar de que hay algunos criterios objetivos para seleccionar una tendencia apropiada, por lo general la elección correcta es cuestión de juicio y, por lo tanto, requiere de la experiencia y el sentido común del analista. Como se analizará en la siguiente sección, la línea de la curva que mejor se ajusta al conjunto de puntos de datos podría carecer de sentido cuando se proyecta como una tendencia del futuro.

### Pronóstico de la tendencia

Suponga que actualmente se está en el tiempo  $t = n$  (fin de la serie) y se desea utilizar un modelo de tendencia para predecir el valor de  $Y$ , para el periodo  $p$  en el futuro. El periodo

**FIGURA 5.7 Curvas de crecimiento en forma de S**



en el que se hace el pronóstico,  $n$  en este caso, se conoce como *origen del pronóstico*. El valor  $p$  se conoce como *tiempo principal*, el periodo futuro pronosticado a partir del origen el pronóstico. Para el modelo de tendencia lineal, se puede producir un pronóstico para evaluar  $\hat{T}_{n+p} = b_0 + b_1(n + p)$ .

Por medio de la línea de tendencia ajustada a los datos del registro vehicular que se presentan en el [ejemplo 5.1](#), un pronóstico de la tendencia para 1993 ( $t = 34$ ) hecho en 1992 ( $t = n = 33$ ) sería el pronóstico un paso adelante  $p = 1$

$$\hat{T}_{33+1} = 7.988 + .0687(33 + 1) = 7.988 + .0687(34) = 10.324$$

De manera semejante,  $p = 2$ , representa dos pasos por delante del pronóstico (1994) y está dado por

$$\hat{T}_{33+2} = 7.988 + .0687(33 + 2) = 7.988 + .0687(35) = 10.393$$

Estos dos pronósticos se muestran en la [figura 5.3](#) como extrapolaciones de la línea de tendencia ajustada.

La [figura 5.5](#) muestra la curva tendencia cuadrática ajustada para los datos del registro vehicular.

Por medio de la ecuación que se muestra en la figura, se pueden calcular los pronósticos de la tendencia para 1993 y 1994 estableciendo  $t = 33 + 1 = 34$  y  $t = 33 + 2 = 35$ . El lector podrá verificar que  $\hat{T}_{33+1} = 8.690$  y  $\hat{T}_{33+2} = 8.470$ . Estos números se graficaron en la [figura 5.5](#) como extrapolaciones de la curva de tendencia cuadrática.

Como los datos de los registros se miden en millones, los dos pronósticos para la tendencia producidos por la curva cuadrática son bastante diferentes de los que se obtienen de la ecuación de tendencia lineal. Es más, se mueven en la dirección opuesta. Si se tuvieran que extraer las tendencias cuadráticas y lineales para periodos adicionales, sus diferencias se magnificarían.

El ejemplo del registro vehicular ilustra por qué se debe tener gran cuidado al usar las curvas de tendencia ajustada para pronosticar tendencias futuras. Dos ecuaciones que podrían representar razonablemente la serie de tiempo observada pueden dar resultados completamente distintos cuando se proyectan para períodos futuros. Estas diferencias pueden ser sustanciales para tiempos principales largos (pronóstico de largo plazo).

Los modelos de la curva de tendencia se basan en los siguientes supuestos:

1. La curva correcta de tendencia ha sido seleccionada.
2. La curva que se ajusta al pasado es indicativa del futuro.

Estos supuestos sugieren que el juicio y la experiencia podrían desempeñar una función significativa en la selección y uso de una curva de tendencia. Para utilizar una curva de tendencia para pronosticar, se debe tener la capacidad de argumentar que se ha seleccionado la tendencia adecuada y que, con toda probabilidad, el futuro será como el pasado.

Hay criterios objetivos para seleccionar una curva de tendencia. Se analizan dos de ellos en capítulos posteriores: el criterio de información de Akaike (AIC, por sus siglas en inglés) y el criterio de información bayesiano (BIC, por sus siglas en inglés). A pesar de que éstos y otros criterios ayudan a determinar un modelo adecuado, no pueden reemplazar al buen juicio.

## Estacionalidad

Un patrón estacional es aquel que se repite cada año. En el caso de los datos anuales, la estacionalidad no es un tema porque no hay posibilidad de modelar un patrón dentro de un año con los datos registrados una sola vez en ese lapso. Sin embargo, las series de tiempo que consisten en observaciones semanales, mensuales o trimestrales con frecuencia muestran estacionalidad.

El análisis del componente estacional de una serie de tiempo tiene implicaciones inmediatas en el corto plazo y es de gran importancia para la administración en los niveles medio y bajo. Los planes de marketing, por ejemplo, han de tomar en cuenta los patrones estacionales esperados de las compras del consumidor.

Se han desarrollado diversos métodos para medir la variación estacional. La idea básica de todos estos métodos es, primero, estimar y eliminar la tendencia de la serie original y, después, suavizar el componente irregular. Teniendo en mente los modelos de descomposición, esto deja datos que contienen sólo variación estacional. Los valores estacionales se recopilan y se resumen para producir un número (generalmente un *número índice*) para cada intervalo observado del año (semana, mes, trimestre y así sucesivamente).

Así, la identificación del componente estacional en una serie de tiempo difiere del análisis de tendencia en dos formas, por lo menos:

1. La tendencia se determina directamente a partir de los datos originales; pero un componente estacional se determina indirectamente después de eliminar los demás componentes de los datos, de manera que sólo permanece la estacionalidad.
2. La tendencia se representa por una curva mejor ajustada, o ecuación; pero tiene que calcularse un valor estacional separado para cada intervalo observado del año (semana, mes o trimestre) y con frecuencia se presenta en la forma de un número índice.

Si se emplea una descomposición aditiva, los estimados de la tendencia y los componentes estacionales e irregulares se suman para producir la serie original. Si se utiliza una descomposición multiplicativa, los componentes individuales deberán multiplicarse para reconstruir la serie original y en esta formulación el componente estacional se representa mediante un conjunto de números índices. Éstos muestran cuáles períodos del año son relativamente bajos y cuáles son relativamente altos. Los índices estacionales describen el patrón estacional.

*Los números índices* son porcentajes que muestran los cambios a través del tiempo.

Con datos mensuales, por ejemplo, un índice estacional de 1.0 para un mes particular significa que el valor esperado para ese lapso es de 1/12 del total del año. Un índice de 1.25 para otro mes quiere decir que la observación para dicho mes está estimada en 25% adicional al 1/12 del total anual. Un índice mensual de 0.80 implica que el nivel esperado de actividad para ese mes es 20% inferior a 1/12 del total del año, y así sucesivamente. Los números índices indican los altibajos en los niveles de actividad en el curso de un año después de los efectos debidos a la tendencia (o ciclo de tendencia) y una vez que los componentes irregulares han sido eliminados.

Para resaltar la estacionalidad, primero se debe estimar y eliminar la tendencia. La tendencia puede estimarse con una de las curvas de tendencia, tal y como se dijo antes, o bien puede estimarse mediante un promedio móvil, como se mencionó en el capítulo 4.

Al suponer un modelo de descomposición multiplicativa, *la proporción del promedio móvil* es un método popular para medir la variación estacional. En este método la tendencia se estima mediante un promedio móvil centrado. Se ilustra el método de proporción del promedio móvil mediante las ventas mensuales de la Cavanaugh Company que se muestran en la figura 5.1 en el siguiente ejemplo.

#### Ejemplo 5.2

Para ilustrar el método de la proporción del promedio móvil para determinar los índices estacionales, se usan los datos correspondientes a las ventas mensuales de dos años de la Cavanaugh Company.<sup>3</sup> La tabla 5.2 proporciona los datos de las ventas mensuales de enero de 2000 a diciembre de 2001 para ejemplificar el comienzo de los cálculos. El primer paso para los datos mensuales consiste en calcular un promedio móvil de 12 meses (para los datos trimestrales se calcula un promedio móvil de cuatro meses). Debido a que todos los meses del año están incluidos en los cálculos de este promedio móvil, se eliminan los efectos derivados de los componentes estacionales y el promedio móvil contiene los componentes irregulares y de tendencia.

---

<sup>3</sup> Las unidades han sido omitidas y las fechas y nombre han sido modificados para proteger la identidad de la compañía.

**TABLA 5.2 Ventas de la Cavanaugh Company  
2000-2001, para el ejemplo 5.2**

| Periodo    | Ventas | Total del promedio móvil a 12 meses | Total del promedio móvil a 2 años | Promedio móvil centrado a 12 meses | Índice estacional |
|------------|--------|-------------------------------------|-----------------------------------|------------------------------------|-------------------|
| 2000       |        |                                     |                                   |                                    |                   |
| Enero      | 518    |                                     |                                   |                                    |                   |
| Febrero    | 404    |                                     |                                   |                                    |                   |
| Marzo      | 300    |                                     |                                   |                                    |                   |
| Abril      | 210    |                                     |                                   |                                    |                   |
| Mayo       | 196    |                                     |                                   |                                    |                   |
| Junio      | 186    |                                     |                                   |                                    |                   |
| Julio      | 247    | 1 4,869                             | 2 9,833                           | 3 409.7                            | 4 0.60            |
| Agosto     | 343    | 4,964                               | 9,916                             | 413.2                              | 0.83              |
| Septiembre | 464    | 4,952                               | 9,877                             | 411.5                              | 1.13              |
| Octubre    | 680    | 4,925                               | 9,962                             | 415.1                              | 1.64              |
| Noviembre  | 711    | 5,037                               | 10,067                            | 419.5                              | 1.69              |
| Diciembre  | 610    | 5,030                               | 10,131                            | 422.1                              | 1.45              |
| 2001       |        |                                     |                                   |                                    |                   |
| Enero      | 613    | 5,101                               | 10,279                            | 428.3                              | 1.43              |
| Febrero    | 392    | 5,178                               | 10,417                            | 434.0                              | 0.90              |
| Marzo      | 273    | 5,239                               | 10,691                            | 445.5                              | 0.61              |
| Abril      | 322    | 5,452                               | 11,082                            | 461.8                              | 0.70              |
| Mayo       | 189    | 5,630                               | 11,444                            | 476.8                              | 0.40              |
| Junio      | 257    | 5,814                               | 11,682                            | 486.8                              | 0.53              |
| Julio      | 324    | 5,868                               |                                   |                                    |                   |
| Agosto     | 404    |                                     |                                   |                                    |                   |
| Septiembre | 677    |                                     |                                   |                                    |                   |
| Octubre    | 858    |                                     |                                   |                                    |                   |
| Noviembre  | 895    |                                     |                                   |                                    |                   |
| Diciembre  | 664    |                                     |                                   |                                    |                   |

Los pasos (identificados en la tabla 5.2) para el cálculo de índices estacionales mediante el método de la proporción del promedio móvil se muestran a continuación.

- Paso 1** Calcule el total móvil a 12 meses y coloque el total para enero de 2000 hasta diciembre de ese año, entre junio y julio de 2000.
- Paso 2** Calcule un total móvil a dos años de manera que los promedios subsecuentes estén centrados en julio, en lugar de que esto ocurra entre uno y otro mes.
- Paso 3** Debido a que el total a dos años contiene datos para 24 meses (enero de 2000, una vez; de febrero a diciembre de 2000, dos veces; y enero de 2001, una vez) este total se centra (enfrente) en julio de 2000.
- Paso 4** Divida el total móvil de dos años entre 24 para obtener el promedio móvil centrado de 12 meses.
- Paso 5** El índice estacional para julio se calcula al dividir el valor real para julio entre el promedio móvil centrado a 12 meses.<sup>4</sup>

Repita los pasos del 1 al 5 comenzando por la segunda serie de tiempo, agosto de 2000, y así sucesivamente. El proceso termina cuando el total móvil a 12 meses ya no se puede calcular.

Puesto que hay varios estimados (que corresponden a distintos años) del índice estacional para cada mes, éstos deberán concentrarse para producir un número único. La mediana, en lugar

<sup>4</sup> Ésta es la operación de la proporción del promedio móvil que da nombre al procedimiento.

**TABLA 5.3 Resumen de los índices estacionales mensuales para la Cavanaugh Company para el ejemplo 5.2**

| Mes        | 1996  | 1997  | 1998  | 1999  | 2000  | 2001  | 2002 | Mediana | Índice estacional ajustado (Mediana X 1.0044) |
|------------|-------|-------|-------|-------|-------|-------|------|---------|-----------------------------------------------|
| Enero      | —     | 1.208 | 1.202 | 1.272 | 1.411 | 1.431 | —    | 1.272   | 1.278                                         |
| Febrero    | —     | 0.700 | 0.559 | 0.938 | 1.089 | 0.903 | —    | 0.903   | 0.907                                         |
| Marzo      | —     | 0.524 | 0.564 | 0.785 | 0.800 | 0.613 | —    | 0.613   | 0.616                                         |
| Abril      | —     | 0.444 | 0.433 | 0.480 | 0.552 | 0.697 | —    | 0.480   | 0.482                                         |
| Mayo       | —     | 0.424 | 0.365 | 0.488 | 0.503 | 0.396 | —    | 0.424   | 0.426                                         |
| Junio      | —     | 0.490 | 0.459 | 0.461 | 0.465 | 0.528 | —    | 0.465   | 0.467                                         |
| Julio      | 0.639 | 0.904 | 0.598 | 0.681 | 0.603 | 0.662 | —    | 0.651   | 0.654                                         |
| Agosto     | 1.115 | 0.913 | 0.889 | 0.799 | 0.830 | 0.830 | —    | 0.860   | 0.864                                         |
| Septiembre | 1.371 | 1.560 | 1.346 | 1.272 | 1.128 | 1.395 | —    | 1.359   | 1.365                                         |
| Octubre    | 1.792 | 1.863 | 1.796 | 1.574 | 1.638 | 1.771 | —    | 1.782   | 1.790                                         |
| Noviembre  | 1.884 | 2.012 | 1.867 | 1.697 | 1.695 | 1.846 | —    | 1.857   | 1.865                                         |
| Diciembre  | 1.519 | 1.088 | 1.224 | 1.282 | 1.445 | —     | —    | 1.282   | 1.288                                         |
|            |       |       |       |       |       |       |      | 11.948  | 12.002                                        |

de la media, se emplea para esta medida. El uso de la mediana elimina la influencia de los datos inusualmente grandes o pequeños para un mes en un año determinado. En la tabla 5.3 se presenta un resumen de las proporciones estacionales, así como del valor de la mediana para cada mes.

Los índices estacionales mensuales de cada año deben sumar 12, de manera que las medianas para cada mes deberán ajustarse para obtener la serie final de los índices estacionales.<sup>5</sup> Puesto que este multiplicador debe ser mayor que 1 si el total de las proporciones de la mediana antes del ajuste es menor a 12, y menor que 1 si el total fuera mayor que 12, el multiplicador se define como

$$\text{Multiplicador} = \frac{12}{\text{Total actual}}$$

Haciendo uso de la información contenida en la tabla 5.3

$$\text{Multiplicador} = \frac{12}{11.948} = 1.0044$$

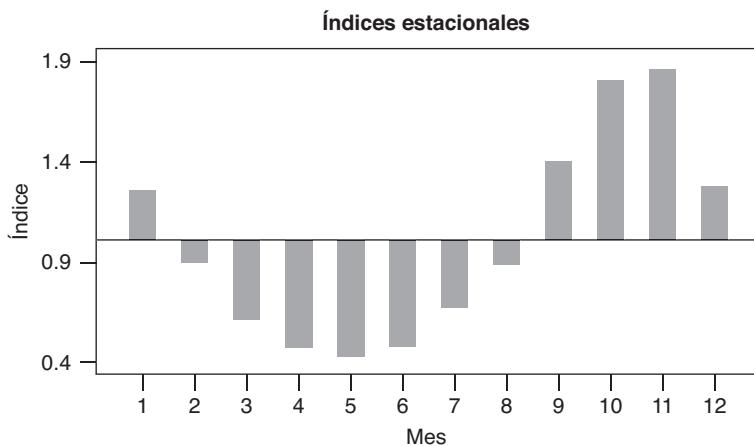
La columna final de la tabla 5.3 contiene los índices estacionales finales para cada mes, determinados mediante el ajuste (multiplicar por 1.0044) para cada una de las proporciones de la mediana.<sup>6</sup> Los índices estacionales finales, que se muestran en la figura 5.8, representan el componente estacional en una descomposición multiplicativa de las ventas en las series de tiempo de Cavanaugh Company.

La estacionalidad en las ventas es evidente a partir de la figura 5.8. Las ventas de esta compañía son periódicas, a finales de la primavera son relativamente bajas y al final del otoño, relativamente altas.

El análisis de las series de ventas del ejemplo 5.2 supone que el patrón estacional permaneció sin movimiento de un año a otro. Si éste cambiara con el tiempo, quizás la estimación del componente estacional con todos los datos podría arrojar resultados engañosos. En este caso, es mejor (1) utilizar únicamente los datos recientes (de los últimos años) para estimar el componente estacional, o (2) utilizar un modelo de serie de tiempo que permita analizar la evolución de la estacionalidad. En capítulos sucesivos, se analizan los modelos que permiten estudiar la evolución de la estacionalidad.

<sup>5</sup> Los índices mensuales deben sumar 12 de tal forma que el total anual esperado iguale al total real del año.

<sup>6</sup> Algunas veces los índices estacionales se multiplican por 100 y se expresan como porcentajes.



**FIGURA 5.8** Índices estacionales para la Cavanaugh Company del ejemplo 5.2

El análisis estacional que se ilustra en el ejemplo 5.2 es apropiado para un modelo de descomposición multiplicativa. Sin embargo, el método general descrito en los pasos 1 al 5 funciona para una descomposición aditiva si, en el paso 5, la estacionalidad se estima al restar la tendencia de la serie original en lugar de dividir entre la tendencia (promedio móvil) para obtener un índice. En una descomposición aditiva, el componente estacional se expresa en las mismas unidades que la serie original.

Además, a partir del ejemplo de ventas es evidente que determinar la tendencia mediante un promedio móvil centrado produce algunos valores faltantes al final de la serie. Esto es particularmente problemático si el objetivo es obtener un pronóstico. Para pronosticar valores futuros mediante un método de descomposición, deberán emplearse métodos alternos para estimar la tendencia.

Los resultados de un análisis estacional pueden utilizarse para (1) eliminar la estacionalidad en los datos, (2) pronosticar valores futuros, (3) evaluar posiciones actuales en ventas, inventario y envíos, por ejemplo, y (4) programar la producción.

## DATOS AJUSTADOS A LA ESTACIONALIDAD

Después de que el componente estacional ha sido aislado, se puede utilizar para calcular los *datos ajustados a la estacionalidad*. Para una descomposición aditiva, los datos ajustados a la estacionalidad se calculan al restar el componente estacional

$$Y_t - S_t = T_t + I_t$$

Para una descomposición multiplicativa, los datos ajustados a la estacionalidad se calculan al dividir las observaciones originales entre el componente estacional

$$\frac{Y_t}{S_t} = T_t \times I_t \quad (5.7)$$

La mayor parte de las series económicas publicadas por las agencias gubernamentales están ajustadas estacionalmente porque la variación estacional no es el interés primordial. En lugar de esto, lo que importa es el patrón general de la actividad económica, que es independiente de las fluctuaciones estacionales normales. Por ejemplo, los registros de

vehículos nuevos pueden incrementarse 10% de mayo a junio; pero, ¿es esto una indicación de que las ventas de automóviles nuevos están aumentando? La respuesta es “no”, si un incremento de 10% es típico en esta época del año debido a factores estacionales.

En una encuesta acerca de la adquisición de los datos ajustados a la estacionalidad, Bell y Hillmer (1984) encontraron que una gran variedad de usuarios valoran los ajustes estacionales y se identificaron tres motivos para tal ajuste:

1. El ajuste estacional permite una comparación confiable de los valores en distintos puntos en el tiempo.
2. Es más fácil comprender las relaciones entre las variables económicas o de negocios después de que el factor complicado de la estacionalidad ha sido eliminado de los datos.
3. El ajuste estacional puede ser un elemento útil en la producción de pronósticos de corto plazo de valores futuros de una serie de tiempo.

Bell y Hillmer concluyeron que “el ajuste estacional se hace para simplificar los datos, de tal manera que los usuarios que no son expertos en estadística los interpreten con mayor facilidad, sin una pérdida significativa de información” (p. 301).

### Variaciones cíclicas e irregulares

Los ciclos son fluctuaciones de largo plazo en forma de ola que ocurren de manera más frecuente en los macroindicadores de la actividad económica. Como se ha comentado, hasta el grado que se pueden medir, los ciclos no tienen un patrón consistente. No obstante, puede obtenerse una perspectiva de la conducta cíclica de una serie de tiempo al eliminar los componentes de tendencia y estacional mediante una descomposición multiplicativa.<sup>7</sup>

$$\frac{Y_t}{T_t \times S_t} = \frac{T_t \times C_t \times S_t \times I_t}{T_t \times S_t} = C_t \times I_t \quad (5.8)$$

Puede utilizarse un promedio móvil para suavizar las irregularidades,  $I_t$ , para obtener el componente cíclico,  $C_t$ . Para eliminar el problema centrado que se encuentra cuando se usa un promedio móvil con un número par de períodos, las irregularidades se suavizan mediante un promedio móvil con un número impar de períodos. Para los datos mensuales, funcionará un promedio móvil de 5, 7, 9, o incluso uno de 11 meses. Para los datos trimestrales, un estimado de  $C$  puede calcularse utilizando un promedio móvil de tres períodos de los valores.<sup>8</sup>

Finalmente, el componente irregular se calcula mediante

$$I_t = \frac{C_t \times I_t}{C_t} \quad (5.9)$$

El componente irregular representa la variabilidad en las series de tiempo después de que los demás componentes han sido eliminados. Algunas veces se le conoce como *residual* o *error*. Por medio de una descomposición multiplicativa, los componentes cíclico e irregular se expresan como índices.

Una razón para descomponer una serie de tiempo es aislarla y examinar los componentes de la serie. Cuando el analista tiene la capacidad de ver uno por uno los componentes de tendencia, estacionales, cíclicos e irregulares de una serie, pueden obtenerse puntos de vista acerca de los patrones en los valores de los datos originales. También, una vez que los componentes han sido aislados, pueden recombinarse o sintetizarse para producir pronósticos de valores futuros de las series de tiempo.

<sup>7</sup> Observe que hemos añadido el componente cíclico  $C$  a la descomposición multiplicativa mostrada en la ecuación 5.2.

<sup>8</sup> Para los datos anuales no existe componente estacional y el componente irregular  $\times$  cíclico se obtiene simplemente al eliminar la tendencia de la serie original.

TABLA 5.4 Descomposición multiplicativa para las ventas de Outboard Marine

|    | <i>Año</i> | <i>Trimestre</i> | <i>Ventas</i> | <i>T</i> | <i>SCI</i> | <i>S</i> | <i>TCI</i> | <i>CI</i> | <i>C</i> | <i>I</i> |
|----|------------|------------------|---------------|----------|------------|----------|------------|-----------|----------|----------|
| 1  | 1990       | 1                | 232.7         | 255.026  | 0.912      | 0.780    | 298.486    | 1.170     | —        | —        |
| 2  |            | 2                | 309.2         | 256.310  | 1.206      | 1.016    | 304.297    | 1.187     | 1.146    | 1.036    |
| 3  |            | 3                | 310.7         | 257.594  | 1.206      | 1.117    | 278.175    | 1.080     | 1.103    | 0.979    |
| 4  |            | 4                | 293.0         | 258.878  | 1.132      | 1.087    | 269.459    | 1.041     | 1.044    | 0.997    |
| 5  | 1991       | 1                | 205.1         | 260.162  | 0.788      | 0.780    | 263.084    | 1.011     | 0.978    | 1.034    |
| 6  |            | 2                | 234.4         | 261.446  | 0.897      | 1.016    | 230.683    | 0.882     | 0.955    | 0.924    |
| 7  |            | 3                | 285.4         | 262.730  | 1.086      | 1.117    | 255.524    | 0.973     | 0.919    | 1.059    |
| 8  |            | 4                | 258.7         | 264.014  | 0.980      | 1.087    | 237.914    | 0.901     | 0.936    | 0.963    |
| 9  | 1992       | 1                | 193.2         | 265.298  | 0.728      | 0.780    | 247.820    | 0.934     | 0.936    | 0.998    |
| 10 |            | 2                | 263.7         | 266.583  | 0.989      | 1.016    | 259.518    | 0.974     | 0.962    | 1.013    |
| 11 |            | 3                | 292.5         | 267.867  | 1.092      | 1.117    | 261.880    | 0.978     | 1.009    | 0.969    |
| 12 |            | 4                | 315.2         | 269.151  | 1.171      | 1.087    | 289.875    | 1.077     | 0.967    | 1.114    |
| 13 | 1993       | 1                | 178.3         | 270.435  | 0.659      | 0.780    | 228.708    | 0.846     | 0.972    | 0.870    |
| 14 |            | 2                | 274.5         | 271.719  | 1.010      | 1.016    | 270.147    | 0.994     | 0.936    | 1.062    |
| 15 |            | 3                | 295.4         | 273.003  | 1.082      | 1.117    | 264.477    | 0.969     | 0.974    | 0.995    |
| 16 |            | 4                | 286.4         | 274.287  | 1.044      | 1.087    | 263.389    | 0.960     | 0.939    | 1.022    |
| 17 | 1994       | 1                | 190.8         | 275.571  | 0.692      | 0.780    | 244.742    | 0.888     | 0.928    | 0.957    |
| 18 |            | 2                | 263.5         | 276.855  | 0.952      | 1.016    | 259.321    | 0.937     | 0.950    | 0.986    |
| 19 |            | 3                | 318.8         | 278.139  | 1.146      | 1.117    | 285.427    | 1.026     | 0.989    | 1.037    |
| 20 |            | 4                | 305.3         | 279.423  | 1.093      | 1.087    | 280.770    | 1.005     | 1.047    | 0.960    |
| 21 | 1995       | 1                | 242.6         | 280.707  | 0.864      | 0.780    | 311.186    | 1.109     | 1.075    | 1.032    |
| 22 |            | 2                | 318.8         | 281.991  | 1.131      | 1.016    | 313.744    | 1.113     | 1.088    | 1.023    |
| 23 |            | 3                | 329.6         | 283.275  | 1.164      | 1.117    | 295.097    | 1.042     | 1.082    | 0.963    |
| 24 |            | 4                | 338.2         | 284.559  | 1.189      | 1.087    | 311.027    | 1.093     | 1.059    | 1.032    |
| 25 | 1996       | 1                | 232.1         | 285.843  | 0.812      | 0.780    | 297.718    | 1.042     | 1.038    | 1.004    |
| 26 |            | 2                | 285.6         | 287.127  | 0.995      | 1.016    | 281.071    | 0.979     | 0.975    | 1.004    |
| 27 |            | 3                | 291.0         | 288.411  | 1.009      | 1.117    | 260.537    | 0.903     | 0.925    | 0.976    |
| 28 |            | 4                | 281.4         | 289.695  | 0.971      | 1.087    | 258.791    | 0.893     | —        | —        |

**Ejemplo 5.3**

En el ejemplo 3.5 (vea la página 69), Perkin Kendell, analista de Outboard Marine Corporation, utilizó el análisis de autocorrelación para determinar que las ventas eran estacionales sobre una base trimestral. Ahora, usa la descomposición para entender la variable de las ventas trimestrales. Perkin usa Minitab (vea “Aplicaciones en Minitab” al final del capítulo) para obtener la tabla 5.4 y la figura 5.9. A fin de mantener el patrón estacional actual, sólo se analizan los datos de ventas (*Y*) de los últimos siete años (de 1990 a 1996).

Los datos originales se muestran en la parte superior izquierda de la figura 5.10. La tendencia se calcula por medio del modelo lineal:  $\hat{T} = 253.742 + 1.284t$ . Debido a que 1 representa el primer trimestre de 1990, la tabla 5.4 muestra que el valor de la tendencia es igual a 255.026 para este periodo y las ventas estimadas (la columna *T*) se incrementaron 1.284 cada trimestre.

La gráfica en la parte superior de la figura 5.10 muestra los datos sin la tendencia. Éstos también se presentan en la columna *SCI* de la tabla 5.4. El valor sin tendencia para el primer trimestre de 1990 fue<sup>9</sup>

$$SCI = \frac{Y}{T} = \frac{232.7}{255.026} = .912$$

<sup>9</sup> Para simplificar la notación de este ejemplo, se omitió el subíndice *t* en los datos originales *Y* y en cada uno de sus componentes *T*, *S*, *C* e *I*. También se omitió el signo de multiplicar,  $\times$ , entre los componentes, ya que está claro que se considera una descomposición multiplicativa.

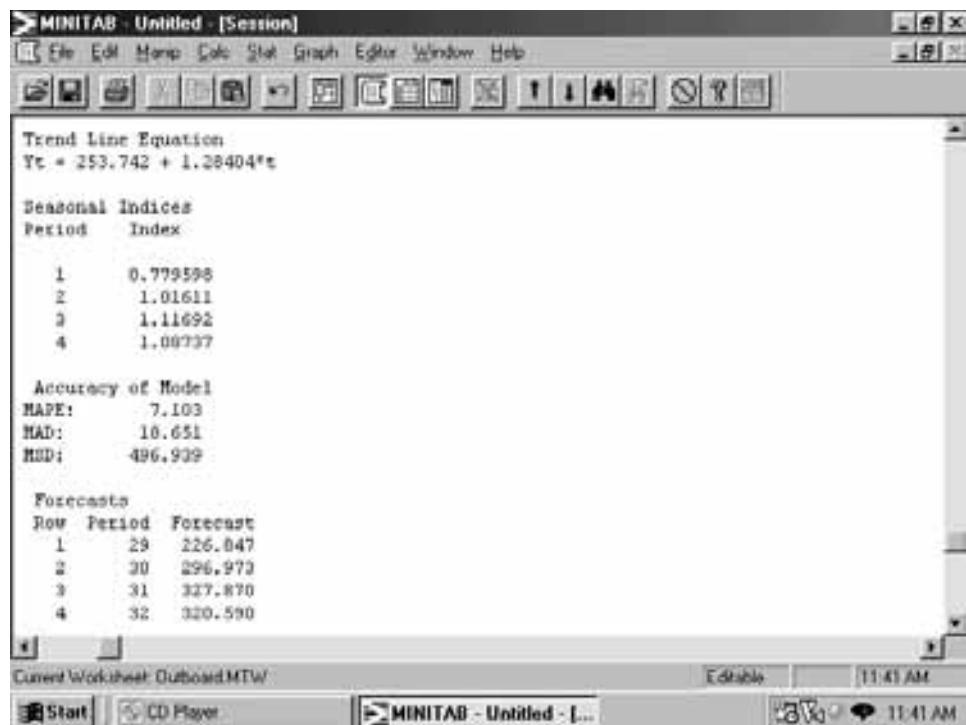


FIGURA 5.9 Resultados en Minitab para la descomposición de las ventas trimestrales de Outboard Marine del ejemplo 5.3

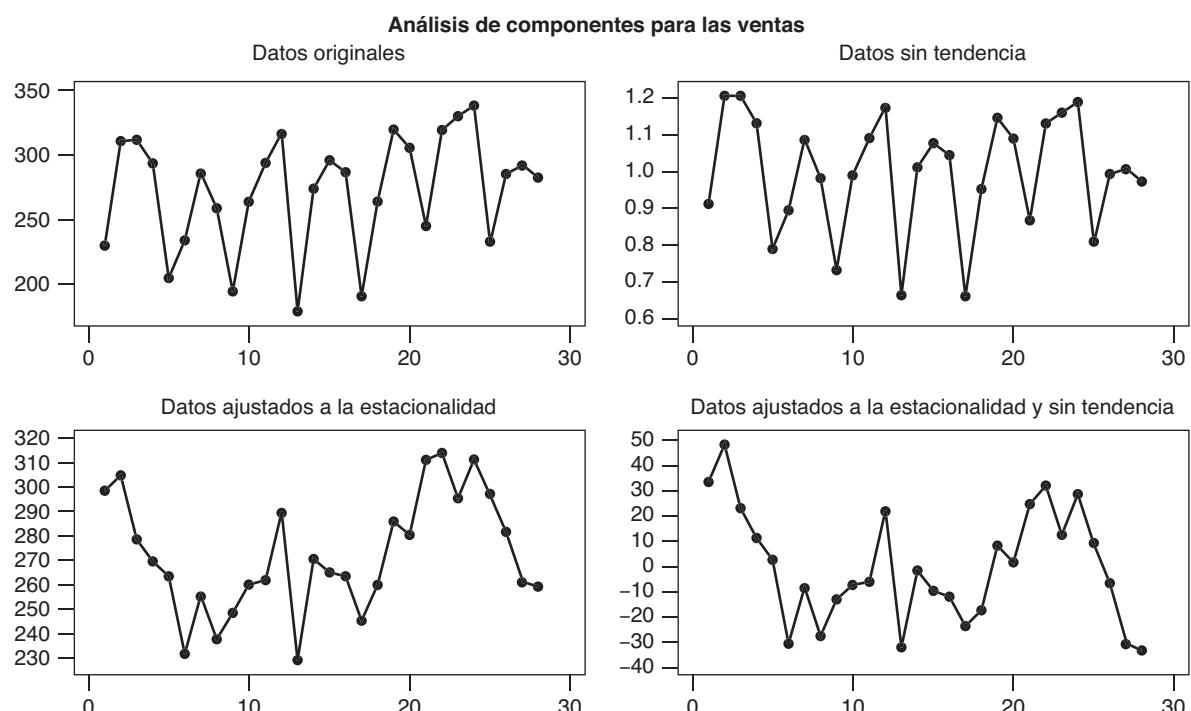
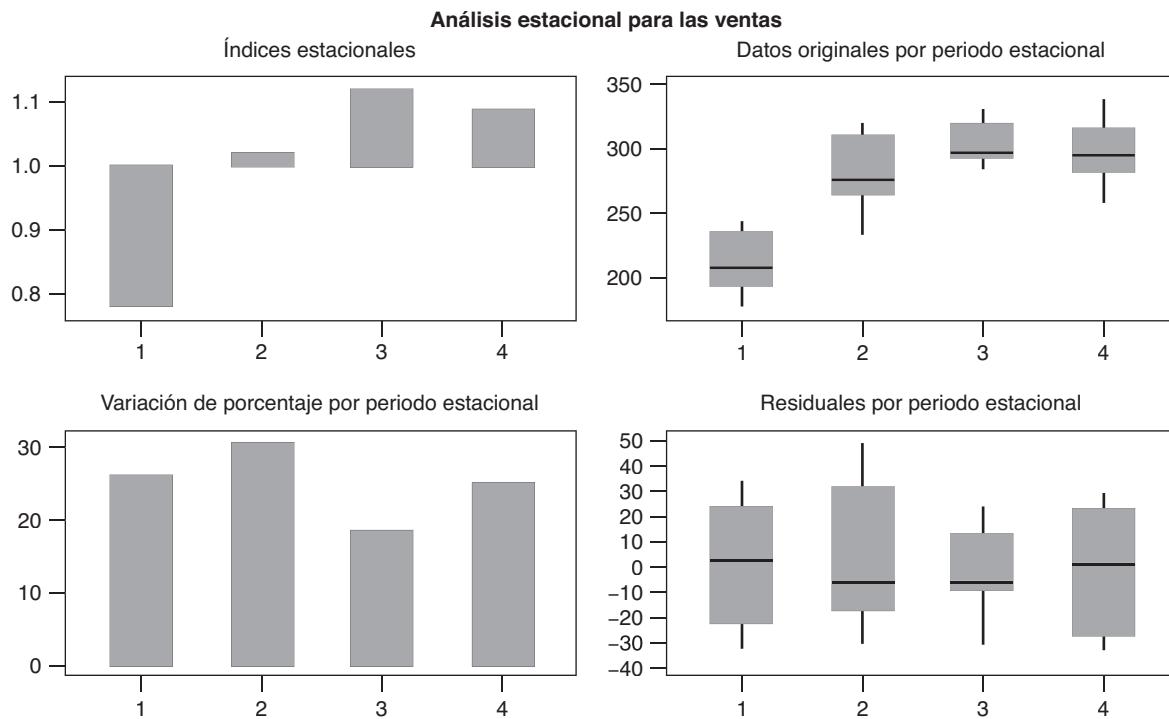


FIGURA 5.10 Análisis de componentes de las ventas de Outboard Marine



Los datos ajustados estacionalmente se muestran en la columna *TCI* en la tabla 5.4 y en la gráfica de la parte inferior izquierda de la figura 5.10. El valor ajustado a la estacionalidad para el primer trimestre de 1990 fue

$$TCI = \frac{232.7}{.7796} = 298.486$$

Las ventas del primer trimestre de 1995 fueron de 242.6. Sin embargo, el análisis de la columna ajustada a la estacionalidad muestra que las ventas de este trimestre fueron altas en realidad al ajustar los datos al hecho de que, por lo general, el primer trimestre es muy débil.

Los índices estacionales en la figura 5.9 fueron

$$\begin{aligned} \text{Primer trimestre} &= .77960 \rightarrow 78.0\% \\ \text{Segundo trimestre} &= 1.01611 \rightarrow 101.6\% \\ \text{Tercer trimestre} &= 1.11692 \rightarrow 111.7\% \\ \text{Cuarto trimestre} &= 1.08737 \rightarrow 108.7\% \end{aligned}$$

La gráfica en la parte superior izquierda de la figura 5.11 muestra los componentes estacionales relativos a 1.0. Se aprecia que las ventas del primer trimestre son 22% inferiores al promedio; las del segundo, cercanas al promedio; las del tercero son casi 12% superiores y las del cuarto trimestre fueron casi 9% superiores.

El valor del componente irregular cíclico (CI) para el primer trimestre de 1990 fue<sup>10</sup>

$$CI = \frac{Y}{TS} = \frac{232.7}{(255.026)(.7796)} = 1.170$$

<sup>10</sup> Minitab calcula el componente cíclico × el irregular (o simplemente el componente irregular si no se considera un componente cíclico) al restar el componente de tendencia × el estacional de los datos originales. En símbolos, Minitab establece que  $CI = Y - TS$ . El componente  $CI$  de Minitab se muestra en la gráfica inferior derecha de la figura 5.10, y a manera de bloques en la gráfica inferior derecha de la figura 5.11.

Para calcular la columna correspondiente al ciclo se calculó un promedio móvil de tres períodos. El valor del segundo trimestre de 1990 fue

$$\begin{array}{r} 1.170 \\ 1.187 \\ \hline 1.080 \\ \hline 3.437 \end{array} \quad 3.437/3 = 1.146$$

Observe qué tan suave es la columna *C* en comparación con la *CI*. La razón para ello es que al usar el promedio móvil se suavizaron las irregularidades. Finalmente, se calcula la columna *I*. Por ejemplo, para el segundo trimestre de 1990

$$I = \frac{CI}{C} = \frac{1.187}{1.146} = 1.036$$

Un examen de la columna *I* muestra que hubo grandes cambios en los componentes irregulares. El índice irregular cayó de 111.4% en el cuarto trimestre de 1992 a 87% en el primer trimestre de 1993 y después creció a 106.2% en el segundo trimestre de ese último año. Esta conducta es el resultado de ventas inusualmente bajas durante el primer trimestre de 1993.

Los índices del ciclo pueden emplearse para responder a las siguientes preguntas:

1. ¿Tiene ciclos la serie?
2. Si eso sucede, ¿cuáles son los extremos del ciclo?
3. ¿Sigue la serie el estado general de la economía (ciclo de negocios)?

Una manera de investigar los patrones cíclicos es mediante el estudio de los indicadores de negocios. Un *índicador de negocios* es una serie de tiempo relacionada con los negocios que se emplea para evaluar el estado general de la economía, particularmente en cuanto al ciclo de negocios. Muchos empresarios y economistas siguen de manera sistemática los movimientos de dichas series estadísticas para obtener información económica y empresarial en la forma de una imagen desdoblada, actualizada, detallada, relativamente objetiva y capaz de ser leída y comprendida en un tiempo mínimo.

Los *índicadores de negocios* son series de tiempo relacionadas con los negocios que se utilizan para evaluar el estado general de la economía.

La lista más importante de indicadores estadísticos se originó durante la recesión empresarial de 1937 y 1938. El secretario del Tesoro, Henry Morgenthau, solicitó que el National Bureau of Economic Research (NBER) diseñase un sistema que indicara cuándo terminaría la recesión. Bajo el liderazgo de Wesley Mitchell y Arthur F. Burns, los economistas de NBER seleccionaron 21 series que, según el comportamiento anterior, prometían ser indicadores muy confiables de la reactivación económica. Desde entonces, la institución ha revisado las series en repetidas ocasiones. La lista actual consta de 22 indicadores: 11 clasificados como principales, cuatro como coincidentes y siete como retrasos.

1. *Indicadores principales*. En la práctica, los componentes de la serie guía se estudian para ayudar a anticipar los momentos decisivos en los que la economía cambia. *Survey of Current Business* publica este listado cada mes junto con los valores reales de cada serie de los últimos meses y del año más reciente. También, se calcula un índice compuesto de los indicadores principales para cada mes y año; frecuentemente, el valor mensual más reciente se da a conocer en la prensa popular para indicar la dirección general de la economía futura. Son ejemplos de los indicadores principales, los nuevos pedidos de los fabricantes y el índice de los precios accionarios.

2. *Indicadores coincidentes.* Los cuatro indicadores coincidentes proporcionan una medida de cómo se desempeña actualmente la economía estadounidense. Cada mes se calcula un índice de estas cuatro series. Los ingresos personales y las ventas son ejemplos de los indicadores coincidentes.
3. *Indicadores de retraso.* Éstos tienen la tendencia a quedar atrás del estado general de la economía, ya sea que ésta esté al alza o a la baja. También se calcula un índice compuesto para esta lista. Las tasas de interés y los créditos comerciales son ejemplos de estos indicadores.

Los ciclos implican puntos de inflexión. Es decir, puntos que representan el inicio de una caída o un crecimiento en el ciclo de negocios. Los indicadores principales cambian de dirección antes que los cambios en la actividad empresarial general, los indicadores coincidentes lo hacen aproximadamente al mismo tiempo en que cambia la economía y los indicadores retrasados se presentan después de los cambios en la economía general. Sin embargo, es difícil identificar los puntos de inflexión cíclicos al momento que ocurren porque no todas las áreas de la economía se expanden al mismo tiempo, ni todas se contraen al mismo tiempo. En consecuencia, podrían pasar varios meses antes de que se identifica con precisión un giro cíclico genuino.

En su artículo titulado “Early Warning Signals for the Economy”, Geoffrey H. Moore y Julius Shiskin (1976) expresan lo siguiente acerca de la utilidad de los indicadores del ciclo de negocios:

Al ver el registro, parece claro que los indicadores del ciclo de negocios son útiles para juzgar el tono de los negocios actuales y de los prospectos de corto plazo. Pero debido a sus limitaciones, los indicadores deben utilizarse de manera conjunta con otros datos y con toda conciencia de los antecedentes de los negocios y de la confianza y expectativas de los consumidores, las políticas gubernamentales y los sucesos internacionales. También debemos anticipar que con frecuencia los indicadores serán difíciles de interpretar, que las interpretaciones variarán entre los analistas y que las señales que ofrecen no se interpretarán correctamente. Los indicadores proporcionan una imagen reveladora y sensible de la marea económica que necesita conocer un hábil analista de la economía, la política y la escena internacional. Con esto en mente, le es posible mejorar sus probabilidades de hacer un pronóstico válido de las tendencias económicas de corto plazo. Si el analista está consciente de sus limitaciones y está al tanto del mundo que le rodea, encontrará que los indicadores son guías para tomar acciones en la economía y en sus tendencias (p. 81).

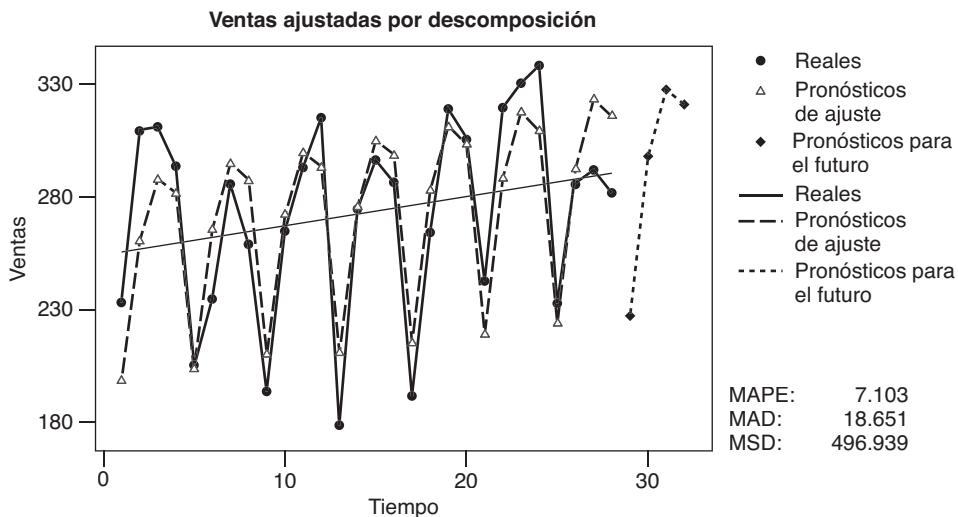
Por lo general, los componentes cíclicos de series de tiempo individuales se ajustan vagamente —y en algunos casos no lo hacen del todo— a un ciclo de negocios identificado con los indicadores NBER. No obstante, si se estima un componente cíclico para una serie de tiempo determinada, siempre deberá graficarse a lo largo del tiempo para obtener una indicación de las magnitudes y longitudes de cualquier ciclo que pudiese existir. Además, la gráfica puede examinarse para ver cualquier relación que exista entre las altas y las bajas de la actividad económica en general.

Hasta ahora, se ha mostrado cómo los factores que crean variación en una serie de tiempo pueden separarse y estudiarse de forma individual. El *análisis* es el proceso de separar las series de tiempo; la *síntesis* es el proceso de juntarlas de nuevo. Se deben reunir nuevamente los componentes de las series de tiempo para realizar el pronóstico.

## **PRONÓSTICO DE UNA SERIE DE TIEMPO ESTACIONAL**

---

Al pronosticar series de tiempo estacionales, el proceso de descomposición se realiza a la inversa. En lugar de separar las series en componentes individuales para su análisis, éstos



**FIGURA 5.12 Pronósticos para el ajuste por descomposición y pronósticos futuros para las ventas de Outboard Marine**

se combinan de nuevo para realizar los pronósticos de los períodos futuros. Se usarán el modelo multiplicativo y los resultados del ejemplo 5.3 para desarrollar pronósticos para las ventas de Outboard Marine.

#### Ejemplo 5.4

Los pronósticos para las ventas de Outboard Marine Corporation para los cuatro trimestres de 1997 pueden desarrollarse por medio de la tabla 5.4.

1. *Tendencia.* La ecuación de la tendencia trimestral es  $\hat{T}_t = 253.742 + 1.284t$ . El pronóstico se realizó en el cuarto trimestre de 1996, o el periodo  $t = n = 28$ . Las ventas para el primer trimestre de 1997 ocurrieron durante el periodo  $t = 28 + 1 = 29$ . Esta notación muestra que se está pronosticando el periodo  $p = 1$  a partir del final de la serie de tiempo. Al establecer  $t = 29$ , la proyección de la tendencia es entonces

$$\hat{T}_{29} = 253.742 + 1.284(29) = 290.978$$

2. *Estacional.* El índice estacional para el primer trimestre, 0.7796, se da en la figura 5.9.
3. *Cíclico.* La proyección cíclica debe determinarse a partir del patrón cíclico estimado (si lo hay) y de otra información generada por los indicadores de la economía general para 1997. La proyección del patrón cíclico para períodos futuros está cargada de incertidumbre y, como ya se dijo, se supone que será incluida en la tendencia para fines de pronóstico. Para mostrar cómo se finaliza este ejemplo, el índice cíclico se establece en 1.0.
4. *Irregular.* Las fluctuaciones irregulares representan la variación aleatoria que no puede explicarse con otros componentes. Para el pronóstico, el componente irregular se fija en el valor promedio de 1.0.<sup>11</sup>

El pronóstico para el primer trimestre de 1997 es

$$\hat{Y}_{29} = T_{29} \times S_{29} \times C_{29} \times I_{29} = (290.978)(0.7796)(1.0)(1.0) = 226.846$$

Los pronósticos para el resto de 1997 son

$$\begin{aligned} \text{Segundo trimestre} &= 296.973 \\ \text{Tercer trimestre} &= 327.870 \\ \text{Cuarto trimestre} &= 320.590 \end{aligned}$$

<sup>11</sup> Para los pronósticos que se generan a partir de un modelo aditivo, el índice irregular se fija en el valor promedio 0.

Los pronósticos ajustados por descomposición multiplicativa de las ventas de Outboard Marine Corporation, junto con los pronósticos para 1997, se muestran en la figura 5.12. Se puede ver ahí que el pronóstico de ajuste, construido a partir de los componentes de tendencia y estacionales, representan bastante bien a los datos reales. Sin embargo, el ajuste no es bueno para los últimos dos trimestres de 1996. Los pronósticos para 1997 imitan el patrón del ajuste.

Los pronósticos producidos por un modelo de descomposición aditiva y multiplicativa reflejan la importancia de los componentes individuales. Si una variable es muy estacional, los pronósticos tendrán un fuerte patrón estacional. Además, si hubiera una tendencia, los pronósticos se regirán por el patrón estacional impuesto en la tendencia extrapolada. Si algún componente domina el análisis, éste podría proporcionar un pronóstico de corto plazo, práctico y preciso.

## MÉTODO DE DESCOMPOSICIÓN DE CENSO II

---

Los métodos de descomposición de series de tiempo tienen una larga historia. En la segunda y tercera décadas del siglo xx, la Federal Reserve Board y el National Bureau of Economic Research estaban muy involucradas en los ajustes estacionales y el suavizamiento de las series de tiempo económicas. No obstante, antes del desarrollo de las computadoras los cálculos por descomposición eran laboriosos y la aplicación práctica de los métodos era limitada. A principios de los cincuenta, Julius Shinkin, estadístico económico y jefe en el Bureau of the Census desarrolló un programa de cómputo de gran escala para descomponer las series de tiempo. El primer programa se aproximaba básicamente a los métodos manuales que se usaron en ese tiempo y, un año después, fue reemplazado por un programa mejorado conocido como Método II. Con el paso de los años, se presentaron variantes mejoradas del Método II. Las variantes actuales del programa de descomposición de las series de tiempo del Census Bureau se conocen como ARIMA-X-12. Este programa está disponible en el Census Bureau, sin costo, y es ampliamente usado por las agencias gubernamentales y las compañías privadas.<sup>12</sup>

La descomposición por Census II es regularmente de índole multiplicativa porque la mayor parte de las series de tiempo económicas tienen una variación estacional que se incrementa con el nivel de las series. La descomposición también considera tres componentes: de ciclo-tendencia, estacional e irregular.

El método de descomposición Census II recorre una serie de pasos hasta que los componentes se ven aislados con éxito. Muchos de los pasos involucran la aplicación de los promedios móviles ponderados a los datos originales, que resulta en una pérdida inevitable de datos al principio y al final de la serie debido al hecho de promediar. La ARIMA en su parte ARIMA-X-12 proporciona la posibilidad de extender la serie original en ambas direcciones con los pronósticos, de tal manera que más observaciones se ajusten por medio de los promedios móviles completamente ponderados. Estos pronósticos se generan desde un modelo de series de tiempo ARIMA (vea el capítulo 9).

A continuación, se describen los pasos para cada una de las iteraciones del método Census II conforme se aplican en ARIMA-X-12. Parecería que los métodos son complicados debido a que se implican muchos pasos. Sin embargo, la idea básica es bastante simple, aislar los componentes del ciclo y de tendencia, los estacionales y los irregulares, uno por uno. Las diversas iteraciones están diseñadas para mejorar el estimado de cada uno de los componentes. Algunas buenas referencias para estudios adicionales son *Forecasting: Methods and Applications* (tercera edición) de Makridakis, Wheelwright y Hyndman (1998), y de Findley *et al.* (1998), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program".

**Paso 1** Un promedio móvil de  $s$  períodos se aplica a los datos originales para obtener un estimado general del ciclo de tendencia. (Para los datos mensuales,  $s = 12$ ; para los datos trimestrales  $s = 4$ ; y así sucesivamente.)

---

<sup>12</sup> La versión del programa ARIMA-X-12 para PC puede descargarse del sitio de Internet del U.S. Census Bureau. Al momento en que este libro fue escrito, ésta era la dirección de tal sitio: [www.census.gov/srd/www-x12a/x12down-pc.html](http://www.census.gov/srd/www-x12a/x12down-pc.html)

**Paso 2** Las proporciones de los datos originales relativos a estos valores de promedio móvil se calculan como en una descomposición multiplicativa clásica que se ilustra en el [ejemplo 5.2](#).

**Paso 3** Las proporciones del paso 2 contienen tanto los componentes estacionales como los irregulares. También se incluyen los valores extremos que resultan de sucesos inusuales, como huelgas o guerras. Las proporciones se dividen entre un estimado del componente estacional para dar un estimado del componente irregular. Un valor demasiado grande del término irregular indica un valor extremo en los datos originales. Estos valores extremos se identifican y las proporciones del paso 2 se ajustan adecuadamente. Esto elimina de manera eficaz los valores que no se ajustan al patrón de los datos restantes. Los valores que faltan al principio y al final de las series también son reemplazados por los estimados en esta etapa.

**Paso 4** Las proporciones creadas a partir de los datos codificados (con valores extremos reemplazados y las estimaciones de los valores faltantes) están suavizadas por un promedio móvil para eliminar la variación irregular. Esto crea un estimado preliminar del componente estacional.

**Paso 5** Los datos originales se dividen entre el componente estacional preliminar del paso 4 para obtener las series preliminares ajustadas estacionalmente. Esta serie así ajustada se expresa matemáticamente como

$$\frac{Y_t}{S_t} = \frac{T_t \times S_t \times I_t}{S_t} = T_t \times I_t$$

**Paso 6** El ciclo-tendencia se estima al aplicar un promedio móvil ponderado a la serie ajustada estacionalmente de forma preliminar. Este promedio móvil elimina la variación irregular y proporciona una curva suave que indica el ciclo-tendencia preliminar en los datos.

**Paso 7** Se repite el paso 2 con este nuevo estimado de ciclo-tendencia. Es decir, las nuevas proporciones, que contienen únicamente los componentes estacionales e irregulares, se obtienen al dividir las observaciones originales entre el ciclo-tendencia del paso 6. Éstas son las proporciones irregulares y estacionales que están al final;  
Matemáticamente,

$$\frac{Y_t}{T_t} = \frac{T_t \times S_t \times I_t}{T_t} = S_t \times I_t$$

**Paso 8** Se repite el paso 3 con las nuevas proporciones que se calcularon en el paso 7.

**Paso 9** Se repite el paso 4 para obtener un estimado nuevo del componente estacional.

**Paso 10** Se repite el paso 5 con el componente estacional del paso 9.

**Paso 11** Se dividen los datos ajustados estacionalmente obtenidos en el paso 10 entre el ciclo-tendencia obtenido en el paso 6 para obtener el componente irregular estimado.

**Paso 12** Los valores extremos del componente irregular se reemplazan como en el paso 3. Una serie de datos codificados se obtiene al multiplicar el componente ciclo-tendencia, el componente estacional y el componente irregular adaptados. Estos datos reproducen los datos originales, a excepción de los valores extremos.

Los 12 pasos anteriores se repiten, comenzando con los datos modificados desde el paso 12, en lugar de utilizar los datos originales. Algunas de las longitudes de los promedios móviles que se utilizan en los diversos pasos se cambian según sea la variabilidad de los datos.

La serie final ajustada estacionalmente se determina al dividir los datos originales entre el componente estacional final. El resultado contiene únicamente el producto del ciclo-tendencia y los componentes irregulares.

Los valores de cada componente final se imprimen y grafican. Existe una serie de pruebas de diagnóstico para determinar si la descomposición fue exitosa.

El programa ARIMA-X-12 contiene varias características adicionales que no se han descrito. Por ejemplo, se pueden hacer ajustes para los diversos números de días comerciales o para efectos de días feriados; también se pueden estimar y reemplazar valores faltantes dentro de las series; los efectos de los datos atípicos se pueden eliminar antes de la descomposición y se pueden modelar otros cambios en la tendencia, como modificaciones en un nivel o efectos de rampa temporales.

## **APLICACIÓN A LA ADMINISTRACIÓN**

---

El análisis de las series de tiempo es una herramienta estadística ampliamente usada para pronosticar acontecimientos futuros relacionados de alguna manera con la economía. Los fabricantes están extremadamente interesados en las fases de crecimiento-disminución del ciclo económico tanto nacional como de las economías extranjeras, de manera que puedan predecir mejor la demanda de sus productos, lo cual en consecuencia afecta los niveles de inventario, las necesidades de empleo, los flujos de efectivo y casi todas las actividades dentro de la empresa.

La complejidad de estos problemas es enorme. Considere, por ejemplo, el problema de predecir la demanda del petróleo y sus productos. A finales de la década de los sesenta del siglo XX, el precio del barril de petróleo era demasiado bajo y parecía haber una demanda mundial insaciable para el crudo y la gasolina. Después vino la crisis en los precios del petróleo durante la primera mitad de los setenta. ¿Cuál sería la demanda futura para el petróleo? ¿Qué sucedería con los precios? Empresas como Shell y General Motors estaban interesadas en estas preguntas. Si los precios del petróleo continuaban su aumento, ¿disminuiría la demanda de automóviles grandes? ¿Cuál sería la demanda de electricidad? En general, los analistas predijeron que la demanda de energía —por lo tanto, del petróleo— sería muy poco elástica; así, los precios continuarían superando a la inflación. No obstante, estas predicciones no consideraron una baja importante en el ciclo de los negocios a principios de los ochenta y una mayor elasticidad en la demanda de energía que la anunciada. Para 1980 el mundo comenzó a ver excedentes de petróleo en el mercado y la caída radical de los precios. Asimismo, parece difícil creer que los consumidores realmente se estaban beneficiando, otra vez, con la guerra de precios de la gasolina.

La demanda de petróleo se afecta no solamente por los sucesos cíclicos de largo plazo, sino también por los hechos estacionales y aleatorios, como sucede con otros pronósticos de la demanda para cualquier otro tipo de producto o servicio; por ejemplo, las industrias de servicios y de venta al menudeo. Hemos sido testigos del movimiento continuo del empleo de los campos de la manufactura a las áreas del menudeo y de los servicios. Así, puesto que la venta al menudeo (en una tienda, por catálogo, por Internet) es un negocio extremadamente estacional y cíclico, y la demanda y las proyecciones de inventario son fundamentales para los minoristas, estos últimos utilizan con mayor frecuencia el análisis cada vez más sofisticado de las series de tiempo.

Los fabricantes tendrán una continua necesidad de proyecciones estadísticas para sucesos futuros. Considere el crecimiento explosivo en la tecnología y los campos de telecomunicaciones durante los años noventa y la contracción sustancial de estas industrias en la primera década del nuevo siglo. Tales crecimiento y contracción fueron resultado, en gran medida, de proyecciones de la demanda que nunca se materializaron por completo. Las preguntas que todos los fabricantes deben hacerse incluyen: ¿Cuál será la tasa de inflación en el futuro? ¿Cómo se verán afectados los ajustes en el costo de la vida que se incluyen en los contratos laborales? ¿Cómo afectarán estos ajustes en los precios y la demanda? ¿Cuáles serán las habilidades gerenciales requeridas para 2020? ¿Cuál será el efecto de los gastos gubernamentales y de las estrategias fiscales?

¿Cómo será la población de personas jóvenes en el futuro? ¿Cómo será la mezcla étnica? Estas cuestiones afectan a casi todos los segmentos de la economía. Los demógrafos

vigilan de cerca la tasa actual de fertilidad y utilizan cualquier técnica de pronósticos de series de tiempo para intentar proyectar las variables de la población. Errores muy pequeños podrían tener importantes repercusiones en todo, desde la producción de juguetes de bebés hasta la solidez financiera o del sistema de seguridad social. Es interesante ver cómo los demógrafos consideran los ciclos empresariales de largo plazo (20 años o más por cada ciclo) para predecir qué hará esta generación de mujeres en edad de procrear en lo concerniente a tener hijos. ¿Tendrán uno o dos hijos, como las familias de los años sesenta y setenta?, o ¿verán a tener dos o tres, como lo hicieron las generaciones anteriores? Estas decisiones determinarán la composición de la edad entre la población para los siguientes 50 o 75 años.

Los estudiosos de las ciencias políticas están interesados en el uso de las series de tiempo para estudiar los patrones cambiantes de los gastos gubernamentales para la defensa o para los programas de bienestar social. Obviamente, estas tendencias tendrán una gran trascendencia en el futuro de industrias enteras.

Finalmente, un microcosmos interesante de aplicaciones del análisis de las series de tiempo se ha presentado en los campos legales. Los abogados están haciendo un uso creciente de testigos expertos para que den testimonio del valor actual de los ingresos futuros de una persona o una empresa, los costos incurridos por causa de la pérdida del trabajo por discriminación o el efecto de una huelga ilegal en un mercado. Con frecuencia, estas cuestiones pueden responderse mejor mediante el uso juicioso de un análisis de series de tiempo.

La tecnología satelital e Internet han hecho casi instantáneas la transmisión y acumulación de información. La proliferación de computadoras personales, el software estadístico de uso fácil y el acceso a bases de datos han llevado el procesamiento de información al escritorio. La supervivencia de un negocio durante períodos de importante cambio competitivo requiere de una rápida toma de decisiones con base en los datos. El análisis de las series de tiempo y los pronósticos desempeñan una función importante en esos procesos de toma de decisiones.

## APÉNDICE: ÍNDICE DE PRECIOS

---

Diversas series empleadas en producción, ventas y otras situaciones económicas contienen datos que solamente están disponibles en valor monetario. Estos datos se ven afectados tanto por la cantidad física de bienes vendidos como por sus precios. A lo largo del tiempo, las grandes variaciones en la inflación y los precios pueden causar problemas de análisis. Por ejemplo, un aumento en el importe podría esconder una caída de las ventas unitarias ocasionada por precios inflados. Así, con frecuencia es necesario saber en qué medida el cambio en valores de dólares representa un cambio verdadero en la cantidad física, y cuánto se debe al cambio en el precio a causa de la inflación. Es deseable que en estos casos se expresen los importes en términos monetarios constantes.

El concepto de *poder de compra* es importante. El poder de compra actual de un dólar se define como

$$\text{Poder actual de compra de 1 dólar} = \frac{100}{\text{Índice de precios al consumidor}} \quad (5.10)$$

Así, si en noviembre de 2001 el índice de precios al consumidor alcanza 150 (si 1997 equivale a 100), el poder actual de compra del dólar del consumidor, en noviembre de 2001, es

$$\text{Poder actual de compra de 1 dólar} = \frac{100}{150} = .67$$

El dólar de 2001 compró únicamente dos terceras partes de los bienes y servicios que pudo haber comprado con un dólar del periodo base (1997).

Para expresar los valores de dólares en términos de *dólares constantes*, se usa la ecuación 5.11.

$$\text{Valor deflactado del dólar} = \text{Valor de dólar} \times \text{Poder actual de compra de 1 dólar} \quad (5.11)$$

Suponga que la venta de automóviles se incrementó de \$300 000 en el 2000 a \$350 000 en el 2001, mientras que el índice de precios de automóviles nuevos (1997 como base) creció de 135 a 155. Las ventas deflactadas para 2000 y 2001 serían

$$\text{Ventas deflactadas para 2000} = (\$300,000) \left( \frac{100}{135} \right) = \$222,222$$

$$\text{Ventas deflactadas para 2001} = (\$350,000) \left( \frac{100}{155} \right) = \$225,806$$

Observe que las ventas en dólares actuales tienen un incremento cuantificable de \$350 000 – \$300 000 = \$50 000. Sin embargo, las ventas deflactadas únicamente crecieron en \$225 806 – \$222 222 = \$3 584.

El propósito de los valores monetarios deflactados es eliminar el efecto del cambio en los precios. Este ajuste se conoce como *deflación del precio* o bien se le refiere como expresión de una serie en monedas constantes.

*La deflación de precios* es el proceso que expresa los términos en una serie en monedas constantes.

El proceso de deflación es relativamente simple. Para ajustar los precios a dólares constantes, se usa un número-índice calculado a partir de los precios de bienes de consumo a cuyos valores se aplicará la deflación. Por ejemplo, las ventas de una zapatería podrían deflacionarse por un índice de precios de zapatos, y no sólo por un índice general de precios al consumidor. Para los valores monetarios deflactados, que representan más de un tipo de bien de consumo, el analista deberá desarrollar un índice de precios al combinar los índices de precios adecuados en una amalgama apropiada.

#### Ejemplo 5.5

El señor Burnham desea estudiar el crecimiento a largo plazo de Burnham Furniture Store. Las tendencias de largo plazo de su negocio deberán ser evaluadas mediante el volumen físico de ventas. Si esta evaluación no se puede realizar, los cambios en los precios reflejados en el importe de las ventas no seguirán un patrón consistente y simplemente oscurecerán el patrón de crecimiento real. Si se usa el importe de ventas, es necesario dividir el importe real de las ventas entre el índice de precios apropiado a fin de obtener las ventas cuantificadas en valor monetario constante.

El índice de precios al consumidor (IPC) no es el adecuado para Burnham porque contiene elementos como rentas, alimentos y servicios personales que no se venden en la tienda; sin embargo, algunos componentes de este índice podrían ser apropiados. Burnham está consciente

**TABLA 5.5 Datos de ventas de Burnham Furniture, 1994-2001, para el ejemplo 5.5**

| Año  | Ventas de Burnham (miles) | Índice de precios para minoristas muebleros (1997 = 100) | Índice de precios para minoristas de aparatos electrodomésticos (1997 = 100) | Índice de precios <sup>a</sup> (1997 = 100) | Ventas deflactadas <sup>b</sup> (miles de dólares de 1997) |
|------|---------------------------|----------------------------------------------------------|------------------------------------------------------------------------------|---------------------------------------------|------------------------------------------------------------|
| 1994 | 42.1                      | 90.1                                                     | 94.6                                                                         | 91.45                                       | 46.0                                                       |
| 1995 | 47.2                      | 95.4                                                     | 97.2                                                                         | 95.94                                       | 49.2                                                       |
| 1996 | 48.4                      | 97.2                                                     | 98.4                                                                         | 97.56                                       | 49.6                                                       |
| 1997 | 50.6                      | 100.0                                                    | 100.0                                                                        | 100.00                                      | 50.6                                                       |
| 1998 | 55.2                      | 104.5                                                    | 101.1                                                                        | 103.48                                      | 53.3                                                       |
| 1999 | 57.9                      | 108.6                                                    | 103.2                                                                        | 106.98                                      | 54.1                                                       |
| 2000 | 59.8                      | 112.4                                                    | 104.3                                                                        | 109.97                                      | 54.4                                                       |
| 2001 | 60.7                      | 114.0                                                    | 105.6                                                                        | 111.48                                      | 54.4                                                       |

<sup>a</sup>Construidos para las mueblerías (peso 70%) y electrodomésticos (peso 30%).

<sup>b</sup>Ventas divididas entre el índice de precios, y el resultado multiplicado por 100.

de que 70% de las ventas son de muebles y 30% corresponden a aparatos electrodomésticos. Por lo tanto, él puede multiplicar el componente IPC de muebles vendidos al menudeo por 0.70, multiplicar el componente de los electrodomésticos por 0.30 y luego sumarlos para obtener un índice de precios combinado. La tabla 5.5 muestra este método en el que los cálculos para 1994 son

$$90.1(0.70) + 94.6(0.30) = 91.45$$

Las ventas están deflactadas para 1994 en términos del poder de compra de 1997, de manera que

$$\text{Ventas deflactadas para 1994} = (42.1) \left( \frac{100}{91.45} \right) = 46.0$$

La tabla 5.5 muestra que aunque las ventas reales fueron mayores en forma constante de 1992 en 1999, el volumen físico de ventas permaneció estable de 1999 a 2001. Es evidente que el incremento en las ventas de 1999 a 2001 se debió a aumentos en los precios ocasionados a su vez por la tendencia inflacionaria de la economía.

## Glosario

|                                                                                                                                                        |                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| <b>Deflación de precios.</b> Proceso mediante el cual se expresan los términos de una serie en dólares constantes.                                     | <b>Números índice.</b> Porcentajes que muestran cambios a través del tiempo. |
| <b>Indicadores de negocios.</b> Series de tiempo relacionadas con los negocios que se utilizan para ayudar a evaluar el estado general de la economía. |                                                                              |

## Fórmulas clave

### Descomposición aditiva de las series de tiempo

$$Y_t = T_t + S_t + I_t \quad (5.1)$$

### Descomposición multiplicativa de las series de tiempo

$$Y_t = T_t \times S_t \times I_t \quad (5.2)$$

### Tendencia lineal

$$\hat{T}_t = b_0 + b_1 t \quad (5.3)$$

### Tendencia cuadrática

$$\hat{T}_t = b_0 + b_1 t + b_2 t^2 \quad (5.5)$$

### Tendencia exponencial

$$\hat{T}_t = b_0 b_1^t \quad (5.6)$$

### Datos ajustados estacionalmente (descomposición multiplicativa)

$$\frac{Y_t}{S_t} = T_t \times I_t \quad (5.7)$$

### Componente irregular-cíclico (descomposición multiplicativa)

$$C_t \times I_t = \frac{Y_t}{T_t \times S_t} \quad (5.8)$$

### Componente irregular (descomposición multiplicativa)

$$I_t = \frac{C_t \times I_t}{C_t} \quad (5.9)$$

**Poder actual de compra de 1 dólar**

$$\frac{100}{\text{Índice de Precios al Consumidor}} \quad (5.10)$$

**Valor de dólar deflactado**

$$\text{Valor del dólar} \times \text{Poder de compra de 1 dólar} \quad (5.11)$$

**Problemas**

1. Explique el concepto de descomponer una serie de tiempo.
2. Explique cuándo una descomposición multiplicativa puede ser más apropiada que una descomposición aditiva.
3. ¿Cuáles son algunas fuerzas básicas que afectan el ciclo de tendencia de la mayoría de las variables?
4. ¿Qué tipo de modelo de tendencia deberá utilizarse para los siguientes casos?
  - a) La variable se incrementa en una tasa constante.
  - b) La variable se incrementa en una tasa constante hasta que alcanza la saturación y entonces se nivela.
  - c) La variable se incrementa por una cantidad constante.
5. ¿Cuáles son algunas de las fuerzas básicas que afectan al componente estacional de la mayoría de las variables?
6. Value Line estima que las ventas y crecimiento de las ganancias para las compañías individuales se deriva de las correlaciones de ventas, utilidades y dividendos con los componentes apropiados de las cuentas nacionales tanto de ingreso como del gasto de capital. Jason Black, analista de Value Line, está examinando la tendencia del gasto de capital variable de 1977 a 1993. Los datos se proporcionan en la tabla P.6.
  - a) Grafique los datos.
  - b) Determine el modelo de tendencia adecuado para el periodo de 1977 a 1993.
  - c) Si el modelo apropiado es lineal, calcule el modelo de tendencia lineal para los años de 1977 a 1993.
  - d) ¿Cuál ha sido el incremento promedio en gastos de capital por año desde 1977?
  - e) Estime el valor de la tendencia para los gastos de capital en 1994.
  - f) Compare su estimado de tendencia con el obtenido por Value Line.
  - g) ¿Qué factores inciden en la tendencia de gastos de capital?
7. Una gran compañía está considerando recortar su publicidad en televisión y a cambio ofrecer videos de negocios a sus clientes. Se está analizando esta acción luego de que

**TABLA P.6 Gastos de capital (miles de millones de dólares), 1977–1993**

| <i>Año</i> | <i>Miles de millones de dólares</i> | <i>Año</i> | <i>Miles de millones de dólares</i> | <i>Año</i> | <i>Miles de millones de dólares</i> |
|------------|-------------------------------------|------------|-------------------------------------|------------|-------------------------------------|
| 1977       | 214                                 | 1983       | 357                                 | 1989       | 571                                 |
| 1978       | 259                                 | 1984       | 416                                 | 1990       | 578                                 |
| 1979       | 303                                 | 1985       | 443                                 | 1991       | 556                                 |
| 1980       | 323                                 | 1986       | 437                                 | 1992       | 566                                 |
| 1981       | 369                                 | 1987       | 443                                 | 1993       | 623                                 |
| 1982       | 367                                 | 1988       | 545                                 | 1994       | 680 <sup>a</sup>                    |

<sup>a</sup>Estimado de Value Line para 1994.Fuente: *The Value Line Investment Survey* (Nueva York, Value Line, 1988, 1990, 1994), pág. 175.

TABLA P.7

| Año  | Y      | Año  | Y      |
|------|--------|------|--------|
| 1980 | 11,424 | 1989 | 26,891 |
| 1981 | 12,811 | 1990 | 29,073 |
| 1982 | 14,566 | 1991 | 28,189 |
| 1983 | 16,542 | 1992 | 30,450 |
| 1984 | 19,670 | 1993 | 31,698 |
| 1985 | 20,770 | 1994 | 35,435 |
| 1986 | 22,585 | 1995 | 37,828 |
| 1987 | 23,904 | 1996 | 42,484 |
| 1988 | 25,686 | 1997 | 44,580 |

Fuente: *Statistical Abstract of the United States*, años diversos.

el presidente de la empresa leyó un artículo reciente en la prensa popular que aclamaba los videos de negocios como el “arma más moderna de ventas”. Antes de realizar esta acción, el presidente de la firma desea investigar la historia de la publicidad televisiva de su país, especialmente el ciclo de tendencia.

La tabla P.7 contiene el importe total gastado en la publicidad televisiva estadounidense, en millones de dólares.

- a) Haga una gráfica de la serie de tiempo del gasto en publicidad televisiva en Estados Unidos.
- b) Ajuste una tendencia lineal a los datos de la publicidad y grafique la línea ajustada sobre la gráfica de la serie de tiempo.
- c) Pronostique el importe de la publicidad televisiva para 1998 en dólares.
- d) Dados los resultados en el inciso b), ¿cree que podría haber un componente cíclico en el importe de la publicidad televisiva? Explique su respuesta.
- 8. Suponga que el siguiente conjunto específico de índices estacionales corresponden a marzo, está dado como porcentajes y se obtuvo por el método de la proporción del promedio móvil:

102.2 105.9 114.3 122.4 109.8 98.9

¿Cuál es el índice estacional para marzo al usar la mediana?

- 9. El valor de tendencia esperado para octubre es de \$850. Si se supone un índice estacional para octubre de 1.12 (112%) y el modelo multiplicativo proporcionado por la ecuación 5.2, ¿cuál sería el pronóstico para octubre?
- 10. Los siguientes porcentajes específicos para los índices estacionales corresponden a diciembre:

75.4 86.8 96.9 72.6 80.0 85.4

Suponga un modelo de descomposición multiplicativa. Si la tendencia esperada para diciembre es de \$900 y se usa el ajuste estacional de la mediana, ¿cuál es el pronóstico para diciembre?

- 11. Un gran centro vacacional situado cerca de Portland, Maine, ha registrado sus ventas mensuales por varios años, pero nunca ha analizado estos datos. El centro turístico calcula los índices estacionales de sus ventas mensuales. De las siguientes afirmaciones acerca de los índices, ¿cuáles son correctas?
- a) La suma de los 12 números índice mensuales, expresados como porcentajes, deberá ser igual a 1200.

TABLA P.12

| Mes        | Ventas<br>(miles de<br>dólares) | Índice<br>estacional<br>ajustado (%) |
|------------|---------------------------------|--------------------------------------|
| Enero      | 125                             | 51                                   |
| Febrero    | 113                             | 50                                   |
| Marzo      | 189                             | 87                                   |
| Abril      | 201                             | 93                                   |
| Mayo       | 206                             | 95                                   |
| Junio      | 241                             | 99                                   |
| Julio      | 230                             | 96                                   |
| Agosto     | 245                             | 89                                   |
| Septiembre | 271                             | 103                                  |
| Octubre    | 291                             | 120                                  |
| Noviembre  | 320                             | 131                                  |
| Diciembre  | 419                             | 189                                  |

Fuente: registros de Kula Department Store.

- b) Un índice de 85 para mayo indica que las ventas son 15% inferiores que el promedio de las ventas mensuales.
- c) Un índice de 130 para enero indica que las ventas son 30% superiores al promedio de las ventas mensuales.
- d) El índice para cualquier mes deberá estar entre cero y 200.
- e) El promedio del índice porcentual para cada uno de los 12 meses deberá ser de 100.
12. Al preparar un informe para June Bancock, gerente de Kula Department Store, usted incluye las estadísticas de las ventas del último año (tabla P.12). Después de verlas, la señora Bancock dice: "Este informe confirma lo que le he estado diciendo: el negocio está mejorando cada vez más." ¿Es correcta esta afirmación? ¿Por qué sí o por qué no?
13. Los niveles de ventas trimestrales cuantificados en millones de dólares para Goodyear Tire se muestran en la tabla P.13. ¿Parece que hay algún efecto estacional significativo en estos niveles de ventas? Analice esta serie de tiempo para obtener los cuatro ín-

TABLA P.13

| Año  | Trimestre |      |                   |                   |
|------|-----------|------|-------------------|-------------------|
|      | 1         | 2    | 3                 | 4                 |
| 1985 | 2292      | 2450 | 2363              | 2477              |
| 1986 | 2063      | 2358 | 2316              | 2366              |
| 1987 | 2268      | 2533 | 2479              | 2625              |
| 1988 | 2616      | 2793 | 2656              | 2746              |
| 1989 | 2643      | 2811 | 2679              | 2736              |
| 1990 | 2692      | 2871 | 2900              | 2811              |
| 1991 | 2497      | 2792 | 2838              | 2780              |
| 1992 | 2778      | 3066 | 3213              | 2928              |
| 1993 | 2874      | 3000 | 2913              | 2916              |
| 1994 | 2910      | 3052 | 3116              | 3210              |
| 1995 | 3243      | 3351 | 3305              | 3267              |
| 1996 | 3246      | 3330 | 3340 <sup>a</sup> | 3300 <sup>a</sup> |

<sup>a</sup>Estimados de Value Line para 1996.

Fuente: *The Value Line Investment Survey* (Nueva York, Value Line, 1988, 1989, 1993, 1994, 1996), p. 126.

**TABLA P.14**

| <b>Mes</b> | <b>1996</b> | <b>1997</b> | <b>1998</b> | <b>1999</b> | <b>2000</b> | <b>2001</b> | <b>2002</b> |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Enero      | 154         | 200         | 223         | 346         | 518         | 613         | 628         |
| Febrero    | 96          | 118         | 104         | 261         | 404         | 392         | 308         |
| Marzo      | 73          | 90          | 107         | 224         | 300         | 273         | 324         |
| Abril      | 49          | 79          | 85          | 141         | 210         | 322         | 248         |
| Mayo       | 36          | 78          | 75          | 148         | 196         | 189         | 272         |
| Junio      | 59          | 91          | 99          | 145         | 186         | 257         |             |
| Julio      | 95          | 167         | 135         | 223         | 247         | 324         |             |
| Agosto     | 169         | 169         | 211         | 272         | 343         | 404         |             |
| Septiembre | 210         | 289         | 335         | 445         | 464         | 677         |             |
| Octubre    | 278         | 347         | 460         | 560         | 680         | 858         |             |
| Noviembre  | 298         | 375         | 488         | 612         | 711         | 895         |             |
| Diciembre  | 245         | 203         | 326         | 467         | 610         | 664         |             |

dices estacionales, y determine la magnitud del componente estacional en las ventas de Goodyear.

- a) ¿Utilizaría los componentes de tendencia o estacional, o ambos para realizar el pronóstico?
  - b) Pronostique los trimestres tercero y cuarto de 1996.
  - c) Compare sus pronósticos con los realizados por Value Line.
14. Las ventas mensuales para la Cavanaugh Company, listadas en la tabla P.14 y graficadas en la figura 5.1, se presentan a continuación.
- a) Realice una descomposición multiplicativa de las series de tiempo de ventas de la Cavanaugh Company suponiendo los componentes de tendencia, estacional e irregular.
  - b) ¿Utilizaría los componentes de tendencia, estacionales, o ambos para realizar el pronóstico?
  - c) Proporcione los pronósticos por el resto de 2002.
15. Construya una tabla semejante a la tabla P.14 con los logaritmos naturales de las ventas mensuales. Por ejemplo, el valor para enero de 1996 es  $\ln(154) = 5.037$ .
- a) Desarrolle una descomposición aditiva de  $\ln(\text{ventas})$  suponiendo el modelo
- $$Y = T + S + I.$$
- b) ¿Utilizaría los componentes de tendencia, estacional, o ambos para realizar el pronóstico?
  - c) Proporcione los pronósticos para  $\ln(\text{ventas})$  para los meses restantes de 2002.
  - d) Tome los antilogaritmos de los pronósticos calculados en el inciso c) para obtener los pronósticos para las ventas reales del resto de 2002.
  - e) Compare los pronósticos del inciso d) con los del problema 14c). ¿Cuál conjunto de pronósticos prefiere? ¿Por qué?
16. La tabla P.16 contiene las ventas trimestrales (en millones de dólares) para la Disney Company, de enero de 1980 a marzo de 1995.
- a) Desarrolle una descomposición multiplicativa de las series de tiempo que consisten en las ventas trimestrales de Disney.
  - b) ¿Parece que hay alguna tendencia significativa? Analice la naturaleza del componente estacional.
  - c) ¿Utilizaría los componentes de tendencia y estacional para realizar el pronóstico?
  - d) Pronostique las ventas para el último trimestre de 1995 y para los cuatro trimestres de 1996.
17. La demanda mensual de gasolina (miles de barriles por día) para la Yukong Oil Company de Corea del Sur, correspondiente al periodo de enero de 1986 a septiembre de 1996, está contenida en la tabla P.17.

TABLA P.16

| Año  | Trimestre |        |        |        |
|------|-----------|--------|--------|--------|
|      | 1         | 2      | 3      | 4      |
| 1980 | 218.1     | 245.4  | 265.5  | 203.5  |
| 1981 | 235.1     | 258.0  | 308.4  | 211.8  |
| 1982 | 247.7     | 275.8  | 295.0  | 270.1  |
| 1983 | 315.7     | 358.5  | 363.0  | 302.2  |
| 1984 | 407.3     | 483.3  | 463.2  | 426.5  |
| 1985 | 451.5     | 546.9  | 590.4  | 504.2  |
| 1986 | 592.4     | 647.9  | 726.4  | 755.5  |
| 1987 | 766.4     | 819.4  | 630.1  | 734.6  |
| 1988 | 774.5     | 915.7  | 1013.4 | 1043.6 |
| 1989 | 1037.9    | 1167.6 | 1345.1 | 1288.2 |
| 1990 | 1303.8    | 1539.5 | 1712.2 | 1492.4 |
| 1991 | 1439.0    | 1511.6 | 1739.4 | 1936.6 |
| 1992 | 1655.1    | 1853.5 | 2079.1 | 2391.4 |
| 1993 | 2026.5    | 1936.8 | 2174.5 | 2727.3 |
| 1994 | 2275.8    | 2353.6 | 2698.4 | 3301.7 |
| 1995 | 2922.8    | 2764.0 | 3123.6 |        |

TABLA P.17

| Mes        | 1986 | 1987 | 1988 | 1989 | 1990 | 1991 | 1992  | 1993  | 1994  | 1995  | 1996  |
|------------|------|------|------|------|------|------|-------|-------|-------|-------|-------|
| Enero      | 15.5 | 20.4 | 26.9 | 36.0 | 52.1 | 64.4 | 82.3  | 102.7 | 122.2 | 145.8 | 170.0 |
| Febrero    | 17.8 | 20.8 | 29.4 | 39.0 | 53.1 | 68.1 | 83.6  | 102.2 | 121.4 | 144.4 | 176.3 |
| Marzo      | 18.1 | 22.2 | 29.9 | 42.2 | 56.5 | 68.5 | 85.5  | 104.7 | 125.6 | 145.2 | 174.2 |
| Abril      | 20.5 | 24.1 | 32.4 | 44.3 | 58.4 | 72.3 | 91.0  | 108.9 | 129.7 | 148.6 | 176.1 |
| Mayo       | 21.3 | 25.5 | 33.3 | 46.6 | 61.7 | 74.1 | 92.1  | 112.2 | 133.6 | 153.7 | 185.3 |
| Junio      | 19.8 | 25.9 | 34.5 | 46.1 | 61.0 | 77.6 | 95.8  | 109.7 | 137.5 | 157.9 | 182.7 |
| Julio      | 20.5 | 26.1 | 34.8 | 48.5 | 65.5 | 79.9 | 98.3  | 113.5 | 143.0 | 169.7 | 197.0 |
| Agosto     | 22.3 | 27.5 | 39.1 | 52.6 | 71.0 | 86.7 | 102.2 | 120.4 | 149.0 | 184.2 | 216.1 |
| Septiembre | 22.9 | 25.8 | 39.0 | 52.2 | 68.1 | 84.4 | 101.5 | 124.6 | 149.9 | 163.2 | 192.2 |
| Octubre    | 21.1 | 29.8 | 36.5 | 50.8 | 67.5 | 81.4 | 98.5  | 116.7 | 139.5 | 155.4 |       |
| Noviembre  | 22.0 | 27.4 | 37.5 | 51.9 | 68.8 | 85.1 | 101.1 | 120.6 | 147.7 | 168.9 |       |
| Diciembre  | 22.8 | 29.7 | 39.7 | 55.1 | 68.1 | 81.7 | 102.5 | 124.9 | 154.7 | 178.3 |       |

- a) Grafique la serie de tiempo para la demanda de gasolina. ¿Cree que sería adecuada una descomposición aditiva o una multiplicativa para esta serie de tiempo? Explique su respuesta.
- b) Desarrolle un análisis de descomposición para la demanda de gasolina.
- c) Interprete los índices estacionales.
- d) Pronostique la demanda de gasolina para los últimos tres meses de 1996.
18. La tabla P.18 contiene los datos que representan las ventas mensuales de todas las tiendas minoristas de Estados Unidos, en miles de millones de dólares. Utilice estos datos a lo largo de 1994 y desarrolle un análisis de descomposición para esta serie. Haga comentarios sobre los tres componentes de la serie. Pronostique las ventas al menudeo para 1995 y compare sus resultados con los valores reales que proporciona la tabla.
19. Los índices estacionales ajustados que se presentan en la tabla P.19 reflejan el volumen cambiante del negocio del Mt. Spokane Resort Hotel, el cual recibe a familias que vacacionan durante el verano y a esquiadores entusiastas durante los meses de invierno. No se esperan variaciones cíclicas repentinas durante 2003.

TABLA P.18

| Mes        | 1988  | 1989  | 1990  | 1991  | 1992  | 1993  | 1994  | 1995  |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Enero      | 113.6 | 122.5 | 132.6 | 130.9 | 142.1 | 148.4 | 154.6 | 167.0 |
| Febrero    | 115.0 | 118.9 | 127.3 | 128.6 | 143.1 | 145.0 | 155.8 | 164.0 |
| Marzo      | 131.6 | 141.3 | 148.3 | 149.3 | 154.7 | 164.6 | 184.2 | 192.1 |
| Abril      | 130.9 | 139.8 | 145.0 | 148.5 | 159.1 | 170.3 | 181.8 | 187.5 |
| Mayo       | 136.0 | 150.3 | 154.1 | 159.8 | 165.8 | 176.1 | 187.2 | 201.4 |
| Junio      | 137.5 | 149.0 | 153.5 | 153.9 | 164.6 | 175.7 | 190.1 | 202.6 |
| Julio      | 134.1 | 144.6 | 148.9 | 154.6 | 166.0 | 177.7 | 185.8 | 194.9 |
| Agosto     | 138.7 | 153.0 | 157.4 | 159.9 | 166.3 | 177.1 | 193.8 | 204.2 |
| Septiembre | 131.9 | 144.1 | 145.6 | 146.7 | 160.6 | 171.1 | 185.9 | 192.8 |
| Octubre    | 133.8 | 142.3 | 151.5 | 152.1 | 168.7 | 176.4 | 189.7 | 194.0 |
| Noviembre  | 140.2 | 148.8 | 156.1 | 155.6 | 167.2 | 180.9 | 194.7 | 202.4 |
| Diciembre  | 171.0 | 176.5 | 179.7 | 181.0 | 204.1 | 218.3 | 233.3 | 238.0 |

Fuente: Survey of Current Business, 1989, 1993, 1996.

TABLA P.19

| Mes     | Índice estacional ajustado | Mes        | Índice estacional ajustado |
|---------|----------------------------|------------|----------------------------|
| Enero   | 120                        | Julio      | 153                        |
| Febrero | 137                        | Agosto     | 151                        |
| Marzo   | 100                        | Septiembre | 95                         |
| Abril   | 33                         | Octubre    | 60                         |
| Mayo    | 47                         | Noviembre  | 82                         |
| Junio   | 125                        | Diciembre  | 97                         |

Fuente: registros del Mt. Spokane Resort Hotel.

TABLA P.24

|         | Volumen de ventas (dólares) | Índice de precios de productos primarios (1992 = 100) |
|---------|-----------------------------|-------------------------------------------------------|
| 1996    |                             |                                                       |
| Enero   | 358,235                     | 118.0                                                 |
| Febrero | 297,485                     | 118.4                                                 |
| Marzo   | 360,321                     | 118.7                                                 |
| Abril   | 378,904                     | 119.2                                                 |
| Mayo    | 394,472                     | 119.7                                                 |
| Junio   | 312,589                     | 119.6                                                 |
| Julio   | 401,345                     | 119.3                                                 |

Fuente: Survey of Current Business.

- a) Si 600 turistas estuvieran en el centro vacacional en enero de 2003, ¿cuál sería un estimado razonable para febrero?
  - b) La ecuación de tendencia mensual es  $\hat{T} = 140 + 5t$  donde  $t = 0$  representa el 15 de enero de 1997. ¿Cuál es el pronóstico para cada mes de 2003?
  - c) ¿Cuál es el número promedio de turistas nuevos por mes?
20. Analice el desempeño del índice compuesto de indicadores principales como barómetro de la actividad empresarial en los últimos años.

21. ¿Cuál es la posición actual del ciclo de negocios? ¿Se expande o se contrae? ¿Cuándo ocurrirá el siguiente punto de inflexión?
22. ¿Cuál es el propósito de la deflación en una serie de tiempo cuantificable en dólares?
23. En el periodo base de junio, el precio de una cantidad seleccionada de bienes fue de \$1 289.73. En el mes más reciente, el índice de precios para estos bienes fue de \$284.7. ¿Cuánto costarían los bienes seleccionados si fueran comprados en el mes más reciente?
24. Desarrolle una deflación en los importes de volúmenes de ventas mediante los valores del índice de precios de productos primarios, que se muestran en la tabla P.14. Estos índices son para todos los productos primarios con 1992 = 100.

## CASOS

---

### CASO 5-1

### SMALL ENGINE DOCTOR<sup>13</sup>

Small Engine Doctor es el nombre de una empresa fundada por Thomas Brown, quien fuera cartero del Servicio Postal estadounidense. Desde niño, había sido un reconstructor que solía tomar los aparatos del hogar en desuso a fin de entender “qué los hacía funcionar”. A medida que fue creciendo, Tom se transformó en un típico habitante de los suburbios y compró diversos equipos para jardinería. Cuando Tom descubrió que se ofrecía un curso de reparación de máquinas pequeñas que ofrecía un colegio comunitario, tomó con entusiasmo la oportunidad. Tom comenzó la reparación de máquinas pequeñas con el desmantelamiento de su propio equipo, lo mejoró y lo volvió a ensamblar. Pronto, después de haber terminado el curso, comenzó a reparar equipo para jardinería, barredoras de nieve y otro tipo de aparatos para sus amigos y vecinos. En el proceso, adquirió varios manuales de los equipos y herramientas especiales.

No pasó mucho tiempo antes de que Tom decidiera transformar su pasatiempo en un negocio de medio tiempo. Bajo el nombre de Small Engine Doctor, colocó un anuncio comercial en un folleto suburbano que se repartía en los centros comerciales. Durante los dos últimos años el negocio creció lo suficiente como para complementar de manera satisfactoria su salario normal. A pesar de que el crecimiento era bienvenido, poco antes de que su negocio entrara a su tercer año de operaciones, ya había algunas preocupaciones. El negocio se manejaba desde la casa de Tom. El sótano se dividió en sala familiar, taller y oficina. Al principio, el área de la oficina se utilizaba para administrar la publicidad, levantar los pedidos y realizar la contabilidad. Todas las reparaciones se hacían en el taller. La política de Tom

es almacenar un número limitado de partes y pedir partes de repuesto conforme se fuesen necesitando. Parecía que ésta era la única manera práctica de controlar la gran variedad de partes necesarias para la reparación de las máquinas fabricadas por una docena, o más, de productores de equipo para jardinería.

Se había comprobado que las refacciones eran el problema más fastidioso en la continuación del negocio. Al principio, Tom adquiría las refacciones con los distribuidores de equipo. Esta práctica ofrecía diversas desventajas. Primero, debía pagar precios al menudeo. En segundo lugar, casi siempre el distribuidor debía hacer un pedido secundario, de una o varias piezas, para alguna reparación. Las partes solicitadas al fabricante tardaban en llegar entre 30 y 120 días. Como resultado, Tom modificó su política y comenzó a pedir las partes directamente a la fábrica. Pronto descubrió que los fletes y los cargos por manejo reducían considerablemente sus utilidades, incluso cuando la refacción costara 60% del precio al menudeo. Sin embargo, los dos problemas más importantes ocasionados por las refacciones eran las ventas perdidas y el espacio para almacenarlas. Tom atraía clientes por la calidad de su servicio y sus precios razonables, que eran posibles debido a sus gastos generales mínimos. Desafortunadamente, muchos clientes potenciales preferían ir con los distribuidores en lugar de esperar varios meses por una reparación. El problema que ocasionaba mayor preocupación era el del almacenamiento. Mientras algún equipo esperaba a que llegaran las partes de repuesto, se le debía almacenar en el local. No pasó mucho tiempo para que su taller y su pequeña cochera se saturaran de equipo, mientras él

<sup>13</sup> Este caso es una contribución de William P. Darrow de la Towson State University, Towson, Maryland.

**TABLA 5.6**  
**Historial de ventas de Small Engine Doctor para el caso 5.1**

| <i>Mes</i> | 2002       |            | 2003       |            |
|------------|------------|------------|------------|------------|
|            | (unidades) | (unidades) | (unidades) | (unidades) |
| Enero      | 5          | 21         | Julio      | 28         |
| Febrero    | 8          | 20         | Agosto     | 20         |
| Marzo      | 10         | 29         | Septiembre | 14         |
| Abril      | 18         | 32         | Octubre    | 8          |
| Mayo       | 26         | 44         | Noviembre  | 6          |
| Junio      | 35         | 58         | Diciembre  | 26         |

esperaba las refacciones. En el segundo año de operación, Tom debió suspender la publicidad como una táctica para limitar la llegada de clientes por la falta de espacio para almacenar.

Tom ha considerado un inventario de existencias para su tercer año de operación. Esta práctica reducirá los costos de compra al hacer posible la obtención de descuentos por volumen y términos de envío más favorables. También espera que le proporcione un mejor tiempo de entrega para los clientes y, por lo tanto, el flujo de efectivo y las ventas mejoren. Los riesgos que se corren con esta estrategia son el descontrol de los costos por manejo de inventario y obsolescencia de partes.

Antes de comprometerse con el almacenamiento de partes, Tom desea tener un pronóstico confiable acerca de la actividad de su negocio para el año entrante. Él confía en su conocimiento de la mezcla de su producto para realizar un pronóstico agregado de las órdenes de reparación, como una base para los pedidos selectivos de refacciones. El pronóstico se complica por los patrones de demanda estacional y la tendencia creciente de las ventas.

Tom planea desarrollar un pronóstico de ventas para el tercer año de operación. En la tabla 5.6 se proporciona un historial de ventas. ■

## PREGUNTAS

1. Grafique los datos en un horizonte de dos años, de 2002 a 2003. Conecte los puntos de los datos para hacer el gráfico de la serie de tiempo.
2. Desarrolle una ecuación de línea de tendencia mediante la regresión lineal y grafique los resultados.
3. Estime los factores de ajuste estacional para cada mes, al dividir la demanda promedio para los meses correspondientes entre el promedio de los pronósticos de línea de tendencia que correspondan. Grafique una línea de tendencia ajustada a la estacionalidad.
4. Suavice la serie de tiempo mediante un suavizamiento exponencial ajustado a la tendencia con tres conjuntos de constantes de suavizamiento: ( $\alpha = 0.1$ ,  $\beta = 0.1$ ), ( $\alpha = 0.25$ ,  $\beta = 0.25$ ), y ( $\alpha = 0.5$ ,  $\beta = 0.5$ ). Grafique los tres conjuntos de valores suavizados en el gráfico de la serie de tiempo. Genere pronósticos para el tercer año con cada una de las posibilidades de ajustar la tendencia mediante suavizamiento exponencial.
5. Calcule los valores de *MAD* para los dos modelos que visualmente parezcan dar los mejores ajustes (pronósticos más precisos de un paso anticipado).
6. Si usted tuviera que limitar su selección a uno de los modelos de las preguntas 2 y 4, identifique el modelo que emplearía para la planeación de su negocio en 2004 y analice por qué seleccionó ese modelo sobre los demás.

## CASO 5-2 MR. TUX

John Mosby ha estado considerando la descomposición de sus series de tiempo mensuales, el importe mensual de las ventas. Él sabe que la serie tiene un efecto de estacionalidad muy grande y quisiera medirlo por dos razones. Primero, su banquero se rehusa a que haga pagos variables al préstamo que tiene. John

le ha explicado que debido a la estacionalidad de sus ventas y su flujo mensual de efectivo, quisiera hacer pagos adicionales en algunos meses y reducir, incluso hasta cero, los pagos en otros. Su banquero desea verificar lo que John dice acerca del efecto de la estacionalidad sobre sus ventas.

En segundo lugar, John quiere ser capaz de predecir sus ventas mensuales. Necesita dichos pronósticos para fines de planeación, en especial porque su negocio está creciendo. Tanto banqueros como inversionistas de capital de riesgo desean pronósticos sólidos en los cuales basar sus decisiones de inversión. John sabe que su negocio está mejorando y que sus perspectivas futuras se ven bien, pero los inversionistas quieren documentos.

El volumen de ventas mensuales de Mr. Tux para el periodo de 1990 a 1996 se introduce en Minitab. Ya que 1989 fue el primer año del negocio, los volúmenes de ventas fueron extremadamente bajos si se comparan con los alcanzados en los demás años. Por esta razón John decide eliminar estos valores de su análisis. Los índices estacionales se muestran en la tabla 5.7. Los demás datos obtenidos por computadora se muestran en la tabla 5.8.

John no está sorprendido de ver que los índices estacionales se presentan en estos datos, pero sí está satisfecho de tener cifras certeras para presentarlas a su banquero. Después de revisar estos números con él, se acuerda que John hará pagos dobles en abril, mayo, junio y agosto y que no pagará en enero, febrero, noviembre y diciembre. Su banquero solicita una copia de los índi-

ces estacionales para mostrársela a su jefe e incluirla en el expediente del préstamo de John.

Cuando es cuestión de pronosticar los primeros seis meses de 1997, John comienza por proyectar valores de tendencia mediante la ecuación de tendencia  $\hat{T}_t = 19,092.3 + 2,861.58t$ . La tendencia estimada para enero de 1997 es

$$\hat{T}_{85} = 19,092.3 + 2,861.58(85) = 262,326.6$$

A continuación, John obtiene el índice estacional de la tabla 5.7. El índice para enero es de 31.73%. De manera regular, John ha estado leyendo el *Wall Street Journal* y ve programas de análisis de negocios, de tal forma que ya tiene una idea sobre la naturaleza global de la economía y de su curso en el futuro. También es miembro de un club de servicios para negocios que participa en pláticas de economistas locales. Conforme John estudia la columna C de los datos que calculó, los cuales muestran la historia cíclica de su serie, piensa en cómo pronosticará este valor para los siguientes seis meses de 1997. Debido a que los pronósticos de los expertos nacionales y locales advierten una lenta mejoría en los negocios durante 1997, y ya que el último valor C

**TABLA 5.7 Resumen de los índices estacionales mensuales de Mr. Tux para el caso 5.2**

*Descomposición de la serie de tiempo*

Ecuación de la línea de tendencia

$$\hat{Y}_t = 19092.3 + 2861.58t$$

*Índice estacional*

| Periodo | Índice |
|---------|--------|
| 1       | 0.3173 |
| 2       | 0.4817 |
| 3       | 0.9013 |
| 4       | 1.8194 |
| 5       | 1.9409 |
| 6       | 1.1913 |
| 7       | 1.0229 |
| 8       | 1.2616 |
| 9       | 0.9016 |
| 10      | 0.7806 |
| 11      | 0.6035 |
| 12      | 0.7780 |

*Precisión del modelo*

|        |            |
|--------|------------|
| MAPE : | 19         |
| MAD :  | 20672      |
| MSD :  | 8.56E + 08 |

**TABLA 5.8 Cálculos de corto plazo para los componentes de Mr. Tux para el caso 5.2**

|    |      | <i>Año</i> | <i>Mes</i> | <i>Ventas</i> | <i>T</i> | <i>SCI</i> | <i>TCI</i> | <i>CI</i> | <i>C</i> | <i>I</i> |
|----|------|------------|------------|---------------|----------|------------|------------|-----------|----------|----------|
| 1  | 1990 | Enero      |            | 16,850        | 21,954   | 0.7675     | 53,106     | 2.4190    | —        | —        |
| 2  |      | Febrero    |            | 12,753        | 24,815   | 0.5139     | 26,476     | 1.0669    | —        | —        |
| 3  |      | Marzo      |            | 26,901        | 27,677   | 0.9720     | 29,848     | 1.0784    | 1.5904   | 0.6781   |
| 4  |      | Abril      |            | 61,494        | 30,539   | 2.0137     | 33,799     | 1.1068    | 1.3751   | 0.8049   |
| 5  |      | Mayo       |            | 147,862       | 33,400   | 4.4270     | 76,183     | 2.2809    | 1.4182   | 1.6084   |
| 6  |      | Junio      |            | 57,990        | 36,262   | 1.5992     | 48,676     | 1.3424    | 1.4049   | 0.9555   |
| 7  |      | Julio      |            | 51,318        | 39,123   | 1.3117     | 50,168     | 1.2823    | 1.2975   | 0.9883   |
| 8  |      | Agosto     |            | 53,599        | 41,985   | 1.2766     | 42,486     | 1.0120    | 1.0636   | 0.9514   |
| 9  |      | Septiembre |            | 23,038        | 44,846   | 0.5137     | 25,553     | 0.5698    | 0.9218   | 0.6181   |
| 10 |      | Octubre    |            | 41,396        | 47,708   | 0.8677     | 53,028     | 1.1115    | 0.7746   | 1.4350   |
| 11 |      | Noviembre  |            | 19,330        | 50,570   | 0.3823     | 32,032     | 0.6334    | 0.7446   | 0.8507   |
| 12 |      | Diciembre  |            | 22,707        | 53,431   | 0.4250     | 29,188     | 0.5463    | 0.8470   | 0.6449   |
| 13 | 1991 | Enero      |            | 15,395        | 56,293   | 0.2735     | 48,521     | 0.8619    | 0.7163   | 1.2034   |
| 14 |      | Febrero    |            | 30,826        | 59,154   | 0.5211     | 63,996     | 1.0819    | 0.7644   | 1.4153   |
| 15 |      | Marzo      |            | 25,589        | 62,016   | 0.4126     | 28,392     | 0.4578    | 0.9558   | 0.4790   |
| 16 |      | Abril      |            | 103,184       | 64,878   | 1.5904     | 56,712     | 0.8742    | 0.9465   | 0.9236   |
| 17 |      | Mayo       |            | 197,608       | 67,739   | 2.9172     | 10,181     | 1.5030    | 0.8363   | 1.7971   |
| 18 |      | Junio      |            | 68,600        | 70,601   | 0.9717     | 57,582     | 0.8156    | 0.9346   | 0.8727   |
| 19 |      | Julio      |            | 39,909        | 73,462   | 0.5433     | 39,015     | 0.5311    | 0.9244   | 0.5745   |
| 20 |      | Agosto     |            | 91,368        | 76,324   | 1.1971     | 72,425     | 0.9489    | 0.8102   | 1.1713   |
| 21 |      | Septiembre |            | 58,781        | 79,185   | 0.7423     | 65,198     | 0.8234    | 0.7776   | 1.0589   |
| 22 |      | Octubre    |            | 59,679        | 82,047   | 0.7274     | 76,449     | 0.9318    | 0.8287   | 1.1244   |
| 23 |      | Noviembre  |            | 33,443        | 84,909   | 0.3939     | 55,420     | 0.6527    | 0.8321   | 0.7844   |
| 24 |      | Diciembre  |            | 53,719        | 87,770   | 0.6120     | 69,051     | 0.7867    | 0.8302   | 0.9477   |
| 25 | 1992 | Enero      |            | 27,773        | 90,632   | 0.3064     | 87,532     | 0.9658    | 0.7616   | 1.2681   |
| 26 |      | Febrero    |            | 36,653        | 93,493   | 0.3920     | 76,093     | 0.8139    | 0.8721   | 0.9333   |
| 27 |      | Marzo      |            | 51,157        | 96,355   | 0.5309     | 56,761     | 0.5891    | 0.9229   | 0.6383   |
| 28 |      | Abril      |            | 217,509       | 99,216   | 2.1923     | 11,955     | 1.2049    | 0.9059   | 1.3301   |
| 29 |      | Mayo       |            | 206,229       | 102,078  | 2.0203     | 106,255    | 1.0409    | 0.9297   | 1.1196   |
| 30 |      | Junio      |            | 110,081       | 104,940  | 1.0490     | 92,401     | 0.8805    | 0.9965   | 0.8836   |
| 31 |      | Julio      |            | 102,893       | 107,801  | 0.9545     | 100,588    | 0.9331    | 0.9602   | 0.9717   |
| 32 |      | Agosto     |            | 128,857       | 110,663  | 1.1644     | 102,141    | 0.9230    | 0.9965   | 0.9263   |
| 33 |      | Septiembre |            | 104,776       | 113,524  | 0.9229     | 116,215    | 1.0237    | 0.9974   | 1.0263   |
| 34 |      | Octubre    |            | 111,036       | 116,386  | 0.9540     | 142,237    | 1.2221    | 0.9848   | 1.2409   |
| 35 |      | Noviembre  |            | 63,701        | 119,247  | 0.5342     | 105,561    | 0.8852    | 0.9587   | 0.9234   |
| 36 |      | Diciembre  |            | 82,657        | 122,109  | 0.6769     | 106,248    | 0.8701    | 0.9110   | 0.9551   |
| 37 | 1993 | Enero      |            | 31,416        | 124,971  | 0.2514     | 99,014     | 0.7923    | 0.8120   | 0.9758   |
| 38 |      | Febrero    |            | 48,341        | 127,832  | 0.3782     | 100,358    | 0.7851    | 0.8347   | 0.9406   |
| 39 |      | Marzo      |            | 85,651        | 130,694  | 0.6554     | 95,033     | 0.7272    | 0.8794   | 0.8269   |
| 40 |      | Abril      |            | 242,673       | 133,555  | 1.8170     | 133,379    | 0.9987    | 0.9190   | 1.0867   |
| 41 |      | Mayo       |            | 289,554       | 136,417  | 2.1226     | 149,186    | 1.0936    | 0.9829   | 1.1126   |
| 42 |      | Junio      |            | 164,373       | 139,279  | 1.1802     | 137,973    | 0.9906    | 1.0300   | 0.9617   |
| 43 |      | Julio      |            | 160,608       | 142,140  | 1.1299     | 157,010    | 1.1046    | 1.0439   | 1.0582   |
| 44 |      | Agosto     |            | 176,096       | 145,002  | 1.2144     | 139,586    | 0.9627    | 1.0205   | 0.9433   |
| 45 |      | Septiembre |            | 142,363       | 147,863  | 0.9628     | 157,905    | 1.0679    | 1.0674   | 1.0005   |
| 46 |      | Octubre    |            | 114,907       | 150,725  | 0.7624     | 147,196    | 0.9766    | 1.0552   | 0.9255   |
| 47 |      | Noviembre  |            | 113,552       | 153,586  | 0.7393     | 188,171    | 1.2252    | 1.0669   | 1.1484   |
| 48 |      | Diciembre  |            | 127,042       | 156,448  | 0.8120     | 163,300    | 1.0438    | 1.0591   | 0.9856   |

**TABLA 5.8** *(Continúa)*

|    |      | <i>Año</i> | <i>Mes</i> | <i>Ventas</i> | <i>T</i> | <i>SCI</i> | <i>TCI</i> | <i>CI</i> | <i>C</i> | <i>I</i> |
|----|------|------------|------------|---------------|----------|------------|------------|-----------|----------|----------|
| 49 | 1994 | Enero      |            | 51,604        | 159,310  | 0.3239     | 162,641    | 1.0209    | 1.1447   | 0.8919   |
| 50 |      | Febrero    |            | 80,366        | 162,171  | 0.4956     | 166,844    | 1.0288    | 1.0724   | 0.9594   |
| 51 |      | Marzo      |            | 208,938       | 165,033  | 1.2660     | 231,826    | 1.4047    | 1.0158   | 1.3828   |
| 52 |      | Abril      |            | 263,830       | 167,894  | 1.5714     | 145,007    | 0.8637    | 1.0240   | 0.8435   |
| 53 |      | Mayo       |            | 252,216       | 170,756  | 1.4771     | 129,949    | 0.7610    | 0.9834   | 0.7739   |
| 54 |      | Junio      |            | 219,566       | 173,617  | 1.2647     | 184,302    | 1.0615    | 0.8915   | 1.1908   |
| 55 |      | Julio      |            | 149,082       | 176,479  | 0.8448     | 145,742    | 0.8258    | 0.9366   | 0.8817   |
| 56 |      | Agosto     |            | 213,888       | 179,341  | 1.1926     | 169,543    | 0.9454    | 0.9694   | 0.9752   |
| 57 |      | Septiembre |            | 178,947       | 182,202  | 0.9821     | 198,483    | 1.0894    | 0.9634   | 1.1308   |
| 58 |      | Octubre    |            | 133,650       | 185,064  | 0.7222     | 171,206    | 0.9251    | 1.0194   | 0.9075   |
| 59 |      | Noviembre  |            | 116,946       | 187,925  | 0.6223     | 193,796    | 1.0312    | 1.0219   | 1.0092   |
| 60 |      | Diciembre  |            | 164,154       | 190,787  | 0.8604     | 211,004    | 1.1060    | 0.9781   | 1.1308   |
| 61 | 1995 | Enero      |            | 58,843        | 193,649  | 0.3039     | 185,456    | 0.9577    | 1.0433   | 0.9180   |
| 62 |      | Febrero    |            | 82,386        | 196,510  | 0.4193     | 171,037    | 0.8704    | 1.0296   | 0.8454   |
| 63 |      | Marzo      |            | 224,803       | 199,372  | 1.1276     | 249,429    | 1.2511    | 0.9733   | 1.2853   |
| 64 |      | Abril      |            | 354,301       | 202,233  | 1.7519     | 194,732    | 0.9629    | 1.0350   | 0.9304   |
| 65 |      | Mayo       |            | 328,263       | 205,095  | 1.6005     | 169,130    | 0.8246    | 1.0599   | 0.7780   |
| 66 |      | Junio      |            | 313,647       | 207,956  | 1.5082     | 263,273    | 1.2660    | 1.0599   | 1.1945   |
| 67 |      | Julio      |            | 214,561       | 210,818  | 1.0178     | 209,754    | 0.9950    | 1.0553   | 0.9429   |
| 68 |      | Agosto     |            | 337,192       | 213,680  | 1.5780     | 267,283    | 1.2509    | 1.0592   | 1.1810   |
| 69 |      | Septiembre |            | 183,482       | 216,541  | 0.8473     | 203,513    | 0.9398    | 1.0144   | 0.9265   |
| 70 |      | Octubre    |            | 144,618       | 219,403  | 0.6591     | 185,256    | 0.8444    | 1.0261   | 0.8229   |
| 71 |      | Noviembre  |            | 139,750       | 222,264  | 0.6288     | 231,585    | 1.0419    | 0.9724   | 1.0715   |
| 72 |      | Diciembre  |            | 184,546       | 225,126  | 0.8198     | 237,216    | 1.0537    | 1.0595   | 0.9946   |
| 73 | 1996 | Enero      |            | 71,043        | 227,987  | 0.3116     | 223,907    | 0.9821    | 1.1285   | 0.8703   |
| 74 |      | Febrero    |            | 152,930       | 230,849  | 0.6625     | 317,490    | 1.3753    | 1.1104   | 1.2385   |
| 75 |      | Marzo      |            | 250,559       | 233,711  | 1.0721     | 278,006    | 1.1895    | 1.0696   | 1.1121   |
| 76 |      | Abril      |            | 409,567       | 236,572  | 1.7313     | 225,108    | 0.9515    | 1.0622   | 0.8958   |
| 77 |      | Mayo       |            | 394,747       | 239,434  | 1.6487     | 203,385    | 0.8494    | 0.9708   | 0.8750   |
| 78 |      | Junio      |            | 272,874       | 242,295  | 1.1262     | 229,048    | 0.9453    | 0.9729   | 0.9717   |
| 79 |      | Julio      |            | 230,303       | 245,157  | 0.9394     | 225,144    | 0.9184    | 0.9554   | 0.9613   |
| 80 |      | Agosto     |            | 375,402       | 248,018  | 1.5136     | 297,571    | 1.1998    | 0.9607   | 1.2489   |
| 81 |      | Septiembre |            | 195,409       | 250,880  | 0.7789     | 216,742    | 0.8639    | 1.0063   | 0.8585   |
| 82 |      | Octubre    |            | 173,518       | 253,742  | 0.6838     | 222,276    | 0.8760    | 1.0790   | 0.8119   |
| 83 |      | Noviembre  |            | 181,702       | 256,603  | 0.7081     | 301,105    | 1.1734    | —        | —        |
| 84 |      | Diciembre  |            | 258,713       | 259,465  | 0.9971     | 332,551    | 1.2817    | —        | —        |

de octubre de 1996 es elevado (107.9%), John decide utilizar los siguientes valores *C* para sus pronósticos:

|      |           |     |
|------|-----------|-----|
| 1996 | Noviembre | 108 |
|      | Diciembre | 109 |
| 1997 | Enero     | 109 |
|      | Febrero   | 110 |
|      | Marzo     | 111 |
|      | Abril     | 113 |
|      | Mayo      | 114 |
|      | Junio     | 115 |

Al cambiar el valor irregular (*I*) para estos meses, John no prevé hechos inusuales, a excepción de marzo

de 1997. En ese mes planea tener una venta especial con precios reducidos en una de sus tiendas, la cual está terminando de remodelar. Debido a esta promoción, la cual estará acompañada por publicidad en radio y televisión, espera que las ventas sean 50% más altas que lo normal. Por sus ventas mensuales globales, espera tener un efecto que resulte en un incremento general de las ventas de 15%.

Mediante el uso de todas las cifras que ha estimado, junto con los datos obtenidos de su computadora, John hace los pronósticos de las ventas de Mr. Tux para los primeros seis meses de 1997, como se muestra en la tabla 5.9.

**TABLA 5.9 Pronósticos de Mr. Tux para el caso 5.2**

|         | <i>Pronóstico<br/>de ventas</i> | = | <i>T</i>  | × | <i>S</i> | × | <i>C</i> | × | <i>I</i> |
|---------|---------------------------------|---|-----------|---|----------|---|----------|---|----------|
| Enero   | 90,727                          | = | 262,326.6 | . | .3173    | . | 1.09     | . | 1.00     |
| Febrero | 140,515                         | = | 265,188.2 | . | .4817    | . | 1.10     | . | 1.00     |
| Marzo   | 308,393                         | = | 268,049.8 | . | .9013    | . | 1.11     | . | 1.15     |
| Abril   | 556,972                         | = | 270,911.4 | . | 1.8194   | . | 1.13     | . | 1.00     |
| Mayo    | 605,757                         | = | 273,773.0 | . | 1.9409   | . | 1.14     | . | 1.00     |
| Junio   | 378,988                         | = | 276,634.5 | . | 1.1913   | . | 1.15     | . | 1.00     |

Después de haber estudiado los pronósticos de 1997, John se preocupa al ver la gran variedad de proyecciones de ventas mensuales: de \$90 727 a \$605 757. A pesar de que sabía que su volumen variaba significativamente, se inquieta por esas fluctuaciones tan amplias. John ha estado pensando en expandir su área de influencia desde Spokane hacia el área de Seattle. Recientemente ha descubierto que ahí existen diversos eventos que requieren de vestimenta formal, distintos a los de su mercado en Spokane. Los bailes formales de regreso a casa son muy populares en Seattle, pero no en Spokane. Ya que estos bailes se realizan durante el otoño, cuando su negocio es más lento en Spokane (vea los índices esta-

cionales para octubre y noviembre), él cree que será una ventaja balancear su negocio al entrar al mercado de Seattle.

No obstante, su mayor preocupación se enfoca en los peores meses: enero y febrero. Ha estado pensando en comprar una máquina que fabrica camisas de esmoquin que vio en una feria comercial, y cree que puede concentrarse en eso durante el invierno. Si recibe una reacción positiva de los compradores potenciales de camisas para este periodo, estará dispuesto a intentarlo. Como están las cosas, los índices estacionales en su computadora han atraído su atención a los cambios extremos que se dan en sus niveles de ventas mensuales. ■

#### PREPREGUNTAS

- Suponga que el banquero de John hubiera solicitado dos afirmaciones, para presentarlas a su jefe, que pudieran justificar la solicitud de John para hacer pagos extraordinarios en algunos meses y ninguno en otros. Escriba estas dos oraciones.
- Suponga que el año entrante, el negocio de John en Seattle sea exactamente el doble del de Spokane.

Determine los índices estacionales de Seattle que serían los ideales para balancear los ingresos mensuales de Mr. Tux.

- Si no considerar a Seattle, ¿cuál sería el volumen que debería sacar John de la máquina de camisas para que enero y febrero fueran meses “promedio”?

## CASO 5-3 CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se describió en el capítulo 1 (caso 1.2). El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de clientes nuevos que se atenderían durante el resto de 1993. Marv proporcionó a Dorothy Mercer los datos men-

suales de los clientes nuevos de CCC durante el periodo de enero de 1985 a marzo de 1993 (vea el caso 3.3).

Dorothy le dio a usted estos datos y le pidió que completara un análisis de descomposición de la serie de tiempo. Ella enfatizó que quería comprender cabalmente los componentes de tendencia y estacionalidad. Dorothy

quería conocer la importancia de cada componente. También quería saber si había alguna irregularidad en los datos. En su instrucción final le solicitó que pronosticara el resto de 1993.

### Tarea

Escriba un informe que proporcione a Dorothy la información que ha solicitado. ■

## CASO 5-4 MURPHY BROTHERS FURNITURE

En el caso 4.4 Julie Murphy desarrolló un modelo informal que combinaba los estimados estacionales y de tendencia (semejante a la ecuación 4.5). Una de las razones más importantes por las que ella optó por un modelo informal fue su simplicidad. Julie sabía que su padre, Glen, necesitaría comprender el modelo de pronóstico usado por la compañía.

Es octubre de 2002 y muchas cosas han cambiado. Glen Murphy se ha retirado. Julie ha terminado varios cursos de negocios, como el de pronósticos empresariales, en la universidad local. Murphy Brothers Furniture construyó una planta en Dallas y comenzó a fabricar su propia línea de muebles en octubre de 1995.

Los datos de las ventas mensuales de Murphy Brothers Furniture, de 1996 a la fecha, se muestran en la tabla 5.10. Tal y como se indica en el patrón de estos datos, mostrados en la figura 5.13, las ventas se han incrementado en una forma espectacular desde 1996. Desafortunadamente, la figura 5.13 también muestra que uno de los problemas con la demanda es que es estacional. La política general de la compañía fue emplear dos turnos durante el verano y los primeros meses del otoño y después trabajar uno solo por el resto del año; por lo tanto, los inventarios crecieron de manera sustancial al final del verano y en los meses de otoño hasta que la

demandas comenzó a elevarse en noviembre y diciembre. Debido a estos requisitos de producción, Julie estaba muy preocupada por preparar pronósticos de corto plazo para la compañía que pudieran basarse en la mejor información disponible concerniente a la demanda.

Para fines de pronóstico, Julie ha decidido utilizar únicamente los datos recopilados desde que Murphy Brothers empezó a producir su propia línea de muebles en 1996 (tabla 5.10). Julie ve (figura 5.13) que sus datos tienen tendencia y estacionalidad. Por esta razón, decidió utilizar un método de descomposición de la serie de tiempo para analizar su variable de ventas.

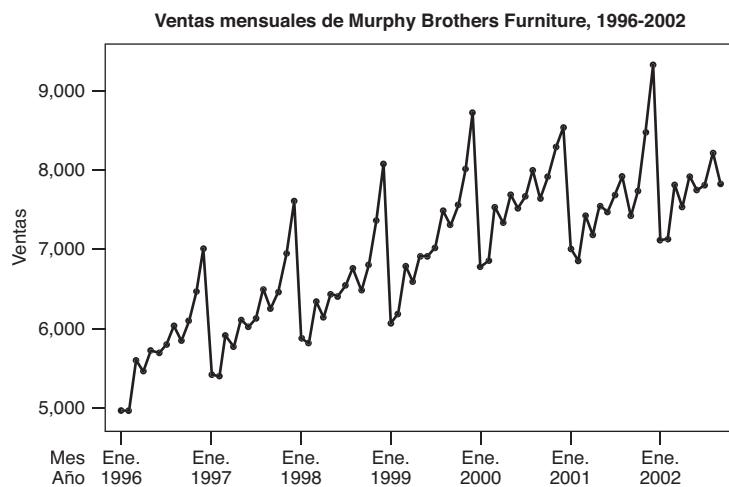
Puesto que la figura 5.13 muestra que la serie de tiempo que está analizando tiene aproximadamente la misma variabilidad a lo largo de la serie, Julie decide utilizar un modelo de componentes aditivos para pronosticar. Ella utiliza el modelo  $Y_t = T_t + S_t + I_t$ . En la tabla 5.11 se muestra un resumen de los resultados. Julie revisa el patrón de autocorrelación de los residuales (vea la figura 5.14) para observar la aleatoriedad. Los residuales no son aleatorios y parece que el modelo no es el adecuado.

Julie está desconcertada. Ha puesto a prueba un modelo informal que combinó estimados estacionales y de tendencia, el suavizamiento exponencial de Winters

**TABLA 5.10 Ventas mensuales de Murphy Brothers Furniture, 1996-2002, caso 5.4**

|      | <i>Ene.</i> | <i>Feb.</i> | <i>Mar.</i> | <i>Abr.</i> | <i>May.</i> | <i>Jun.</i> | <i>Jul.</i> | <i>Ago.</i> | <i>Sep.</i> | <i>Oct.</i> | <i>Nov.</i> | <i>Dic.</i> |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1996 | 4946        | 4968        | 5601        | 5454        | 5721        | 5690        | 5804        | 6040        | 5843        | 6087        | 6469        | 7002        |
| 1997 | 5416        | 5393        | 5907        | 5768        | 6107        | 6016        | 6131        | 6499        | 6249        | 6472        | 6946        | 7615        |
| 1998 | 5876        | 5818        | 6342        | 6143        | 6442        | 6407        | 6545        | 6758        | 6485        | 6805        | 7361        | 8079        |
| 1999 | 6061        | 6187        | 6792        | 6587        | 6918        | 6920        | 7030        | 7491        | 7305        | 7571        | 8013        | 8727        |
| 2000 | 6776        | 6847        | 7531        | 7333        | 7685        | 7518        | 7672        | 7992        | 7645        | 7923        | 8297        | 8537        |
| 2001 | 7005        | 6855        | 7420        | 7183        | 7554        | 7475        | 7687        | 7922        | 7426        | 7736        | 8483        | 9329        |
| 2002 | 7120        | 7124        | 7817        | 7538        | 7921        | 7757        | 7816        | 8208        | 7828        |             |             |             |

Registros de ventas de Murphy Brothers Furniture.



**FIGURA 5.13** Ventas mensuales de Murphy Brothers Furniture, 1996-2002

**TABLA 5.11** Resumen del modelo de descomposición para Murphy Brothers Furniture, caso 5.4

*Descomposición de la serie de tiempo*

Ecuación de la línea de tendencia

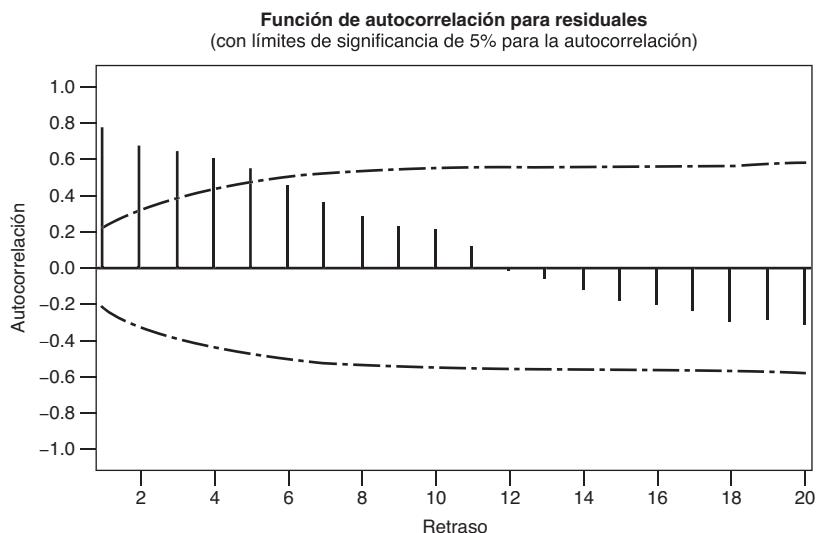
$$\hat{T}_t = 5604.8 + 32.45t$$

*Índice estacional*

| Periodo | Índice  |
|---------|---------|
| 1       | -674.60 |
| 2       | -702.56 |
| 3       | -143.72 |
| 4       | -366.64 |
| 5       | -53.52  |
| 6       | -173.27 |
| 7       | -42.74  |
| 8       | 222.32  |
| 9       | -57.95  |
| 10      | 145.76  |
| 11      | 612.30  |
| 12      | 1234.63 |

*Medidas de precisión*

|       |        |
|-------|--------|
| MAPE: | 1.8    |
| MAD:  | 131.6  |
| MSD:  | 3216.2 |



**FIGURA 5.14 Función de autocorrelación para los residuales utilizando un modelo de descomposición aditiva de series de tiempo**

**TABLA 5.12 Ventas mensuales ajustadas estacionalmente de Murphy Brothers Furniture, 1996-2002**

|      | Ene. | Feb. | Mar. | Abr. | May. | Jun. | Jul. | Ago. | Sep. | Oct. | Nov. | Dic. |
|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1996 | 5621 | 5671 | 5745 | 5821 | 5775 | 5863 | 5847 | 5818 | 5901 | 5941 | 5857 | 5767 |
| 1997 | 6091 | 6096 | 6051 | 6135 | 6161 | 6189 | 6741 | 6277 | 6307 | 6326 | 6334 | 6380 |
| 1998 | 6551 | 6521 | 6486 | 6510 | 6496 | 6580 | 6588 | 6536 | 6543 | 6659 | 6749 | 6844 |
| 1999 | 6736 | 6890 | 6936 | 6954 | 6972 | 7093 | 7073 | 7269 | 7363 | 7425 | 7401 | 7492 |
| 2000 | 7451 | 7550 | 7675 | 7700 | 7739 | 7691 | 7715 | 7770 | 7703 | 7777 | 7685 | 7302 |
| 2001 | 7680 | 7558 | 7564 | 7550 | 7608 | 7648 | 7730 | 7700 | 7484 | 7590 | 7871 | 8094 |
| 2002 | 7795 | 7827 | 7961 | 7905 | 7975 | 7930 | 7859 | 7986 | 7886 |      |      |      |

y la descomposición clásica. Finalmente, Julie decide ajustar la eliminación de la estacionalidad de los datos, de tal manera que se apliquen las técnicas de pronóstico que no pueden manejar los datos estacionales. Julie excluye la estacionalidad de los datos al añadir o restar el índice estacional de manera adecuada en cada mes. Por ejemplo, suma 674.60 a los datos de cada enero y resta 1 234.63 de cada diciembre. La tabla 5.12 muestra los datos ajustados estacionalmente.

### Tarea

- Desarrolle un modelo para pronosticar los datos de ventas ajustados a la estacionalidad.

- A pesar de que los datos estuvieron involucrados en la descomposición, al usar el origen de pronóstico de diciembre de 2001, pronostique las ventas para los primeros nueve meses de 2002. ¿Este pronóstico es preciso cuando se compara con los datos reales?
- Pronostique las ventas para octubre de 2002.
- Compare el patrón de los datos de ventas al menudeo que se presentan en el caso 3.1A con el patrón de las ventas reales de 1992 a 1995, que se presentaron en el caso 4.4 con el patrón de los datos reales de ventas de 1996 a 2001 que se presentaron en este caso. ■

## CASO 5-5 AAA WASHINGTON<sup>14</sup>

AAA Washington es uno de los dos clubes de automovilistas afiliados a la American Automobile Association (AAA o triple A) que opera en el estado de Washington. En 1993, 69% de las personas que pertenecían a clubes de automovilistas eran miembros de esta asociación, por lo que era el club de automovilistas más grande en Estados Unidos. AAA es una asociación nacional que sirve a sus afiliados mediante una federación de aproximadamente 150 clubes regionales que eligieron adherirse a esta asociación nacional. La organización establece ciertos estándares mínimos que los clubes afiliados deben observar para que puedan mantener su inscripción en el organismo. Cada club regional es controlado por su propio consejo de administración y fideicomisarios. La administración local y los fideicomisarios son responsables de reclutar y retener a los miembros dentro de los territorios asignados; además, deben asegurar la salud financiera del club regional. Más allá del cumplimiento con los estándares mínimos establecidos por AAA, cada club regional tiene la libertad de determinar qué productos y servicios adicionales ofrecerá y cómo les fijará los precios.

AAA Washington fue fundada en 1904. Su territorio de servicio abarca a los 26 condados del estado de Washington al occidente del río Columbia. El club ofrece a sus miembros una variedad de servicios relativos a los automóviles y viajes en vehículo. Los beneficios que reciben los miembros, en cooperación con la asociación nacional, incluyen servicios de emergencia en carretera; un servicio de hospedaje; restaurantes; talleres mecánicos; visitas guiadas a los centros de hospedaje aprobados por la AAA; centros para acampar y sitios de interés; además de asesoría legal y de gastos públicos empleados en pro de los automovilistas. Aparte de estos servicios, AAA Washington ofrece a sus miembros planes de protección adicional para el servicio de emergencia en carretera; servicios financieros, que incluyen afiliación a tarjetas de crédito, líneas de crédito personal, cuentas de cheques y de ahorros, depósitos, cheques de viajero American Express sin cuota alguna, acceso a la flotilla de camionetas de diagnóstico móvil para determinar la "salud" del vehículo, una agencia de viajes y una compañía de seguros. El club proporciona estos servicios a través de una red de oficinas ubicadas en Bellevue, Bellingham, Bremerton, Everett, Lynnwood, Olympia, Renton, Seattle, Tacoma, las Tres Ciudades (Pasco, Richland y Kennewick); Vancouver, Wenatchee y Yakima.

La investigación del club ha mostrado, de manera consistente, que el servicio de emergencia en carretera es la principal razón para afiliarse a la AAA. La importancia del servicio de emergencia en carretera para asegurar la permanencia de los integrantes se refleja en los tres tipos de membresías ofrecidas por AAA Washington: Básica, AAA Plus, y AAA Plus RV. La membresía Básica proporciona a los miembros cinco millas de remolque desde el punto donde su vehículo quedó inhabilitado. AAA Plus proporciona a sus miembros 100 millas de remolque desde el punto del percance. AAA plus RV proporciona el servicio de remolque de 100 millas a los miembros que poseen vehículos recreativos, además de los autos de pasajeros y camiones ligeros. La provisión del servicio de emergencia en carretera es el gasto más fuerte del club. Se proyecta que el suministro del servicio de emergencia en carretera costará aproximadamente 9.5 millones de dólares el próximo año fiscal, equivalentes a 37% del presupuesto de operación anual del club.

Michael DeCoria, un CP y graduado de MBA en la Eastern Washington University, recientemente ingresó al equipo de administración del club como vicepresidente de operaciones. Una de las responsabilidades que el señor DeCoria asumió fue la administración de los servicios de emergencia en carretera. En la primera fase de su evaluación del servicio de emergencia en carretera, el señor DeCoria descubrió que los costos de este servicio se habían incrementado a un ritmo mayor de lo que podría justificarse por la tasa de inflación y el crecimiento de la membresía del club. El señor DeCoria analizó la manera en que el club proporciona el servicio de emergencia para determinar si los costos podían controlarse un poco más en esta área.

El servicio de emergencia en carretera se proporciona en una de cuatro formas: por medio de la flotilla de servicios de AAA Washington, compañías contratistas, reembolso recíproco o reembolso directo. La flotilla de vehículos de AAA Washington responde a las llamadas de emergencia de miembros cuyos vehículos han quedado inhabilitados en el área de Seattle. Dentro del área de servicio de esta empresa, pero fuera del centro de Seattle, las compañías comerciales de servicio de grúa que han sido contratadas por la organización para proporcionar este servicio responden a las llamadas del servicio de emergencia. Los miembros aceptan que ambos tipos de servicio sean proporcionados mediante llamadas al centro de envíos del club.

<sup>14</sup> Este caso es una contribución de Steve Branton, exalumno y graduado de M.B.A. graduate, Eastern Washington University.

Si un miembro quedara inhabilitado fuera del área de atención de AAA Washington, puede llamar al club local afiliado para recibir el servicio de emergencia en carretera. El club regional paga por este servicio y después factura a AAA Washington para un reembolso recíproco a través de un servicio de compensación proporcionado por la asociación nacional. Finalmente, los miembros pueden contactar directamente a una compañía de grúas de su elección, pagar por el servicio de remolque y después solicitar un reembolso al club. AAA Washington reintegra directamente al miembro el costo real del servicio de grúa o 50 dólares, el que sea menor. Después de un cuidadoso examen de los cuatro tipos de servicios proporcionados por el club, el señor DeCoria concluyó que éste controlaba

el costo del servicio tan estrechamente como fuera práctico.

Otra posible fuente del incremento en los costos fue el alza en el servicio de emergencia en carretera. El número de las membresías había crecido constantemente por varios años, pero el aumento del costo era mayor de lo que podría atribuirse a un simple crecimiento de la membresía. Enseguida, el señor DeCoria revisó si había aumentado el uso del servicio de emergencia en carretera de manera individual entre sus miembros. Descubrió que entre el año fiscal de 1990 y el de 1991 el número promedio de llamadas de servicio de emergencia por miembro había crecido 3.28% de un promedio de 0.61 a 0.63 llamadas individuales (el año fiscal de

**TABLA 5.13 Volumen de llamadas para el servicio de emergencia en carretera para el caso 5.5**

| <i>Año</i> | <i>Mes</i> | <i>Llamadas</i> | <i>Año</i> | <i>Mes</i> | <i>Llamadas</i> |
|------------|------------|-----------------|------------|------------|-----------------|
| 1988       | Mayo       | 20,002          | 1991       | Enero      | 23,441          |
|            | Junio      | 21,591          |            | Febrero    | 19,205          |
|            | Julio      | 22,696          |            | Marzo      | 20,386          |
|            | Agosto     | 21,509          |            | Abril      | 19,988          |
|            | Septiembre | 22,123          |            | Mayo       | 19,077          |
|            | Octubre    | 21,449          |            | Junio      | 19,141          |
|            | Noviembre  | 23,475          |            | Julio      | 20,883          |
|            | Diciembre  | 23,529          |            | Agosto     | 20,709          |
|            | Enero      | 23,327          |            | Septiembre | 19,647          |
|            | Febrero    | 24,050          |            | Octubre    | 22,013          |
|            | Marzo      | 24,010          |            | Noviembre  | 22,375          |
|            | Abrial     | 19,735          |            | Diciembre  | 22,727          |
| 1989       | Mayo       | 20,153          | 1992       | Enero      | 22,367          |
|            | Junio      | 19,512          |            | Febrero    | 21,155          |
|            | Julio      | 19,892          |            | Marzo      | 21,209          |
|            | Agosto     | 20,326          |            | Abril      | 19,286          |
|            | Septiembre | 19,378          |            | Mayo       | 19,725          |
|            | Octubre    | 21,263          |            | Junio      | 20,276          |
|            | Noviembre  | 21,443          |            | Julio      | 20,795          |
|            | Diciembre  | 23,366          |            | Agosto     | 21,126          |
|            | Enero      | 23,836          |            | Septiembre | 20,251          |
|            | Febrero    | 23,336          |            | Octubre    | 22,069          |
|            | Marzo      | 22,003          |            | Noviembre  | 23,268          |
|            | Abrial     | 20,155          |            | Diciembre  | 26,039          |
| 1990       | Mayo       | 20,070          | 1993       | Enero      | 26,127          |
|            | Junio      | 19,588          |            | Febrero    | 20,067          |
|            | Julio      | 20,804          |            | Marzo      | 19,673          |
|            | Agosto     | 19,644          |            | Abril      | 19,142          |
|            | Septiembre | 17,424          |            |            |                 |
|            | Octubre    | 20,833          |            |            |                 |
|            | Noviembre  | 22,490          |            |            |                 |
|            | Diciembre  | 24,861          |            |            |                 |

AAA Washington comienza el 1 de julio). Preocupado porque la continuidad de esta tendencia tendrá una repercusión negativa en las finanzas del club, el señor

DeCoria ha recopilado los datos del volumen de llamadas del servicio de emergencia en carretera tal como se presentan en la tabla 5.13. ■

### PREGUNTAS

- Desarrolle una descomposición de serie de tiempo de los datos de llamadas del servicio de emergencia en carretera de AAA.
- Escriba un informe al señor DeCoria en el que resuma las consideraciones importantes sobre los

cambios en el volumen de llamadas del servicio de emergencia en carretera que ha descubierto, a partir de su propio análisis de descomposición de la serie de tiempo.

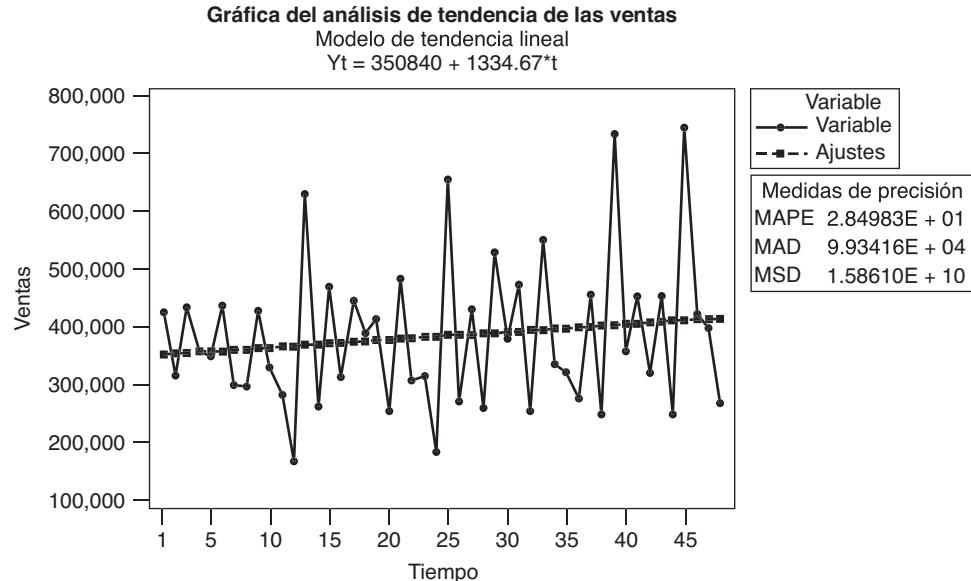
## CASO 5-6 ALOMEGA FOOD STORES

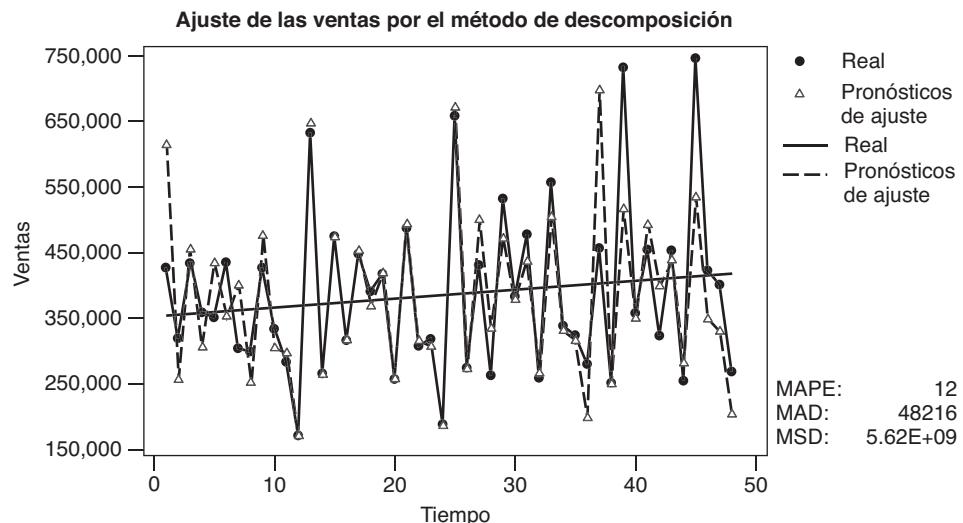
En el ejemplo 1.1, Julie Ruth, presidenta de Alomega Food Stores, había recopilado datos acerca de las ventas mensuales de su compañía junto con otras variables que ella pensó que podrían estar relacionadas con las ventas (revise el ejemplo 1.1 del capítulo 1). El caso 2.3 en el capítulo 2 explicó la forma en que Julie utilizó Minitab para calcular una ecuación de regresión simple utilizando la mejor forma de predecir las ventas mensuales (vea las páginas 52 y 53).

Después de haber revisado los resultados de este análisis de regresión, incluyendo el valor bajo de la  $r$  cuadrada (36%), ella decidió probar la descomposición de

series de tiempo con una sola variable, las ventas mensuales. La figura 5.15 muestra las gráficas que obtuvo para los datos de ventas (vea el caso 3.4). Pareciera que las ventas estaban demasiado dispersas alrededor de la línea de tendencia para poder elaborar pronósticos precisos. Esta impresión se confirmó cuando ella pudo ver el valor de MAPE de 28. Ella interpretó que el error porcentual promedio entre los valores reales y la línea de tendencia fue de 28%, un valor que consideró demasiado alto. A continuación intentó una descomposición multiplicativa de los datos. Los resultados se muestran en la figura 5.16.

**FIGURA 5.15 Análisis de tendencia de las ventas de Alomega Food Stores**





**FIGURA 5.16 Ajuste de descomposición de las ventas de Alomega Food Store**

Además de la ecuación de tendencia mostrada en el listado, Julie estaba interesada en los índices estacionales (mensuales) que el programa calculó. Observó que el mes de ventas más bajas fue diciembre (mes 12, índice = 0.49) y el de ventas más elevadas fue enero (mes 1, índice = 1.74). Estaba consciente de la gran variación entre diciembre y enero, pero no se había dado cuenta de cuán extremosa era.

También observó que *MAPE* había caído a 12%, una mejoría definitiva sobre valor obtenido mediante la ecuación de tendencia sola.

Finalmente, Julie hizo que el programa proporcionara pronósticos para los siguientes 12 meses median-

te las proyecciones de la ecuación de tendencia modificada por los índices estacionales. Pensó que podría utilizar éstos como pronósticos para sus propósitos de planeación, pero se preguntaba si otro método podría producir mejores pronósticos. También estaba preocupada por lo que diría su gerente de producción, Jackson Tilson, acerca de sus pronósticos, especialmente desde que él había expresado cierta preocupación por el uso de la computadora para realizar predicciones (vea sus comentarios al final del ejemplo 1.1 en la página 8). ■

#### PREGUNTA

1. ¿Qué podría decir Jackson Tilson acerca de los pronósticos de Julie?

#### Aplicaciones en Minitab

**El problema.** En el ejemplo 5.1 se desarrolló una ecuación de tendencia para el registro anual de vehículos de pasajeros en Estados Unidos, de 1960 a 1992.

#### Solución en Minitab

1. Después de haber introducido los datos del registro de vehículos para pasajeros en C1 de la hoja de trabajo, dé clic en los siguientes menús para ejecutar el análisis de la tendencia:

Stat>Time Series>Trend Analysis

2. Aparece el cuadro de diálogo del análisis de tendencia.
  - a) La *Variable* es Registros.
  - b) Dé clic en *Linear* (lineal) para seleccionar el tipo de modelo (*Model Type*).
  - c) Dé clic en *Generate forecasts* (generar pronósticos) y coloque un 1 en el cuadro *number of forecasts* (número de pronósticos) para pronosticar 1993.
  - d) Dé clic en *Options* (opciones). En el espacio proporcionado para un *Title* (título) coloque *Linear Trend for Car Registration Time Series* (tendencia lineal de la serie de tiempo de registro de automóviles).
  - e) Dé clic en OK en el cuadro de diálogo de *Options*. Dé clic en OK y aparecerá la gráfica que se mostró en la figura 5.3.

**El problema.** La tabla 5.1 se construyó para mostrar los estimados de tendencia y los errores calculados para los datos del registro de automóviles nuevos de pasajeros (vea la página 162).

#### **Solución en Minitab**

1. La columna C1 está etiquetada como *Year* (año), la C2 como Y, la C3 con t, la C4 con *Estimates* (estimados) y la C5 con *Error*. Al dar clic en los siguientes menús se crean los años:
 

```
Calc>Make Patterned Data>Simple Set of Numbers
```
2. Aparece el cuadro de diálogo de *Simple Set of Numbers* (serie simple de números).
  - a) Se proporcionan las siguientes respuestas:
 

Guardar los datos del patrón en: C1  
Del primer valor: 1960  
Al último valor: 1992  
En los pasos de: 1
  - b) Dé clic en OK y los años aparecen en la columna C1.
  - c) Los nuevos datos del registro de automóviles de pasajeros se introducen en la columna C2.
3. Los datos con código de tiempo *t* se introducen en C3 mediante el cuadro de diálogo de *Simple Set of Numbers*.
4. Los estimados de la tendencia se introducen en C4 al dar clic en los mismos menús del análisis de la tendencia utilizados para resolver el [ejemplo 5.1](#), con un paso adicional.
5. Dé clic en el menú de *Storage* (almacenamiento) para que salga el cuadro de diálogo de *Trend-Analysis Storage* (almacenamiento del análisis de la tendencia)
  - a) Bajo *Storage* dé clic en *Fits* (*trend line*) (ajustes a la línea de tendencia) y *Residuals* (*detrended data*) (residuales, datos sin tendencia).
  - b) Dé clic en OK en este cuadro de diálogo y después en el cuadro de diálogo de *Trend Analysis*. Los estimados de tendencia aparecerán en C4 y los errores (residuales) aparecerán en C5.

**El problema.** En los ejemplos 5.3 y 5.4 Perkin Kendell, el analista de Outboard Marine Corporation quería pronosticar las ventas trimestrales para 1997.

#### **Solución en Minitab**

1. Introduzca los años que correspondan en C1, los trimestres en C2 y los datos en C3. Para ejecutar un modelo de descomposición dé clic en el siguiente menú:
 

```
Stat>Time Series>Decomposition
```
2. El cuadro de diálogo *Decomposition* (descomposición) que se muestra en la figura 5.17 aparece.
  - a) La *Variable* es C3 o Sales (ventas).
  - b) Ya que los datos son trimestrales, la longitud estacional (*Seasonal length*) es 4.

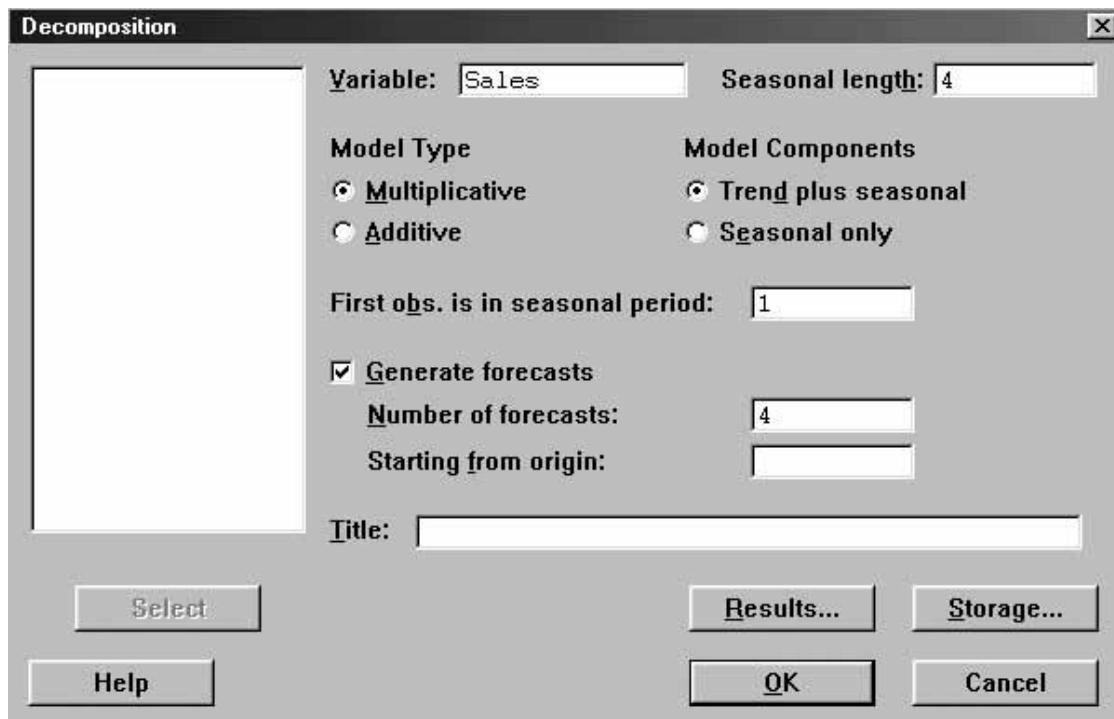


FIGURA 5.17 Cuadro de diálogo de descomposición en Minitab

- c) El tipo de modelo (*Model type*) es multiplicativo (*Multiplicative*) y los componentes del modelo (*Model components*) son de tendencia y estacional (*Trend plus seasonal*).  
d) Dé clic en *Options* (opciones). El periodo inicial es 1. Dé clic en **OK**.  
e) Dé clic en *Generate forecasts* (generar pronósticos) e introduzca 4 en *Number of forecasts* (número de pronósticos).
3. Dé clic en el menú de *Storage* (almacenamiento) para que aparezca el cuadro de diálogo de *Decomposition-Storage* (descomposición-almacenamiento) como se muestra en la figura 5.18.
- a) Bajo *Trend Line* (línea de tendencia), dé clic en *Trend line* (línea de tendencia), *Deseasonalized data* (datos sin estacionalidad), *Seasonals* (estacionales) y *Seasonally adjusted data* (datos ajustados estacionalmente).  
b) Dé clic en **OK** en ambos cuadros de diálogo. La tabla 5.4 muestra los estimados de tendencia en C4 (etiquetada como *T*), los datos sin tendencia en C5 (etiquetada como *SCI*), los estacionales en C6 (etiquetada como *S*) y los datos ajustados estacionalmente en C7 (etiquetada como *TCI*).
4. Las figuras 5.10 y 5.11 aparecen en la pantalla y pueden imprimirse una por una mediante los siguientes menús:  
**File>Print Graph**
5. Después de haber impreso las gráficas dé clic en  
**File>Print Session Window**  
y los pronósticos que se muestran en la figura 5.9 se imprimirán.  
Las columnas C1, C e I podrían calcularse con Minitab, pero eso se mostrará en la siguiente sección, “Aplicaciones en Excel”.

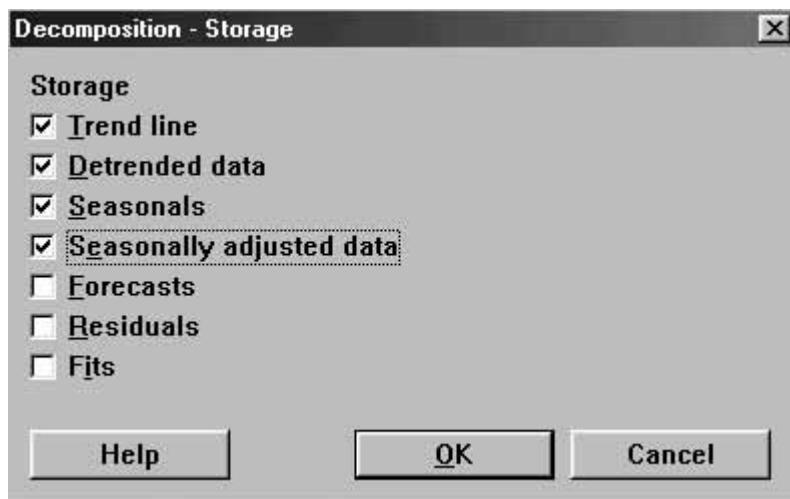


FIGURA 5.18 Cuadro de diálogo de descomposición en Minitab

### Aplicaciones en Excel

**El problema.** La figura 5.6 muestra los datos y la gráfica para los vendedores de fondos mutuos (vea la página 165). Se necesita un modelo de tendencia exponencial para ajustar esos datos.

#### Solución en Excel

1. Introduzca  $Y$  en la celda A1 y los datos de los vendedores en A2:A8.
2. Introduzca  $X$  en la celda B1 y la variable codificada  $X$  (1, 2, 3, 4, 5, 6, 7) en B2:B8.
3. Introduzca  $\log Y$  en C1 y haga los logaritmos (a la base 10) de la variable  $Y$  introduciendo la fórmula =LOG10(A2) en la celda C2. Copie esta fórmula en el resto de la columna.
4. Dé clic en los siguientes menús para calcular la tendencia exponencial  
Tools>Data Analysis
5. Aparece el cuadro de diálogo de *Data Analysis* (análisis de datos). Bajo *Analysis Tools* (herramientas de análisis) seleccione *Regression* (regresión) y dé clic en OK. Aparecerá el cuadro de diálogo de *Regression* tal y como se muestra en la figura 5.19.
  - a) Introduzca C1:C8 como el rango  $Y$  de entrada.
  - b) Introduzca B1:B8 como el rango  $X$  de entrada.
  - c) Seleccione el cuadro de verificación *Labels* (rótulos).
  - d) Introduzca *Figure 5.20* como el nombre de *New Worksheet Ply* (en una hoja nueva). Dé clic en OK.

La figura 5.10 representa los resultados de Excel para un modelo exponencial de los vendedores de fondos mutuos. La ecuación es:

$$\log \hat{T} = 1.00069 + 0.11834t$$

Los antilogaritmos de los coeficientes de regresión en esta ecuación son:

$$b_0 = \text{antilog } 1.000692 = 10.016$$

$$b_1 = \text{antilog } 0.118338 = 1.313$$

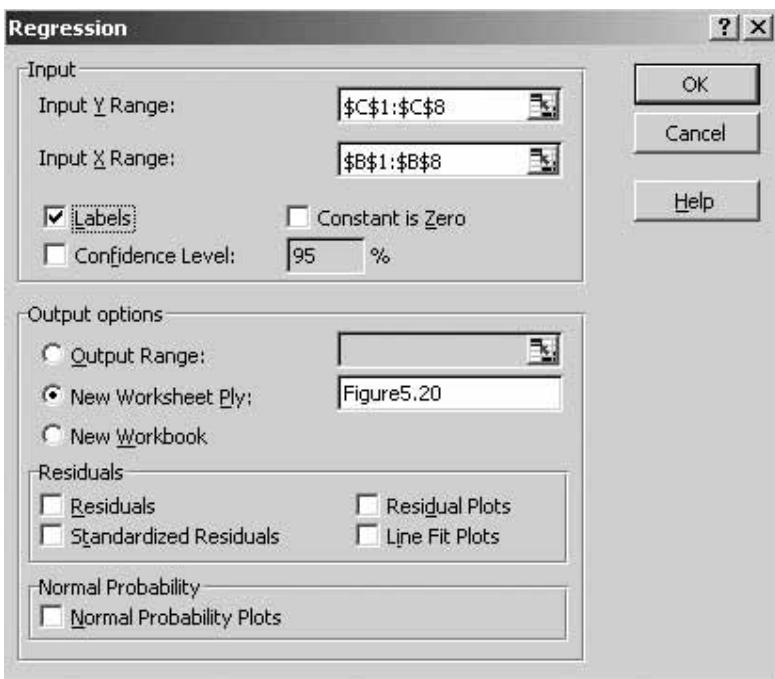


FIGURA 5.19 Cuadro de diálogo de regresión en Excel

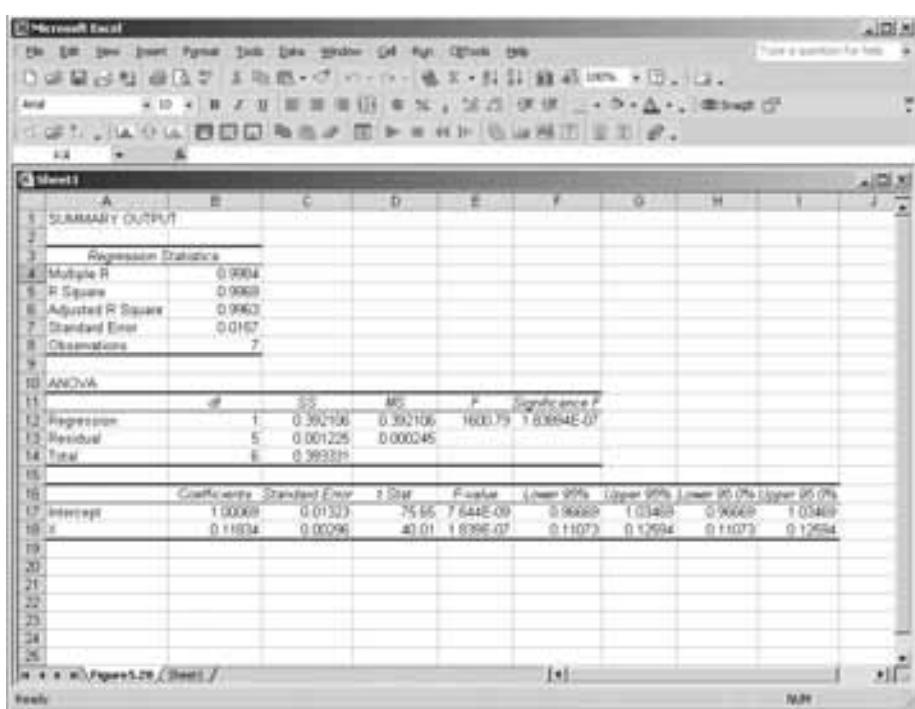


FIGURA 5.20 Resultados de la regresión en Excel para el ejemplo de los vendedores de fondos mutuos

De esta forma, la ecuación de la tendencia exponencial es

$$\hat{T} = (10.016)(1.313)^t$$

**El problema.** En la situación de Outboard Marine Corporation, en los ejemplos 5.3 y 5.4, la columna  $CI$ , la  $C$  (promedio móvil de tres períodos) y la  $I$  se calculan para la tabla 5.4.

#### Solución en Excel

1. Abra el archivo de Minitab que contiene los datos para la tabla 5.4 y resalte la columna  $T$ , C4. Dé clic en los siguientes menús:

Edit>Copy Cells

2. Abra su hoja de cálculo de Excel, resalte A1 y dé clic en los siguientes menús:

Edit>Paste

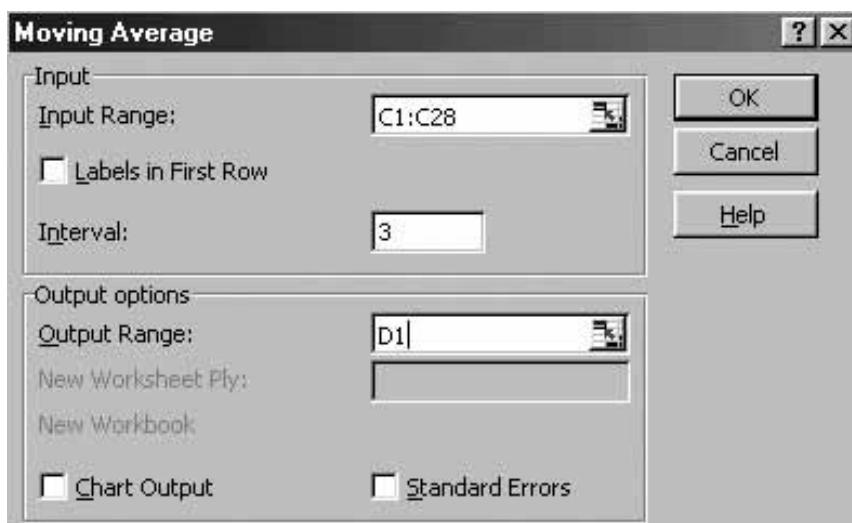
3. Los datos de la columna  $T$  aparecerán en la columna A. Repita el proceso al copiar la columna  $TCI$ , C7 en la B1 de la hoja de cálculo.
4. Para crear la columna  $CI$ , coloque el cursor en la columna C1 y escriba la fórmula  $=B1/A1$ . Copie esta fórmula en el resto de la columna C.
5. Dé clic en los siguientes menús para calcular la columna  $C$  mediante un promedio móvil de tres períodos.

Tools>Data Analysis

6. Aparece el cuadro de diálogo de *Data Analysis* (análisis de datos). Bajo *Analysis Tools* (herramientas de análisis) seleccione *Moving Average* (media móvil) y dé clic en OK. En la figura 5.21 se muestra el cuadro de diálogo correspondiente.

- a) Introduzca C1:C28 en el *Input Range* (rango de entrada)
- b) El intervalo es igual a 3.

FIGURA 5.21 Cuadro de diálogo del promedio móvil en Excel



- c) Introduzca D1 en *Output Range* (rango de salida). Excel ha colocado de manera inadecuada el primer promedio móvil en D3 en lugar de D2.
- d) Resalte la primera celda en el rango de resultados, D1.
- e) Dé clic en

**Edit>Delete**

y en el cuadro de diálogo *Delete* (eliminar) seleccione la opción *Shift Cells Up* (desplazar las celdas hacia arriba).

- f) Dé clic en OK.
- 7. Para crear la columna *I*, coloque el cursor en E2 e introduzca la fórmula =C2/D2. Copie esta fórmula en el resto de la columna E.
- 8. Para poder transferir las columnas *C*, *I* e *I* a la hoja de cálculo de Minitab, coloque el cursor en C1 y resalte hasta E28. Dé clic en los siguientes menús:

**Edit>Copy**

- 9. Abra la hoja de cálculo de Minitab que contiene la tabla 5.4 y resalte la fila 1 de C8. Dé clic en:

**Edit>Paste>Insert Cells**

Ahora, la tabla 5.4 está completa.

## Referencias

- Bell, W.R., y S.C. Hillmer, “Issues Involved with the Seasonal Adjustment of economic Time Series”, *Journal of Business and Economic Statistics* (1984): 291-320.
- Diebold, F. X., *Elements of Forecasting*, tercera edición, Cincinnati, OH, South Western, 2004.
- Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, y B. Chen, “New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program”, *Journal of Business and Economic Statistics* (1998): 127-152.
- Johnson, R. A. y D. W. Wichern, *Business Statistics: Decision Making with Data*, Nueva York, John Wiley & Sons, 1997.
- Makridakis, S., S. C. Wheelwright y R. J. Hyndman, *Forecasting Methods and Applications*, tercera edición, Nueva York, John Wiley & Sons, 1998.
- Moore, G. H., y J. Shiskin, “Early Warning Signals for the Economy”, en *Statistics: A Guide to Business and Economics*, eds. J. M. Tanur *et al.*, San Francisco: Holden-Day, 1976.
- Wichern, D. W., “Lagging indicators”, en *Encyclopedia of Statistical Sciences*, eds. S. Kotz *et al.*, vol. 4, Nueva York: John Wiley & Sons, 1983, 439-440.
- Wichern, D. W., “Leading Indicators”, en *Encyclopedia of Statistical Sciences*, eds. S. Kotz *et al.*, vol. 4, Nueva York, John Wiley & Sons, 1983, 582-585.



## REGRESIÓN LINEAL SIMPLE

En el capítulo 2 se analizó la asociación lineal (correlación) entre dos variables numéricas. La asociación lineal implica una relación en línea recta, y se mostró cómo ajustar una línea recta a pares de observaciones en las dos variables mediante el método de mínimos cuadrados. En este capítulo se estudian los modelos de regresión lineal simple (línea recta). Una vez que se establece una relación lineal, se puede utilizar el conocimiento de la variable independiente para pronosticar la variable dependiente.

Para repasar el análisis de la relación entre dos variables que se presentó en el capítulo 2, se considera el siguiente ejemplo.

**Ejemplo 6.1**

Suponga que el señor Bump observa el precio y el volumen de venta de leche durante 10 semanas seleccionadas al azar. Los datos que ha recabado se muestran en la tabla 6.1.

Primero elabora un diagrama de dispersión de datos, el cual se muestra en la figura 6.1. De acuerdo con este diagrama de dispersión, parece que existe una relación lineal negativa entre  $Y$ , el número de galones de leche vendidos, y  $X$ , el precio por galón. Aparentemente, cuando el precio sube, el volumen de ventas baja.

Ahora, Bump quiere medir el grado de esta aparente relación mediante el cálculo del coeficiente de correlación de la muestra  $r$ . Por medio de la ecuación 2.9 y los cálculos de la tabla 6.2, encuentra que

$$\begin{aligned} r &= \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2} \sqrt{n\sum Y^2 - (\sum Y)^2}} \\ &= \frac{10(149.3) - (14.4)(112)}{\sqrt{10(21.56) - (14.4)^2} \sqrt{10(1,488) - (112)^2}} \\ &= \frac{-119.8}{138.7} = -.86 \end{aligned}$$

TABLA 6.1 Datos de galones de leche para el ejemplo 6.1

| <i>Semana</i> | <i>Nivel semanal de ventas, Y<br/>(miles de galones)</i> | <i>Precio de venta<br/>X</i> |
|---------------|----------------------------------------------------------|------------------------------|
| 1             | 10                                                       | \$1.30                       |
| 2             | 6                                                        | 2.00                         |
| 3             | 5                                                        | 1.70                         |
| 4             | 12                                                       | 1.50                         |
| 5             | 10                                                       | 1.60                         |
| 6             | 15                                                       | 1.20                         |
| 7             | 5                                                        | 1.60                         |
| 8             | 12                                                       | 1.40                         |
| 9             | 17                                                       | 1.00                         |
| 10            | 20                                                       | 1.10                         |

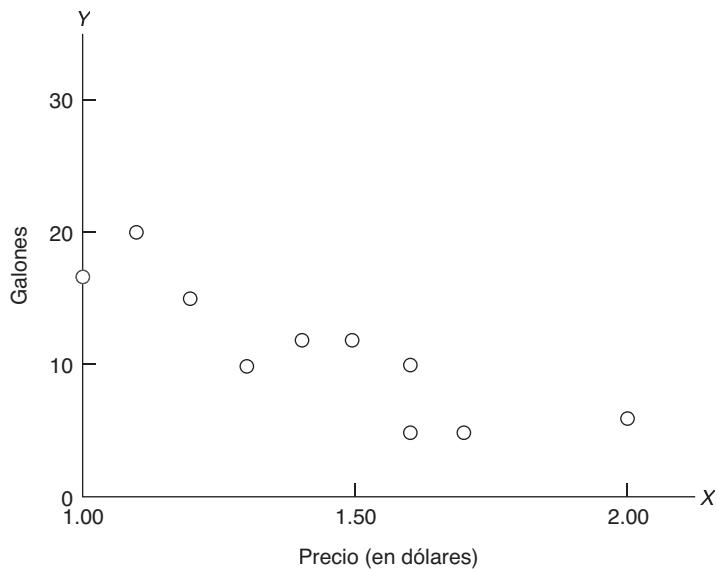


FIGURA 6.1 Diagrama de dispersión para el ejemplo 6.1

TABLA 6.2 Cálculos necesarios para ejemplo 6.1

|          | $Y$ | $X$   | $XY$  | $X^2$ | $Y^2$ |
|----------|-----|-------|-------|-------|-------|
| $n = 10$ | 10  | 1.30  | 13.0  | 1.69  | 100   |
|          | 6   | 2.00  | 12.0  | 4.00  | 36    |
|          | 5   | 1.70  | 8.5   | 2.89  | 25    |
|          | 12  | 1.50  | 18.0  | 2.25  | 144   |
|          | 10  | 1.60  | 16.0  | 2.56  | 100   |
|          | 15  | 1.20  | 18.0  | 1.44  | 225   |
|          | 5   | 1.60  | 8.0   | 2.56  | 25    |
|          | 12  | 1.40  | 16.8  | 1.96  | 144   |
|          | 17  | 1.00  | 17.0  | 1.00  | 289   |
|          | 20  | 1.10  | 22.0  | 1.21  | 400   |
| Totales  | 112 | 14.40 | 149.3 | 21.56 | 1,488 |

El coeficiente de correlación de la muestra de  $-0.86$  indica una relación bastante negativa entre  $Y$  y  $X$ , —mientras el precio del galón de leche aumenta, el número de galones vendidos disminuye.

La pregunta que podría surgir enseguida es: ¿qué tanto desciende el volumen a medida que aumenta el precio? Esta pregunta sugiere dibujar una línea recta a través de los puntos de datos expuestos en el diagrama de dispersión. Despues de haber dibujado esta línea, la pendiente de la línea mostrará la disminución del promedio en el volumen  $Y$  para cada incremento de dólar en  $X$ .

## LÍNEA DE REGRESIÓN

El señor Bump puede dibujar visualmente una línea recta a través de los puntos de datos tratando de ajustar la línea a los puntos tan estrechamente como sea posible. No obstante, alguien más puede dibujar una línea diferente. Un procedimiento más adecuado es encontrar la mejor línea recta utilizando un criterio que, para un conjunto de datos determinado, produzca la misma línea sin importar quién realice el ajuste.

Como se dijo en el capítulo 2, un criterio más utilizado para determinar la línea recta de mejor ajuste es el *criterio de mínimos cuadrados*.

La *línea que mejor se ajusta* a un conjunto de datos  $X$ - $Y$  es aquella que minimiza la suma del cuadrado de las distancias desde los puntos hasta la línea, midiéndola en dirección vertical, o  $Y$ . Se conoce como *recta de mínimos cuadrados* o *recta de regresión ajustada* y su ecuación se llama *ecuación de regresión ajustada*.

La línea recta ajustada es de la forma  $\hat{Y} = b_0 + b_1 X$ . El primer término,  $b_0$ , es la *intersección con eje Y*, y el segundo término,  $b_1$ , es la *pendiente*. Cabe recordar que la pendiente representa la cantidad de cambio en  $Y$  cuando  $X$  se incrementa en una unidad. El objetivo inmediato es determinar los valores de  $b_0$  y  $b_1$ .

El *método de mínimos cuadrados* elige los valores para  $b_0$  y  $b_1$  para reducir el error de la suma de los cuadrados (distancias):

$$SSE = \sum(Y - \hat{Y})^2 = \sum(Y - b_0 - b_1 X)^2 \quad (6.1)$$

Si se usa un poco de cálculo diferencial (vea el apéndice A), se pueden hacer derivaciones sobre expresiones algebraicas específicas para obtener los valores de mínimos cuadrados. Especialmente,

$$b_1 = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} \quad (6.2)$$

$$b_0 = \frac{\sum Y}{n} - \frac{b_1 \sum X}{n} = \bar{Y} - b_1 \bar{X} \quad (6.3)$$

Como se puede suponer, el coeficiente de la pendiente de mínimos cuadrados está relacionado con el coeficiente de correlación de la muestra. Específicamente,

$$b_1 = \frac{\sqrt{\sum(Y - \bar{Y})^2}}{\sqrt{\sum(X - \bar{X})^2}} r \quad (6.4)$$

De manera que  $b_1$  y  $r$  son proporcionales el uno con el otro y tienen el mismo signo.

La diferencia entre los valores  $Y$  realmente observados y los correspondientes valores ajustados  $\hat{Y}$  se llaman *residuales*. Los residuales son las distancias verticales (positiva o negativa) de los datos de puntos de la recta de mínimos cuadrados.

Se tiene la identidad

$$\text{Observación} = \text{Ajuste} + \text{Residual}$$

o en símbolos,

$$Y = \hat{Y} + (Y - \hat{Y}) \quad (6.5)$$

En este contexto, el ajuste representa el patrón general en los datos y los residuales representan las desviaciones de este patrón. El desdoblamiento en ajustes más residuales se aplica también a patrones que no sean las líneas rectas, lo que utilizaremos repetidamente en capítulos posteriores.

**Ejemplo 6.2**

Los coeficientes de mínimos cuadrados para una línea recta ajustados a los datos del señor Bump (vea la figura 6.1) se calculan fácilmente mediante las ecuaciones 6.2 y 6.3. Así, se tiene que

$$b_1 = \frac{10(149.3) - (14.4)(112)}{10(21.56) - (14.4)^2} = \frac{-119.8}{8.24} = -14.54$$

$$b_0 = \frac{112}{10} - (-14.54)\frac{14.4}{10} = 11.2 + 14.54(1.44) = 32.14$$

La línea de mínimos cuadrados ajustada tiene la siguiente ecuación

$$\begin{aligned}\hat{Y} &= b_0 + b_1 X \\ \hat{Y} &= 32.14 - 14.54X\end{aligned}\tag{6.6}$$

A esta ecuación se le llama *ecuación de regresión ajustada*.

Tal vez, el señor Bump quiera interpretar los valores en esta ecuación. La intersección con el eje  $Y$ ,  $b_0$ , es el valor  $Y$  cuando  $X$  es igual a cero. Una rigurosa interpretación sugiere que el número promedio de galones vendidos cuando  $X = 0$  (esto es, si el precio de un galón de leche fuera de cero) es de 32 140 galones. Esta interpretación no concuerda con el sentido común, pues uno esperaría que se “venda” más leche si fuera gratuita. El problema ilustrado aquí no supone predecir un valor para  $Y$  basado en un valor  $X$  sobre el cual ninguna evidencia de muestra ha sido recopilada. Es decir, ninguno de los puntos de muestra tiene un valor  $X$  en o cerca de cero. En este caso, como en muchos casos de análisis de regresión, no es posible obtener una interpretación útil de la intersección con el eje  $Y$ .

En términos más generales, a menudo no es muy conveniente predecir los valores  $Y$  para ninguna  $X$  más allá del rango de las  $X$  recopiladas en los datos de muestra. Una función de regresión debe ser interpretada como una aproximación útil al comportamiento del mundo real *sobre la región para la cual existen datos*. La extrapolación más allá de los datos requiere una fuerte suposición de que la naturaleza de la relación entre  $Y$  y  $X$  no cambia.<sup>1</sup>

El coeficiente de pendiente,  $b_1$ , puede ser interpretado como el cambio promedio en  $Y$  que ocurre cuando  $X$  aumenta una unidad. En este ejemplo,  $Y$  disminuye en un promedio de 14.54 (es decir, se venden 14 540 galones menos) cuando  $X$  aumenta una unidad (el costo de un galón de leche se eleva un dólar). Cada incremento de dólar en un galón de leche reduce la cantidad comprada a un promedio de 14 540 galones o, para poner esta afirmación en unidades más elocuentes, la evidencia empírica indica que cada aumento de un centavo por cada galón de leche reduce la cantidad comprada a un promedio de 145.4 galones.

La relación  $X$ - $Y$  se puede ilustrar al trazar una línea recta que ajuste mejor los datos de puntos sobre el diagrama de dispersión. El resultado se muestra en la figura 6.2. Nótese que las distancias verticales de los puntos a la línea se muestran como líneas punteadas. Si estas distancias fueran cuadradas y se sumaran, la suma resultante sería inferior a la de cualquier otra línea que pudiera dibujarse a través de los puntos.<sup>2</sup> De acuerdo con el procedimiento de mínimos cuadrados, esta línea representa el mejor ajuste posible para los 10 datos de puntos de la muestra.

Las ideas estadísticas deben ser introducidas en el estudio de la relación entre dos variables cuando los puntos en la gráfica de dispersión no descansan exactamente en una línea, como en la figura 6.2. Se piensa en los datos de la gráfica de dispersión como una muestra de observaciones en una relación subyacente que contiene la población de los valores  $X$ - $Y$ . El modelo estadístico de regresión lineal simple supone que por cada valor de  $X$ , los valores observados de la variable dependiente  $Y$  son, por lo general, distribuidos so-

<sup>1</sup> Se encontró una situación similar cuando se analizó, en el capítulo 5, la utilidad de extrapolar curvas de tendencia para predecir valores futuros de series de tiempo.

<sup>2</sup> Para los datos de Bump, esta pequeña suma de distancias cuadradas es  $SSE = 59.14$ .

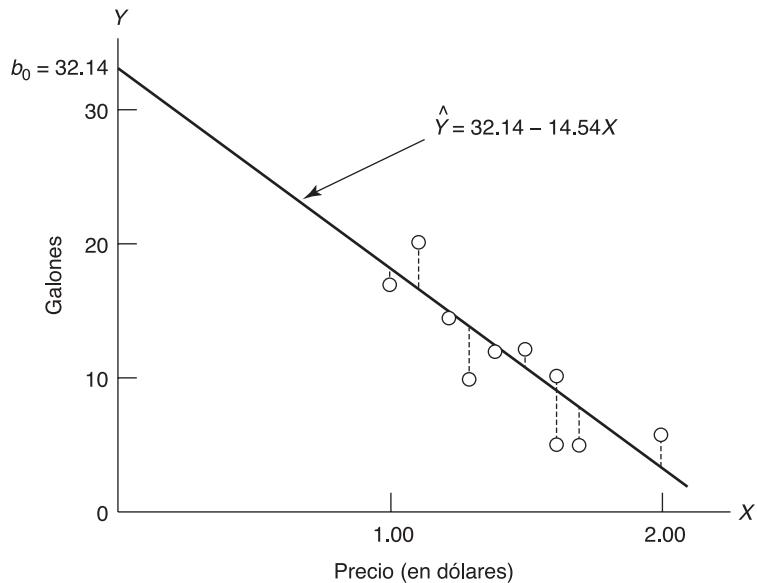


FIGURA 6.2 Línea de regresión ajustada para los datos de Bump

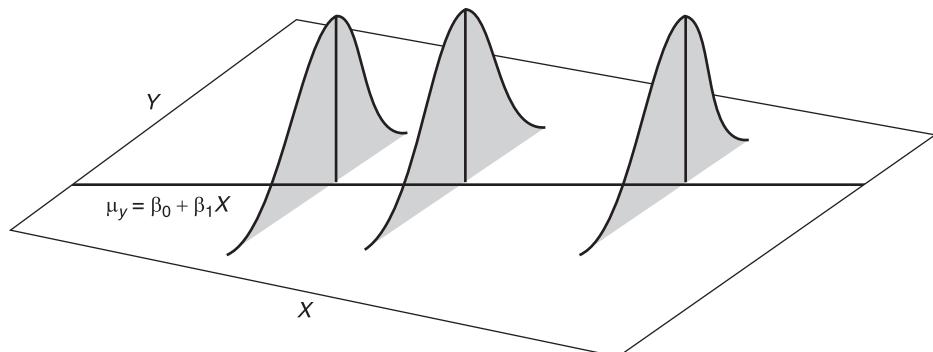


FIGURA 6.3 Modelo estadístico de regresión lineal simple

bre un valor promedio,  $\mu_y = \beta_0 + \beta_1 X$ , que depende linealmente de  $X$ . Esto es, mientras  $X$  cambia, los valores promedio de la distribución de los posibles valores  $Y$  descansan sobre una línea recta. Esta línea se conoce como *línea de regresión de la población*. Las  $Y$  observadas variarán respecto de estos valores promedio por la influencia de los factores que no se pueden medir. El modelo supone que esta variación, medida por la desviación estándar  $\sigma$ , es la misma para todos los valores de  $X$ . Finalmente, la desviación (distancia) entre un valor  $Y$  y su valor promedio es conocida como *error* y se representa con  $\epsilon$ , la letra griega épsilon.

En el modelo de regresión lineal simple, una respuesta  $Y$  es la suma de su valor promedio y una desviación aleatoria  $\epsilon$  del promedio. Las desviaciones  $\epsilon$  representan variación en  $Y$  debido a otros factores inadvertidos que impiden que los valores  $X-Y$  descansen exactamente sobre una línea recta en la gráfica de dispersión.

El modelo estadístico para la regresión lineal simple se ilustra de manera esquemática en la figura 6.3.

### Modelo estadístico de regresión en línea recta

La variable de respuesta, o variable dependiente  $Y$ , se relaciona con la variable de control, o independiente  $X$ , por

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Aquí  $\beta_0 + \beta_1 X$  es la respuesta promedio para una  $X$  determinada. Se supone que las desviaciones  $\epsilon$  son independientes y por lo general se distribuyen con media 0 y desviación estándar  $\sigma$ . Las constantes desconocidas son  $\beta_0$ ,  $\beta_1$ , y  $\sigma$ .

La línea de regresión ajustada (muestral)  $\hat{Y} = b_0 + b_1 X$  puede considerarse un estimado de la línea de regresión de la población  $\mu_y = \beta_0 + \beta_1 X$  y los residuales  $e = Y - \hat{Y}$  pueden considerarse como estimados de los componentes de error  $\epsilon$ . Esto implica la siguiente correspondencia

$$\begin{aligned} \text{Población: } Y &= \beta_0 + \beta_1 X + \epsilon \\ \text{Muestra: } Y &= b_0 + b_1 X + e \end{aligned}$$

La mayoría de los paquetes de software realizan análisis de correlación y regresión. Las instrucciones para utilizar Excel para correr los ejemplos 6.1 y 6.2 se encuentran en la sección “Aplicaciones de Excel” al final de este capítulo.

## ERROR ESTÁNDAR DE LA ESTIMACIÓN

Después de haber calculado la línea recta ajustada, ahora el señor Bump puede estar interesado en medir la dispersión de los datos muestrales en torno a la función de regresión ajustada. Se puede desarrollar una medida de dispersión análoga a la desviación estándar muestral. Esta medida, llamada el *error estándar de la estimación*, mide la dispersión de los datos respecto a la línea ajustada en la dirección  $Y$ . El error estándar de la estimación se indica con  $s_{y \cdot x}$  y se determina por

$$s_{y \cdot x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}} \quad (6.7)$$

El *error estándar de la estimación* mide la cantidad por la cual los valores verdaderos  $Y$  difieren de los valores estimados  $\hat{Y}$ . Para muestras relativamente grandes, se esperaría que alrededor de 67% de las diferencias  $Y - \hat{Y}$  estuvieran a una distancia inferior a  $s_{y \cdot x}$  del 0 y aproximadamente el 95% de estas diferencias estarían a  $2s_{y \cdot x}$  del 0.

El error estándar de la estimación es similar a la desviación estándar de la muestra presentada en el capítulo 2. Puede ser usado para estimar la desviación estándar de la población. De hecho,  $s_{y \cdot x}$  estima la desviación estándar  $\sigma$  del término de error  $\epsilon$  en el modelo estadístico de regresión lineal simple. De manera equivalente,  $s_{y \cdot x}$  estima la desviación estándar común  $\sigma$  de la distribución normal de los valores  $Y$  sobre la línea de regresión de la población  $\mu_y = \beta_0 + \beta_1 X$  en cada  $X$  (vea la figura 6.3).

Un análisis de regresión con un pequeño error estándar de la estimación significa que todos los datos descansan muy cerca de la línea de regresión ajustada.<sup>3</sup> Si el error estándar de la estimación es grande, los datos se dispersan considerablemente respecto de la línea ajustada.

Para fines de cálculo, la ecuación 6.7 se puede convertir a

$$s_{y \cdot x} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum XY}{n - 2}} \quad (6.8)$$

Para el ejemplo de Bump, el error estándar de la estimación es

$$\begin{aligned} s_{y \cdot x} &= \sqrt{\frac{1,488 - (32.14)(112) - (-14.54)(149.3)}{10 - 2}} \\ &= \sqrt{\frac{59.14}{8}} = \sqrt{7.39} = 2.72 \end{aligned}$$

Con los valores  $\hat{Y}$  en el rango de 3 a 18 (vea la figura 6.2),  $s_{y \cdot x}$  es moderadamente grande e indica que una cantidad considerable de la variación en  $\hat{Y}$  (galones vendidos) no se explica con  $X$  (precio). Más adelante investigaremos esta afirmación con más detalle en la sección “Coeficiente de determinación”.

## PRONÓSTICO DE $Y$

---

A continuación, la línea de regresión ajustada puede usarse para estimar el valor de  $Y$  para un valor determinado  $X$ . Para obtener una *predicción puntual* o pronóstico, para un valor determinado  $X$ , simplemente evalúe la función de regresión estimada en  $X$ .

### Ejemplo 6.3

Suponga que el señor Bump desea pronosticar la cantidad de leche vendida si el precio por galón fuera de 1.63 dólares. El pronóstico de la ecuación 6.6 es

$$\begin{aligned} \hat{Y} &= 32.14 - 14.54X \\ \hat{Y} &= 32.14 - 14.54(1.63) = 8.440 \end{aligned}$$

u 8 440 galones. Nótese que este pronóstico es un valor de  $\hat{Y}$ ; es decir, el pronóstico es la coordenada  $Y$  del punto sobre la línea de regresión ajustada donde  $X = 1.63$ .

Por supuesto, el señor Bump está consciente de que es poco probable que los valores reales de  $Y$  correspondientes a las posiciones de las  $X$  descansen exactamente en la línea de regresión. De hecho, éstos se dispersarán sobre la línea de acuerdo con el valor  $s_{y \cdot x}$ ; además, la línea de regresión de la muestra (ajustada) es un estimado de la línea de regresión de la población con base en una muestra de 10 puntos de datos. Otras muestras aleatorias de 10 producirían diferentes líneas de regresión ajustada, similares al caso donde muchas muestras dibujadas desde la misma población tienen diferentes medias muestrales.

Existen dos fuentes de incertidumbre asociadas con una predicción puntual generada por la ecuación de regresión adaptada:

1. Incertidumbre debida a la dispersión de los datos respecto a la línea de regresión de la muestra.
2. Incertidumbre debida a la dispersión de la regresión de la muestra respecto a la línea de regresión de la población

---

<sup>3</sup> Si todos los datos descansan exactamente en la línea ajustada,  $Y = \hat{Y}$  para todas las  $X$ , y  $s_{y \cdot x} = 0$ .

Es posible elaborar un *intervalo de predicción* de  $Y$  que tome en cuenta estas dos fuentes de incertidumbre.

El *error estándar del pronóstico*,  $s_f$ , mide la variabilidad de la  $Y$  prevista sobre la  $Y$  real para un valor determinado de  $X$ . El error estándar del pronóstico es

$$\begin{aligned} s_f &= \sqrt{s_{y \cdot x}^2 + s_{y \cdot x}^2 \left( \frac{1}{n} + \frac{(X - \bar{X})^2}{\Sigma(X - \bar{X})^2} \right)} \\ s_f &= s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\Sigma(X - \bar{X})^2}} \end{aligned} \quad (6.9)$$

El primer término,  $s_{y \cdot x}^2$ , bajo el primer signo radical en la ecuación 6.9 mide la dispersión de los puntos de datos sobre la línea de regresión de la muestra (primera fuente de incertidumbre). El segundo término bajo el signo radical mide la dispersión de la línea de regresión de la muestra sobre la línea de regresión de la población (segunda fuente de incertidumbre). Nótese que el error estándar del pronóstico depende de  $X$ , el valor de  $X$  por el cual se desea un pronóstico  $Y$ . También observe que  $s_f$  es menor cuando  $X = \bar{X}$  porque entonces el numerador en el tercer término bajo la radical en la ecuación 6.9 será  $(\bar{X} - \bar{X})^2 = 0$ .<sup>4</sup> Asimismo, mientras más lejos esté  $X$  de  $\bar{X}$ , mayor será el error estándar del pronóstico.

Si el modelo estadístico para la regresión lineal simple es el adecuado, se determina un intervalo de predicción para  $Y$  por

$$\hat{Y} \pm ts_f \quad (6.10)$$

donde  $t$  es un punto porcentual de la distribución  $t$  de Student con  $df = n - 2$  grados de libertad. Si el tamaño de la muestra es más grande ( $n \geq 30$ ), el punto porcentual  $t$  puede reemplazarse con el punto porcentual correspondiente  $Z$  de la distribución normal estándar. Por ejemplo, un intervalo de predicción de muestra grande de 95% para  $Y$  es esencialmente

$$\hat{Y} \pm 2s_f \quad (6.11)$$

#### Ejemplo 6.4

De manera gráfica, el intervalo de predicción de 95% del señor Bump para  $Y$  con varios valores  $X$  se vería como en la figura 6.4.

Al usar los resultados de la tabla 6.3 y la ecuación 6.9 con  $\bar{X} = 1.44$ , el error estándar del pronóstico en  $X = 1.63$  es

$$s_f = 2.72 \sqrt{1 + \frac{1}{10} + \frac{(1.63 - 1.44)^2}{.824}} = 2.72(1.069) = 2.91$$

Del ejemplo 6.3,  $\hat{Y} = 8.440$  cuando  $X = 1.63$  y, mediante la ecuación 6.10, un intervalo de predicción de 95% para  $Y$  es

$$\hat{Y} \pm ts_f = 8.44 \pm 2.306(2.91) = 8.44 \pm 6.71$$

o (1.73, 15.15); esto es, desde 1 730 galones hasta 15 150 galones. Aquí 2.306 =  $t_{0.025}$  es el punto que deja un área de 2.5% a la derecha en una distribución de  $t$  con  $df = 8$  grados de libertad.

El intervalo de predicción es muy extenso; tanto, que es prácticamente inútil pronosticar  $Y$ , y esto ocurre por el tamaño tan pequeño de la muestra y el valor relativamente grande de  $s_f$ . La cuantía de la incertidumbre reflejada por el amplio intervalo de predicción no es evidente en el

---

<sup>4</sup> Para la elección,  $X = \bar{X}$ , la  $Y$  pronosticada es  $\hat{Y} = \bar{Y}$ .

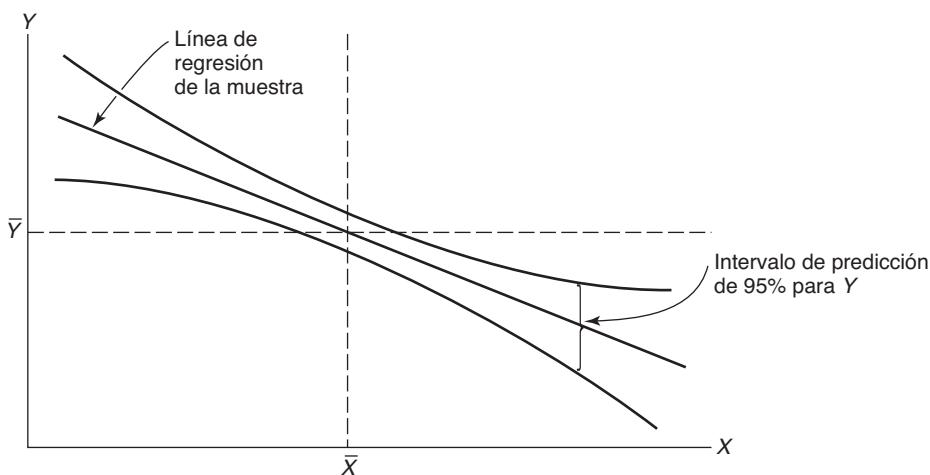


FIGURA 6.4 Intervalo de predicción para los datos de Bump

TABLA 6.3 Cálculo de  $\sum(X - \bar{X})^2$  para el ejemplo 6.4

| $X$                             | $(X - \bar{X})^2$ |
|---------------------------------|-------------------|
| 1.30                            | .0196             |
| 2.00                            | .3136             |
| 1.70                            | .0676             |
| 1.50                            | .0036             |
| 1.60                            | .0256             |
| 1.20                            | .0576             |
| 1.60                            | .0256             |
| 1.40                            | .0016             |
| 1.00                            | .1936             |
| 1.10                            | .1156             |
| $\Sigma(X - \bar{X})^2 = .8240$ |                   |

pronóstico puntual calculado a partir de la función de regresión ajustada. La mayor ventaja del intervalo estimado es que explica categóricamente las incertidumbres asociadas al pronóstico.

En general, es peligroso utilizar la función de regresión ajustada para predecir valores de  $Y$  más allá del rango de los datos disponibles. Se puede justificar que el señor Bump trate de predecir  $Y$  cuando  $X = 1.63$  porque algunos de los valores originales de  $X$  están cerca de 1.63. Por otro lado, podría ser imprudente predecir  $Y$  cuando  $X = 3.00$ . No se han recopilado datos para valores  $X$  tan grandes, y por esta razón cualquier predicción que implique tal valor  $X$  sería muy dudosa.<sup>5</sup> Para calcular la cantidad de leche vendida cuando el precio por galón es de 3.00 dólares, el señor Bump tiene que suponer que el modelo de línea recta es aún válido. Puede tener buenas razones para pensar así, pero no tiene evidencia directa para respaldarlo.

Es bastante útil terminar esta sección revisando los supuestos subyacentes del modelo estadístico de regresión lineal.

<sup>5</sup> Por otra parte, el error estándar del pronóstico sería muy grande porque la cantidad  $(X - \bar{X})^2$  sería relativamente grande.

1. Para un valor determinado  $X$ , la población de valores  $Y$  es normalmente distribuida sobre la línea de regresión de la población. Esta condición se muestra en la figura 6.3. En la práctica, se obtienen resultados razonablemente precisos mientras los valores  $Y$  se distribuyan de manera aproximadamente normal.
2. La dispersión de los datos poblacionales alrededor de la línea de regresión de la población permanece constante a lo largo de la línea. Es decir, la varianza de la población no se hace ni más grande ni más pequeña, aunque los valores  $X$  de los puntos de datos aumenten. Una violación a este supuesto es denominada *heteroscedasticidad*; se puede encontrar un ejemplo de esta condición y su remedio en el capítulo 8.
3. Los términos de error ( $\epsilon$ ) son independientes uno del otro. Este supuesto implica que la muestra de datos  $X-Y$  es aleatoria. Cuando los datos  $X-Y$  son registrados a lo largo del tiempo, a menudo se viola este supuesto. Más que ser independientes, las observaciones consecutivas están correlacionadas serialmente. En el capítulo 8 se consideran los métodos para manejar el problema de la correlación serial.
4. En la población existe una relación lineal entre  $X$  y  $Y$ . También hay extensiones de la regresión lineal simple para tratar las relaciones  $X-Y$  que no son lineales; más adelante se hablará de algunas de ellas en este capítulo.

## DESCOMPOSICIÓN DE LA VARIANZA

---

De la ecuación 6.5,

$$Y = \hat{Y} + (Y - \hat{Y})$$

o

$$Y = b_0 + b_1 X + (Y - b_0 - b_1 X)$$

↑                      ↑                      ↑

Valor observado  $Y$     Explicado por la relación lineal    Residual o desviación de la relación lineal

En una situación ideal, donde todos los puntos descansan exactamente en la línea ajustada, todos los residuales están en cero y los valores  $Y$  están completamente explicados por la relación lineal con  $X$ .

Al restar  $\bar{Y}$  de ambos lados de la expresión previa,

$$(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$$

Se puede mostrar con un poco de álgebra que las sumas de los cuadrados tiene como resultado:

$$\Sigma(Y - \bar{Y})^2 = \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2 \quad (6.12)$$

o

$$SST = SSR + SSE$$

donde

$$SST = \Sigma(Y - \bar{Y})^2$$

$$SSR = \Sigma(\hat{Y} - \bar{Y})^2$$

$$SSE = \Sigma(Y - \hat{Y})^2$$

Aquí  $SS$  significa *suma de los cuadrados* y  $T, R$  y  $E$  representan *total, regresión* y *error*, respectivamente. Las sumas de los cuadrados tienen los siguientes grados de libertad asociados

$$df(SST) = n - 1$$

$$df(SSR) = 1$$

$$df(SSE) = n - 2$$

En correspondencia con las sumas de los cuadrados, los grados de libertad se relacionan por

$$n - 1 = 1 + (n - 2) \quad (6.13)$$

Si no existe una relación lineal, la  $Y$  no depende de  $X$  y la variación en  $Y$  se describe por la varianza de la muestra

$$s_y^2 = \frac{1}{n-1} \sum (Y - \bar{Y})^2$$

Si, por otro lado, la  $Y$  se relaciona con  $X$ , algo de la diferencia en los valores de  $Y$  se deben a esta relación.

La regresión de la suma de los cuadrados,  $SSR$ , mide esa parte de la variación en  $Y$  explicada por la relación lineal. El error de la suma de los cuadrados,  $SSE$ , es la variación restante en  $Y$ , o la variación en  $Y$  no explicada por la relación lineal.

### Descomposición de la variabilidad

|                              |   |                                                     |   |                                            |
|------------------------------|---|-----------------------------------------------------|---|--------------------------------------------|
| <b>SST</b>                   | = | <b>SSR</b>                                          | + | <b>SSE</b>                                 |
| Variabilidad<br>total de $Y$ |   | Variabilidad<br>explicada por la<br>relación lineal |   | Residual o<br>variabilidad<br>no explicada |

Las sumas de los cuadrados asociados a la descomposición de la variabilidad de  $Y$  y sus grados de libertad correspondientes se pueden arreglar como se muestra en la tabla 6.4, la cual es conocida como *tabla ANOVA (analysis of variance, análisis de varianza)*.

La columna final en la tabla ANOVA es la columna del *cuadrado medio*. El cuadrado medio de la regresión,  $MSR$ , es la suma de los cuadrados de la regresión dividida entre su número de grados de libertad. Asimismo, el error cuadrado medio,  $MSE$ , es la suma de los cuadrados de los errores dividida entre su número de grados de libertad.

**TABLA 6.4 Tabla ANOVA para regresión en línea recta**

| <i>Fuente</i> | <i>Suma de los cuadrados</i> | <i>df</i> | <i>Cuadrado medio</i> |
|---------------|------------------------------|-----------|-----------------------|
| Regresión     | $SSR$                        | 1         | $MSR = SSR/1$         |
| Error         | $SSE$                        | $n - 2$   | $MSE = SSE/(n - 2)$   |
| Total         | $SST$                        | $n - 1$   |                       |

**TABLA 6.5 Residuales para los datos de Bump con  $\bar{Y}$  como el pronóstico**

| <i>Y Real</i> | <i>Y Pronosticada (<math>\bar{Y}</math>)</i> | <i>Residual (<math>Y - \bar{Y}</math>)</i> | <i>(<math>Y - \bar{Y}</math>)<sup>2</sup></i> |
|---------------|----------------------------------------------|--------------------------------------------|-----------------------------------------------|
| 10            | 11.2                                         | -1.2                                       | 1.44                                          |
| 6             | 11.2                                         | -5.2                                       | 27.04                                         |
| 5             | 11.2                                         | -6.2                                       | 38.44                                         |
| 12            | 11.2                                         | .8                                         | .64                                           |
| 10            | 11.2                                         | -1.2                                       | 1.44                                          |
| 15            | 11.2                                         | 3.8                                        | 14.44                                         |
| 5             | 11.2                                         | -6.2                                       | 38.44                                         |
| 12            | 11.2                                         | .8                                         | .64                                           |
| 17            | 11.2                                         | 5.8                                        | 33.64                                         |
| 20            | 11.2                                         | 8.8                                        | 77.44                                         |
| Totales       |                                              | 0.0                                        | 233.60                                        |

Note que en la ecuación 6.7 el error cuadrático medio es

$$MSE = \frac{SSE}{n-2} = \frac{\Sigma(Y - \hat{Y})^2}{n-2} = s_{y \cdot x}^2$$

el cuadrado del error estándar de la estimación. Las proporciones de los cuadrados medios serán usadas para otros propósitos más adelante en este capítulo.

### Ejemplo 6.5

El señor Bump empezó su análisis al conocer las cantidades de los volúmenes de venta de sólo 10 semanas (la variable  $Y$ ). Si no hubiera ninguna información adicional, el señor Bump podría utilizar el promedio de muestra  $\bar{Y} = 11.2$  como un pronóstico de los galones de leche vendidos cada semana. Los errores o residuales, asociados con este pronóstico son  $Y - \bar{Y}$  y la suma de los errores cuadráticos es  $\Sigma(Y - \bar{Y})^2$ . Esta última cantidad,  $\Sigma(Y - \bar{Y})^2$ , es exactamente  $SST$ , la suma total de los cuadrados introducidos en la ecuación 6.12. De este modo,  $SST$  mide la variabilidad de  $Y$  sobre un pronóstico que utiliza solamente los valores de  $Y$  en su cálculo.<sup>6</sup> Los pronósticos  $\bar{Y}$ , los residuales  $Y - \bar{Y}$ , y la suma total de los cuadrados  $SST = \Sigma(Y - \bar{Y})^2$  se muestran en la tabla 6.5.<sup>7</sup>

Asimismo, el señor Bump tiene información sobre una variable  $X$ , el precio por galón de la leche, que está relacionada con  $Y$ , el volumen semanal de leche vendida. (Recuerde el ejemplo 6.1 en donde  $r = -0.86$ .) Con esta variable adicional, él espera poder explicar algunas de estas variaciones (diferencias) en los valores  $Y$  más allá de lo explicado por el pronóstico  $\bar{Y}$ .

En el ejemplo 6.2, la línea ajustada a la gráfica de dispersión de las observaciones  $X$ - $Y$  tiene la ecuación  $\hat{Y} = 32.14 - 14.54X$ . Se puede elaborar una tabla similar a la tabla 6.5 si se usa  $\hat{Y}$  para pronosticar  $Y$ . El resultado es la tabla 6.6.<sup>8</sup>

Al comparar las tablas 6.5 y 6.6, se observa que el uso de  $\hat{Y}$  como pronóstico de  $Y$  ha resultado generalmente en residuales más pequeños (en valor absoluto) y una suma de cuadrados residual (error) considerablemente menor que cuando se utiliza  $\bar{Y}$  como pronóstico. El uso de la variable relacionada  $X$  reduce los errores de pronóstico o predicción. En otras palabras, el conocimiento de  $X$  ayuda a explicar las diferencias en los datos  $Y$ . ¿Qué tanto ayuda  $X$ ? La descomposición de la variabilidad responde esta pregunta.

Al emplear las tablas 6.5 y 6.6, y la ecuación 6.12,

$$SST = \Sigma(Y - \bar{Y})^2 = 233.60$$

$$SSE = \Sigma(Y - \hat{Y})^2 = 59.41$$

<sup>6</sup> Si el análisis se terminara en este punto, la variabilidad en  $Y$  sería medida por la varianza de la muestra  $s_y^2 = \Sigma(Y - \bar{Y})^2/(n - 1)$  más que por  $SST = \Sigma(Y - \bar{Y})^2$ . La varianza de la muestra es la medida usual de variabilidad para mediciones en una variable sencilla.

<sup>7</sup> Los residuales  $Y - \bar{Y}$  siempre suman cero porque el promedio  $\bar{Y}$  es el centro matemático de los valores  $Y$ .

<sup>8</sup> Si se incluye un término de intersección con el eje  $Y$  en la ecuación de regresión, la suma de los residuales  $\Sigma(Y - \hat{Y})$  será siempre cero.

**TABLA 6.6 Residuales para los datos de Bump con  $\hat{Y}$  como el pronóstico**

| <i>X</i> | <i>Y</i> | <i>Y Pronosticada (<math>\hat{Y}</math>) Usando<br/><math>\hat{Y} = 32.14 - 14.54X</math></i> | <i>Residual<br/>(<math>Y - \hat{Y}</math>)</i> | <i>(<math>Y - \hat{Y}</math>)<sup>2</sup></i> |
|----------|----------|-----------------------------------------------------------------------------------------------|------------------------------------------------|-----------------------------------------------|
| 1.30     | 10       | 13.238                                                                                        | -3.238                                         | 10.48                                         |
| 2.00     | 6        | 3.060                                                                                         | 2.940                                          | 8.64                                          |
| 1.70     | 5        | 7.422                                                                                         | -2.422                                         | 5.87                                          |
| 1.50     | 12       | 10.330                                                                                        | 1.670                                          | 2.79                                          |
| 1.60     | 10       | 8.876                                                                                         | 1.124                                          | 1.26                                          |
| 1.20     | 15       | 14.692                                                                                        | .308                                           | .09                                           |
| 1.60     | 5        | 8.876                                                                                         | -3.876                                         | 15.02                                         |
| 1.40     | 12       | 11.784                                                                                        | .216                                           | .05                                           |
| 1.00     | 17       | 17.600                                                                                        | -.600                                          | .36                                           |
| 1.10     | 20       | 16.146                                                                                        | 3.854                                          | 14.85                                         |
| Totales  |          | 0.000                                                                                         | 59.41                                          |                                               |

y en consecuencia,

$$SSR = \Sigma(\hat{Y} - \bar{Y})^2 = 233.60 - 59.41 - 174.19$$

La descomposición de la variabilidad es

|                    |   |                        |   |                           |
|--------------------|---|------------------------|---|---------------------------|
| <i>SST</i>         | = | <i>SSR</i>             | + | <i>SSE</i>                |
| 233.60             | = | 174.19                 | + | 59.41                     |
| Variación<br>total |   | Variación<br>explicada |   | Variación<br>no explicada |

El señor Bump observa que de la variabilidad restante después de pronosticar *Y* con  $\bar{Y}$ , una proporción de éste

$$\frac{SSR}{SST} = \frac{174.19}{233.60} = .75$$

ha sido explicada por la relación de *Y* con *X*. Una proporción  $1 - 0.75 = 0.25$  de la variación en *Y* en torno a  $\bar{Y}$  permanece sin explicación. Desde esta perspectiva, el conocimiento de la variable relacionada *X* resulta en pronósticos de *Y* que son mejores que los obtenidos de  $\bar{Y}$ , una cantidad que no depende de *X*.

La descomposición de la variabilidad para los datos del señor Bump se puede presentar en una tabla de análisis de varianza, como lo demuestra el siguiente ejemplo.

#### Ejemplo 6.6

El señor Bump forma una tabla de análisis de varianza (ANOVA) para sus datos para la que usa el formato de la tabla 6.4 y los cálculos de las sumas de los cuadrados del ejemplo 6.5, como se muestra en la tabla 6.7.

La descomposición de la variabilidad se muestra claramente en la columna donde se suman los cuadrados. Nótese que, dentro del redondeo del error,

$$MSE = 7.43 = (2.72)^2 = s_{y \cdot x}^2$$

**TABLA 6.7 Tabla ANOVA para los datos de Bump**

| <i>Fuente</i> | <i>Suma de los cuadrados</i> | <i>df</i> | <i>Cuadrado medio</i> |
|---------------|------------------------------|-----------|-----------------------|
| Regresión     | 174.19                       | 1         | 174.19                |
| Error         | 59.41                        | 8         | 7.43                  |
| Total         | 233.60                       | 9         |                       |

## COEFICIENTE DE DETERMINACIÓN

La identidad

$$(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$$

conduce a la descomposición de la varianza determinada por la ecuación 6.12 y se muestra gráficamente para la recta de regresión ajustada del señor Bump y un punto de datos hipotético en la figura 6.5.

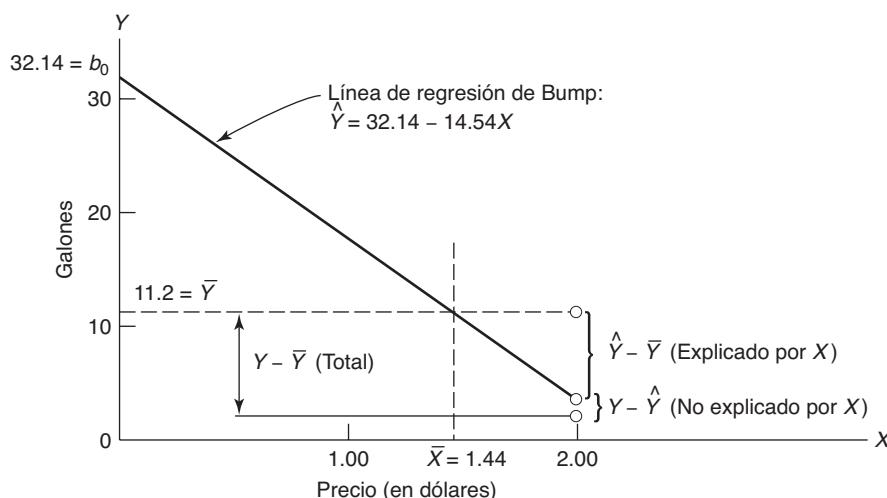
Si  $Y$  no dependiera de  $X$ , el señor Bump esperaría que  $Y$  estuviera cerca de  $\bar{Y}$  y que la desviación  $Y - \bar{Y}$  simplemente reflejara variación aleatoria. Sin embargo,  $Y$  depende de  $X$  en la manera sugerida por la función de regresión ajustada. En la figura, el valor observado  $X$  es mayor que  $\bar{X}$ , y se sabe que  $X$  y  $Y$  tienen una alta correlación negativa ( $r = -0.86$ ). De la distancia vertical total de  $Y - \bar{Y}$ , una cantidad  $\hat{Y} - \bar{Y}$  es, por lo tanto, “explicada” por el movimiento en  $X$ , mientras que la distancia vertical restante  $Y - \hat{Y}$  “no se explica” por el movimiento en  $X$ .

Como se indica en la sección previa,  $SST$  mide la variación total en torno a  $\bar{Y}$ , y esa parte del total que se explicó por el movimiento en  $X$  es  $SSR$ . La variación restante o no explicada es  $SSE$ . El cociente de lo explicado sobre la variación total se llama *coeficiente de determinación* de la muestra y se indica con  $r^2$ .

$$\begin{aligned} r^2 &= \frac{\text{Variación explicada}}{\text{Variación total}} = \frac{SSR}{SST} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} \\ &= 1 - \frac{\text{Variación no explicada}}{\text{Variación total}} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} \end{aligned} \quad (6.14)$$

El *coeficiente de determinación* mide el porcentaje de variabilidad en  $Y$  que se puede explicar con el conocimiento de la variabilidad (diferencias) en la variable independiente  $X$ .

**FIGURA 6.5 Variaciones explicadas y no explicadas en los datos de Bump**



**Ejemplo 6.7**

El coeficiente de determinación  $r^2$  para los datos del señor Bump se calculó en el ejemplo 6.5, aunque en ese punto no estaba reconocido como tal. El coeficiente de determinación también está disponible para la tabla ANOVA, tabla 6.7, en el ejemplo 6.6. Recuerde que

$$\begin{aligned} SST &= \Sigma(Y - \bar{Y})^2 = 233.60 \\ SSR &= \Sigma(\hat{Y} - \bar{Y})^2 = 174.19 \\ SSE &= \Sigma(Y - \hat{Y})^2 = 59.41 \end{aligned}$$

y

$$r^2 = \frac{174.19}{233.60} = .746$$

De manera alterna,  $r^2$  se puede calcular como

$$r^2 = 1 - \frac{59.41}{233.60} = 1 - .254 = .746$$

Alrededor de 75% de la variabilidad en los galones de leche vendidos ( $Y$ ) se puede explicar por las diferencias en el precio por galón ( $X$ ). Cerca de 25% de la variabilidad en la cantidad de leche vendida no se puede explicar por el precio. Esta porción de la variabilidad debe ser explicada por factores que no han sido considerados en este análisis de regresión (por ejemplo, el monto de la publicidad, existencia de productos sustitutos y la calidad de la leche).

Para la regresión en línea recta,  $r^2$  es el cuadrado de  $r$ , el coeficiente de correlación.

$$\text{Coeficiente de determinación} = (\text{Coeficiente de correlación})^2$$

$$r^2 = (r)^2$$

Por lo tanto, para los datos de Bump, e ignorando el redondeo del error,

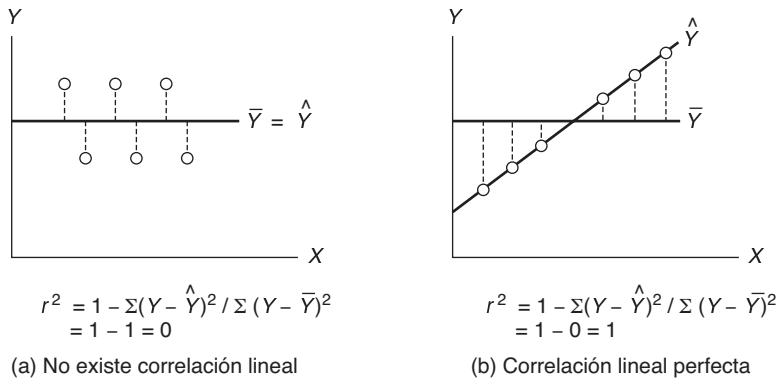
$$0.746 = (-0.863)^2$$

¿Por qué es necesario identificar específicamente  $r$  y  $r^2$  en un análisis de regresión? Porque transmiten información diferente.

El coeficiente de correlación revela la fuerza y dirección de la relación lineal. En el caso de los datos que el señor Bump ha recopilado, existe una relación negativa ( $r = -0.86$ ). En otros casos, un valor positivo de  $r$  puede tener una relación positiva. Como se verá en el capítulo siguiente, es útil identificar la naturaleza de las relaciones que existen entre ciertos pares de variables cuando tratan con una serie grande de variables. Observe que cuando el coeficiente de correlación se eleva al cuadrado, el valor es siempre positivo, por lo cual se pierde la naturaleza de la relación.

El coeficiente de determinación  $r^2$  mide la fuerza de la relación entre  $Y$  y  $X$  de manera distinta que en  $r$ . El valor para  $r^2$  mide el *porcentaje* de la variabilidad en  $Y$  que se explica por las diferencias en  $X$ . Esta útil interpretación se puede generalizar a las relaciones entre  $Y$  y varias variables  $X$  en vez de una sola  $X$ .

La figura 6.6 ilustra los dos casos extremos para  $r^2$ ,  $r^2 = 0$  y  $r^2 = 1$ . En el caso anterior, *ninguna* variabilidad en  $Y$  se explica por  $X$ : el diagrama de dispersión no sugiere *ninguna* relación lineal entre  $X$  y  $Y$ . Cuando  $r^2 = 1$ , *todas* las variabilidades en  $Y$  se explican cuando se conoce  $X$ : todos los puntos de datos de la muestra descansan en la línea de regresión ajustada.



**FIGURA 6.6** Valores extremos para  $r^2$

## PRUEBA DE HIPÓTESIS

La línea de regresión ajustada se produce por una *muestra* de valores  $X$ - $Y$ . El modelo estadístico de la línea de regresión ajustada sugiere que la relación de línea recta entre  $Y$  y  $X$  se mantiene en todas las opciones de pares  $X$ - $Y$ . Es decir, existe una auténtica relación entre  $X$  y  $Y$  de la forma  $\mu_y = \beta_0 + \beta_1 X$ . Dada la evidencia muestral, ¿se puede concluir que la relación es verdadera para todas las  $X$  y  $Y$ ?

Considere esta hipótesis

$$H_0: \beta_1 = 0$$

donde  $\beta_1$  es la pendiente de la línea de regresión de la población. Observe que si esta hipótesis es verdadera, no existe una relación entre  $Y$  y  $X$  en la población. Si no se rechaza  $H_0$  significa que a pesar de que el muestreo ha producido una línea ajustada con un valor distinto a cero para  $b_1$ , se debe concluir que no hay evidencia suficiente que indique que  $Y$  está relacionada con  $X$ . Es decir, no se puede descartar la posibilidad de que la recta de regresión de la población es plana (horizontal).<sup>9</sup>

¿Cómo puede  $\beta_1$  ser cero mientras que  $b_1$  es distinto a cero? Considere la figura 6.7 donde se muestran los puntos que representan datos de la población de la cual se tomó una muestra de cinco (los datos de la muestra se indican con una  $X$ ). Como sugiere este diagrama de dispersión, si se seleccionan suficientes puntos de datos de la muestra, será obvio que los puntos de datos de la población tendrán una recta de regresión con una pendiente de cero. Sin embargo, en la muestra aleatoria de cinco puntos de datos se seleccionaron puntos que descansan bastante cerca de una línea de regresión que tiende a subir. Con esta evidencia, es posible concluir erróneamente que  $X$  y  $Y$  se relacionan linealmente de manera positiva. No obstante, si la hipótesis  $\beta_1 = 0$  se pone a prueba con los datos de la muestra, seguramente el analista no podrá rechazarla.

Si  $H_0: \beta_1 = 0$  es verdadero, la prueba estadística  $t$  con el valor  $t = \frac{b_1}{s_{b_1}}$  tiene una distribución  $t$  con  $df = n - 2$  grados de libertad. Aquí,  $s_{b_1}$  es la desviación estándar estimada (o error estándar) de  $b_1$ , determinado por  $s_{b_1} = s_{\text{res}} / \sqrt{\sum(X - \bar{X})^2}$ .

Este resultado proporciona una manera de probar  $H_0: \beta_1 = 0$ , como lo ilustra el siguiente ejemplo.

<sup>9</sup> Una línea horizontal de regresión de la población (por ejemplo,  $\beta_1 = 0$ ) también es un equivalente a la afirmación  $H_0: \rho = 0$ , donde  $\rho$  es el coeficiente de correlación de la población.

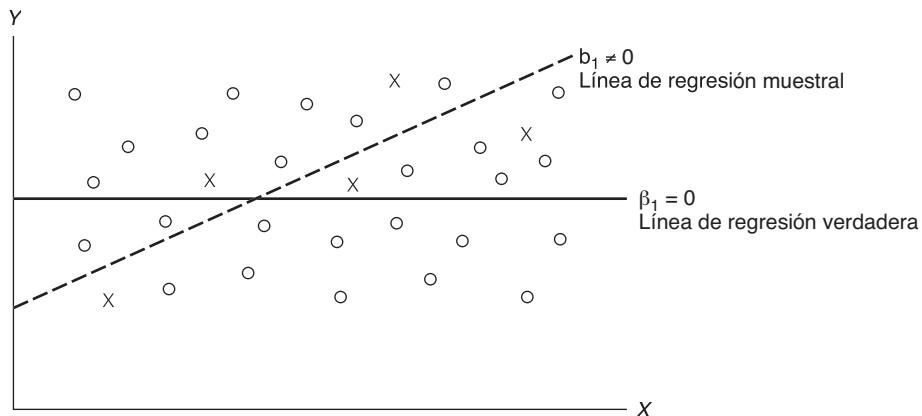


FIGURA 6.7 Datos de la población y de la muestra

**Ejemplo 6.8**

El señor Bump quiere probar

$$H_0: \beta_1 = 0 \text{ (no existe relación lineal)}$$

contra

$$H_1: \beta_1 \neq 0 \text{ (una relación lineal con pendiente diferente de cero)}$$

Primero calcula<sup>10</sup>

$$s_{b_1} = s_{y \cdot x} / \sqrt{\sum (X - \bar{X})^2} = 2.72 / \sqrt{.824} = 3.00$$

y luego forma el cociente (estadística de prueba)

$$t = \frac{b_1}{s_{b_1}} = \frac{-14.54}{3.00} = -4.8$$

¿Es el valor  $t = -4.8$  un resultado inusual si  $H_0$  es verdadera? El señor Bump verifica la tabla  $t$  con  $n - 2 = 8$  grados de libertad y encuentra que

$$\cdot 0.005 = 3.355$$

$$\cdot 0.025 = 2.306$$

Puesto que  $|t| = 4.8 > 3.355$ , el señor Bump rechaza  $H_0$  al nivel 1% de significancia. El señor Bump concluye que su regresión tiene significancia porque  $t = -4.8$  es altamente improbable si  $H_0$  es verdadera. Un valor  $t$  de esta magnitud ocurrirá menos de una vez en 100 si no existiera relación entre  $Y$  y  $X$ .

Para muestras de gran tamaño, es posible rechazar  $H_0$  y concluir que existe una relación lineal entre  $Y$  y  $X$  aunque  $r^2$  pueda ser menor, por ejemplo, 10%. Del mismo modo, para muestras pequeñas y un  $r^2$  muy grande, como de 95%, es posible concluir que la regresión es significativa. Un  $r^2$  pequeño significa que es improbable que la ecuación de regresión adaptada tenga mucho poder de predicción. Por otro lado, un  $r^2$  alto con una

<sup>10</sup> Los valores utilizados en el siguiente cálculo fueron hechos con anterioridad en este capítulo.

muestra muy pequeña puede dejar al analista insatisfecho y requiere más evidencia muestral antes de poder utilizar la función ajustada para realizar un pronóstico. Puede existir una diferencia entre la relevancia estadística y la relevancia práctica. Algunas veces se requiere de criterio y conocimiento del tema para determinar si la función de regresión ajustada va a ser una herramienta de pronóstico útil.

Una prueba alterna de  $H_0: \beta_1 = 0$  está en la tabla ANOVA.

Si los supuestos del modelo estadístico de regresión lineal son apropiadas y si la hipótesis nula  $H_0: \beta_1 = 0$  es verdadera, el cociente

$$F = \frac{\text{Cuadrado promedio de la regresión}}{\text{Error cuadrático medio}} = \frac{MSR}{MSE} \quad (6.15)$$

tiene una distribución  $F$  con  $df = 1, n - 2$  grados de libertad del numerador y del denominador, respectivamente. Cuando  $H_0$  es verdadera,  $MSR$  y  $MSE$  son, cada uno, estimadores de  $\sigma^2$ , la varianza del término de error  $\epsilon$  en el modelo estadístico de regresión de línea recta. Por otro lado, si  $H_1: \beta_1 \neq 0$  es verdadera, el numerador en el cociente que define a  $F$  tiende a ser más grande que el denominador. Los valores grandes  $F$ , son entonces, coherentes con la hipótesis alternativa.

En el modelo de regresión de línea recta, la prueba de la hipótesis  $H_0: \beta_1 = 0$  contra  $H_1: \beta_1 \neq 0$  se puede basar en la estadística  $F$  dada por  $F = \frac{MSR}{MSE}$  con  $df = 1, n - 2$  grados de libertad del numerador y del denominador, respectivamente. En el nivel de significancia  $\alpha$ , la región de rechazo de la hipótesis nula  $H_0: \beta_1 = 0$  es  $F > F_\alpha$ .

Como se analizará en el siguiente capítulo, la prueba  $F$  se puede extender para verificar la significancia en modelos de regresión con más de una variable independiente.

### Ejemplo 6.9

La tabla 6.7 es la tabla ANOVA para los datos del señor Bump. De dicha tabla,

$$F = \frac{MSR}{MSE} = \frac{174.19}{7.43} = 23.4$$

y con  $\delta_1 = 1$  y  $\delta_2 = 8$  grados de libertad

$$F_{0.05} = 5.32$$

$$F_{0.01} = 11.26$$

Puesto que  $F = 23.4 > 11.26$ ,  $H_0: \beta_1 = 0$  se rechaza en el nivel de 1%. Por lo tanto, a ese nivel la regresión es significativa.

No es accidental que los resultados de la prueba  $F$  sean coherentes con los resultados de la prueba  $t$  en el ejemplo 6.8. De hecho, dentro del redondeo del error,

$$F = 23.4 = (-4.8)^2 = t^2$$

Además,

$$F_{0.01} = 11.26 = (3.555)^2 = (t_{0.005})^2$$

Así, para un nivel de significancia determinado, la prueba  $t$  rechaza  $H_0: \beta_1 = 0$  cuando la prueba  $F$  rechaza y viceversa. Esta relación entre la  $t$  y la  $F$  solamente funciona para el modelo de regresión lineal con una sola variable  $X$ .

La estadística  $F$  puede expresarse también en términos del coeficiente de determinación  $r^2$ :

$$F = \frac{r^2(n - 2)}{1 - r^2} \quad (6.16)$$

Al ser igual todo lo demás, el valor de  $F$  se incrementa cuando  $r^2$  aumenta. Mientras más se explica la variación en  $Y$  por la función de regresión ajustada, la prueba  $F$  calificará a la regresión como más significativa.

## ANÁLISIS DE RESIDUALES

---

El hecho de ajustar un modelo por cuadrados mínimos, construir intervalos de predicción y probar las hipótesis no completa el estudio de la regresión. Estos pasos son solamente la mitad de la historia: las inferencias que se pueden hacer *cuando los supuestos del modelo se cumplen de manera adecuada*. En la mayoría de los estudios no es obvio que un modelo en particular sea correcto. Las inferencias pueden ser seriamente engañosas si los supuestos elaborados en la formulación del modelo son extremadamente incompatibles con los datos. Es esencial verificar cuidadosamente los datos para evitar violaciones de los supuestos. Recuerde que los supuestos para el modelo de regresión lineal simple son:

1. La relación entre las variables es lineal.
2. Los errores son independientes.
3. Los errores tienen varianza constante.
4. Los errores se distribuyen normalmente.

La información sobre la variación que no se puede explicar por la función de regresión ajustada se encuentra en los residuales  $e = Y - \hat{Y}$ . Para verificar los méritos de un modelo tentativo, se pueden examinar diversas gráficas de residuales.

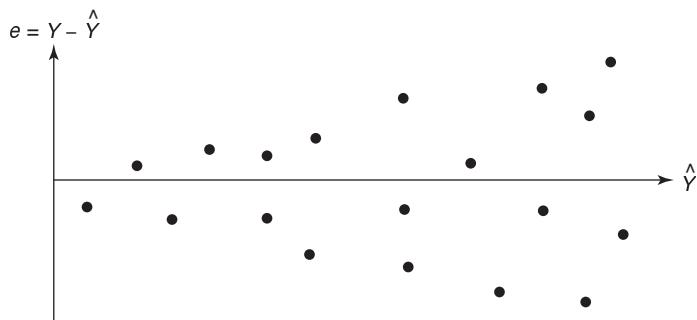
1. Gráfica como histograma de los residuales.
2. Gráfica de dispersión, residuales contra los valores ajustados.
3. Gráfica de dispersión, residuales contra la variable explicativa.
4. Gráfica de residuales contra el tiempo si los datos son cronológicos.

Un histograma de los residuales constituye una manera de verificar el supuesto de normalidad. En la práctica, las desviaciones moderadas de una curva normal o una curva en forma de campana no afectan las conclusiones en pruebas de hipótesis o en intervalos de confianza para la predicción cuando se basan en la distribución  $t$ , particularmente si las series de datos son grandes. Una violación del supuesto de normalidad, por sí sola, ordinariamente no es tan grave como una violación de cualquiera de los demás supuestos.

Si una gráfica de dispersión de residuales contra valores ajustados a la recta indica que la naturaleza general de la relación entre  $Y$  y  $X$  forma una curva en vez de una línea recta, una transformación adecuada de los datos puede reducir una relación no lineal a otra que se aproxime a la lineal. La transformación de variables se considera en una sección posterior de este capítulo.

Una transformación también puede ayudar a estabilizar la varianza. En la figura 6.8 hay una gráfica residual que indica que la distribución de los residuales aumenta cuando la magnitud de los valores ajustados se incrementa. Esto quiere decir que la variabilidad de los puntos de datos sobre la línea de mínimos cuadrados es mayor para valores ajustados de  $Y$  grandes que para valores pequeños, lo cual implica que el supuesto de varianzas constantes no se cumple. En esta situación, al relacionar el logaritmo de  $Y$  contra  $X$  podría producir una variación residual más coherente con el supuesto de varianza constante.

El supuesto de independencia es el más importante. La falta de independencia puede tergiversar drásticamente las conclusiones extraídas de las pruebas  $t$ . Este supuesto corre el riesgo de ser violado particularmente para datos de series de tiempo, que son frecuentes en problemas de pronóstico económico o pronóstico de negocios.



**FIGURA 6.8** Gráfica en que se observa un incremento en el tamaño de los residuales

Para residuales de series de tiempo, es decir, residuales producidos utilizando métodos de regresión en datos ordenados por tiempo, se puede verificar la independencia mediante una gráfica de los residuales sobre el tiempo. Si hay independencia no debe haber un patrón sistemático, tal como un conjunto de valores consecutivos altos seguido por otro de valores bajos. Además, cuando se calculan las autocorrelaciones muestrales de los residuales se puede verificar la independencia:

$$r_k(e) = \frac{\sum_{t=k+1}^n e_t e_{t-k}}{\sum_{t=1}^n e_t^2} \quad k = 1, 2, \dots, K \quad (6.17)$$

donde  $n$  es el número de residuales y  $K$  típicamente es  $n/4$ . Se indica independencia si los coeficientes residuales de autocorrelación son uniformemente pequeños (cada uno en el intervalo  $0 \pm 2/\sqrt{n}$ ) para todos los retrasos  $k$ .

#### Ejemplo 6.10

Al usar los valores ajustados y residuales mostrados en la tabla 6.6, el señor Bump elaboró un histograma de los residuales y una gráfica de los residuales contra los valores ajustados. Los resultados se pueden observar en la figura 6.9.

El histograma está centrado en cero y, aunque es simétrico, parece que no tiene forma de campana. No obstante, con sólo 10 observaciones, un histograma como el que se muestra en la figura 6.9 no es raro para los datos distribuidos normalmente. Al parecer, el supuesto normal es razonable.

Los puntos en la gráfica de probabilidad normal, ubicados en la primera fila de la figura 6.9, casi descansan sobre una línea recta. Como se dijo en el capítulo 2, este comportamiento similar al de la línea recta sugiere una buena coherencia entre los datos (en este caso, los residuales) y una distribución normal. La gráfica de probabilidad normal sugiere que no existe ninguna razón para dudar el supuesto normal.

La segunda gráfica en la primera fila de la figura 6.9 también parece ser adecuada. Cuando se traza una gráfica de los residuales contra los valores ajustados, la distribución de puntos en la dirección vertical con centro en el cero debe ser aproximadamente la misma para todos los valores a lo largo del eje horizontal. En otras palabras, las magnitudes de los residuales para valores ajustados pequeños deben ser aproximadamente las mismas que las magnitudes de los residuales para valores ajustados intermedios y casi las mismas que las magnitudes de los residuales para valores ajustados grandes. Este comportamiento ideal sugiere dos cosas: (1) la relación fundamental entre  $Y$  y  $X$  es lineal, y (2) la variabilidad del error es constante (las  $Y$  en diferentes valores de  $X$  tienen la misma distribución sobre la línea de regresión).

El señor Bump está contento porque la gráfica de los residuales contra los valores ajustados no está “arqueada”, por ejemplo, no ocurre que una cadena de residuales positivos esté seguida por otra de residuales negativos, y a su vez, seguida por otra cadena de residuales positivos. Este

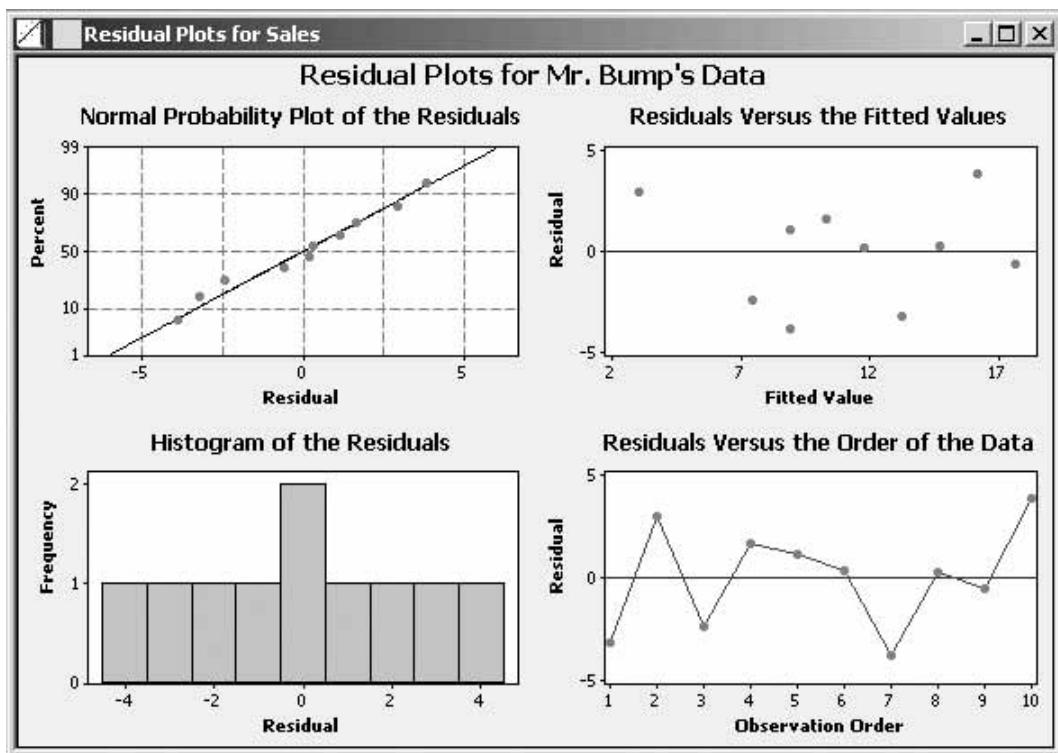


FIGURA 6.9 Gráfica de residuales para los datos de Bump

comportamiento sugeriría una relación no lineal entre  $Y$  y  $X$ . También se siente satisfecho porque su gráfica no tenía la apariencia en forma de cono de la gráfica en la figura 6.8 que indica variabilidad inconstante (creciente).

Aunque las  $Y$  son ventas semanales de leche, las semanas se eligieron al azar y no fueron ordenadas en el tiempo. Por consiguiente, no era apropiado graficar los residuales sobre tiempo o calcular las autocorrelaciones residuales.

El señor Bump está complacido con su análisis residual. Siente que su modelo de regresión de líneas describe adecuadamente la relación entre los volúmenes semanales de venta y el precio.

## RESULTADOS EN COMPUTADORA AL USAR MINITAB

El problema de análisis de regresión del señor Bump (los datos de la tabla 6.1) se realiza en Minitab (para instrucciones, vea la sección “Aplicaciones de Minitab” al final de este capítulo) y el reporte de los resultados se presenta en la tabla 6.8.

Para explicar la terminología utilizada en este reporte de resultados, en la siguiente lista se presentan definiciones y cálculos que aparecen en la tabla 6.8.

1. Correlación =  $-0.86$ . El coeficiente de correlación de muestra ( $r$ ) señala la relación entre  $X$  y  $Y$ , o precio y ventas, respectivamente.
2. Coeficiente de regresión (Coef) =  $-14.54$ . Este valor ( $b_1$ ) es el cambio en  $Y$  (ventas) cuando  $X$  (precio) aumenta en una unidad. Cuando el precio aumenta en \$1, las ventas estimadas disminuyen 14 540 unidades.

**TABLA 6.8 Resultado en Minitab para los datos de Bump**

| Correlations (Pearson)                                  |                |                     |           |            |       |
|---------------------------------------------------------|----------------|---------------------|-----------|------------|-------|
| Y                                                       |                |                     |           |            |       |
| $X$ - .863 (1) and (12)                                 |                |                     |           |            |       |
| Regression Analysis                                     |                |                     |           |            |       |
| The regression equation is                              |                |                     |           |            |       |
| $Sales (Y) = 32.1 - 14.5 \text{ Price (X)}$ (5) and (2) |                |                     |           |            |       |
| Predictor                                               | Coef           | SE Coef             | T         | P          |       |
| Constant                                                | 32.136 (5)     | 4.409               | 7.29      | 0.000      |       |
| Price (X)                                               | -14.539 (2)    | 3.002 (3)           | -4.84 (4) | 0.001 (7)  |       |
| S=2.725 (6)                                             | R-Sq=74.6% (8) | R-Sq(adj)=71.4% (9) |           |            |       |
| Analysis of Variance                                    |                |                     |           |            |       |
| Source                                                  | DF             | SS                  | MS        | F          | P     |
| Regression                                              | 1              | 174.18              | 174.18    | 23.45 (13) | 0.001 |
| Residual Error                                          | 8              | 59.42 (10)          | 7.43 (6)  |            |       |
| Total                                                   | 9              | 233.60 (11)         |           |            |       |

3. El coeficiente del error estándar de la regresión (SE Coef) = 3.0. Este valor es la desviación estándar estimada del valor de coeficiente de regresión ( $b_1$ ).

$$s_{b_1} = s_{y,x} \sqrt{\sum(X - \bar{X})^2} = 2.725 \sqrt{.824} = 2.725/.908 = 3.002$$

4. El valor  $t$  calculado (T) = -4.84. El valor calculado  $t$  se usa para probar si el coeficiente de regresión de la población  $\beta_1$  es significativamente diferente de cero.

$$s_{b_1} = 3.002$$

$$t = \frac{b_1}{s_{b_1}} = \frac{-14.54}{3.002} = -4.84$$

5. Constante (Coef) = 32.14. Este valor es la intersección ( $b_0$ ), de la línea de regresión con el eje vertical  $Y$ . Por lo tanto, la ecuación de regresión completa es

$$\hat{Y} = 32.14 - 14.54X$$

6. El error estándar de la estimación (S) = 2.725. El error estándar de la estimación indica que los valores  $Y$  típicamente caen aproximadamente a 2.725 unidades de la línea de regresión.

$$s_{y,x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{SSE}{n - 2}} = \sqrt{MSE} = \sqrt{7.43} = 2.725$$

7. El valor  $p$  (P), 0.001, es la probabilidad de obtener un valor  $t$  tan grande como  $|t| = |-4.84| = 4.84$  por casualidad si  $H_0: \beta_1 = 0$  es verdadera. Ya que el valor  $p$  es extremadamente pequeño, se concluye que la pendiente de coeficiente de regresión es significativamente diferente de 0.

8.  $r$  cuadrado = 0.746. La línea de regresión ajustada explica el 74.6% de la varianza del volumen de venta.

$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{59.42}{233.60} = 1 - .254 = .746$$

9.  $r$  cuadrado ajustado = 0.714. El  $r^2$  se ajusta a los grados de libertad apropiados.

$$\begin{aligned}\bar{r}^2 &= 1 - \frac{\Sigma(Y - \hat{Y})^2/(n - 2)}{\Sigma(Y - \bar{Y})^2/(n - 1)} = 1 - \frac{SSE/(n - 2)}{SST/(n - 1)} = 1 - \frac{59.42/8}{233.60/9} \\ &= 1 - \frac{7.428}{25/956} = 1 - .286 = .714\end{aligned}$$

10. Suma de los cuadrados (error) residuales = 59.42. La suma de cuadrados de los residuales es la suma de las desviaciones entre las  $Y$  reales y la  $Y$  pronosticada ( $\hat{Y}$ ) elevadas al cuadrado.

$$SSE = \Sigma(Y - \hat{Y})^2 = 59.42$$

11. Suma total de los cuadrados = 233.60. Este valor es la suma de las desviaciones cuadradas de las  $Y$  respecto a su valor promedio.

$$SST = \Sigma(Y - \bar{Y})^2 = 233.60$$

12. Matriz de correlación. Esta matriz proporciona el coeficiente de correlación entre todos los pares de variables en el análisis. Como sólo se involucran dos variables,  $X$  y  $Y$ , existe un solo coeficiente de correlación (-0.863).
13. Análisis de varianza y estadística  $F$ . El cociente de varianzas  $F$  ( $23.45 = 174.18/7.43$ ) en esta tabla ANOVA (análisis de varianza) prueba la hipótesis nula de que la regresión  $H_0: \beta_1 = 0$  no es significativa. Un valor  $F$  grande permitirá el rechazo de dicha hipótesis y será una evidencia de que la regresión es *significativa*. El valor  $F$  (23.45) se vuelve más grande, a medida que una mayor porción de la suma total de las desviaciones cuadradas ( $SST$ ) se explica por la regresión. En este caso, el valor tabulado  $F$  ( $df = 1, 8; \alpha = 0.01$ ) es 11.26. De este modo, la hipótesis de regresión no significativa se rechaza al nivel de significancia de 1% porque  $F = 23.45 > 11.26$  (vea el ejemplo 6.9).

## TRANSFORMACIÓN DE VARIABLES

---

Aunque el modelo de regresión lineal simple supone una relación de línea recta entre  $Y$  y  $X$ , en general, un modelo lineal de regresión se refiere a un modelo que es lineal en las  $\beta$  desconocidas. Mientras la función de regresión sea lineal en las  $\beta$  (por ejemplo,  $\beta_1^2$  no está presente), las variables de predicción (las  $X$ ) pueden tomar varias formas y la metodología de regresión lineal sigue siendo apropiada. Los modelos de regresión pueden ser usados para modelar relaciones complejas entre  $Y$  y  $X$  (o muchas  $X$ ) o para modelar una relación de línea recta entre  $Y$  y alguna función (transformación) de  $X$ .

Cuando un diagrama de dispersión indica que hay una relación no lineal entre  $Y$  y  $X$ , existen dos enfoques básicos para tratar este caso. El primero es ajustar los datos con una función de regresión que grafique como una curva y utilizar la relación ajustada para propósitos de pronóstico. El segundo enfoque involucra la transformación de la variable  $X$  en otra fórmula, de manera que la relación resultante con  $Y$  sea lineal.

Cuatro de las transformaciones (funciones) más comunes que se utilizan para generar nuevas variables explicativas son la recíproca, el logaritmo, la raíz cuadrada y el cuadrado:

$$\frac{1}{X}, \quad \log(X), \quad \sqrt{X}, \quad X^2$$

Cuando cada una de estas variables son graficadas contra  $Y$ , se espera que la relación no lineal entre  $Y$  y  $X$  se convierta en una relación lineal entre  $Y$  y una de las  $X$  transformadas. Si es así,  $Y$  y su nueva variable pueden ser tratadas con la aplicación del modelo de línea recta analizado en este capítulo, incluyendo el cálculo del coeficiente de correlación y la ecuación de regresión adaptada.

En el siguiente ejemplo se utiliza Minitab (vea “Aplicaciones en Minitab” al final del capítulo) para graficar una relación simple de  $X$  con  $Y$  que en apariencia es no lineal. Luego, se le ordena al programa que calcule las cuatro transformaciones descritas previamente. Enseguida, estas variables se grafican contra  $Y$  para producir la gráfica de datos mostrada.

### Ejemplo 6.11

Gilbert García es dueño de una cadena de ferreterías en Chicago, Illinois, y está interesado en pronosticar sus ventas mensuales en función de los gastos mensuales de publicidad. Gil sospecha que las ventas aumentarán a medida que aumente la cantidad gastada en publicidad. Sin embargo, también cree que después de cierto punto las ventas empezarán a aumentar a un ritmo más lento. Gil piensa que después de gastar cierta cantidad en publicidad, llegará un momento en que habrá pocas ganancias en ventas con más gastos en publicidad.

Gil tomó al azar una muestra de 14 meses de datos de compañías de discos. Los datos aparecen en la figura 6.10 y un diagrama de dispersión de los datos se muestra en la figura 6.11. Gil nota que las ventas parecen dejar de crecer después de haber gastado cierta cantidad en publicidad. Luego, ajusta la ecuación de regresión lineal mostrada en la figura 6.12 y percibe que la ecuación explica 77.7% de la variabilidad en ventas.

En la figura 6.13 se muestra una gráfica de los residuales contra los valores ajustados mediante la línea recta. Esta gráfica indica que una línea recta no describe adecuadamente la relación entre las ventas y los gastos de publicidad. Todos los residuales son negativos para los valores pronosticados más bajos y para los más altos, pero son positivos para valores centrales.

**FIGURA 6.10** Datos para el ejemplo 6.11

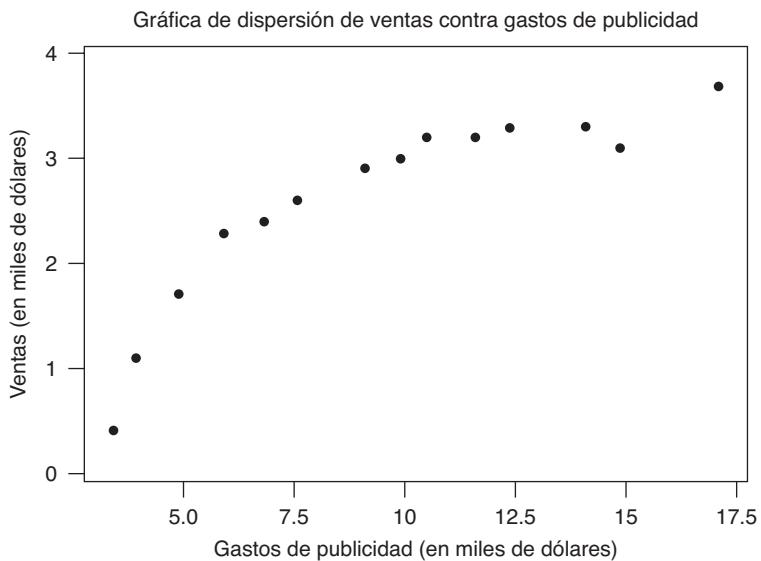


FIGURA 6.11 Diagrama de dispersión para el ejemplo 6.11

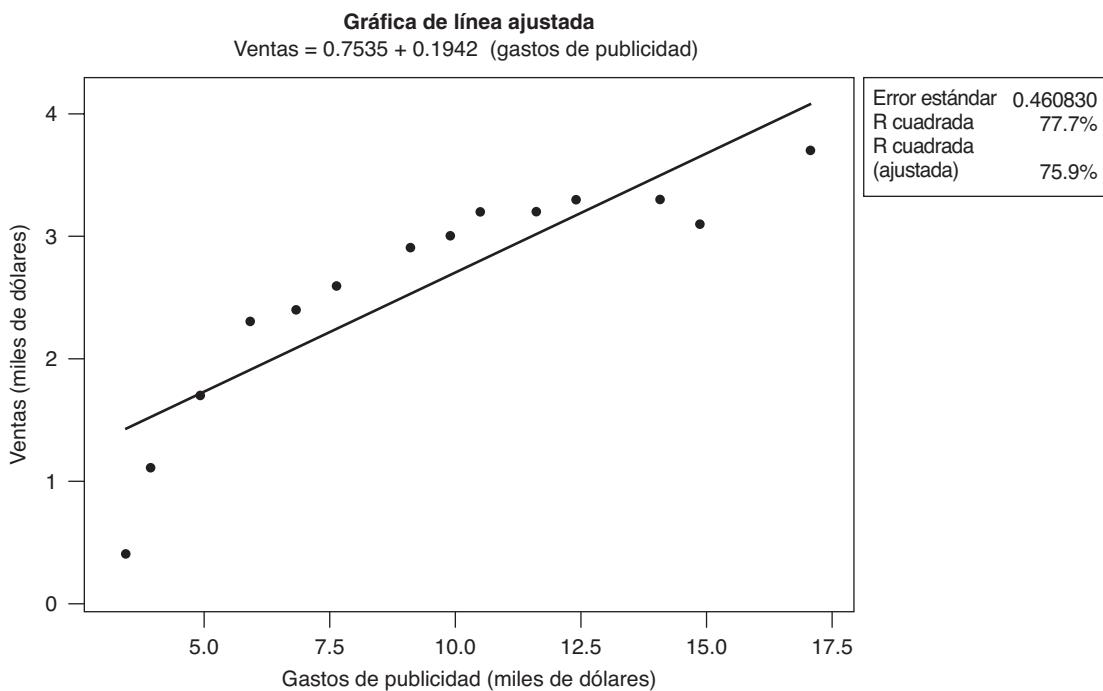
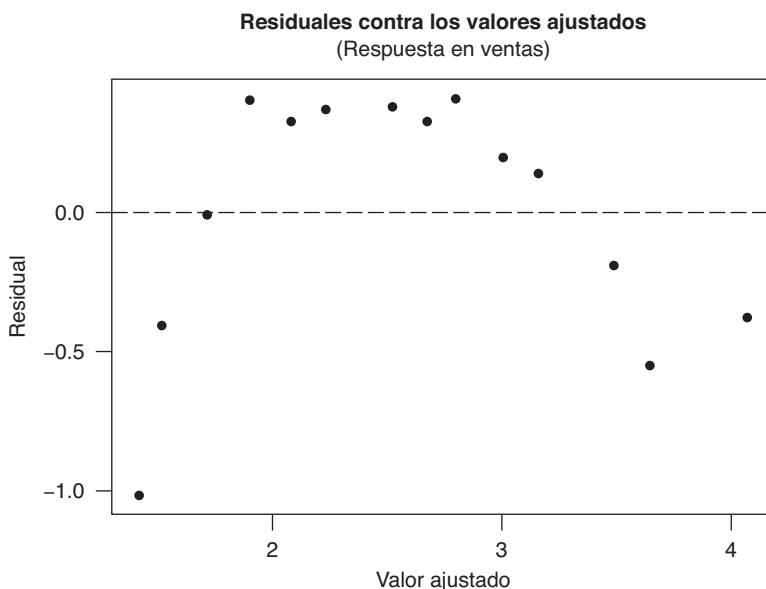


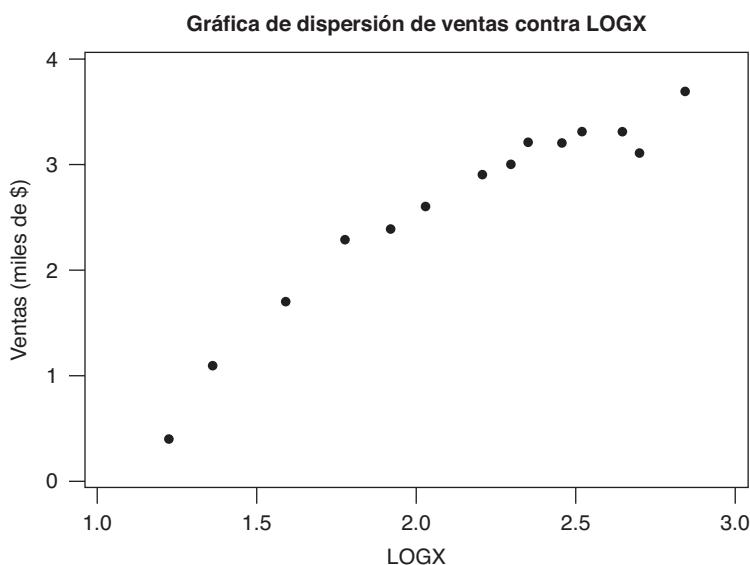
FIGURA 6.12 Línea de regresión ajustada para el ejemplo 6.11

Los residuales no están distribuidos de manera uniforme sobre la línea de regresión estimada. Se observa claramente que la línea recta no captura la curvatura en los datos.

A continuación, Gil transforma la variable de gastos de publicidad mediante logaritmo de  $X$  (vea la figura 6.10). Luego grafica la relación entre las ventas y el logaritmo de gastos de publicidad en la figura 6.14. La relación es aún ligeramente curvilínea.



**FIGURA 6.13** Gráfica de residuales contra valores ajustados mediante la línea recta

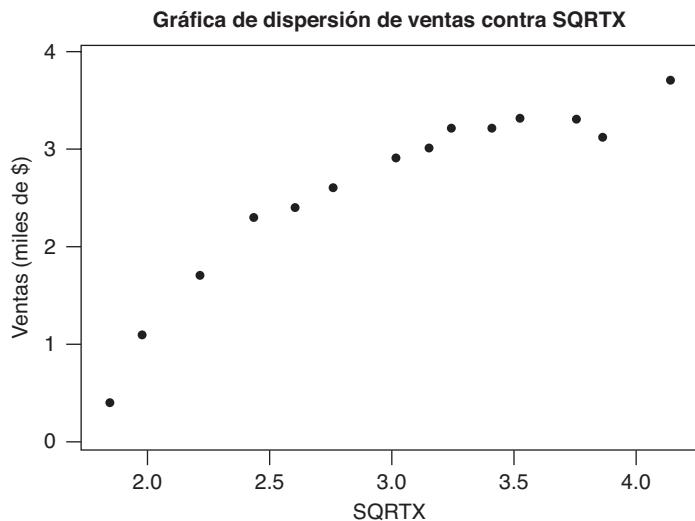


**FIGURA 6.14** Diagrama de dispersión contra el logaritmo de los gastos de publicidad de García

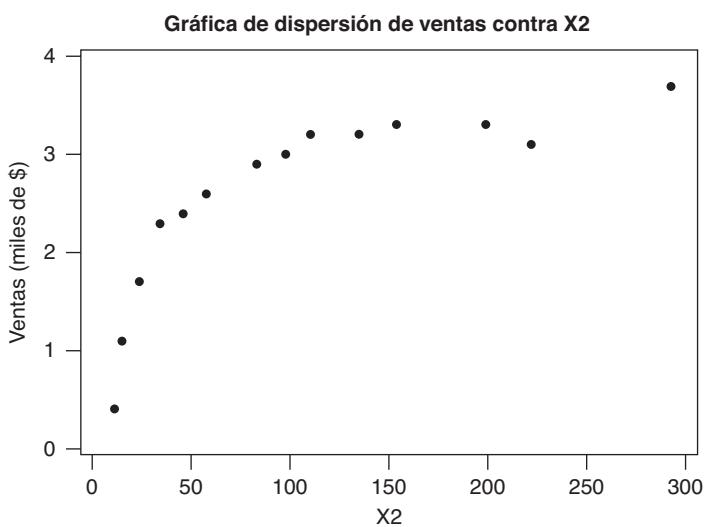
Gil transforma la variable de gastos de publicidad en la raíz cuadrada de  $X$  (vea la figura 6.10). Después de graficar esta relación, como se muestra en la figura 6.15, concluye que el ajuste aún no es lineal.

La figura 6.16 muestra la relación entre las ventas y el cuadrado de la variable de gastos de publicidad. De nuevo, no se ha encontrado una relación lineal.

Finalmente, Gil transforma la variable de gastos de publicidad en la recíproca de  $X$  (vea la figura 6.10). La figura 6.17 muestra que las ventas y la recíproca de gastos se relacionan lineal-



**FIGURA 6.15** Diagrama de dispersión para ventas contra raíz cuadrada de los gastos de publicidad de García



**FIGURA 6.16** Diagrama de dispersión para ventas contra el cuadrado de los gastos de publicidad de García

mente. Utilizando Minitab, determina la ecuación ajustada  $\hat{Y} = 4.29 - 12.7(1/X)$  como se muestra en la tabla 6.9, y observa que  $r^2$  es igual a 98.1%. Un análisis residual indica que el modelo apropiado es el que relaciona las ventas a la recíproca de gastos de publicidad.

## CURVAS DE CRECIMIENTO

Las curvas de crecimiento se presentaron en el capítulo 5 (vea la figura 5.7) cuando se habló sobre modelar la tendencia en la descomposición de las series de tiempo. A menudo, la tendencia (cambio de largo plazo) es de interés en sí misma. Por ejemplo, un pronosticador puede estar interesado en proyectar el costo a futuro por cada transacción de computadoras

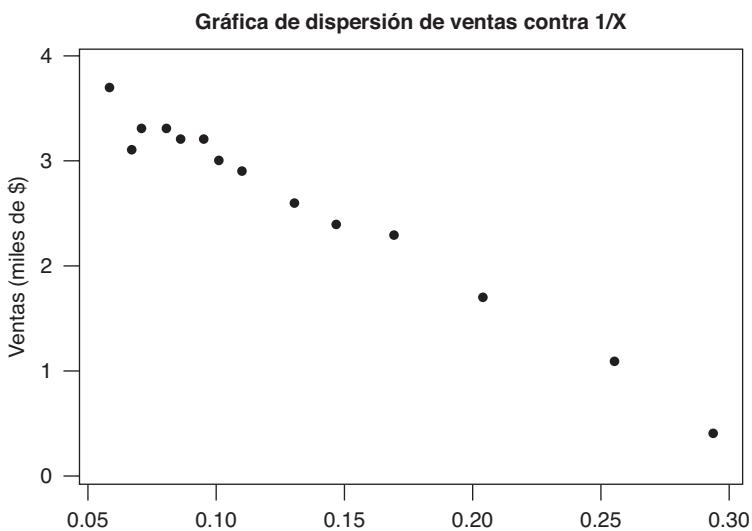


FIGURA 6.17 Diagrama de dispersión para ventas contra la recíproca de los gastos de publicidad de García

TABLA 6.9 Resultado en Minitab para el ejemplo 6.11

Regression Analysis: Sales versus  $1/X$

The regression equation is  
 $Sales = 4.29 - 12.7 \cdot 1/X$

| Predictor    | Coef            | SE Coef                     | T      | P     |
|--------------|-----------------|-----------------------------|--------|-------|
| Constant     | 4.28587         | 0.07695                     | 55.69  | 0.000 |
| $1/X$        | -12.7132        | 0.5092                      | -24.97 | 0.000 |
| $S = 0.1342$ | $R-Sq = 98.1\%$ | $R-Sq(\text{adj}) = 98.0\%$ |        |       |

Analysis of Variance

| Source         | DF | SS     | MS     | F      | P     |
|----------------|----|--------|--------|--------|-------|
| Regression     | 1  | 11.221 | 11.221 | 623.44 | 0.000 |
| Residual Error | 12 | 0.216  | 0.018  |        |       |
| Total          | 13 | 11.437 |        |        |       |

sin tomar en cuenta la tecnología que podría traer cambios en su costo. Otro ejemplo podría ser el de un analista de seguros de vida que se interese en proyecciones a largo plazo de la expectativa de vida de la población de Estados Unidos, desde el nacimiento y sin tomar en cuenta las condiciones económicas y ambientales que podrían cambiar las expectativas de vida.

Las curvas de crecimiento son relaciones curvilíneas entre una variable de interés y el tiempo. Es común que las curvas de crecimiento se ajusten a datos anuales cuando se requiere un pronóstico de largo plazo. Aunque es probable que se obtengan predicciones imprecisas cuando las curvas de crecimiento ajustadas a datos históricos son extrapoladas para predecir el futuro, este método de pronóstico puede ser benéfico para los administradores porque concentra la atención en los aspectos a largo plazo del negocio. Además, las curvas de crecimiento indican el índice de crecimiento anual que se debe mantener para alcanzar niveles proyectados a futuro. Este índice de crecimiento anual

puede, o no, ser razonable y se puede debatir en reuniones de ejecutivos o sesiones de “comités de expertos”.

Si una variable medida sobre el tiempo se incrementa en el mismo *porcentaje* cada periodo, se dice que exhibe un *crecimiento exponencial*. Si una variable aumenta en la misma cantidad cada periodo, se dice que exhibe un *crecimiento lineal*. A veces, una simple transformación convertirá a la variable con crecimiento exponencial en una variable con crecimiento lineal. Si éste es el caso, se pueden utilizar los métodos de regresión tratados en este capítulo para modelar el crecimiento exponencial.

Por ejemplo, suponga que una variable  $Y$ , medida anualmente, sigue la curva de tendencia exponencial (vea la ecuación 5.6):

$$Y = b_0 b_1^t$$

Aquí,  $100(b_1 - 1)\%$  es el incremento de porcentaje de cada año en  $Y$ . Al tomar los logaritmos de ambos lados de la tendencia de ecuación exponencial se obtiene

$$\log Y = \log(b_0 b_1^t) = \log b_0 + t \log b_1 = \tilde{b}_0 + \tilde{b}_1 t$$

De manera que  $\log Y$  contra tiempo ( $t$ ) se grafica como una línea recta con pendiente  $\tilde{b}_1 = \log(b_1)$  e intersección  $\tilde{b}_0 = \log(b_0)$ . El logaritmo  $Y$  se incrementa por una cantidad constante  $b_1$  de un año al siguiente. Como lo ilustra el siguiente ejemplo, se pueden obtener los pronósticos de  $Y$  y un estimado del índice de crecimiento  $b_1$  por medio de la ecuación de la recta que relaciona el logaritmo de  $Y$  con el tiempo  $t$ .

### Ejemplo 6.12

Jill Johnson, asistente administrativa de una importante universidad, está muy complacida por la aprobación de un sistema de registro de cursos por Internet. Ha recopilado los datos de la tabla 6.10 y observa que el crecimiento en el uso del sistema ha sido fenomenal. Si el índice de crecimiento actual continúa, en un año todos los estudiantes se registrarán en línea para los cursos y el viejo sistema de registro telefónico se podrá eliminar por completo.

Los datos de la tabla 6.10 muestran el número de alumnos que utilizan el sistema de registro de cursos por Internet para el periodo 1998-2002. La última columna de la tabla contiene los logaritmos en base 10 del número de usuarios.

Jill ha tenido alguna experiencia con la regresión lineal simple y se pregunta si podrá usar esta metodología para pronosticar el número de usuarios para 2003 para obtener un estimado del índice de crecimiento anual. Mediante una gráfica del número de usuarios contra en tiempo en años (vea la figura 6.18a) Jill rápidamente se da cuenta de que un modelo de regresión lineal simple no es el adecuado para sus datos. Usa el crecimiento exponencial con un incremento del porcentaje anual de alrededor de 70% por año. No obstante, el incremento en el número de usuarios de un año a otro es bastante diferente. Los incrementos en magnitud abarcan el rango que va desde 2 500 como el valor más bajo y que corresponde al periodo 1998-1999 a un valor alto de 10 000 del periodo 2001-2002.

Jill siempre ha sido buena manipulando logaritmos y recuerda que un cambio constante en el porcentaje de una variable sobre periodos equivale a un incremento constante en magnitud del logaritmo de la variable sobre el mismo periodo. Con lo anterior en mente, Jill transforma mediante logaritmos el número de usuarios que se registran en línea, y grafica el logaritmo del número de usuarios contra el tiempo en años (vea la figura 6.18b).

**TABLA 6.10 Datos para el ejemplo 6.12**

| Año  | Tiempo | Usuarios | Log(Usuarios) |
|------|--------|----------|---------------|
| 1998 | 1      | 3000     | 3.477         |
| 1999 | 2      | 5500     | 3.740         |
| 2000 | 3      | 9000     | 3.954         |
| 2001 | 4      | 16500    | 4.217         |
| 2002 | 5      | 26500    | 4.423         |

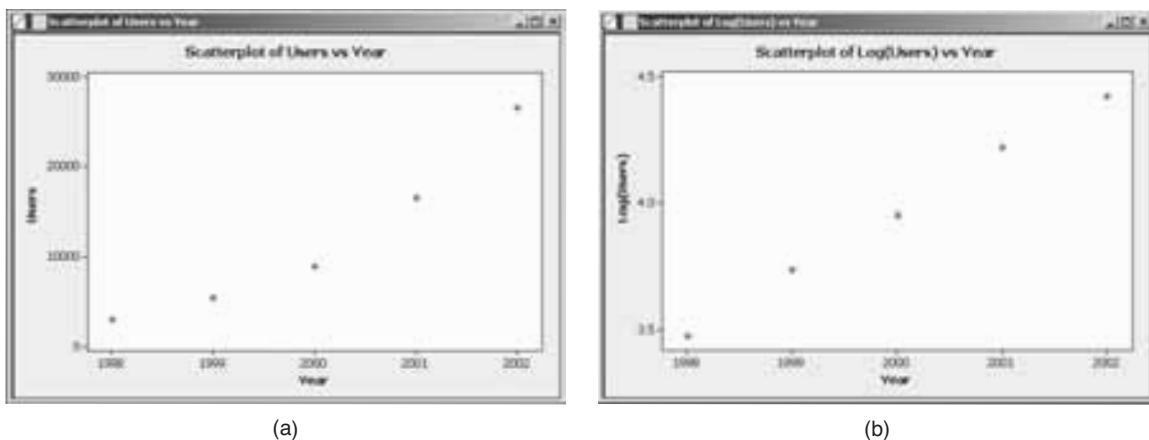


FIGURA 6.18 Diagramas de dispersión de los datos de Jill Johnson para el ejemplo 6.12

Como lo sospechó, cuando el número de usuarios ha sido transformado tomando sus logaritmos, el crecimiento exponencial se convierte en crecimiento lineal. La curvatura que representa el crecimiento exponencial en la figura 6.18(a) ha sido enderezada en la figura 6.18(b) por la transformación logarítmica. En la figura 6.18(b), el incremento en magnitud de logaritmo(usuarios) es casi el mismo año con año, alrededor de 0.23 por año. Jill ajusta una línea recta a los datos en la figura 6.18(b) con logaritmo(usuarios) como la variable dependiente  $Y$  y el tiempo ( $t$ ) como la variable independiente. Los resultados se muestran en la figura 6.19.

Mediante la ecuación ajustada  $\log \hat{Y} = 3.252 + 0.2369 t$  con  $t = 6$  para el año 2003, Jill calcula el pronóstico

$$\begin{aligned}\log \hat{Y}_{2003} &= 3.252 + .2369(6) = 4.673 \\ \hat{Y}_{2003} &= 10^{4.673} = 47,100\end{aligned}$$

Jill pronostica que 47 100 estudiantes usarán el sistema de registro por Internet en el 2003. El total de inscripciones de la universidad tiene un límite de 44 mil estudiantes, así que Jill concluye que es probable que todos los estudiantes usen el nuevo sistema y que el viejo sistema telefónico se puede eliminar.

Finalmente, Jill estima que el índice de crecimiento anual de usuarios del sistema será de

$$100(\text{antilog}(.2369) - 1)\% = 100(10^{.2369} - 1)\% = 100(1.725 - 1)\% = 72.5\%$$

Jill reconoce que con el límite de inscripciones de la universidad, no tiene sentido extrapolar el crecimiento de usuarios más allá del 2003.

En ciertas situaciones, la curva de crecimiento en línea recta se puede ajustar directamente a los datos originales.

### Ejemplo 6.13

Una compañía interesada en atender las necesidades de salud de las personas mayores desea realizar en Estados Unidos una proyección a largo plazo de la longevidad, desde el nacimiento, de tal manera que pueda formular una estrategia corporativa a largo plazo. La firma encuentra proyecciones para esta variable, de 1970 a 1997, en el *Statistical Abstract of the United States*. Estos datos se muestran en la tabla 9.11. Después de codificar los datos en  $1 = 1970, 2 = 1971, \dots, 28 = 1997$ , y teclear los datos en Minitab, la recta de regresión ajustada produce la gráfica mostrada en la figura 6.20.

La compañía no está interesada en examinar las razones del obvio aumento en la longevidad de años pasados, sino en proyectar la recta de regresión ajustada hacia el futuro. Tampoco desea especular sobre si la tendencia ascendente continuará, o por qué medios se puede lograr este aumento. Solamente quiere completar la siguiente afirmación: "Si las tendencias actuales continúan, en el futuro la expectativa de vida desde el nacimiento llegará a  $Y$  en  $X$  años." El objetivo a futuro escogido es el año 2050 ( $X = 53$  años después de 1997). Al emplear  $28 + 53 = 81$

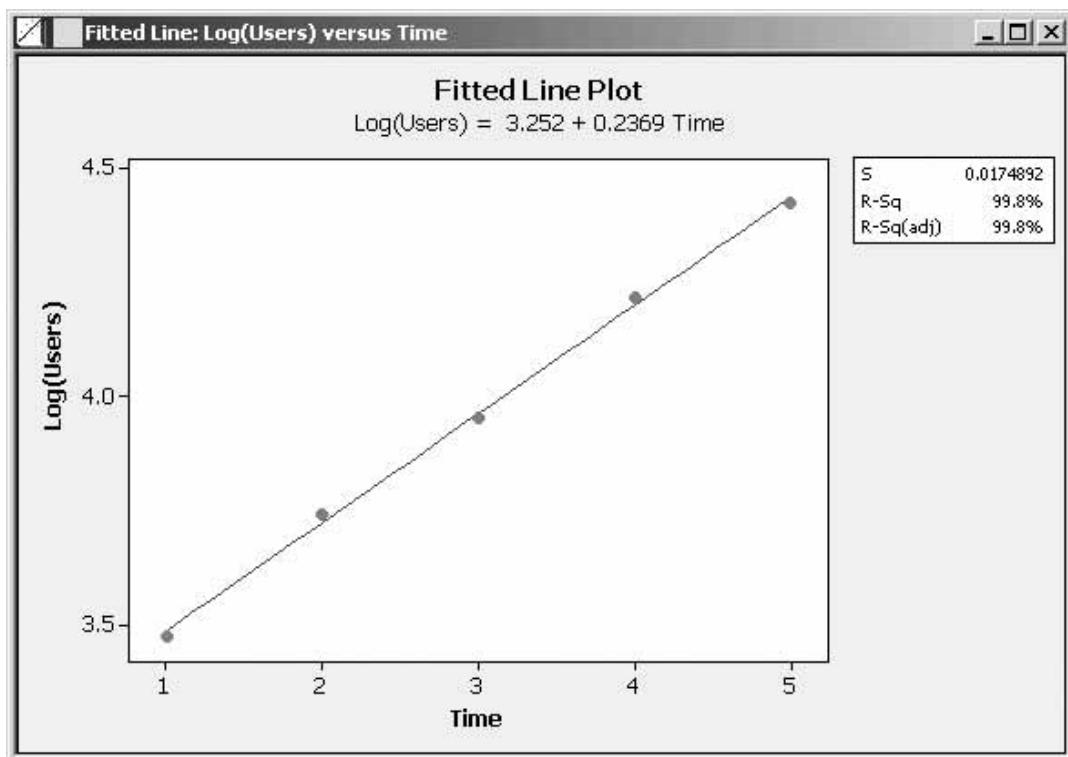


FIGURA 6.19 Línea de regresión ajustada para el ejemplo 6.12

**TABLA 6.11 Expectativa de vida al nacer, población total, Estados Unidos**

|      |       |      |      |
|------|-------|------|------|
| 1970 | 70.8  | 1984 | 74.7 |
| 1971 | 71.2* | 1985 | 74.7 |
| 1972 | 71.5* | 1986 | 74.7 |
| 1973 | 71.9* | 1987 | 74.9 |
| 1974 | 72.4* | 1988 | 74.9 |
| 1975 | 72.6  | 1989 | 75.1 |
| 1976 | 72.8* | 1990 | 75.4 |
| 1977 | 73.0* | 1991 | 75.5 |
| 1978 | 73.3* | 1992 | 75.8 |
| 1979 | 73.5* | 1993 | 75.5 |
| 1980 | 73.7  | 1994 | 75.7 |
| 1981 | 74.1* | 1995 | 75.8 |
| 1982 | 74.5  | 1996 | 76.1 |
| 1983 | 74.6  | 1997 | 76.5 |

\*Interpolado

Fuente: *Statistical Abstract of the United States*, edición 119, 1999.

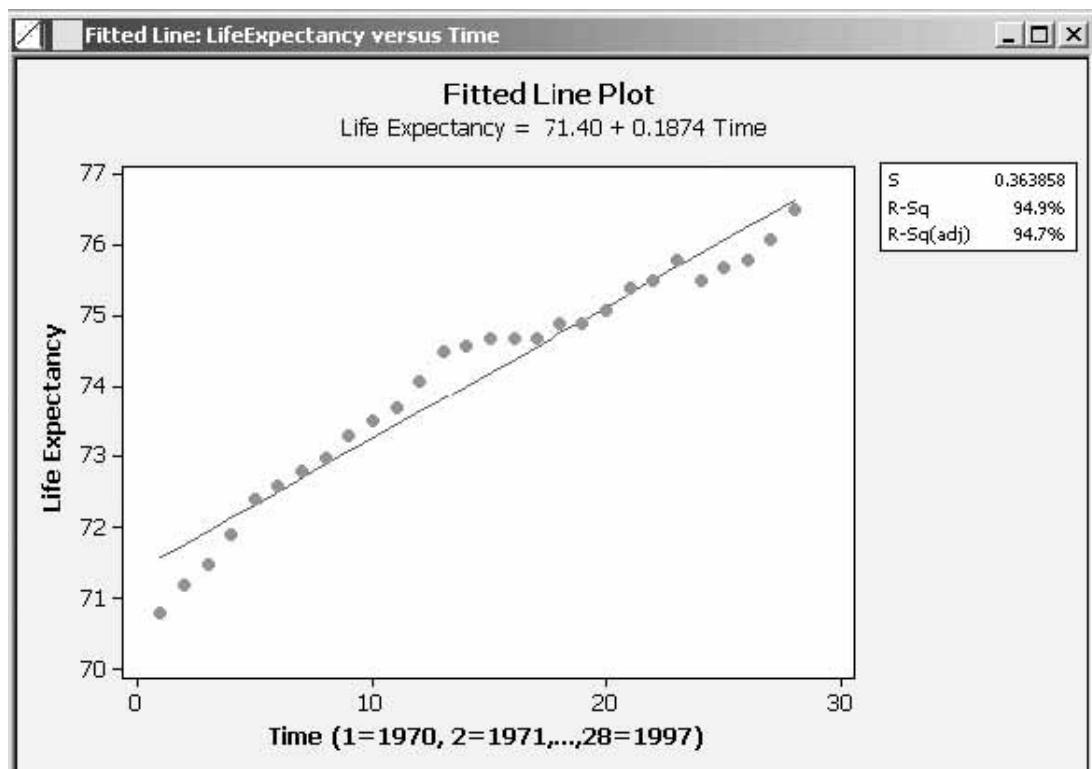


FIGURA 6.20 Línea de regresión ajustada para el ejemplo 6.13

como el valor para Tiempo en la ecuación de regresión mostrada en la figura 6.20 y al resolver para  $Y$ , la compañía obtiene una expectativa de vida de 86.6 años. Esta edad obviamente representa un aumento sustancial en longevidad para los próximos años y la compañía empieza a trabajar en planes estratégicos para tomar ventaja de este incremento en la expectativa de vida.

## APLICACIÓN A LA ADMINISTRACIÓN

El análisis de regresión es la herramienta estadística más utilizada por la administración cuando se requiere evaluar el impacto de una sola variable independiente en una variable dependiente. El análisis de regresión, junto con el análisis de correlación, ayuda al pronosticador a caracterizar las relaciones entre variables. Éste puede determinar tanto la significancia como la dirección de la relación entre variables.

Al utilizar el análisis de regresión, la mayoría de los problemas involucra una versión más sofisticada llamada *análisis de regresión múltiple* (el cual se abordará más adelante en este capítulo) porque la mayoría de las relaciones implica el estudio de la relación entre una variable dependiente y más de una variable independiente. Sin embargo, los análisis de regresión simple y de correlación a menudo son útiles. A continuación se presentan algunos ejemplos de situaciones que involucran la regresión múltiple:

- *Consumo del producto.* Un fabricante puede tratar de pronosticar cuánta cerveza bebe una persona por semana observando variables como ingreso, edad, educación y clase social.
- *Ventas.* Un minorista puede tratar de pronosticar las ventas de un producto para una tienda en contraste con otra, con base en las diferencias de precios, los ingresos relati-

vos de la comunidad, la amabilidad del personal de la tienda y el número y fuerza de los competidores en cada mercado.

- *Precios de las acciones.* Un analista accionario de una firma regional de corretaje puede tratar de pronosticar el precio de una nueva emisión para una firma local basándose en la economía regional, ingresos, población y posición de la firma.
- *Deudas incobrables.* Un contador puede tratar de pronosticar las deudas incobrables con las que se puede enfrentar una empresa en el siguiente trimestre fiscal, con base en el número de personas desempleadas, créditos pendientes, tasas de interés y ventas esperadas.
- *Necesidades de empleo.* El director de personal de una importante fábrica puede tratar de pronosticar los requerimientos de personal para el año próximo tomando como base la edad promedio de sus empleados, la escala de ingresos comparada con los de la localidad, nuevos contratos de venta anticipados y la existencia de empleos competitivos.
- *Demanda de un centro comercial.* El gerente de un nuevo centro comercial puede tratar de anticipar la demanda mediante el análisis del ingreso de la localidad, el tamaño de la población y la proximidad y tamaño de los centros comerciales competidores.

Una vez que se establece la relación entre las variables independiente y dependiente, la administración puede, en algunos casos, tratar de controlar la variable dependiente con este conocimiento. Por ejemplo, suponga que un gerente de marketing determina que existe una relación positiva relevante entre los gastos de publicidad y las ventas. La ecuación de regresión podría ser

$$\text{Ventas} = \$43,000 + 0.3 (\text{gastos de publicidad})$$

Con esta ecuación, el gerente de marketing puede tratar de controlar las ventas al aumentar o disminuir la publicidad en la cantidad que podría maximizar las ganancias. Siempre que el gerente pueda controlar la variable independiente, existe la posibilidad de un control parcial de la variable dependiente. La ecuación de regresión y el coeficiente de determinación ayudan a la administración a establecer si tal control es o no valioso.

## Glosario

**Coeficiente de determinación.** El coeficiente de determinación mide el porcentaje de la variabilidad en  $Y$  que se puede explicar con el conocimiento de la variabilidad (diferencias) en la variable independiente  $X$ .

**Error estándar de estimación.** El error estándar de estimación mide la cantidad por la que los valores reales  $Y$  difieren de los valores estimados  $\hat{Y}$ . Es una

estimación de la desviación estándar del término de error  $\epsilon$  en el modelo de regresión lineal simple.

**Línea de regresión ajustada.** Una línea de regresión ajustada es la que mejor ajusta una colección de puntos de datos  $X-Y$ . Minimiza la suma de las distancias elevadas al cuadrado desde los puntos hasta la recta y medidas en dirección vertical, o  $Y$ .

## Fórmulas clave

### Método de mínimos cuadrados: fórmula de la pendiente

$$b_1 = \frac{n \Sigma XY - \Sigma X \Sigma Y}{n \Sigma X^2 - (\Sigma X)^2} = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (X - \bar{X})^2} \quad (6.2)$$

### Método de mínimos cuadrados: fórmula para la intersección con eje $Y$

$$b_0 = \frac{\Sigma Y}{n} - \frac{b_1 \Sigma X}{n} = \bar{Y} - b_1 \bar{X} \quad (6.3)$$

**Relación entre la pendiente de la línea de regresión y el coeficiente de correlación**

$$b_1 = \frac{\sqrt{\sum(Y - \bar{Y})^2}}{\sqrt{\sum(X - \bar{X})^2}} r \quad (6.4)$$

**Ecuación de regresión ajustada**

$$\hat{Y} = b_0 + b_1 X \quad (6.6)$$

**Error estándar de la estimación: fórmula que lo define**

$$s_{y \cdot x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}} \quad (6.7)$$

**Error estándar de la estimación: fórmula para calcularlo**

$$s_{y \cdot x} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum XY}{n - 2}} \quad (6.8)$$

**Error estándar del pronóstico**

$$s_f = s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2}} \quad (6.9)$$

**Intervalo para la predicción**

$$\hat{Y} \pm t s_f \quad (6.10)$$

**Intervalo para la predicción del 95% para muestra grande**

$$\hat{Y} \pm 2 s_f \quad (6.11)$$

**Descomposición de la suma de cuadrados y grados de libertad**

$$SST = SSR + SSE \quad (6.12)$$

$$\Sigma(Y - \bar{Y})^2 = \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2 \quad (6.12)$$

$$df: n - 1 = 1 + (n - 2) \quad (6.13)$$

**Coeficiente de determinación**

$$r^2 = \frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2} = 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2} \quad (6.14)$$

**Estadística  $t$  para probar  $H_0: \beta_1 = 0$** 

$$t = \frac{b_1}{s_{b_1}} \quad (\text{vea el recuadro en pantalla, p. 226})$$

**Error estándar para el coeficiente de regresión**

$$s_{b_1} = s_{y \cdot x} / \sqrt{\sum(X - \bar{X})^2} \quad (\text{vea el recuadro en pantalla, p. 226})$$

**Estadística  $F$** 

$$F = \frac{\text{Cuadrado promedio de la regresión}}{\text{Error cuadrático medio}} = \frac{MSR}{MSE} \quad (6.15)$$

**Relación de la estadística  $F$  con el coeficiente de determinación**

$$F = \frac{r^2(n - 2)}{1 - r^2} \quad (6.16)$$

**Coefficiente de autocorrelación de los residuales**

$$r_k(e) = \frac{\sum_{t=k+1}^n e_t e_{t-k}}{\sum_{t=1}^n e_t^2} \quad k = 1, 2, \dots, K \quad (6.17)$$

**Problemas**

*Nota:* la mayoría de los siguientes problemas contiene datos que se deben manipular mediante los procedimientos de análisis de regresión. Aunque es posible, y hasta útil, resolver uno o dos a mano, es importante que usted aprenda a utilizar el software de la computadora para resolver dichos problemas. En el siguiente capítulo aprenderá sobre el análisis de regresión múltiple, en el cual no es posible resolver los problemas a mano. Debe familiarizarse con el software de análisis de regresión mientras resuelve los problemas. Si tiene acceso a Excel o Minitab, vaya a la sección que está al final de este capítulo para ver las instrucciones de su uso.

1. ¿Cuál de las siguientes situaciones es inconsistente?
  - a)  $\hat{Y} = 499 + 0.21X$  y  $r = 0.75$
  - b)  $\hat{Y} = 100 + 0.9X$  y  $r = -0.70$
  - c)  $\hat{Y} = -20 + 1X$  y  $r = 0.40$
  - d)  $\hat{Y} = -7 - 4X$  y  $r = -0.90$
2. Las ganancias de AT&T (American Telephone and Telegraph) en miles de millones de dólares se estiman mediante el GNP (*gross national product*, Producto Nacional Bruto). La ecuación de regresión es  $\hat{Y} = 0.078 + 0.06X$  donde el GNP se mide en miles de millones de dólares.
  - a) Interprete la pendiente.
  - b) Interprete la intersección con el eje  $Y$ .
3. Considere los datos en la tabla P.3, donde  $X$  = gastos semanales de publicidad e  $Y$  = ventas semanales.
  - a) ¿Existe una relación significativa entre los gastos de publicidad y las ventas?
  - b) Determine la ecuación de predicción.
  - c) Pronostique las ventas para un gasto de publicidad de 50 dólares.
  - d) ¿Qué porcentaje de la variación en ventas se puede explicar con la ecuación de predicción?
  - e) Mencione la cantidad de la variación no explicada.

TABLA P.3

| <b><i>Y</i></b> | <b><i>X</i></b> | <b><i>Y</i></b> | <b><i>X</i></b> |
|-----------------|-----------------|-----------------|-----------------|
| \$1250          | \$41            | \$1300          | \$46            |
| 1380            | 54              | 1400            | 62              |
| 1425            | 63              | 1510            | 61              |
| 1425            | 54              | 1575            | 64              |
| 1450            | 48              | 1650            | 71              |

TABLA P.4

| <i>Tiempo requerido para pagar en caja (minutos)</i> | <i>Valor de la compra (dólares)</i> | <i>Tiempo requerido para pagar en caja (minutos)</i> | <i>Valor de la compra (dólares)</i> |
|------------------------------------------------------|-------------------------------------|------------------------------------------------------|-------------------------------------|
| 3.6                                                  | 30.6                                | 1.8                                                  | 6.2                                 |
| 4.1                                                  | 30.5                                | 4.3                                                  | 40.1                                |
| .8                                                   | 2.4                                 | .2                                                   | 2.0                                 |
| 5.7                                                  | 42.2                                | 2.6                                                  | 15.5                                |
| 3.4                                                  | 21.8                                | 1.3                                                  | 6.5                                 |

f) Obtenga la cantidad de la variación total.

4. El tiempo requerido para que los clientes paguen en el supermercado y los valores correspondientes de las compras se muestran en la tabla P.4. Conteste los incisos a), b), e) y f) del problema 3 utilizando estos datos. Proporcione un estimado puntual y un intervalo estimado de 99% para  $Y$  si  $X = 3.0$ .
5. Lori Franz, supervisora de mantenimiento de Baltimore Transit Authority, quisiera determinar si existe una relación positiva entre los costos anuales de mantenimiento de un autobús y su edad. Si existe tal relación, Lori cree que puede hacer un mejor trabajo si pronostica el presupuesto anual para el autobús. Ella recopiló los datos de la tabla P.5.
  - a) Grafique un diagrama de dispersión.
  - b) ¿Qué clase de relación existe entre estas dos variables?
  - c) Calcule el coeficiente de correlación.
  - d) Determine la recta de mínimos cuadrados.
  - e) Pruebe la significancia de la pendiente de la regresión en el nivel de significancia 0.05. ¿Esta regresión es significativa? Explique.
  - f) Pronostique el costo anual del mantenimiento de un autobús de 5 años.

TABLA P.5

| <i>Autobús</i> | <i>Costo del mantenimiento (\$)</i> |          | <i>Edad (años)</i> |
|----------------|-------------------------------------|----------|--------------------|
|                | <i>Y</i>                            | <i>X</i> |                    |
| 1              | 859                                 |          | 8                  |
| 2              | 682                                 |          | 5                  |
| 3              | 471                                 |          | 3                  |
| 4              | 708                                 |          | 9                  |
| 5              | 1094                                |          | 11                 |
| 6              | 224                                 |          | 2                  |
| 7              | 320                                 |          | 1                  |
| 8              | 651                                 |          | 8                  |
| 9              | 1049                                |          | 12                 |

TABLA P.6

| <i>Semana</i> | <i>Número de libros vendidos</i> | <i>Espacio en pies de los anaqueles</i> |
|---------------|----------------------------------|-----------------------------------------|
|               | <i>Y</i>                         | <i>X</i>                                |
| 1             | 275                              | 6.8                                     |
| 2             | 142                              | 3.3                                     |
| 3             | 168                              | 4.1                                     |
| 4             | 197                              | 4.2                                     |
| 5             | 215                              | 4.8                                     |
| 6             | 188                              | 3.9                                     |
| 7             | 241                              | 4.9                                     |
| 8             | 295                              | 7.7                                     |
| 9             | 125                              | 3.1                                     |
| 10            | 266                              | 5.9                                     |
| 11            | 200                              | 5.0                                     |

TABLA P.7

| <i>Ciudad</i> | <i>Número de órdenes postales recibidas (miles)</i> | <i>Número de catálogos repartidos (miles)</i> | <i>Ciudad</i> | <i>Número de órdenes postales recibidas (miles)</i> | <i>Número de catálogos repartidos (miles)</i> |
|---------------|-----------------------------------------------------|-----------------------------------------------|---------------|-----------------------------------------------------|-----------------------------------------------|
|               | <i>Y</i>                                            | <i>X</i>                                      |               | <i>Y</i>                                            | <i>X</i>                                      |
| A             | 24                                                  | 6                                             | G             | 18                                                  | 15                                            |
| B             | 16                                                  | 2                                             | H             | 18                                                  | 3                                             |
| C             | 23                                                  | 5                                             | I             | 35                                                  | 11                                            |
| D             | 15                                                  | 1                                             | J             | 34                                                  | 13                                            |
| E             | 32                                                  | 10                                            | K             | 15                                                  | 2                                             |
| F             | 25                                                  | 7                                             | L             | 32                                                  | 12                                            |

6. Andrew Vazsonyi es el gerente de la cadena de supermercados Spendwise, a quien le gustaría poder pronosticar las ventas de libros (ejemplares por semana) basándose en la cantidad de espacio de exhibición (en pies) de los anaqueles. Andrew reúne los datos para una muestra de 11 semanas, como se mostró en la tabla P.6.
- Grafeque un diagrama de dispersión.
  - ¿Qué clase de relación existe entre estas dos variables?
  - Calcule el coeficiente de correlación.
  - Determine la línea de mínimos cuadrados.
  - Pruebe la significancia de la pendiente de correlación en el nivel de significancia de 0.10. ¿Esta correlación es significativa? Explique.
  - Grafeque los residuales contra los valores ajustados. Basándose en esta gráfica, ¿es apropiado el modelo de regresión lineal simple para estos datos?
  - Pronostique las ventas de libros para una semana durante la cual se proveen 4 pies de espacio en anaqueles.
7. En la tabla P.7 se muestra la información proporcionada por un negocio de órdenes por correo para 12 ciudades.
- Determine si existe una relación lineal significativa entre estas dos variables. (Pruébelo con un nivel de 0.05 de significancia.)
  - Determine la línea de regresión ajustada.
  - Calcule el error estándar de la estimación.
  - Elabore una tabla ANOVA.
  - ¿Qué porcentaje de la variación en las órdenes por correo se explica por el número de catálogos repartidos?

TABLA P.8

| <i>Inversión anual<br/>(miles de dólares)</i> | <i>Tasa promedio<br/>de inversión (%)</i> | <i>Inversión anual<br/>(miles de dólares)</i> | <i>Tasa promedio<br/>de inversión (%)</i> |
|-----------------------------------------------|-------------------------------------------|-----------------------------------------------|-------------------------------------------|
| 1060                                          | 4.8                                       | 2050                                          | 3.8                                       |
| 940                                           | 5.1                                       | 2070                                          | 3.7                                       |
| 920                                           | 5.9                                       | 2030                                          | 4.5                                       |
| 1110                                          | 5.1                                       | 1780                                          | 4.9                                       |
| 1590                                          | 4.8                                       | 1420                                          | 6.2                                       |

- f) Realice la prueba de hipótesis para determinar si la pendiente o coeficiente de regresión es significativamente diferente de cero. (Utilice el nivel 0.01 de significancia.)
- g) Pruebe la significancia de la regresión usando la estadística  $F$  de la tabla ANOVA. (Use el nivel de significancia de 0.01) ¿Es el resultado consistente en el punto f? ¿Debe serlo?
- h) Pronostique el número de órdenes por correo recibidas cuando se han repartido 10 mil catálogos con un intervalo para la predicción de 90% de confianza.
8. En un estudio de inversiones y tasas de interés se recabaron los datos de la tabla P.8, durante 10 años.
- a) ¿Es significativa la relación entre estas variables?
  - b) ¿Se puede desarrollar una ecuación de predicción efectiva?
  - c) Si en cinco años el promedio de la tasa de interés es de 4%, ¿se puede pronosticar la inversión anual?
  - d) Calcule e interprete  $r^2$ .
  - e) Analice la correlación y causalidad de este ejemplo.
9. La ABC Investment Company está en el negocio de subastar inversiones ofrecidas por varias firmas que desean financiamiento adicional. ABC ha tabulado sus subastas en las últimas 25 emisiones de ofertas de acuerdo con el porcentaje del valor nominal de la subasta. La oferta del principal competidor de ABC, como un porcentaje del valor nominal, también está tabulada en estas emisiones. Ahora, ABC se pregunta si está usando la misma base lógica que su competidor al preparar las pujas. En otras palabras, ¿se podrán usar las subastas de ABC para pronosticar las de su competidor? Si no, entonces el competidor debe estar evaluando las emisiones de diferente manera. Los datos se proporcionan en la tabla P-9.
- a) ¿Hasta qué punto las dos firmas están utilizando la misma base lógica para preparar las subastas?
  - b) Pronostique la puja del competidor si ABC ofrece 101% del valor nominal. Deles a ambos un estimado puntual y un intervalo de la predicción.
  - c) En el inciso b), ¿cuál es la probabilidad de que, en particular, ABC gane esta subasta (la oferta más baja gana)?
10. Evalúe las siguientes afirmaciones:
- a) Un  $r^2$  alto significa una regresión significante.
  - b) Una muestra muy grande en un problema de regresión siempre producirá resultados útiles.
11. Ed Bogdanski, dueño de la American Precast Company, lo ha contratado a usted como un analista de medio tiempo. Ed estaba muy complacido cuando usted determinó que había una relación positiva entre el número de permisos de construcción emitidos y la cantidad de trabajo disponible para esta compañía. Ahora, él se pregunta si es posible utilizar el conocimiento de las tasas de interés para las primeras hipotecas a fin de predecir el número de permisos de construcción que serán emitidos cada mes. Usted recopiló una muestra de datos aleatoria de nueve meses, como se muestra en la tabla P.11.

TABLA P.9

| <i>Emisión</i> | <i>Subasta de ABC</i> | <i>Subasta del competidor</i> | <i>Emisión</i> | <i>Subasta de ABC</i> | <i>Subasta del competidor</i> |
|----------------|-----------------------|-------------------------------|----------------|-----------------------|-------------------------------|
| 1              | 99.035                | 100.104                       | 14             | 100.542               | 99.936                        |
| 2              | 104.358               | 105.032                       | 15             | 96.842                | 95.834                        |
| 3              | 99.435                | 99.517                        | 16             | 99.200                | 99.863                        |
| 4              | 96.932                | 95.808                        | 17             | 101.614               | 102.010                       |
| 5              | 98.904                | 98.835                        | 18             | 99.501                | 99.432                        |
| 6              | 101.635               | 101.563                       | 19             | 100.898               | 99.965                        |
| 7              | 100.001               | 101.237                       | 20             | 97.001                | 96.838                        |
| 8              | 98.234                | 99.123                        | 21             | 100.025               | 100.804                       |
| 9              | 93.849                | 94.803                        | 22             | 103.014               | 104.300                       |
| 10             | 99.412                | 100.063                       | 23             | 98.702                | 99.010                        |
| 11             | 99.949                | 99.564                        | 24             | 101.834               | 100.936                       |
| 12             | 104.012               | 103.889                       | 25             | 102.903               | 103.834                       |
| 13             | 99.473                | 99.348                        |                |                       |                               |

TABLA P.11

| <i>Mes</i> | <i>Permisos de construcción</i> | <i>Tasa de interés (%)</i> |
|------------|---------------------------------|----------------------------|
|            | <i>Y</i>                        | <i>X</i>                   |
| 1          | 786                             | 10.2                       |
| 2          | 494                             | 12.6                       |
| 3          | 289                             | 13.5                       |
| 4          | 892                             | 9.7                        |
| 5          | 343                             | 10.8                       |
| 6          | 888                             | 9.5                        |
| 7          | 509                             | 10.9                       |
| 8          | 987                             | 9.2                        |
| 9          | 187                             | 14.2                       |

- a) Grafique los datos en un diagrama de dispersión.
- b) Determine la función de regresión ajustada.
- c) Pruebe la significancia del coeficiente de que corresponde a la pendiente en un nivel de significancia de 0.05.
- d) Cuando la tasa de interés aumenta 1%, ¿cuál es el promedio de disminución en el número de permisos de construcción emitidos?
- e) Calcule el coeficiente de determinación.
- f) Escriba una oración que Ed pueda entender interpretando el número calculado en el punto 3.
- g) Escriba un informe a Ed explicando los resultados de su análisis.
12. Considere la población de 140 observaciones que se presentan en la tabla P.12. La Marshall Printing Company desea estimar la relación entre el número de copias producido por una técnica de impresión en offset (*X*) y el costo de producción asociado a la mano de obra (*Y*).
- Seleccione al azar una muestra de 20 observaciones.
- a) Elabore un diagrama de dispersión.
- b) Calcule el coeficiente de correlación.
- c) Determine la línea de regresión ajustada.
- d) Grafique la línea ajustada en el diagrama de dispersión.
- e) Calcule el error estándar de la estimación.
- f) Calcule el coeficiente de determinación e interprete su valor.
- g) Pruebe la hipótesis de que la pendiente,  $\beta_1$ , de la línea de regresión de la población es cero.

TABLA P.12

| OBS. | Y   | X   | OBS. | Y   | X   | OBS.  | Y   | X   |
|------|-----|-----|------|-----|-----|-------|-----|-----|
| (1)  | 1.0 | 10  | (48) | 2.2 | 180 | (95)  | 2.0 | 330 |
| (2)  | 0.9 | 10  | (49) | 2.4 | 180 | (96)  | 2.4 | 340 |
| (3)  | 0.8 | 10  | (50) | 1.6 | 180 | (97)  | 2.2 | 340 |
| (4)  | 1.3 | 20  | (51) | 1.8 | 190 | (98)  | 2.0 | 340 |
| (5)  | 0.9 | 20  | (52) | 4.1 | 190 | (99)  | 2.5 | 350 |
| (6)  | 0.6 | 30  | (53) | 2.0 | 190 | (100) | 2.8 | 350 |
| (7)  | 1.1 | 30  | (54) | 1.5 | 200 | (101) | 2.3 | 350 |
| (8)  | 1.0 | 30  | (55) | 2.1 | 200 | (102) | 2.7 | 350 |
| (9)  | 1.4 | 40  | (56) | 2.5 | 200 | (103) | 2.8 | 360 |
| (10) | 1.4 | 40  | (57) | 1.7 | 220 | (104) | 3.1 | 360 |
| (11) | 1.2 | 40  | (58) | 2.0 | 220 | (105) | 2.5 | 370 |
| (12) | 1.7 | 50  | (59) | 2.3 | 220 | (106) | 2.9 | 370 |
| (13) | 0.9 | 50  | (60) | 1.8 | 220 | (107) | 2.6 | 370 |
| (14) | 1.2 | 50  | (61) | 1.3 | 230 | (108) | 3.0 | 380 |
| (15) | 1.3 | 50  | (62) | 1.6 | 230 | (109) | 3.2 | 380 |
| (16) | 0.7 | 60  | (63) | 2.8 | 230 | (110) | 2.9 | 390 |
| (17) | 1.0 | 60  | (64) | 2.2 | 230 | (111) | 2.6 | 390 |
| (18) | 1.3 | 70  | (65) | 2.6 | 230 | (112) | 2.5 | 390 |
| (19) | 1.5 | 70  | (66) | 1.4 | 240 | (113) | 2.7 | 400 |
| (20) | 2.0 | 70  | (67) | 1.6 | 240 | (114) | 3.1 | 400 |
| (21) | 0.8 | 80  | (68) | 1.7 | 240 | (115) | 2.4 | 400 |
| (22) | 0.6 | 80  | (69) | 1.5 | 250 | (116) | 3.0 | 400 |
| (23) | 1.8 | 80  | (70) | 2.2 | 250 | (117) | 3.4 | 420 |
| (24) | 1.0 | 90  | (71) | 2.5 | 250 | (118) | 3.5 | 420 |
| (25) | 2.0 | 100 | (72) | 2.4 | 260 | (119) | 3.1 | 420 |
| (26) | 0.5 | 100 | (73) | 2.0 | 260 | (120) | 2.9 | 420 |
| (27) | 1.5 | 100 | (74) | 2.7 | 260 | (121) | 2.8 | 430 |
| (28) | 1.3 | 110 | (75) | 2.0 | 270 | (122) | 3.3 | 430 |
| (29) | 1.7 | 110 | (76) | 2.2 | 270 | (123) | 2.5 | 440 |
| (30) | 1.2 | 110 | (77) | 2.4 | 270 | (124) | 2.8 | 440 |
| (31) | 0.8 | 110 | (78) | 1.8 | 280 | (125) | 2.4 | 450 |
| (32) | 1.0 | 120 | (79) | 2.8 | 290 | (126) | 2.6 | 450 |
| (33) | 1.8 | 120 | (80) | 2.2 | 290 | (127) | 3.0 | 450 |
| (34) | 2.1 | 120 | (81) | 2.4 | 290 | (128) | 3.4 | 460 |
| (35) | 1.5 | 130 | (82) | 2.1 | 290 | (129) | 3.0 | 460 |
| (36) | 1.9 | 130 | (83) | 1.9 | 290 | (130) | 3.3 | 470 |
| (37) | 1.7 | 140 | (84) | 2.4 | 300 | (131) | 3.4 | 470 |
| (38) | 1.2 | 150 | (85) | 2.5 | 300 | (132) | 3.1 | 470 |
| (39) | 1.4 | 150 | (86) | 2.9 | 300 | (133) | 3.6 | 480 |
| (40) | 2.1 | 150 | (87) | 2.0 | 300 | (134) | 3.0 | 480 |
| (41) | 0.9 | 160 | (88) | 1.9 | 310 | (135) | 2.9 | 480 |
| (42) | 1.1 | 160 | (89) | 2.5 | 310 | (136) | 3.2 | 480 |
| (43) | 1.7 | 160 | (90) | 2.6 | 310 | (137) | 2.6 | 490 |
| (44) | 2.0 | 160 | (91) | 3.2 | 320 | (138) | 3.8 | 490 |
| (45) | 1.6 | 170 | (92) | 2.8 | 320 | (139) | 3.3 | 490 |
| (46) | 1.9 | 170 | (93) | 2.4 | 320 | (140) | 2.9 | 500 |
| (47) | 1.7 | 170 | (94) | 2.5 | 320 |       |     |     |

TABLA P.13

| <i>Lote</i> | <i>Número de defectuosos</i> | <i>Tamaño del lote</i> |
|-------------|------------------------------|------------------------|
|             | <i>Y</i>                     | <i>X</i>               |
| 1           | 4                            | 25                     |
| 2           | 8                            | 50                     |
| 3           | 6                            | 75                     |
| 4           | 16                           | 100                    |
| 5           | 22                           | 125                    |
| 6           | 27                           | 150                    |
| 7           | 36                           | 175                    |
| 8           | 49                           | 200                    |
| 9           | 53                           | 225                    |
| 10          | 70                           | 250                    |
| 11          | 82                           | 275                    |
| 12          | 95                           | 300                    |
| 13          | 109                          | 325                    |

- h)* Calcule una estimación puntual y un intervalo de predicción de 90% para el costo en mano de obra para la producción, si el proyecto requiere 250 copias.
- i)* Examine los residuales. ¿Un modelo de regresión lineal simple será el apropiado para estos datos? Explique.
13. Harry Daniels es un ingeniero de control de calidad para la Specific Electric Corporation, empresa que fabrica motores eléctricos. Uno de los pasos del proceso de fabricación implica el uso de una máquina fresadora automática para hacer las aberturas en los ejes de los motores. Cada lote de ejes es evaluado y se descartan todos los ejes que no cumplen con las dimensiones requeridas. Antes de fabricar un nuevo lote, la máquina fresadora debe reajustarse porque su cabeza cortadora se desgasta ligeramente durante la producción. A Harry se le ordenó pronosticar cómo afecta el tamaño de los lotes al número de ejes defectuosos en cada conjunto, a fin de que pueda seleccionar el mejor tamaño de lote. Luego, reúne los siguientes datos para el tamaño promedio de los lotes, de 13 —el cual se muestra en la tabla P.13—, y le asigna a usted la tarea de analizarlo.
- a)* Grafique los datos en un diagrama de dispersión.
  - b)* Ajuste un modelo de regresión lineal simple.
  - c)* Pruebe la significancia del coeficiente de la pendiente.
  - d)* Examine los residuales.
  - e)* Desarrolle un modelo curvilíneo y mediante alguna transformación de la variable independiente ajústelo a un modelo de regresión lineal simple.
  - f)* Pruebe la significancia del coeficiente de pendiente de la variable transformada.
  - g)* Examine los residuales.
  - h)* Pronostique el número de defectuosos para un lote de 300 piezas.
  - i)* ¿Cuál de los modelos en los incisos *b)* y *e)* prefiere?
  - j)* Escriba a Harry un informe en el que resuma sus resultados.
14. Los datos en la tabla P.14 fueron recopilados como parte de un estudio de evaluación de propiedades de bienes raíces. Los números son observaciones en  $X$  = valor tasado (en miles de dólares) de los libros del valuador de la ciudad, y  $Y$  = valor en el mercado (precio de venta en miles de dólares) para  $n = 30$  parcelas de tierra que se venden en un año particular y en un área geográfica determinada.
- a)* Grafique el valor en el mercado contra el valor tasado en libros como un diagrama de dispersión.
  - b)* Con un modelo de regresión lineal simple, determine la línea de mínimos cuadrados que relaciona el valor en el mercado con el valor tasado en libros.

TABLA P.14

| Parcela | Valor en libros | Mercado | Parcela | Valor en libros | Mercado |
|---------|-----------------|---------|---------|-----------------|---------|
| 1       | 68.2            | 87.4    | 16      | 74.0            | 88.4    |
| 2       | 74.6            | 88.0    | 17      | 72.8            | 93.6    |
| 3       | 64.6            | 87.2    | 18      | 80.4            | 92.8    |
| 4       | 80.2            | 94.0    | 19      | 74.2            | 90.6    |
| 5       | 76.0            | 94.2    | 20      | 80.0            | 91.6    |
| 6       | 78.0            | 93.6    | 21      | 81.6            | 92.8    |
| 7       | 76.0            | 88.4    | 22      | 75.6            | 89.0    |
| 8       | 77.0            | 92.2    | 23      | 79.4            | 91.8    |
| 9       | 75.2            | 90.4    | 24      | 82.2            | 98.4    |
| 10      | 72.4            | 90.4    | 25      | 67.0            | 89.8    |
| 11      | 80.0            | 93.6    | 26      | 72.0            | 97.2    |
| 12      | 76.4            | 91.4    | 27      | 73.6            | 95.2    |
| 13      | 70.2            | 89.6    | 28      | 71.4            | 88.8    |
| 14      | 75.8            | 91.8    | 29      | 81.0            | 97.4    |
| 15      | 79.2            | 94.8    | 30      | 80.6            | 95.4    |

- c) Determine  $r^2$  e interprete su valor.
- d) ¿La regresión es significativa? Explique.
- e) Pronostique el valor en el mercado de una propiedad con un valor tasado en libros de 90.5. ¿Hay algún peligro al hacer esta predicción?
- f) Examine los residuales. ¿Puede identificar alguna observación atípica que tenga una gran influencia en la ubicación de la línea de mínimos cuadrados?
15. En la tabla P-15 se proporcionan los costos de los jugadores ( $X$ ) y los gastos operativos ( $Y$ ) por  $n = 26$  de los equipos de grandes ligas para la temporada 1990-1991. (Todos los datos están en millones de dólares.)
- a) Con un modelo de regresión lineal simple, determine la ecuación para la línea recta ajustada.
  - b) Determine  $r^2$  y hable sobre la fuerza de la relación lineal.
  - c) Pruebe la significancia de la regresión con un nivel de significancia de 0.10.
  - d) ¿Podemos concluir que, como regla general, los gastos operativos son casi el doble que los costos de los jugadores? Analícelo.
  - e) Pronostique los gastos operativos, mediante un intervalo de 95% de confianza para muestra grande, si los costos de los jugadores son de 30.5 millones de dólares.
  - f) Utilizando los residuales como guía, identifique cualquier observación atípica. Es decir, ¿algunos equipos tienen costos de jugadores inusitadamente bajos o altos en comparación con los gastos operativos?
16. La tabla P.16 contiene datos de consumo de papel de prensa ( $Y$ ) durante un año en particular y un número de familias en la ciudad ( $X$ ) para un muestreo de  $n = 23$  ciudades.
- a) Grafique los consumos de papel de prensa contra el número de familias en un diagrama de dispersión.
  - b) ¿Un modelo de regresión lineal simple es adecuado para los datos de la tabla P-16? Asegúrese de que su respuesta contenga un análisis de los residuales.
  - c) Considere una transformación logarítmica para el consumo de papel de prensa y un modelo de regresión lineal simple relacionado con  $Y = \log(\text{consumo de papel de prensa})$  a  $X = \text{número de familias}$ . Ajuste este modelo.
  - c) Examine los residuales de la regresión en el inciso c). ¿Qué modelo es mejor, el del inciso b) o el del c)? Justifique su respuesta.
  - e) Usando la función adaptada en el inciso c), pronostique la cantidad de papel de prensa consumido en un año si la ciudad tiene 10 mil familias.

TABLA P.15

| <i>Equipo</i> | <i>Costos de los jugadores</i> |          | <i>Gastos operativos</i> |
|---------------|--------------------------------|----------|--------------------------|
|               | <i>X</i>                       | <i>Y</i> |                          |
| 1             | 29.8                           | 59.6     |                          |
| 2             | 36.0                           | 72.0     |                          |
| 3             | 35.2                           | 70.4     |                          |
| 4             | 29.7                           | 62.4     |                          |
| 5             | 35.4                           | 70.8     |                          |
| 6             | 15.8                           | 39.5     |                          |
| 7             | 18.0                           | 60.0     |                          |
| 8             | 23.2                           | 46.4     |                          |
| 9             | 29.0                           | 58.0     |                          |
| 10            | 20.7                           | 47.6     |                          |
| 11            | 30.4                           | 60.8     |                          |
| 12            | 21.7                           | 43.4     |                          |
| 13            | 39.2                           | 66.6     |                          |
| 14            | 34.3                           | 61.7     |                          |
| 15            | 33.3                           | 53.3     |                          |
| 16            | 27.1                           | 48.8     |                          |
| 17            | 24.4                           | 48.8     |                          |
| 18            | 12.1                           | 31.5     |                          |
| 19            | 24.9                           | 49.8     |                          |
| 20            | 31.1                           | 54.4     |                          |
| 21            | 20.4                           | 40.8     |                          |
| 22            | 24.1                           | 48.2     |                          |
| 23            | 17.4                           | 41.8     |                          |
| 24            | 26.4                           | 50.2     |                          |
| 25            | 19.5                           | 46.8     |                          |
| 26            | 21.8                           | 43.6     |                          |

TABLA P.16

| <i>Ciudad</i> | <i>Consumo de papel</i> | <i>Número de</i> | <i>Ciudad</i> | <i>Consumo de papel</i> | <i>Número de</i> |
|---------------|-------------------------|------------------|---------------|-------------------------|------------------|
|               | <i>de prensa</i>        | <i>familias</i>  |               | <i>Y</i>                | <i>X</i>         |
| 1             | 961                     | 8,600            | 13            | 878                     | 8,330            |
| 2             | 469                     | 6,870            | 14            | 637                     | 9,010            |
| 3             | 556                     | 9,880            | 15            | 3,291                   | 11,790           |
| 4             | 1,252                   | 12,370           | 16            | 2,470                   | 18,910           |
| 5             | 902                     | 6,920            | 17            | 916                     | 8,550            |
| 6             | 1,399                   | 13,760           | 18            | 525                     | 8,850            |
| 7             | 1,877                   | 7,450            | 19            | 1,159                   | 8,540            |
| 8             | 921                     | 6,700            | 20            | 1,138                   | 6,910            |
| 9             | 494                     | 7,420            | 21            | 979                     | 7,060            |
| 10            | 530                     | 6,930            | 22            | 1,899                   | 10,920           |
| 11            | 488                     | 7,400            | 23            | 5,022                   | 14,800           |
| 12            | 1,253                   | 7,420            |               |                         |                  |

- f) ¿Puede mencionar otras variables que puedan incidir en la cantidad de papel de prensa consumido en un año?
17. Outback Steakhouse creció increíblemente durante su primer año de operación. El número de sucursales de Outback Steakhouse para el periodo 1988-1993 se muestra en la siguiente tabla.

TABLA P.18

| Año  | Tiempo | Centros de copiado |
|------|--------|--------------------|
| 1990 | 1      | 1                  |
| 1991 | 2      | 2                  |
| 1992 | 3      | 2                  |
| 1993 | 4      | 6                  |
| 1994 | 5      | 10                 |
| 1995 | 6      | 16                 |
| 1996 | 7      | 25                 |
| 1997 | 8      | 41                 |
| 1998 | 9      | 60                 |
| 1999 | 10     | 97                 |
| 2000 | 11     | 150                |
| 2001 | 12     | 211                |
| 2002 | 13     | 382                |
| 2003 | 14     | 537                |

| Años                 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 |
|----------------------|------|------|------|------|------|------|
| Número de sucursales | 2    | 9    | 23   | 49   | 87   | 137  |

Fuente: Outback Steakhouse.

- a) ¿Parece haber un crecimiento lineal o exponencial en el número de establecimientos de Steakhouse?
- b) Estime la tasa de crecimiento anual para Outback Steakhouse en el periodo 1988-1993.
- c) Pronostique el número de sucursales de Outback para 2003. ¿Parece razonable este número? Explique.
18. La cadena de centros de copiado On The Double comenzó sus operaciones con una sola tienda en 1990. El número de centros de copiado en operación,  $Y$ , se ha registrado para 14 años consecutivos en la tabla P.18.
- a) Grafique el número de centros de copiado contra el año. ¿On The Double ha experimentado crecimiento lineal o exponencial?
- b) Determine la tasa de crecimiento anual para On The Double.
- c) Pronostique el número de centros de copiado que podrían operar en 2010. ¿Parece razonable este número? ¿Por qué sí o por qué no?

## CASOS

### CASO 6-1

### TIGER TRANSPORT

Tiger Transport Company se dedica al traslado de productos para el hogar localmente y a través del país. Su preocupación actual involucra el precio que cobra por transportar pequeñas cargas a grandes distancias. Están contentos con sus tarifas para camiones con carga

completa, las cuales se basan en los costos variables del conductor, combustible y mantenimiento, además de gastos generales y ganancias. Sin embargo, hay dudas sobre la tarifa apropiada que deben cobrar por las mercancías adicionales requeridas para llenar un camión. Con

el propósito de pronosticar las necesidades futuras de combustible y preparar presupuestos de largo alcance, a Tiger le gustaría determinar el costo de agregar carga a un camión parcialmente lleno.

Tiger cree que el único costo adicional en el que se incurre, si se le agrega carga al camión, es el costo adicional del combustible porque las millas por galón bajarían. Como uno de los factores usados para determinar tarifas para cargas pequeñas, a la compañía le gustaría saber el monto de los costos asociados al consumo adicional de combustible.

Usted se graduó en la escuela de administración y trabaja en el departamento de contabilidad de costos. Se le ha asignado la tarea de investigar este problema y aconsejar a la administración sobre las consideraciones necesarias para tomar una decisión en cuanto a costos. Usted empezó por suponer que todos los camiones son iguales; de hecho, casi son idénticos en tamaño, capacidad en peso bruto y tamaño del motor. También supone que cada conductor obtendrá el mismo millaje en un viaje largo. El contador en jefe de Tiger está de acuerdo con estas suposiciones.

Entonces, usted se queda con una sola variable que puede afectar las millas por galón para camiones de distancia grande: el peso de la carga. Encuentra que el departamento de contabilidad tiene registros de todos los viajes hechos por uno de los camiones Tiger en años pasados. Estos registros incluyen el peso total de la carga, la distancia cubierta y el número de galones de diesel utilizado. El cociente entre estas dos cantidades mide las millas recorridas por galón.

Usted selecciona como población los viajes realizados en los cuatro años anteriores; hay un total de 5 428 viajes. Entonces, selecciona al azar 40 números de una tabla de números aleatorios, y como los viajes están registrados uno tras otro, les asigna el número 1 al primer

viaje registrado, el número 2 al segundo y así sucesivamente. Sus 40 números aleatorios producen una selección al azar de 40 viajes para examinarlos. El peso de la carga y las millas por galón de estos viajes se registran en la tabla 6.12.

Su computadora personal tiene un software con un paquete de análisis de regresión, así que puede ajustar un modelo lineal de regresión simple a los datos de la tabla 6.12. La tabla de resultados aparece en la tabla 6.13.

Al estudiar los resultados en la tabla 6.13, usted decide que los datos muestrales han producido una ecuación de regresión útil. Esta conclusión se basa en un  $r^2$  (76%) relativamente alto, un valor  $t$  negativo grande ( $-10.9$ ) y un valor  $F$  alto (119). A partir de la tabla de resultados, escribe la ecuación de regresión ajustada

$$\hat{Y} = 8.8484 - 0.0604X$$

donde  $Y$  se mide en millas por galón y  $X$  se mide en miles de libras. La pendiente de la ecuación de regresión ( $-0.0604$ ) se interpreta como sigue: *cada mil libras adicionales de carga reducen el millaje de un camión en 0.0604 millas por galón, en promedio.*

Actualmente, Tiger paga alrededor de 1.25 dólares por galón de diesel. Así, usted puede calcular el costo de transportar mil libras de carga a lo largo de cien millas como sigue:

De la tabla 6.12, la media de las millas por galón = 4.7

$$\text{El costo de 100 millas} = \frac{100(1.25)}{4.7} = \$26.60$$

El costo del mismo viaje con 1 000 libras adicionales

$$\frac{100(1.25)}{(4.7 - .0604)} = \$26.94$$

**TABLA 6.12** Datos para el peso de la carga por viaje (miles de libras) y millas por galón de Tiger Transport para el caso 6.1

| Peso | Millas por galón | | | | | | |
|---|---|---|---|---|---|---|---|
| 60   | 5.3              | 58   | 4.9              | 63   | 5.0              | 63   | 5.0              |
| 55   | 5.0              | 60   | 5.1              | 65   | 4.9              | 62   | 4.9              |
| 80   | 4.0              | 74   | 4.5              | 72   | 4.6              | 77   | 4.6              |
| 72   | 4.2              | 80   | 4.3              | 81   | 4.0              | 76   | 4.5              |
| 75   | 4.5              | 53   | 5.9              | 64   | 5.3              | 51   | 5.7              |
| 63   | 5.1              | 61   | 5.5              | 78   | 4.4              | 74   | 4.2              |
| 48   | 7.2              | 80   | 3.5              | 62   | 4.9              | 78   | 4.3              |
| 79   | 3.9              | 68   | 4.1              | 83   | 3.8              | 50   | 6.1              |
| 82   | 3.8              | 76   | 4.5              | 79   | 4.1              | 79   | 4.3              |
| 72   | 4.4              | 75   | 4.4              | 61   | 4.8              | 55   | 4.7              |

**TABLA 6.13 Resultado de análisis de regresión para Tiger Transport**

**Regression Analysis: MPG versus Weight**

The regression equation is

$$\text{MPG} = 8.85 - 0.0604 \text{ Weight}$$

| Predictor | Coef      | SE Coef  | T      | P     |
|-----------|-----------|----------|--------|-------|
| Constant  | 8.8484    | 0.3840   | 23.04  | 0.000 |
| Weight    | -0.060399 | 0.005538 | -10.91 | 0.000 |

$$S=0.3534 \quad R-\text{Sq}=75.8\% \quad R-\text{Sq}(\text{adj})=75.1\%$$

**Analysis of Variance**

| Source         | DF | SS     | MS     | F      | P     |
|----------------|----|--------|--------|--------|-------|
| Regression     | 1  | 14.853 | 14.853 | 118.93 | 0.000 |
| Residual Error | 38 | 4.746  | 0.125  |        |       |
| Total          | 39 | 19.599 |        |        |       |

Por lo tanto,

Aumento de costo de 1 000 libras transportadas en 100 millas = \$0.34

Usted cree que ya ha completado parte de su tarea, concretamente, determinar los costos de los gastos en efectivo asociados con agregar peso de carga a un camión medio lleno. Entonces, se da cuenta de que existen otros

factores a considerar para tomar una decisión de costos en cargas pequeñas.

**Tarea**

Prepare un informe para la administración de Tiger donde resuma el análisis. Incluya comentarios sobre cómo su trabajo mejorará los pronósticos para necesidades de combustible e ingresos en los camiones. ■

## CASO 6-2 BUTCHER PRODUCTS, INC.

Gene Butcher es dueño y presidente de Butcher Products, Inc., una pequeña compañía que fabrica sistemas de conductos de fibra de vidrio para instalaciones de cableado eléctrico. Gene ha estado estudiando la cantidad de conductos fabricados por día en los últimos dos y medio años y está preocupado sobre la amplia variabilidad de esta cifra. Para pronosticar apropiadamente el rendimiento de producción, costos, ingresos y alguna otra variable, Gene necesita establecer una relación entre el resultado y alguna otra variable.

Con base en su experiencia en la compañía, no encuentra ninguna razón para la variabilidad en el rendimiento hasta que empieza a pensar en las condiciones climáticas. Su razonamiento es que la temperatura exterior puede influir en la productividad de su fuerza laboral y el rendimiento diario obtenido.

Gene selecciona al azar algunos días de sus registros y anota la cantidad de conductos producidos en cada

uno de esos días. Luego se dirige a la agencia climática local y registra las temperaturas de cada uno de los días seleccionados. Gene está listo para realizar un estudio de correlación entre estas dos variables cuando se da cuenta de que, probablemente, el rendimiento se relacione con la desviación de la temperatura ideal y no por la temperatura misma. En otras palabras, piensa que un día que es muy caluroso o muy frío tendrá un efecto negativo en la producción cuando se compara con un día con temperatura ideal. Decide convertir sus lecturas de temperatura a desviaciones de 65 grados Fahrenheit, a la que considera ideal para generar altos rendimientos laborales. Sus datos aparecen como sigue:  $Y$  representa el número de unidades producidas, mientras que  $X$  representa la diferencia absoluta (se eliminan los signos negativos) entre la temperatura del día y 65 grados:

**TABLA 6.14 Resultado de análisis de regresión de Butcher Products, Inc.**

## Regression Analysis: Y versus X

The regression equation is

$$Y = 552 - 8.91 \cdot X$$

| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | 552.04 | 22.85   | 24.16 | 0.000 |
| x         | -8.911 | 1.453   | -6.13 | 0.000 |

S=59.41 R-Sq=64.2% R-Sq(adj)=62.5%

## Analysis of Variance

| Source         | DF | SS     | MS     | F     | P     |
|----------------|----|--------|--------|-------|-------|
| Regression     | 1  | 132758 | 132758 | 37.62 | 0.000 |
| Residual Error | 21 | 74109  | 3529   |       |       |
| Total          | 22 | 206866 |        |       |       |

| <b>Y</b> | <b>X</b> | <b>Y</b> | <b>X</b> |
|----------|----------|----------|----------|
| 485      | 12       | 327      | 15       |
| 512      | 10       | 308      | 25       |
| 625      | 3        | 603      | 8        |
| 585      | 4        | 321      | 35       |
| 318      | 27       | 426      | 5        |
| 405      | 10       | 410      | 12       |
| 379      | 18       | 515      | 2        |
| 497      | 12       | 498      | 7        |
| 316      | 27       | 357      | 17       |
| 351      | 20       | 429      | 8        |
| 525      | 4        | 401      | 12       |
| 395      | 11       |          |          |

Gene realiza un análisis de regresión con la computadora de su compañía y el software Minitab. Gene está contento al ver los resultados de su análisis de regresión como se presenta en la tabla 6.14. Los valores  $t$  son altos, lo cual indica que ambos coeficientes en la ecuación de regresión de muestra (552 y -8.9) son significativos: para ambos, los valores  $p$  para la prueba  $t$  son muy pequeños.

Cambiando a  $r$  cuadrada, Gene está un poco decepcionado al apreciar que su valor, aunque satisfactorio, no es tan alto como él esperaba (64.2%). Sin embargo, decide que es lo suficientemente alto como para pensar en la manera de incrementar los niveles de producción diarios. ■

## PREGUNTAS

1. ¿Cuántas unidades pronosticaría para un día en el que la temperatura fuera de 89 grados?
  2. ¿Cuántas unidades pronosticaría para un día en el que la temperatura fuera de 41 grados?
  3. Con base en los resultados del análisis de regresión mostrado anteriormente, ¿qué acción le recomienda-  
ría tomar a Gene para incrementar el rendimiento diario?
  4. ¿Cree usted que Gene ha desarrollado una herramienta efectiva de pronóstico?

CASO 6-3 ACE MANUFACTURING

Ace Manufacturing Company emplea varios miles de personas en la fabricación de teclados, estuches para equipo y cables para la pequeña industria de la computación. El presidente de Ace ha empezado a preocuparse por el índice de ausentismo de los empleados y ha soli-

citado al departamento de personal que revise este asunto. Éste comprende que un método efectivo para pronosticar el ausentismo reforzaría su habilidad de planear apropiadamente.

Bill McGone, director de personal, emprende la revisión de algunos expedientes de los empleados para tratar de medir el problema. Selecciona al azar 15 expedientes y registra el número de días de ausencia durante el pasado año fiscal, así como la edad del empleado. Después de leer un artículo en una revista sobre recursos humanos, él cree que la edad puede tener un efecto significativo en el ausentismo. Si encuentra que la edad y los días de ausencia muestran una buena correlación en su pequeña muestra, McGone pretende elaborar una muestra de 200 o 300 expedientes para formular una buena ecuación de predicción.

La siguiente tabla contiene los valores de datos recopilados en la muestra inicial.  $Y$  representa el número

de días de ausencia durante el pasado año fiscal, mientras que  $X$  representa la edad del empleado.

| $Y$ | $X$ | $Y$ | $X$ |
|-----|-----|-----|-----|
| 3   | 25  | 9   | 56  |
| 4   | 36  | 12  | 60  |
| 7   | 41  | 8   | 51  |
| 4   | 27  | 5   | 33  |
| 3   | 35  | 6   | 37  |
| 3   | 31  | 2   | 31  |
| 5   | 35  | 2   | 29  |
| 7   | 41  |     |     |

### PREGUNTAS

1. ¿Qué tan bien se correlacionan los días de ausencia y la edad? ¿Se puede generalizar esta correlación con la fuerza laboral completa?
2. ¿Cuál es la ecuación de pronóstico de los días de ausencia al usar la edad como una variable explicativa?
3. ¿Qué porcentaje de la variabilidad en días de ausencia se puede explicar al conocer la edad del empleado?
4. ¿Existe una relación significativa entre los días de ausencia y la edad? Cuando responda esta pregunta, utilice procedimientos estadísticos apropiados para avalar su respuesta.
5. Suponga que una persona recientemente contratada tiene 24 años de edad. ¿Cuántos días de ausencia pronosticaría para este empleado durante el año fiscal?
6. ¿Bill McGone debe realizar la muestra más grande de los empleados de la compañía con base en los resultados preliminares de su muestra?
7. ¿Ha desarrollado un método de pronóstico efectivo?

## CASO 6-4 MR. TUX

John Mosby ha escuchado que a menudo el análisis de regresión es utilizado para pronosticar variables en series de tiempo y, como tiene una computadora con un software de regresión, decide probarlo. Los volúmenes de ventas mensuales de Mr. Tux para los años de 1989 a 1996 constituyen la variable dependiente. Como primer intento, usa el número de periodo como la variable explicativa  $X$ . Su primer valor de venta  $Y$ , \$6028 corresponderá un valor de 1 en  $X$ , el segundo tendrá un valor 2 en  $X$  y así sucesivamente. Su razonamiento es que la tendencia ascendente que él sabe que existe en estos datos será considerada al utilizar siempre creciente valor  $X$  para explicar sus datos de venta.

Después de realizar el análisis de regresión en su computadora, John registra los siguientes valores:

$$\begin{aligned} t &= 11.01 & r^2 &= 0.563 \\ F &= 121.14 & \hat{Y} &= -6\,495 + 2\,729.2X \end{aligned}$$

El alto valor  $t$  le indica a John que la pendiente de su recta de regresión ajustada (2 729.2) es significativa; es decir, rechaza la hipótesis nula de que la pendiente de la recta de regresión de la población es cero. El alto valor  $F$  es coherente con este resultado (John recuerda que  $F = t^2$  para la regresión en línea recta) y la hipótesis nula de que la regresión no es significativa puede ser rechazada.

John está desilusionado con la relativamente baja  $r^2$  (56.3%). Esperaba un valor más alto para que su ecuación de regresión simple fuera útil para pronosticar con precisión sus ventas. John comprende que su bajo valor se debe a las ventas estacionales mensuales, un hecho que él conocía antes de empezar sus pronósticos. Las variaciones considerables en las estaciones resultarán en puntos de datos mensuales que no se agrupen en la recta de regresión lineal, teniendo como resultado un valor  $r$  cuadrado poco satisfactorio.

Los resultados de su regresión también preocupan a John. En la tabla de resultados se puede observar que: Durbin-Watson = 0.99. John no comprende esta afirmación y llama al profesor de estadística que tenía en la universidad. Después de describirle los valores de su

copia impresa de regresión, el profesor dice: "Por ahora tengo una clase, pero tu baja estadística Durbin-Watson significa que uno de los supuestos en el análisis de regresión no se cumple." ■

## PREGUNTAS

1. Analice la creencia de John de que sus ventas mensuales se relacionan con la estación y, por lo tanto, conducen a un valor  $r$ -cuadrado “bajo”.
  2. ¿Opina que el método de pronóstico de John es adecuado?
  3. ¿Cómo es que los datos de John incumplen uno de los supuestos del análisis de regresión?

## CASO 6-5 CONSUMER CREDIT COUNSELING

La operación del Consumer Credit Counseling (CCC) se describió en el capítulo 1 (caso 1.2). El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba para pronosticar era el número de nuevos clientes que tendrían para el resto de 1993. Marv le dio a Dorothy Mercer datos mensuales del número de nuevos clientes previstos por CCC para el periodo de enero de 1985 a marzo de 1993 (vea el caso 3.3). En el caso 3.3 Dorothy empleó el análisis de autocorrelación para explorar el patrón de datos. En el caso 4.3, ella usó el promedio móvil y métodos de suavización exponencial para pronosticar el resto de 1993.

Dorothy se preguntaba si el análisis de regresión se podría usar para desarrollar un buen modelo de pronóstico. Ella le preguntó a Marv si podía pensar en alguna variable explicativa potencial. Marv creyó que el número de personas con estampillas de alimentos (*food stamps*) se podría relacionar el número de nuevos clientes.

Dorothy sólo pudo encontrar datos del número de personas que utilizan estampillas de alimentos de enero 1989 a diciembre 1992:

|      |        |        |        |        |        |        |
|------|--------|--------|--------|--------|--------|--------|
| 1989 | 24,450 | 24,761 | 25,397 | 25,617 | 25,283 | 25,242 |
|      | 25,163 | 25,184 | 25,417 | 25,411 | 25,565 | 26,543 |
| 1990 | 26,784 | 27,044 | 27,567 | 28,080 | 28,142 | 28,412 |
|      | 28,161 | 27,936 | 28,423 | 28,366 | 29,029 | 29,035 |
| 1991 | 29,254 | 29,962 | 30,499 | 30,879 | 30,995 | 31,356 |
|      | 30,863 | 31288  | 31,492 | 31,577 | 31,912 | 32,050 |
| 1992 | 32,383 | 32,625 | 33,499 | 34,076 | 34,191 | 33,788 |
|      | 33,556 | 33,751 | 33,777 | 33,769 | 34,077 | 34,232 |

Marv también estaba familiarizado con el índice de actividad de negocios calculado para la circunscripción por el Consejo de Desarrollo Económico local. El índice de actividad de negocios era un indicador de los cambios relativos en las condiciones generales de negocios para la región. Los datos de este índice aparecen a continuación. ■

## PREGUNTAS

1. Determine si existe una relación significativa entre el número de clientes nuevos atendidos y el número de personas que utilizan estampillas de alimentos o el índice de actividad de negocios. No olvide la posibilidad de las transformaciones de datos.
2. Desarrolle una ecuación de regresión y úsela para pronosticar el número de nuevos clientes para los primeros tres meses de 1993.
3. Compare los resultados de su pronóstico con las observaciones reales para los primeros tres meses de 1993.
4. ¿El índice de actividad de negocios sería un buen factor de pronóstico del número de clientes nuevos?
5. Los datos consisten en una serie de tiempo: ¿significa esto que se ha violado el supuesto de independencia?
6. Suponga que ha desarrollado una ecuación de regresión apropiada. ¿Sería capaz de utilizar esta ecuación para pronosticar el resto de 1993? Explique su respuesta.

## CASO 6-6 AAA WASHINGTON

En el caso 5.5 se mostró una perspectiva de AAA Washington. Para ello, los estudiantes hicieron una descomposición de series de tiempo de las llamadas a los servicios de emergencia en carretera recibidas por el club en los últimos cinco años. La descomposición de series de tiempo realizada en el caso 5.5 mostró que el patrón que había observado el señor DeCoria en el volumen de las llamadas de emergencia era, de alguna manera, de naturaleza cíclica. El señor DeCoria quisiera poder pronosticar los volúmenes de estas llamadas en años futuros.

Otra investigación hecha por el club reveló muchos factores que tienen trascendencia en el volumen de llamadas. Entre ellos, se encuentra la temperatura promedio diaria y la cantidad de lluvia del día. Esta investigación mostró que las llamadas se incrementan cuando aumentan las lluvias y la temperatura promedio baja. El club también cree que el número total de llamadas que recibe depende de la cantidad de miembros afiliados. Finalmente, el señor DeCoria piensa que el número de llamadas recibidas se relaciona con el ciclo económico general. La tasa de desempleo para el estado de Washington se utiliza como una medida que sustituye al dato sobre el estado general de la economía de Washington. En la tabla 6.15 se presentan los datos de la tasa de desempleo, la temperatura promedio mensual, las lluvias mensuales y el número de miembros del club.

Una conversación con el gerente del centro de llamadas de emergencia en carretera conduce a dos importantes observaciones: (1) parece que los fabricantes

diseñan automóviles que trabajan mejor en temperaturas de 65 grados Fahrenheit, y (2) en apariencia, el volumen de llamadas se incrementa cuando la temperatura promedio baja algunos grados desde 30 grados, que cuando una baja similar ocurre pero desde los 60 grados de temperatura. Esta información sugiere que el efecto de la temperatura en el servicio de emergencia no es lineal.

### Tarea

1. Resuelva cuatro modelos de regresión lineal simple usando el número total de llamadas de emergencia como variable dependiente, la tasa de desempleo, temperatura, lluvia y número de miembros como las cuatro variables independientes. ¿Alguna de estas variables independientes será útil para pronosticar el número total de llamadas de emergencia?
2. Elabore una nueva variable que sea función de la temperatura y relacionela con el servicio de emergencia en carretera. Recuerde que la temperatura es una escala relativa y que la selección del punto cero es arbitraria. Si los vehículos son diseñados para operar mejor a temperaturas de 65 grados Fahrenheit, entonces cada grado arriba o abajo de los 65 grados les restará confiabilidad a los vehículos. Para realizar una transformación de los datos de temperatura que estimule este efecto, empiece por sustraer 65 grados de los valores mensuales de temperaturas, lo cual coloca al “cero” en 65 grados Fahrenheit. ¿Se deben usar los valores absolutos de esta nueva variable de temperatura?

TABLA 6.15 Datos de AAA Washington para el caso 6.6

| <i>Año</i> | <i>Mes</i> | <i>Llamadas</i> | <i>Tasa</i> | <i>Temp.</i> | <i>Lluvia</i> | <i>Miembros</i> |
|------------|------------|-----------------|-------------|--------------|---------------|-----------------|
| 1988       | Mayo       | 20002           | 5.7867      | 55.1         | 3.75          | —               |
|            | Junio      | 21591           | 5.7592      | 59.0         | 1.95          | —               |
|            | Julio      | 22696           | 5.5718      | 63.8         | 0.89          | —               |
|            | Agosto     | 21509           | 5.2939      | 63.8         | 0.51          | 384746          |
|            | Septiembre | 22123           | 5.4709      | 59.1         | 2.31          | 388652          |
|            | Octubre    | 21449           | 5.5049      | 54.6         | 3.12          | 392241          |
|            | Noviembre  | 23475           | 5.8630      | 45.4         | 8.42          | 393115          |
|            | Diciembre  | 23529           | 6.1349      | 41.0         | 4.44          | 392631          |
|            | Enero      | 23327           | 7.5474      | 40.3         | 4.30          | 396975          |
|            | Febrero    | 24050           | 7.8157      | 34.3         | 3.18          | 395186          |
|            | Marzo      | 24010           | 7.1390      | 43.2         | 6.57          | 397791          |
|            | Abril      | 19735           | 6.2637      | 52.5         | 2.39          | 397764          |
| 1989       | Mayo       | 20153           | 5.8332      | 55.3         | 2.83          | 399348          |
|            | Junio      | 19512           | 5.8077      | 62.4         | 1.30          | 401949          |
|            | Julio      | 19892           | 5.6713      | 62.9         | 0.83          | 404866          |
|            | Agosto     | 20326           | 5.4977      | 63.5         | 1.53          | 405341          |
|            | Septiembre | 19378           | 5.2989      | 60.9         | 0.32          | 407479          |
|            | Octubre    | 21263           | 5.6028      | 51.9         | 3.44          | 405430          |
|            | Noviembre  | 21443           | 5.9143      | 46.2         | 7.24          | 412134          |
|            | Diciembre  | 23366           | 5.0000      | 41.8         | 4.72          | 415342          |
|            | Enero      | 23836           | 6.1917      | 41.8         | 9.55          | 416255          |
|            | Febrero    | 23336           | 6.3775      | 38.9         | 5.73          | 423001          |
|            | Marzo      | 22003           | 5.7234      | 46.3         | 3.40          | 428559          |
|            | Abril      | 20155           | 4.7792      | 51.7         | 2.91          | 431429          |
| 1990       | Mayo       | 20070           | 4.5715      | 54.9         | 2.15          | 434675          |
|            | Junio      | 19588           | 4.3899      | 59.8         | 3.55          | 435864          |
|            | Julio      | 20804           | 4.2559      | 66.7         | 0.59          | 437969          |
|            | Agosto     | 19644           | 3.9359      | 66.4         | 1.33          | 440565          |
|            | Septiembre | 17424           | 3.9048      | 61.9         | 0.24          | 441936          |
|            | Octubre    | 20833           | 4.4294      | 50.4         | 1.17          | 448595          |
|            | Noviembre  | 22490           | 5.1523      | 45.8         | 10.66         | 446291          |
|            | Diciembre  | 24861           | 5.5102      | 33.9         | 7.93          | 446455          |
|            | Enero      | 23441           | 6.8901      | 37.9         | 4.40          | 445392          |
|            | Febrero    | 19205           | 7.0308      | 46.9         | 5.42          | 445787          |
|            | Marzo      | 20386           | 6.7186      | 43.4         | 4.35          | 445746          |
| 1991       | Abril      | 19988           | 6.1280      | 49.1         | 5.69          | 446430          |
|            | Mayo       | 19077           | 5.8146      | 54.3         | 2.12          | 450001          |
|            | Junio      | 19141           | 5.9480      | 58.2         | 1.61          | 452303          |
|            | Julio      | 20883           | 5.9026      | 65.4         | 0.51          | 456551          |
|            | Agosto     | 20709           | 5.7227      | 66.0         | 2.80          | 455747          |
|            | Septiembre | 19647           | 5.6877      | 60.9         | 0.20          | 456764          |
|            | Octubre    | 22013           | 6.2922      | 51.0         | 1.70          | 462340          |
|            | Noviembre  | 22375           | 7.0615      | 46.2         | 6.50          | 460492          |
|            | Diciembre  | 22727           | 7.4370      | 42.4         | 3.45          | 465361          |
|            | Enero      | 22367           | 8.4513      | 43.0         | 7.26          | 465492          |
|            | Febrero    | 21155           | 8.7699      | 46.0         | 3.59          | 466775          |
| 1992       | Marzo      | 21209           | 8.0728      | 48.9         | 1.47          | 467168          |
|            | Abril      | 19286           | 7.2392      | 52.7         | 4.35          | 464575          |
|            | Mayo       | 19725           | 7.0461      | 58.3         | 0.60          | 459019          |
|            | Junio      | 20276           | 7.0478      | 63.6         | 1.84          | 463665          |

**TABLA 6.15 (Continuación)**

| <i>Año</i>      | <i>Mes</i> | <i>Llamadas</i> | <i>Tasa</i> | <i>Temp.</i> | <i>Lluvia</i> | <i>Miembros</i> |
|-----------------|------------|-----------------|-------------|--------------|---------------|-----------------|
| 1992<br>(cont.) | Julio      | 20795           | 7.1080      | 64.9         | 1.41          | 463775          |
|                 | Agosto     | 21126           | 6.7824      | 65.0         | 1.01          | 466230          |
|                 | Septiembre | 20251           | 6.7691      | 58.4         | 2.16          | —               |
|                 | Octubre    | 22069           | 7.5896      | 53.2         | 2.55          | —               |
|                 | Noviembre  | 23268           | 7.9908      | 44.8         | 6.23          | —               |
|                 | Diciembre  | 26039           | 8.2460      | 37.8         | 4.38          | —               |
| 1993            | Enero      | 26127           | 9.5301      | 34.9         | 4.08          | —               |
|                 | Febrero    | 20067           | 9.2790      | —            | —             | —               |
|                 | Marzo      | 19673           | 8.6802      | —            | —             | —               |
|                 | Abril      | 19142           | 7.7815      | —            | —             | —               |

3. Desarrolle un diagrama de dispersión. ¿Existe una relación lineal entre las llamadas y la nueva variable de temperatura?
4. En caso de que exista una relación lineal entre las llamadas y la nueva variable de temperatura, desarrolle el modelo más adecuado. ■

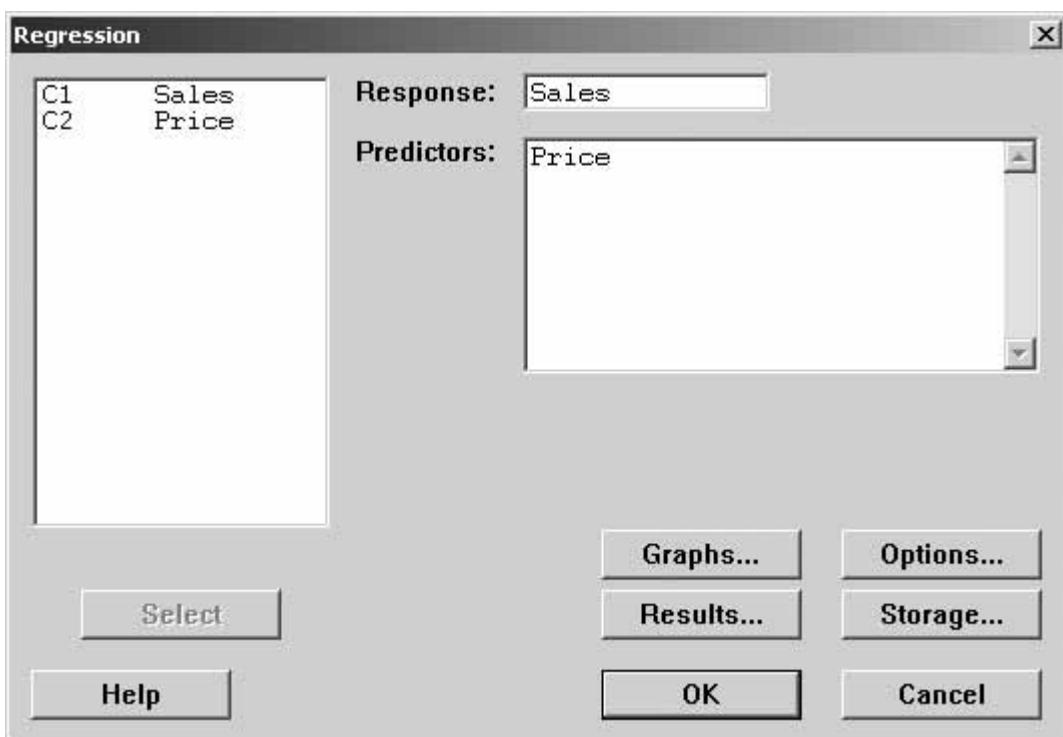
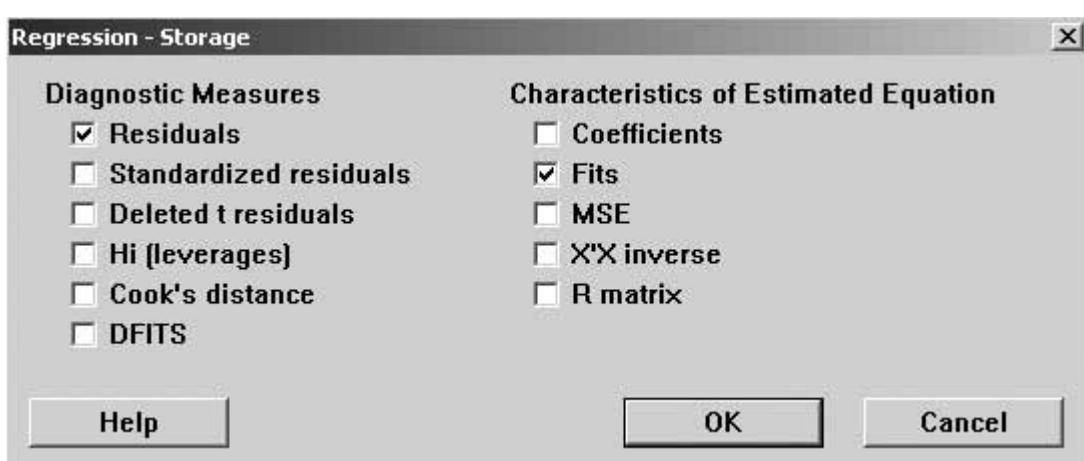
## Aplicaciones en Minitab

**El problema.** En el ejemplo 6.2, el señor Bump quería correr un análisis de regresión con los datos mostrados en la tabla 6.1.

### Solución con Minitab

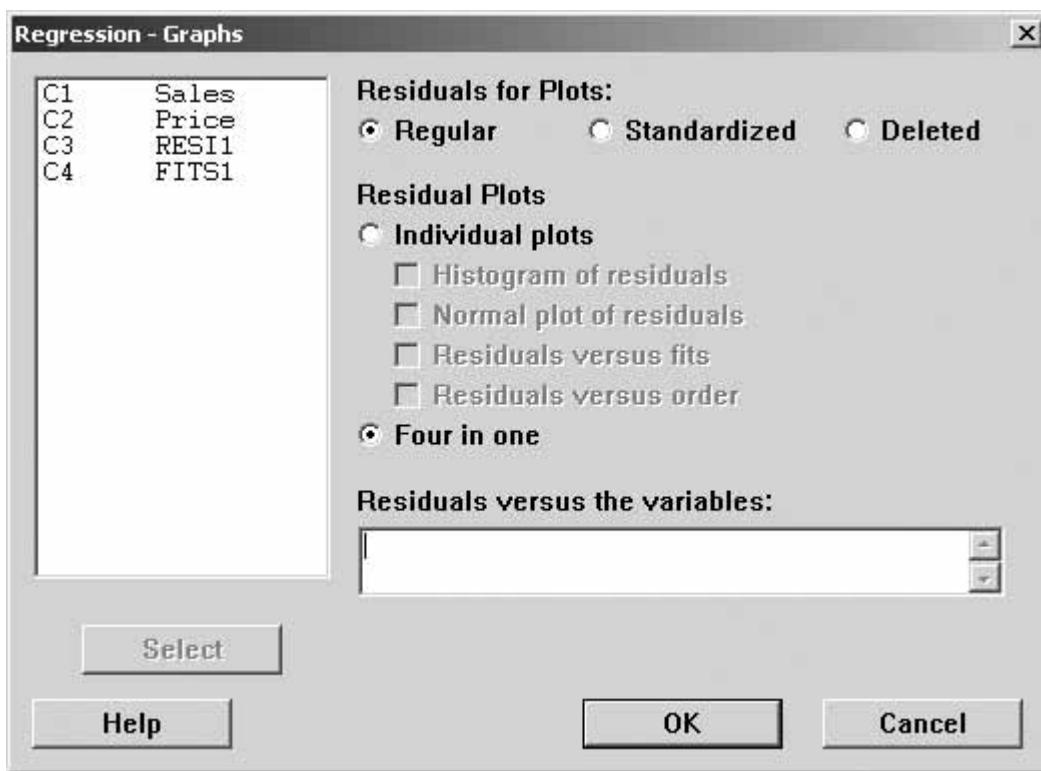
1. Coloque los datos de la tabla 6.1 en la hoja: las ventas de galones de leche en la columna C1 y el precio de venta en la columna C2.
2. Para ejecutar un modelo de regresión, dé clic en los siguientes menús:  
`Stat>Regression>Regression`
3. Aparece el cuadro de diálogo *Regression* (regresión) mostrado en la figura 6.21.
  - a) Se selecciona *Sales* (ventas) como la *Response* (respuesta) o variable dependiente.
  - b) Se selecciona *Price* (precio) como *Predictor* (variable explicativa) o variable independiente.
4. Para guardar residuales y estimaciones *Y* dé clic en *Storage* (guardar). Aparece el cuadro de diálogo *Regression – Storage* (regresión – guardar) mostrado en la figura 6.22.
  - a) Dé clic en *Residuals* (residuales) bajo *Diagnostic Measures* (medidas de diagnóstico) para guardar los residuales en C3.
  - b) Dé clic en *Fits* (ajustes) bajo *Characteristics of Estimated Equation* (características de ecuación estimada) para guardar los valores pronosticados de *Y* en C4.
  - c) Dé clic en *OK* para cerrar el cuadro de diálogo *Regression – Storage* (regresión – guardar).
5. Para ejecutar gráficas de residuales, dé clic en *Graphs* (gráficas) en el cuadro de diálogo *Regression*. Aparece el cuadro de diálogo *Regression – Graphs* (regresión – gráficas) mostrado en la figura 6.23.
  - a) Dé clic en *Four in one* (cuatro en uno) para incluir las cuatro gráficas.
  - b) Dé clic en *OK* para cerrar el cuadro de diálogo *Regression – Graphs*.
  - c) Dé clic en *OK* en el cuadro de diálogo *Regression*; el análisis de regresión de la tabla 6.8 se presenta en la ventana de sesión (vea la página 232) y la gráfica mostrada en la figura 6.9 (vea la página 231) aparece en la pantalla.

**El problema.** En el ejemplo 6.11, Gilbert García quería pronosticar las ventas usando los costos de publicidad.

FIGURA 6.21 Cuadro de diálogo *Regression* de MinitabFIGURA 6.22 Cuadro de diálogo *Regression – Storage* de Minitab

### *Solución con Minitab*

1. Coloque los datos de la figura 6.10 en la hoja (vea la página 234): las ventas en la columna C1 y los costos de publicidad en la columna C2.
2. Para obtener el diagrama de dispersión mostrado en la figura 6.11, dé clic en: Graph>Scatterplot
3. Aparece una lista de *Scatterplots* (gráficas de dispersión). Seleccione *simple* y dé clic en OK.

FIGURA 6.23 Cuadro de diálogo de *Regression – Graphs* de Minitab

4. Aparece el cuadro de diálogo *Scatterplot – Data Source* (gráficas de dispersión – fuente de datos).

a) Seleccione C1, *Sales* (ventas), como la variable *Y*, y C2, *Advertising Expenditures* (gastos de publicidad), como la variable *X*.

b) Haga clic en OK y aparecerá el diagrama de dispersión mostrado en la figura 6.11.

5. Para ejecutar un modelo adaptado como el que se muestra en la figura 6.12, dé clic en:

*Stat>Regression>Fitted Line Plot*

6. Aparece el cuadro de diálogo *Fitted Line Plot* (gráfico de línea ajustada).

a) La variable dependiente o *Response (Y)* es *Sales* (ventas).

b) La variable independiente o *Predictor (X)* es *Advertising Expenditures* (gastos de publicidad).

c) El tipo de modelo de regresión escogido es *Linear* (lineal).

d) Dé clic en OK. Los resultados se muestran en la figura 6.12.

7. A continuación, convierta la variable *X* al logaritmo natural de *X* dando clic en los siguientes menús:

*Calc>Calculator*

8. Aparece una calculadora en el cuadro de diálogo que se muestra en la figura 6.24.

a) Ponga C3 en el espacio al lado de *Store result in variable* (guardar el resultado en la variable).

b) Para realizar la transformación el logaritmo natural está resaltado en la pantalla *Functions* (funciones).

c) Dé clic en *Select* (seleccionar) y aparece LOGE (number) bajo *Select* aparece en el espacio *Expression* (expresión).

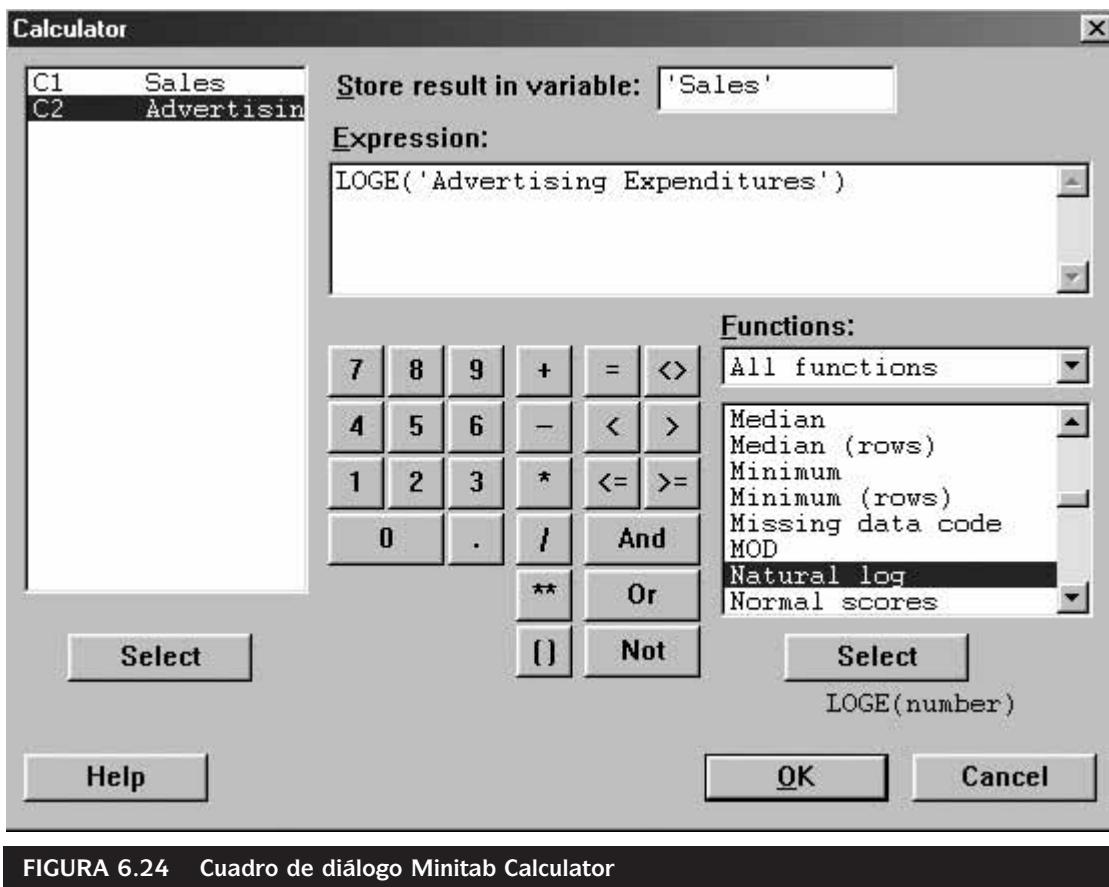


FIGURA 6.24 Cuadro de diálogo Minitab Calculator

- d) Como *Advertising Expenditures* (gastos publicitarios) es la variable a transformar, reemplaza a *number* (número) en esta expresión.
- e) Dé clic en OK y el logaritmo natural de *X* aparece en la columna C3 de la hoja de datos.
9. Se pueden realizar transformaciones para la raíz cuadrada de *X*, el cuadrado de *X* y la recíproca de *X* utilizando el cuadro de diálogo que contiene la calculadora.
10. La hoja Minitab completa se muestra en la figura 6.10.

### Aplicaciones en Excel

**El problema.** En la situación del señor Bump, ejemplo 6.1, se usó el análisis de regresión para determinar si el precio de venta podría ser usado para pronosticar las ventas semanales de galones de leche.

#### Solución en Excel

1. Escriba en la hoja las ventas semanales (vea la tabla 6.1 en la pagina 211) de A1 hasta A10 y el precio de venta de B1 hasta B10 (vea la figura 6.26).
2. Dé clic en los siguientes menús para realizar el análisis de regresión:  
Tools>Data Analysis
3. Aparece el cuadro de diálogo *Data Analysis* (análisis de datos). En *Analysis Tools* (herramientas de análisis), seleccione *Regression* (regresión) y dé clic en OK. Aparece el cuadro de diálogo *Regression* mostrado en la figura 6.25.

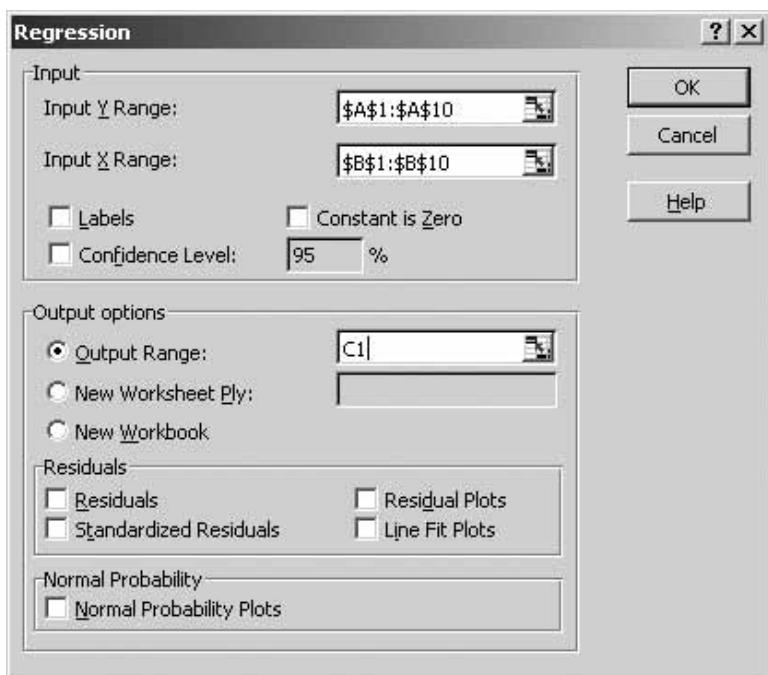


FIGURA 6.25 Cuadro de diálogo de regresión en Excel

- Escriba A1:A10 en *Input Y Range* (rango Y de entrada).
- Introduzca B1:B10 en *Input X Range* (rango X de entrada).
- Dé clic en *Output Range* (rango de salida) e introduzca C1 en el espacio siguiente.
- Dé clic en OK y aparecerá el resultado presentado en la figura 6.26.

FIGURA 6.26 Resultados de regresión de Excel para el ejemplo 6.1

| SUMMARY OUTPUT        |         |         |         |        |                |        |
|-----------------------|---------|---------|---------|--------|----------------|--------|
| <hr/>                 |         |         |         |        |                |        |
| Regression Statistics |         |         |         |        |                |        |
| <hr/>                 |         |         |         |        |                |        |
| 1. R                  | 0.863   |         |         |        |                |        |
| 2. R Square           | 0.746   |         |         |        |                |        |
| 3. Adjusted R Square  | 0.714   |         |         |        |                |        |
| 4. Standard Error     | 2.726   |         |         |        |                |        |
| 5. Observaciones      | 10      |         |         |        |                |        |
| <hr/>                 |         |         |         |        |                |        |
| ANOVA                 |         |         |         |        |                |        |
| <hr/>                 |         |         |         |        |                |        |
| 6. df                 | 1       | 2       | MS      | F      | Significance F |        |
| 7. Regression         | 1       | 174.176 | 174.176 | 23.488 | 0.0003         |        |
| 8. Residual           | 8       | 59.425  | 7.428   |        |                |        |
| 9. Total              | 9       | 233.600 |         |        |                |        |
| <hr/>                 |         |         |         |        |                |        |
| Coefficients          |         |         |         |        |                |        |
| <hr/>                 |         |         |         |        |                |        |
| 10. Intercept         | 32.136  | 4.409   | 7.389   | 0.0001 | 21.970         | 40.302 |
| 11. X Variable 1      | -18.539 | 3.002   | -4.842  | 0.0013 | -21.462        | -7.615 |
| 12.                   |         |         |         |        |                |        |
| 13.                   |         |         |         |        |                |        |
| 14.                   |         |         |         |        |                |        |
| 15.                   |         |         |         |        |                |        |
| 16.                   |         |         |         |        |                |        |
| 17.                   |         |         |         |        |                |        |
| 18.                   |         |         |         |        |                |        |
| 19.                   |         |         |         |        |                |        |
| 20.                   |         |         |         |        |                |        |
| 21.                   |         |         |         |        |                |        |
| 22.                   |         |         |         |        |                |        |
| 23.                   |         |         |         |        |                |        |
| 24.                   |         |         |         |        |                |        |
| 25.                   |         |         |         |        |                |        |
| 26.                   |         |         |         |        |                |        |
| 27.                   |         |         |         |        |                |        |
| 28.                   |         |         |         |        |                |        |
| 29.                   |         |         |         |        |                |        |
| 30.                   |         |         |         |        |                |        |
| 31.                   |         |         |         |        |                |        |
| 32.                   |         |         |         |        |                |        |
| 33.                   |         |         |         |        |                |        |
| 34.                   |         |         |         |        |                |        |
| 35.                   |         |         |         |        |                |        |
| 36.                   |         |         |         |        |                |        |
| 37.                   |         |         |         |        |                |        |
| 38.                   |         |         |         |        |                |        |
| 39.                   |         |         |         |        |                |        |
| 40.                   |         |         |         |        |                |        |
| 41.                   |         |         |         |        |                |        |
| 42.                   |         |         |         |        |                |        |
| 43.                   |         |         |         |        |                |        |
| 44.                   |         |         |         |        |                |        |
| 45.                   |         |         |         |        |                |        |
| 46.                   |         |         |         |        |                |        |
| 47.                   |         |         |         |        |                |        |
| 48.                   |         |         |         |        |                |        |
| 49.                   |         |         |         |        |                |        |
| 50.                   |         |         |         |        |                |        |
| 51.                   |         |         |         |        |                |        |
| 52.                   |         |         |         |        |                |        |
| 53.                   |         |         |         |        |                |        |
| 54.                   |         |         |         |        |                |        |
| 55.                   |         |         |         |        |                |        |
| 56.                   |         |         |         |        |                |        |
| 57.                   |         |         |         |        |                |        |
| 58.                   |         |         |         |        |                |        |
| 59.                   |         |         |         |        |                |        |
| 60.                   |         |         |         |        |                |        |
| 61.                   |         |         |         |        |                |        |
| 62.                   |         |         |         |        |                |        |
| 63.                   |         |         |         |        |                |        |
| 64.                   |         |         |         |        |                |        |
| 65.                   |         |         |         |        |                |        |
| 66.                   |         |         |         |        |                |        |
| 67.                   |         |         |         |        |                |        |
| 68.                   |         |         |         |        |                |        |
| 69.                   |         |         |         |        |                |        |
| 70.                   |         |         |         |        |                |        |
| 71.                   |         |         |         |        |                |        |
| 72.                   |         |         |         |        |                |        |
| 73.                   |         |         |         |        |                |        |
| 74.                   |         |         |         |        |                |        |
| 75.                   |         |         |         |        |                |        |
| 76.                   |         |         |         |        |                |        |
| 77.                   |         |         |         |        |                |        |
| 78.                   |         |         |         |        |                |        |
| 79.                   |         |         |         |        |                |        |
| 80.                   |         |         |         |        |                |        |
| 81.                   |         |         |         |        |                |        |
| 82.                   |         |         |         |        |                |        |
| 83.                   |         |         |         |        |                |        |
| 84.                   |         |         |         |        |                |        |
| 85.                   |         |         |         |        |                |        |
| 86.                   |         |         |         |        |                |        |
| 87.                   |         |         |         |        |                |        |
| 88.                   |         |         |         |        |                |        |
| 89.                   |         |         |         |        |                |        |
| 90.                   |         |         |         |        |                |        |
| 91.                   |         |         |         |        |                |        |
| 92.                   |         |         |         |        |                |        |
| 93.                   |         |         |         |        |                |        |
| 94.                   |         |         |         |        |                |        |
| 95.                   |         |         |         |        |                |        |
| 96.                   |         |         |         |        |                |        |
| 97.                   |         |         |         |        |                |        |
| 98.                   |         |         |         |        |                |        |
| 99.                   |         |         |         |        |                |        |
| 100.                  |         |         |         |        |                |        |
| 101.                  |         |         |         |        |                |        |
| 102.                  |         |         |         |        |                |        |
| 103.                  |         |         |         |        |                |        |
| 104.                  |         |         |         |        |                |        |
| 105.                  |         |         |         |        |                |        |
| 106.                  |         |         |         |        |                |        |
| 107.                  |         |         |         |        |                |        |
| 108.                  |         |         |         |        |                |        |
| 109.                  |         |         |         |        |                |        |
| 110.                  |         |         |         |        |                |        |
| 111.                  |         |         |         |        |                |        |
| 112.                  |         |         |         |        |                |        |
| 113.                  |         |         |         |        |                |        |
| 114.                  |         |         |         |        |                |        |
| 115.                  |         |         |         |        |                |        |
| 116.                  |         |         |         |        |                |        |
| 117.                  |         |         |         |        |                |        |
| 118.                  |         |         |         |        |                |        |
| 119.                  |         |         |         |        |                |        |
| 120.                  |         |         |         |        |                |        |
| 121.                  |         |         |         |        |                |        |
| 122.                  |         |         |         |        |                |        |
| 123.                  |         |         |         |        |                |        |
| 124.                  |         |         |         |        |                |        |

## Referencias

- Dielman, T., *Applied Regression Analysis for Business and Economics*, tercera edición, Pacific Grove, CA, Duxbury, 2001.
- Draper, N., y H. Smith, *Applied Regression Analysis*, tercera edición, Nueva York, John Wiley & Sons, 1998.
- Flaherty, W. P., "Using Regression Analysis to Pick the Best Targets", *M&A*, marzo-abril de 1991, pp. 47-49.
- Frees, E. E., *Data Analysis Using Regression Models: The Business Perspective*, Upper Saddle River, NJ, Prentice Hall, 1996.
- Johnson, R. A., y D. W. Wichern, *Business Statistics: Decision Making with Data*, Nueva York, John Wiley & Sons, 1997.
- Neter, J., W. Wasserman, M. Kutner, y C. Nachtsheim, *Applied Linear Regression Models*, tercera edición, Homewood, IL, Richard D. Irwin, 1996.



## ANÁLISIS DE REGRESIÓN MÚLTIPLE

En la regresión lineal simple se investiga la relación entre las variables independiente y dependiente. En ocasiones, la relación entre dos variables permite predecir con precisión la variable dependiente mediante el conocimiento de la independiente. Desafortunadamente, muchas situaciones de predicción no son tan simples en la práctica. Es habitual que se necesite más de una variable independiente para predecir con precisión la variable dependiente. Los modelos de regresión con más de una variable independiente se llaman *modelos de regresión múltiple*. La mayoría de los conceptos presentados en la regresión lineal simple se mantienen en la regresión múltiple. Sin embargo, también surgen algunos conceptos nuevos debido a que se usa más de una variable independiente para predecir la variable dependiente.

La *regresión múltiple* implica el uso de más de una variable independiente para predecir una variable dependiente.

### ALGUNAS VARIABLES EXPLICATIVAS

A manera de ejemplo, regresemos al problema donde se pronostican volúmenes de venta de galones de leche con el conocimiento del precio por galón. El señor Bump tiene el problema de realizar una predicción que no es totalmente precisa. Puede explicar casi 75% de las diferencias en galones de leche vendidos mediante la variable independiente. Así, 25% ( $1 - r^2$ ) de la variación total queda sin explicar. En otras palabras, con la evidencia de la muestra, el señor Bump conoce 75% de lo que debe saber para predecir los volúmenes de venta. Para realizar un trabajo más preciso, necesita encontrar otra variable explicativa que le permita explicar mejor la variación total. Si el señor Bump puede reducir la variación no explicada, su pronóstico será más preciso y no tan incierto.

Se debe buscar otra variable independiente que se relacione con los volúmenes de ventas de los galones de leche. Sin embargo, esta variable independiente o explicativa nueva, no se puede relacionar mucho con la variable independiente (precio por galón) que ya se está utilizando. Si las dos variables independientes están altamente relacionadas entre sí, explicarán la misma variación, y el pronóstico no mejorará si se agrega la segunda variable.<sup>1</sup> Acerca de este problema de intercorrelación entre variables independientes, a menudo llamado *multicolinealidad*, existe una gran preocupación en campos como la econometría y las estadísticas aplicadas. La solución simple al problema de dos variables independientes

<sup>1</sup> Esencialmente, las variables explicativas interrelacionadas contienen mucha de la misma información y, por lo tanto, no aportan información “nueva” a la conducta de la variable dependiente. Lo ideal sería que los efectos de las variables explicativas separadas en la variable dependiente no se relacionaran entre sí.

altamente relacionadas es *no* usarlas juntas. El problema de la multicolinealidad será tratado más adelante en este capítulo.

## MATRIZ DE CORRELACIÓN

---

El señor Bump decide que los costos de publicidad pueden mejorar su pronóstico de volúmenes de ventas semanales. Luego, investiga la relación entre los gastos de publicidad, el volumen de ventas y el precio por galón mediante el examen de una matriz de correlación. La *matriz de correlación* se construye calculando los coeficientes de correlación simples para cada combinación de pares de variables.

En la tabla 7.1 se ilustra un ejemplo de la matriz de correlación. El coeficiente de correlación que indica la relación entre variables 1 y 2 está representado con  $r_{12}$ . Observe que el primer subíndice, 1, también se refiere a la fila y el segundo subíndice, 2, se refiere a la columna en la tabla. Este enfoque permite determinar, a simple vista, la relación entre dos variables. Por supuesto, la correlación entre las variables 1 y 2 es exactamente la misma que la correlación entre las variables 2 y 1; es decir,  $r_{12} = r_{21}$ . Por lo tanto, sólo se necesita la mitad de la matriz de correlación. Además, la correlación de una variable consigo misma es siempre 1, de manera que, por ejemplo,  $r_{11} = r_{22} = r_{33} = 1$ .

El señor Bump corre estos datos en la computadora y obtiene la matriz de correlación mostrada en la tabla 7.2. Una investigación de las relaciones entre los gastos de publicidad, volumen de ventas y precio por galón indica que la nueva variable independiente debe contribuir a mejorar el pronóstico. La matriz de correlación indica que los gastos de publicidad tienen una correlación muy positiva ( $r_{13} = 0.89$ ) con la variable dependiente —volumen de ventas— y una correlación moderadamente negativa ( $r_{23} = -0.65$ ) con la variable independiente —el precio por galón. Esta combinación de relaciones permite que algunas de las variaciones totales del volumen de ventas, que no se habían explicado por el precio del galón, encuentren su esclarecimiento en los gastos de publicidad. Como se verá, cuando se utilizan tanto el precio por galón como los gastos de publicidad para estimar el volumen de ventas,  $R^2$  aumenta a 93.2%.

El análisis de la matriz de correlación es un paso inicial importante en la solución de cualquier problema que involucre variables independientes múltiples.

**TABLA 7.1 Matriz de correlación**

| <i>Variables</i> | <i>Variable s</i> |          |          |
|------------------|-------------------|----------|----------|
|                  | <i>1</i>          | <i>2</i> | <i>3</i> |
| 1                | $r_{11}$          | $r_{12}$ | $r_{13}$ |
| 2                | $r_{21}$          | $r_{22}$ | $r_{23}$ |
| 3                | $r_{31}$          | $r_{32}$ | $r_{33}$ |

**TABLA 7.2 Matriz de correlación para los datos de Bump**

| <i>Variable s</i> | <i>Variable s</i> |               |                   |
|-------------------|-------------------|---------------|-------------------|
|                   | <i>Ventas</i>     | <i>Precio</i> | <i>Publicidad</i> |
| <i>1</i>          | <i>2</i>          | <i>3</i>      |                   |
| Ventas, 1         | 1.00              | -.86          | .89               |
| Precio 2          |                   | 1.00          | -.65              |
| Publicidad, 3     |                   |               | 1.00              |

## MODELO DE REGRESIÓN MÚLTIPLE

En la regresión simple, la variable dependiente se puede representar con  $Y$  y la variable independiente con  $X$ . En el análisis de regresión múltiple, se utilizan las  $X$  con subíndices para representar las variables independientes. La variable dependiente todavía se representa con  $Y$ , y las variables independientes se indican con  $X_1, X_2, \dots, X_k$ . Una vez que se ha determinado la serie inicial de variables independientes, la relación entre  $Y$  y dichas  $X$  se puede expresar como un *modelo de regresión múltiple*.

En el modelo de regresión múltiple, se toma el valor esperado como una función lineal de las variables explicativas.

$$\mu_Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad (7.1)$$

Esta expresión es la *función de regresión múltiple de la población*. Como fue el caso con la regresión lineal simple, no se puede observar directamente la función de regresión de la población porque los valores observados de  $Y$  varían en sus valores promedio. Cada combinación de valores para todas las  $X$  define la media para una subpoblación de respuestas  $Y$ . Se supone que en cada una de estas subpoblaciones las  $Y$  se distribuyen normalmente sobre sus medias con la misma desviación estándar  $\sigma$ .

Los datos para esta regresión lineal simple consisten en observaciones  $(X_i, Y_i)$  sobre las dos variables. En la regresión múltiple, los datos para cada caso consisten en una observación en la respuesta y una observación en cada variable independiente. La observación  $i$ -ésima en la variable explicativa  $j$ -ésima se indica con  $X_{ij}$ . Con esta anotación, los datos para la regresión múltiple toman la forma mostrada en la tabla 7.3. Es conveniente referirse a los datos para el caso  $i$ -ésimo simplemente como la observación  $i$ -ésima. Con esta convención,  $n$  es el número de observaciones y  $k$ , el número de variables explicativas.

Modelo estadístico para regresión múltiple

La respuesta  $Y$  es una variable aleatoria que se relaciona con las variables independientes (explicativas)  $X_1, X_2, \dots, X_k$  mediante

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

donde:

- Para la observación  $i$ -ésima,  $Y = Y_i$  y  $X_1, X_2, \dots, X_k$  están dispuestas en valores  $X_{i1}, X_{i2}, \dots, X_{ik}$ .

**TABLA 7.3 Estructura de datos para regresión múltiple**

|             | <i>Variables explicativas</i> |          |       |          |       | <i>Resposta</i> |
|-------------|-------------------------------|----------|-------|----------|-------|-----------------|
| <i>Caso</i> | $X_1$                         | $X_2$    | ..... | $X_k$    | $Y$   |                 |
| 1           | $X_{11}$                      | $X_{12}$ | ..... | $X_1$    | $Y_1$ |                 |
| 2           | $X_{21}$                      | $X_{22}$ | ..... | $X_2$    | $Y_2$ |                 |
| .           | .                             | .        | ..... | .        | .     |                 |
| .           | .                             | .        | ..... | .        | .     |                 |
| $i$         | $X_{i1}$                      | $X_{i2}$ | ..... | $X_1$    | $Y_i$ |                 |
| .           | .                             | .        | ..... | .        | .     |                 |
| .           | .                             | .        | ..... | .        | .     |                 |
| $n$         | $X_{n1}$                      | $X_n$    | ..... | $X_{nk}$ | $Y_n$ |                 |

2. Los  $\epsilon$  son componentes del error que representan las desviaciones de la respuesta respecto a la relación verdadera. Son variables aleatorias no observables que explican los efectos de otros factores en la respuesta. Se supone que los errores son independientes y que cada uno se distribuye normalmente con la media 0 y una desviación estándar  $\sigma$ .
3. Se desconocen los coeficientes de regresión,  $\beta_0, \beta_1, \dots, \beta_k$ , que juntos definen la función de regresión.

Dados los datos, los coeficientes de regresión se pueden estimar por el método de mínimos cuadrados. Los estimadores de mínimos cuadrados se indican con  $b_0, b_1, \dots, b_k$ , y la función de regresión estimada con

$$\hat{Y} = b_0 + b_1 X_1 + \dots + b_k X_k \quad (7.2)$$

Los residuales  $e = Y - \hat{Y}$  son estimados del componente del error y son similares a los correspondientes errores en la regresión lineal simple; la correspondencia entre la población y la muestra es

$$\begin{aligned} \text{Población: } & Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \\ \text{Muestra: } & Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k + e \end{aligned}$$

De manera ordinaria, los cálculos en el análisis de regresión múltiple se realizan mediante el empleo de paquetes de software como Excel y Minitab (vea las secciones de “Aplicaciones” al final de este capítulo).

### Ejemplo 7.1

Para los datos mostrados en la tabla 7.4, el señor Bump considera un modelo de regresión múltiple que relacione el volumen de venta ( $Y$ ) con el precio ( $X$ ) y la publicidad ( $X_2$ )

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

El señor Bump determina la función de regresión ajustada

$$\hat{Y} = 16.41 - 8.25 X_1 + 59 X_2$$

Los valores de mínimos cuadrados  $b_0 = 16.41$ ,  $b_1 = -8.25$  y  $b_2 = 0.59$  son los que hacen mínima la suma de los cuadrados del error

$$SSE = \sum_i (Y_i - b_0 - b_1 X_{i1} - b_2 X_{i2})^2 = \sum_i (Y_i - \hat{Y}_i)^2$$

para todas las posibles elecciones de  $b_0, b_1$  y  $b_2$ . En esta ocasión, la función que mejor se adapta es un plano (vea la figura 7.1). Los puntos de los datos se grafican en tres dimensiones junto con los ejes  $Y$ ,  $X_1$  y  $X_2$ . Los puntos caen encima y debajo del plano, de tal manera que  $\Sigma(Y - \hat{Y})^2$  es un mínimo.

La función de regresión ajustada se puede usar para predecir las ventas de las próximas semanas. Si los planos requieren un precio por galón de 1.50 dólares y costos de publicidad por mil dólares, el pronóstico es de 9.935 miles de galones; esto es

$$\begin{aligned} \hat{Y} &= 16.41 - 8.25 X_1 + 0.59 X_2 \\ \hat{Y} &= 16.41 - 8.25(1.5) + 0.59(10) = 9.935 \end{aligned}$$

TABLA 7.4 Datos de Bump para el ejemplo 7.1

| Semana    | Ventas<br>(miles)<br>Y | Precio por<br>galón<br>X <sub>1</sub> | Publicidad<br>(cientos de<br>dólares)<br>X <sub>2</sub> |
|-----------|------------------------|---------------------------------------|---------------------------------------------------------|
| 1         | 10                     | \$1.30                                | 9                                                       |
| 2         | 6                      | 2.00                                  | 7                                                       |
| 3         | 5                      | 1.70                                  | 5                                                       |
| 4         | 12                     | 1.50                                  | 14                                                      |
| 5         | 10                     | 1.60                                  | 15                                                      |
| 6         | 15                     | 1.20                                  | 12                                                      |
| 7         | 5                      | 1.60                                  | 6                                                       |
| 8         | 12                     | 1.40                                  | 10                                                      |
| 9         | 17                     | 1.00                                  | 15                                                      |
| 10        | 20                     | 1.10                                  | 21                                                      |
| Totales   | 112                    | 14.40                                 | 114                                                     |
| Promedios | 11.2                   | 1.44                                  | 11.4                                                    |

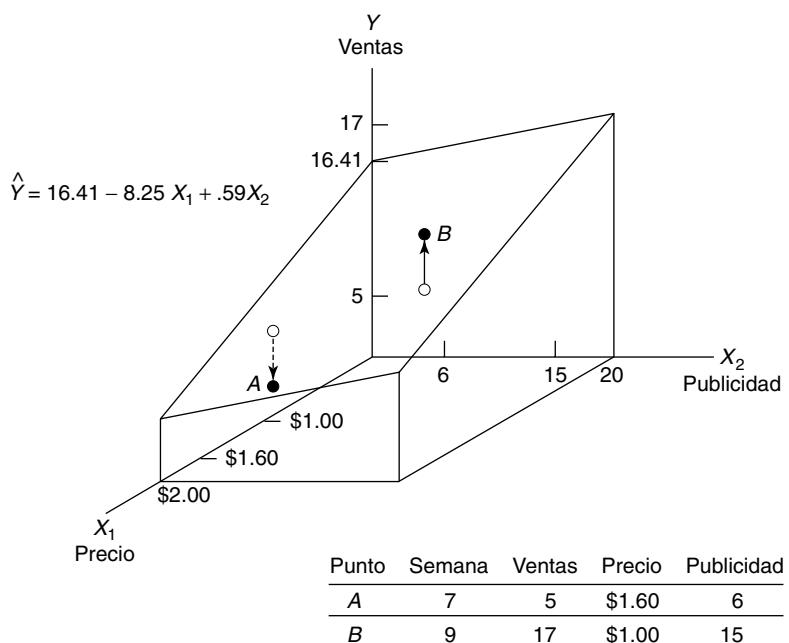


FIGURA 7.1 Plano de regresión ajustado para los datos de Bump

## INTERPRETACIÓN DE LOS COEFICIENTES DE REGRESIÓN

Reflexionemos sobre la interpretación  $b_0$ ,  $b_1$  y  $b_2$  de la función de regresión adaptada del señor Bump. El valor  $b_0$  es de nuevo la intercepción con el eje  $Y$ . Sin embargo, ahora se interpreta como el valor de  $\hat{Y}$  cuando  $X_1$  y  $X_2$  son iguales a cero. Los coeficientes  $b_1$  y  $b_2$  son llamados *coeficientes de regresión parcial, o neta*. Cada uno mide el cambio de promedio en  $Y$  que se debe al cambio en una unidad en las variables independientes pertinentes. De cualquier modo, como la influencia simultánea de todas las variables independientes en  $Y$

es medida por la función de regresión, el efecto parcial o neto de  $X_1$  (o cualquier otra  $X$ ) no debe tener ninguna influencia de otras variables. Por lo tanto, se dice que  $b_1$  mide el cambio de promedio en  $Y$  por cambio de unidad en  $X_1$ , cuando se mantienen constantes las otras variables independientes.

El *coeficiente de regresión parcial, o neto*, mide el cambio promedio en la variable dependiente por cambio en una unidad en la variable independiente que corresponde, manteniendo constantes las otras variables independientes.

En este ejemplo, el valor  $b_1$  de  $-8.25$  indica que cada aumento de un centavo por galón de leche, cuando los costos de publicidad se mantienen *constantes*, reduce la cantidad comprada en un promedio de 82.5 galones. Del mismo modo, el valor  $b_2$  de 0.59 significa que si los costos de publicidad se incrementan 100 dólares cuando el precio por galón se mantiene constante, el volumen de ventas se incrementará en un promedio de 590 galones.

### Ejemplo 7.2

Para mostrar los efectos netos de las  $X$  individuales en la respuesta, considere la situación en la que el precio es de 1.00 dólar por galón y se gastarán 1000 dólares en publicidad. Entonces

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25X_1 + .59X_2 \\ &= 16.41 - 8.25(1.00) + .59(10) \\ &= 16.41 - 8.25 + 5.9 = 14.06\end{aligned}$$

Se pronostica que las ventas serán de 14 060 galones de leche.

¿Cuál es el efecto en las ventas del incremento en un centavo en el precio si los gastos en publicidad permanecen constantes en 1000 dólares?

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25(1.01) + .59(10) \\ &= 16.41 - 8.3325 + 5.9 = 13.9775\end{aligned}$$

Observe que las ventas disminuyen 82.5 galones ( $14.06 - 13.9775 = 0.0825$ ).

¿Cuál es el efecto en las ventas de un incremento en publicidad de \$100 si el precio permanece constante en \$1.00?

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25(1.00) + .59(11) \\ &= 16.41 - 8.25 + 6.49 = 14.65\end{aligned}$$

Observe que las ventas se incrementan 590 galones ( $14.65 - 14.06 = 0.59$ )

## INFERENCIAS EN MODELOS DE REGRESIÓN MÚLTIPLE

La inferencia en los modelos de regresión múltiple es análoga al modelo de regresión lineal simple. Los estimados de mínimos cuadrados de los parámetros modelo, sus errores estándar de la estimación, las estadísticas  $t$  usadas para examinar la significancia de términos individuales en el modelo de regresión, y una estadística  $F$  para revisar la significancia de la regresión, aparecen todos en los cuadros de resultados del software estándar para estadística. No es práctico en absoluto determinar a mano las cantidades de un análisis de regresión múltiple de cualquier tamaño; en su lugar, debe utilizarse la computadora para realizar los cálculos.

Como en el capítulo 6, se puede escribir cualquier observación  $Y$

$$\text{Observación} = \text{Ajuste} + \text{Residual}$$

o

$$Y = \hat{Y} + (Y - \hat{Y})$$

donde

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

es la función de regresión ajustada. Recuerde que  $\hat{Y}$  es un estimado de la función de regresión de la población, además, representa esa parte de  $Y$  explicada por la relación de  $Y$  con las  $X$ . El residual  $Y - \hat{Y}$  es un estimado del componente del error del modelo; también representa esa parte de  $Y$  que no se explica con las variables explicativas.

La suma separada de cuadrados y los grados de libertad asociados son

$$\begin{aligned}\Sigma(Y - \bar{Y})^2 &= \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2 \\ SST &= SSR + SSE \\ df: n - 1 &= k + n - k - 1\end{aligned}\tag{7.3}$$

La variación total en la respuesta,  $SST$ , consiste en dos componentes:  $SSR$ , la variación explicada por las variables explicativas a través de la función de regresión estimada, y  $SSE$ , el error o variación no explicada. La información de la ecuación 7.3 se puede presentar en una tabla de análisis de varianza (ANOVA), esto se analizará en una sección posterior.

### Error estándar de la estimación

El error estándar de la estimación es la desviación estándar de los residuales. Por lo tanto, es una medida de la dispersión de los valores  $Y$  respecto a la función de regresión ajustada.<sup>2</sup> El error estándar de la estimación es

$$s_{y \cdot x's} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - k - 1}} = \sqrt{\frac{SSE}{n - k - 1}} = \sqrt{MSE}\tag{7.4}$$

donde

$n$  = número de observaciones

$k$  = número de variables independientes en la función de regresión

$SSE = \Sigma(Y - \hat{Y})^2$  = suma de cuadrados de los residuales

$MSE = SSE/(n - k - 1)$  = cuadrado medio de residuales

El *error estándar de la estimación* mide la cantidad en la que los valores reales ( $Y$ ) que difieren de los valores estimados ( $\hat{Y}$ ). Para muestras relativamente grandes, esperaríamos alrededor de 67% de diferencias  $Y - \hat{Y}$  dentro de  $S_{y \cdot x's}$  de 0 y aproximadamente 95% de estas diferencias dentro de  $2s_{y \cdot x's}$  de 0.

### Ejemplo 7.3

Las cantidades requeridas para calcular el error estándar de la estimación para los datos del señor Bump se proporcionan en la tabla 7.5.

---

<sup>2</sup> El error estándar de la estimación es un estimado de  $\sigma$ , la desviación estándar del término del error en el modelo de regresión múltiple.

**TABLA 7.5 Residuales del modelo para los datos de Bump en el ejemplo 7.1**

| <i>Y</i> | <i>X</i> <sub>1</sub> | <i>X</i> <sub>2</sub> | <i>Y</i> ( $\hat{Y}$ ) pronosticado usando<br>$\hat{Y} = 16.406 - 8.248X_1 + .585X_2$ | <i>Residual</i><br>$(Y - \hat{Y})$ | $(Y - \hat{Y})^2$ |
|----------|-----------------------|-----------------------|---------------------------------------------------------------------------------------|------------------------------------|-------------------|
| 10       | 1.30                  | 9                     | 10.95                                                                                 | -.95                               | .90               |
| 6        | 2.00                  | 7                     | 4.01                                                                                  | 1.99                               | 3.96              |
| 5        | 1.70                  | 5                     | 5.31                                                                                  | -.31                               | .10               |
| 12       | 1.50                  | 14                    | 12.23                                                                                 | -.23                               | .05               |
| 10       | 1.60                  | 15                    | 11.99                                                                                 | -1.99                              | 3.96              |
| 15       | 1.20                  | 12                    | 13.53                                                                                 | 1.47                               | 2.16              |
| 5        | 1.60                  | 6                     | 6.72                                                                                  | -1.72                              | 2.96              |
| 12       | 1.40                  | 10                    | 10.71                                                                                 | 1.29                               | 1.66              |
| 17       | 1.00                  | 15                    | 16.94                                                                                 | .06                                | .00               |
| 20       | 1.10                  | 21                    | 19.62                                                                                 | .38                                | .14               |
| Totales  |                       |                       |                                                                                       | .00                                | 15.90             |

El error estándar de la estimación es

$$s_{y \cdot x's} = \sqrt{\frac{15.90}{10 - 2 - 1}} = \sqrt{2.27} = 1.51$$

Con una sola variable explicativa,  $X_1$  = precio, el error estándar de la estimación fue  $S_{y \cdot x} = 2.72$  (vea el capítulo 6). Con la variable explicativa adicional,  $X_2$  = publicidad, el señor Bump ha reducido el error estándar de la estimación casi 50%. Las diferencias entre los volúmenes de leche vendida reales y los pronósticos obtenidos de la ecuación de regresión ajustada son considerablemente más pequeñas con las dos variables explicativas que con una sola. Es decir, la ecuación de dos variables explicativas se acerca más a reproducir las  $Y$  reales que la ecuación con una sola variable explicativa.

### Significancia de la regresión

En la tabla 7.6 se muestra la tabla de análisis de la varianza (ANOVA) utilizada en la descomposición de la variación total en  $Y$  ( $SST$ ) en sus partes explicada ( $SSR$ ) e inexplicada ( $SSE$ ) (vea la ecuación 7.3).

Considere la hipótesis  $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ . Esta hipótesis significa que  $Y$  no se relaciona con ninguna de las  $X$  (el coeficiente agregado a cada  $X$  es cero). La prueba  $H_0$  es llamada prueba de *significación de la regresión*. Si las suposiciones del modelo de regresión son apropiadas, y  $H_0$  es verdadero, entonces la razón

$$F = \frac{MSR}{MSE}$$

tiene una distribución  $F$  con  $df = k, n - k - 1$  grados de libertad. De este modo, la razón  $F$  se puede usar para probar la significancia de la regresión.

**TABLA 7.6 Tabla ANOVA para regresión múltiple**

| <i>Fuente</i> | <i>Suma de cuadrados</i> | <i>Grados de libertad</i><br><i>df</i> | <i>Cuadrado medio</i>   | <i>Razón F</i>        |
|---------------|--------------------------|----------------------------------------|-------------------------|-----------------------|
| Regresión     | $SSR$                    | $k$                                    | $MSR = SSR/k$           | $F = \frac{MSR}{MSE}$ |
| Error         | $SSE$                    | $n - k - 1$                            | $MSE = SSE/(n - k - 1)$ |                       |
| Total         |                          | $n - 1$                                |                         |                       |

En la regresión lineal simple hay una sola variable explicativa. Por consiguiente, poner a prueba la significancia de la regresión usando la estadística  $F$  de la tabla ANOVA equivale a la prueba  $t$  de dos colas para la hipótesis de que la pendiente de la recta de regresión es 0. Para la regresión múltiple, las pruebas  $t$  (que se presentarán dentro de poco) examinan la significancia de las  $X$  individuales en la función de regresión, y la prueba  $F$  examina colectivamente la significancia conjunta de todas las  $X$ .

### Prueba $F$ para la significancia de la regresión

En el modelo de regresión múltiple, las hipótesis

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1: \text{por lo menos una } \beta_j \neq 0$$

son probadas por la razón  $F$

$$F = \frac{MSR}{MSE}$$

con  $df = k, n - k - 1$  grados de libertad. En un nivel de significancia de  $\alpha$ , la región de rechazo es

$$F > F_\alpha$$

donde  $F_\alpha$  es el punto que deja un área  $\alpha$  a la derecha en una distribución  $F$  con  $\delta_1 = k, \delta_2 = n - k - 1$  grados de libertad del numerador y del denominador, respectivamente.

El coeficiente de determinación  $R^2$  es dado por

$$\begin{aligned} R^2 &= \frac{SSR}{SST} = \frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2} \\ &= 1 - \frac{SSE}{SST} = 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2} \end{aligned} \tag{7.5}$$

y tiene la misma forma e interpretación que  $r^2$  para la regresión lineal simple. Representa la razón de la variación en la respuesta  $Y$  explicada por su relación con las  $X$ .

Un valor de  $R^2 = 1$  dice que todas las  $Y$  observadas caen exactamente en la función de regresión ajustada. Toda la variación en la respuesta  $Y$  se explica por la regresión. Un valor de  $R^2 = 0$  dice que  $\hat{Y} = \bar{Y}$ , es decir,  $SSR = 0$ , y nada de la variación en  $\hat{Y}$  se explica por la regresión. En la práctica,  $0 \leq R^2 \leq 1$ , y el valor de  $R^2$  debe interpretarse en relación con los extremos, 0 y 1.

La cantidad

$$R = \sqrt{R^2} \tag{7.6}$$

se llama coeficiente de correlación múltiple y es la correlación entre las respuestas  $Y$  y los valores adaptados  $\hat{Y}$ . Debido a que los valores ajustados predicen las respuestas,  $R$  siempre es positiva, de manera que  $0 \leq R \leq 1$ .

Para la regresión múltiple

$$F = \frac{R^2}{1 - R^2} \left( \frac{n - k - 1}{k} \right) \quad (7.7)$$

así que, siendo todo igual, las regresiones significativas (valores calculados de  $F$  grandes) se asocian con valores relativamente grandes para  $R^2$ .

El coeficiente de determinación ( $R^2$ ) siempre se puede incrementar al agregar una variable independiente  $X$  adicional a la función de regresión, aun si esta variable adicional no es importante.<sup>3</sup> Por esta razón, algunos analistas prefieren calcular el valor de  $R^2$  ajustado al número de términos en la función de regresión. El *coeficiente ajustado de determinación*,  $\bar{R}^2$ , se da por

$$\bar{R}^2 = 1 - (1 - R^2) \left( \frac{n - 1}{n - k - 1} \right) \quad (7.8)$$

Al igual que  $R^2$ , el coeficiente ajustado  $\bar{R}^2$  es una medida de la proporción de variabilidad en la respuesta  $Y$  explicada por la regresión. Se puede mostrar que  $0 \leq \bar{R}^2 \leq R^2$ . Cuando el número de observaciones  $n$  es relativamente grande comparado con el número de variables independientes  $k$ ,  $\bar{R}^2 \approx R^2$ . Si  $k = 0$ ,  $\hat{Y} = \bar{Y}$  y  $\bar{R}^2 = R^2$ . En muchas situaciones prácticas, no hay mucha diferencia entre las magnitudes de  $R^2$  y  $\bar{R}^2$ .

#### Ejemplo 7.4

Al usar la suma total de los cuadrados en la tabla 6.7 y la suma residual de los cuadrados del ejemplo 7.3, la descomposición de la suma de cuadrados del problema del señor Bump es

$$\begin{aligned} SST &= SST + SSE \\ \Sigma(Y - \bar{Y})^2 &= \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2 \\ 233.6 &= 217.7 + 15.9 \end{aligned}$$

De este modo, se pueden usar las dos formas de la ecuación 7.5 para los cálculos,

$$R^2 = \frac{217.7}{233.6} = 1 - \frac{15.9}{233.6} = .932$$

y el coeficiente de correlación múltiple es  $R = \sqrt{R^2} = \sqrt{.932} = .965$ .

Aquí, alrededor de 93% de la variación en el volumen de ventas se explica por la regresión; es decir, la relación de las ventas con el precio y el costo de la publicidad. Además, la correlación entre las ventas y las ventas ajustadas es de 0.965, lo que indica una concordancia cercana entre los valores reales y los pronosticados. En la tabla 7.7 se proporciona un resumen del análisis de los datos de Bump.

#### Variables explicativas individuales

En la función de regresión, el coeficiente de una  $X$  individual mide el efecto parcial, o neto, de esa  $X$  en la respuesta  $Y$ , manteniendo a las otras  $X$  en la ecuación constante. Si la regresión es considerada significativa, es de interés examinar la significancia de las variables explicativas individuales. La pregunta es: dadas las otras  $X$ , ¿el efecto de una  $X$  particular es importante o se puede eliminar esta  $X$  de la función de regresión? Esta pregunta tiene respuesta al examinar un valor  $t$  apropiado.

<sup>3</sup> Aquí, “no importante” significa no significativo. Es decir, el coeficiente de  $X$  no es significativamente diferente de 0 (vea la sección “Variables explicativas individuales” que sigue).

**TABLA 7.7 Resumen del análisis de los datos de Bump**

| Variables usadas para explicar la variabilidad en $Y$ | $R^2$ | $\Sigma(Y - \hat{Y})^2$ |
|-------------------------------------------------------|-------|-------------------------|
| Ninguna                                               | .00   | 233.6                   |
| Precio                                                | .75   | 59.4                    |
| Precio y costo de publicidad                          | .93   | 15.9                    |

Si  $H_0: \beta_j = 0$  es verdadera, la estadística de prueba  $t$  con valor  $t = b_j/s_{bj}$  tiene una distribución  $t$  con  $df = n - k - 1$  grados de libertad.<sup>4</sup>

Para determinar la significancia del término  $j$ -ésimo,  $j = 0, 1, \dots, k$ , en la función de regresión, la estadística de prueba  $t$  se compara con un punto de una distribución  $t$  con grados de libertad  $n - k - 1$ . Una prueba en el nivel de significancia  $\alpha$  de

$$\begin{aligned} H_0: \beta_j &= 0 \\ H_1: \beta_j &\neq 0 \end{aligned}$$

se rechaza  $H_0$  si  $|t| > t_{\alpha/2}$ . Aquí,  $t_{\alpha/2}$  es el punto que deja un área  $\alpha/2$  a la derecha en una distribución  $t$  con  $df = n - k - 1$  grados de libertad. Se debe tener cuidado cuando se eliminan variables explicativas de la función de regresión que son consideradas insignificantes por la prueba  $t$  ( $H_0: \beta_j = 0$  no puede ser rechazada). Si las  $X$  están relacionadas (caso de multicolinealidad), los coeficientes de los mínimos cuadrados y los valores correspondientes  $t$  pueden cambiar, algunas veces considerablemente, si se elimina una sola  $X$  de la función de regresión. Por ejemplo, una  $X$  que antes era insignificante puede volverse significativa. En consecuencia, si existen varios valores  $t$  pequeños (insignificantes), las variables explicativas deben ser eliminadas una por una (empezando con la variable con el valor  $t$  más pequeño) en lugar de en grupo. El proceso se detiene cuando la regresión es significativa y todas las variables tienen estadísticas  $t$  grandes (significativas).

### Pronóstico de una respuesta futura

Un pronóstico  $\hat{Y}^*$  de una respuesta futura  $Y$  para valores nuevos de las  $X$ , por ejemplo  $X_1 = X_{1*}, X_{2*} = \dots, X_k = X_k$ , se determina al evaluar la función de regresión ajustada en las  $X^*$

$$\hat{Y}^* = b_0 + b_1 X_{1*}^* + b_2 X_{2*}^* + \dots + b_k X_k^* \quad (7.9)$$

con el nivel de confianza  $1 - \alpha$ , el intervalo para la predicción de  $Y$  toma la forma

$$\hat{Y}^* \pm t_{\alpha/2} \times (\text{error estándar de pronóstico})$$

El error estándar de pronóstico es una expresión complicada, pero el error estándar de la estimación  $S_{y \cdot x^*}$  es un componente importante. De hecho, si  $n$  es grande y todas las  $X$  son muy variables, el intervalo *aproximado* para la predicción  $100(1 - \alpha)\%$  para una nueva respuesta  $Y$  es

$$(\hat{Y}^* - t_{\alpha/2} s_{y \cdot x^*}, \hat{Y}^* + t_{\alpha/2} s_{y \cdot x^*}) \quad (7.10)$$

<sup>4</sup> Aquí  $b_j$  es el coeficiente de mínimos cuadrados para la variable explicativa  $j$ -ésima,  $X_j$ , y  $s_{bj}$  es la desviación estandarizada (error estándar). Por lo general, estas dos estadísticas se obtienen con un software, como Minitab.

## RESULTADO EN COMPUTADORA

---

En la tabla 7.8 se presenta el resultado en computadora para el problema del señor Bump. El examen de este resultado conduce a las siguientes observaciones (explicaciones acerca de la tabla 7.8).

1. Los coeficientes de regresión son  $-8.25$  para precio y  $0.585$  para gastos de publicidad. La ecuación de regresión ajustada es  $\hat{Y} = 16.4 - 8.25X_1 + 0.585X_2$ .
2. La ecuación de regresión ajustada explica  $93.2\%$  de la variación en el volumen de ventas.
3. El error estándar de la estimación es de  $1.507$  galones. Este valor es una medida de la diferencia entre la cantidad de valores reales y los valores ajustados.
4. En el capítulo 6 el coeficiente de regresión correspondiente a la pendiente se sometió a prueba para determinar si era diferente de cero. En este caso, el tamaño de la estadística  $t$  que es de  $-3.76$  para la variable de precio  $X_1$  y su valor pequeño valor  $p$  ( $0.007$ ) indican que el coeficiente del precio es significativamente diferente de cero (se rechaza  $H_0: \beta_1 = 0$ ). Dada la variable de publicidad  $X_2$  en la función de regresión, el precio no se puede descartar de la función de regresión. Asimismo, el tamaño de la estadística  $t$  de  $4.38$  para la variable de publicidad  $X_2$  y su pequeño valor  $p$  ( $0.003$ ) indican que el coeficiente de publicidad es significativamente diferente de cero (se rechaza  $H_0: \beta_2 = 0$ ). Dada la variable de precio  $X_1$  en la función de regresión, la variable de publicidad no se puede descartar de la función de regresión. (Como punto de referencia para la magnitud de los valores  $t$ , con siete grados de libertad, la tabla del apéndice C-3 da  $t_{0.01} = 2.998$ ). En resumen, los coeficientes de ambas variables explicativas son significativamente diferentes de cero.
5. El valor  $p$ ,  $0.007$ , es la probabilidad de obtener un valor  $t$  al menos tan grande como  $-3.76$ , si la hipótesis  $H_0: \beta_1 = 0$  es verdadera. Como esta probabilidad es muy pequeña, es improbable que  $H_0$  sea verdadero y se rechaza. El coeficiente del precio es significativamente diferente de cero. El valor  $p$ ,  $0.003$ , es la probabilidad de obtener un valor  $t$  al menos tan grande como  $4.38$ , si  $H_0: \beta_2 = 0$  es verdadera. Ya que es extremadamente improbable obtener un valor  $t$  de esta magnitud, se rechaza  $H_0$ . El coeficiente de publicidad es significativamente diferente de cero.

**TABLA 7.8 Resultado en Minitab para los datos de Bump**

Correlations (Pearson)

|    | Y      | X1         |
|----|--------|------------|
| X1 | -0.863 |            |
| X2 | 0.891  | -0.654 (6) |

Regression Analysis

The regression equation is

$$\hat{Y} = 16.4 - 8.25X_1 + 0.585X_2 \quad (1)$$

| Predictor | Coef       | SE Coef | T         | P         |
|-----------|------------|---------|-----------|-----------|
| Constant  | 16.406 (1) | 4.343   | 3.78      | 0.007     |
| X1        | -8.248 (1) | 2.196   | -3.76 (4) | 0.007 (5) |
| X2        | 0.5851 (1) | 0.1337  | 4.38 (4)  | 0.003 (5) |

$$S = 1.507 \quad (3) \quad R-Sq = 93.2\% \quad (2) \quad R-Sq(\text{adj}) = 91.2\% \quad (9)$$

Analysis of Variance

| Source         | DF | SS         | MS     | F         | P     |
|----------------|----|------------|--------|-----------|-------|
| Regression     | 2  | 217.70 (7) | 108.85 | 47.92 (8) | 0.000 |
| Residual Error | 7  | 15.90 (7)  | 2.27   |           |       |
| Total          | 9  | 233.60 (7) |        |           |       |

6. En la tabla 7.2 se mostró la matriz de correlación.
7. En el ejemplo 7.4 se mostró la suma de la descomposición de cuadrados,  $SST = SSR + SSE$  (suma del total de los cuadrados = suma de cuadrados de la regresión + suma de cuadrados de los errores).
8. El valor calculado  $F$  (47.92) se usa para probar la significancia de la regresión. El gran valor calculado de  $F$  y su pequeño valor  $p$  (0.000) muestran que la regresión es significativa (se rechaza  $H_0$ ;  $\beta_1 = \beta_2 = 0$ ). La  $F$  se calcula del cociente

$$F = \frac{MSR}{MSE} = \frac{108.85}{2.27} = 47.92$$

Como referencia para la magnitud de la proporción  $F$ , la tabla C-5 del apéndice da el punto superior de 1% de una distribución  $F$  con 2 y 7 grados de libertad como  $F_{0.01} = 9.55$ . La función de regresión explica una cantidad significativa de la variabilidad en las ventas  $Y$ .

9. El cálculo para la  $R^2$  corregido o ajustado,  $\bar{R}^2$  es

$$\begin{aligned}\bar{R}^2 &= 1 - (1 - R^2) \left( \frac{n - 1}{n - k - 1} \right) \\ &= 1 - (1 - .932) \left( \frac{10 - 1}{10 - 2 - 1} \right) = 1 - (.068)(1.286) = .912\end{aligned}$$

## VARIABLES FICTICIAS

---

Considere el siguiente ejemplo:

### Ejemplo 7.5

Suponga que un analista desea investigar la eficacia de una prueba de aptitud para predecir el desempeño en el trabajo. Esta prueba mide la destreza de las manos con objetos minúsculos; ocho mujeres y siete hombres se sometieron a ella. Durante un mes, los participantes recibieron capacitación intensiva como ensambladores de piezas electrónicas, enseguida, practicaron en una ensambladora verdadera durante otro mes. En este periodo se evaluó su productividad con un índice de valores que tenía una escala desde 0 hasta 10 (aquí, cero significa improductivo).

Los datos se muestran en la tabla 7.9. Asimismo, se presenta un diagrama de dispersión en la figura 7.2. Cada mujer se representa con 0 y cada hombre con 1.

Al observar la figura 7.2, es evidente de inmediato que la relación de esta prueba de aptitud con el desempeño en el trabajo sigue dos patrones diferentes, uno que se aplica a las mujeres y el otro a los hombres.

A veces es necesario determinar de qué manera se relaciona una variable dependiente con una independiente cuando un factor *cualitativo* tiene alguna influencia en la situación. Esta relación se logra con la creación de una *variable ficticia*. Existen muchas maneras de señalar cuantitativamente las clases de variables cualitativas. Por ello, en este texto se utilizan los valores 0 y 1.

Las *variables ficticias*, o indicadoras, se utilizan para determinar la relación entre variables cualitativas independientes y una variable dependiente.

La técnica de la variable ficticia se ilustra en la figura 7.3. Para mostrar los puntos de datos de las mujeres se utiliza el 0; para los hombres se usa el 1. Se construyen dos líneas paralelas para el diagrama de dispersión. La línea superior adapta los datos de las mujeres; la inferior ajusta los datos de los hombres.

Cada una de estas líneas se obtuvo de una función de regresión ajustada, de la forma

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

**TABLA 7.9 Variable ficticia y demás datos de los ensambladores de electrónicos para el ejemplo 7.5**

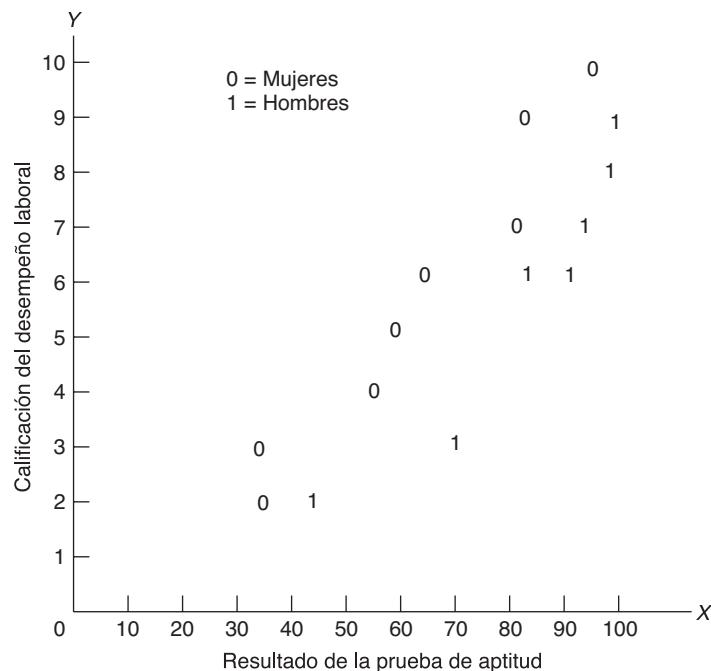
| Sujeto | Calificación del desempeño laboral<br><i>Y</i> | Resultado de la prueba de aptitud<br><i>X</i> <sub>1</sub> | Género<br><i>X</i> <sub>2</sub> |
|--------|------------------------------------------------|------------------------------------------------------------|---------------------------------|
| 1      | 5                                              | 60                                                         | 0 (F)                           |
| 2      | 4                                              | 55                                                         | 0 (F)                           |
| 3      | 3                                              | 35                                                         | 0 (F)                           |
| 4      | 10                                             | 96                                                         | 0 (F)                           |
| 5      | 2                                              | 35                                                         | 0 (F)                           |
| 6      | 7                                              | 81                                                         | 0 (F)                           |
| 7      | 6                                              | 65                                                         | 0 (F)                           |
| 8      | 9                                              | 85                                                         | 0 (F)                           |
| 9      | 9                                              | 99                                                         | 1 (M)                           |
| 10     | 2                                              | 43                                                         | 1 (M)                           |
| 11     | 8                                              | 98                                                         | 1 (M)                           |
| 12     | 6                                              | 91                                                         | 1 (M)                           |
| 13     | 7                                              | 95                                                         | 1 (M)                           |
| 14     | 3                                              | 70                                                         | 1 (M)                           |
| 15     | 6                                              | 85                                                         | 1 (M)                           |
| Total  | 87                                             | 1093                                                       |                                 |

$\bar{Y}_F$  = Media de la calificación del desempeño laboral de las mujeres = 5.75

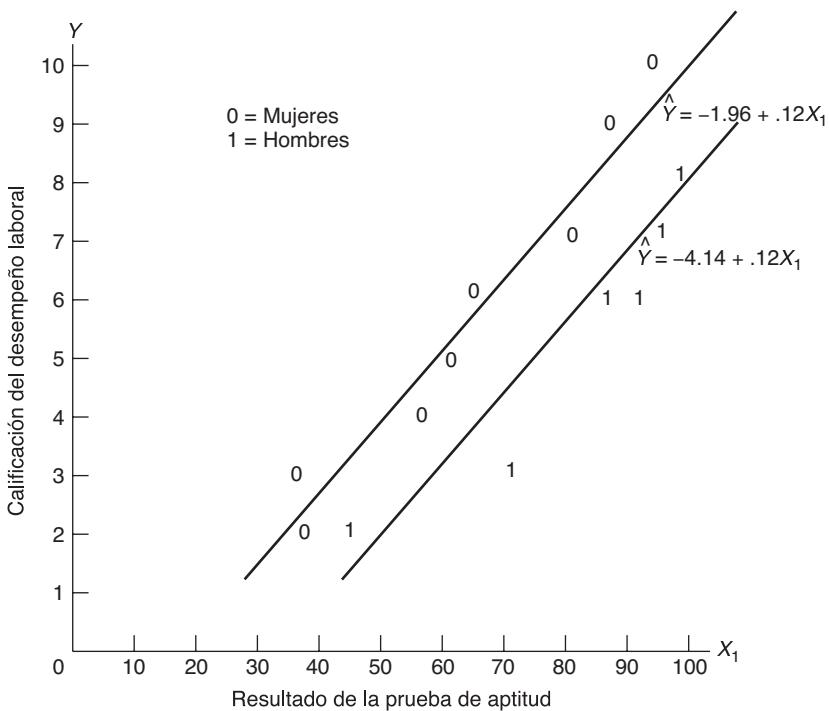
$\bar{Y}_M$  = Media de la calificación del desempeño laboral de los hombres = 5.86

$\bar{X}_F$  = Media del resultado de la prueba de aptitud de las mujeres = 64

$\bar{X}_M$  = Media del resultado de la prueba de aptitud de los hombres = 83



**FIGURA 7.2 Diagrama de dispersión para los datos del ejemplo 7.5**



**FIGURA 7.3** Líneas de regresión correspondientes a las variables ficticias del ejemplo 7.5

donde

$X_1$  = resultado de la prueba

$$X_2 = \begin{cases} 0 & \text{para mujeres} \\ 1 & \text{para hombres} \end{cases} \quad \text{variable ficticia}$$

La ecuación simple es equivalente a las siguientes dos ecuaciones

$$\begin{aligned} \hat{Y} &= b_0 + b_1 X_1 && \text{para mujeres} \\ \hat{Y} &= b_0 + b_1 X_1 + b_2 = (b_0 + b_2) + b_1 X_1 && \text{para hombres} \end{aligned}$$

Observe que  $b_2$  representa el efecto de ser un hombre en el desempeño laboral y que  $b_1$  representa el efecto de las diferencias en los resultados de la prueba de aptitud (se supone que el valor  $b_1$  es el mismo para hombres y mujeres). Lo importante es que una ecuación de regresión múltiple producirá las dos líneas estimadas que se muestran en la figura 7.3. La línea superior es la relación estimada para mujeres y la inferior es la relación estimada para hombres. Uno podría imaginarse que  $X_2$  es una variable “cambiante”, que se “activa” cuando se hace una observación para un hombre y se “desactiva” cuando se hace una indicación para una mujer.

#### Ejemplo 7.6

La ecuación estimada de regresión múltiple para los datos del ejemplo 7.5 se muestra en el resultado de Minitab en la computadora y aparece en la tabla 7.10. Es éste

$$\hat{Y} = -1.96 + 0.12X_1 - 2.18X_2$$

**TABLA 7.10 Resultado de Minitab para el ejemplo 7.6**

| Correlations: Ratings, Test, Gender             |              |                   |        |       |       |
|-------------------------------------------------|--------------|-------------------|--------|-------|-------|
| Rating      Test                                |              |                   |        |       |       |
| Test                                            | 0.876        |                   |        |       |       |
| Gender                                          | 0.021        | 0.428             |        |       |       |
| Regression Analysis: Rating versus Test, Gender |              |                   |        |       |       |
| The regression equation is                      |              |                   |        |       |       |
| Rating = -1.96 + 0.120Test -2.18Gender          |              |                   |        |       |       |
| Predictor                                       | Coef         | SE Coef           | T      | P     |       |
| Constant                                        | -1.9565      | 0.7068            | -2.77  | 0.017 |       |
| Test                                            | 0.12041      | 0.01015           | 11.86  | 0.000 |       |
| Gender                                          | -2.1807      | 0.4503            | -4.84  | 0.000 |       |
| S = 0.7863                                      | R-Sq = 92.1% | R-Sq(adj) = 90.8% |        |       |       |
| Analysis of Variance                            |              |                   |        |       |       |
| Source                                          | DF           | SS                | MS     | F     | P     |
| Regression                                      | 2            | 86.981            | 43.491 | 70.35 | 0.000 |
| Residual Error                                  | 12           | 7.419             | 0.618  |       |       |
| Total                                           | 14           | 94.400            |        |       |       |

Para los dos valores (0 y 1) de  $X_2$ , la ecuación ajustada se convierte en

$$\hat{Y} = -1.96 + 0.12X_1 - 2.18(0) = -1.96 + 0.12X_1 \quad \text{para mujeres}$$

y

$$\hat{Y} = -1.96 + 0.12X_1 - 2.18(1) = -4.14 + 0.12X_1 \quad \text{para hombres}$$

Estas ecuaciones pueden interpretarse de la siguiente manera: el valor del coeficiente de regresión  $b_1 = 0.12$  —que es la pendiente para cada una de las líneas— es el aumento promedio estimado en la calificación de desempeño para cada aumento de una unidad en el resultado de la prueba de aptitud. Este coeficiente se aplica tanto a hombres como a mujeres.

El otro coeficiente de regresión,  $b_2 = -2.18$ , se aplica sólo a los hombres. Cuando un hombre realiza la prueba, la calificación estimada del desempeño laboral se reduce 2.18 unidades —con respecto a las mujeres— cuando la de aptitud es la misma sean hombres o mujeres.

Un examen de los promedios de las variables  $Y$  y  $X_1$ , clasificadas por género, ayudará a entender este resultado. La tabla 7.9 muestra que los promedios de las calificaciones del desempeño fueron aproximadamente iguales para hombres, 5.86, y mujeres, 5.75. Sin embargo, los hombres obtuvieron un resultado más alto (83) en la prueba de aptitud que las mujeres (64). Por lo tanto, si dos candidatos, un hombre y una mujer, realizan la prueba de aptitud y ambos obtienen 70, la calificación estimada del desempeño de la mujer sería 2.18 puntos más alta que la del hombre, porque

$$\text{Mujer: } \hat{Y} = -1.96 + 0.12X_1 = -1.96 + .12(70) = 6.44$$

$$\text{Hombre: } \hat{Y} = -4.14 + 0.12X_1 = -4.14 + .12(70) = 4.26$$

Se logra un mejor entendimiento si se observa la matriz de correlación de la tabla 7.10. Existe una fuerte relación lineal entre el desempeño y la prueba de aptitud porque  $r_{12} = 0.876$ . Si solamente se utilizaran los resultados de la prueba de aptitud para predecir el desempeño, se explicaría alrededor de 77% ( $0.876^2 = 0.767$ ) de la variación de los resultados en la práctica laboral.

El coeficiente de correlación  $r_{12} = 0.02$  no indica virtualmente relación alguna entre el género y el desempeño en el trabajo. Esta conclusión también es evidente porque los promedios de calificaciones del desempeño para hombres y mujeres casi son iguales (5.86 contra 5.75). A primera vista, se podría suponer que el conocimiento del género del aspirante no constituye una información útil. Sin embargo, la relación moderada,  $r_{23} = 0.428$ , entre el género y el resultado

de la prueba de aptitud indica que ésta puede discriminar a los sexos. Al parecer, los hombres tienen mejor desempeño en la prueba que las mujeres (83 contra 64). Tal vez se requiera un elemento de fuerza en la prueba que no se pide en el trabajo.

Cuando se utilizan los resultados de la prueba y el género para pronosticar el desempeño laboral, se explica 92% de la varianza. Este resultado sugiere que ambas variables contribuyen, valiosamente, a pronosticar el desempeño. La prueba de aptitud explica 77% de la varianza y, al usarlos de manera conjunta, el género y los resultados de la prueba de aptitud agregan otro 15%. La estadística  $t$ , 11.86 y  $-4.84$ , calculada para los resultados de la prueba de aptitud y género, indica que ambas variables explicativas se deben incluir en la función de regresión final.

## MULTICOLINEALIDAD

---

En muchos problemas de regresión, los datos son registrados rutinariamente en vez de ser generados por posiciones previamente elegidas de las variables independientes. En estos casos, las variables independientes a menudo son linealmente dependientes. Por ejemplo, en el trabajo de avalúos, el precio de venta de las casas se puede relacionar con variables explicativas como la edad, el espacio de la vivienda en metros cuadrados, el número de baños, la cantidad de recámaras y número de cuartos que no sean baños, el tamaño del lote y un índice cualitativo de la construcción. Ciertamente, el espacio de vivienda, el número de cuartos y la cantidad de baños deben “moverse juntos”. Si una de estas variables se incrementa, por lo general, las otras también.

Si esta dependencia lineal es menos que perfecta, aún se pueden obtener los estimados de mínimos cuadrados de los coeficientes del modelo de regresión. Sin embargo, estos estimados tienden a ser inestables e inflados —sus valores pueden cambiar dramáticamente mediante ligeros cambios en los datos—; sus valores son más grandes de lo esperado. En especial, los coeficientes individuales pueden tener el signo equivocado y las estadísticas  $t$  para juzgar la significancia de términos individuales pueden ser insignificantes; aun así, la prueba  $F$  indicará que la regresión es significativa. Finalmente, el cálculo de los estimados de mínimos cuadrados es susceptible a errores de redondeo.

La relación lineal entre dos o más variables independientes se llama *multicolinealidad*. La fuerza de la multicolinealidad se mide por el *factor de expansión de la varianza* (*VIF*, por sus siglas en inglés).<sup>5</sup>

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, 2, \dots, k \quad (7.11)$$

Aquí  $R_j^2$  es el coeficiente de determinación de la regresión de la variable *independiente*  $j$ -ésima en las  $k - 1$  variables independientes restantes. Para dos variables independientes  $k = 2$ ,  $R_j^2$  es el cuadrado de la correlación muestral  $r$  entre estas variables.

Si la variable explicativa  $j$ -ésimo  $X_j$  no se relaciona con las  $X$  restantes,  $R_j^2 = 0$  y  $VIF_j = 1$ . Si existe una relación, entonces  $VIF_j > 1$ . Por ejemplo, cuando  $R_j^2$  es igual a 0.90,  $VIF_j = 1/(1 - 0.90) = 10$ .

Un *VIF* que se acerque a 1 sugiere que la multicolinealidad no es un problema para esa variable independiente. Su coeficiente estimado y valor asociado  $t$  no cambiarán mucho mientras se agregan o eliminan otras variables independientes de la ecuación de regresión. Un *VIF* mucho mayor que 1 indica que el coeficiente estimado sujeto a esa variable independiente es inestable. Su valor y estadística  $t$  asociada pueden cambiar considerablemente mientras se agregan o eliminan otras variables independientes de la ecuación de regresión. Un *VIF* grande significa, en esencia, que existe información redundante entre las variables explicativas. Las variables explicativas restantes ya explicaron la información transmitida por la variable con un *VIF* grande. De esta manera, la multicolinealidad difi-

<sup>5</sup> El *factor de expansión de la varianza* (*VIF*) obtiene su nombre del hecho de que  $S_{bj} \propto VIF_j$ . La desviación estándar estimada (error estándar) del coeficiente de mínimos cuadrados  $b_j$  se incrementa al mismo tiempo que aumenta  $VIF_j$ .

culta la interpretación del efecto de una variable explicativa individual en la respuesta (variable dependiente).

### Ejemplo 7.7

Un componente importante de los costos para el propietario de un diario es el costo del papel de periódico. Los editores están interesados en los factores que determinan el consumo anual de papel. En un estudio (véanse Jonson y Wichern, 1997) se recopilaron los datos del consumo anual de papel ( $Y$ ), el número de periódicos en una ciudad ( $X_1$ ), el logaritmo<sup>6</sup> del número de familias en una ciudad ( $X_2$ ) y el logaritmo de las ventas totales al menudeo en una ciudad ( $X_3$ ) para  $n = 15$  ciudades. En la tabla 7.11 se muestra la serie de correlación para las tres variables explicativas y el resultado en Minitab de un análisis de regresión relacionado con el consumo de papel de periódico con las variables explicativas.

La estadística  $F$  (18.54) y su valor  $p$  (0.000) claramente indican que la regresión es significativa. La estadística  $t$  para cada una de las variables independientes es pequeña con un valor  $p$  relativamente grande. Se debe concluir, por ejemplo, que la variable LnFamily (logaritmo del número de familias) no es significativa siempre que otras variables explicativas permanezcan en la función de regresión. Esto sugiere que el término  $\beta_2 X_2$  se puede descartar de la función de regresión si se retienen los términos restantes  $\beta_1 X_1$  y  $\beta_3 X_3$ . De modo parecido,  $\beta_3 X_3$  se puede eliminar si  $\beta_1 X_1$  y  $\beta_2 X_2$  permanecen en la función de regresión. El valor  $t$  (1.69) asociado con los papeles es ligeramente significativo, pero el término  $\beta_1 X_1$  se puede eliminar si las otras variables explicativas permanecen en la ecuación. Aquí la regresión es significativa pero cada una las variables explicativas no lo son. ¿Por qué?

La respuesta está en la columna *VIF* en la tabla 7.11. Porque *VIF* = 1.7 para Papers (papeles), esta variable explicativa es muy débil relacionada con (*VIF* cerca de 1) las variables explicativas restantes, LnFamily y LnRetSales (logaritmo de las ventas totales). El *VIF* = 7.4 para LnFamily es relativamente grande e indica que esta variable está relacionada linealmente con las variables explicativas restantes. Además, el *VIF* = 8.1 para LnRetSales indica que LnRetSales está relacionado con las variables explicativas restantes. Porque Papers está débilmente relacionado con LnFamily y LnRetSales, la relación entre la variable explicativa es, en esencia, la relación

**TABLA 7.11 Resultado en Minitab para el ejemplo 7.7 – Tres variables explicativas**

| Correlations: Papers, LnFamily, LnRetSales                         |        |                      |          |       |       |  |
|--------------------------------------------------------------------|--------|----------------------|----------|-------|-------|--|
|                                                                    |        | Papers      LnFamily |          |       |       |  |
| LnFamily                                                           | 0.600  |                      |          |       |       |  |
| LnRetSal                                                           | 0.643  | 0.930                |          |       |       |  |
| Regression Analysis: Newsprint versus Papers, LnFamily, LnRetSales |        |                      |          |       |       |  |
| The regression equation is                                         |        |                      |          |       |       |  |
| Newsprint = -56388 + 2385Papers + 1859LnFamily + 3455LnRetSales    |        |                      |          |       |       |  |
| Predictor                                                          | Coef   | SE Coef              | T        | P     | VIF   |  |
| Constant                                                           | -56388 | 13206                | -4.27    | 0.001 |       |  |
| Papers                                                             | 2385   | 1410                 | 1.69     | 0.119 | 1.7   |  |
| LnFamily                                                           | 1859   | 2346                 | 0.79     | 0.445 | 7.4   |  |
| LnRetSales                                                         | 3455   | 2590                 | 1.33     | 0.209 | 8.1   |  |
| $S = 1849 \quad R-Sq = 83.5\% \quad R-Sq(\text{adj}) = 79.0\%$     |        |                      |          |       |       |  |
| Analysis of Variance                                               |        |                      |          |       |       |  |
| Source                                                             | DF     | SS                   | MS       | F     | P     |  |
| Regression                                                         | 3      | 190239371            | 63413124 | 18.54 | 0.000 |  |
| Residual Error                                                     | 11     | 37621478             | 3420134  |       |       |  |
| Total                                                              | 14     | 227860849            |          |       |       |  |

<sup>6</sup> Los logaritmos del número de familias y las ventas totales al menudeo se utilizan para hacer que los números tengan menos sesgo positivo y sean más manejables.

**TABLA 7.12 Resultado en Minitab para ejemplo 7.7 – Dos variables explicativas**

| Regression Analysis: Newsprint versus Papers, LnRetSales |              |                   |          |       |       |  |
|----------------------------------------------------------|--------------|-------------------|----------|-------|-------|--|
| The regression equation is                               |              |                   |          |       |       |  |
| Newsprint = -59766 + 2393 Papers + 5279 LnRetSales       |              |                   |          |       |       |  |
| Predictor                                                | Coef         | SE Coef           | T        | P     | VIF   |  |
| Constant                                                 | -59766       | 12304             | -4.86    | 0.000 |       |  |
| Papers                                                   | 2393         | 1388              | 1.72     | 0.110 | 1.7   |  |
| LnRetSales                                               | 5279         | 1171              | 4.51     | 0.001 | 1.7   |  |
| S = 1820                                                 | R-Sq = 82.5% | R-Sq(adj) = 79.6% |          |       |       |  |
| Analysis of Variance                                     |              |                   |          |       |       |  |
| Source                                                   | DF           | SS                | MS       | F     | P     |  |
| Regression                                               | 2            | 188090489         | 94045244 | 28.38 | 0.000 |  |
| Residual Error                                           | 12           | 39770360          | 3314197  |       |       |  |
| Total                                                    | 14           | 227860849         |          |       |       |  |

entre LnFamily y LnRetSales. De hecho, la correlación muestral entre LnFamily y LnRetSales es  $r = 0.93$ , lo cual indica una fuerte asociación lineal.

Las dos variables LnFamily y LnRetSales son muy similares en su habilidad de explicar el consumo de papel de periódico. En la función de regresión nada más se necesita una, no las dos. El resultado en Minitab de un análisis de regresión con LnFamily (la estadística  $t$  más pequeña) eliminado de la función de regresión se muestra en la tabla 7.12.

Observe que el coeficiente de Papers es casi el mismo para las dos regresiones. No obstante, los coeficientes de LnRetSales son bastante diferentes (3455 para  $k = 3$  explicativas y 5279 para  $k = 2$  explicativas). Asimismo, para la segunda regresión, la variable LnRetSales es claramente significativa ( $t = 4.51$  con valor  $p = 0.001$ ). Con Papers en el modelo, LnRetSales es una importante variable explicativa adicional del consumo de papel de prensa. Las dos  $R^2$  para cada una de las dos regresiones son casi la misma, aproximadamente 0.83, como son los errores estándar de la estimación,  $s_{y,x's} = 1849$  y  $s_{y,x's} = 1820$ , respectivamente. Finalmente, el  $VIF = 1.7$  común para las dos explicativas en el segundo modelo indica que la multicolinealidad ya no es un problema. Como lo confirma el análisis residual, para las variables consideradas, la regresión de papel de periódico en las variables Papers y LnRetSales es totalmente adecuada.

¿Qué debemos hacer si es importante estimar los efectos separados de las variables explicativas, pero la multicolinealidad parece ser un problema? A continuación presentamos varias maneras de manejar la multicolinealidad severa. Quizás ninguna de ellas sea en su totalidad satisfactoria o viable.

- Cree nuevas variables  $X$ , y llámelas  $\bar{X}$ , luego escale todas las variables independientes de acuerdo con la fórmula

$$\tilde{X}_j = \frac{X_{ij} - \bar{X}_j}{\sqrt{\sum_i (X_{ij} - \bar{X}_j)^2}} \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, n \quad (7.12)$$

Estas nuevas variables tendrán una media muestral de 0 cada una y la misma desviación estándar de muestra. Los cálculos de la regresión con las nuevas  $X$  son menos susceptibles al redondeo del error en presencia de severa multicolinealidad.

- Identifique y elimine una o más variables independientes redundantes de la función de regresión. (Utilizamos este mismo enfoque en el ejemplo 7.7.)

- Considere otros procedimientos de estimación que no sean los mínimos cuadrados.<sup>7</sup>
- Realice una regresión de la respuesta  $Y$  tomando como variables explicativas a las combinaciones lineales de las  $X$  que no están correlacionadas.<sup>8</sup>
- Con cuidado, seleccione variables independientes potenciales al principio del estudio. Trate de evitar variables que “digan la misma cosa”.

## **SELECCIÓN DE LA “MEJOR” ECUACIÓN DE REGRESIÓN**

---

¿Cómo desarrollamos la mejor ecuación de regresión múltiple para pronosticar la variable de interés? El *primer paso* es la selección de una serie completa de variables explicativas potenciales. Se debe incluir cualquier variable que pueda contribuir a la precisión del pronóstico. En la selección de una ecuación final, generalmente uno se enfrenta al dilema de escoger variables explicativas para incluirlas en la ecuación final, el analista debe evaluarlas mediante los siguientes criterios *opuestos*:

1. El analista desea incluir tantas variables explicativas útiles como le sea posible.<sup>9</sup>
2. Dado que es costoso obtener y monitorear información en un gran número de  $X$ , la ecuación debe incluir menos variables explicativas hasta donde sea posible. Por lo general, la mejor ecuación es la más simple.

La selección de la mejor ecuación de regresión generalmente involucra una selección intermedia entre estos extremos, y el criterio personal también es una parte vital para cualquier solución.

Después de haber recopilado, en apariencia, una lista completa de variables explicativas potenciales, el *segundo paso* es someter a una revisión las variables independientes que no parezcan apropiadas. Es posible que una variable independiente (1) no sea fundamental para el problema (debe existir alguna relación convincente entre la variable dependiente y una variable independiente), o (2) puede estar sujeta a grandes errores de medición, o (3) puede duplicar otras variables independientes (multicolinealidad) o (4) puede ser difícil medirla con exactitud (los datos precisos son costosos o no están disponibles).

El *tercer paso* es acortar la lista de las explicativas para obtener una “mejor” selección de variables independientes. En la siguiente sección se analizan técnicas actualmente utilizadas. No se puede decir que alguno de estos procedimientos de búsqueda produce la “mejor” serie de variables independientes. De hecho, a menudo no existe una serie “mejor”. Para confundirlo más, no todas las diferentes técnicas conducen, necesariamente, a la misma ecuación de predicción final. Todo el proceso de selección de variable es extremadamente subjetivo. La ventaja primordial de los procedimientos de búsqueda automática es que el analista puede concentrarse en las áreas fundamentales del problema.

Para demostrar varios procedimientos de búsqueda, se presenta un ejemplo simple que tiene cinco variables independientes potenciales.

### **Ejemplo 7.8**

Pam Weigand, gerente de personal de Zurenko Pharmaceutical Company, está interesada en pronosticar si un aspirante en particular se convertirá en un buen vendedor. Ella decide usar las ventas del primer mes como la variable dependiente ( $Y$ ), y elige las siguientes variables independientes para analizarlas:

<sup>7</sup> Los procedimientos alternos para estimar los parámetros de regresión están más allá del alcance de este libro. Si el lector está interesado, puede consultar Draper y Smith (1998).

<sup>8</sup> De nuevo, los procedimientos para crear combinaciones lineales de las  $X$  que no están correlacionadas están más allá del alcance de este libro. Draper y Smith (1998) tratan estas técnicas.

<sup>9</sup> Recuerde que cuando una nueva variable explicativa se agrega a la ecuación de regresión múltiple,  $R^2$  aumenta. Por lo tanto, es importante que la nueva variable explicativa contribuya *significativamente* a la ecuación de regresión.

$X_1$  = prueba de aptitud de ventas

$X_2$  = edad, en años

$X_3$  = resultado de prueba de ansiedad

$X_4$  = experiencia, en años

$X_5$  = GPA (promedio de la calificación) de la escuela secundaria

Pam recopila los datos mostrados en la tabla 7.13 y asigna a su analista la tarea de obtener el “mejor” conjunto de variables independientes para pronosticar las habilidades de venta.

El primer paso es obtener, mediante un programa de computadora, la matriz de correlación para todas las variables. Esta matriz aportará el conocimiento esencial sobre las relaciones básicas entre las variables.

El examen de la matriz de correlación de la tabla 7.14 revela que la prueba de aptitud de ventas, edad, experiencia y GPA están relacionadas positivamente con la habilidad de ventas y tienen potencial como variables explicativas adecuadas. El resultado de la prueba de ansiedad muestra una baja correlación negativa con ventas y, probablemente, no sea una variable explicativa importante. Un análisis más completo indica que el GPA y la edad, y la experiencia y la edad están moderadamente correlacionados. Hay que abordar estas interrelaciones cuando se pretende encontrar la mejor serie posible de variables explicativas.

Se muestran dos procedimientos: todas las regresiones posibles y la regresión por pasos.

**TABLA 7.13 Datos de Zurenko Pharmaceutical Company para el ejemplo 7.8**

| Ventas de un mes (unidades) | Resultado de prueba de aptitud | Edad (años) | Resultado de prueba de ansiedad | Experiencia (años) | GPA de la secundaria |
|-----------------------------|--------------------------------|-------------|---------------------------------|--------------------|----------------------|
| 44                          | 10                             | 22.1        | 4.9                             | 0                  | 2.4                  |
| 47                          | 19                             | 22.5        | 3.0                             | 1                  | 2.6                  |
| 60                          | 27                             | 23.1        | 1.5                             | 0                  | 2.8                  |
| 71                          | 31                             | 24.0        | .6                              | 3                  | 2.7                  |
| 61                          | 64                             | 22.6        | 1.8                             | 2                  | 2.0                  |
| 60                          | 81                             | 21.7        | 3.3                             | 1                  | 2.5                  |
| 58                          | 42                             | 23.8        | 3.2                             | 0                  | 2.5                  |
| 56                          | 67                             | 22.0        | 2.1                             | 0                  | 2.3                  |
| 66                          | 48                             | 22.4        | 6.0                             | 1                  | 2.8                  |
| 61                          | 64                             | 22.6        | 1.8                             | 1                  | 3.4                  |
| 51                          | 57                             | 21.1        | 3.8                             | 0                  | 3.0                  |
| 47                          | 10                             | 22.5        | 4.5                             | 1                  | 2.7                  |
| 53                          | 48                             | 22.2        | 4.5                             | 0                  | 2.8                  |
| 74                          | 96                             | 24.8        | .1                              | 3                  | 3.8                  |
| 65                          | 75                             | 22.6        | .9                              | 0                  | 3.7                  |
| 33                          | 12                             | 20.5        | 4.8                             | 0                  | 2.1                  |
| 54                          | 47                             | 21.9        | 2.3                             | 1                  | 1.8                  |
| 39                          | 20                             | 20.5        | 3.0                             | 2                  | 1.5                  |
| 52                          | 73                             | 20.8        | .3                              | 2                  | 1.9                  |
| 30                          | 4                              | 20.0        | 2.7                             | 0                  | 2.2                  |
| 58                          | 9                              | 23.3        | 4.4                             | 1                  | 2.8                  |
| 59                          | 98                             | 21.3        | 3.9                             | 1                  | 2.9                  |
| 52                          | 27                             | 22.9        | 1.4                             | 2                  | 3.2                  |
| 56                          | 59                             | 22.3        | 2.7                             | 1                  | 2.7                  |
| 49                          | 23                             | 22.6        | 2.7                             | 1                  | 2.4                  |
| 63                          | 90                             | 22.4        | 2.2                             | 2                  | 2.6                  |
| 61                          | 34                             | 23.8        | .7                              | 1                  | 3.4                  |
| 39                          | 16                             | 20.6        | 3.1                             | 1                  | 2.3                  |
| 62                          | 32                             | 24.4        | .6                              | 3                  | 4.0                  |
| 78                          | 94                             | 25.0        | 4.6                             | 5                  | 3.6                  |

**TABLA 7.14 Correlaciones: Ventas, aptitud, edad, ansiedad, Exp., GPA**

|          | Sales           | Aptitude        | Age             | Anxiety         | Exp.           |                                                  |
|----------|-----------------|-----------------|-----------------|-----------------|----------------|--------------------------------------------------|
| Aptitude | 0.676<br>0.000  |                 |                 |                 |                | Cell Contents:<br>Pearson correlation<br>P-Value |
| Age      |                 | 0.228<br>0.000  | 0.226           |                 |                |                                                  |
| Anxiety  | -0.296<br>0.112 | -0.222<br>0.238 | -0.287<br>0.124 |                 |                |                                                  |
| Exp.     | 0.550<br>0.002  | 0.350<br>0.058  | 0.540<br>0.002  | -0.279<br>0.136 |                |                                                  |
| GPA      | 0.622<br>0.000  | 0.318<br>0.087  | 0.695<br>0.000  | -0.244<br>0.193 | 0.312<br>0.093 |                                                  |

### Todas las regresiones posibles

El procedimiento requiere la investigación de todas las ecuaciones de regresión posibles que a las variables independientes potenciales. El analista empieza con una ecuación que no contiene ninguna variable independiente y analiza todas las combinaciones posibles para seleccionar la mejor serie de variables explicativas.

Se pueden utilizar diferentes criterios para comparar las diferentes ecuaciones de regresión con *todos los enfoques de regresión posibles*. Aquí, sólo se hablará de la técnica de cuatro pasos  $R^2$ .

Primero, el procedimiento requiere ajustar cada modelo de regresión posible que involucre la variable dependiente y cualquier número de variables independientes. Cada variable independiente puede aparecer o no en la ecuación (dos posibles resultados), y esto se aplica para cada variable independiente. En total, existen  $2^k$  ecuaciones (donde  $k$  es igual al número de variables independientes). De manera que si tenemos ocho variables independientes que considerar ( $k = 8$ ), entonces  $2^8 = 256$  ecuaciones que se deben examinar.

El segundo paso en el procedimiento es dividir las ecuaciones en conjuntos de acuerdo con el número de parámetros que se estimarán.

#### Ejemplo 7.9

Los resultados de todas las regresiones posibles para el ejemplo de Zurenko Pharmaceutical Company se presentan en la tabla 7.15. Observe que la tabla 7.15 está dividida en seis series de resultados de la ecuación de regresión. Este desglose coincide con el número de parámetros que cada ecuación contiene.

El tercer paso requiere la selección de la mejor variable independiente (o variables) para cada grupo de parámetros. Se considera que la ecuación que tiene el más alto  $R^2$  es la mejor. La mejor ecuación de cada serie enlistada en la tabla 7.15 se presenta en la tabla 7.16.

El cuarto paso consiste en tomar la decisión subjetiva: “¿cuál ecuación es la mejor?” Por un lado, el analista desea el  $R^2$  más alto posible; por otro, él o ella desean la ecuación más simple posible. El enfoque de todas las regresiones posibles supone que el número de puntos de datos  $n$  es mayor que el número de parámetros  $k + 1$ .

#### Ejemplo 7-10

El analista intenta encontrar el punto donde no valga la pena agregar variables independientes adicionales al problema de Zurenko Pharmaceutical, porque esto conduce a un aumento muy pequeño en  $R^2$ . Los resultados de la tabla 7.16 indican con claridad que no es necesario agregar

**TABLA 7.15** Valores  $R^2$  para todas las regresiones posibles de Zurenko Pharmaceutical, ejemplo 7.9

| Variables independientes utilizadas | Número de parámetros | Grados de libertad de los errores | $R^2$ |
|-------------------------------------|----------------------|-----------------------------------|-------|
| Ninguna                             | 1                    | 29                                | 0000  |
| $X_1$                               | 2                    | 28                                | .4570 |
| $X_2$                               | 2                    | 28                                | .6370 |
| $X_3$                               | 2                    | 28                                | .0880 |
| $X_4$                               | 2                    | 28                                | .3020 |
| $X_5$                               | 2                    | 28                                | .3870 |
| $X_1, X_2$                          | 3                    | 27                                | .8948 |
| $X_1, X_3$                          | 3                    | 27                                | .4790 |
| $X_1, X_4$                          | 3                    | 27                                | .5690 |
| $X_1, X_5$                          | 3                    | 27                                | .6410 |
| $X_2, X_3$                          | 3                    | 27                                | .6420 |
| $X_2, X_4$                          | 3                    | 27                                | .6570 |
| $X_2, X_5$                          | 3                    | 27                                | .6460 |
| $X_3, X_4$                          | 3                    | 27                                | .3240 |
| $X_3, X_5$                          | 3                    | 27                                | .4090 |
| $X_4, X_5$                          | 3                    | 27                                | .5270 |
| $X_1, X_2, X_3$                     | 4                    | 26                                | .8951 |
| $X_1, X_2, X_4$                     | 4                    | 26                                | .8948 |
| $X_1, X_2, X_5$                     | 4                    | 26                                | .8953 |
| $X_1, X_3, X_4$                     | 4                    | 26                                | .5750 |
| $X_1, X_3, X_5$                     | 4                    | 26                                | .6460 |
| $X_1, X_4, X_5$                     | 4                    | 26                                | .7010 |
| $X_2, X_3, X_4$                     | 4                    | 26                                | .6590 |
| $X_2, X_3, X_5$                     | 4                    | 26                                | .6500 |
| $X_2, X_4, X_5$                     | 4                    | 26                                | .6690 |
| $X_3, X_4, X_5$                     | 4                    | 26                                | .5310 |
| $X_1, X_2, X_3, X_4$                | 5                    | 25                                | .8951 |
| $X_1, X_2, X_3, X_5$                | 5                    | 25                                | .8955 |
| $X_1, X_2, X_4, X_5$                | 5                    | 25                                | .8953 |
| $X_1, X_3, X_4, X_5$                | 5                    | 25                                | .7010 |
| $X_2, X_3, X_4, X_5$                | 5                    | 25                                | .6710 |
| $X_1, X_2, X_3, X_4, X_5$           | 6                    | 24                                | .8955 |

**TABLA 7.16** Mejores ecuaciones de regresión para Zurenko Pharmaceutical, ejemplo 7.9

| Número de parámetros | Variables independientes  | Grados de libertad de los errores | $R^2$ |
|----------------------|---------------------------|-----------------------------------|-------|
| 1                    | Ninguna                   | 29                                | 0000  |
| 2                    | $X_2$                     | 28                                | .6370 |
| 3                    | $X_1, X_2$                | 27                                | .8948 |
| 4                    | $X_1, X_2, X_5$           | 26                                | .8953 |
| 5                    | $X_1, X_2, X_3, X_5$      | 25                                | .8955 |
| 6                    | $X_1, X_2, X_3, X_4, X_5$ | 24                                | .8955 |

variables después de la prueba ( $X_1$ ) y edad ( $X_2$ ). Por lo tanto, la ecuación de regresión ajustada final es la siguiente

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

y explica 89.48% de la variación en  $Y$ .

Draper y Smith (1998) comentan sobre el procedimiento de todas las regresiones posibles:

En general, el análisis de todas las regresiones no está suficientemente justificado. Mientras dice que el investigador ha “buscado todas las posibilidades”, también señala que ha examinado un gran número de ecuaciones de regresión que a menudo se rechazarían. La cantidad de tiempo utilizado en la computadora es un desperdicio y el esfuerzo físico al examinar todas las copias impresas es enorme cuando se exploran muchas variables. Es preferible utilizar alguno de los procedimientos de selección que acorten esta tarea (página 333).

### Regresión por pasos

El procedimiento de regresión por pasos agrega al modelo una variable independiente a la vez, paso a paso. Cuando se utiliza este procedimiento, se puede manejar en la computadora un gran número de variables independientes en una sola corrida.

La regresión por pasos se puede describir mejor enlistando los pasos básicos (algoritmos) involucrados en los cálculos.

1. Se consideran todas las regresiones *simples* posibles. La primera variable explicativa que entra en la ecuación de regresión es la que explica la proporción significativa más grande de la variación en  $Y$  (tiene la más grande correlación con la respuesta).
2. La siguiente variable que entra en la ecuación es (entre las que no están incluidas) la que aporta la contribución más grande a la suma de cuadrados de la regresión y que resulta significativa. La significancia de la contribución se determina por medio de una prueba  $F$ . El valor de la estadística  $F$  que debe sobrepasarse antes de considerar significativa a la variable, a menudo es llamado *F de entrada* (*F to enter*).
3. Una vez que la variable adicional ha sido incluida en la ecuación, se verifican las contribuciones individuales a la suma de cuadrados de la regresión de otras variables, *ya incluidas en la ecuación* usando las pruebas  $F$ . Si la estadística  $F$  es menor que el valor llamado *F por eliminar* (*F to remove*), la variable de la ecuación de regresión es eliminada.
4. Se deben repetir los pasos 2 y 3 hasta que todas las adiciones posibles son no significativas y todas las eliminaciones posibles son significativas. En este punto, la selección se detiene.

La *regresión por pasos* permite que las variables explicativas entren o salgan de la función de regresión en diferentes etapas de su desarrollo. Una variable independiente se elimina del modelo si ya no tiene una contribución significativa cuando se agrega una nueva variable.

El usuario de un programa de regresión por pasos suministra los valores que deciden cuándo debe entrar una variable y cuándo se debe eliminar. Como las estadísticas  $F$  utilizadas en la regresión por pasos son tales que  $F = t^2$ , donde  $t$  es la estadística  $t$  para verificar la significancia de una variable explicativa,  $F = 4$  (correspondiente a  $|t| = 2$ ) es una elección común para ambas, la *F de entrada* y la *F de salida*. Una *F de entrada* de 4 es equi-

valente a probar la significancia de una variable explicativa en el nivel de 5%. El programa por pasos de Minitab permite al usuario escoger un nivel alfa para agregar y eliminar variables en lugar de  $F$ . Una alfa de 0.05 es equivalente a usar un  $F = 4$ .

Como resultado del procedimiento de regresión por pasos se obtiene un modelo que sólo contiene variables independientes con valores  $t$  que son significativos en el nivel especificado. Sin embargo, debido al desarrollo paso a paso no existe garantía de que la regresión por pasos seleccionará, por ejemplo, las tres mejores variables para el pronóstico. Además, un método de selección automática no es capaz de indicar cuándo las transformaciones de variables son útiles, tampoco evita necesariamente el problema de la multicolinealidad. Finalmente, la regresión por pasos no puede crear variables importantes que no son suministradas por el usuario. Es necesario pensar cuidadosamente sobre la recopilación de variables independientes que se le proporcionan al programa de regresión por pasos.

El procedimiento de regresión por pasos se ilustra en el ejemplo 7.11.

### Ejemplo 7.11

“Resolvamos” el problema de Zurenko usando la regresión por pasos.

Pam examina la matriz de correlación mostrada en la tabla 7.14 y decide que, cuando corra el análisis por pasos, la variable de edad entrará primero al modelo porque tiene la correlación más grande con ventas ( $r_{1,3} = 0.798$ ) y explicará 63.7% ( $0.798^2$ ) para la variación en ventas.

Pam nota que el resultado de la prueba de aptitud probablemente será la segunda variable explicativa en entrar en el modelo porque se relaciona mucho con las ventas ( $r_{1,2} = 0.676$ ), pero no tanto con la variable de edad ( $r_{2,3} = 0.228$ ) que ya está en el modelo.

También se da cuenta de que otras variables probablemente no calificarán como buenas variables explicativas. El resultado de la prueba de ansiedad no será una buena variable explicativa porque no se relaciona bien con las ventas ( $r_{1,4} = 0.296$ ). Las variables de experiencia y GPA pueden tener potencial como buenas variables explicativas ( $r_{1,5} = 0.550$ ) y ( $r_{1,6} = 0.622$ ). Sin embargo, ambas variables explicativas tienen un problema potencial de multicolinealidad con la variable de edad ( $r_{3,5} = 0.540$ ) y ( $r_{3,6} = 0.695$ ).

Los comandos de Minitab para correr un análisis de regresión por pasos para este ejemplo se muestran en la sección “Aplicaciones de Minitab”, al final de este capítulo. El resultado de esta regresión por pasos se muestra en la tabla 7.17. El análisis por pasos se realiza de acuerdo con lo siguiente.

**TABLA 7.17 Regresión por pasos: Ventas contra aptitud, edad, ansiedad, exp., GPA**

| Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05     |         |        |
|------------------------------------------------|---------|--------|
| Response is Sales on 5 predictors, with N = 30 |         |        |
| Step                                           | 1       | 2      |
| Constant                                       | -100.85 | -86.79 |
| Age                                            | 6.97    | 5.93   |
| T-Value                                        | 7.01    | 10.60  |
| P-Value                                        | 0.000   | 0.000  |
| Aptitude                                       |         | 0.200  |
| T-Value                                        |         | 8.13   |
| P-Value                                        |         | 0.000  |
| S                                              | 6.85    | 3.75   |
| R-Sq                                           | 63.70   | 89.48  |
| R-Sq(adj)                                      | 62.41   | 88.70  |
| C-p                                            | 57.4    | 0.2    |

**Paso 1.** El modelo después del paso 1 es

$$\text{Ventas} = -100.85 + 6.97 \text{ (edad)}$$

Como Pam pensó, la variable de edad primero entró al modelo y explica 63.7% de la varianza en ventas. Como el valor  $p$  de 0.000 es menor que el valor alfa de edad de 0.05, la edad entra al modelo. Recuerde que el valor  $p$  es la probabilidad de obtener una estadística  $t$  tan grande como 7.01 por casualidad. La regla de decisión de Minitab es agregar una variable si su valor  $p$  es menor que alfa.

Note que  $t = 7.01 > 2.084$ , el punto superior que deja a la derecha un área de 0.025 en una distribución  $t$  con 28 ( $n - k - 1 = 30 - 1 - 1$ ) grados de libertad. De manera que, al nivel de significancia de 0.05, la hipótesis  $H_0: \beta_1 = 0$  se rechaza a favor de  $H_1: \beta_1 \neq 0$ . Porque  $t^2 = F$  o  $2.048^2 = 4.19$ , una  $F$  de entrada de 4 es también equivalente a probar la significancia de una variable explicativa en el nivel de 5%. En este caso, como el coeficiente de la variable de edad es claramente diferente de cero, la edad entra en la ecuación de regresión y el procedimiento sigue con el paso 2.

**Paso 2** El modelo después del paso 2 es

$$\text{Ventas} = -86.79 + 5.93 \text{ (edad)} + 0.200 \text{ (aptitud)}$$

Este modelo explica 89.48% de la variación en ventas.

La hipótesis nula y la alternativa para determinar si el coeficiente de regresión del resultado de la prueba de aptitud es significativamente diferente de cero, son

$$\begin{aligned} H_0: \beta_2 &= 0 \\ H_1: \beta_2 &\neq 0 \end{aligned}$$

De nuevo, el valor  $p$  de 0.000 es menor que el valor alfa de 0.05 y la aptitud se agrega al modelo. El coeficiente de regresión del resultado de la prueba de aptitud es significativamente diferente de cero, y la probabilidad de que esto haya ocurrido por un error de muestreo casual es de aproximadamente cero. Este resultado dice que el resultado de la prueba de aptitud es una variable importante cuando se usa junto con la edad.

La estadística crítica  $t$  basada en 27 ( $n - k - 1 = 30 - 2 - 1$ ) grados de libertad es 2.052.<sup>10</sup> El valor de la  $t$  calculada que se ve en el resultado en Minitab es de 8.13, el cual es más grande que 2.052. Mediante el enfoque de la prueba  $t$  también se rechaza la hipótesis nula.

Observe que el valor  $p$  para la estadística  $t$  de la variable de edad 0.000 sigue siendo muy pequeño. La edad continúa siendo una variable significativa de ventas. El procedimiento ahora continúa en el paso 3.

**Paso 3** La computadora ahora está considerando agregar una tercera variable explicativa dado que  $X_1$  (edad) y  $X_2$  (aptitud) están en la ecuación de regresión. Ninguna de las variables independientes restantes es significativa (tiene un valor  $p$  menor que 0.05) cuando se corre en combinación con  $X_1$  y  $X_2$ , así se completa el procedimiento por pasos de regresión por pasos.

El modelo final de Pam seleccionado con el procedimiento por pasos es el modelo con dos variables explicativas del paso 2.

### Notas finales sobre la regresión por pasos

La técnica de regresión por pasos es muy fácil de usar. Desgraciadamente, es también muy fácil emplearla mal. Los analistas que están desarrollando un modelo de regresión a menudo producen una serie grande de variables independientes potenciales y luego dejan que el procedimiento por pasos determine cuáles son significativas. El problema es que cuando se analiza una serie grande se realizan muchas pruebas  $t$ , y se puede obtener como resultado un error tipo I (que agrega una variable no significativa). Es decir, el modelo final puede contener una variable que no está linealmente relacionada con la variable dependiente y sólo entró en el modelo por casualidad.

<sup>10</sup> De nuevo, como  $2.052^2 = 4.21$ , por lo tanto, el criterio de entrada  $F = 4$  equivale aproximadamente a seleccionar una variable explicativa con el nivel significancia de 0.05.

Como ya se dijo, otro problema es la selección inicial de variables independientes potenciales. Cuando se seleccionan estas variables, a veces se omiten los términos de orden más alto (curvilíneo, no lineal, e interacción) para mantener un número de variables de fácil manejo. En consecuencia, algunas variables importantes pueden ser inicialmente omitidas del modelo. Obviamente, es muy importante la intuición del analista en la selección de las variables independientes iniciales para el desarrollo de un modelo de regresión exitoso.

## DIAGNÓSTICOS DE REGRESIÓN Y ANÁLISIS RESIDUAL

---

Un análisis de regresión no está completo hasta que uno se haya convencido de que el modelo es una representación adecuada de los datos. Es muy importante examinar la suficiencia del modelo *antes* de que se convierta en una parte del aparato de la toma de decisiones.

Un componente crucial de la determinación de la suficiencia del modelo es el examen de los residuales. En el capítulo 6 se mostraron varias gráficas de residuales diseñadas para verificar las suposiciones del modelo de regresión. Además, si los modelos de regresión se utilizan con series de datos de tiempo (esto se verá en el capítulo 8), es importante calcular las autocorrelaciones residuales para verificar la suposición de independencia. Las conclusiones (y decisiones) hechas con modelos que no cumplen con los supuestos de regresión pueden ser terriblemente engañosas. Por ejemplo, se puede concluir que una variable explicativa produce cambios específicos en la respuesta cuando, de hecho, no lo hará. Se puede concluir que un pronóstico tiene muchas posibilidades (95% de confianza) de estar dentro de 2% de la respuesta futura cuando, de hecho, la confianza real es mucho menor, y así sucesivamente.

En esta sección, se muestra cómo se pueden utilizar algunas herramientas adicionales para evaluar un modelo de regresión. Estas herramientas están diseñadas para identificar observaciones atípicas o extremas (observaciones que distan mucho del resto de los datos). A menudo, las observaciones atípicas no se pueden ver durante el proceso de ajuste, tampoco se pueden detectar con un examen de la gráfica de residuales. Incluso, pueden tener un papel relevante para determinar la función de regresión ajustada. Es importante estudiar las observaciones atípicas para decidir si deben mantenerse o eliminarse y, si se mantienen, decidir si se reduce su influencia en el proceso de ajuste o se revisa la función de regresión.

La influencia de los datos  $i$ -ésimos en la posición de la función de regresión ajustada se puede medir mediante el *apalancamiento*  $h_{ii}$ . Éste depende solamente de las explicativas; no depende de las respuestas  $Y$ . Para una regresión lineal simple con la variable explicativa  $X$ ,

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum(X_i - \bar{X})^2}$$

Con  $k$  variables explicativas, la expresión para el apalancamiento  $i$ -ésimo es más complicada; sin embargo, se puede mostrar que  $0 < h_{ii} < 1$ , y que la media de apalancamiento es  $\bar{h} = (k + 1)/n$ .

Si el  $i$ -ésimo dato tiene un apalancamiento grande ( $h_{ii}$  está cerca de 1), la respuesta ajustada  $\hat{Y}_i$  en estas  $X$  se determina casi completamente con  $Y_i$ , y los datos restantes tienen muy poca influencia. El dato con alto apalancamiento es también un dato atípico entre las  $X$  (lejos de otras combinaciones de valores  $X$ ).<sup>11</sup> La regla general sugiere que  $h_{ii}$  es lo bastante grande como para verificarlo si  $h_{ii} \geq 3(k + 1)/n$ .

El descubrimiento de valores  $Y$  atípicos o extremos se basa en el tamaño de los residuales  $e = Y - \hat{Y}$ . Residuales grandes indican un valor  $Y$  que está “lejos” del valor ajustado o

---

<sup>11</sup> Lo opuesto no es necesariamente cierto. Esto quiere decir que un dato atípico entre las  $X$  puede no ser un dato de alto apalancamiento.

pronosticado  $\hat{Y}$ . Un residual grande aparecerá en un histograma de los residuales como un valor lejos de cero (en cualquier dirección). Un residual grande aparecerá en una gráfica de los residuales contra los valores ajustados como un punto muy por arriba o por debajo del eje horizontal.

Los paquetes de software como Minitab marcan los puntos de datos con valores extremos  $Y$  al calcular los residuales “estandarizados” y al identificar puntos con residuales estandarizados grandes.

Una estandarización se basa en el hecho de que los residuales tienen desviaciones estándar estimadas

$$s_{e_i} = s_{y \cdot x'} s \sqrt{1 - h_{ii}}$$

donde  $s_{y \cdot x'} s = \sqrt{MSE}$  es el error estándar de la estimación y  $h_{ii}$  es el apalancamiento asociado con el  $i$ -ésimo dato. Entonces, el *residual estandarizado*<sup>12</sup> es

$$\frac{e_i}{s_{e_i}} = \frac{e_i}{s_{y \cdot x'} s \sqrt{1 - h_{ii}}} \quad (7.13)$$

Todos los residuales estandarizados tienen una varianza de 1. Un residual estandarizado es considerado grande (la respuesta extrema) si

$$\left| \frac{e_i}{s_{e_i}} \right| > 2$$

Los valores  $Y$  correspondientes a los puntos de datos con residuales estandarizados grandes pueden tener una gran influencia en la posición de la función de regresión ajustada.

### Ejemplo 7.2

Los salarios del CEO o presidente de las compañías estadounidenses son de interés por su comparación respecto a los salarios en firmas internacionales y a los salarios de los profesionales más importantes que no trabajan para corporaciones de ese país. Asimismo, para una firma individual, la compensación del CEO tiene una influencia directa o indirecta sobre los salarios de los gerentes en posiciones inferiores a la del CEO. El salario de este último varía mucho de una firma a otra, pero los datos sugieren que el salario se puede explicar en términos de las ventas de una firma, así como la experiencia, nivel de educación, y participación en la propiedad que el CEO tiene en la firma. En un estudio se usaron 50 firmas para desarrollar un modelo de regresión múltiple que conectaba la compensación del CEO con algunas variables explicativas como las ventas, ganancias, edad, experiencia, formación profesional, nivel de educación y participación en la propiedad.

Después de eliminar variables explicativas sin importancia, la función de regresión final ajustada fue

$$\hat{Y} = 5.52 - .467X_1 + .263X_2$$

donde

$Y$  = Logaritmo de pago o compensación al CEO

$X_1$  = Indicador de la variable del nivel de educación

$X_2$  = Logaritmo de las ventas de la compañía

En este análisis de regresión Minitab identificó tres observaciones que tienen residuales estandarizados grandes o un gran apalancamiento.

---

<sup>12</sup> Algunos paquetes de software suelen llamar *Studentized* al residual estandarizado proporcionado en la ecuación 7.13.

**Unusual Observations**

| Obs | Educate | LnComp | Fit    | StDev Fit | Residual | St Resid |
|-----|---------|--------|--------|-----------|----------|----------|
| 14  | 1.00    | 6.0568 | 7.0995 | 0.0949    | -1.0427  | -2.09R   |
| 25  | 0.00    | 8.1342 | 7.9937 | 0.2224    | 0.1405   | 0.31X    |
| 33  | 0.00    | 6.3969 | 7.3912 | 0.2032    | -0.9943  | -2.13R   |

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large influence.

Las observaciones 14 y 33 tienen residuales estandarizados grandes. La función de regresión ajustada predice una compensación (en logaritmo) demasiado grande para estos dos CEO. Un examen de las series de datos completos muestra que cada uno de estos CEO posee un porcentaje relativamente grande de las acciones de su compañía. El caso 14 posee más de 10% de las acciones de la compañía, y el caso 33 más de 17%. Estos individuos reciben una gran parte de su remuneración en compensaciones de largo plazo, como incentivos en acciones, no de su salario anual o bonificaciones. Como la cantidad de acciones que poseen (o valor de la acción) no se incluye como una variable en la función de regresión, esto no se puede utilizar para ajustar el pronóstico de compensación determinado por la educación del CEO y las ventas de la compañía. Aunque educación y ventas (en logaritmo) no pronostican la compensación de estos dos CEO tan bien como los demás, no parece haber una razón para eliminarlos de la consideración.

Se señaló la observación 25 porque el apalancamiento de estos puntos de datos es mayor que  $3(k + 1)/n = 3(3)/50 = 0.18$ . Este CEO no tiene un título universitario (educación = 0), pero su compañía tiene ventas relativamente grandes (LnSales = 9.394). La combinación (0, 9.394) está lejos del punto  $(\bar{X}_1, \bar{X}_2)$ ; por consiguiente, es un dato atípico entre los pares de  $X$ . La respuesta asociada con estas  $X$  tendrá una gran influencia en la determinación de la función de regresión ajustada. (Observe que el residual estandarizado para este punto de datos es pequeño e indica que el logaritmo de la compensación pronosticada, o ajustado, está más cerca del valor real.) Este CEO en particular tiene 30 años de experiencia como tal, solamente otro CEO tiene más experiencia que él en la serie de datos. Esta observación tiene alguna influencia pero no existe razón para eliminarla.

El apalancamiento indica si una observación tiene un dato inusual en las variables explicativas, y un residual estandarizado indica si la observación tiene una respuesta fuera de lo común. Estas cantidades se pueden combinar en una medida de influencia global conocida como la *distancia de Cook*. Las distancias de Cook se pueden imprimir en la mayoría de los paquetes de software de estadística; pero un análisis adicional de este tema está más allá del alcance de este libro.<sup>13</sup>

## ADVERTENCIAS EN EL PRONÓSTICO

Finaliza esta discusión sobre la regresión múltiple con algunos comentarios generales. Estos comentarios están orientados a la aplicación práctica del análisis de regresión.

### Sobreajuste

El *sobreajuste* se refiere al desarrollo de un modelo de regresión que, en gran medida, sigue a las perturbaciones de los datos de muestra que se analizan.

Cuando se aplica un modelo sobreajustado a nuevos conjuntos de datos seleccionados entre la misma población, su pronóstico no es tan bueno como sugeriría el ajuste inicial.

Es más probable que ocurra un sobreajuste cuando la muestra es muy pequeña, especialmente si en el modelo se incluye un gran número de variables independientes. Algunos profesionales han sugerido que deben existir por lo menos 10 observaciones por cada va-

<sup>13</sup> Draper y Smith (1998) aportan un buen análisis sobre la distancia de Cook.

riable independiente. (Si hay cuatro variables independientes, se sugiere una muestra  $n$  de al menos 40.)

Una manera de evitar el sobreajuste es desarrollar la función de regresión con una parte de los datos y luego aplicarlo a una porción de los datos. Luego se utiliza la función de regresión ajustada para pronosticar las respuestas de otra porción de datos y se calculan los errores de pronóstico. Si los errores de pronóstico son considerablemente mayores que los errores del ajuste equiparables, por ejemplo los errores cuadráticos medios, entonces el sobreajuste ha ocurrido.

### Regresiones útiles, valores grandes calculados de $F$

Una regresión que estadísticamente es significativa, no necesariamente es útil. Con una muestra relativamente grande ( $n$  grande relativo a  $k$ , el número de variables explicativas), no es extraño obtener una estadística  $F$  significativa y una  $R^2$  pequeña. Es decir, la regresión es significativa y aun así explica solamente una pequeña proporción de la variación en la respuesta. La regla general sugiere que con un nivel de significancia de 0.05, la proporción  $F$  debe ser de por lo menos cuatro veces el valor crítico correspondiente a fin de que la regresión sea útil para los pronósticos de regresión.<sup>14</sup>

El criterio de las “cuatro veces” viene de la discusión de que el rango de las predicciones (sobre todas las  $X$ ) debe ser cuatro veces el error de pronóstico (promedio) para que la regresión produzca una interpretación útil.<sup>15</sup>

Como ejemplo, con  $k = 3$  variables explicativas,  $n = 25$  observaciones, y un nivel de significancia de 0.05, la  $F$  calculada de la tabla ANOVA tendrá que superar al valor crítico  $F = 3.07$  (vea la tabla C-6 con  $\delta_1 = k = 3$ ,  $\delta_2 = n - k - 1 = 21$  grados de libertad) para que la regresión sea significativa. (Mediante la ecuación 7.7, la  $F = 3.07$  crítica corresponde a un  $R^2$  de alrededor de 30%, un número no muy grande.) No obstante, la regla de las “cuatro veces” sugiere que la  $F$  calculada debe sobrepasar  $4(3.07) = 12.28$  para que la regresión sea útil desde un punto de vista práctico.

---

## APLICACIONES A LA ADMINISTRACIÓN

---

El análisis de regresión múltiple ha sido extensamente usado para pronosticar la actividad económica de varios segmentos de la economía. Muchos de los reportes y pronósticos sobre el futuro de la economía que aparecen en *The Wall Street Journal*, *Fortune*, *Business Week* y otras publicaciones similares se basan en modelos económétricos (de regresión). El gobierno de Estados Unidos utiliza ampliamente los análisis de regresión para predecir ingresos futuros, gastos, niveles de ingreso, tasas de interés, índices de nacimiento y de desempleo, y requerimientos de los beneficios de seguridad social, así como un gran número de acontecimientos. De hecho, casi todos los departamentos más importantes del gobierno estadounidense utilizan las herramientas descritas en este capítulo.

Igualmente, las entidades de negocios han adoptado y, cuando es necesario modificado, el análisis de regresión para ayudarlos a predecir acontecimientos futuros. Pocas firmas han sobrevivido al entorno actual sin un pronóstico bastante exacto de las ventas, gastos, requerimientos de capital y flujos de caja a futuro. Aunque las firmas pequeñas o menos sofisticadas pueden arreglárselas con pronósticos intuitivos, las firmas más grandes o más sofisticadas han recurrido al análisis de regresión para estudiar las relaciones entre algunas variables y determinar cómo estas variables pueden afectar su futuro.

Desgraciadamente, la reputación que el análisis de regresión tiene por su utilidad como una herramienta para pronosticar el futuro opaca una de sus cualidades igualmente importante: su habilidad para ayudar a evaluar y controlar el presente. Como la ecuación

<sup>14</sup> Algunos autores debaten el hecho de que la regla de las “cuatro veces” no es suficiente y se debe reemplazar por un criterio de “diez veces”.

<sup>15</sup> Esto supone que no se detectó ningún otro defecto en el ajuste.

de regresión ajustada le proporciona al investigador información de *fuerza y dirección*, la administración puede evaluar y cambiar sus estrategias actuales.

Suponga, por ejemplo, que un fabricante de jaleas quiere saber hacia dónde dirigir sus esfuerzos de mercadeo cuando presente un nuevo sabor. El análisis de regresión se puede utilizar para ayudar a determinar el perfil de los consumidores de jaleas. Por ejemplo, una compañía puede tratar de pronosticar el número de sabores de jalea que un hogar puede tener en un tiempo determinado, basándose en un número de variables independientes como:

- Número de niños que viven en la casa
- Edad de los niños
- Género de los niños
- Propiedad de la casa contra renta
- Tiempo que pasan comprando
- Ingresos

El ejemplo de la jalea, aun mediante una reflexión superficial, rápidamente le indica al investigador que el análisis de regresión contiene numerosas posibilidades que se pueden usar en los estudios de segmentación del mercado. De hecho, muchas compañías usan la regresión para estudiar segmentos de mercado y determinar qué variables podrían tener un impacto en la participación del mercado, frecuencia de compra, participación de la propiedad, la fidelidad del producto y de la marca, así como muchas otras áreas.

Los científicos agrícolas utilizan el análisis de regresión para explorar la relación entre el rendimiento del producto (por ejemplo, el número de fanegas de maíz por acre) y el tipo y cantidad de los fertilizantes, lluvia, temperatura, días soleados y plagas de insectos. Las granjas modernas están equipadas con varias computadoras con paquetes de software para ayudarlos en este proceso.

Los investigadores médicos utilizan el análisis de regresión para buscar vínculos entre la presión arterial y variables independientes como la edad, clase social, peso, tabaquismo y raza. Los doctores exploran el impacto de las comunicaciones, número de contactos y edad del paciente, con la satisfacción del paciente con el servicio.

Los directores de personal exploran las relaciones de los niveles de salario de los empleados con la ubicación geográfica, tasas de desempleo, crecimiento de la industria, afiliación a sindicato, tipo de industria o salarios competitivos. Los analistas financieros buscan las causas de los elevados precios en las acciones al analizar el reparto de dividendos, ganancias por acción, división en tipos de acciones, expectativas del consumidor en las tasas de interés, niveles de ahorro y tasas de inflación.

Frecuentemente, los gerentes de publicidad tratan de estudiar el impacto de los presupuestos de publicidad, selección de medios, copia de mensajes, frecuencia de publicación, o a quién eligen como portavoz para cambiar la actitud del consumidor. De modo parecido, los vendedores intentan determinar las ventas al tomar los gastos de publicidad, niveles de precios, gastos de mercadeo competitivo y el ingreso disponible del consumidor, así como una gran variedad de diversas variables.

Un ejemplo final ilustra la versatilidad del análisis de regresión. Los analistas de ubicación de bienes raíces han descubierto que el análisis de regresión puede ser muy útil para determinar con exactitud las áreas geográficas con una alta o baja penetración de cierta clase de tiendas que venden al menudeo. Por ejemplo, una cadena de ferreterías puede buscar una ciudad potencial donde abrir una nueva tienda mediante el desarrollo de un modelo de regresión diseñado para predecir las ventas de ferretería en cualquier urbe. Los investigadores pueden concentrar sus esfuerzos en las ciudades donde el modelo pronosticó ventas más altas que las logradas en realidad (como se puede determinar mediante muchas fuentes). La hipótesis es que las ventas de ferretería no tienen potencial en estas ciudades.

En resumen, el análisis de regresión le ha proporcionado a la administración una versátil y poderosa herramienta para estudiar las relaciones entre una variable dependiente y múltiples variables independientes. La meta es entender mejor y controlar los sucesos actuales, así como poder predecir mejor los acontecimientos futuros.

## Glosario

**Coeficiente de regresión parcial, o neto.** El coeficiente de regresión parcial, o neto, mide el cambio promedio en la variable dependiente por unidad de cambio en la variable independiente pertinente cuando se mantienen constantes a las otras variables independientes.

**Error estándar de la estimación.** El error estándar de la estimación es la desviación estándar de los residuales. Mide la cantidad en que los valores reales ( $Y$ ) difieren de los valores estimados ( $\hat{Y}$ ).

**Multicolinealidad.** La multicolinealidad es la situación donde las variables independientes están sumamente intercorrelacionadas en una ecuación de regresión múltiple.

**Regresión múltiple.** La regresión múltiple involucra el uso de más de una variable independiente para predecir una variable dependiente.

**Regresión por pasos.** La regresión por pasos es un procedimiento para seleccionar la “mejor” función de regresión, agregando o eliminando variables independientes en las diferentes etapas de su desarrollo.

**Sobreajuste.** El sobreajuste consiste en agregar demasiadas variables independientes a la función de regresión que su cantidad se acerque al número de datos de muestra bajo análisis.

**Variables ficticias.** Las variables ficticias se usan para determinar las relaciones entre variables independientes cualitativas y una variable dependiente.

## Fórmulas clave

### Función de regresión múltiple de la población

$$\mu_Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \quad (7.1)$$

### Función de regresión estimada (ajustada)

$$\hat{Y} = b_0 + b_1 X_1 + \cdots + b_k X_k \quad (7.2)$$

### Descomposición de la suma de cuadrados y grados de libertad asociados

$$\begin{aligned} \Sigma(Y - \bar{Y})^2 &= \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2 \\ SST &= SSR + SSE \\ df: n - 1 &= k + n - k - 1 \end{aligned} \quad (7.3)$$

### Error estándar de la estimación

$$s_{y \cdot x's} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - k - 1}} = \sqrt{\frac{SSE}{n - k - 1}} = \sqrt{MSE} \quad (7.4)$$

### Estadística $F$ para probar la significancia de la regresión

$$F = \frac{MSR}{MSE} \quad (\text{ver recuadro en la página 277})$$

### Coeficiente de determinación

$$\begin{aligned} R^2 &= \frac{SSR}{SST} = \frac{\Sigma(\hat{Y} - \bar{Y})^2}{\Sigma(Y - \bar{Y})^2} \\ &= 1 - \frac{SSE}{SST} = 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2} \end{aligned} \quad (7.5)$$

**Coefficiente de correlación múltiple**

$$R = \sqrt{R^2} \quad (7.6)$$

**Relación entre la estadística  $F$  y  $R^2$** 

$$F = \frac{R^2}{1 - R^2} \left( \frac{n - k - 1}{k} \right) \quad (7.7)$$

**Coefficiente ajustado de determinación**

$$\bar{R}^2 = 1 - (1 - R^2) \left( \frac{n - 1}{n - k - 1} \right) \quad (7.8)$$

**Estadística  $t$  para probar  $H_0: \beta_j = 0$** 

$$t = \frac{b_j}{s_{b_j}} \quad (\text{ver recuadro sombreado en la página 279})$$

**Pronóstico de un valor futuro**

$$\hat{Y}^* = b_0 + b_1 X_1^* + b_2 X_2^* + \cdots + b_k X_k^* \quad (7.9)$$

**Intervalo de confianza para la predicción al utilizar una muestra grande para una respuesta futura.**

$$(\hat{Y}^* - t_{\alpha/2} s_{y \cdot x' s}, \hat{Y}^* + t_{\alpha/2} s_{y \cdot x' s}) \quad (7.10)$$

**Factor de expansión de la varianza**

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, 2, \dots, k \quad (7.11)$$

**Valores estandarizados de la variable independiente**

$$\tilde{X}_j = \frac{X_{ij} - \bar{X}_j}{\sqrt{\sum_i (X_{ij} - \bar{X}_j)^2}} \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, n \quad (7.12)$$

**Residual estandarizado**

$$\frac{e_i}{s_{e_i}} = \frac{e_i}{s_{y \cdot x' s} \sqrt{1 - h_{ii}}} \quad (7.13)$$

**Problemas**

1. ¿Cuáles son las características de una variable explicativa adecuada?
2. ¿Cuáles son los supuestos asociados al modelo de regresión múltiple?
3. ¿Qué mide el coeficiente de regresión parcial, o neto, en la regresión múltiple?
4. ¿Qué mide el error estándar de la estimación en la regresión múltiple?
5. La ecuación estimada de regresión múltiple es  $\hat{Y} = 7.52 + 3X_1 - 12.2X_2$ . Pronostique el valor de  $Y$  si  $X_1 = 20$  y  $X_2 = 7$ .

TABLA P.7

| Número de la variable | Número de la variable |      |      |      |      |      |
|-----------------------|-----------------------|------|------|------|------|------|
|                       | 1                     | 2    | 3    | 4    | 5    | 6    |
| 1                     | 1.00                  | .55  | .20  | -.51 | .79  | .70  |
| 2                     |                       | 1.00 | .27  | .09  | .39  | .45  |
| 3                     |                       |      | 1.00 | .04  | .17  | .21  |
| 4                     |                       |      |      | 1.00 | -.44 | -.14 |
| 5                     |                       |      |      |      | 1.00 | .69  |
| 6                     |                       |      |      |      |      | 1.00 |

6. Explique cada uno de los siguientes conceptos:
- Matriz de correlación
  - $R^2$
  - Multicolinealidad
  - Residual
  - Variable ficticia
  - Regresión por pasos
7. La mayoría de las soluciones en computadora para la regresión múltiple empiezan con una matriz de correlación. Cuando se analiza un problema de regresión que involucra más de una variable dependiente, el primer paso generalmente es examinar esta matriz. Conteste las siguientes preguntas sobre la matriz de correlación mostrada en la tabla P.7.
- ¿Por qué todos los datos en la diagonal principal son igual a 1?
  - ¿Por qué la mitad inferior de la matriz está debajo de la diagonal principal está en blanco?
  - Si la variable 1 es la variable dependiente, ¿cuáles variables independientes tienen el grado más alto de asociación lineal con la variable 1?
  - ¿Qué tipo de asociación existe entre las variables 1 y 4?
  - ¿Existe alguna evidencia de multicolinealidad en esta matriz de correlación?
  - En su opinión, ¿qué variable o variables deben ser incluidas en el mejor modelo de pronóstico?
  - Si los datos proporcionados en la matriz de correlación anterior se corren en un programa por pasos, ¿qué variable independiente (2, 3, 4, 5, o 6) será la primera en entrar a la función de regresión?
8. Jennifer Dahl, supervisora de la cadena de tiendas de descuento Circle O, quiere pronosticar el tiempo que le toma a un cliente pagar en la caja. Decide usar las siguientes variables independientes: número de artículos comprados y la cantidad total de la compra. Entonces, recolecta los datos de muestra de 18 clientes, como se ve en la tabla P.8.
- Determine la mejor ecuación de regresión.
  - Cuando se compra un artículo adicional, ¿cuál es el aumento promedio en el tiempo que tardan en pagar en la caja?
  - Calcule el residual del cliente 18.
  - Calcule el error estándar de la estimación.
  - Interprete el inciso d en los términos de las variables utilizadas en este problema.
  - Calcule un pronóstico del tiempo para pagar si un cliente compra 14 artículos y paga una cantidad de 70 dólares.
  - Calcule el intervalo estimado de 95% para su predicción del inciso f.
  - ¿A qué conclusión debe llegar Jennifer?
9. La tabla P.9 contiene datos de gastos en alimentos, ingreso anual y el tamaño de la familia para un ejemplo de 10 familias.

TABLA P.8

| <i>Cliente</i> | <i>Tiempo para pagar<br/>(minutos)</i> | <i>Cantidad<br/>(dólares)</i> | <i>Número de<br/>artículos</i> |
|----------------|----------------------------------------|-------------------------------|--------------------------------|
|                | <i>Y</i>                               | <i>X<sub>1</sub></i>          | <i>X<sub>2</sub></i>           |
| 1              | 3.0                                    | 36                            | 9                              |
| 2              | 1.3                                    | 13                            | 5                              |
| 3              | .5                                     | 3                             | 2                              |
| 4              | 7.4                                    | 81                            | 14                             |
| 5              | 5.9                                    | 78                            | 13                             |
| 6              | 8.4                                    | 103                           | 16                             |
| 7              | 5.0                                    | 64                            | 12                             |
| 8              | 8.1                                    | 67                            | 11                             |
| 9              | 1.9                                    | 25                            | 7                              |
| 10             | 6.2                                    | 55                            | 11                             |
| 11             | .7                                     | 13                            | 3                              |
| 12             | 1.4                                    | 21                            | 8                              |
| 13             | 9.1                                    | 121                           | 21                             |
| 14             | .9                                     | 10                            | 6                              |
| 15             | 5.4                                    | 60                            | 13                             |
| 16             | 3.3                                    | 32                            | 11                             |
| 17             | 4.5                                    | 51                            | 15                             |
| 18             | 2.4                                    | 28                            | 10                             |

TABLA P.9

| <i>Familia</i> | <i>Gastos anuales<br/>en alimentos<br/>(\$100's)</i> | <i>Ingreso<br/>anual<br/>(\$1000's)</i> | <i>Tamaño de<br/>la familia</i> |
|----------------|------------------------------------------------------|-----------------------------------------|---------------------------------|
|                | <i>Y</i>                                             | <i>X<sub>1</sub></i>                    | <i>X<sub>2</sub></i>            |
| A              | 24                                                   | 11                                      | 6                               |
| B              | 8                                                    | 3                                       | 2                               |
| C              | 16                                                   | 4                                       | 1                               |
| D              | 18                                                   | 7                                       | 3                               |
| E              | 24                                                   | 9                                       | 5                               |
| F              | 23                                                   | 8                                       | 4                               |
| G              | 11                                                   | 5                                       | 2                               |
| H              | 15                                                   | 7                                       | 2                               |
| I              | 21                                                   | 8                                       | 3                               |
| J              | 20                                                   | 7                                       | 2                               |

- a) Construya la matriz de correlación para las tres variables de la tabla P.9. Interprete las correlaciones en la matriz.
- b) Ajuste un modelo de regresión múltiple al relacionar los gastos en alimentos con el ingreso y el tamaño de la familia. Interprete los coeficientes de regresiones parciales del ingreso y tamaño de la familia. ¿Parecen razonables?
- c) Calcule los factores de expansión de la varianza (VIF) para las variables independientes. Con estos datos, ¿existe un problema de multicolinealidad?, si es así, ¿cómo modificaría el modelo de regresión?
10. Las ventas de cerveza de la Shapiro One-Stop Store se analizan usando como variables independientes la temperatura y el número de personas (de 21 años de edad o más) que están en la calle. Se seleccionó en forma aleatoria 20 días para medir estas tres variables.

**TABLA P.10 Resultado en Minitab**

Correlations (Pearson)

|    | Y     | X1    |
|----|-------|-------|
| X1 | 0.827 |       |
| X2 | 0.822 | 0.680 |

## Regression Analysis

The regression equation is  
 $Y = -26.7 + .782 X_1 + .068 X_2$

| Predictor | Coef    | SE Coef | T | P |
|-----------|---------|---------|---|---|
| Constant  | -26.706 |         |   |   |
| X1        | .78207  | .22694  |   |   |
| X2        | .06795  | .02026  |   |   |

S = R-Sq = R-Sq(adj)=

## Analysis of Variance

| Source         | DF | SS        | MS       | F     |
|----------------|----|-----------|----------|-------|
| Regression     | 2  | 11589.035 | 5794.516 | 36.11 |
| Residual Error | 17 | 2727.914  | 160.466  |       |
| Total          | 19 | 14316.949 |          |       |

 $Y$  = Número de paquetes de seis cervezas vendidos al día $X_1$  = Temperatura máxima diaria. $X_2$  = Conteo diario de tráfico

Los datos son analizados mediante un análisis de regresión múltiple. El resultado parcial en computadora aparece en la tabla P.10.

- Analice la matriz de correlación.
  - Pruebe la significancia de los coeficientes parciales de regresión en un nivel de significancia de 0.01.
  - Pronostique el volumen de cerveza vendida un día en que la temperatura más alta es de 60 grados y el conteo de tráfico es de 500 personas.
  - Calcule  $R^2$  e interprete su significado en los términos de este problema.
  - Calcule el error estándar de la estimación.
  - Explique cómo son afectadas las ventas de cerveza con el incremento de un grado en la temperatura máxima.
  - Manifieste sus conclusiones para este análisis sobre la exactitud de la ecuación de pronóstico y también sobre las contribuciones de las variables independientes.
11. Una compañía de taxis está interesada en la relación entre el millaje, calculado en millas por galón, y la edad de su flota de automóviles. Los 12 vehículos de la flota son del mismo tamaño y fabricante, y se encuentran en buenas condiciones de operación como resultado de un mantenimiento regular. La compañía emplea a conductores de sexo indistinto y se cree que una parte de la variabilidad en el millaje se debe a las diferencias en las técnicas de manejo entre los grupos de conductores del sexo opuesto. De hecho, siendo otras cosas iguales, las mujeres tienden a obtener mejor millaje que los hombres. Los datos se generan al azar asignando los 12 autos a cinco mujeres y siete hombres, y calculando las millas por galón después de 300 millas. Los datos aparecen en la tabla P.11.
- Construya un diagrama de dispersión con  $Y$  como el eje vertical y  $X_1$  como el eje horizontal. Identifique los puntos correspondientes a conductores femeninos y masculinos.

TABLA P.11

| <b><i>Y</i></b><br><b>(MPG)</b> | <b><i>X</i><sub>1</sub></b><br><b>(Edad del auto)</b> | <b><i>X</i><sub>2</sub></b><br><b>(Género: 0 = masculino, 1 = femenino)</b> |
|---------------------------------|-------------------------------------------------------|-----------------------------------------------------------------------------|
| 22.3                            | 3                                                     | 0                                                                           |
| 22.0                            | 4                                                     | 1                                                                           |
| 23.7                            | 3                                                     | 1                                                                           |
| 24.2                            | 2                                                     | 0                                                                           |
| 25.5                            | 1                                                     | 1                                                                           |
| 21.1                            | 5                                                     | 0                                                                           |
| 20.6                            | 4                                                     | 0                                                                           |
| 24.0                            | 1                                                     | 0                                                                           |
| 26.0                            | 1                                                     | 1                                                                           |
| 23.1                            | 2                                                     | 0                                                                           |
| 24.8                            | 2                                                     | 1                                                                           |
| 20.2                            | 5                                                     | 0                                                                           |

TABLA P.12

| <b>Región</b> | <b>Ventas anuales<br/>(millones)</b> | <b>Número de<br/>puntos de venta</b> | <b>Número de automóviles<br/>registrados (millones)</b> |
|---------------|--------------------------------------|--------------------------------------|---------------------------------------------------------|
|               | <b>Y</b>                             | <b>X</b> <sub>1</sub>                | <b>X</b> <sub>2</sub>                                   |
| 1             | 52.3                                 | 2011                                 | 24.6                                                    |
| 2             | 26.0                                 | 2850                                 | 22.1                                                    |
| 3             | 20.2                                 | 650                                  | 7.9                                                     |
| 4             | 16.0                                 | 480                                  | 12.5                                                    |
| 5             | 30.0                                 | 1694                                 | 9.0                                                     |
| 6             | 46.2                                 | 2302                                 | 11.5                                                    |
| 7             | 35.0                                 | 2214                                 | 20.5                                                    |
| 8             | 3.5                                  | 125                                  | 4.1                                                     |
| 9             | 33.1                                 | 1840                                 | 8.9                                                     |
| 10            | 25.2                                 | 1233                                 | 6.1                                                     |
| 11            | 38.2                                 | 1699                                 | 9.5                                                     |

b) Ajuste el modelo de regresión

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

e interprete el coeficiente de mínimos cuadrados  $\beta_2$ .

- c) Calcule los valores ajustados para cada par  $(X_1, X_2)$ , y grafique los valores adaptados en el diagrama de dispersión. Dibuje líneas rectas a través de los valores ajustados para conductores masculinos y femeninos, respectivamente. Especifique las ecuaciones para estas dos líneas rectas.
- d) Suponga que ignoramos el género. Ajuste el modelo de regresión lineal simple  $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ , y grafique la línea recta ajustada en el diagrama de dispersión. ¿Es importante incluir los efectos del género en este caso? Explique.
12. El gerente de ventas de un importante distribuidor de partes automotrices, Hartman Auto Supplies, quiere desarrollar un modelo para predecir las ventas anuales totales de una región desde mayo. Si se pueden pronosticar las ventas regionales, entonces también se podrán pronosticar las ventas totales de la compañía. Las dos variables independientes investigadas son el número de puntos de venta en la región que tienen en existencia productos de la compañía y el número de automóviles registrado para cada región desde el 1 de mayo. Los datos aparecen en la tabla P.12.

- a) Analice la matriz de correlación.
- b) ¿Cuál es el nivel de error en la predicción para la región 1?
- c) Pronostique las ventas anuales de la región 12, con 2500 puntos de venta y 20.2 millones de automóviles registrados.
- d) Discuta la precisión del pronóstico realizado en el inciso c.
- e) Muestre cómo se calculó el error estándar de la estimación.
- f) Proporcione una interpretación del coeficiente de regresión parcial. ¿Estos coeficientes de regresión son razonables?
- g) ¿Cómo se puede mejorar esta ecuación de regresión?
13. El gerente de ventas de Hartman Auto Supplies decide investigar una nueva variable dependiente, la del ingreso personal por región (vea el problema 12). Los datos para esta nueva variable se presentan en la tabla P.13.
- a) ¿El ingreso personal por región tiene alguna contribución en el pronóstico de ventas?
- b) Pronostique las ventas anuales de la región 12 con un ingreso personal de 40 mil millones de dólares, usando las tres variables independientes.
- c) Discuta sobre la exactitud del pronóstico realizado en el inciso b.
- d) ¿Qué variables independientes incluiría en su modelo de pronóstico final? ¿Por qué?
14. Nelson Corporation decide desarrollar una ecuación de regresión múltiple para pronosticar el desempeño en ventas. Se entrevista a 14 vendedores seleccionados al azar y se les realiza una prueba de aptitud. Además, se calcula un índice de esfuerzo consumido para cada uno de los vendedores con base en la proporción del millaje del automóvil de estos vendedores con el millaje total proyectado para una cobertura adecuada de territorio. El análisis de regresión produce los siguientes resultados:

$$\hat{Y} = 16.57 + .65 X_1 + 20.6 X_2$$

(.05)                    (1.69)

Las cantidades entre paréntesis son los errores estándar de los coeficientes de regresión parciales. El error estándar de la estimación es de 3.56. La desviación estándar de las variables de ventas es  $S_y = 16.57$ . Las variables son

$Y$  = Desempeño en ventas, en miles

$X_1$  = Resultado de la prueba de aptitud

$X_2$  = Índice de esfuerzo

- a) ¿Los coeficientes de regresión parcial son significativamente diferentes de cero, en un nivel de significancia de 0.01?
- b) Interprete el coeficiente de regresión parcial del índice de esfuerzo.

TABLA P.13

| Región | Ingreso personal<br>(miles de millones) | Región | Ingreso personal<br>(miles de millones) |
|--------|-----------------------------------------|--------|-----------------------------------------|
| 1      | \$98.5                                  | 7      | 67.6                                    |
| 2      | 31.1                                    | 8      | 19.7                                    |
| 3      | 34.8                                    | 9      | 67.9                                    |
| 4      | 32.7                                    | 10     | 61.4                                    |
| 5      | 68.8                                    | 11     | 85.6                                    |
| 6      | 94.7                                    |        |                                         |

- c) Pronostique el desempeño en ventas de un vendedor que tiene un resultado en la prueba de aptitud de 75 y un índice de esfuerzo de 0.5.
- d) Calcule la suma de los cuadrados residuales  $\sum(Y - \hat{Y})^2$ .
- e) Calcule la suma total de los cuadrados  $\sum(Y - \bar{Y})^2$ .
- f) Calcule  $R^2$  e interprételo con los términos de este problema.
- g) Calcule el coeficiente de determinación ajustado  $\bar{R}^2$ .
15. Se puede esperar que en una misma tienda las compras con tarjeta de crédito difieran de las compras en efectivo. La tabla P.15 contiene las ventas diarias brutas en efectivo y el número de artículos vendidos para cada compra en efectivo, y las ventas diarias brutas con tarjeta de crédito y el número de artículos vendidos en compras con tarjeta de crédito en la misma tienda por 25 días consecutivos.
- a) Elabore un diagrama de dispersión de las ventas diarias brutas  $Y$  contra los artículos vendidos  $X_1$  en compras en efectivo. Use un símbolo o color diferente en la gráfica para agregar las ventas diarias brutas y los artículos adquiridos con tarjeta de crédito. Compare visualmente la relación entre las ventas y el número de artículos comprados en efectivo con las compras con tarjeta de crédito.
- b) Defina la variable ficticia

$$X_2 = \begin{cases} 1 & \text{si para compras en efectivo} \\ 0 & \text{si para compras con tarjeta de crédito} \end{cases}$$

y ajuste el modelo de regresión

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

TABLA P.15

| Día | Efectivo bruto | Número de artículos | Ventas brutas con tarjeta de crédito | Número de artículos |
|-----|----------------|---------------------|--------------------------------------|---------------------|
| 1   | \$348          | 55                  | \$148                                | 4                   |
| 2   | 42             | 8                   | 111                                  | 6                   |
| 3   | 61             | 9                   | 62                                   | 7                   |
| 4   | 94             | 16                  | 0                                    | 0                   |
| 5   | 60             | 11                  | 39                                   | 5                   |
| 6   | 165            | 26                  | 7                                    | 1                   |
| 7   | 126            | 27                  | 143                                  | 26                  |
| 8   | 111            | 19                  | 27                                   | 5                   |
| 9   | 26             | 5                   | 14                                   | 2                   |
| 10  | 109            | 18                  | 71                                   | 12                  |
| 11  | 180            | 27                  | 116                                  | 21                  |
| 12  | 212            | 36                  | 50                                   | 9                   |
| 13  | 58             | 10                  | 13                                   | 2                   |
| 14  | 115            | 20                  | 105                                  | 16                  |
| 15  | 15             | 8                   | 19                                   | 3                   |
| 16  | 97             | 15                  | 44                                   | 14                  |
| 17  | 61             | 10                  | 0                                    | 0                   |
| 18  | 85             | 15                  | 24                                   | 3                   |
| 19  | 157            | 24                  | 144                                  | 10                  |
| 20  | 88             | 15                  | 63                                   | 11                  |
| 21  | 96             | 19                  | 0                                    | 0                   |
| 22  | 202            | 33                  | 14                                   | 3                   |
| 23  | 108            | 23                  | 0                                    | 0                   |
| 24  | 158            | 21                  | 24                                   | 4                   |
| 25  | 176            | 43                  | 253                                  | 28                  |

- c) Analice el ajuste en el inciso *b*. Asegúrese de incluir un análisis de los residuales. ¿Está usted contento con su modelo?
- d) Mediante el modelo ajustado del inciso *b*, genere un pronóstico de ventas diarias para un individuo que compra 25 artículos y paga en efectivo. Construya, para muestra grande, un intervalo de 95% para la predicción para las ventas diarias.
- e) Describa la naturaleza de la función ajustada del inciso *b*. ¿Cree usted que es mejor ajustar dos líneas rectas separadas a los datos de la tabla P.15, una para las ventas en efectivo y otra para las ventas con tarjeta de crédito? Discútalo.
16. Cindy Lawson acaba de comprar un equipo de béisbol de las grandes ligas. Ha recibido consejos sobre lo que debe hacer para crear un equipo ganador. Cindy le pide a usted estudiar el problema y escribirle un reporte. Usted decide usar el análisis de regresión múltiple para determinar cuáles son las estadísticas importantes para desarrollar un equipo ganador (medido por el número de juegos ganados durante la temporada de 1991). Usted reúne las siguientes estadísticas del *Sporting News 1992 Baseball Yearbook*, como se muestra en la tabla P.16, y corre un programa de regresión por pasos.
- Discuta la importancia de cada variable.
  - ¿Qué ecuación debe utilizar Cindy para predecir las victorias?
  - Escriba un reporte para Cindy.
  - Recolete los datos del más reciente *Sporting News Baseball Yearbook* o de otra fuente de estadísticas de béisbol. Corra una regresión por pasos y compare sus resultados.
17. La señora Haight, una corredora de bienes raíces, desea predecir la importancia de cuatro factores para determinar los precios de los lotes. Ella ha acumulado datos de precio, área, altitud y pendiente, y establece un índice para la vista de 50 lotes. Corre

TABLA P.16

| Equipo         | Victorias | ERA  | SO   | BA   | Carreras | HR  | SB  |
|----------------|-----------|------|------|------|----------|-----|-----|
| Gigantes       | 75        | 4.03 | 905  | .246 | 649      | 141 | 95  |
| Mets           | 77        | 3.56 | 1028 | .244 | 640      | 117 | 153 |
| Cachorros      | 77        | 4.03 | 927  | .253 | 695      | 159 | 123 |
| Rojos          | 74        | 3.83 | 997  | .258 | 689      | 164 | 124 |
| Piratas        | 98        | 3.44 | 919  | .263 | 768      | 126 | 124 |
| Cardenales     | 84        | 3.69 | 822  | .255 | 651      | 68  | 202 |
| Phillies       | 78        | 3.86 | 988  | .241 | 629      | 111 | 92  |
| Astros         | 65        | 4.00 | 1033 | .244 | 605      | 79  | 125 |
| Dodgers        | 93        | 3.06 | 1028 | .253 | 665      | 108 | 126 |
| Expos          | 71        | 3.64 | 909  | .246 | 579      | 95  | 221 |
| Bravos         | 94        | 3.49 | 969  | .258 | 749      | 141 | 165 |
| Padres         | 84        | 3.57 | 921  | .244 | 636      | 121 | 101 |
| Medias Rojas   | 84        | 4.01 | 999  | .269 | 731      | 126 | 59  |
| Medias Blancas | 87        | 3.79 | 923  | .262 | 758      | 139 | 134 |
| Yankees        | 71        | 4.42 | 936  | .256 | 674      | 147 | 109 |
| Tigres         | 84        | 4.51 | 739  | .247 | 817      | 209 | 109 |
| Orioles        | 67        | 4.59 | 868  | .254 | 686      | 170 | 50  |
| Cerveceros     | 83        | 4.14 | 859  | .271 | 799      | 116 | 106 |
| Indios         | 57        | 4.23 | 862  | .254 | 576      | 79  | 84  |
| Azulejos       | 91        | 3.50 | 971  | .257 | 684      | 133 | 148 |
| Marineros      | 83        | 3.79 | 1003 | .255 | 702      | 126 | 97  |
| Rangers        | 85        | 4.47 | 1022 | .270 | 829      | 177 | 102 |
| Atléticos      | 84        | 4.57 | 892  | .248 | 760      | 159 | 151 |
| Reales         | 82        | 3.92 | 1004 | .264 | 727      | 117 | 119 |
| Serafines      | 81        | 3.69 | 990  | .255 | 653      | 115 | 94  |
| Mellizos       | 95        | 3.69 | 876  | .280 | 776      | 140 | 107 |

TABLA P.17

| Variable  | Variable |      |         |           |       |
|-----------|----------|------|---------|-----------|-------|
|           | Precio   | Área | Altitud | Pendiente | Vista |
| Precio    | 1.00     | .59  | .66     | .68       | .88   |
| Área      |          | 1.00 | .04     | .64       | .41   |
| Altitud   |          |      | 1.00    | .13       | .76   |
| Pendiente |          |      |         | 1.00      | .63   |
| Vista     |          |      |         |           | 1.00  |

TABLA P.18

| $X_1$ | $X_2$ | $X_3$ | $Y$ | $X_1$ | $X_2$ | $X_3$ | $Y$ |
|-------|-------|-------|-----|-------|-------|-------|-----|
| 87    | 85    | 2.7   | 91  | 93    | 60    | 3.2   | 54  |
| 100   | 84    | 3.3   | 90  | 92    | 69    | 3.1   | 63  |
| 91    | 82    | 3.5   | 83  | 100   | 86    | 3.6   | 96  |
| 85    | 60    | 3.7   | 93  | 80    | 87    | 3.5   | 89  |
| 56    | 64    | 2.8   | 43  | 100   | 96    | 3.8   | 97  |
| 81    | 48    | 3.1   | 75  | 69    | 51    | 2.8   | 50  |
| 77    | 67    | 3.1   | 63  | 80    | 75    | 3.6   | 74  |
| 86    | 73    | 3.0   | 78  | 74    | 70    | 3.1   | 58  |
| 79    | 90    | 3.8   | 98  | 79    | 66    | 2.9   | 87  |
| 96    | 69    | 3.7   | 99  | 95    | 83    | 3.3   | 57  |

los datos en un programa de correlación y obtiene la matriz de correlación que está en la tabla P.17. Entonces, la señora Haight corre los datos en un programa de regresión múltiple por pasos.

- Determine en qué orden entrarían las variables en el modelo, primero, segundo, tercero y último.
  - ¿Qué variable o variables se incluirían en la mejor ecuación de pronóstico?
18. En la tabla P.18 se muestran los resultados  $X_1$  y  $X_2$  de dos exámenes parciales aplicados en un periodo escolar, el puntaje GPA (GPA, *grade point average*)  $X_3$ , y el resultado del examen final  $Y$  para 20 estudiantes en una clase de estadísticas de negocios.
- Ajuste un modelo de regresión múltiple para pronosticar el resultado del examen final mediante los resultados de los exámenes parciales del periodo escolar y el GPA. ¿La regresión es significativa? Explique.
  - Pronostique el resultado del examen final de un estudiante con resultados de los exámenes parciales del periodo escolar de 86 y 77, y un GPA de 3.4.
  - Calcule los VIF y examine las estadísticas  $t$  para verificar la significancia de las variables explicativas individuales. ¿Existe un problema de multicolinealidad? Explique.
  - Calcule el apalancamiento medio. ¿Alguna de las observaciones tiene un punto de datos de alto apalancamiento?
  - Calcule los residuales estandarizados. Identifique cualquier observación con un residual estandarizado grande. ¿El modelo ajustado predice una respuesta muy alta o muy baja de estas observaciones?
19. Con referencia a los datos de la tabla P.18. Localice el “mejor” modelo de regresión utilizando el procedimiento de regresión por pasos y todos los procedimientos de regresión posibles. Compare los resultados. ¿Se siente cómodo al usar un modelo de regresión para pronosticar el resultado del examen final con menos variables independientes de las tres originales?
20. Recuerde el ejemplo 7.12. El apéndice D (vea la página 517) contiene la serie completa de datos relacionada con la compensación del CEO. Utilice la regresión por pasos para seleccionar el “mejor” modelo con  $k = 3$  variables explicativas. Ajuste el modelo por pa-

sos e interprete los coeficientes estimados. Examine los residuales. Identifique y explique cualquier observación que tenga alguna influencia. Si tuviera que escoger entre este modelo y el modelo explicativo  $k = 2$  del ejemplo 7.12, ¿Cuál escogería? ¿Por qué?

## CASOS

### CASO 7-1 EL MERCADO DE BONOS<sup>16</sup>

Judy Johnson, vicepresidenta de finanzas de una gran empresa privada de servicios públicos de inversionistas del noroeste, se enfrentaba a un problema de financiamiento. La compañía necesitaba dinero para pagar sus deudas de corto plazo que estaban por vencer y para continuar la construcción de una planta de carbón.

La mayor preocupación de Judy era estimar el mercado de bonos a 10 o 30 años; la compañía tenía que decidir si debía optar por el financiamiento emitiendo acciones o deuda de largo plazo. Para tomar esta decisión, la empresa necesitaba un pronóstico confiable de la tasa de interés que pagaría al momento de la emisión de los bonos.

Judy convocó a su personal financiero a una junta para discutir el problema del mercado de bonos. Un miembro de su personal, Ron Peterson, recientemente graduado como MBA, sugirió que se podría desarrollar un modelo de regresión múltiple para pronosticar las tasas de interés de los bonos. La vicepresidenta no estaba familiarizada con la regresión múltiple, así que dirigió la conversación hacia otro lado. Después de una hora de interacción improductiva, Judy le pidió a Ron que le presentara un reporte el lunes siguiente.

Ron sabía que la clave para el desarrollo de un buen modelo de pronóstico es identificar las variables independientes que se relacionan con la tasa de interés pagada por las empresas al momento de emitir los bonos. Después de discutir el problema con varias personas de la empresa, Ron decidió investigar las siguientes variables: un índice (Moody's) para la calidad de los bonos de una empresa de servicios públicos, la tasa de ganancias de renta fija de una empresa de servicios públicos, la tasa de interés de los bonos del tesoro de Estados Unidos (Treasury Bond), el vencimiento de los bonos (de 10 a 30 años) y la prima de interés preferencial al momento de la emisión.

Ron reunió los datos que se podían correlacionar con las tasas de interés de los bonos para la emisión de bonos de la empresa durante los dos años anteriores. Primero no estaba seguro de cómo manejar las tasas de los bonos. Finalmente decidió considerar sólo empresas con la misma tasa, o ligeramente más alta, que su compañía. Esta decisión le proporcionó una muestra

de 93 emisiones para analizar. Pero él aún estaba preocupado sobre la validez de utilizar la calidad de los bonos como datos de tipo intervalo. Ron llamó a su antiguo profesor de estadística y entendió que las variables ficticias podrían resolver el problema. Por esto, codificó las tasas de los bonos de la siguiente manera:

$$\begin{aligned} X_1 &= 1 \text{ si la calificación del bono es A; de lo contrario, } 0 \\ X_2 &= 1 \text{ si la calificación del bono es AA; de lo contrario, } 0 \end{aligned}$$

Si la tasa de los bonos de la empresa es BAA,  $X_1$  y  $X_2$  son 0. El siguiente paso era seleccionar un programa de computadora de regresión múltiple e introducir los datos. Las siguientes variables fueron incluidas en el modelo completo de la ecuación:

Variable 1:  $Y$  = Tasa de interés pagada por la empresa en el momento de la emisión de los bonos

Variable 2:  $X_1 = 1$  si la calificación de los bonos es A

Variable 3:  $X_2 = 1$  si la calificación de los bonos es AA

Variable 4:  $X_3$  = Tasa de ganancias de la empresa contra cargos fijos

Variable 5:  $X_4$  = Tasa de los bonos del tesoro (Treasury Bond, de 10 a 30 años) al momento de la emisión de los bonos

Variable 6:  $X_5$  = Vencimiento del bono (10 a 30 años)

Variable 7:  $X_6$  = Prima de interés preferencial al momento de la emisión

Los datos reales se muestran en el Apéndice B.

Ron decide analizar la matriz de correlación de la tabla 7.18. No se sorprendió al encontrar una alta relación positiva entre la tasa de interés pagada por la empresa al momento de la emisión de los bonos y la tasa de los bonos del Treasury Bond,  $r = 0.883$ . Él también esperaba una relación bastante positiva entre la variable dependiente y la prima de interés preferencial ( $r = 0.596$ ). Tampoco se sorprendió demasiado al descubrir que estas dos variables explicativas también se relacionaban entre sí (multicolinealidad potencial,  $r = 0.713$ ).

<sup>16</sup> Los datos para este caso fueron proporcionados por Dorothy Mercer, estudiante de MBA de la Eastern Washington University. El análisis fue realizado por los estudiantes de MBA Tak Fu, Ron Hand, Dorothy Mercer, Mary Lou Redmond y Harold Wilson.

**TABLA 7.18 Matriz de correlación para el estudio del mercado de bonos**

| Correlaciones (Pearson)            |                 |        |       |                                    |                   |             |
|------------------------------------|-----------------|--------|-------|------------------------------------|-------------------|-------------|
|                                    | Tasa de interés | A      | AA    | Tasa de ganancias<br>en renta fija | Tasa de los bonos | Vencimiento |
| A                                  | -0.347          |        |       |                                    |                   |             |
| AA                                 | -0.173          | -0.399 |       |                                    |                   |             |
| Tasa de ganancias<br>en renta fija | 0.097           | 0.037  | 0.577 |                                    |                   |             |
| Tasa de los bonos                  | 0.883           | -0.256 | 0.291 | 0.253                              |                   |             |
| Vencimiento                        | -0.221          | 0.278  | 0.010 | 0.094                              | -0.477            |             |
| Prima                              | 0.596           | -0.152 | 0.342 | 0.255                              | 0.713             | -0.314      |

El resultado de la relación negativa entre la variable dependiente y la duración del vencimiento de los bonos (de 10 a 30 años),  $r = -0.221$  tenía sentido para Ron.

Ahora, Ron corre un modelo completo que contiene todas las variables explicativas. El examen de los valores calculados  $t$  y los valores  $p$ , que se presentan en la tabla 7.19, indicó que quizás la variable de la tasa de ganancias por renta fija y seguramente la prima de interés preferencial no contribuyan al pronóstico de la tasa de interés pagada por la empresa al momento de la emisión de los bonos.

Ron concluyó que la tasa de ganancias de la empresa por cargos fijos no se relacionaba con la variable dependiente ( $r = 0.097$ ) y que la prima de interés preferencial era colineal con las tasas del Treasury Bond ( $r = 0.713$ ). Ron decidió entonces eliminar la tasa y las variables de la prima. Los resultados de la computadora para este modelo se muestran en la tabla 7.20.

Los valores  $p$  fueron significativos para cada una de las variables independientes.

Ron incluyó los siguientes comentarios en su reporte para Judy:

1. El mejor modelo, tasa de interés =  $-1.28 - 0.929A - 1.18AA + 1.23$  (tasa de los bonos) + 0.0615 (vencimiento), explicó 90.6% de la variación en la tasa de interés.
2. El error estándar de la estimación es 0.53. Por lo tanto, 95% de los valores reales de la variable dependiente caerán dentro de  $2(0.53) = 1.06$  de un pronóstico determinado.
3. Los coeficientes de regresión son significativos y parecen ser confiables.

Ron estaba muy complacido con su esfuerzo y pensó que Judy también lo estaría. ■

## PREGUNTA

1. ¿Qué preguntas cree usted que Judy hará a Ron?

**TABLA 7.19 Corrida del modelo completo para el estudio del mercado de bonos**

### Regression Analysis

The regression equation is

$$\text{Interest Rates} = -1.03 - 0.829 A - 0.889 AA - 0.242 \text{ Ratio} + 1.26 \text{ Bond R} \\ + 0.0628 \text{ Maturity} - 0.0031 \text{ Prime}$$

| Predictor | Coef     | SE Coef  | T     | P     |
|-----------|----------|----------|-------|-------|
| Constant  | -1.0263  | 0.6572   | -1.56 | 0.122 |
| A         | -0.8285  | 0.1342   | -6.17 | 0.000 |
| AA        | -0.8894  | 0.2248   | -3.96 | 0.000 |
| Ratio     | -0.2417  | 0.1135   | -2.13 | 0.036 |
| Rates     | 1.25753  | 0.05964  | 21.08 | 0.000 |
| Maturity  | 0.062839 | 0.006589 | 9.54  | 0.000 |
| Prime     | -0.00313 | 0.02720  | -0.12 | 0.909 |

**TABLA 7.19 (Continúa)**

S = 0.5235      R-Sq = 91.1%      R-Sq(adj) = 90.5%

## Analysis of Variance

| Source         | DF | SS      | MS     | F      | P     |
|----------------|----|---------|--------|--------|-------|
| Regression     | 6  | 240.907 | 40.151 | 146.52 | 0.000 |
| Residual Error | 86 | 23.567  | 0.274  |        |       |
| Total          | 92 | 264.474 |        |        |       |

## Unusual Observations

| Obs | A    | Rate    | Fit     | St Dev Fit | Residual | St Resid |
|-----|------|---------|---------|------------|----------|----------|
| 64  | 1.00 | 15.5000 | 13.7645 | 0.1627     | 1.7355   | 3.49R    |
| 66  | 1.00 | 11.3000 | 13.4780 | 0.1357     | -2.1780  | -4.31R   |
| 67  | 1.00 | 12.3800 | 13.7897 | 0.1297     | -1.4097  | -2.78R   |
| 68  | 1.00 | 12.1500 | 13.7097 | 0.0947     | -1.5597  | -3.03R   |

R denotes an observation with a large standardized residual.

---

**TABLA 7.20 Modelo final para estudio del mercado de bonos**

## Regression Analysis

The regression equation is

Interest Rates = -1.28 - 0.929 A - 1.18 AA + 1.23 Bond Rates +  
0.0615 Maturity

| Predictor | Coef     | SE Coef  | T     | P     |
|-----------|----------|----------|-------|-------|
| Constant  | -1.2765  | 0.6554   | -1.95 | 0.055 |
| A         | -0.9293  | 0.1264   | -7.35 | 0.000 |
| AA        | -1.1751  | 0.1781   | -6.60 | 0.000 |
| Rates     | 1.23308  | 0.05964  | 26.81 | 0.000 |
| Maturity  | 0.061474 | 0.006649 | 9.25  | 0.000 |

S = 0.5310      R-Sq = 90.6%      R-Sq(adj) = 90.2%

## Analysis of Variance

| Source         | DF | SS      | MS     | F      | P     |
|----------------|----|---------|--------|--------|-------|
| Regression     | 4  | 239.665 | 59.916 | 212.53 | 0.000 |
| Residual Error | 88 | 24.809  | 0.282  |        |       |
| Total          | 92 | 264.474 |        |        |       |

## Unusual Observations

| Obs | A    | Rate    | Fit     | St Dev Fit | Residual | St Resid |
|-----|------|---------|---------|------------|----------|----------|
| 64  | 1.00 | 15.5000 | 13.6956 | 0.0894     | 1.8044   | 3.45R    |
| 66  | 1.00 | 11.3000 | 13.6462 | 0.0898     | -2.3462  | -4.48R   |
| 67  | 1.00 | 12.3800 | 13.6462 | 0.0898     | -1.2662  | -2.42R   |
| 68  | 1.00 | 12.1500 | 13.6462 | 0.0898     | -1.4962  | -2.86R   |

R denotes an observation with a large standardized residual.

---

## CASO 7-2 AAA WASHINGTON

En el caso 5.5 se proporciona una visión general de AAA Washington, donde se les pidió a los estudiantes preparar una descomposición de series de tiempo de las llamadas al servicio de emergencia en carretera recibidas por el club en cinco años. El análisis realizado en el caso 5.5 mostró que el patrón en el volumen de llamadas al servicio de emergencia era probablemente de naturaleza cíclica. En el caso 6.6 se investigaron cuatro variables: tasa de desempleo, promedio de temperatura diaria, cantidad de lluvia y el número de miembros del club. Se determinó que los promedios de temperatura diaria y tasa de desempleo eran variables significativas. La cantidad de lluvia y el número de socios del club no eran variables significativas.

Una conversación con el gerente del centro de llamadas de emergencia condujo a dos importantes

observaciones: (1) los fabricantes de automóviles parecen diseñar vehículos que operan mejor a 65 grados Fahrenheit, y (2) el volumen de llamadas parece incrementarse cuando la temperatura promedio baja algunos grados del promedio de 30, con respecto a una baja similar con temperaturas de 60. Esta información sugirió que el efecto de la temperatura en los servicios de emergencia en carretera no era lineal.

El señor DeCoria ha observado que la tendencia cíclica de las series de tiempo parece retrasarse del ciclo económico general. Él ha sugerido que la tasa de desempleo para el estado de Washington sería una buena medida sustituta para el estado general de la economía de Washington. En la tabla 7.21 se presentan los datos de la temperatura promedio mensual y la tasa de desempleo del estado de Washington.

**TABLA 7.21 Datos para AAA Washington para el caso 7.2**

| Año  | Mes        | Llamadas | Tasa   | Temp. |
|------|------------|----------|--------|-------|
| 1987 | Junio      |          | 6.9940 |       |
|      | Julio      |          | 6.8137 |       |
|      | Agosto     |          | 6.3648 |       |
|      | Septiembre |          | 6.5435 |       |
|      | Octubre    |          | 6.7826 |       |
|      | Noviembre  |          | 6.9225 |       |
|      | Diciembre  |          | 7.1560 |       |
|      | Enero      | 20002    | 7.9143 | 55.1  |
|      | Febrero    | 21591    | 7.7576 | 59.0  |
|      | Marzo      | 22696    | 5.7867 | 63.8  |
|      | Abril      | 21509    | 5.7592 | 63.8  |
|      | Mayo       | 21223    | 5.5718 | 59.1  |
| 1988 | Junio      | 21449    | 5.5049 | 54.6  |
|      | Julio      | 23475    | 5.8630 | 45.4  |
|      | Agosto     | 23529    | 5.4709 | 41.0  |
|      | Septiembre | 23327    | 7.5474 | 40.3  |
|      | Octubre    | 24050    | 7.8157 | 34.3  |
|      | Noviembre  | 24010    | 7.1390 | 43.2  |
|      | Diciembre  | 19735    | 6.2637 | 52.5  |
|      | Enero      | 20153    | 5.8332 | 55.3  |
|      | Febrero    | 19512    | 5.8077 | 62.4  |
|      | Marzo      | 19892    | 5.6713 | 62.9  |
|      | Abril      | 20326    | 5.4977 | 63.5  |
|      | Mayo       | 19378    | 5.2989 | 60.9  |
| 1989 | Junio      | 21263    | 5.6028 | 51.9  |
|      | Julio      |          |        |       |
|      | Agosto     |          |        |       |
|      | Septiembre |          |        |       |

TABLA 7.21 (Continúa)

| Año          | Mes        | Llamadas | Tasa   | Temp. |
|--------------|------------|----------|--------|-------|
| 1989 (cont.) | Noviembre  | 21443    | 5.9143 | 46.2  |
|              | Diciembre  | 23366    | 5.0000 | 41.8  |
| 1990         | Enero      | 23836    | 6.1917 | 41.8  |
|              | Febrero    | 23336    | 6.3775 | 38.9  |
|              | Marzo      | 22003    | 5.7234 | 46.3  |
|              | Abril      | 20155    | 4.7792 | 51.7  |
|              | Mayo       | 20070    | 4.5715 | 54.9  |
|              | Junio      | 19588    | 4.3899 | 59.8  |
|              | Julio      | 20804    | 4.2559 | 66.7  |
|              | Agosto     | 19644    | 3.9359 | 66.4  |
|              | Septiembre | 17424    | 3.9048 | 61.9  |
|              | Octubre    | 20833    | 4.4294 | 50.4  |
|              | Noviembre  | 22490    | 5.1523 | 45.8  |
|              | Diciembre  | 24861    | 5.5102 | 33.9  |
| 1991         | Enero      | 23441    | 6.8901 | 37.9  |
|              | Febrero    | 19205    | 7.0308 | 46.9  |
|              | Marzo      | 20386    | 6.7186 | 43.4  |
|              | Abril      | 19988    | 6.1280 | 49.1  |
|              | Mayo       | 19077    | 5.8146 | 54.3  |
|              | Junio      | 19141    | 5.9480 | 58.2  |
|              | Julio      | 20883    | 5.9026 | 65.4  |
|              | Agosto     | 20709    | 5.7227 | 66.0  |
|              | Septiembre | 19647    | 5.6877 | 60.9  |
|              | Octubre    | 22013    | 6.2922 | 51.0  |
|              | Noviembre  | 22375    | 7.0615 | 46.2  |
|              | Diciembre  | 22727    | 7.4370 | 42.4  |
| 1992         | Enero      | 22367    | 8.4513 | 43.0  |
|              | Febrero    | 21155    | 8.7699 | 46.0  |
|              | Marzo      | 21209    | 8.0728 | 48.9  |
|              | Abril      | 19286    | 7.2392 | 52.7  |
|              | Mayo       | 19725    | 7.0461 | 58.3  |
|              | Junio      | 20276    | 7.0478 | 63.6  |
|              | Julio      | 20795    | 7.1080 | 64.9  |
|              | Agosto     | 21126    | 6.7824 | 65.0  |
|              | Septiembre | 20251    | 6.7691 | 58.4  |
|              | Octubre    | 22069    | 7.5896 | 53.2  |
|              | Noviembre  | 23268    | 7.9908 | 44.8  |
|              | Diciembre  | 26039    | 8.2460 | 37.8  |
| 1993         | Enero      | 26127    | 9.5301 | 34.9  |

**Tarea**

- Desarrolle una ecuación de regresión múltiple usando la tasa de desempleo y la temperatura promedio mensual para pronosticar las llamadas al servicio de emergencia en carretera.
- Cree una nueva variable de temperatura y relacionela con el servicio de emergencia. Recuerde que la temperatura es una escala relativa y que la selección del punto cero es arbitraria. Si los vehículos están diseñados para operar mejor en temperaturas de 65 grados Fahrenheit, entonces cada grado arriba o debajo de los 65 grados debe causar que los vehículos operen con menor confiabilidad. Para realizar la transformación de los datos de temperatura que estimule este efecto, empiece por restar 65 de los valores promedio de temperatura mensual. Esto coloca a “cero” en 65 grados Fahrenheit. ¿Se deben utilizar los valores absolutos de esta nueva variable de temperatura?
- Transforme la nueva variable de temperatura promedio de manera que se relacione linealmente con las llamadas de servicio.

4. Elabore una nueva variable de tasa de desempleo y relaciónela con el servicio de emergencia en carretera. Dé al desempleo un efecto retrasado en el servicio de emergencia mediante la tasa de desempleo para el mes: (1) tres meses previos al presente mes y (2) 11 meses previos al presente mes como los datos para la variable independiente de desempleo. ¿Cuál es el mejor modelo para el pronóstico? ¿Los signos en los coeficientes para las variables

independientes son los que esperaba? Los coeficientes de las variables independientes son significativamente diferentes de cero?

5. Desarrolle una ecuación de regresión múltiple mediante la variable de temperatura promedio transformada que se creó en el punto 3 y la variable retrasada creada en el punto 4. ¿Este es un buen modelo? ¿Se ha incumplido alguna de las suposiciones fundamentales? ■

## CASO 7-3

## BEISBOL DE FANTASÍA (A)

Era febrero y John Hanke, profesor de estadística retirado, se preparaba para otra temporada de beisbol de fantasía. En años anteriores sus compañeros jugadores siempre lo molestaron porque utilizaba su conocimiento de estadística para obtener ventaja. Desgraciadamente, esto nunca ha ocurrido. La enseñanza, investigación, publicación y su trabajo en comités lo han mantenido demasiado ocupado. Ahora, habiéndose retirado recientemente, tiene el tiempo para aplicar sus conocimientos de estadística durante el reclutamiento anual de jugadores. En este tipo de liga de fantasía, cada manager tiene 260 dólares con que proponer la compra de 23 jugadores (14 bateadores y 9 lanzadores). Entonces, cada equipo será clasificado (con base en las estadísticas actuales de los jugadores en la temporada) en ocho categorías estadísticas. El doctor Hanke estaba muy preocupado por elegir a los jugadores que se desempeñaran bien en tres de las cuatro categorías de lanzamiento. En años pasados, su personal de lanzamiento, especialmente los peloteros abridores, han sido el hazmerreír de la liga. La temporada 2000 iba a ser diferente. Él pretendía desarrollar modelos para pro-

nosticar con exactitud el desempeño de lanzamiento de los lanzadores abridores.

Las tres categorías que Hanke deseaba investigar eran las victorias (WINS), promedio de carreras admitidas (ERA), las bases por bolas y el porcentaje otorgado por entradas lanzadas (Br/IP). Ha pasado un tiempo considerable descargando estadísticas de beisbol para lanzadores abridores de la temporada 1999.<sup>17</sup> El doctor Hanke pretende desarrollar un modelo de regresión múltiple para pronosticar cada una de las categorías que le interesan. A menudo ha recomendado a sus estudiantes que la selección inicial de variables era el aspecto más importante para desarrollar un modelo de regresión. Él sabía que si no tenía buenas variables explicativas, no podría obtener ecuaciones de pronóstico útiles. Después de bastante trabajo, el doctor Hanke escogió las siguientes seis variables. También decidió incluir sólo peloteros abridores que habían lanzado por lo menos 100 entradas durante la temporada. Los datos para los 138 abridores seleccionados se presentan en la tabla 7.22.

**TABLA 7.22** Estadísticas de lanzamiento para los 139 peloteros abridores del caso 7.3

| Lanzador       | Tm  | ERA  | Br/IP | Ctl | K/9 | HR/9 | OBA |
|----------------|-----|------|-------|-----|-----|------|-----|
| 1 Álvarez, W.  | TAM | 4.22 | 1.49  | 1.6 | 7.2 | 1.2  | 254 |
| 2 Anderson, B. | ARI | 4.57 | 1.32  | 2.7 | 5.2 | 1.2  | 275 |
| 3 Appier, K.   | OAK | 5.17 | 1.50  | 1.6 | 5.6 | 1.2  | 274 |
| 4 Arrojo, R.   | TAM | 5.20 | 1.58  | 1.8 | 6.9 | 1.5  | 284 |
| 5 Ashby, A.    | SD  | 3.80 | 1.25  | 2.4 | 5.8 | 1.1  | 253 |
| 6 Astacio, P.  | COL | 5.04 | 1.44  | 2.8 | 8.1 | 1.5  | 276 |
| 7 Baldwin, J.  | CHW | 5.11 | 1.51  | 1.5 | 5.6 | 1.5  | 274 |
| 8 Batista, M.  | MON | 4.90 | 1.52  | 1.6 | 6.4 | 0.7  | 271 |
| 9 Belcher, T.  | ANA | 6.74 | 1.62  | 1.1 | 3.5 | 1.8  | 303 |
| 10 Benes, A.   | ARI | 4.82 | 1.50  | 1.7 | 6.4 | 1.5  | 272 |

<sup>17</sup> Las estadísticas de lanzamiento de beisbol fueron descargadas del sitio de Internet Ron Shandler's BaseballHQ en [www.BaseballHQ.com](http://www.BaseballHQ.com).

TABLA 7.22 (Continúa)

| <i>Lanzador</i> |                 | <i>Tm</i> | <i>ERA</i> | <i>Br/IP</i> | <i>Ctl</i> | <i>K/9</i> | <i>HR/9</i> | <i>OBA</i> |
|-----------------|-----------------|-----------|------------|--------------|------------|------------|-------------|------------|
| 11              | Benson, K.      | PIT       | 4.08       | 1.36         | 1.7        | 6.4        | 0.7         | 243        |
| 12              | Bergman, S.     | ATL       | 5.22       | 1.56         | 1.5        | 3.8        | 0.8         | 306        |
| 13              | Blair, W.       | DET       | 6.85       | 1.59         | 1.9        | 5.5        | 1.9         | 302        |
| 14              | Bohanon, B.     | COL       | 6.21       | 1.66         | 1.3        | 5.5        | 1.4         | 291        |
| 15              | Bottenfield, K. | STL       | 3.98       | 1.50         | 1.4        | 5.9        | 1.0         | 262        |
| 16              | Brock, C.       | SF        | 5.51       | 1.55         | 1.9        | 6.4        | 1.5         | 286        |
| 17              | Brown, K.       | LA        | 3.00       | 1.07         | 3.7        | 7.9        | 0.7         | 222        |
| 18              | Burba, D.       | CLE       | 4.25       | 1.40         | 1.8        | 7.1        | 1.2         | 247        |
| 19              | Burkett, J.     | TEX       | 5.63       | 1.56         | 2.1        | 5.9        | 1.1         | 300        |
| 20              | Byrd, P.        | PHI       | 4.61       | 1.38         | 1.5        | 4.8        | 1.5         | 261        |
| 21              | Carpenter, C.   | TOR       | 4.38       | 1.50         | 2.2        | 6.4        | 1.0         | 288        |
| 22              | Clemens, R.     | NYY       | 4.62       | 1.47         | 1.8        | 7.8        | 1.0         | 253        |
| 23              | Clement, M.     | SD        | 4.50       | 1.53         | 1.6        | 6.7        | 0.9         | 265        |
| 24              | Colón, B.       | CLE       | 3.95       | 1.27         | 2.1        | 7.1        | 1.1         | 236        |
| 25              | Cone, D.        | NYY       | 3.45       | 1.32         | 2.0        | 8.2        | 1.0         | 225        |
| 26              | Córdova, F.     | PIT       | 4.44       | 1.40         | 1.7        | 5.5        | 0.9         | 262        |
| 27              | Daal, O.        | ARI       | 3.66       | 1.25         | 1.9        | 6.2        | 0.9         | 231        |
| 28              | Dempster, R.    | FLA       | 4.71       | 1.63         | 1.4        | 7.7        | 1.3         | 254        |
| 29              | Dreifort, D.    | LA        | 4.80       | 1.42         | 1.8        | 7.1        | 1.0         | 254        |
| 30              | Elarton, S.     | HOU       | 3.48       | 1.24         | 2.8        | 8.8        | 0.6         | 235        |
| 31              | Erickson, S.    | BAL       | 4.81       | 1.49         | 1.1        | 4.1        | 1.1         | 266        |
| 32              | Escobar, K.     | TOR       | 5.69       | 1.63         | 1.6        | 6.7        | 1.0         | 285        |
| 33              | Estes, S.       | SF        | 4.92       | 1.58         | 1.4        | 7.0        | 0.9         | 261        |
| 34              | Farnsworth, K.  | CHC       | 5.05       | 1.48         | 1.3        | 4.8        | 1.9         | 269        |
| 35              | Fassero, J.     | TEX       | 7.21       | 1.86         | 1.4        | 6.6        | 2.0         | 313        |
| 36              | Fernández, A.   | FLA       | 3.38       | 1.25         | 2.2        | 5.8        | 0.6         | 247        |
| 37              | Finley, C.      | ANA       | 4.43       | 1.37         | 2.1        | 8.4        | 1.0         | 240        |
| 38              | García, F.      | SEA       | 4.07       | 1.47         | 1.9        | 7.6        | 0.8         | 259        |
| 39              | Gardner, M.     | SF        | 6.47       | 1.43         | 1.5        | 5.6        | 1.7         | 259        |
| 40              | Glavine, T.     | ATL       | 4.12       | 1.46         | 1.7        | 5.3        | 0.7         | 275        |
| 41              | Gooden, D.      | CLE       | 6.26       | 1.69         | 1.3        | 6.9        | 1.4         | 274        |
| 42              | Guzmán, J.      | CIN       | 3.74       | 1.40         | 1.8        | 7.0        | 1.3         | 249        |
| 43              | Halama, J.      | SEA       | 4.22       | 1.39         | 1.9        | 5.3        | 1.0         | 270        |
| 44              | Halladay, R.    | TOR       | 3.92       | 1.58         | 1.0        | 4.9        | 1.1         | 264        |
| 45              | Hampton, M.     | HOU       | 2.90       | 1.28         | 1.8        | 6.7        | 0.5         | 228        |
| 46              | Harnisch, P.    | CIN       | 3.68       | 1.25         | 2.1        | 5.5        | 1.1         | 247        |
| 47              | Hawkins, L.     | MIN       | 6.67       | 1.71         | 1.7        | 5.3        | 1.5         | 319        |
| 48              | Haynes, J.      | OAK       | 6.34       | 1.68         | 1.2        | 5.9        | 1.3         | 276        |
| 49              | Helling, R.     | TEX       | 4.85       | 1.43         | 1.5        | 5.4        | 1.7         | 263        |
| 50              | Hentgen, P.     | TOR       | 4.79       | 1.46         | 1.8        | 5.3        | 1.4         | 279        |
| 51              | Heredia, G.     | OAK       | 4.81       | 1.31         | 3.4        | 5.3        | 1.0         | 281        |
| 52              | Hermanson, D.   | MON       | 4.21       | 1.36         | 2.1        | 6.0        | 0.8         | 263        |
| 53              | Hernández, O.   | NYY       | 4.12       | 1.28         | 1.8        | 6.6        | 1.0         | 230        |
| 54              | Hernández, L.   | SF        | 4.65       | 1.52         | 1.9        | 6.5        | 1.0         | 281        |
| 55              | Hershiser, O.   | NYM       | 4.58       | 1.41         | 1.2        | 4.5        | 0.7         | 251        |
| 56              | Hill, K.        | ANA       | 4.78       | 1.60         | 1.0        | 5.3        | 1.0         | 256        |
| 57              | Hitchcock, S.   | SD        | 4.12       | 1.35         | 2.6        | 8.5        | 1.3         | 252        |
| 58              | Holt, C.        | HOU       | 4.66       | 1.52         | 2.0        | 6.3        | 0.7         | 287        |
| 59              | Hudson, T.      | OAK       | 3.24       | 1.34         | 2.1        | 8.7        | 0.5         | 233        |
| 60              | Irabu, H.       | NYY       | 4.84       | 1.34         | 2.9        | 7.1        | 1.4         | 267        |
| 61              | Jiménez, J.     | STL       | 5.85       | 1.50         | 1.6        | 6.2        | 0.9         | 267        |
| 62              | Johnson, R.     | ARI       | 2.49       | 1.02         | 5.2        | 12.1       | 1.0         | 207        |

TABLA 7.22 (Continúa)

|     | <i>Lanzador</i> | <i>Tm</i> | <i>ERA</i> | <i>Br/IP</i> | <i>Ctl</i> | <i>K/9</i> | <i>HR/9</i> | <i>OBA</i> |
|-----|-----------------|-----------|------------|--------------|------------|------------|-------------|------------|
| 63  | Johnson, J.     | BAL       | 5.47       | 1.52         | 1.3        | 5.6        | 1.3         | 263        |
| 64  | Karl, S.        | MIL       | 4.79       | 1.60         | 1.1        | 3.4        | 1.0         | 299        |
| 65  | Kile, D.        | COL       | 6.62       | 1.76         | 1.1        | 5.5        | 1.6         | 288        |
| 66  | Leiter, A.      | NYM       | 4.23       | 1.42         | 1.7        | 6.8        | 0.8         | 252        |
| 67  | Lieber, J.      | CHC       | 4.08       | 1.34         | 4.0        | 8.2        | 1.2         | 276        |
| 68  | Lima, J.        | HOU       | 3.58       | 1.22         | 4.3        | 6.8        | 1.1         | 263        |
| 69  | Loaiza, E.      | TEX       | 4.57       | 1.40         | 1.9        | 5.8        | 0.7         | 267        |
| 70  | Maddux, G.      | ATL       | 3.57       | 1.35         | 3.7        | 5.6        | 0.7         | 287        |
| 71  | Martínez, P.    | BOS       | 2.07       | 0.92         | 8.5        | 13.2       | 0.4         | 205        |
| 72  | Mays, J.        | MIN       | 4.37       | 1.44         | 1.7        | 6.1        | 1.3         | 264        |
| 73  | Meadows, B.     | FLA       | 5.61       | 1.52         | 1.3        | 3.6        | 1.6         | 292        |
| 74  | Mercker, K.     | BOS       | 4.81       | 1.64         | 1.3        | 5.6        | 1.1         | 282        |
| 75  | Millwood, K.    | ATL       | 2.68       | 1.00         | 3.5        | 8.1        | 0.9         | 201        |
| 76  | Milton, E.      | MIN       | 4.50       | 1.23         | 2.6        | 7.1        | 1.2         | 240        |
| 77  | Mlicki, D.      | DET       | 4.61       | 1.46         | 1.7        | 5.4        | 1.1         | 274        |
| 78  | Moehler, B.     | DET       | 5.05       | 1.47         | 1.8        | 4.9        | 1.0         | 286        |
| 79  | Morgan, M.      | TEX       | 6.24       | 1.66         | 1.3        | 3.9        | 1.6         | 310        |
| 80  | Moyer, J.       | SEA       | 3.87       | 1.24         | 2.9        | 5.4        | 0.9         | 261        |
| 81  | Mulholland, T.  | ATL       | 4.39       | 1.45         | 1.8        | 4.4        | 1.1         | 288        |
| 82  | Mussina, M.     | BAL       | 3.50       | 1.28         | 3.3        | 7.6        | 0.7         | 259        |
| 83  | Nagy, C.        | CLE       | 4.95       | 1.47         | 2.1        | 5.6        | 1.2         | 287        |
| 84  | Navarro, J.     | CHW       | 6.11       | 1.74         | 1.0        | 4.2        | 1.6         | 307        |
| 85  | Neagle, D.      | CIN       | 4.29       | 1.21         | 1.9        | 6.2        | 1.9         | 226        |
| 86  | Nomo, H.        | MIL       | 4.55       | 1.43         | 2.1        | 8.2        | 1.4         | 252        |
| 87  | Núñez, V.       | FLA       | 4.08       | 1.38         | 1.6        | 7.2        | 0.9         | 231        |
| 88  | Ogea, C.        | PHI       | 5.63       | 1.51         | 1.3        | 4.1        | 1.9         | 281        |
| 89  | Olivares, O.    | OAK       | 4.17       | 1.45         | 1.0        | 3.7        | 0.8         | 266        |
| 90  | Oliver, D.      | STL       | 4.27       | 1.38         | 1.6        | 5.5        | 0.7         | 256        |
| 91  | Oquist, M.      | OAK       | 5.39       | 1.58         | 1.4        | 5.7        | 1.2         | 278        |
| 92  | Ortiz, R.       | SF        | 3.82       | 1.52         | 1.3        | 7.1        | 1.0         | 238        |
| 93  | Park, C.        | LA        | 5.24       | 1.59         | 1.7        | 8.1        | 1.4         | 268        |
| 94  | Parque, J.      | CHW       | 5.14       | 1.67         | 1.4        | 5.8        | 1.2         | 293        |
| 95  | Parris, S.      | CIN       | 3.51       | 1.37         | 1.7        | 6.0        | 1.1         | 249        |
| 96  | Pavano, C.      | MON       | 5.63       | 1.46         | 2.0        | 6.1        | 0.7         | 278        |
| 97  | Person, R.      | PHI       | 4.68       | 1.51         | 1.6        | 8.5        | 1.5         | 243        |
| 98  | Pettitte, A.    | NYY       | 4.71       | 1.60         | 1.4        | 5.7        | 0.9         | 279        |
| 99  | Ponson, S.      | BAL       | 4.71       | 1.46         | 1.4        | 4.8        | 1.5         | 270        |
| 100 | Portugal, M.    | BOS       | 5.52       | 1.47         | 1.9        | 4.7        | 1.7         | 290        |
| 101 | Radke, B.       | MIN       | 3.75       | 1.30         | 2.8        | 5.0        | 1.2         | 273        |
| 102 | Rapp, P.        | BOS       | 4.13       | 1.48         | 1.3        | 5.5        | 0.8         | 256        |
| 103 | Reed, R.        | NYM       | 4.59       | 1.41         | 2.2        | 6.3        | 1.4         | 272        |
| 104 | Reynolds, S.    | HOU       | 3.85       | 1.24         | 5.3        | 7.7        | 0.9         | 270        |
| 105 | Reynoso, A.     | ARI       | 4.37       | 1.47         | 1.2        | 4.3        | 1.1         | 267        |
| 106 | Ritchie, T.     | PIT       | 3.50       | 1.30         | 2.0        | 5.6        | 0.9         | 252        |
| 107 | Rogers, K.      | NYM       | 4.20       | 1.41         | 1.8        | 5.8        | 0.7         | 266        |
| 108 | Rosado, J.      | KC        | 3.85       | 1.29         | 2.0        | 6.1        | 1.0         | 245        |
| 109 | Rueter, K.      | SF        | 5.42       | 1.49         | 1.7        | 4.6        | 1.4         | 289        |
| 110 | Rupe, R.        | TAM       | 4.56       | 1.36         | 1.7        | 6.1        | 1.1         | 247        |
| 111 | Saberhagen, B.  | BOS       | 2.95       | 1.12         | 7.4        | 6.1        | 0.8         | 260        |
| 112 | Schilling, C.   | PHI       | 3.55       | 1.13         | 3.5        | 7.6        | 1.2         | 232        |
| 113 | Schmidt, J.     | PIT       | 4.20       | 1.43         | 1.7        | 6.3        | 1.0         | 261        |

TABLA 7.22 (Continúa)

| <i>Lanzador</i> |                 | <i>Tm</i> | <i>ERA</i> | <i>Br/IP</i> | <i>Ctl</i> | <i>K/9</i> | <i>HR/9</i> | <i>OBA</i> |
|-----------------|-----------------|-----------|------------|--------------|------------|------------|-------------|------------|
| 114             | Schourek, P.    | PIT       | 5.34       | 1.57         | 1.9        | 7.5        | 1.6         | 280        |
| 115             | Sele, A.        | TEX       | 4.79       | 1.53         | 2.7        | 8.2        | 0.9         | 290        |
| 116             | Sirotka, M.     | CHW       | 4.00       | 1.40         | 2.2        | 5.4        | 1.0         | 279        |
| 117             | Smoltz, J.      | ATL       | 3.19       | 1.12         | 3.9        | 7.5        | 0.7         | 236        |
| 118             | Snyder, J.      | CHW       | 6.69       | 1.67         | 1.4        | 4.7        | 1.9         | 307        |
| 119             | Sparks, S.      | ANA       | 5.44       | 1.68         | 0.9        | 4.5        | 1.3         | 277        |
| 120             | Stottlemyre, T. | ARI       | 4.09       | 1.44         | 1.9        | 6.6        | 1.1         | 264        |
| 121             | Suppan, J.      | KC        | 4.54       | 1.36         | 1.7        | 4.5        | 1.2         | 267        |
| 122             | Tapani, K.      | CHC       | 4.83       | 1.35         | 2.2        | 4.8        | 0.8         | 275        |
| 123             | Thompson, J.    | DET       | 5.13       | 1.48         | 1.4        | 5.3        | 1.5         | 268        |
| 124             | Thurman, M.     | MON       | 4.06       | 1.31         | 1.6        | 5.2        | 1.0         | 247        |
| 125             | Tomko, B.       | CIN       | 4.92       | 1.37         | 2.2        | 6.9        | 1.6         | 258        |
| 126             | Trachsel, S.    | CHC       | 5.57       | 1.41         | 2.3        | 6.5        | 1.4         | 274        |
| 127             | Valdés, I.      | LA        | 3.99       | 1.33         | 2.5        | 6.3        | 1.4         | 264        |
| 128             | Vázquez, J.     | MON       | 5.02       | 1.34         | 2.2        | 6.6        | 1.2         | 255        |
| 129             | Villone, R.     | CIN       | 4.24       | 1.32         | 1.3        | 6.1        | 0.5         | 215        |
| 130             | Weaver, J.      | DET       | 5.57       | 1.42         | 2.0        | 6.3        | 1.5         | 270        |
| 131             | Wells, D.       | TOR       | 4.83       | 1.33         | 2.7        | 6.6        | 1.2         | 267        |
| 132             | Williams, W.    | SD        | 4.41       | 1.37         | 1.9        | 5.9        | 1.4         | 260        |
| 133             | Witasick, J.    | KC        | 5.58       | 1.73         | 1.2        | 5.8        | 1.3         | 293        |
| 134             | Witt, B.        | TAM       | 5.85       | 1.72         | 1.3        | 6.1        | 1.1         | 288        |
| 135             | Wolf, R.        | PHI       | 5.57       | 1.59         | 1.7        | 8.6        | 1.5         | 263        |
| 136             | Woodard, S.     | MIL       | 4.52       | 1.38         | 3.3        | 5.8        | 1.1         | 288        |
| 137             | Wright, J.      | CLE       | 6.08       | 1.66         | 1.2        | 6.1        | 1.2         | 270        |
| 138             | Yoshii, M.      | NYM       | 4.40       | 1.30         | 1.8        | 5.4        | 1.3         | 248        |

TABLA 7.23 Correlaciones: ERA, Br/IP, Ctl, K/9, HR/9, OBA

|       | ERA             | Br/IP           | Ctl             | K/9             | HR/9           |                                               |
|-------|-----------------|-----------------|-----------------|-----------------|----------------|-----------------------------------------------|
| Br/IP | 0.819<br>0.000  |                 |                 |                 |                |                                               |
| Ctl   | -0.798<br>0.000 | -0.701<br>0.000 |                 |                 |                | Cell Contents: Pearson correlation<br>P-Value |
| K/9   | -0.429<br>0.000 | -0.416<br>0.000 | 0.573<br>0.000  |                 |                |                                               |
| HR/9  | 0.649<br>0.000  | 0.438<br>0.000  | -0.305<br>0.000 | -0.245<br>0.004 |                |                                               |
| OBA   | 0.745<br>0.000  | 0.747<br>0.000  | -0.329<br>0.000 | -0.576<br>0.000 | 0.432<br>0.000 |                                               |

ERA: Promedio de carreras admitidas o el número de carreras permitidas por juego (lanzando nueve entradas).

Br/IP: Corredores en base por entrada o el número de bases por bolas más los imparables permitidos por entrada lanzada.

Ctl: Cociente de control (ponches/bases por bolas).

K/9: Cantidad de bateadores ponchados por juego (lanzando nueve entradas).

HR/9: Cuadrangulares del equipo contrario por partido (lanzando nueve entradas).

OBA: Promedio de bateo del equipo contrario.

**TABLA 7.24 Resultado de la regresión en Minitab usando todas las variables explicativas: beisbol de fantasía (A) caso 7.3**

Regression Analysis: ERA versus Br/IP, Ctl, K/9, HR/9, OBA

The regression equation is

$$\text{ERA} = -3.06 + 2.56 \text{ Br/IP} - 0.0610 \text{ Ctl} + 0.0184 \text{ K/9} + 0.899 \text{ HR/9} + 0.0114 \text{ OBA}$$

| Predictor | Coef     | SE Coef  | T     | P     | VIF |
|-----------|----------|----------|-------|-------|-----|
| Constant  | -3.0583  | 0.6704   | -4.56 | 0.000 |     |
| Br/IP     | 2.5578   | 0.6555   | 3.90  | 0.000 | 8.2 |
| Ctl       | -0.06105 | 0.08060  | -0.76 | 0.450 | 5.5 |
| K/9       | 0.01837  | 0.04709  | 0.39  | 0.697 | 3.4 |
| HR/9      | 0.8990   | 0.1225   | 7.34  | 0.000 | 1.3 |
| OBA       | 0.011355 | 0.004160 | 2.73  | 0.007 | 6.4 |

$$S = 0.4276 \quad R-\text{Sq} = 79.8\% \quad R-\text{Sq}(\text{adj}) = 79.0\%$$

Analysis of Variance

| Source         | DF  | SS      | MS     | F      | P     |
|----------------|-----|---------|--------|--------|-------|
| Regression     | 5   | 95.186  | 19.037 | 104.13 | 0.000 |
| Residual Error | 132 | 24.132  | 0.183  |        |       |
| Total          | 137 | 119.317 |        |        |       |

**TABLA 7.25 Resultado final de la regresión en Minitab para pronosticar el ERA del beisbol de fantasía (A) caso 7.3**

Regression Analysis: ERA versus Ctl, K/9, HR/9, OBA

The regression equation is

$$\text{ERA} = -3.26 - 0.327 \text{ Ctl} + 0.137 \text{ K/9} + 0.902 \text{ HR/9} + 0.0252 \text{ OBA}$$

| Predictor | Coef     | SE Coef  | T     | P     | VIF |
|-----------|----------|----------|-------|-------|-----|
| Constant  | -3.2624  | 0.7032   | -4.64 | 0.000 |     |
| Ctl       | -0.32655 | 0.04546  | -7.18 | 0.000 | 1.6 |
| K/9       | 0.13651  | 0.03795  | 3.60  | 0.000 | 2.0 |
| HR/9      | 0.9023   | 0.1289   | 7.00  | 0.000 | 1.3 |
| OBA       | 0.025224 | 0.002275 | 11.09 | 0.000 | 1.7 |

$$S = 0.4499 \quad R-\text{Sq} = 77.4\% \quad R-\text{Sq}(\text{adj}) = 76.8\%$$

Analysis of Variance

| Source         | DF  | SS      | MS     | F      | P     |
|----------------|-----|---------|--------|--------|-------|
| Regression     | 4   | 92.402  | 23.100 | 104.13 | 0.000 |
| Residual Error | 133 | 26.915  | 0.202  |        |       |
| Total          | 137 | 119.317 |        |        |       |

El siguiente paso del análisis es la creación de la matriz de correlación mostrada en la tabla 7.23. El doctor Hanke encontró extremadamente interesante la relación entre ERA y Br/IP, 0.819. Si él pudiera adquirir lanzadores que se desempeñaran bien en una de estas medidas, también deberían desempeñarse bien en la otra. La tabla 7.24 muestra el resultado de la regresión cuando todas las variables independientes, incluyendo Br/IP, se utilizaron para predecir ERA. El VIF = 8.2

para la variable Br/IP es grande, lo cual indica que esta variable se relaciona linealmente con las variables explicativas restantes y existe un problema de multicolinealidad. El examen de la tabla 7.25 muestra el resultado cuando la variable Br/IP se excluye del modelo. La  $R^2$  es 77.4%, y la ecuación se ve bien. La estadística  $t$  para cada una de las variables explicativas es grande con un valor  $p$  muy pequeño.

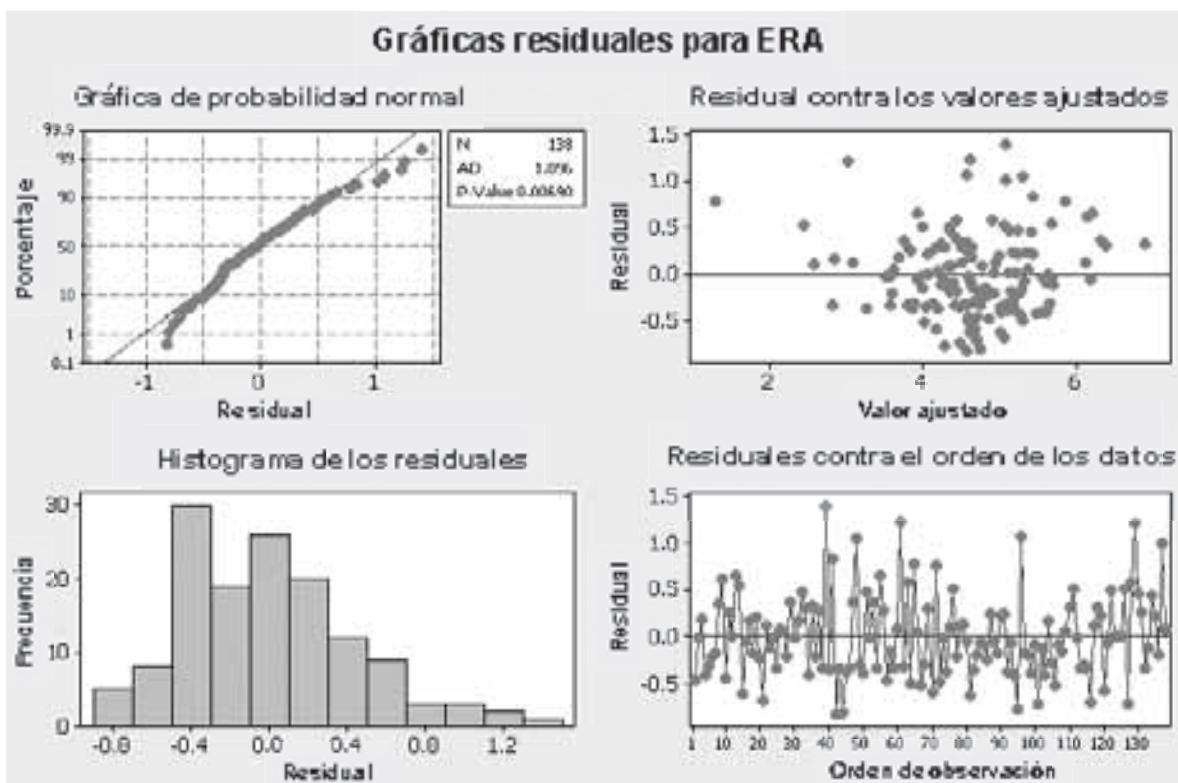


FIGURA 7.4 Gráficas de residuales para el pronóstico de ERA: Beisbol de fantasía (A) caso 7.3

Los *VIF* son relativamente pequeños para las cuatro variables explicativas, lo cual indica que la multicolinealidad ya no es un problema.

El doctor Hanke decide que tiene un buen modelo y desarrolla las gráficas de residuales mostradas en la figura 7.4.■

#### PREGUNTAS

1. Comente sobre el modelo que el doctor Hanke desarrolló para pronosticar el promedio de carreras admitidas (ERA). Examine las gráficas de residuales mostradas en la figura 7.4 y determine si el modelo es válido.
2. ¿Existe alguna relación lineal entre las variables explicativas y el promedio de carreras admitidas? Si es así, desarrolle un nuevo modelo incluyendo la transformación apropiada de la variable.
3. Desarrolle un modelo para pronosticar Br/IP.

## CASO 7-4

## BEISBOL DE FANTASÍA (B)<sup>18</sup>

Ahora que el doctor Hanke siente que ha desarrollado exitosamente modelos para pronosticar el ERA y Br/IP, está listo para manejar las WINS(victorias).<sup>19</sup> Sin embargo, el consenso de los expertos indica que el

proyecto está destinado a fracasar. Algunos comentarios del sitio de Internet Ron Shandler's BaseballHQ dicen: “No existe una manera de proyectar a los lanzadores con exactitud año con año” (Bill James); “Su

<sup>18</sup> El caso 7.4 está basado en un modelo desarrollado por Internet Ron Shandler's BaseballHQ.com.

<sup>19</sup> Las estadísticas de lanzamiento de béisbol fueron descargadas del sitio de Internet Ron Shandler's BaseballHQ.com.

producto más valioso es un lanzador abridor en quien pueda confiar. El único problema es que no puede confiar en ninguno de ellos” (Peter Golenbock); “¿Dónde más en el reino de los deportes de fantasía puede tener las peores probabilidades de éxito que en el maravilloso mundo de los lanzadores?” (Ron Beaton); “Los lanzadores abridores son el grupo de personas más inestable, imprevisible, y desagradable del mundo, hablando estadísticamente, por supuesto” (John Benson).

El doctor Hanke investigó y encontró un modelo estadístico que podía utilizarse. De acuerdo con Ron Shandler’s BaseballHQ.com cuatro variables tenían un impacto importante en las victorias totales de lanzamiento. Éstas eran:

1. La ofensiva del equipo
2. La efectividad de los lanzamientos
3. El apoyo del *bullpen*
4. La defensiva del equipo

Él estaba listo para desarrollar una base de datos de predicción.

Para planear la ofensiva de un equipo, escogió las carreras anotadas por otro (RUNS)<sup>20</sup> como la variable más importante. Para indicar qué tan bueno era un equipo en general, el doctor Hanke escogió las victorias (TmWINS) del equipo. Se usaron las seis variables del caso 7.3 para indicar la efectividad del lanzador, el promedio de carreras admitidas o el número de carreras permitidas por juego (lanzando nueve entradas) (ERA), corredores en base por entrada o el número de bases por bolas más los imparables recibidos por entradas lanzadas (Br/IP), cociente de control (ponches/bases por bolas) (Ctl), cuadrangulares del equipo opuesto por partido (lanzando nueve entradas) (HR/9), promedio de bateo del equipo contrario (OBA), y la cantidad de bateadores ponchados por el lanzador por partido (lanzando nueve entradas) (K/9). Para la defensiva del equipo, utilizó el número total de errores del equipo (ERR). Para el apoyo del *bullpen* trató cuatro variables: carreras prevenidas (ARP),<sup>21</sup> apoyo del bullpen (Bsupp), salvamentos (Sv) y oportunidades de salvamento (SvOpp).

Los datos para las seis variables de lanzamiento son presentados en la tabla 7.22. Los datos para el resto de las variables se muestran en la tabla 7.26. ■

**TABLA 7.26 Estadísticas del equipo para 138 lanzadores abridores del caso 7.4**

| Lanzador           | Wins | Tm Wins | Runs  | Errors | ARP  | Bsupp | Saves | Sv Opp |
|--------------------|------|---------|-------|--------|------|-------|-------|--------|
| 1 Álvarez, W.      | 9    | 69      | 772   | 135    | 17.1 | 0.95  | 45    | 63     |
| 2 Anderson, B.     | 8    | 100     | 908   | 104    | 37.3 | 0.14  | 42    | 65     |
| 3 Appier, K.       | 16   | 87      | 893   | 122    | 11.4 | 0.96  | 48    | 76     |
| 4 Arrojo, R.       | 7    | 69      | 772   | 135    | 17.1 | 0.95  | 45    | 63     |
| 5 Ashby, A.        | 14   | 74      | 710   | 129    | 41.3 | 0.17  | 43    | 60     |
| 6 Astacio, P.      | 17   | 72      | 906   | 118    | 10.7 | 0.66  | 33    | 51     |
| 7 Baldwin, J.      | 12   | 75      | 777   | 136    | 21.3 | 0.58  | 39    | 58     |
| 8 Batista, M.      | 8    | 68      | 718   | 160    | 18.1 | 1.08  | 44    | 66     |
| 9 Belcher, T.      | 6    | 70      | 711   | 106    | 34.7 | 0.40  | 37    | 58     |
| 10 Benes, A.       | 13   | 100     | 908   | 104    | 37.3 | 0.14  | 42    | 65     |
| 11 Benson, K.      | 11   | 78      | 775   | 147    | 23.7 | 0.08  | 34    | 52     |
| 12 Bergman, S.     | 5    | 103     | 840   | 111    | 47.5 | 0.18  | 45    | 63     |
| 13 Blair, W.       | 3    | 69      | 747   | 106    | 7.9  | 0.66  | 33    | 44     |
| 14 Bohanon, B.     | 12   | 72      | 906   | 118    | 10.7 | 0.66  | 33    | 51     |
| 15 Bottenfield, K. | 18   | 75      | 809   | 132    | 44.6 | 0.90  | 38    | 62     |
| 16 Brock, C.       | 6    | 86      | 872   | 105    | 1.8  | 0.38  | 42    | 67     |
| 17 Brown, K.       | 18   | 77      | 793   | 137    | 13.9 | 0.47  | 37    | 53     |
| 18 Burba, D.       | 15   | 97      | 1,009 | 106    | 14.3 | 0.90  | 46    | 70     |
| 19 Burkett, J.     | 9    | 95      | 945   | 119    | 50.5 | 0.38  | 47    | 67     |
| 20 Byrd, P.        | 15   | 77      | 841   | 100    | 11.5 | 0.06  | 32    | 47     |
| 21 Carpenter, C.   | 9    | 84      | 883   | 106    | 5.4  | 0.09  | 39    | 60     |

<sup>20</sup> Las estadísticas para RUNS, TmWINS, Sv y SvOpp se tomaron de SportsTicker® y aparecen en [www.sportsline.com/baseball/mlb/stats](http://www.sportsline.com/baseball/mlb/stats).

<sup>21</sup> Las estadísticas para ARP y Bsupp se encontraron en una sección de Michael Wolverton en [www.Baseballprospectus.com](http://www.Baseballprospectus.com).

TABLA 7.26 (Continúa)

| <i>Lanzador</i> |                | <i>WINS</i> | <i>Tm WINS</i> | <i>Runs</i> | <i>Errors</i> | <i>ARP</i> | <i>Bsupp</i> | <i>Saves</i> | <i>Sv Opp</i> |
|-----------------|----------------|-------------|----------------|-------------|---------------|------------|--------------|--------------|---------------|
| 22              | Clemens, R.    | 14          | 98             | 900         | 111           | 40.7       | 0.21         | 50           | 67            |
| 23              | Clement, M.    | 10          | 74             | 710         | 129           | 41.3       | 0.17         | 43           | 60            |
| 24              | Colón, B.      | 18          | 97             | 1,009       | 106           | 14.3       | 0.90         | 46           | 70            |
| 25              | Cone, D.       | 12          | 98             | 900         | 111           | 40.7       | 0.21         | 50           | 67            |
| 26              | Córdova, F.    | 8           | 78             | 775         | 147           | 23.7       | 0.08         | 34           | 52            |
| 27              | Daal, O.       | 16          | 100            | 908         | 104           | 37.3       | 0.14         | 42           | 65            |
| 28              | Dempster, R.   | 7           | 64             | 691         | 127           | 71.2       | 0.73         | 33           | 53            |
| 29              | Dreifort, D.   | 13          | 77             | 793         | 137           | 13.9       | 0.47         | 37           | 53            |
| 30              | Elarton, S.    | 9           | 97             | 823         | 106           | 35.4       | 0.38         | 48           | 63            |
| 31              | Erickson, S.   | 15          | 78             | 851         | 89            | 16.9       | 0.32         | 33           | 58            |
| 32              | Escobar, K.    | 14          | 84             | 883         | 106           | 5.4        | 0.09         | 39           | 60            |
| 33              | Estes, S.      | 11          | 86             | 872         | 105           | 1.8        | 0.38         | 42           | 67            |
| 34              | Farnsworth, K. | 5           | 67             | 747         | 139           | 49.3       | 0.32         | 32           | 58            |
| 35              | Fassero, J.    | 5           | 95             | 945         | 119           | 50.5       | 0.38         | 47           | 67            |
| 36              | Fernández, A.  | 7           | 64             | 691         | 127           | 71.2       | 0.73         | 33           | 53            |
| 37              | Finley, C.     | 12          | 70             | 711         | 106           | 34.7       | 0.40         | 37           | 58            |
| 38              | García, F.     | 17          | 79             | 859         | 113           | 64.1       | 0.57         | 40           | 60            |
| 39              | Gardner, M.    | 5           | 86             | 872         | 105           | 1.8        | 0.38         | 42           | 67            |
| 40              | Glavine, T.    | 14          | 103            | 840         | 111           | 47.5       | 0.18         | 45           | 63            |
| 41              | Gooden, D.     | 3           | 97             | 1,009       | 106           | 14.3       | 0.90         | 46           | 70            |
| 42              | Guzmán, J.     | 11          | 96             | 865         | 105           | 84.8       | 0.07         | 55           | 78            |
| 43              | Halama, J.     | 11          | 79             | 859         | 113           | 64.1       | 0.57         | 40           | 60            |
| 44              | Halladay, R.   | 8           | 84             | 883         | 106           | 5.4        | 0.09         | 39           | 60            |
| 45              | Hampton, M.    | 22          | 97             | 823         | 106           | 35.4       | 0.38         | 48           | 63            |
| 46              | Harnisch, P.   | 16          | 96             | 865         | 105           | 84.8       | 0.07         | 55           | 78            |
| 47              | Hawkins, L.    | 10          | 63             | 686         | 92            | 13.2       | 0.12         | 34           | 52            |
| 48              | Haynes, J.     | 7           | 87             | 893         | 122           | 11.4       | 0.96         | 48           | 76            |
| 49              | Helling, R.    | 13          | 95             | 945         | 119           | 50.5       | 0.38         | 47           | 67            |
| 50              | Hentgen, P.    | 11          | 84             | 883         | 106           | 5.4        | 0.09         | 39           | 60            |
| 51              | Heredia, G.    | 13          | 87             | 893         | 122           | 11.4       | 0.96         | 48           | 76            |
| 52              | Hermanson, D.  | 9           | 68             | 718         | 160           | 18.1       | 1.08         | 44           | 66            |
| 53              | Hernández, O.  | 17          | 98             | 900         | 111           | 40.7       | 0.21         | 50           | 67            |
| 54              | Hernández, L.  | 8           | 86             | 872         | 105           | 1.8        | 0.38         | 42           | 67            |
| 55              | Hershiser, O.  | 13          | 97             | 853         | 68            | 58.8       | 0.15         | 49           | 69            |
| 56              | Hill, K.       | 4           | 70             | 711         | 106           | 34.7       | 0.40         | 37           | 58            |
| 57              | Hitchcock, S.  | 12          | 74             | 710         | 129           | 41.3       | 0.17         | 43           | 60            |
| 58              | Holt, C.       | 5           | 97             | 823         | 106           | 35.4       | 0.38         | 48           | 63            |
| 59              | Hudson, T.     | 11          | 87             | 893         | 122           | 11.4       | 0.96         | 48           | 76            |
| 60              | Irabu, H.      | 11          | 98             | 900         | 111           | 40.7       | 0.21         | 50           | 67            |
| 61              | Jiménez, J.    | 5           | 75             | 809         | 132           | 44.6       | 0.90         | 38           | 62            |
| 62              | Johnson, R.    | 17          | 100            | 908         | 104           | 37.3       | 0.14         | 42           | 65            |
| 63              | Johnson, J.    | 8           | 78             | 851         | 89            | 16.9       | 0.32         | 33           | 58            |
| 64              | Karl, S.       | 11          | 74             | 815         | 127           | 5.0        | 0.75         | 40           | 69            |
| 65              | Kile, D.       | 8           | 72             | 906         | 118           | 10.7       | 0.66         | 33           | 51            |
| 66              | Leiter, A.     | 13          | 97             | 853         | 68            | 58.8       | 0.15         | 49           | 69            |
| 67              | Lieber, J.     | 10          | 67             | 747         | 139           | 49.3       | 0.32         | 32           | 58            |
| 68              | Lima, J.       | 21          | 97             | 823         | 106           | 35.4       | 0.38         | 48           | 63            |
| 69              | Loaiza, E.     | 9           | 95             | 945         | 119           | 50.5       | 0.38         | 47           | 67            |
| 70              | Maddux, G.     | 19          | 103            | 840         | 111           | 47.5       | 0.18         | 45           | 63            |
| 71              | Martínez, P.   | 23          | 94             | 836         | 127           | 59.9       | 0.75         | 50           | 76            |
| 72              | Mays, J.       | 6           | 63             | 686         | 92            | 13.2       | 0.12         | 34           | 52            |

TABLA 7.26 (Continúa)

|     | <i>Lanzador</i> | <i>WINS</i> | <i>Tm WINS</i> | <i>Runs</i> | <i>Errors</i> | <i>ARP</i> | <i>Bsupp</i> | <i>Saves</i> | <i>Sv Opp</i> |
|-----|-----------------|-------------|----------------|-------------|---------------|------------|--------------|--------------|---------------|
| 73  | Meadows, B.     | 11          | 64             | 691         | 127           | 71.2       | 0.73         | 33           | 53            |
| 74  | Mercker, K.     | 8           | 94             | 836         | 127           | 59.9       | 0.75         | 50           | 76            |
| 75  | Millwood, K.    | 18          | 103            | 840         | 111           | 47.5       | 0.18         | 45           | 63            |
| 76  | Milton, E.      | 7           | 63             | 686         | 92            | 13.2       | 0.12         | 34           | 52            |
| 77  | Mlicki, D.      | 14          | 69             | 747         | 106           | 7.9        | 0.66         | 33           | 44            |
| 78  | Moehler, B.     | 10          | 69             | 747         | 106           | 7.9        | 0.66         | 33           | 44            |
| 79  | Morgan, M.      | 13          | 95             | 945         | 119           | 50.5       | 0.38         | 47           | 67            |
| 80  | Moyer, J.       | 14          | 79             | 859         | 113           | 64.1       | 0.57         | 40           | 60            |
| 81  | Mulholland, T.  | 10          | 103            | 840         | 111           | 47.5       | 0.18         | 45           | 63            |
| 82  | Mussina, M.     | 18          | 78             | 851         | 89            | 16.9       | 0.32         | 33           | 58            |
| 83  | Nagy, C.        | 17          | 97             | 1,009       | 106           | 14.3       | 0.90         | 46           | 70            |
| 84  | Navarro, J.     | 8           | 75             | 777         | 136           | 21.3       | 0.58         | 39           | 58            |
| 85  | Neagle, D.      | 9           | 96             | 865         | 105           | 84.8       | 0.07         | 55           | 78            |
| 86  | Nomo, H.        | 12          | 74             | 815         | 127           | 5.0        | 0.75         | 40           | 69            |
| 87  | Núñez, V.       | 7           | 64             | 691         | 127           | 71.2       | 0.73         | 33           | 53            |
| 88  | Ogea, C.        | 6           | 77             | 841         | 100           | 11.5       | 0.06         | 32           | 47            |
| 89  | Olivares, O.    | 15          | 87             | 893         | 122           | 11.4       | 0.96         | 48           | 76            |
| 90  | Oliver, D.      | 9           | 75             | 809         | 132           | 44.6       | 0.90         | 38           | 62            |
| 91  | Oquist, M.      | 9           | 87             | 893         | 122           | 11.4       | 0.96         | 48           | 76            |
| 92  | Ortiz, R.       | 18          | 86             | 872         | 105           | 1.8        | 0.38         | 42           | 67            |
| 93  | Park, C.        | 13          | 77             | 793         | 137           | 13.9       | 0.47         | 37           | 53            |
| 94  | Parque, J.      | 9           | 75             | 777         | 136           | 21.3       | 0.58         | 39           | 58            |
| 95  | Parris, S.      | 11          | 96             | 865         | 105           | 84.8       | 0.07         | 55           | 78            |
| 96  | Pavano, C.      | 6           | 68             | 718         | 160           | 18.1       | 1.08         | 44           | 66            |
| 97  | Person, R.      | 10          | 77             | 841         | 100           | 11.5       | 0.06         | 32           | 47            |
| 98  | Pettitte, A.    | 14          | 98             | 900         | 111           | 40.7       | 0.21         | 50           | 67            |
| 99  | Ponson, S.      | 12          | 78             | 851         | 89            | 16.9       | 0.32         | 33           | 58            |
| 100 | Portugal, M.    | 7           | 94             | 836         | 127           | 59.9       | 0.75         | 50           | 76            |
| 101 | Radke, B.       | 12          | 63             | 686         | 92            | 13.2       | 0.12         | 34           | 52            |
| 102 | Rapp, P.        | 6           | 94             | 836         | 127           | 59.9       | 0.75         | 50           | 76            |
| 103 | Reed, R.        | 11          | 97             | 853         | 68            | 58.8       | 0.15         | 49           | 69            |
| 104 | Reynolds, S.    | 16          | 97             | 823         | 106           | 35.4       | 0.38         | 48           | 63            |
| 105 | Reynoso, A.     | 10          | 100            | 908         | 104           | 37.3       | 0.14         | 42           | 65            |
| 106 | Ritchie, T.     | 15          | 78             | 775         | 147           | 23.7       | 0.08         | 34           | 52            |
| 107 | Rogers, K.      | 10          | 97             | 853         | 68            | 58.8       | 0.15         | 49           | 69            |
| 108 | Rosado, J.      | 10          | 64             | 856         | 125           | 59.6       | 1.17         | 29           | 60            |
| 109 | Rueter, K.      | 15          | 86             | 872         | 105           | 1.8        | 0.38         | 42           | 67            |
| 110 | Rupe, R.        | 8           | 69             | 772         | 135           | 17.1       | 0.95         | 45           | 63            |
| 111 | Saberhagen, B.  | 10          | 94             | 836         | 127           | 59.9       | 0.75         | 50           | 76            |
| 112 | Schilling, C.   | 15          | 77             | 841         | 100           | 11.5       | 0.06         | 32           | 47            |
| 113 | Schmidt, J.     | 13          | 78             | 775         | 147           | 23.7       | 0.08         | 34           | 52            |
| 114 | Schourek, P.    | 4           | 78             | 775         | 147           | 23.7       | 0.08         | 34           | 52            |
| 115 | Sele, A.        | 18          | 95             | 945         | 119           | 50.5       | 0.38         | 47           | 67            |
| 116 | Sirotna, M.     | 11          | 75             | 777         | 136           | 21.3       | 0.58         | 39           | 58            |
| 117 | Smoltz, J.      | 11          | 103            | 840         | 111           | 47.5       | 0.18         | 45           | 63            |
| 118 | Snyder, J.      | 9           | 75             | 777         | 136           | 21.3       | 0.58         | 39           | 58            |
| 119 | Sparks, S.      | 5           | 70             | 711         | 106           | 34.7       | 0.40         | 37           | 58            |
| 120 | Stottlemyre, T. | 6           | 100            | 908         | 104           | 37.3       | 0.14         | 42           | 65            |
| 121 | Suppan, J.      | 10          | 64             | 856         | 125           | 59.6       | 1.17         | 29           | 60            |
| 122 | Tapani, K.      | 6           | 67             | 747         | 139           | 49.3       | 0.32         | 32           | 58            |
| 123 | Thompson, J.    | 9           | 69             | 747         | 106           | 7.9        | 0.66         | 33           | 44            |

**TABLA 7.26 (Continúa)**

| <i>Lanzador</i> |              | <i>WINS</i> | <i>Tm WINS</i> | <i>Runs</i> | <i>Errors</i> | <i>ARP</i> | <i>Bsupp</i> | <i>Saves</i> | <i>Sv Opp</i> |
|-----------------|--------------|-------------|----------------|-------------|---------------|------------|--------------|--------------|---------------|
| 124             | Thurman, M.  | 7           | 68             | 718         | 160           | 18.1       | 1.08         | 44           | 66            |
| 125             | Tomko, B.    | 5           | 96             | 865         | 105           | 84.8       | 0.07         | 55           | 78            |
| 126             | Trachsel, S. | 8           | 67             | 747         | 139           | 49.3       | 0.32         | 32           | 58            |
| 127             | Valdés, I.   | 9           | 77             | 793         | 137           | 13.9       | 0.47         | 37           | 53            |
| 128             | Vázquez, J.  | 9           | 68             | 718         | 160           | 18.1       | 1.08         | 44           | 66            |
| 129             | Villone, R.  | 9           | 96             | 865         | 105           | 84.8       | 0.07         | 55           | 78            |
| 130             | Weaver, J.   | 9           | 69             | 747         | 106           | 7.9        | 0.66         | 33           | 44            |
| 131             | Wells, D.    | 17          | 84             | 883         | 106           | 5.4        | 0.09         | 39           | 60            |
| 132             | Williams, W. | 12          | 74             | 710         | 129           | 41.3       | 0.17         | 43           | 60            |
| 133             | Witasick, J. | 9           | 64             | 856         | 125           | 59.6       | 1.17         | 29           | 60            |
| 134             | Witt, B.     | 7           | 69             | 772         | 135           | 17.1       | 0.95         | 45           | 63            |
| 135             | Wolf, R.     | 6           | 77             | 841         | 100           | 11.5       | 0.06         | 32           | 47            |
| 136             | Woodard, S.  | 11          | 74             | 815         | 127           | 5.0        | 0.75         | 40           | 69            |
| 137             | Wright, J.   | 8           | 97             | 1,009       | 106           | 14.3       | 0.90         | 46           | 70            |
| 138             | Yoshii, M.   | 12          | 97             | 853         | 68            | 58.8       | 0.15         | 49           | 69            |

**PREGUNTA**

- El consenso de los expertos indica que el proyecto está destinado a fallar. ¿Los expertos están en lo cierto?

**Aplicaciones en Minitab**

**El problema.** En el ejemplo 7.11, Pam Weigand quería correr una regresión por pasos con los datos de Zurenko Pharmaceutical Company para pronosticar qué aspirante se convertiría en un buen vendedor.

**Solución con Minitab**

- Los datos están en un archivo, abra el archivo (TAB 7-13) usando los siguientes menús:

File>Open Worksheet

Si los datos no están en el archivo, escríbalos en la hoja de cálculo. Para ejecutar la regresión por pasos, dé clic en los siguientes menús:

Stat>Regression>Stepwise

- Aparece el cuadro de diálogo *Stepwise Regression* (regresión por pasos) mostrado en la figura 7.5.
  - La variable dependiente o *Response* (respuesta) es C1 o *Sales* (ventas).
  - Las variables explicativas son C2-C6 o *Aptitude-GPA*.
  - Dé clic en *Methods* (métodos) para proporcionar el nivel alfa.
- Aparece el cuadro de diálogo *Stepwise-Methods* (métodos por pasos) mostrado en la figura 7.6.
  - Dé clic en *Use alpha values* (usar valores alfa).
  - Dé clic en *Stepwise –forward and backwardk* (Pasos –hacia adelante y hacia atrás).
  - Cambie *Alpha to enter* (alfa por entrar) y *Alpha to remove* (alfa por eliminar) de 0.15 a 0.05.
  - Dé clic en OK y luego en OK en el cuadro de diálogo *Stepwise Regression* (regresión por pasos). Los resultados se muestran en la tabla 7.17.

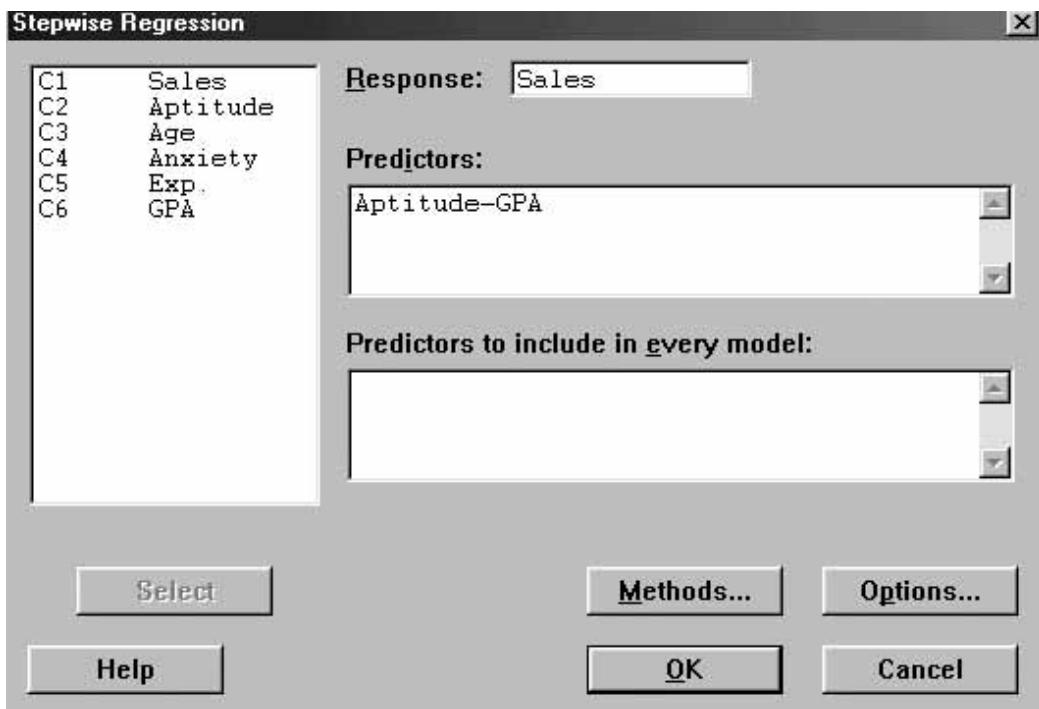


FIGURA 7.5 Cuadro de diálogo Minitab de regresión por pasos

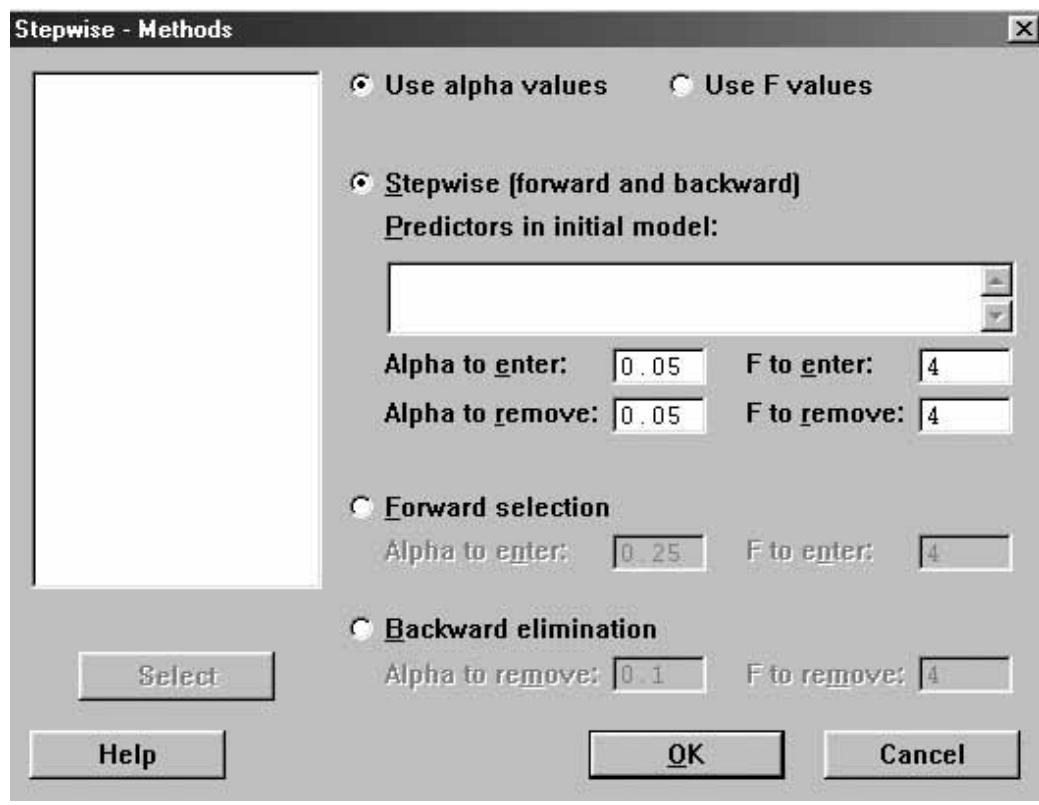


FIGURA 7.6 Cuadro de diálogo Minitab métodos por pasos

## Aplicaciones en Excel

**El problema.** En el ejemplo 7.1 se utilizó el análisis de regresión múltiple para determinar si el precio de venta y costos de publicidad podían ser usados para pronosticar las ventas semanales de galones de leche (vea la tabla 7.1 en la página 273).

### Solución de Excel

1. Escriba en la hoja de cálculo la descripción Precio en A1, Gastos de Publicidad en B1 y Ventas en C1. Introduzca el precio de venta desde A3 hasta A12, gastos de publicidad desde B3 hasta B12, y ventas semanales desde C3 hasta C12. Las dos variables explicativas deben estar en columnas adyacentes a la izquierda.
2. Dé clic en los siguientes menús:

Tools>Data Analysis

3. Aparece el cuadro de diálogo *Data Analysis* (análisis de datos). En *Analysis Tools* (herramientas de análisis) seleccione *Regression* (regresión). Aparecerá el cuadro de diálogo *Regression* mostrado en la figura 6.25 (vea la página 266).
  - a) Introduzca A3 a A12 en *Input Y Range* (rango Y de entrada).
  - b) Introduzca B3 a C12 en *Input X Range* (rango X de entrada) para las dos variables explicativas, precio de venta y costos de publicidad.
  - c) El *Output Range* (rango de salida) empieza en D1.
  - d) Dé clic en OK y aparecerá el resultado.

## Referencias

- Belsley, D. A., *Conditioning, Diagnostics, Collinearity and Weak Data in Regression*, Nueva York, John Wiley & Sons, 1991.
- Dielman, T., *Applied Regression Analysis for Business and Economics*, tercera edición, Pacific Grove, CA, Duxbury, 2001.
- Draper, N. R. y H. Smith, *Applied Regression Analysis*, tercera edición, Nueva York, John Wiley & Sons, 1998.
- Frees, E. E., *Data Analysis Using Regression Models: The Business Perspective*, Upper Saddle River, NJ, Prentice Hall, 1996.
- Johnson, R. A. y D. W. Wichern, *Business Statistics: Decision Making with Data*, Nueva York, John Wiley & Sons, 1997.
- Neter, J., W. Wasserman, M. Kutner y C. Nachtsheim, *Applied Linear Regression Models*, tercera edición, Homewood, IL, Richard D. Irwin, 1996.

## CAPÍTULO

# 8

## REGRESIÓN CON DATOS DE SERIES DE TIEMPO

Muchas aplicaciones comerciales y económicas de los pronósticos involucran datos de series de tiempo. Los modelos de regresión pueden ajustarse para datos mensuales, trimestrales o anuales por medio de las técnicas descritas en capítulos previos. Sin embargo, debido a que los datos recopilados a través del tiempo suelen mostrar tendencias, patrones estacionales y demás, las observaciones en distintos períodos están relacionadas o autocorrelacionadas. Es decir, para los datos de las series de tiempo, la muestra de observaciones no puede considerarse como aleatoria. Pueden surgir problemas de interpretación cuando los métodos de regresión estándar se aplican a las observaciones que se relacionan entre sí con el tiempo. Los modelos de regresión ajustados a los datos de las series de tiempo deben realizarse con sumo cuidado.

### DATOS DE SERIES DE TIEMPO Y EL PROBLEMA DE LA AUTOCORRELACIÓN

Los modelos de regresión analizados en los capítulos 6 y 7 suponen que los errores,  $\epsilon$ , son variables aleatorias independientes (o no correlacionadas). Esto significa que los distintos valores de la variable respuesta  $Y$  pueden relacionarse con los valores de las variables explicativas, las  $X$ , pero no entre ellas. Las interpretaciones habituales de los resultados de un análisis de regresión dependen ampliamente del supuesto de independencia.

Con datos de series de tiempo, el supuesto de independencia en raras ocasiones se cumple. Considere el precio base anual para un determinado modelo de un automóvil reciente. ¿Puede usted imaginar el caos que existiría si los precios de los automóviles nuevos de un año a otro no estuvieran relacionados (independientes) entre sí? En dicho mundo, los precios se determinarían como números sacados de una tabla de números aleatorios. El conocimiento del precio en un año no le diría nada acerca del precio del año siguiente. En el mundo real, los precios del año actual se relacionan (correlacionan) con el precio del año anterior y quizás con el de hace dos años, y así sucesivamente. Es decir, los precios de los distintos años están autocorrelacionados; no son independientes.

*La autocorrelación* existe cuando observaciones sucesivas a través del tiempo están relacionadas unas con otras.

La autocorrelación puede suceder debido a que el efecto de una variable explicativa en la respuesta se distribuye a través del tiempo. Por ejemplo, quizás un incremento en el salario afecte su consumo (o sus ahorros) no sólo en el tiempo presente sino también por varios períodos futuros. Un cambio en el precio puede afectar las ventas en el período actual y en períodos futuros. Un contrato laboral vigente puede afectar el costo de producción por algún tiempo. Con el paso del tiempo, las relaciones tienden a ser dinámicas (evolutivas), no estáticas.

Desde una perspectiva de pronóstico, la autocorrelación no es del todo mala. Si los valores de una respuesta  $Y$  en un periodo se relacionan con los valores  $Y$  en periodos anteriores, entonces las  $Y$  previas pueden emplearse para predecir las  $Y$  futuras.<sup>1</sup> En un marco de regresión, la autocorrelación se administra al “arreglar” el modelo de regresión estándar. Para acomodar la autocorrelación, a veces es necesario cambiar la mezcla de las variables explicativas o la forma de la función de regresión. Sin embargo, de forma más típica, la autocorrelación se maneja al cambiar la naturaleza del término del error.

Un tipo común de autocorrelación, que con frecuencia se llama *correlación serial de primer orden*, es aquel donde el término del error en el periodo actual se relaciona de manera directa con el término del error en el periodo previo. En este caso, con el subíndice  $t$  que representa al tiempo, el modelo de regresión lineal simple adopta la forma

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (8.1)$$

con

$$\varepsilon_t = \rho \varepsilon_{t-1} + \nu_t \quad (8.2)$$

donde

$\varepsilon_t$  = el error en el tiempo  $t$

$\rho$  = el parámetro (retraso 1 del coeficiente de autocorrelación) que mide la correlación entre los términos de error adyacentes.

$\nu_t$  = término de error independiente distribuido normalmente con media 0 y varianza  $\sigma_\nu^2$ .

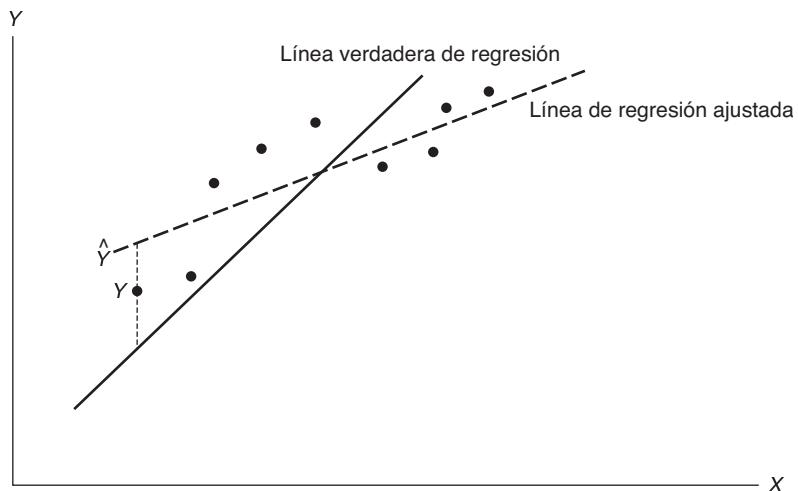
La ecuación 8.2 dice que el nivel de un término de error ( $\varepsilon_{t-1}$ ) afecta directamente el nivel del siguiente término de error ( $\varepsilon_t$ ). La magnitud del coeficiente de autocorrelación  $\rho$ , donde  $-1 \leq \rho < 1$ , indica la fuerza de la correlación serial. Si  $\rho$  es cero, entonces no existe correlación serial y los términos del error son independientes ( $\varepsilon_t = \nu_t$ ).

La figura 8.1 ilustra el efecto de una correlación serial positiva en un modelo de regresión lineal simple. Suponga la verdadera relación entre  $Y$  y  $X$ , que indica la línea sólida en la figura, crece con el tiempo. Si el primer valor de  $Y$  está por arriba de la verdadera línea de regresión, entonces los siguientes valores de  $Y$  tienen la probabilidad de estar por encima de la línea debido a la autocorrelación positiva (el primer error es positivo, de tal manera que el segundo tiene la posibilidad de ser positivo también, y así sucesivamente). Eventualmente, es probable que exista una secuencia de  $Y$  debajo de la línea verdadera de regresión (un error negativo tiene la posibilidad de preceder a un error negativo). Los datos están “inclinados” con relación a la verdadera relación  $X-Y$ . Sin embargo, la línea de mínimos cuadrados, por su verdadera naturaleza, pasará a través de las observaciones, como se indica en la línea punteada de la figura. Al usar la línea punteada para hacer inferencias acerca de la línea sólida o para generar pronósticos acerca de las futuras  $Y$  podría ser muy engañoso.

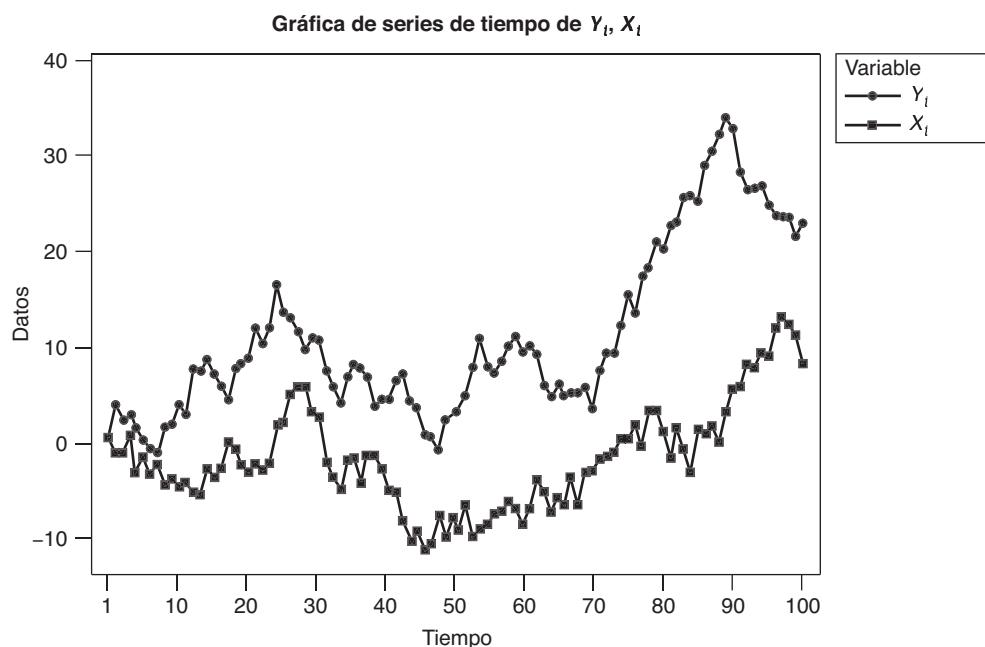
También a partir de la figura 8.1 debería quedar muy claro que la dispersión de la línea de los mínimos cuadrados es más estrecha de lo que es la dispersión sobre la línea de regresión verdadera. En consecuencia, el error estándar del estimado  $s_{y-x}$  subestimará la variabilidad de las  $Y$  en relación con la verdadera línea de regresión o, de manera equivalente, subestimará las desviaciones estándar  $\sigma$  del término de error  $\varepsilon$ .

Una fuerte autocorrelación puede hacer que dos variables que no están relacionadas aparezcan como si lo estuvieran. Los procedimientos de regresión estándar aplicados a las observaciones de estas variables pueden producir una regresión significativa. En este caso, la relación estimada es falsa y, de manera ordinaria, un examen de los residuales revelará el problema. Sin embargo, con una aplicación poco crítica de los procedimientos estándares, es posible que la regresión falsa no sea detectada y se tenga una mala interpretación de los resultados.

<sup>1</sup> La idea se considera de nuevo en una sección posterior de este capítulo cuando se analizan los modelos autorregresivos, y se verá con mayor detalle en el capítulo 9.



**FIGURA 8.1** Correlación serial positiva y el procedimiento de los mínimos cuadrados.

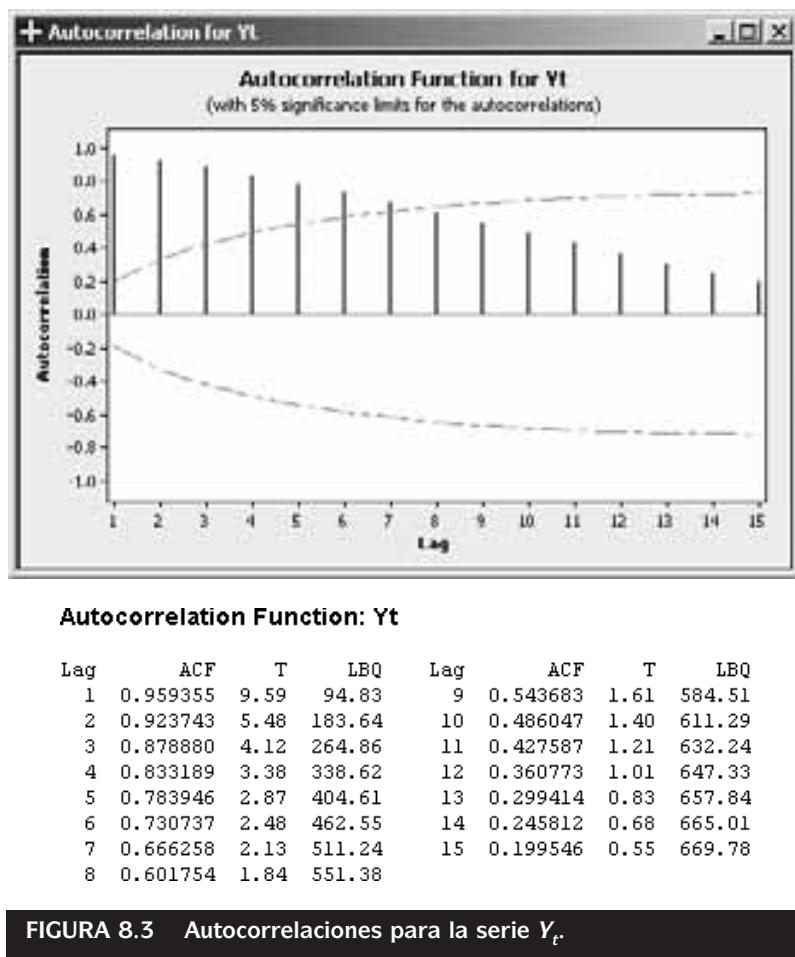


**FIGURA 8.2** Gráficas de series de tiempo de dos series no relacionadas,  $Y_t$  (superior) y  $X_t$  (inferior).

#### Ejemplo 8.1

La figura 8.2 contiene gráficas de dos series de tiempo generadas por computadora,  $Y_t$  y  $X_t$ . Estas dos series de tiempo se formaron de manera tal que la primera serie ( $Y_t$ ) no está relacionada con la segunda ( $X_t$ ).<sup>2</sup> Al mismo tiempo, cada secuencia de observaciones está altamente

<sup>2</sup> La primera serie se construyó al seleccionar una muestra aleatoria de 100 valores de una distribución normal con una media de 0 y una desviación estándar de 2, y después al formar sumas parciales. Por ejemplo,

FIGURA 8.3 Autocorrelaciones para la serie  $Y_t$ 

autocorrelacionada. Las autocorrelaciones para la primera serie se muestran en la figura 8.3. Las autocorrelaciones para la segunda serie (no mostradas) son muy similares.

La figura 8.2 indica que las dos series de tiempo aparentan moverse juntas. De hecho, esto podría permitir que se relacionen la primera serie (superior) con la segunda (inferior) por medio de un modelo de regresión lineal. En la figura 8.4 se muestra un diagrama de dispersión de datos junto con una línea de mínimos cuadrados. La regresión estimada es significativa ( $F = 83.08$ , valor  $p = .000$ ), con  $X_t$  que explica aproximadamente 46% de la variabilidad en  $Y_t$ . Así, la serie  $Y_t$  se generó independientemente de la serie  $X_t$ ; es decir, no se usaron las  $X$  para generar las  $Y$ . En este caso, la regresión estimada es falsa.

El análisis de los residuales revelaría los problemas de este análisis de regresión. Por ejemplo, las autocorrelaciones de residuales son grandes para varios retrasos, lo que indica que el supuesto de los errores independientes es erróneo y que se debe modificar el modelo de regresión inicial. En este caso, si la modificación se hace correctamente, desaparecería la relación falsa entre  $Y_t$  y  $X_t$ .

Si los modelos de regresión se usan con datos autocorrelacionados (serie de tiempo) es muy importante examinar los residuales. Si esto no es así, es posible alcanzar conclusiones que no están justificadas. La falla está en la aplicación del modelo de regresión estándar a

---

la primera observación en la serie correspondió al primer valor seleccionado en la muestra; la segunda observación en la serie fue la suma de los dos primeros valores; la tercera observación fue la suma de los tres primeros valores, y así sucesivamente. La segunda serie fue construida de la misma manera, comenzando con una muestra aleatoria *distinta* de 100 valores de la misma distribución normal.



**FIGURA 8.4 Resultados de la regresión lineal simple para dos series de tiempo altamente autocorrelacionadas,  $Y_t$  y  $X_t$**

una situación que no corresponde a los habituales supuestos de regresión. Entre los problemas técnicos que surgen están:

1. El error estándar del estimado puede subestimar seriamente la variabilidad en los términos de error.
2. Las inferencias usuales que se basan en las estadísticas de  $t$  y  $F$  ya no son estrictamente aplicables.
3. Los errores estándar de los coeficientes de regresión subestiman la variabilidad de los coeficientes estimados de regresión. Pueden resultar regresiones falsas.

## PRUEBA DE DURBIN-WATSON PARA CORRELACIÓN SERIAL

Un método que se emplea con frecuencia para determinar si una correlación serial está presente, es la *prueba Durbin-Watson*.<sup>3</sup> Esta prueba involucra la determinación de si un parámetro de correlación  $\rho$ , mostrado en la ecuación 8.2, es igual a cero. A partir de

$$\varepsilon_t = \rho \varepsilon_{t-1} + \nu_t$$

Las hipótesis que se desea probar son

$$\begin{aligned} H_0: \rho &= 0 \\ H_1: \rho &> 0 \end{aligned}$$

La hipótesis alternativa es  $\rho > 0$  porque las series de tiempo económicas y de negocios tienden a mostrar una autocorrelación positiva.

<sup>3</sup> Vea Durbin y Watson (1951). Esta prueba no se puede aplicar de forma directa si la ecuación de regresión no contiene un término constante.

Si un modelo de regresión no se considera adecuadamente para la autocorrelación, los residuales quedarán autocorrelacionados. De esta manera, la prueba Durbin-Watson se llevó a cabo mediante los residuales del análisis de regresión.

La estadística Durbin-Watson se define como:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (8.3)$$

Donde

$$e_t = Y_t - \hat{Y}_t = \text{residual del periodo } t$$

$$e_{t-1} = Y_{t-1} - \hat{Y}_{t-1} = \text{residual del periodo } t-1$$

Para una correlación serial positiva, los residuales sucesivos tienden a parecerse y la suma de las diferencias cuadradas en el numerador de la estadística Durbin-Watson serán relativamente pequeñas. Los valores mínimos de la estadística Durbin-Watson son coherentes con la correlación serial positiva.

El coeficiente de autocorrelación  $\rho$  puede estimarse por la autocorrelación del residual con el retraso 1  $r_1(e)$  (vea la ecuación 6.17) y, con un poco de cálculos matemáticos, la estadística Durbin-Watson puede relacionarse con  $r_1(e)$ . Para muestras con tamaño de moderado a grande

$$DW = 2(1 - r_1(e)) \quad (8.4)$$

Debido a que  $-1 < r_1(e) < 1$  la ecuación 8.4 muestra que  $0 < DW < 4$ . Para  $r_1(e)$  será cercana a 0, la estadística  $DW$  estará cerca de 2. Una autocorrelación del residual positiva con el retraso 1 se asocia con los valores  $DW$  menores que 2, y una autocorrelación del residual negativa con el retraso 1 se asocia con los valores  $DW$  mayores que 2.

Una prueba útil, aunque no definitiva en ciertas ocasiones, para la correlación serial se puede realizar al comparar el valor calculado de la estadística Durbin-Watson con un límite inferior ( $d_L$ ) y otro superior ( $d_U$ ). Las reglas de decisión son:

1. Cuando la estadística Durbin-Watson es mayor que el límite superior ( $d_U$ ), el coeficiente de autocorrelación  $\rho$  es igual a cero (no hay autocorrelación positiva).
2. Cuando la estadística Durbin-Watson es menor que el enlace superior ( $d_L$ ), el coeficiente de autocorrelación  $\rho$  es mayor que cero (hay autocorrelación positiva).
3. Cuando la estadística Durbin-Watson se encuentra entre los límites inferior y superior, la prueba no es concluyente (no se sabe si hay una autocorrelación positiva).

La prueba de *Durbin-Watson* se utiliza para determinar si hay una autocorrelación positiva

Si  $DW > d_U$ , se concluye que  $H_0: \rho = 0$

Si  $DW < d_L$ , se concluye que  $H_0: \rho > 0$

Si  $DW$  está entre los límites inferior y superior ( $d_L \leq DW \leq d_U$ ), la prueba no es concluyente.

Los límites críticos para  $d_L$  y  $d_U$  se proporcionan en la tabla C-6 del Apéndice. Para encontrar los  $d_L$  y  $d_U$  apropiados, el analista necesita conocer el tamaño de la muestra, el

nivel de significancia y el número de variables independientes. En la tabla de Durbin-Watson que se encuentra en el Apéndice C, el tamaño de la muestra se proporciona en la columna de la izquierda y el número de variables independientes se determina a partir del encabezado de cada columna. Si se usaron tres variables independientes, por ejemplo, uno buscaría en la columna  $k = 3$ .<sup>4</sup>

Como se indica en la ecuación 8.4, se puede inferir el signo y la magnitud del coeficiente de autocorrelación residual con el retraso 1 a partir de una estadística  $DW$  y viceversa. Así, para las situaciones en que la prueba Durbin-Watson no es concluyente, la significancia de la correlación serial puede ser investigada al comparar  $r_1(e)$  con  $\pm 2\sqrt{n}$ .<sup>5</sup> Si  $r_1(e)$  cae en el intervalo de  $0 \pm 2/\sqrt{n}$ ; entonces, se concluye que la autocorrelación es pequeña y puede ser ignorada.

### Ejemplo 8.2

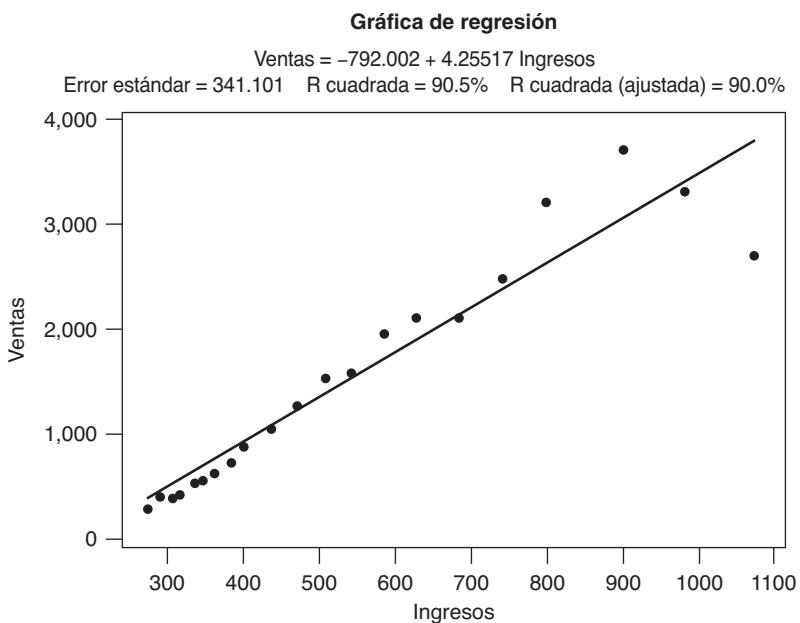
Suponga que un analista trabaja con la planeación a futuro de Reynolds Metals Company—productor de aluminio—y desea establecer una base cuantitativa para proyectar las ventas futuras. Debido a que la compañía vende regionalmente, una medida de ingresos personales disponibles para la región deberá estar íntimamente relacionada con las ventas. La tabla 8.1 muestra las ventas y los ingresos del periodo de 1976 a 1996. También en esta tabla se muestran las columnas necesarias para calcular la estadística  $DW$  (vea la sección “Aplicaciones en Minitab” que aparece al final de este capítulo). Los residuales vienen de una línea de cuadrados mínimos ajustada a los datos, como se muestran en la figura 8.5.

**TABLA 8.1 Datos de las ventas de Reynolds Metals y los cálculos de Durbin-Watson**

| Año  | Ventas<br>$Y_t$ | Ingresos<br>$X_t$ | Residuales<br>$e_t$ | $e_t - e_{t-1}$ | $(e_t - e_{t-1})^2$ | $e_t^2$      |
|------|-----------------|-------------------|---------------------|-----------------|---------------------|--------------|
| 1976 | 295             | 273.4             | -76.36              | —               | —                   | 5,830.85     |
| 1977 | 400             | 291.3             | -47.53              | 28.83           | 831.17              | 2,259.10     |
| 1978 | 390             | 306.9             | -123.91             | -76.38          | 5,833.90            | 15,353.69    |
| 1979 | 425             | 317.1             | -132.32             | -8.41           | 70.73               | 17,508.58    |
| 1980 | 547             | 336.1             | -91.16              | 41.16           | 1,694.15            | 8,310.15     |
| 1981 | 555             | 349.4             | -139.76             | -48.60          | 2,361.96            | 19,532.86    |
| 1982 | 620             | 362.9             | -132.20             | 7.56            | 57.15               | 17,476.84    |
| 1983 | 720             | 383.9             | -121.56             | 10.64           | 113.21              | 14,776.83    |
| 1984 | 880             | 402.8             | -41.98              | 79.58           | 6,332.98            | 1,762.32     |
| 1985 | 1,050           | 437.0             | -17.51              | 24.47           | 598.78              | 306.60       |
| 1986 | 1,290           | 472.2             | 72.71               | 90.22           | 8,139.65            | 5,286.74     |
| 1987 | 1,528           | 510.4             | 148.16              | 75.45           | 5,692.70            | 21,951.39    |
| 1988 | 1,586           | 544.5             | 61.06               | -87.10          | 7,586.41            | 3,728.32     |
| 1989 | 1,960           | 588.1             | 249.53              | 188.47          | 35,520.94           | 62,265.22    |
| 1990 | 2,118           | 630.4             | 227.54              | -21.99          | 483.56              | 51,774.45    |
| 1991 | 2,116           | 685.9             | -10.62              | -238.16         | 56,720.19           | 112.78       |
| 1992 | 2,477           | 742.8             | 108.26              | 118.88          | 14,132.45           | 11,720.23    |
| 1993 | 3,199           | 801.3             | 581.33              | 473.07          | 223,795.22          | 337,944.57   |
| 1994 | 3,702           | 903.1             | 651.16              | 69.83           | 4,876.23            | 424,009.35   |
| 1995 | 3,316           | 983.6             | -77.38              | -728.54         | 530,770.53          | 5,987.66     |
| 1996 | 2,702           | 1,076.7           | -1,087.54           | -1,010.16       | 1,020,423.23        | 1,182,743.25 |
|      |                 |                   |                     | Totals          | 1,926,035.14        | 2,210,641.78 |

<sup>4</sup> También es posible probar la existencia de una autocorrelación negativa. En este caso  $H_1: \rho < 0$  y la prueba estadística  $DW$  se compara con  $4 - d_L$  y  $4 - d_U$ . Se rechaza la hipótesis nula  $H_0: \rho < 0$  se rechaza si  $DW > 4 - d_L$  y no se rechaza si  $DW < 4 - d_U$ . La prueba no es concluyente cuando  $DW$  está entre  $4 - d_U$  y  $4 - d_L$ .

<sup>5</sup> Si no hubiera autocorrelación, el error estándar de  $r_1(e)$  es aproximadamente  $1/\sqrt{n}$  (vea el análisis de la autocorrelación en el capítulo 3 y el análisis de la autocorrelación residual en el capítulo 6).



**FIGURA 8.5 Gráfica de regresión para los datos de Reynolds Metals**

Antes de usar la línea de mínimos cuadrados para el pronóstico, el analista lleva a cabo una prueba de Durbin-Watson para una correlación serial positiva. Los cálculos de las últimas tres columnas para 1977 son:

$$\begin{aligned} e_t - e_{t-1} &= -47.53 - (-76.36) = 28.83 \\ (e_t - e_{t-1})^2 &= 28.83^2 = 831.17 \\ e_t^2 &= (-47.53)^2 = 2,259.1 \end{aligned}$$

La estadística de Durbin-Watson se calcula como

$$DW = \frac{\sum_{t=2}^{21} (e_t - e_{t-1})^2}{\sum_{t=1}^{21} e_t^2} = \frac{1,926,035.14}{2,210,641.78} = .87$$

Al utilizar un nivel de 0.01 de significancia para una muestra de  $n = 21$  y una variable independiente  $k = 1$ , se obtiene

$$\begin{aligned} d_L &= .97 \\ d_U &= 1.16 \end{aligned}$$

$DW = 0.87$  ahora está por debajo de  $d_L = .97$ , de manera que la hipótesis nula  $H_0: \rho = 0$  es rechazada, y se concluye que los errores están correlacionados positivamente ( $\rho > 0$ ). El modelo de regresión deberá modificarse antes de que se utilice para realizar pronósticos.

## **SOLUCIONES PARA LOS PROBLEMAS DE AUTOCORRELACIÓN**

Después que se encontró una autocorrelación en una regresión de datos de una serie de tiempo, es necesario eliminarla o modelarla antes de que la función de regresión pueda

evaluarse por su eficacia. El método apropiado para tratar una correlación serial depende de lo que, en primer lugar, la ocasionó. La autocorrelación puede surgir debido a un error de especificación, como cuando se omite una variable, o porque los términos independientes del error están correlacionados en un modelo especificado correctamente de otra manera.

La solución al problema de la correlación serial comienza con una evaluación de la especificación del modelo. ¿Es correcta la forma funcional? ¿Se omitieron algunas variables importantes? ¿Hay efectos que pudieron tener algún patrón sobre el tiempo y que hayan introducido una autocorrelación en los errores?

Debido a que la causa más importante de los errores autocorrelacionados en un modelo de regresión es la omisión de una o más variables básicas, el mejor método para resolver el problema es encontrarlas. En ocasiones, a este esfuerzo se le conoce como mejoría de la especificación del modelo. La especificación del modelo involucra no sólo encontrar las variables explicativas importantes, sino también introducirlas correctamente en la función de regresión. Desafortunadamente no siempre se puede mejorar la especificación del modelo porque no sería posible cuantificar una variable importante perdida, y si pudiera hacerse, los datos no estarían disponibles. Por ejemplo, uno puede sospechar que la inversión futura en un negocio está relacionada con la actitud de los inversionistas potenciales. Sin embargo, es difícil cuantificar la variable “actitud”. No obstante, cuando es posible, el modelo deberá especificarse de acuerdo con una pretendida perspicacia.

Sólo después de que se ha revisado cuidadosamente la especificación de la ecuación, debería considerarse la posibilidad de un ajuste. Se analizarán diversas técnicas para eliminar la autocorrelación.

Un método para eliminar la autocorrelación es agregar a la función de regresión una variable omitida que explique la asociación en la respuesta de un periodo al siguiente.

Otro método involucra la noción general de diferenciar. En éste, el modelo de regresión se especifica en términos de *cambios* en lugar de *niveles*. Por ejemplo, si se usan los datos de Reynolds Metals, el cambio en las ventas de un año al siguiente (de 105 para el periodo de 1976 a 1977) se relaciona con el cambio correspondiente en los ingresos (17.9 para el periodo de 1976 en 1977). A veces, las variables originales pueden expresarse en términos de logaritmos y cambios en los logaritmos que se usan en la regresión. Este procedimiento es equivalente a realizar una regresión utilizando el cambio porcentual en la respuesta y cambios porcentuales en las variables explicativas. Finalmente, en lugar de utilizar diferencias simples o primarias en el modelo de regresión, podría darse el caso de que se requieran *diferencias generalizadas* para eliminar la autocorrelación.

El método del modelo autorregresivo para eliminar la autocorrelación genera variables explicativas mediante la variable de respuesta *Y* retrasada en uno o más periodos. En el caso del modelo autorregresivo de primer orden, la única variable explicativa es la variable *Y* atrasada un periodo. Al utilizar de nuevo los datos de Reynolds Metals, el valor usado para predecir las ventas de 1977 son las ventas de 1976 (295).

Los siguientes ejemplos ilustran estos métodos para eliminar la autocorrelación.

### Error de especificación del modelo (omisión de una variable)

El ejemplo 8.3 muestra la manera en que al incorporar una variable perdida se puede eliminar la correlación serial.

#### Ejemplo 8.3

La Novak Corporation desea desarrollar un modelo de pronóstico para la proyección de las ventas futuras. Ya que la corporación tiene tiendas de fábrica a lo largo de la región, se eligen los ingresos personales disponibles, sobre la base de una región amplia, como una variable explicativa posible. La tabla 8.2 muestra las ventas de Novak de 1980 a 1996. También muestra los ingresos personales disponibles y el desempleo existente en la región.

**TABLA 8.2** Datos de las ventas de Novak Corporation para los ejemplos 8.3 y 8.6

| Fila | Año  | Ventas | Ingresos | Índice | Y atrasada |
|------|------|--------|----------|--------|------------|
| 1    | 1980 | 8.0    | 336.1    | 5.5    | —          |
| 2    | 1981 | 8.2    | 349.4    | 5.5    | 8.0        |
| 3    | 1982 | 8.5    | 362.9    | 6.7    | 8.2        |
| 4    | 1983 | 9.2    | 383.9    | 5.5    | 8.5        |
| 5    | 1984 | 10.2   | 402.8    | 5.7    | 9.2        |
| 6    | 1985 | 11.4   | 437.0    | 5.2    | 10.2       |
| 7    | 1986 | 12.8   | 472.2    | 4.5    | 11.4       |
| 8    | 1987 | 13.6   | 510.4    | 3.8    | 12.8       |
| 9    | 1988 | 14.6   | 544.5    | 3.8    | 13.6       |
| 10   | 1989 | 16.4   | 588.1    | 3.6    | 14.6       |
| 11   | 1990 | 17.8   | 630.4    | 3.5    | 16.4       |
| 12   | 1991 | 18.6   | 685.9    | 4.9    | 17.8       |
| 13   | 1992 | 20.0   | 742.8    | 5.9    | 18.6       |
| 14   | 1993 | 21.9   | 801.3    | 5.6    | 20.0       |
| 15   | 1994 | 24.9   | 903.1    | 4.9    | 21.9       |
| 16   | 1995 | 27.3   | 983.6    | 5.6    | 24.9       |
| 17   | 1996 | 29.1   | 1076.7   | 8.5    | 27.3       |

**TABLA 8.3** Resultados de Minitab para las ventas de Novak y el ingreso personal disponible para el ejemplo 8.3**Regression Analysis: Sales versus Income**

The regression equation is  
 Sales = -1.50 + 0.0292 Income

| Predictor | Coef      | SE Coef   | T     | P     |
|-----------|-----------|-----------|-------|-------|
| Constant  | -1.5046   | 0.3290    | -4.57 | 0.000 |
| Income    | 0.0291916 | 0.0005129 | 56.92 | 0.000 |

S = 0.4767      R-Sq = 99.5%      R-Sq(adj) = 99.5%

**Analysis of Variance**

| Source         | DF | SS     | MS     | F       | P     |
|----------------|----|--------|--------|---------|-------|
| Regression     | 1  | 736.15 | 736.15 | 3239.89 | 0.000 |
| Residual Error | 15 | 3.41   | 0.23   |         |       |
| Total          | 16 | 739.56 |        |         |       |

Durbin-Watson statistic = 0.72

En la tabla 8.3, la estadística Durbin-Watson es de 0.72 y al usar un nivel de significancia de 0.01 con  $n = 17$  y  $k = 1$ , la tabla C-6 del Apéndice da

$$\begin{aligned} d_L &= .87 \\ d_U &= 1.10 \end{aligned}$$

Ya a que  $DW = .72 < d_L = .87$  se indica una correlación serial positiva. Una variable importante que explica la asociación restante en las ventas de un año al siguiente, podría faltar en el modelo. El resultado puede ser verdadero, incluso cuando el resultado de Minitab indica que el ingreso disponible es explicativo de 99.5% de la variabilidad en las ventas.

**TABLA 8.4 Resultados de Minitab para las ventas de Novak, el ingreso personal disponible y la tasa de desempleo para el ejemplo 8.3**

```

Regression Analysis: Sales versus Income, Rate
The regression equation is
Sales = -0.014 + 0.0297 Income - 0.350 Rate

Predictor          Coef        SE Coef      T      P
Constant          -0.0140     0.2498    -0.06  0.956
Income            0.0297492  0.0002480   119.96 0.000
Rate              -0.34987    0.04656    -7.51  0.000

S = 0.2199      R-Sq = 99.9%      R-Sq(adj) = 99.9%
Analysis of Variance

Source           DF       SS       MS      F      P
Regression       2     738.88  369.44  7637.91 0.000
Residual Error  14      0.68     0.05
Total            16     739.56

Durbin-Watson statistic = 1.98

```

La tasa de desempleo puede ser una variable explicativa faltante e importante para las ventas. La tabla 8.4 muestra los resultados del análisis de regresión cuando la tasa de desempleo se añade al modelo.

Ahora, el modelo ajustado explica 99.9% de la variabilidad en las ventas. Aunque la intercepción Y no es significativa, el reajuste del modelo sin la intercepción deja sin cambios esenciales a la estadística Durbin-Watson, que es de 1.98.

Con un nivel de 0.01 de significancia, y con  $n = 17$  y  $k = 2$ , la tabla C-6 del Apéndice proporciona

$$\begin{aligned} d_L &= .77 \\ d_U &= 1.25 \end{aligned}$$

Puesto que  $DW = 1.98 > d_U = 1.25$  no hay evidencia alguna de que exista una correlación serial.

La función  $\hat{Y} = -.014 + .03X_1 - .35X_2$  puede utilizarse para predecir las ventas de Novak a sabiendas de que los errores son independientes.<sup>6</sup> Los estimados expertos del ingreso personal (\$1,185 millones) y la tasa de desempleo (7.8%) para la región se utilizan para generar un pronóstico de las ventas de Novak para 1997. El pronóstico es

$$\hat{Y} = -.014 + .03(1,185) - .35(7.8) = 32.8$$

o \$32.8 millones.

### Regresión con diferencias

Para los datos que están altamente correlacionados, modelar los cambios y no los niveles, frecuentemente eliminan la correlación serial. Es decir, en lugar de formular la ecuación de regresión en términos de Y y  $X_1, X_2, \dots, X_k$ , la ecuación de regresión se escribe en términos de las diferencias,  $Y'_t = Y_t - Y_{t-1}$  y  $X'_{t1} = X_{t1} - X_{t-1,1}, X'_{t2} = X_{t2} - X_{t-1,2}$ , y así sucesivamente. Hay que considerar las diferencias cuando la estadística Durbin-Watson, que está asociada con la regresión que involucra a las variables originales, es cercana a 0.<sup>7</sup>

<sup>6</sup> Las gráficas residuales indican que no hay razón para dudar de los supuestos del modelo de regresión.

<sup>7</sup> Un patrón de autocorrelación para la variable Y o para las variables X, como el que se ilustra en la figura 8.3 (vea la página 330) también indica que la función de regresión con diferencias podría eliminar (o en gran medida reducir) los problemas ocasionados por una correlación serial.

Una base racional para la diferenciación procede del siguiente argumento. Suponga que las ecuaciones 8.1 y 8.2 sostienen que

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

con

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t$$

donde

$\rho$  = correlación entre los errores consecutivos

$v_t$  = error aleatorio

$v_t = \varepsilon_t$  cuando  $\rho = 0$

El modelo se cumple para cualquier periodo, de manera que

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + \varepsilon_{t-1}$$

Al multiplicar esta ecuación en ambos lados por  $\rho$ , y después restar y reacomodar términos en la ecuación 8.1 se obtiene

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (\text{ecuación 8.1})$$

$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{t-1} + \rho \varepsilon_{t-1} \quad (\text{multiplicar } Y_{t-1} \text{ por } \rho)$$

$$Y_t - \rho Y_{t-1} = \beta_0 - \rho \beta_0 + (\beta_1 X_t - \rho \beta_1 X_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}) \quad (\text{restar})$$

o

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t \quad (8.5)$$

donde la “prima” indica las *diferencias generalizadas*

$$Y'_t = Y_t - \rho Y_{t-1}$$

$$X'_t = X_t - \rho X_{t-1} \quad (8.6)$$

El modelo de la ecuación 8.5 tiene errores  $v_t$  que se distribuyen de manera independiente con un promedio igual a cero, en una varianza constante. Así, los métodos usuales de regresión pueden aplicarse a este modelo.

Si la correlación entre los errores consecutivos es fuerte ( $\rho$  cercana a 1), las diferencias generalizadas son en esencia diferencias simples, o primeras diferencias

$$Y'_t = Y_t - Y_{t-1}$$

$$X'_t = X_t - X_{t-1} \quad (8.7)$$

y la constante de intercepción en el modelo (8.5) es cercano a cero (desaparece).

A menudo, mediante el uso de modelos de regresión construidos con diferencias generalizadas se puede eliminar la correlación serial. Si la correlación serial es en particular fuerte, se pueden utilizar las diferencias simples. El ejemplo 8.1 muestra lo que puede suceder si se ignora una autocorrelación fuerte.

**Ejemplo 8.4**

Fred Gardner está a cargo del pronóstico de las ventas en importe para la región occidental de Sears Roebuck. Eligió como su variable independiente los ingresos personales disponibles en la región. Al relacionar las ventas con el ingreso disponible mediante un modelo de regresión lineal logarítmica, Fred podrá estimar la elasticidad de los ingresos por ventas. La elasticidad mide el cambio del porcentaje en las ventas que se deben a un cambio de 1% en el ingreso.

El modelo de regresión lineal logarítmica supone que el ingreso está relacionado con las ventas por la ecuación

$$\text{Ventas} = \gamma(\text{Ingreso})^{81}$$

Al tomar logaritmos naturales en ambos lados de la ecuación se obtiene

$$\ln(\text{Ventas}) = \ln \gamma + \beta_1 \ln(\text{Ingreso})$$

Al añadir un término de error y considerarlo para la influencia de las variables, a excepción del ingreso en las ventas, la expresión anterior se vuelve un *modelo de regresión lineal logarítmica* de la forma

$$\ln Y_t = \beta_0 + \beta_1 \ln X_t + \varepsilon_t \quad (8.8)$$

donde

$\ln Y_t = \ln (\text{Ventas})$  = logaritmo natural de ventas

$\ln X_t = \ln (\text{Ingreso})$  = logaritmo natural del ingreso

$\varepsilon_t$  = el término del error

$\beta_0$  =  $\ln \gamma$  el coeficiente de intercepción

$\beta_1$  = el coeficiente de la pendiente = la elasticidad del ingreso de las ventas

La tabla 8.5 muestra las ventas de Sears, el ingreso disponible, sus logaritmos y las diferencias en los logaritmos de las ventas y el ingreso disponible para el periodo de 1976 a 1996.

**TABLA 8.5 Ventas de Sears y el ingreso disponible en EU de 1976 a 1996, junto con datos transformados para el ejemplo 8.4**

| Año  | Ventas<br>(miles) | Ingreso<br>(millones) | Diferencias |           |       |        |
|------|-------------------|-----------------------|-------------|-----------|-------|--------|
|      | $Y_t$             | $X_t$                 | $\ln Y_t$   | $\ln X_t$ | $Y'$  | $X'_t$ |
| 1976 | 3,307             | 273.4                 | 8.1038      | 5.6109    | —     | —      |
| 1977 | 3,556             | 291.3                 | 8.1764      | 5.6744    | .0726 | .0634  |
| 1978 | 3,601             | 306.9                 | 8.1890      | 5.7265    | .0126 | .0522  |
| 1979 | 3,721             | 317.1                 | 8.2218      | 5.7592    | .0328 | .0327  |
| 1980 | 4,036             | 336.1                 | 8.3030      | 5.8174    | .0813 | .0582  |
| 1981 | 4,134             | 349.4                 | 8.3270      | 5.8562    | .0240 | .0388  |
| 1982 | 4,268             | 362.9                 | 8.3589      | 5.8941    | .0319 | .0379  |
| 1983 | 4,578             | 383.9                 | 8.4290      | 5.9504    | .0701 | .0563  |
| 1984 | 5,093             | 402.8                 | 8.5356      | 5.9984    | .1066 | .0481  |
| 1985 | 5,716             | 437.0                 | 8.6510      | 6.0799    | .1154 | .0815  |
| 1986 | 6,357             | 472.2                 | 8.7573      | 6.1574    | .1063 | .0775  |
| 1987 | 6,769             | 510.4                 | 8.8201      | 6.2352    | .0628 | .0778  |
| 1988 | 7,296             | 544.5                 | 8.8951      | 6.2999    | .0750 | .0647  |
| 1989 | 8,178             | 588.1                 | 9.0092      | 6.3769    | .1141 | .0770  |
| 1990 | 8,844             | 630.4                 | 9.0875      | 6.4464    | .0783 | .0695  |
| 1991 | 9,251             | 685.9                 | 9.1325      | 6.5307    | .0450 | .0844  |
| 1992 | 10,006            | 742.8                 | 9.2109      | 6.6104    | .0785 | .0797  |
| 1993 | 11,200            | 801.3                 | 9.3237      | 6.6862    | .1127 | .0758  |
| 1994 | 12,500            | 903.1                 | 9.4335      | 6.8058    | .1098 | .1196  |
| 1995 | 13,101            | 983.6                 | 9.4804      | 6.8912    | .0470 | .0854  |
| 1996 | 13,640            | 1,076.7               | 9.5208      | 6.9817    | .0403 | .0904  |

Una parte de los resultados de Minitab que muestra la regresión de  $\ln(\text{ventas})$  en  $\ln(\text{ingresos})$  se ilustra en la tabla 8.6. Fred se da cuenta de que 99.2% de la variabilidad en los logaritmos de las ventas de Sears para la región occidental, puede explicarse por su relación con el logaritmo del ingreso disponible para la misma región. La regresión es altamente significativa. Además, la elasticidad del ingreso se estima en  $b_1 = 1.117$  con un error estándar de  $sb_1 = .023$ . Sin embargo, la estadística Durbin-Watson de 0.50 es pequeña y menor que  $d_L = .97$  el más bajo nivel del valor crítico 0.01 para  $n = 21$  y  $k = 1$ . Fred concluye que la correlación entre los errores sucesivos es positiva y grande (cercana a 1).

Debido a la gran correlación serial, Fred decide modelar los cambios (diferencias) en los logaritmos de las ventas y del ingreso, respectivamente. Entiende que el coeficiente de la pendiente del modelo para las diferencias es el mismo en el modelo original, que involucra los logaritmos; por lo tanto, todavía puede estimar la elasticidad del ingreso de manera directa. Es probable que el coeficiente de intercepción en el modelo de regresión para las diferencias sea pequeño y se le omita. Los resultados en Minitab para los cambios se muestran en la tabla 8.7.

La tabla 8.7 muestra que la regresión es significativa. La elasticidad del ingreso se estima en  $b_1 = 1.010$  con un error estándar de  $sb_1 = .093$ . El estimado de la elasticidad  $b_1$  no cambió demasiado desde la primera regresión (un incremento de 1% en el ingreso disponible lleva a un incremento aproximado de 1% en las ventas anuales para ambos casos), pero su error estándar actual ( $sb_1 = .093$ ) es aproximadamente cuatro veces mayor que el error estándar anterior ( $sb_1 = .023$ ). Es probable que el error estándar previo subestime el error estándar verdadero debido a la correlación serial.

Al revisar la estadística Durbin-Watson para  $n = 20, k = 1$  y un nivel de significancia de 0.05, Fred descubre que  $d_L = 1.20 < DW = 1.28 < d_U = 1.41$  de manera que la prueba para la correlación serial positiva no es concluyente. No obstante, una revisión de las autocorrelaciones residuales, como se muestra en la figura 8.6, indica que todas están dentro de sus dos límites de error estándar (las líneas punteadas de la figura) para los primeros y escasos retrasos. Fred concluye que la correlación serial ha sido eliminada y que utilizará la ecuación ajustada para realizar los pronósticos.

Con el fin de utilizar el modelo final para pronosticar, Fred escribe

$$\hat{Y}'_t = 1.01X'_t \quad \text{donde} \quad \begin{aligned} \hat{Y}'_t &= \ln \hat{Y}_t - \ln \hat{Y}_{t-1} \\ X'_t &= \ln X_t - \ln X_{t-1} \end{aligned}$$

**TABLA 8.6 Resultados de Minitab para la regresión de los logaritmos de las ventas de Sears en los logaritmos del ingreso disponible para el ejemplo 8.4**

**Regression Analysis: Ln(Sales) versus Ln(Income)**

The regression equation is

$$\ln(\text{Sales}) = 1.82 + 1.12 \ln(\text{Income})$$

| Predictor  | Coef    | SE Coef | T     | P     |
|------------|---------|---------|-------|-------|
| Constant   | 1.8232  | 0.1434  | 12.71 | 0.000 |
| Ln(Income) | 1.11727 | 0.02305 | 48.47 | 0.000 |

$$S = 0.04368 \quad R-\text{Sq} = 99.2\% \quad R-\text{Sq}(\text{adj}) = 99.2\%$$

#### Analysis of Variance

| Source         | DF | SS     | MS     | F       | P     |
|----------------|----|--------|--------|---------|-------|
| Regression     | 1  | 4.4821 | 4.4821 | 2349.13 | 0.000 |
| Residual Error | 19 | 0.0363 | 0.0019 |         |       |
| Total          | 20 | 4.5184 |        |         |       |

Durbin-Watson statistic = 0.50

**TABLA 8.7** Resultados de Minitab para la regresión de los cambios en los logaritmos de las ventas de Sears en los cambios de los logaritmos del ingreso disponible para el ejemplo 8.4

```
Regression Analysis: Change in Ln(Sales) versus Change in Ln(Income)
The regression equation is
Change in Ln(Sales) = 1.01 Change in Ln(Income)

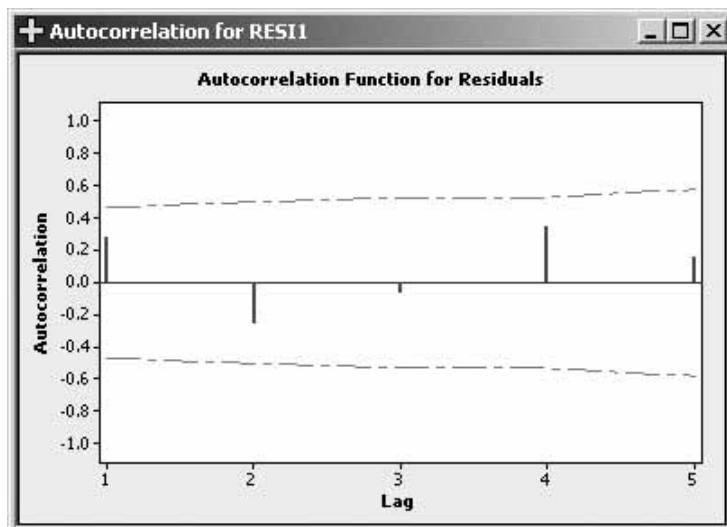
Predictor          Coef    SE Coef      T      P
Noconstant
Change in Ln(Income)  1.00989  0.09304  10.85  0.000
S = 0.02975

Analysis of Variance

Source           DF      SS      MS      F      P
Regression        1  0.10428  0.10428  117.83  0.000
Residual Error   19  0.01681  0.00088
Total            20  0.12109

Durbin-Watson statistic = 1.28
```

---



**Autocorrelation Function: RESI1**

| Lag | ACF       | T     | LBQ  |
|-----|-----------|-------|------|
| 1   | 0.274432  | 1.23  | 1.74 |
| 2   | -0.257852 | -1.08 | 3.37 |
| 3   | -0.065264 | -0.26 | 3.48 |
| 4   | 0.349965  | 1.38  | 6.85 |
| 5   | 0.160779  | 0.58  | 7.61 |

**FIGURA 8.6** Las autocorrelaciones residuales para el análisis de regresión en la tabla 8.7

Al sustituir  $\hat{Y}'_t$  y  $X'_t$  y acomodar nuevamente los términos

$$\ln \hat{Y}'_t = \ln \hat{Y}'_{t-1} + 1.01(\ln X_t - \ln X_{t-1}) \quad (8.9)$$

El pronóstico para las ventas de Sears en 1997 se obtiene al establecer  $t = 22$ :

$$\ln \hat{Y}'_{22} = \ln \hat{Y}'_{21} + 1.01(\ln X_{22} - \ln X_{21})$$

Las ventas en 1996 son conocidas, entonces  $\hat{Y}'_{21} = T_{21} = 13,640$ . También se conoce el ingreso disponible, de manera que  $X_{21} = 1,076.7$ . Para continuar, Fred necesita el ingreso disponible para 1997. Un economista familiarizado con la región occidental le manda a Fred un estimado de \$1,185 millones como ingreso disponible para 1997. Fred emplea el estimado del experto y establece  $X_{22} = 1,185$ . La ecuación del pronóstico se vuelve

$$\begin{aligned}\ln \hat{Y}'_{22} &= \ln (13,640) + 1.01(\ln (1,185) - \ln (1,076.7)) \\ &= 9.5208 + 1.01(7.0775 - 6.9817) = 9.6176\end{aligned}$$

o, mediante antilogaritmos

$$\hat{Y}_{22} = e^{9.6176} = 15,027$$

El pronóstico de Fred de las ventas de Sears para la región occidental es de \$15,027 miles. Fred puede emplear la ecuación 8.9 y el procedimiento descrito con anterioridad para generar pronósticos para los años de 1998, 1999 y así sucesivamente, pero para hacerlos necesita los estimados de los ingresos personales disponibles para esos años.

### Errores autocorrelacionados y diferencias generalizadas

El objetivo consiste en describir adecuadamente la naturaleza de la relación que existe entre las variables  $Y$  y  $X$  cuando hay una correlación serial.

Considere nuevamente el modelo de regresión con errores correlacionados serialmente (vea las ecuaciones 8.1 y 8.2):

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t$$

Se dice que los errores  $\varepsilon_t$  siguen un modelo autorregresivo de primer orden o modelo AR(1).<sup>8</sup>

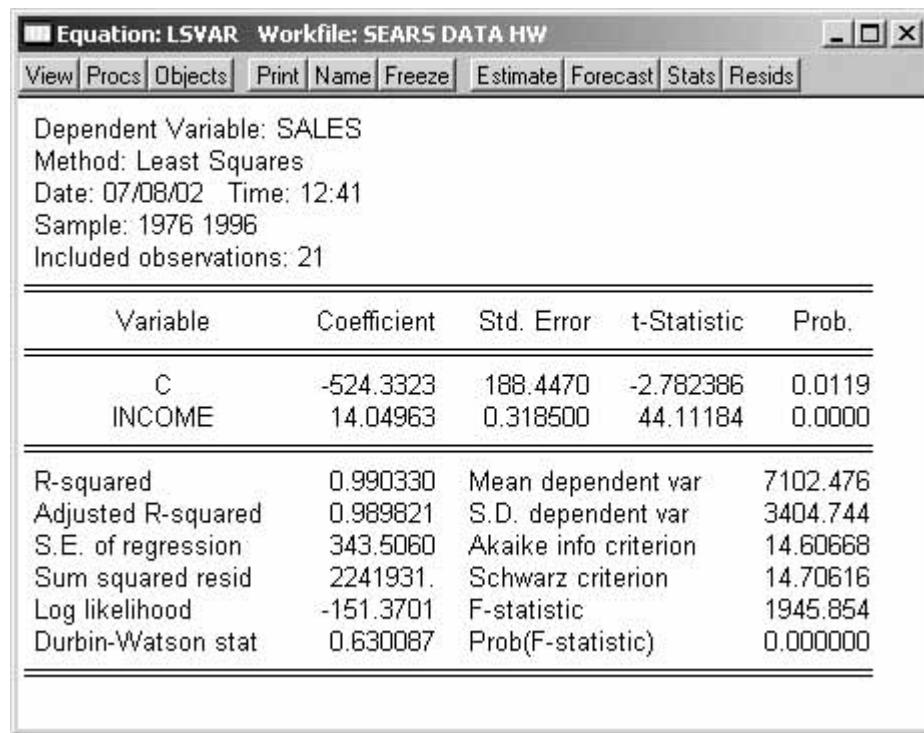
Recuerde que después de algunos cálculos algebraicos, el sistema de ecuaciones mencionado puede escribirse con un modelo de regresión lineal simple que involucre diferencias generalizadas  $Y'_t = Y_t - \rho Y_{t-1}$  y  $X'_t = X_t - \rho X_{t-1}$  (vea la ecuación 8.5):

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t$$

El modelo de regresión que involucra las diferencias generalizadas se construye específicamente para eliminar la correlación serial en los errores. Se supone que los errores  $v_t$  son independientes y se distribuyen normalmente con una media de cero y una varianza constante de  $\sigma_v^2$ . En consecuencia, parece razonable ajustar este modelo directamente por la aplicación del principio de mínimos cuadrados. Sin embargo, no se conoce el coeficiente de autocorrelación  $\rho$  de manera que  $Y'_t$  y  $X'_t$  tampoco pueden determinarse. Por lo tanto, el modelo no puede ajustarse por medio de mínimos cuadrados ordinarios.

Hay dos métodos generales para estimar los parámetros  $\beta_0$  y  $\beta_1$  de interés primario. Un método consiste en utilizar el modelo corregido para la correlación serial (ecuación 8.5) y estimar los parámetros  $\beta_0$ ,  $\beta_1$  y  $\rho$  directamente, por medio de una técnica numérica conocida como *mínimos cuadrados no lineales*. Este método utiliza una rutina de búsqueda

<sup>8</sup> Los modelos autorregresivos de primer orden se introducen formalmente en la siguiente sección de este capítulo y se analizan nuevamente en el capítulo 9.



**FIGURA 8.7** Resultados de E-Views para la regresión de las ventas de Sears con base en el ingreso disponible para el ejemplo 8.5

para encontrar los valores de los parámetros que minimizan el error en la suma de los cuadrados  $\sum v_t^2$ . El otro método consiste en estimar  $\rho$ , usar el estimado de  $\hat{\rho}$  para construir las diferencias generalizadas, y después ajustar el modelo con estas diferencias generalizadas mediante el método de los mínimos cuadrados ordinarios.<sup>9</sup>

El siguiente ejemplo ilustra el método de los mínimos cuadrados no lineales con el uso de los resultados de E-Views, un programa de computación popular para el modelado econométrico.

#### Ejemplo 8.5

Los datos de sears se proporcionan en la tabla 8.5. se usa el método de mínimos cuadrados ordinarios para ajustar un modelo de regresión lineal simple y así relacionar las ventas con el ingreso personal disponible. El resultado de E-views se muestra en la figura 8.7. como se observa en la figura 8.7, la ecuación de regresión ajustada es

$$\hat{Y}_t = -524.33 + 14.05X_t$$

donde

$$b_1 = 14.04$$

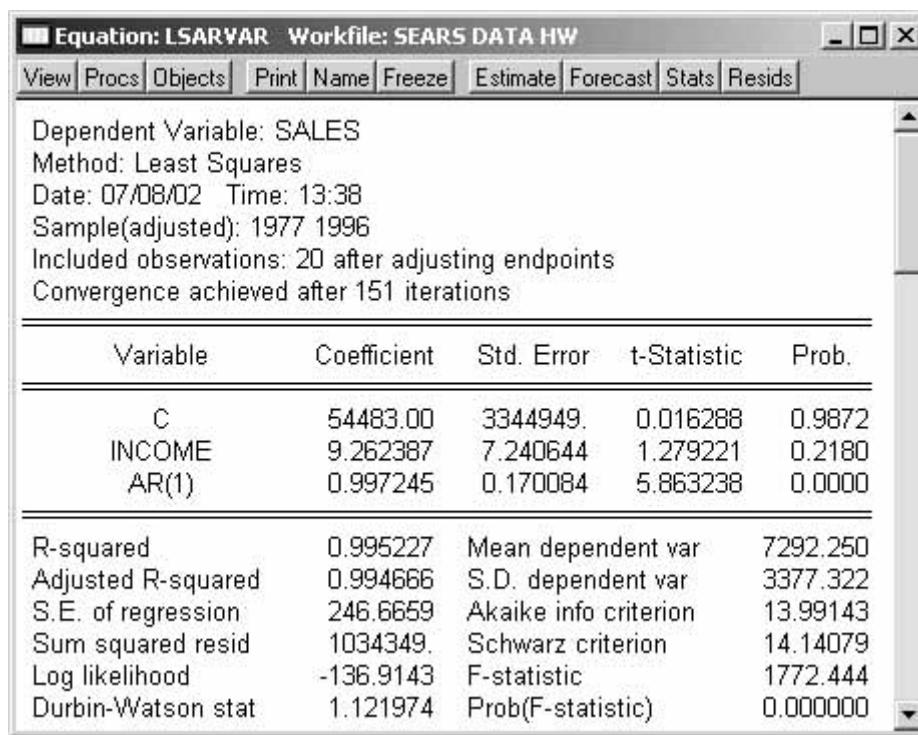
$$s_{b_1} = .319$$

$$t = b_1/s_{b_1} = 44.11$$

$$r^2 = .99$$

$$DW = .63$$

<sup>9</sup> En Pindyck y Rubinfeld (1998) está disponible un análisis de las técnicas para estimar  $\rho$  y considerar una correlación serial.



**FIGURA 8.8 Resultados de E-Views de la regresión de las diferencias generalizadas de las ventas de Sears en función de las diferencias generalizadas del ingreso disponible para el ejemplo 8.5**

La estadística  $DW$  es cercana a cero, lo cual indica la presencia de una autocorrelación positiva.<sup>10</sup> En este punto, se permite la presencia de los errores seriales correlacionados [AR(1)] y se ajusta el modelo que involucra las diferencias generalizadas

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t$$

con  $Y'_t = Y_t - \rho Y_{t-1}$  y  $X'_t = X_t - \rho X_{t-1}$ . Se emplea E-Views para estimar los parámetros directamente en el modelo. El resultado de E-Views se muestra en la figura 8.8. La función de regresión ajustada es

$$\hat{Y}'_t = 54,483(1 - .997) + 9.26X'_t$$

donde

$$\begin{aligned} \hat{\rho} &= .997 \\ \hat{Y}'_t &= Y_t - .997Y_{t-1}, \quad X'_t = X_t - .997X_{t-1} \\ b_0 &= 54,483, \quad b_1 = 9.26 \\ s_{b_1} &= 7.241 \\ t &= b_1/s_{b_1} = 1.28 \\ DW &= 1.12 \end{aligned}$$

El valor de la estadística Durbin-Watson, 1.12, está en la región “no concluyente” en el nivel 0.01. Observe que los estimados del coeficiente de la pendiente en las dos regresiones son se-

<sup>10</sup> Para  $n = 21, k = 1$  y  $\alpha = 0.01, d_L = 0.97$ . Puesto que  $DW = 0.63 < d_L = 0.97$ , se rechaza  $H_0: \rho = 0$  y se elige  $H_1: \rho > 0$ .

mejantes (14.05 y 9.26). Sin embargo, el error estándar asociado con  $b_1$  en la segunda regresión es considerablemente mayor que el correspondiente error estándar en la primera regresión (7.241 contra 0.319). Así, la estadística  $t$  para probar la significancia del coeficiente de la pendiente en la segunda regresión es mucho más pequeña que en la primera (1.28 contra 44.11). De hecho, el valor  $p$  asociado con estadística  $t$  en la segunda regresión es 0.218 lo que indica un coeficiente de la pendiente que no es significativamente distinto de cero. La fuerte correlación serial tiene poco efecto en el estimado del coeficiente de la pendiente en la relación entre  $Y$  y  $X$ . No obstante, la fuerte (positiva) correlación serial resulta en una subestimación severa del error estándar del estimado del coeficiente de la pendiente. Sin duda, una razón para ajustar dicha correlación serial es evitar los errores de inferencia debido a que los valores de  $t$  son demasiado grandes.

Finalmente  $\hat{\rho} = .997$  es muy cercana a 1, lo cual sugiere que la relación entre  $Y$  y  $X$  podría estar representada por un modelo de regresión lineal simple en las diferencias  $Y'_t = Y_t - Y_{t-1}$  y  $X'_t = X_t - X_{t-1}$ . Esta cuestión se explora con mayor detalle en el problema 8.17.

### Modelos autorregresivos

La autocorrelación implica que los valores en la variable dependiente en un periodo están relacionados linealmente con los valores de la variable dependiente en otro periodo. Así, una manera de resolver el problema de la correlación serial es modelar directamente mediante la asociación en donde los distintos períodos de retraso en la variable dependiente se toman como variables independientes o explicativas en un modelo de regresión. Los modelos de regresión formulados de esta forma se conocen como *modelos autorregresivos*. El modelo autorregresivo de primer orden se escribe

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \quad (8.10)$$

donde se considera que los errores  $\varepsilon_t$  tienen propiedades habituales del modelo de regresión. Después de que este modelo ha sido ajustado a los datos por mínimos cuadrados, la ecuación para el pronóstico se vuelve

$$\hat{Y}_t = b_0 + b_1 Y_{t-1} \quad (8.11)$$

Un *modelo autorregresivo* expresa un pronóstico como una función de los valores previos de una serie de tiempo.

Los modelos autorregresivos son un subconjunto de los modelos de promedios móviles integrados autorregresivos (ARIMA) que se analizan con mayor detalle en el capítulo 9.

#### Ejemplo 8.6

Los datos de ventas de la Novak Corporation introducidos en el ejemplo 8.3 y que se muestran en la tabla 8.2 (vea la página 336) se utilizarán para demostrar el desarrollo de un modelo autorregresivo. Observe que luego que las ventas fueron retrasadas como se muestran en la tabla 8.2, se pierde un año de datos porque las ventas de Novak para 1979 se desconocen. El tamaño de la muestra es  $n = 16$  en lugar de  $n = 17$ . Un modelo autorregresivo de primer orden se desarrolla con las ventas de Novak atrasadas por un año ( $Y$ -retrasada) como la variable explicativa.

El resultado de Minitab que se obtiene luego de ajustar el modelo autorregresivo a los gastos de ventas de Novak se muestra en la tabla 8.8. También hay un pronóstico para las ventas de 1997 en un de intervalo de confianza de 95% para la predicción.

El pronóstico de ventas de Novak para 1997 es de \$31 722 000. Esta cifra se calcula dentro de un error redondeado de la siguiente forma:

$$\begin{aligned}\hat{Y} &= b_0 + b_1 Y_{t-1} = -.109 + 1.094 Y_{t-1} \\ \hat{Y}_{18} &= -.109 + 1.094 Y_{18-1} = -.109 + 1.094 Y_{17} \\ \hat{Y}_{18} &= -.109 + 1.094(29.1) = 31.726\end{aligned}$$

**TABLA 8.8 Resultados de Minitab para las ventas de Novak Corporation:  
ajuste del modelo autorregresivo para el ejemplo 8.6**

**Regression Analysis: Sales versus Y-Lagged**

The regression equation is

$$\text{Sales} = -0.109 + 1.09 \text{ Y-Lagged}$$

| Predictor | Coef    | SE Coef | T     | P     |
|-----------|---------|---------|-------|-------|
| Constant  | -0.1093 | 0.3367  | -0.32 | 0.750 |
| Y-Lagged  | 1.09388 | 0.02063 | 53.01 | 0.000 |

$$S = 0.4875 \quad R-Sq = 99.5\% \quad R-Sq(\text{adj}) = 99.5\%$$

**Analysis of Variance**

| Source         | DF | SS     | MS     | F       | P     |
|----------------|----|--------|--------|---------|-------|
| Regression     | 1  | 667.73 | 667.73 | 2810.16 | 0.000 |
| Residual Error | 14 | 3.33   | 0.24   |         |       |
| Total          | 15 | 671.05 |        |         |       |

**Predicted Values for New Observations**

| New Obs | Fit    | SE Fit | 95% CI           | 95% PI           |
|---------|--------|--------|------------------|------------------|
| 1       | 31.722 | 0.311  | (31.055, 32.390) | (30.482, 32.963) |

La prueba de Durbin-Watson no puede utilizarse en este ejemplo. Cuando se incluye una variable dependiente retrasada en la regresión como una variable explicativa, la estadística de Durbin-Watson se inclina hacia el 2. En su lugar, una prueba de correlación serial puede basarse en la estadística  $h$  de Durbin-Watson.<sup>11</sup>

El coeficiente de intercepción en esta regresión es pequeño y no es significativamente diferente de cero. Al correr de nuevo la regresión sin un término de intercepción, se deja esencialmente sin cambio al estimado del coeficiente de la pendiente y a los pronósticos subsecuentes.

Cuando un análisis de regresión se aplica a los datos de una serie de tiempo, con frecuencia los residuales están altamente correlacionados. En ocasiones, el término *correlación serial* se utiliza para referirse esta situación. El análisis de regresión supone que los errores son independientes, lo cual puede causar problemas. La  $R^2$  para una regresión con datos que contienen una correlación serial puede ser artificialmente elevada; es más, los errores estándar de los coeficientes de regresión pueden ser subestimados seriamente y la correspondiente estadística  $t$ , inflada.

Una causa de los residuales autocorrelacionados es la omisión de una o más variables explicativas importantes. Esta omisión significa que una parte importante de la variación de la variable dependiente no ha sido explicada adecuadamente. Una solución para este problema es buscar las variables faltantes para incluirlas en el modelo. Otras soluciones consisten en considerar los modelos de regresión con datos diferenciados, o modelos autorregresivos.

## DATOS DE LAS SERIES DE TIEMPO Y EL PROBLEMA DE LA HETEROSCEDASTICIDAD

La variabilidad en algunas series de tiempo tiende a incrementarse con el nivel de la serie. Ejemplo de esto es la serie de tiempo de las ventas de la Cavanaugh Company que se muestran en la figura 5.1 y se presentan en el ejemplo 5.2. La variabilidad puede in-

<sup>11</sup> En Pindyck y Rubinfeld (1998) se describe una prueba  $h$  para la correlación serial.

crementarse si la variable está creciendo a un ritmo constante, en lugar de una cantidad constante a través del tiempo. A la variabilidad que no es constante se le llama *heteroscedasticidad*.

En un marco de regresión, la heteroscedasticidad se presenta si la varianza del término de error,  $\varepsilon$ , no es constante. Si la variabilidad de períodos recientes es mayor de lo que era para los períodos anteriores, el error estándar del estimado  $s_{y \cdot x'}$ , subestima la desviación estándar actual del término de error. Si la desviación estándar del estimado se utiliza para fijar límites de pronóstico para observaciones futuras, estos límites pueden ser demasiado estrechos para el nivel de confianza establecido.

Algunas veces el problema de la heteroscedasticidad puede resolverse por una simple transformación en los datos. Por ejemplo, en el caso de dos variables, el modelo lineal logarítmico que se mostró la ecuación 8.8 puede utilizarse para reducir la heteroscedasticidad. Asimismo, si las variables se expresan como cantidades en importe, convertir los dólares actuales a dólares constantes (vea el análisis de la deflación de precios en el capítulo 5) podría superar el problema de incrementar la variabilidad del error.

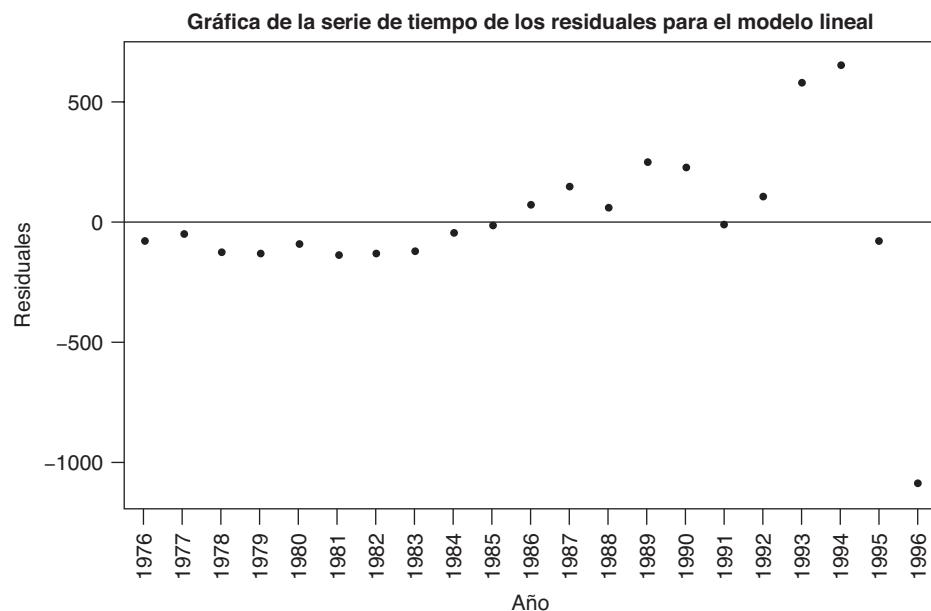
### Ejemplo 8.7

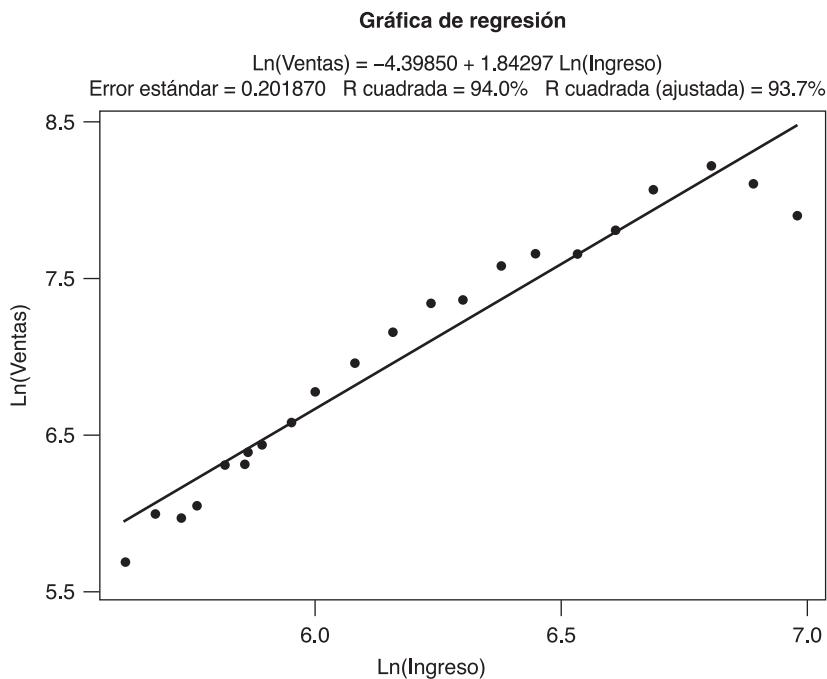
Considere nuevamente los datos de ventas de Reynolds Metals que se presentaron en el ejemplo 8.2 y que se mostraron en la tabla 8.1 (vea la página 333). El resultado de una regresión lineal simple de las ventas en el ingreso personal disponible se proporciona en la figura 8.5. Una gráfica de la secuencia de tiempo de los residuales para esta regresión se muestra en la figura 8.9.

Además de una autocorrelación positiva de los residuales (una serie de residuales seguida de una serie de residuales positivos; vea el análisis del ejemplo 8.2), queda claro, con base en la figura 8.9, que el tamaño de los residuales se incrementa a través del tiempo. Un método para enfrentar a este problema es intentar un modelo lineal logarítmico (ecuación 8.8) para los datos de Reynolds Metals.

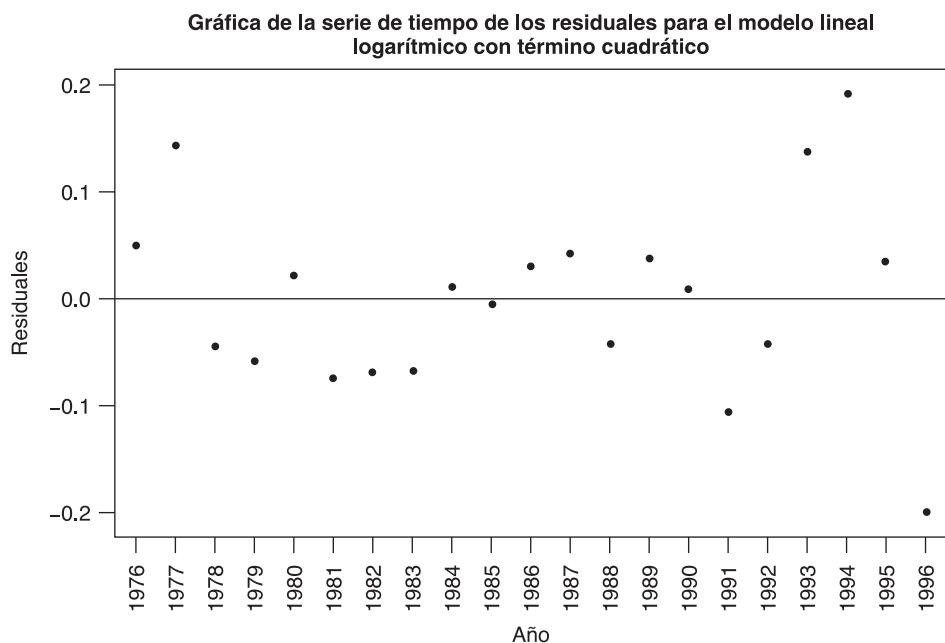
Los resultados del ajuste al modelo lineal logarítmico se proporciona en la figura 8.10. Al comparar la figura 8.10 con la figura 8.5 se puede observar que los residuales (desviaciones de la línea ajustada) para la regresión lineal logarítmica son más uniformes en tamaño a lo largo del

**FIGURA 8.9 Gráfica de secuencia de tiempo de los residuales de la regresión de las ventas en el ingreso disponible: Datos de Reynolds Metals.**





**FIGURA 8.10** Gráfica de la regresión ajustada del modelo lineal logarítmico para los datos de Reynolds Data



**FIGURA 8.11** Gráfica de la secuencia de tiempo de los residuales obtenidos del ajuste del modelo lineal logarítmico con el término cuadrático para los datos de Reynolds Metals

periodo que se estudia, pero la línea recta ajustada no capta la curvatura de los datos. La variable explicativa  $X_2 = X_1^2 = (\ln \text{Ingreso})^2$  se añadió al modelo<sup>12</sup>

$$(\ln \text{Ventas}) = \beta_0 + \beta_1 \ln \text{ingreso} + \beta_2 (\ln \text{ingreso})^2 + \varepsilon$$

se ajustó a los datos. Una gráfica de secuencia de tiempo de los residuales de esta regresión se presenta en la figura 8.11.

En apariencia, los residuales de la figura 8.11 están distribuidos aleatoriamente cerca de cero con una variabilidad constante. Es como si la regresión final representara adecuadamente los datos de Reynolds Metals. Para este modelo no existe razón alguna para dudar de los supuestos acerca del término de error.

## USO DE LA REGRESIÓN PARA PRONOSTICAR DATOS ESTACIONALES

---

Los modelos de descomposición para la serie de tiempo con patrones estacionales se analizaron en el capítulo 5. Enseguida se presenta un modelo de regresión para representar los datos estacionales que están alineados muy cerca de la descomposición aditiva. En este modelo, la estacionalidad se maneja por medio de las *variables ficticias* en la función de regresión.

Un modelo estacional para los datos trimestrales con una tendencia de tiempo es

$$Y_t = \beta_0 + \beta_1 t + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \varepsilon_t \quad (8.12)$$

Donde

$Y_t$  = la variable que se va a pronosticar

$t$  = el índice de tiempo

$S_2$  = una variable ficticia que es 1 para el segundo trimestre del año, y 0 en cualquier otro caso

$S_3$  = una variable ficticia que es 1 para el tercer trimestre del año, y 0 en cualquier otro caso

$S_4$  = una variable ficticia que es 1 para el segundo cuarto del año, y 0 en cualquier otro caso

$\varepsilon_t$  = errores considerados como independientes y normalmente distribuidos con una media de cero y una varianza constante

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ , = coeficientes a ser estimados

Observe que los cuatro niveles de la variable cualitativa (trimestre o estación) se describen únicamente con tres variables ficticias. El último trimestre (el primer trimestre en este caso) se maneja por medio del término de intercepción  $\beta_0$ . Para ver esto, los datos del primer trimestre  $S_2 = S_3 = S_4 = 0$  y el nivel esperado es

$$E(Y_t) = \beta_0 + \beta_1 t$$

Para los datos de segundo trimestre  $S_2 = 1, S_3 = S_4 = 0$  y el nivel esperado es

$$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 = (\beta_0 + \beta_2) + \beta_1 t$$

Patrones similares surgen del tercero y cuarto trimestres.

En el modelo (8.12), trimestres diferentes proporcionan distintos términos de intercepción. La tendencia y el patrón estacional se modelan como series de líneas rectas, una para cada trimestre. Se considera que las cuatro líneas tienen la misma pendiente, pero distintas intercepciones. Los coeficientes de regresión para las variables ficticias representan

---

<sup>12</sup> Un modelo de regresión con las variables explicativas  $X, X^2, X^3, \dots$  se conoce como *modelo de regresión polinomial*.

cambios en la intercepción relativa a la intercepción para el primer trimestre  $\beta_0$ . Si existiera un patrón estacional, pero no así una tendencia, la ecuación 8.12 se aplica con  $\beta_1 = 0$ .

### Ejemplo 8.8

James Brown, encargado de los pronósticos en Washington Water Company, intenta prever el uso de la energía eléctrica de los clientes residenciales para los trimestres tercero y cuarto de 1996. Él sabe que los datos son estacionales y decide utilizar la ecuación 8.12 para desarrollar una ecuación de pronóstico. Entonces, recopila datos trimestrales desde 1980 hasta el segundo trimestre de 1996. Los datos correspondientes al consumo de energía eléctrica, medida en millones de kilowatt-horas, se proporcionan en la tabla 8.9.

James crea variables ficticias  $S_2$ ,  $S_3$  y  $S_4$  que representan los trimestres segundo, tercero y cuarto, respectivamente. Los datos para los cuatro trimestres de 1980 se proporcionan en la tabla 8.10.

**TABLA 8.9 Consumo de energía eléctrica para Washington Water Power Company, 1980-1996, para el ejemplo 8.8**

| <i>Año</i> | <i>Trimestre</i> | <i>Kilowatt-horas (millones)</i> | <i>Año</i> | <i>Trimestre</i> | <i>Kilowatt-horas (millones)</i> |
|------------|------------------|----------------------------------|------------|------------------|----------------------------------|
| 1980       | 1                | 1071                             | 1989       | 1                | 1036                             |
|            | 2                | 648                              |            | 2                | 612                              |
|            | 3                | 480                              |            | 3                | 503                              |
|            | 4                | 746                              |            | 4                | 710                              |
| 1981       | 1                | 965                              | 1990       | 1                | 952                              |
|            | 2                | 661                              |            | 2                | 628                              |
|            | 3                | 501                              |            | 3                | 534                              |
|            | 4                | 768                              |            | 4                | 733                              |
| 1982       | 1                | 1065                             | 1991       | 1                | 1085                             |
|            | 2                | 667                              |            | 2                | 692                              |
|            | 3                | 486                              |            | 3                | 568                              |
|            | 4                | 780                              |            | 4                | 783                              |
| 1983       | 1                | 926                              | 1992       | 1                | 928                              |
|            | 2                | 618                              |            | 2                | 655                              |
|            | 3                | 483                              |            | 3                | 590                              |
|            | 4                | 757                              |            | 4                | 814                              |
| 1984       | 1                | 1047                             | 1993       | 1                | 1018                             |
|            | 2                | 667                              |            | 2                | 670                              |
|            | 3                | 495                              |            | 3                | 566                              |
|            | 4                | 794                              |            | 4                | 811                              |
| 1985       | 1                | 1068                             | 1994       | 1                | 962                              |
|            | 2                | 625                              |            | 2                | 647                              |
|            | 3                | 499                              |            | 3                | 630                              |
|            | 4                | 850                              |            | 4                | 803                              |
| 1986       | 1                | 975                              | 1995       | 1                | 1002                             |
|            | 2                | 623                              |            | 2                | 887                              |
|            | 3                | 496                              |            | 3                | 615                              |
|            | 4                | 728                              |            | 4                | 828                              |
| 1987       | 1                | 933                              | 1996       | 1                | 1003                             |
|            | 2                | 582                              |            | 2                | 706                              |
|            | 3                | 490                              |            |                  |                                  |
|            | 4                | 708                              |            |                  |                                  |
| 1988       | 1                | 953                              |            |                  |                                  |
|            | 2                | 604                              |            |                  |                                  |
|            | 3                | 508                              |            |                  |                                  |
|            | 4                | 708                              |            |                  |                                  |

Fuente: Informe anual de Washington Water Power, varios años.

**TABLA 8.10 Datos para los cuatro trimestres de 1980, del ejemplo 8.8**

| $Y_t$ | $S_2$ | $S_3$ | $S_4$ |
|-------|-------|-------|-------|
| 1071  | 0     | 0     | 0     |
| 648   | 1     | 0     | 0     |
| 480   | 0     | 1     | 0     |
| 746   | 0     | 0     | 1     |

**TABLA 8.11 Resultados en computadora para Washington Water Power, ejemplo 8.8**

The regression equation is  
 $\text{Hours} = 968 + 0.938 \text{ Time} - 342 \text{ 2nd Qt.} - 472 \text{ 3rd Qt.} - 230 \text{ 4th Qt.}$

| Predictor | Coef    | SE Coef | T      | P     |
|-----------|---------|---------|--------|-------|
| Constant  | 968.39  | 16.88   | 57.38  | 0.000 |
| Time      | 0.9383  | 0.3377  | 2.78   | 0.007 |
| 2nd Qt.   | -341.94 | 17.92   | -19.08 | 0.000 |
| 3rd Qt.   | -471.60 | 18.20   | -25.91 | 0.000 |
| 4th Qt.   | -230.23 | 18.20   | -12.65 | 0.000 |

$S = 52.25$        $R-Sq = 92.4\%$        $R-Sq(\text{adj}) = 91.9\%$

Analysis of Variance

| Source         | DF | SS      | MS     | F      | P     |
|----------------|----|---------|--------|--------|-------|
| Regression     | 4  | 2012975 | 503244 | 184.34 | 0.000 |
| Residual Error | 61 | 166526  | 2730   |        |       |
| Total          | 65 | 2179502 |        |        |       |

Durbin-Watson statistic = 1.48

| New Obs | Fit    | SE Fit | 95.0% CI         | 95.0% PI         |
|---------|--------|--------|------------------|------------------|
| 1       | 559.65 | 17.39  | (524.87, 594.43) | (449.54, 669.76) |

Values of Predictors for New Observations

| New Obs | Time | 2nd Qt. | 3rd Qt. | 4th Qt. |
|---------|------|---------|---------|---------|
| 1       | 67.0 | 0       | 1       | 0       |

Los comandos en Minitab para correr el análisis estacional se muestran en la sección “Aplicaciones en Minitab” que aparece al final de este capítulo.

Los resultados se muestran en la tabla 8.11.

El modelo de regresión estacional ajustado es

$$\hat{Y}_t = 968 + .938t - 342S_2 - 472S_3 - 230S_4$$

donde  $S_2$  = segundo trim.,  $S_3$  = tercer trim.,  $S_4$  = cuarto trim. James observa que el modelo explica 92.4% de la variabilidad en el uso de la energía eléctrica en el caso de los consumidores residenciales, medido en kilowatts·hora. El pronóstico para el tercer trimestre de 1996 es

$$\text{Tercer trimestre: Horas} = 968 + .938(67) - 342(0) - 472S_3 - 230S_4$$

El pronóstico para el cuarto trimestre de 1996 es

$$\text{Cuarto trimestre: Horas} = 968 + .938(68) - 342(0) - 472(0) - 230(1) = 802$$

James observa que los pronósticos para los distintos trimestres quedarán situados a lo largo de cuatro líneas rectas. Éstas tendrán la misma pendiente (0.938), pero las intercepciones cambiarán con cada trimestre. Los pronósticos del primer trimestre estarán en una línea con una inter-

cepción de 968. Los pronósticos del segundo trimestre estarán en una línea con intercepciones  $968 - 342 = 626$ . La intercepción para el tercer trimestre será  $968 - 472 = 496$ , y la intercepción para el cuarto trimestre será  $968 - 230 = 738$ . A James le da gusto que el modelo de pronósticos capte el patrón estacional y la tendencia ligeramente ascendente que se puede observar en la serie. En un año determinado, los pronósticos para el uso de electricidad serán los más elevados en el primer trimestre, menores en el segundo, aún más bajos en el tercero y durante el cuarto trimestre serán los segundos más elevados.

## PRONÓSTICO ECONOMÉTRICO

---

Cuando se aplica un análisis de regresión a los datos económicos, las predicciones que se llevan a cabo a partir de dichos modelos se conocen como *pronósticos económicos*. Sin embargo, debido a que con frecuencia la teoría económica sugiere que los valores que toman las cantidades de interés se determinan a través de la interacción simultánea de distintas fuerzas económicas, podría ser necesario modelar esta interacción con un sistema de ecuaciones simultáneas. Esta idea lleva a la construcción de *modelos econométricos de ecuaciones simultáneas*, los cuales involucran ecuaciones individuales del tipo de ecuaciones de regresión. Además, en un sistema simultáneo las ecuaciones individuales están acopladas y el modelo econométrico permite la *determinación conjunta* de una *serie de variables dependientes* en términos de diversas variables independientes. Esto contrasta con la situación habitual de regresión en la que una sola ecuación determina el valor esperado de una variable dependiente en términos de las variables independientes.

Un modelo econométrico de ecuaciones simultáneas determina de forma conjunta los valores de un conjunto de variables dependientes, conocidas por los economistas como *variables endógenas*, y en términos de los valores de variables independientes a las que se nombran *variables exógenas*. Se considera que los valores de las variables exógenas ejercen influencia en las variables endógenas, pero no al revés. Un modelo completo de ecuaciones simultáneas implica el mismo número de ecuaciones que las variables endógenas.

En un sistema econométrico la simultaneidad genera algunos problemas que requieren un tratamiento estadístico especial. Un tratamiento completo de los modelos econométricos está fuera del alcance de este libro.<sup>13</sup> Sin embargo, un modelo de dos ecuaciones mostrará algunos de los conceptos.

La teoría económica establece que, en equilibrio, la cantidad de la oferta es igual a la cantidad demandada en un precio particular. Es decir, la cantidad que se demanda, la cantidad que se oferta y el precio se determinan de forma simultánea. En un estudio de la elasticidad de la demanda respecto a los precios, el modelo se especificó como:

$$\begin{aligned} Q_t &= \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 T_t + \varepsilon_t \\ P_t &= \beta_0 + \beta_1 Q_t + \beta_2 L_t + v_t \end{aligned}$$

donde

$Q_t$  = medida de la demanda (cantidad vendida)

$P_t$  = medida del precio (dólares deflactados)

$I_t$  = medida del ingreso *per cápita*

$T_t$  = medida de la temperatura

$L_t$  = medida del costo por mano de obra

$\varepsilon_t, v_t$  = términos de error independientes o que no están correlacionados entre sí

Observe que en este modelo, el precio y las variables de cantidad  $P_t$  y  $Q_t$  aparecen en ambas ecuaciones. En la primera ecuación la cantidad vendida se determina parcialmente

---

<sup>13</sup> Un recuento introductorio de los modelos econométricos de ecuaciones simultáneas está disponible en Pindyck y Rubinfeld (1998).

por el precio, y en la segunda el precio se determina parcialmente por la cantidad vendida. El precio y la cantidad son variables endógenas, cuyos valores se determinan dentro del sistema. Las variables restantes, ingreso y temperatura en la primera ecuación y costo por mano de obra en la segunda, son variables exógenas. Dados los estimados adecuados para los coeficientes en el modelo (el problema de la identificación), pueden generarse los pronósticos de, por ejemplo, la demanda futura (ventas). Por supuesto, para estimar la demanda futura es necesario especificar desde afuera del sistema los valores futuros de las variables exógenas. Además, los valores futuros de la variable del precio deberán determinarse.

Los modelos econométricos a gran escala se emplean hoy día para modelar el comportamiento de empresas específicas dentro de una industria, de ciertas industrias dentro de toda la economía o el comportamiento de la economía total. Los modelos econométricos pueden incluir cualquier cantidad de ecuaciones simultáneas múltiples parecidas a las de la regresión. Los modelos econométricos se utilizan para comprender cómo funciona la economía y para generar pronósticos de las variables económicas claves. Este tipo de modelos es importante para la formulación de políticas económicas.

## **APLICACIÓN A LA ADMINISTRACIÓN**

---

Las aplicaciones descritas en el capítulo 5 también son apropiadas para esta sección. Las técnicas descritas en este capítulo permiten que el analista detecte y corrija el problema de la correlación serial y, por lo tanto, desarrolle mejores modelos de pronósticos. El resultado neto es que los administradores o los economistas pueden manejar una mayor variedad de datos dependientes del tiempo y confiar en que las predicciones son sólidas. Algunas áreas donde estas técnicas son particularmente útiles incluyen las siguientes:

- Pronóstico de ventas
- Proyecciones de precios de acciones y bonos
- Proyecciones del costo de la materia prima
- Proyecciones de penetración de nuevos productos
- Estimados de personal requerido
- Estudios de la relación entre publicidad y ventas
- Control de inventarios

Puesto que estas aplicaciones involucran variables que evolucionan con el tiempo, esas variables tienen la probabilidad de estar autocorrelacionadas. Los modelos de pronósticos que están basados en las técnicas presentadas en este capítulo deberán proporcionar pronósticos más confiables que algunas técnicas consideradas antes, las cuales ignoran la autocorrelación.

## Glosario

**Autocorrelación (correlación serial).** La autocorrelación existe cuando las observaciones sucesivas que se presentan en el tiempo están relacionadas entre sí.

**Modelo autorregresivo.** Un modelo autorregresivo expresa un pronóstico con una función de valores previos de la serie de tiempo.

**Prueba de Durbin-Watson.** Un tipo de prueba utilizada para determinar si está presente la autocorrelación.

## Fórmulas clave

### **Modelo de regresión lineal simple**

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (8.1)$$

**Correlación serial de primer orden**

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \quad (8.2)$$

**Estadística de Durbin-Watson**

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (8.3)$$

**Relación de la estadística de Durbin-Watson con la autocorrelación de los residuales de retraso 1 ( $n$  grande)**

$$DW = 2(1 - r_1(e)) \quad (8.4)$$

**Modelo de regresión lineal simple transformado**

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t \quad (8.5)$$

**Diferencias generalizadas**

$$\begin{aligned} Y'_t &= Y_t - \rho Y_{t-1} \\ X'_t &= X_t - \rho X_{t-1} \end{aligned} \quad (8.6)$$

**Diferencias simples o primeras diferencias**

$$\begin{aligned} Y'_t &= Y_t - Y_{t-1} \\ X'_t &= X_t - X_{t-1} \end{aligned} \quad (8.7)$$

**Modelo logarítmico de regresión lineal**

$$\ln Y_t = \beta_0 + \beta_1 \ln X_t + \varepsilon_t \quad (8.8)$$

**Ecuación de pronósticos para la forma diferenciada del modelo logarítmico de regresión lineal**

$$\ln \hat{Y}'_t = \ln \hat{Y}'_{t-1} + b_1 (\ln X_t - \ln X_{t-1}) \quad (8.9)$$

**Modelo autorregresivo de primer orden**

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \quad (8.10)$$

**Ecuación de pronóstico para el modelo autorregresivo de primer orden**

$$\hat{Y}_t = b_0 + b_1 Y_{t-1} \quad (8.11)$$

**Modelo estacional con variables ficticias para los datos trimestrales**

$$Y_t = \beta_0 + \beta_1 t + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \varepsilon_t \quad (8.12)$$

**Coefficientes estandarizados (vea el caso 8.2)**

$$B = b \left( \frac{s_x}{s_y} \right) \quad (8.13)$$

**Problemas**

1. ¿Por qué la correlación serial es un problema cuando se analizan los datos de la serie de tiempo?
2. ¿Cuál es la causa más frecuente de la correlación serial?
3. ¿Qué supuesto implícito de la regresión se incumple con mayor frecuencia cuando se analizan las variables de serie de tiempo?
4. ¿Qué estadística se utiliza comúnmente para detectar la correlación serial?
5. Usted realiza una prueba para saber si hay alguna correlación serial en el nivel 0.01 con 32 residuales de una regresión con dos variables independientes. Si la estadística Durbin-Watson que se ha calculado es igual a 1.0, ¿cuál es su conclusión?
6. Usted realiza una prueba para saber si hay alguna correlación serial en el nivel 0.05 con 65 residuales de una regresión con dos variables independientes. Si la estadística Durbin-Watson que se ha calculado es igual a 1.6, ¿cuál es su conclusión?
7. Sugiera formas para resolver el problema de la correlación serial.
8. ¿Cómo funciona el modelo autorregresivo?
9. Tamson Russell, un economista que trabaja para el gobierno, intenta determinar la demanda para el combustible de los motores de los vehículos de pasajeros en Estados Unidos. Tamson desarrolló un modelo que utilizó el precio verdadero de un galón de gasolina regular para predecir el consumo anual de combustible. Con esto, ella pudo explicar 83.5% de la variación en el consumo de combustible con su modelo. Tamson ha decidido añadir a su modelo una variable que represente la población de Estados Unidos. Determine si la correlación serial sería un problema. Los datos se presentan en la tabla P.9.
10. A la organización Decision Science Associates se le ha pedido que realice un estudio de factibilidad para una propuesta del destino turístico que habrá de localizarse a media milla de la presa Grand Coulee. Mark Craze no está satisfecho con el modelo de regresión que se utilizó para fijar el precio de un galón de gasolina y con éste predecir el número de visitantes al centro de visitantes de la presa Grand Coulee. Después de graficar los datos en el diagrama de dispersión, Mark decide utilizar una variable ficticia para representar las celebraciones importantes en el área general. Mark utiliza el 1 para indicar una celebración y el 0 para indicar que no hay celebración. Observe que el 1, en 1974, representa la Expo Feria Mundial 74 que se llevó a cabo en Spokane, Washington; el 1, en 1983, representa la celebración del quincuagésimo aniversario de la construcción de la presa Grand Coulee; el 1, en 1986, representa la Feria Mundial que se celebró en Vancouver, Canadá. Mark también decide utilizar el tiempo como una variable explicativa. Los datos se muestran en la tabla P.10. Escriba un informe para que Mark lo presente a su jefe. Indique si la correlación serial es un problema. También señale qué información adicional podría ser importante para decidir si es recomendable construir el destino turístico.
11. Jim Jackson, un analista de precios para la Washington Water Power Company, está preparando una tarifa y necesita pronosticar los ingresos provenientes de los consumidores de energía eléctrica residenciales para el año 1996. Jim decide investigar tres variables explicativas posibles: el uso residencial por kilowatt-hora (kWh), el cargo residencial por kWh (centavos/kWh) y el número de clientes residenciales que consumen electricidad. Él recopila los datos del periodo de 1968 a 1995 y éstos se muestran en la tabla P.11. Jim compareció previamente ante la comisión de tarifas de Idaho

TABLA P.9

| Año  | Combustible<br>consumido<br>por automóviles<br>(miles de millones<br>de galones)<br>Y | Precio de      | Población<br>estadounidense<br>(millones) |
|------|---------------------------------------------------------------------------------------|----------------|-------------------------------------------|
|      |                                                                                       | X <sub>1</sub> |                                           |
| 1973 | 78.8                                                                                  | .39            | 211.9                                     |
| 1974 | 75.1                                                                                  | .53            | 213.9                                     |
| 1975 | 76.4                                                                                  | .57            | 216.0                                     |
| 1976 | 79.7                                                                                  | .59            | 218.0                                     |
| 1977 | 80.4                                                                                  | .62            | 220.2                                     |
| 1978 | 81.7                                                                                  | .63            | 222.6                                     |
| 1979 | 77.1                                                                                  | .86            | 225.1                                     |
| 1980 | 71.9                                                                                  | 1.19           | 227.7                                     |
| 1981 | 71.0                                                                                  | 1.33           | 230.1                                     |
| 1982 | 70.1                                                                                  | 1.26           | 232.5                                     |
| 1983 | 69.9                                                                                  | 1.22           | 234.8                                     |
| 1984 | 68.7                                                                                  | 1.21           | 236.3                                     |
| 1985 | 69.3                                                                                  | 1.16           | 238.5                                     |
| 1986 | 71.4                                                                                  | .92            | 240.7                                     |
| 1987 | 70.6                                                                                  | .95            | 242.8                                     |
| 1988 | 71.7                                                                                  | .95            | 245.1                                     |

Fuente: *Statistical Abstract of the United States*, varios años.

TABLA P.10

| Año  | Número<br>de visitantes | Tiempo         | Precio de      | Celebración |
|------|-------------------------|----------------|----------------|-------------|
|      | Y                       | Y <sub>1</sub> | X <sub>2</sub> |             |
| 1973 | 268,528                 | 1              | .39            | 0           |
| 1974 | 468,136                 | 2              | .53            | 1           |
| 1975 | 390,129                 | 3              | .57            | 0           |
| 1976 | 300,140                 | 4              | .59            | 0           |
| 1977 | 271,140                 | 5              | .62            | 0           |
| 1978 | 282,752                 | 6              | .63            | 0           |
| 1979 | 244,006                 | 7              | .86            | 0           |
| 1980 | 161,524                 | 8              | 1.19           | 0           |
| 1981 | 277,134                 | 9              | 1.31           | 0           |
| 1982 | 382,343                 | 10             | 1.22           | 0           |
| 1983 | 617,737                 | 11             | 1.16           | 1           |
| 1984 | 453,881                 | 12             | 1.13           | 0           |
| 1985 | 471,417                 | 13             | .86            | 0           |
| 1986 | 654,147                 | 14             | .90            | 1           |

Fuente: Grand Coulee Dam Visitors Center y *Statistical Abstract of the United States*, 1988.

donde se le pidió que explicara si la correlación serial era un problema. Él no supo qué responder y le ha pedido a usted que escriba la respuesta para la pregunta que realiza la comisión.

12. Paul Raymond, presidente de Washington Water Power Company, está preocupado por una eventual operación de absorción de acciones comunes con el fin de tomar el control de la empresa y de que en efecto, el número de accionistas comunes esté dis-

TABLA P.11

| Año  | Ingresos<br>(millones) | Uso por<br>kWh | Cargo<br>(centavos /kWh) | Número<br>de clientes |
|------|------------------------|----------------|--------------------------|-----------------------|
|      | Y                      | X <sub>1</sub> | X <sub>2</sub>           | X <sub>3</sub>        |
| 1968 | 19.3                   | 10,413         | 1.33                     | 139,881               |
| 1969 | 20.4                   | 11,129         | 1.29                     | 142,806               |
| 1970 | 20.9                   | 11,361         | 1.25                     | 146,616               |
| 1971 | 21.9                   | 11,960         | 1.21                     | 151,640               |
| 1972 | 23.4                   | 12,498         | 1.19                     | 157,205               |
| 1973 | 24.5                   | 12,667         | 1.19                     | 162,328               |
| 1974 | 25.8                   | 12,857         | 1.21                     | 166,558               |
| 1975 | 30.5                   | 13,843         | 1.29                     | 170,317               |
| 1976 | 33.3                   | 14,223         | 1.33                     | 175,536               |
| 1977 | 37.2                   | 14,427         | 1.42                     | 181,553               |
| 1978 | 42.5                   | 14,878         | 1.52                     | 188,325               |
| 1979 | 48.8                   | 15,763         | 1.59                     | 194,237               |
| 1980 | 55.4                   | 15,130         | 1.84                     | 198,847               |
| 1981 | 64.3                   | 14,697         | 2.17                     | 201,465               |
| 1982 | 78.9                   | 15,221         | 2.55                     | 203,444               |
| 1983 | 86.5                   | 14,166         | 2.97                     | 205,533               |
| 1984 | 114.6                  | 14,854         | 3.70                     | 208,574               |
| 1985 | 129.7                  | 14,997         | 4.10                     | 210,811               |
| 1986 | 126.1                  | 13,674         | 4.34                     | 212,865               |
| 1987 | 132.0                  | 13,062         | 4.71                     | 214,479               |
| 1988 | 138.1                  | 13,284         | 4.82                     | 215,610               |
| 1989 | 141.2                  | 13,531         | 4.81                     | 217,164               |
| 1990 | 143.7                  | 13,589         | 4.81                     | 219,968               |
| 1991 | 149.2                  | 13,800         | 4.84                     | 223,364               |
| 1992 | 146.1                  | 13,287         | 4.83                     | 227,575               |
| 1993 | 153.9                  | 13,406         | 4.91                     | 233,795               |
| 1994 | 146.9                  | 12,661         | 4.84                     | 239,733               |
| 1995 | 156.8                  | 12,434         | 4.98                     | 253,364               |

Fuente: "Financial and Operating Supplement", Informe anual de Washington Water Power, varios años.

minuyendo desde 1983. Él le instruye a usted para que estudie el número de accionistas comunes desde 1968 y realice un pronóstico para 1996. Usted decide investigar tres variables explicativas potenciales: ganancias por acción (comunes), dividendos por acción (comunes) y dividendos pagados, y recopila los datos de 1968 a 1995, como se muestran en la tabla P.12.

- a) Corra estos datos en la computadora y descubra el mejor modelo de predicción.
  - b) ¿La correlación serial es un problema en este modelo?
  - c) Si la correlación serial representa un problema, escriba un informe para Paul en el que se analicen las diversas soluciones para el problema de autocorrelación y que incluya su recomendación final.
13. Thompson Airlines ha determinado que 5% del número total de pasajeros estadounidenses que utilizan líneas aéreas viaja por Thompson. Se le asigna la tarea de pronosticar el número de pasajeros que volarán por Thompson Airlines en 2004. Los datos se presentan en la tabla P.13.
- a) Desarrolle un modelo de regresión de serie de tiempo; use el tiempo como variable independiente y el número de pasajeros como variable dependiente.
  - b) ¿Están dispersos aleatoriamente los términos de error de este modelo?

TABLA P.12

| <i>Año</i> | <i>Y</i> | <i>Accionistas comunes</i> | <i>Ganancias por acción</i> | <i>Dividendos por acción</i> | <i>Pago de dividendos</i> |
|------------|----------|----------------------------|-----------------------------|------------------------------|---------------------------|
|            |          |                            | <i>X</i> <sub>1</sub>       | <i>X</i> <sub>2</sub>        | <i>X</i> <sub>3</sub>     |
| 1968       | 26,472   | 1.68                       | 1.21                        | 72                           |                           |
| 1969       | 28,770   | 1.70                       | 1.28                        | 73                           |                           |
| 1970       | 29,681   | 1.80                       | 1.32                        | 73                           |                           |
| 1971       | 30,481   | 1.86                       | 1.36                        | 72                           |                           |
| 1972       | 30,111   | 1.96                       | 1.39                        | 71                           |                           |
| 1973       | 31,052   | 2.02                       | 1.44                        | 71                           |                           |
| 1974       | 30,845   | 2.11                       | 1.49                        | 71                           |                           |
| 1975       | 32,012   | 2.42                       | 1.53                        | 63                           |                           |
| 1976       | 32,846   | 2.79                       | 1.65                        | 55                           |                           |
| 1977       | 32,909   | 2.38                       | 1.76                        | 74                           |                           |
| 1978       | 34,593   | 2.95                       | 1.94                        | 61                           |                           |
| 1979       | 34,359   | 2.78                       | 2.08                        | 75                           |                           |
| 1980       | 36,161   | 2.33                       | 2.16                        | 93                           |                           |
| 1981       | 38,892   | 3.29                       | 2.28                        | 69                           |                           |
| 1982       | 46,278   | 3.17                       | 2.40                        | 76                           |                           |
| 1983       | 47,672   | 3.02                       | 2.48                        | 82                           |                           |
| 1984       | 45,462   | 2.46                       | 2.48                        | 101                          |                           |
| 1985       | 45,599   | 3.03                       | 2.48                        | 82                           |                           |
| 1986       | 41,368   | 2.06                       | 2.48                        | 120                          |                           |
| 1987       | 38,686   | 2.31                       | 2.48                        | 107                          |                           |
| 1988       | 37,072   | 2.54                       | 2.48                        | 98                           |                           |
| 1989       | 36,968   | 2.70                       | 2.48                        | 92                           |                           |
| 1990       | 34,348   | 3.46                       | 2.48                        | 72                           |                           |
| 1991       | 34,058   | 2.68                       | 2.48                        | 93                           |                           |
| 1992       | 34,375   | 2.74                       | 2.48                        | 91                           |                           |
| 1993       | 33,968   | 2.88                       | 2.48                        | 86                           |                           |
| 1994       | 34,120   | 2.56                       | 2.48                        | 97                           |                           |
| 1995       | 33,138   | 2.82                       | 2.48                        | 88                           |                           |

Fuente: "Financial and Operating Supplement", *Informe Anual de Washington Water Power*, varios años

- c) Transforme la variable del número de pasajeros de manera que los términos de error se dispersen aleatoriamente.
  - d) Corra un programa de computadora para el modelo transformado en el inciso c.
  - e) Si los términos de error son dependientes, ¿qué problemas implica el uso de este modelo?
  - f) Pronostique el número de pasajeros de Thompson Airlines para 2004.
14. Thomas Furniture Company concluye que la producción programada puede mejorarse al desarrollar un método preciso para predecir las ventas trimestrales. El señor Estes, analista de la compañía, decide investigar la relación existente entre los permisos de construcción de casas y las ventas de muebles en el área de Springfield. Estes cree que los permisos guían las ventas por uno o dos trimestres. Además, se pregunta si las estaciones afectan las ventas de los muebles. Estes decide considerar otra variable independiente:

$$X_2 = \begin{cases} 0 & \text{para las ventas en el primero o segundo trimestres} \\ 1 & \text{para las ventas en el tercero o cuarto trimestres} \end{cases}$$

TABLA P.13

| Año  | Número<br>de pasajeros<br>(miles) | | | | | | |
|---|---|---|---|---|---|---|---|
| 1979 | 22.8                              | 1986 | 45.4                              | 1993 | 79.9                              | 2000 | 144.8                             |
| 1980 | 26.1                              | 1987 | 46.3                              | 1994 | 96.3                              | 2001 | 147.9                             |
| 1981 | 29.4                              | 1988 | 45.8                              | 1995 | 109.0                             | 2002 | 150.1                             |
| 1982 | 34.5                              | 1989 | 48.0                              | 1996 | 116.0                             | 2003 | 151.9                             |
| 1983 | 37.6                              | 1990 | 54.6                              | 1997 | 117.2                             |      |                                   |
| 1984 | 40.3                              | 1991 | 61.9                              | 1998 | 124.9                             |      |                                   |
| 1985 | 39.5                              | 1992 | 69.9                              | 1999 | 136.6                             |      |                                   |

TABLA P.14

| Año  | Trimestre | Ventas<br>(miles)<br>Y | Permisos<br>$X_1$ | Año  | Trimestre | Ventas<br>(miles)<br>Y | Permisos<br>$X_1$ |
|------|-----------|------------------------|-------------------|------|-----------|------------------------|-------------------|
| 1999 | 3         |                        | 19                | 2002 | 1         | 120                    | 72                |
|      | 4         |                        | 3                 |      | 2         | 150                    | 31                |
|      | 1         | 120                    | 35                |      | 3         | 660                    | 19                |
|      | 2         | 80                     | 11                |      | 4         | 270                    | 14                |
|      | 3         | 400                    | 11                |      | 1         | 200                    | 75                |
|      | 4         | 200                    | 16                |      | 2         | 280                    | 41                |
|      | 1         | 75                     | 32                |      | 3         | 800                    | 17                |
|      | 2         | 120                    | 10                |      | 4         | 320                    | 10                |
| 2001 | 3         | 270                    | 12                | 2004 |           |                        |                   |
|      | 4         | 155                    | 21                |      |           |                        |                   |

Los datos se proporcionan en la tabla P.14.

- a) Desarrolle un modelo de regresión que utilice los permisos de construcción de casas como variable explicativa.
  - b) Pruebe la presencia de la autocorrelación en este modelo.
  - c) Desarrolle un modelo de regresión que utilice los permisos y la estacionalidad ficticia como variables explicativas.
  - d) ¿Hay algún patrón de estacionalidad significativo en estos datos? (Pruebe al nivel 0.05.)
  - e) ¿Hay algún problema de autocorrelación con el modelo de regresión múltiple desarrollado en el inciso c? Si lo hay, ¿cómo puede corregirse?
  - f) Por medio del modelo del inciso c, pronostique las ventas para Thomas Furniture Company para los cuatro trimestres de 2004.
15. National Presto es fabricante de aparatos electrodomésticos pequeños, tales como ollas de presión, calentadores, abrelatas, sartenes, parrillas, rosticerías, freidoras, máquinas para hacer palomitas de maíz, cafeteras, rebanadoras, batidoras, etc. Sus ventas trimestrales en millones de dólares, se muestran en la tabla P.15. Presto tiene un negocio principalmente navideño, en donde existe un fuerte efecto estacional. Desarrolle un modelo de regresión múltiple mediante variables ficticias para pronosticar las ventas para el tercero y cuarto trimestres de 1996. Redacte un informe con un resumen de sus resultados.
  16. Los datos en la tabla P.16 muestran las ventas trimestrales ajustadas estacionalmente para Dickson Corporation y para la industria completa por un periodo de 20 trimestres.
    - a) Ajuste un modelo de regresión lineal, obtenga los residuales y grafíquelos en contra el tiempo. ¿Qué descubrió?

TABLA P.15

| Año  | Trimestre |      |      |      |
|------|-----------|------|------|------|
|      | 1         | 2    | 3    | 4    |
| 1985 | 16.3      | 17.7 | 28.1 | 34.3 |
| 1986 | 17.3      | 16.7 | 32.2 | 42.3 |
| 1987 | 17.4      | 16.9 | 30.9 | 36.5 |
| 1988 | 17.5      | 16.5 | 28.6 | 45.5 |
| 1989 | 24.3      | 24.2 | 33.8 | 45.2 |
| 1990 | 20.6      | 18.7 | 28.1 | 59.6 |
| 1991 | 19.5      | 22.5 | 38.3 | 81.2 |
| 1992 | 24.9      | 17.5 | 26.8 | 59.1 |
| 1993 | 22.4      | 14.3 | 24.7 | 57.2 |
| 1994 | 16.2      | 16.5 | 35.5 | 59.8 |
| 1995 | 18.0      | 15.9 | 28.0 | 57.3 |
| 1996 | 17.1      | 17.0 |      |      |

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1996), pág. 32.

TABLA P.16

| Año  | Trimestre | Ventas de Dickinson (miles) |                | Ventas de la industria (millones) |                | Año   | Trimestre | Ventas de Dickinson (miles) |                | Ventas de la industria (millones) |  |
|------|-----------|-----------------------------|----------------|-----------------------------------|----------------|-------|-----------|-----------------------------|----------------|-----------------------------------|--|
|      |           | Y                           | X <sub>2</sub> | X <sub>2</sub>                    | X <sub>1</sub> |       |           | Y                           | X <sub>1</sub> | X <sub>1</sub>                    |  |
| 1998 | 1         | 83.8                        | 31.8           | 2000                              | 3              | 98.2  | 37.1      |                             |                |                                   |  |
|      | 2         | 85.6                        | 32.5           |                                   | 4              | 97.2  | 36.6      |                             |                |                                   |  |
|      | 3         | 87.8                        | 33.2           |                                   | 1              | 100.1 | 37.6      |                             |                |                                   |  |
|      | 4         | 86.1                        | 32.4           |                                   | 2              | 102.6 | 38.3      |                             |                |                                   |  |
| 1999 | 1         | 89.6                        | 33.8           | 2002                              | 3              | 105.4 | 39.3      |                             |                |                                   |  |
|      | 2         | 91.0                        | 34.3           |                                   | 4              | 107.9 | 40.2      |                             |                |                                   |  |
|      | 3         | 93.9                        | 35.3           |                                   | 1              | 110.1 | 41.1      |                             |                |                                   |  |
|      | 4         | 94.6                        | 35.7           |                                   | 2              | 111.1 | 41.4      |                             |                |                                   |  |
| 2000 | 1         | 96.4                        | 36.4           |                                   | 3              | 110.1 | 41.1      |                             |                |                                   |  |
|      | 2         | 96.0                        | 36.3           |                                   | 4              | 111.1 | 41.4      |                             |                |                                   |  |

- b) Calcule la estadística de Durbin-Watson y determine si existe alguna autocorrelación.
- c) Estime el coeficiente de regresión  $\beta_1$  por medio del uso de diferencias generalizadas. (Estime  $\rho$  con el retraso 1 del coeficiente de autocorrelación residual.)
- d) Compare los errores estándar de los dos estimados de  $\beta_1$  por medio de los datos originales y las diferencias generalizadas. ¿Cuál estimado es más preciso? Explique su respuesta.
17. Refiérase al ejemplo 8.5. Al usar los datos de Sears que aparecen en la tabla 8.5, convierta las ventas y los valores de ingresos disponibles a diferencias simples. Es decir, crear los números  $Y_t' = Y_t - Y_{t-1}$  y  $X_t' = X_t - X_{t-1}$ . Ajuste un modelo de regresión lineal simple. Compare sus resultados con los resultados obtenidos por el método de diferencias generalizadas en el ejemplo 8.5. ¿Esperaba que fueran distintos? Explique su respuesta.

TABLA P.18

| Año  | Ahorros personales |    | Ingreso personal |    | Año | Ahorros personales |    | Ingreso personal |   | Año | Ahorros personales |   | Ingreso personal |   |
|------|--------------------|----|------------------|----|-----|--------------------|----|------------------|---|-----|--------------------|---|------------------|---|
|      | Y                  | X  | Y                | X  |     | Y                  | X  | Y                | X |     | Y                  | X | Y                | X |
| 1935 | 2                  | 60 | 1942             | 28 | 123 | 1949               | 9  | 207              |   |     |                    |   |                  |   |
| 1936 | 4                  | 69 | 1943             | 33 | 151 | 1950               | 13 | 279              |   |     |                    |   |                  |   |
| 1937 | 4                  | 74 | 1944             | 37 | 165 | 1951               | 18 | 257              |   |     |                    |   |                  |   |
| 1938 | 1                  | 68 | 1945             | 30 | 171 | 1952               | 19 | 273              |   |     |                    |   |                  |   |
| 1939 | 3                  | 73 | 1946             | 15 | 179 | 1953               | 20 | 288              |   |     |                    |   |                  |   |
| 1940 | 4                  | 78 | 1947             | 7  | 191 | 1954               | 19 | 290              |   |     |                    |   |                  |   |
| 1941 | 11                 | 96 | 1948             | 13 | 210 |                    |    |                  |   |     |                    |   |                  |   |

18. Se realiza un estudio en un intento por relacionar los ahorros personales con el ingreso personal (en miles de millones de dólares) para el periodo de 1935 a 1954. Los datos se proporcionan en la tabla P.18.
- Ajuste un modelo de regresión lineal simple a los datos de la tabla P.18 haciendo uso del ingreso personal para predecir los ahorros personales. De manera específica: (1) Realice una prueba de significancia para el coeficiente de la pendiente de regresión en el nivel  $\alpha = 0.01$ ; (2) realice una prueba de significancia para la regresión por medio de la prueba  $F$  (en  $\alpha = 0.01$ ); (3) calcule  $r^2$  e interprete esta cantidad; y (4) realice una prueba de autocorrelación (en  $\alpha = 0.05$ ). ¿Modificaría sus conclusiones de las partes 1 y 2? ¿Cómo se podría mejorar el modelo?
  - Construya una variable ficticia  $X_2$  para los años de guerra. Sea  $X_2 = 0$  para los tiempos de paz y  $X_2 = 1$  para el periodo de guerra. Los años de guerra son de 1941 a 1945. Ajuste un modelo de regresión lineal simple usando el ingreso personal y la variable ficticia de los años de guerra como variables explicativas. Evalúe los resultados. De manera específica: (1) evalúe si el conocimiento de los años de guerra aporta una contribución significativa a la predicción de los ahorros personales, más allá de la proporcionada por el ingreso personal (sea  $\alpha = 0.01$ ); y (2) evalúe la presencia de la autocorrelación. ¿Es mejor el modelo de regresión lineal simple que el inciso a)? Comente su respuesta.
19. Circuit City Inc. es un distribuidor de equipo de audio y video, y otros artículos electrónicos y de oficina de uso común. Recientemente, las ventas han sido débiles, con un descenso total de 5% en diciembre. Entre las razones se encuentra una competencia feroz entre los vendedores de productos electrónicos, la deflación del precio en muchos productos, una caída en el tráfico comercial y una disminuida demanda por productos de cómputo. Jim Lowe tiene la tarea de pronosticar las ventas para el año fiscal 2003. Jim tiene acceso a los estimados proporcionados por *The Value Line Investment Survey* (vea la tabla P.19); sin embargo, él teme que sean demasiado optimistas.
- ¿Cuáles modelos de pronósticos debe investigar Jim?
  - ¿Sería apropiado un modelo autorregresivo?
  - ¿Cómo determinaría Jim si la correlación serial representa un problema para un modelo autorregresivo?
  - Desarrolle un modelo autorregresivo para Jim.
  - De qué manera los estimados para 2003 de este modelo se comparan con los de *Value Line*?
  - Puede usted decir si la correlación serial es un problema en este modelo?

TABLA P.19

| Año fiscal | May 31                  | Ago 31                  | Nov 30                  | Feb 28                  |
|------------|-------------------------|-------------------------|-------------------------|-------------------------|
| 1996       | 1742                    | 1806                    | 1999                    | 2897 <sup>b</sup>       |
| 1997       | 1851                    | 1948                    | 2039                    | 3156                    |
| 1998       | 2078                    | 2369                    | 2222                    | 3289                    |
| 1999       | 2205                    | 2423                    | 2495                    | 3476                    |
| 2000       | 2449                    | 2566                    | 2336                    | 3177                    |
| 2001       | 1882                    | 2037                    | 2280                    | 3391                    |
| 2002       | 2118                    | 2221                    | 2422                    | 3239                    |
| 2003       | <b>2150<sup>a</sup></b> | <b>2350<sup>a</sup></b> | <b>2600<sup>a</sup></b> | <b>3400<sup>a</sup></b> |

<sup>a</sup>Estimados de Value Line para 2003.<sup>b</sup>El año fiscal termina el último día de febrero del siguiente año calendario.Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 2002), pág. 1725.

## CASOS

### CASO 8-1 COMPAÑÍA DE SU ELECCIÓN

La salud de una compañía puede examinarse cada mes, trimestre o año con mediciones que cuentan con gran cantidad de variables. Es posible que para cada una de éstas existan otras variables que proporcionen información sobre su conducta y que pueden emplearse como variables explicativas en una ecuación de pronóstico.

El propósito de este caso es simular la identificación de una variable de serie de tiempo importante para la compañía de su elección y, enseguida, analizar los patrones que existen en los datos por medio de un análisis de autocorrelación. Además, es posible hacer uso del programa de computadora adecuado para desarrollar una ecuación que pueda emplearse para prever valores futuros de su variable de la serie de tiempo.

#### Tarea

- Identifique una compañía u organización que sea de su interés. Ésta puede ser local o nacional y que haya publicado registros, incluyendo las mediciones de las variables de las series de tiempo. Identifique una variable importante de la empresa elegida y registre sus valores para varios años, trimestres o meses.

- Ya sea a mano o por medio de una computadora (vea la aplicación CB Predictor en el capítulo 9), calcule diversos coeficientes de autocorrelación y grafique la función de autocorrelación.
- Con base en el patrón de la función de autocorrelación, describa los patrones de su serie de tiempo.
- Calcule las primeras diferencias de sus datos y construya la función de autocorrelación para los datos diferenciados. Describa los patrones resultantes en la serie de tiempo de las primeras diferencias.
- Identifique las diversas variables explicativas potenciales para su variable dependiente. Usted puede utilizar los registros de la compañía y otras fuentes de datos en este proceso.
- Desarrolle una ecuación de pronóstico para su variable dependiente a través de uno o más variables explicativas identificadas.
- Examine los residuales de su modelo ajustado. En particular, revise la autocorrelación. Una vez que esté satisfecho con su ecuación de pronóstico, genere pronósticos para los siguientes seis períodos. De ser posible, compare sus pronósticos con los valores reales. ■

## CASO 8-2

## ÍNDICE DE ACTIVIDAD EMPRESARIAL EN EL CONDADO DE SPOKANE

Antes de 1973 el condado de Spokane, en Washington, no tenía mediciones actualizadas de la actividad empresarial. No obstante, lo que en general sucede en esta área afecta a cada negocio local, agencia gubernamental o individuo. Los planes y políticas realizados por una unidad económica estarían incompletas si no tuviesen un conocimiento confiable acerca del desempeño reciente de la economía, a la cual pertenece la unidad. Un índice de actividad empresarial para la circunscripción de Spokane debe servir como información vital en la formulación de estrategias y decisiones, tanto para organismos públicos como privados.

Un índice de actividad económica es el indicador de los cambios relativos en las condiciones generales de los negocios existentes en la región. En el plano nacional, el Producto Nacional Bruto (medido por el Departamento de Comercio de Estados Unidos) y el Índice de Producción Industrial (medido por la Federal Reserve Board) por lo general son considerados como excelentes indicadores. Cada una de estas series está basada en miles de piezas de información, así como en la recolección, la edición y el cálculo, los cuales son procesos costosos que consumen tiempo. Para un área local como Spokane, una versión simplificada que sea capaz de proporcionar información razonablemente precisa y actual, a costos moderados, es una opción muy deseable.

Se puede usar la regresión múltiple para construir un índice de actividad empresarial. La construcción de dicho índice debe tratar con tres cuestiones esenciales; a saber:

- ¿Cuáles son los componentes de ese índice?
- ¿En realidad estos componentes representan adecuadamente los cambios de las condiciones generales de los negocios?
- ¿Qué peso debe asignarse a cada uno de los componentes elegidos?

Las respuestas a estas preguntas pueden obtenerse por medio de un análisis de regresión.

El doctor Shik Chun Young, profesor de economía en Eastern Washington University, está intentando desarrollar un índice de actividad empresarial para Spokane. Young selecciona el ingreso personal como la variable dependiente. En el nivel de esa jurisdicción, el ingreso personal se juzga como el mejor indicador disponible de las condiciones locales para los negocios. Los

ingresos personales miden los ingresos totales de las familias, antes de impuestos. Debido a que, habitualmente, las actividades productivas son remuneradas por medios monetarios, sin duda el ingreso personal puede considerarse como un aproximado del desempeño económico general. Entonces, ¿es necesario construir otro índice si el ingreso personal puede funcionar como buen indicador de la actividad empresarial? Desafortunadamente, el Departamento de Comercio estadounidense estima una vez al año los datos del ingreso personal relativos al condado, y sólo después de 16 meses los hace públicos. En consecuencia, estos datos son poco útiles para una planeación de corto plazo. La tarea de Young es organizar un índice de actividad empresarial actualizado.

Las variables independientes se obtienen a partir de los datos locales que ya están disponibles cada mes. Actualmente, alrededor de 50 series de dichos datos mensuales están disponibles y van desde el empleo, las actividades bancarias y las transacciones de bienes raíces hasta el consumo de energía eléctrica. Si cada serie fuera incluida en el análisis de regresión, el esfuerzo proporcionaría una baja productividad debido a que tan sólo un puñado de estas series sería estadísticamente significativo. Por lo tanto, es necesario cierto conocimiento acerca de la relación entre el ingreso personal y los datos disponibles para determinar cuáles variables independientes habrá que incluir en la ecuación de regresión. A partir de lo que el profesor Young conoce acerca de la economía de Spokane, se han seleccionado las siguientes 10 series:

- $X_1$ , empleo total
- $X_2$ , empleo en la manufactura
- $X_3$ , empleo en la construcción
- $X_4$ , empleo en las ventas al mayoreo y al menudeo
- $X_5$ , empleo en los servicios
- $X_6$ , débitos bancarios
- $X_7$ , Demanda de depósitos bancarios
- $X_8$ , licencias de construcción expedidas
- $X_9$ , hipotecas sobre bienes raíces
- $X_{10}$ , consumo total de energía eléctrica

El primer paso en el análisis consiste en ajustar el modelo

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{10} X_{10}$$

donde

$$\begin{aligned} Y &= \text{ingreso personal} \\ \beta_0 &= \text{intercepción con el eje } Y \\ \beta_1, \beta_2 \dots \beta_{10} &= \text{coeficientes de las variables independientes respectivas} \end{aligned}$$

Cuando el modelo precedente se ajusta a los datos  $\bar{R}^2$  es 0.96, lo cual significa que las 10 variables utilizadas explican 96% de la variabilidad en la variable dependiente, el ingreso personal. Sin embargo, otras estadísticas de regresión indican problemas. En primer lugar, de estas diez variables independientes únicamente tres tienen un valor  $t$  significativo en el nivel 0.05; a saber, la del empleo total, la del empleo en servicios y la de los débitos bancarios totales. En segundo lugar, la matriz de correlación muestra un alto grado de interdependencia entre las diversas variables a independientes, multicolinealidad.<sup>14</sup> Por ejemplo, el empleo total y los débitos bancarios tienen coeficientes de correlación de 0.88; el consumo total de energía eléctrica y los depósitos de demanda bancaria, 0.76, y las licencias de construcción expedidas y las hipotecas de bienes raíces, 0.68. En tercer lugar, una prueba de autocorrelación que usa la estadística de Durbin-Watson, de 0.91, indica que los valores sucesivos de la variable dependiente están correlacionados de manera positiva. Desde luego, la autocorrelación es muy común en los datos de la serie de tiempo; en general, las observaciones en estas series tienden relacionarse entre sí.

Debido a que uno de los supuestos básicos en el análisis de regresión es que las observaciones de la variable dependiente son aleatorias, Young prefiere tratar primero con el problema de la autocorrelación. Decide calcular las primeras diferencias, o cambios, en un intento por minimizar la interdependencia entre las observaciones en cada una de las series de tiempo. Ahora, las 10 variables independientes se miden por la diferencia que existe entre los períodos, en lugar de medirse por el valor absoluto de cada período. Para permitir que las series de datos se distingan, se utiliza una nueva designación para las variables independientes.

$\Delta X_1$ , cambio en el empleo total  
 $\Delta X_2$ , cambio en el empleo de la manufactura  
 $\Delta X_3$ , cambio en el empleo de la construcción  
 $\Delta X_4$ , cambio en el empleo de la venta al mayoreo y al menudeo

$$\begin{aligned} \Delta X_5, &\text{cambio en el empleo de los servicios} \\ \Delta X_6, &\text{cambio en los débitos bancarios} \\ \Delta X_7, &\text{cambio en la demanda de depósitos} \\ \Delta X_8, &\text{cambio en las licencias para construcción expedidas} \\ \Delta X_9, &\text{cambio en las hipotecas sobre bienes raíces} \\ \Delta X_{10}, &\text{cambio en el consumo total de la energía eléctrica} \end{aligned}$$

El modelo de regresión se presenta ahora como

$$E(\Delta Y) = \beta_0 + \beta_1 \Delta X_1 + \beta_2 \Delta X_2 + \dots + \beta_{10} \Delta X_{10}$$

donde

$$\begin{aligned} \Delta Y &= \text{cambio en el ingreso personal} \\ \beta_0 &= \text{intercepción con} \\ \beta_1, \beta_2 \dots \beta_{10} &= \text{coeficientes de las respectivas variables independientes} \end{aligned}$$

Una corrida de regresión utilizando este modelo basada en los datos de la primera diferencia produce una estadística Durbin-Watson de 1.71. Esto indica que la autocorrelación ya no es un problema.

El próximo paso es determinar cuál de las 10 variables es una variable explicativa significativa de la variable dependiente. La variable dependiente  $\Delta Y$  se entra en una regresión con las diversas combinaciones posibles en las 10 variables explicativas potenciales para poder seleccionar la mejor ecuación. El criterio que se usa en la selección es:

- Una  $\bar{R}^2$  ajustada altamente satisfactoria
- Coeficientes de correlación bajos entre las variables independientes
- Coeficientes significativos (en el nivel 0.05) para cada una de las variables independientes

Después de un escrutinio cuidadoso de los resultados de la regresión, Young descubre que la ecuación que contiene  $\Delta X_4$ ,  $\Delta X_5$ , y  $\Delta X_{10}$  como variables independientes es la que mejor cubre con el criterio señalado.

Sin embargo, Young razona que (además de los usos comerciales e industriales) el uso total de energía eléctrica incluye el consumo residencial, el cual no debería tener una relación significativa con la actividad empresarial en el corto plazo. Para probar esta hipótesis, Young subdivide el consumo total de energía eléctrica en cuatro variables:

<sup>14</sup> Se debe tener cierto cuidado al interpretar una correlación entre dos variables de series de tiempo, debido a que la autocorrelación en las series individuales puede producir una correlación falsa; vea el ejemplo 8.1.

**TABLA 8.12 Variables de regresión de Young**

| Ecuación | Variables independientes                 | Variabile dependiente |
|----------|------------------------------------------|-----------------------|
| A        | $\Delta X_4, \Delta X_5, \Delta X_{11},$ | $\Delta Y$            |
| B        | $\Delta X_4, \Delta X_5, \Delta X_{12},$ | $\Delta Y$            |
| C        | $\Delta X_4, \Delta X_5, \Delta X_{13},$ | $\Delta Y$            |
| D        | $\Delta X_4, \Delta X_5, \Delta X_{14},$ | $\Delta Y$            |

- $\Delta X_{11}$ , cambio en el consumo residencial de la electricidad  
 $\Delta X_{12}$ , cambio en el consumo comercial de la electricidad  
 $\Delta X_{13}$ , cambio en el consumo industrial de la electricidad  
 $\Delta X_{14}$ , cambio en el consumo comercial e industrial de la electricidad

Las cuatro variables combinadas con  $\Delta X_4$  y  $\Delta X_5$ , se utilizan para producir las cuatro nuevas ecuaciones de regresión (vea la tabla 8.12).

El análisis estadístico indica que la ecuación D en la tabla 8.12 es la mejor. Comparada con la ecuación previa que contiene  $\Delta X_4$ ,  $\Delta X_5$  y  $\Delta X_{10}$ , como variables independientes, la ecuación A es la única que muestra un deterioro de significancia estadística. Este resultado confirma la noción que tenía Young acerca de los usos comerciales industriales como mejores variables explicativas del ingreso personal que el consumo total de energía eléctrica, el cual incluye también el consumo residencial de la electricidad.

Por lo tanto, la ecuación D se selecciona como la ecuación final de regresión y los resultados son

$$\hat{\Delta Y} = -1.86 + 17.10\Delta X_4 + 23.01\Delta X_5 + .007\Delta X_{14}$$

(4.07)                    (5.61)                    (.002)

$$n = 15 \quad F = 26.26$$

$$DW = 1.77 \quad \bar{R}^2 = .835$$

Las cifras que están entre paréntesis y debajo de los coeficientes de regresión son los errores estándar de los coeficientes estimados. Los valores  $t$  de los coeficientes son 4.20, 4.10 y 3.50 para  $\Delta X_4$ ,  $\Delta X_5$  y  $\Delta X_{14}$ , respectivamente. La  $\bar{R}^2$  indica que casi 84% de la varianza en el cambio del ingreso personal se explica por las tres variables independientes. La estadística Durbin-Watson ( $DW$ ) muestra que la autocorrelación no es un problema. Además, la matriz del coeficiente de relación de

**TABLA 8.13 Matriz de correlación**

|                 | $\Delta X_4$ | $\Delta X_5$ | $\Delta X_{14}$ |
|-----------------|--------------|--------------|-----------------|
| $\Delta X_4$    | 1.00         | .45          | .11             |
| $\Delta X_5$    | .45          | 1.00         | .12             |
| $\Delta X_{14}$ | .11          | .12          | 1.00            |

la tabla 8.13 muestra un bajo nivel de interdependencia entre las tres variables dependientes.

Para propósitos de construcción de índices, las variables independientes en la ecuación final de regresión se vuelven componentes del índice. Se pueden determinar los pesos de los componentes por los coeficientes de regresión. (Recuerde que el coeficiente de regresión representa el cambio promedio en la variable dependiente para un incremento de una unidad en la variable independiente.) Sin embargo, debido a que las variables en la ecuación de regresión no se miden por las mismas unidades (por ejemplo,  $\Delta Y$  se mide en miles de dólares, y  $\Delta X_{14}$  se mide en miles de kilowatt-horas), los coeficientes de regresión deben transformarse a valores relativos. Esta transformación se logra mediante el cálculo de los *coeficientes estandarizados*, o *coeficientes B*.

$$B = b \left( \frac{s_x}{s_y} \right)$$

donde

$b$  = el coeficiente de regresión de la variable independiente

$s_x$  = la desviación estándar de la variable independiente

$s_y$  = la desviación estándar de la variable dependiente

Es común que los valores de todas estas estadísticas estén disponibles en los resultados de la regresión proporcionados por la computadora. Por lo tanto, los coeficientes estandarizados en las tres variables independientes son:

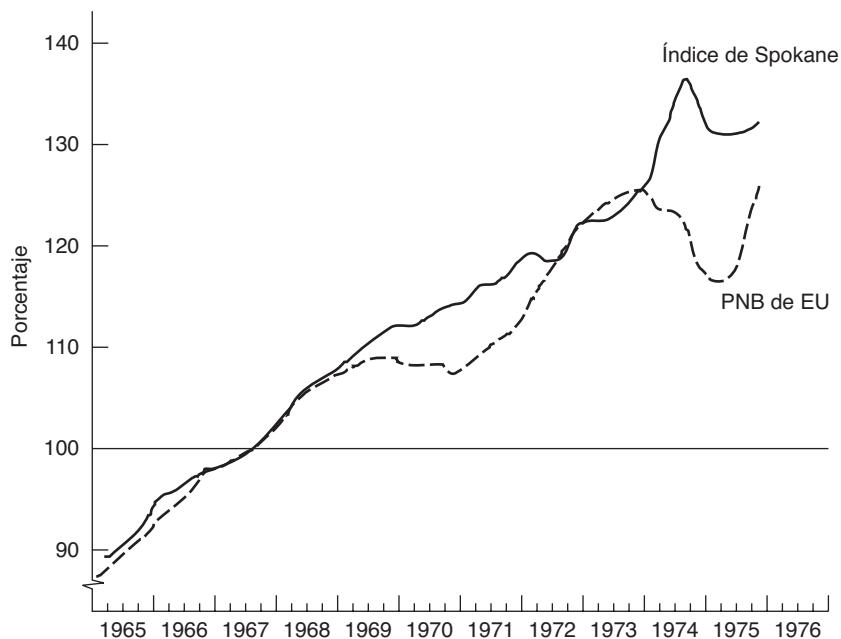
$$B_4 = .4959$$

$$B_5 = .4833$$

$$B_{14} = .3019$$

$$\text{Total} = 1.2811$$

Debido a que la suma de pesos en un índice debe equivaler a 100%, por lo general, los coeficientes estandarizados se normalizan de la manera en que lo indica la tabla 8.14.



**FIGURA 8.12** Índice de actividad empresarial del condado de Spokane y el PNB estadounidense en dólares constantes (1967 = 100).

**TABLA 8.14** Coeficientes estandarizados

| Componente      | Peso                         |
|-----------------|------------------------------|
| $\Delta X_4$    | $\frac{4959}{1\ 2811} .3871$ |
| $\Delta X_5$    | $\frac{4833}{1\ 2811} .3772$ |
| $\Delta X_{14}$ | $\frac{3019}{1\ 2811} .2357$ |
| Total           | 1.000                        |

Después de que los componentes y sus plazos respectivos han sido determinados, los siguientes pasos proporcionan el índice:

1. Calcule el cambio porcentual de cada uno de los componentes a partir del periodo base.
2. Multiplique el cambio de porcentaje por el peso apropiado.
3. Sume los cambios de porcentaje sopesados y que se obtuvieron en el paso 2.

El índice de actividad empresarial de Spokane se comparan con el Producto Nacional Bruto de Estados Unidos en dólares constantes (1967 = 100); esto aparece en la figura 8.12.■

## PREGUNTAS

1. ¿Por qué escogió Young resolver primero el problema de autocorrelación?
2. ¿Habría sido mejor eliminar primero la multicolinealidad y después atacar directamente a la autocorrelación?
3. ¿De qué manera afecta el tamaño de una muestra pequeña al análisis?
4. ¿La regresión hecha en las primeras diferencias debió atravesar el origen?
5. ¿Existe alguna factibilidad para el uso de los datos retrasados?
6. ¿Qué conclusiones pueden obtenerse de una comparación del índice de actividad empresarial del condado de Spokane y del PNB?

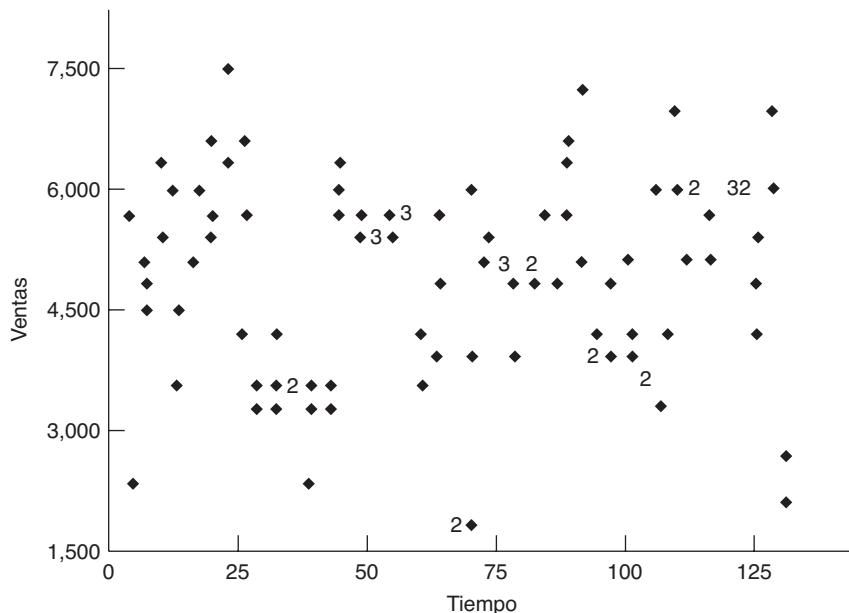


FIGURA 8.12 Ventas del restaurante, enero de 1981 - diciembre de 1982

### CASO 8-3 VENTAS DE UN RESTAURANTE<sup>15</sup>

Jim Price, quien estaba trabajando en la obtención de su grado de maestría en administración, laboraba en un pequeño restaurante cerca de Marquette University, situada en Milwaukee, Wisconsin. Certo día, la gerente del restaurante le pidió a Jim que se presentara en su oficina, pues estaba muy interesada en pronosticar las ventas semanales y quería saber si Jim podría ayudarle. Él acababa de tomar un curso de estadística avanzada y dijo que le encantaría enfrentar el reto.

Jim pidió a la gerente que le proporcionara cualquier registro histórico que tuviera. Ella le indicó que el restaurante había recopilado las ventas de las semanas pasadas cada lunes. Jim comenzó su análisis al obtener los datos de las ventas semanales desde el domingo 1 de enero de 1981 hasta el final de la semana el lunes 29 de diciembre de 1982, lo que sumaba un total de 104 observaciones.

La media de las ventas semanales para las 104 semanas resultó ser de \$4862. La figura 8.13 es una gráfica de las ventas semanales que se presentaron a lo largo de ese periodo. La gráfica indica que las ventas semanales fueron muy volátiles, que iban de \$1870 a \$7548, con una tendencia mínima. Debido a que Jim apenas había completado el curso del análisis de regre-

sión, decidió utilizar las ventas semanales como la variable dependiente y ver si podía encontrar variables independientes, o explicativas, útiles.

Jim evaluó tres elementos de predicción. El primero fue el tiempo. El segundo fue una variable ficticia que indicaba si Marquette University estaba en sesión plena esa semana o no (0 = no había sesión plena; 1 = sí había sesión plena). El examen de los datos de ventas que se presentan en la figura 8.13 reveló que las ventas semanales siempre caían cuando Marquette no tenía sesión plena, principalmente durante las vacaciones de Navidad, las de Semana Santa y las de verano. Jim no estaba sorprendido, el restaurante está localizado cerca del campus de Marquette y la mayor parte de sus clientes son miembros de la comunidad universitaria. El tercer factor de predicción que Jim probó fue el de las ventas atrasadas una semana, ya que el examen de la figura 8.13 indicó que las ventas para dos semanas adyacentes eran frecuentemente similares.

Jim calculó las correlaciones simples entre los tres factores potenciales de predicción y la variable dependiente, las ventas semanales. Los resultados se presentan en la matriz de correlación que se muestra en la tabla 8.15. Como Jim lo esperaba, casi no hubo tenden-

<sup>15</sup> El caso 8.3 fue proporcionado por Frank G. Frost, de Marquette University, Milwaukee, Wisconsin.

**TABLA 8.15 Matriz de la correlación de ventas para el caso 8.3**

|                                 | Ventas actuales | Tiempo | Variable ficticia | Ventas atrasadas por una semana |
|---------------------------------|-----------------|--------|-------------------|---------------------------------|
| Ventas actuales                 | 1.000           | .049   | .772              | .580                            |
| Tiempo                          |                 | 1.000  | .048              | .120                            |
| Variable ficticia               |                 |        | 1.000             | .490                            |
| Ventas atrasadas por una semana |                 |        |                   | 1.000                           |

**TABLA 8.16 Modelos de regresión del restaurante para el caso 8.3**

| Variables explicativas del modelo        | R <sup>2</sup> | Estadística de Durbin-Watson | ¿Autocorrelación significativa en el nivel 0.05? | Evidencia de colinealidad |
|------------------------------------------|----------------|------------------------------|--------------------------------------------------|---------------------------|
| (1) Tiempo                               | .0024          | .81                          | Sí                                               | Ninguna                   |
| (2) Ficticia                             | .5960          | 1.30                         | Sí                                               | Ninguna                   |
| (3) Ventas atrasadas                     | .3360          | 1.89                         |                                                  | Ninguna                   |
| (4) Tiempo y ventas ficiticias           | .6030          | 1.32                         |                                                  | Muy poca                  |
| (5) Tiempo y ventas atrasadas            | .3360          | 1.89                         | No                                               | Poca                      |
| (6) Ventas ficticias y atrasadas         | .6490          | 1.74                         | No                                               | Poca                      |
| (7) Tiempo, ventas ficticias y atrasadas | .6510          | 1.73                         | No                                               | Moderada                  |

cia en las ventas semanales como se indica en el coeficiente de correlación de 0.049. Sin embargo, la variable ficticia estaba correlacionada fuertemente con las ventas actuales  $r = 0.722$ , es decir, ya fuera que Marquette tuviera mucha asistencia o no, tenía un buen potencial para que se le considerara como un elemento confiable de predicción para las ventas semanales. Las ventas actuales fueron relacionadas moderadamente,  $r = 0.580$ , con las ventas retrasadas una semana. No obstante, Jim también observó que la variable ficticia estaba relacionada moderadamente,  $r = 0.490$ , con las ventas atrasadas por una semana.

Jim experimentó con diversos modelos de regresión. Los resultados de los diversos modelos de regresión se presentan en la tabla 8.16. Ya que los datos de las ventas casi no tienen tendencia, el elemento de predicción “tiempo” añade muy poco valor predictivo al modelo de regresión. Observe que el modelo 4 tiene únicamente una  $R^2$  ligeramente más alta que el modelo 2, y que los errores de ambos modelos parecen tener una cantidad significativa de autocorrelación. Los modelos 3 y 5 tienen la misma  $R^2$ ; mientras que el modelo 7 tiene una  $R^2$  ligeramente más alta que la del modelo 6. Por otra parte, el elemento de predicción “ventas atrasadas” añade una cantidad clara de poder de predicción al modelo de regresión. Finalmente, el modelo 6 tiene una  $R^2$  más alta que el modelo 2 sin una cantidad significativa de autocorrelación.

Jim decidió seleccionar el modelo de regresión 6 para pronosticar las ventas semanales, por las siguientes razones:

1. El modelo 6 tuvo la segunda  $R^2$  más grande, únicamente 0.002 por debajo del modelo 7.
2. Los parámetros de modelo 6 fueron significativamente distintos de cero en el nivel 0.01.
3. La autocorrelación no es un problema para el modelo 6.
4. El modelo 6 es más simple que el modelo 7 y no tiene tanta multicolinealidad.

La función de regresión ajustada que Jim utilizó fue:

$$\hat{Y} = 2,614.3 + 1,610.7 (\text{variable ficticia}) \\ + .2605 (\text{ventas retrasadas})$$

$R^2$  significa que 64.9% de la variación en las ventas semanales puede explicarse por el hecho de que Marquette tenga asistencia completa y por las ventas de la semana anterior. La ecuación de regresión implica que el promedio de las ventas semanales es de aproximadamente \$1611 más elevado, cuando Marquette tiene asistencia completa, si se consideran como constantes las ventas de la semana anterior.

Jim quedó satisfecho con su esfuerzo, pero se preguntaba si algún otro tipo de modelo de pronósticos podría resultar más eficaz. Por esta razón, decidió tomar un curso de pronósticos.

**PREGUNTAS**

1. ¿Es significativa la autocorrelación para los modelos 3 y 4? Evalúe en el nivel 0.05.
2. ¿Fue correcto el uso que Jim dio a las variables ficticias?
3. ¿Fue correcto que Jim utilizara las ventas retrasadas con una variable explicativa?

4. ¿Está usted de acuerdo con las conclusiones obtenidas por Jim?
5. ¿Podría otro tipo de modelo de pronósticos para pronosticar las ventas semanales?

**CASO 8-4 MR. TUX**

John Mosby ha corrido un análisis de regresión simple utilizando el tiempo como la variable explicativa; también ha observado un decepcionante  $r^2$  de 0.563 (vea el caso Mr. Tux del capítulo 6). Debido a que sabe que sus datos de ventas tienen un componente estacional importante, decide ajustar un modelo de regresión múltiple estacional.

John generó 11 variables ficticias para el efecto estacional. Debido a que los datos de Mr. Tux son

mensuales, codificó un dólar S1 como 1, si los datos eran enero; 0 si sucedía cualquier otra cosa. John hizo esto para cada mes y terminó con S11, que equivalía a 1 para noviembre y 0 para cualquier otra situación.

John corrió el modelo utilizando las 12 variables explicativas, una para el tiempo y las otras 11 representando el efecto estacional mensual. Los resultados se muestran en la tabla 8.17. El modelo explica 87.7% de

**TABLA 8.17 Resultados de Minitab para el modelo de regresión estacional de Mr. Tux para el caso 8.4**

**Regression Analysis: Sales Versus Time, S1,...**

The regression equation is

$$\begin{aligned} \text{Sales} = & -35023 + 2752 \text{ Time} = 48459 \text{ S1} - 29808 \text{ S2} + 21681 \text{ S3} + 199019 \text{ S4} \\ & + 139212 \text{ S5} + 57713 \text{ S6} + 21689 \text{ S7} + 74014 \text{ S8} + 7872 \text{ S9} - 9009 \text{ S10} \\ & - 25050 \text{ S11} \end{aligned}$$

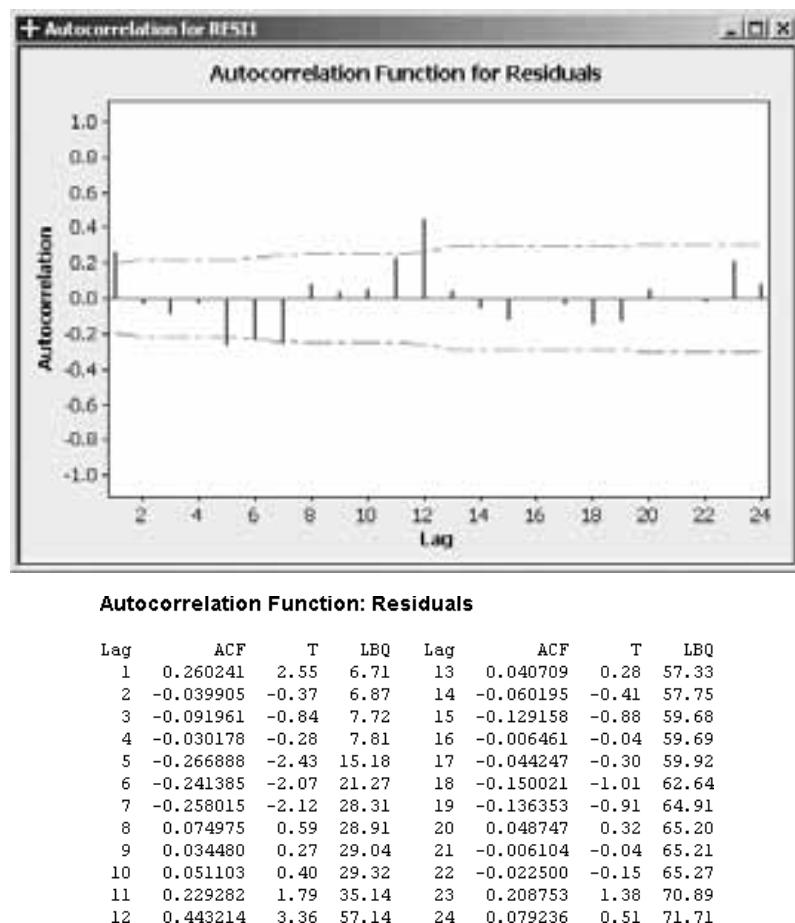
| Predictor | Coef   | SE Coef | T     | P     |
|-----------|--------|---------|-------|-------|
| Constant  | -35023 | 15441   | -2.27 | 0.026 |
| Time      | 2752.4 | 141.0   | 19.52 | 0.000 |
| S1        | 48459  | 19059   | -2.54 | 0.013 |
| S2        | -29808 | 19048   | -1.56 | 0.121 |
| S3        | 21681  | 19039   | 1.14  | 0.258 |
| S4        | 119019 | 19030   | 6.25  | 0.000 |
| S5        | 139212 | 19022   | 7.32  | 0.000 |
| S6        | 57713  | 19015   | 3.04  | 0.003 |
| S7        | 21689  | 19009   | 1.14  | 0.257 |
| S8        | 74014  | 19005   | 3.89  | 0.000 |
| S9        | 7872   | 19001   | 0.41  | 0.680 |
| S10       | -9009  | 18998   | -0.47 | 0.637 |
| S11       | -25050 | 18997   | -1.32 | 0.191 |

S = 37992.3 R-Sq = 87.7% R-Sq(adj) = 85.9%

Analysis of Variance

| Source         | DF | SS          | MS          | F     | P     |
|----------------|----|-------------|-------------|-------|-------|
| Regression     | 12 | 8.55392E+11 | 71282630871 | 49.38 | 0.000 |
| Residual Error | 83 | 1.19804E+11 | 1443416727  |       |       |
| Total          | 95 | 9.75195E+11 |             |       |       |

Durbin-Watson statistic = 1.41



**FIGURA 8.14 Función de autocorrelación para los residuales del modelo de regresión estacional de Mr. Tux**

la varianza de la variable ventas. Sin embargo, un MAPE superior a 20% parece elevado. A continuación, John generó las autocorrelaciones para los residuales del modelo que se muestran en la figura 8.14.

Las autocorrelaciones de los residuales tienen una punta en el retraso 12, el retraso estacional. De alguna manera insatisfecho con los resultados de su modelo de regresión estacional, John probó el modelo autorregresivo para pronosticar sus ventas mensuales (los datos aparecen en el caso de Mr. Tux al final del capítulo 2).

John sabe que estos datos tienen un componente estacional fuerte, de manera que decide intentar modelar este componente utilizando un modelo autorregresivo con los valores  $Y$  retrasados 12 meses. John no

puede utilizar los primeros 12 meses de sus datos de ventas, pero debido que comenzó con 96 períodos, esto aún le deja una muestra de 84. El resultado de Minitab se muestra la tabla 8.18.

### Tarea

Escriba un informe a John con un análisis cuidadoso de los resultados de sus intentos para desarrollar un modelo de pronósticos estacional. ¿Qué modelo es el mejor? Asegúrese de que su análisis incluya una evaluación del ajuste del modelo, una precisión potencial de los pronósticos y cualquier otro problema que pudiera surgir, por ejemplo, la autocorrelación.

**TABLA 8.18 Resultados de Minitab para el modelo autorregresivo estacional de Mr. Tux para el caso 8.4**

```

Regression Analysis: Sales versus Y-Lagged

The regression equation is
Sales = 24786 + 1.07 Y-Lagged

84 cases used, 12 cases contain missing values

Predictor      Coef    SE Coef      T      P
Constant      24786     5322     4.66  0.000
Y-Lagged      1.06807   0.03803   28.08  0.000

S = 30784.9  R-Sq = 90.6%  R-Sq(adj) = 90.5%

Analysis of Variance

Source        DF          SS          MS          F      P
Regression     1  7.47470E+11  7.47470E+11  788.71  0.000
Residual Error 82  77712386803   947712034
Total         83  8.25183E+11

Durbin-Watson statistic = 1.829

```

## CASO 8-5 CONSUMER CREDIT COUNSELING

La operación del Consumer Credit Counseling (CCC) se describió en los capítulos 1 (caso 1.2) y 3 (caso 3.3).

Su director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de nuevos clientes que se atenderían durante el resto de 1993. Marv proporcionó a Dorothy Mercer los datos mensuales del número de nuevos clientes de la CCC en el periodo de enero de 1985 a marzo de 1993 (vea el caso 3.3). En el caso 3.3, Dorothy utilizó el análisis de autocorrelación para explorar los patrones de los datos. En el caso 6.5, trató tanto el número de personas que tenían vales de comida como el índice de la actividad empresarial para desarrollar un modelo de regresión para pronosticar el resto de 1993.

Dorothy estaba insatisfecha con los resultados de su modelo de regresión. Decidió intentar la regresión múltiple y le pidió a Marv que pensara en otras variables que pudieran estar relacionadas con el número de clientes vistos. Marv indicó que podría intentar con el número de quiebras registradas y la cantidad de licencias de construcción emitidas.

Los datos del número de bancarrotas registradas de enero de 1986 a diciembre de 1992 se muestran en la tabla 8.19, y el número de licencias de construcción emitidas de enero de 1986 a diciembre de 1992 se muestran en la tabla 8.20.

Dorothy desarrolló un modelo de regresión múltiple que utilizaba el número de personas con vales de comida, el índice de actividad empresarial, el número de quiebras registradas y el número de licencias de construcción emitidas. También creó un modelo basado únicamente en el supuesto de que los datos eran estacionales.

Recientemente, Dorothy recibió información acerca de que la correlación serial con frecuencia es un problema cuando la regresión se desarrolla con datos de series de tiempo. Está preocupada porque algunos modelos de regresión, que se desarrollaron para predecir el número de clientes por atender, fueran afectados por este problema.

Ya que le gustó el informe que usted le presentó en el cual aplicó un análisis de descomposición de las series de tiempo, ella le ha asignado la labor de revisar esta situación.

**TABLA 8.19 Quiebras de enero de 1986 a diciembre de 1992**

|      |     |     |     |     |     |     |     |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1986 | 160 | 170 | 140 | 173 | 162 | 160 | 150 | 145 | 134 | 155 | 145 | 152 |
| 1987 | 171 | 206 | 173 | 195 | 165 | 177 | 168 | 165 | 131 | 169 | 166 | 157 |
| 1988 | 174 | 162 | 196 | 178 | 169 | 170 | 143 | 177 | 192 | 195 | 164 | 146 |
| 1989 | 180 | 149 | 200 | 165 | 168 | 177 | 143 | 180 | 169 | 170 | 160 | 161 |
| 1990 | 172 | 146 | 185 | 143 | 156 | 173 | 140 | 160 | 131 | 169 | 152 | 135 |
| 1991 | 136 | 167 | 179 | 181 | 166 | 151 | 165 | 129 | 132 | 162 | 140 | 140 |
| 1992 | 130 | 165 | 172 | 145 | 129 | 166 | 146 | 143 | 127 | 186 | 157 | 94  |

**TABLA 8.20 Licencias de construcción emitidas entre enero de 1986 y diciembre de 1992**

|      |     |     |     |     |     |     |     |     |     |     |     |     |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1986 | 49  | 60  | 149 | 214 | 191 | 193 | 161 | 174 | 168 | 203 | 111 | 64  |
| 1987 | 44  | 76  | 137 | 145 | 150 | 168 | 168 | 170 | 158 | 146 | 104 | 48  |
| 1988 | 32  | 32  | 127 | 149 | 128 | 178 | 132 | 164 | 152 | 92  | 121 | 77  |
| 1989 | 61  | 48  | 108 | 161 | 200 | 204 | 164 | 175 | 200 | 171 | 153 | 79  |
| 1990 | 111 | 92  | 166 | 189 | 245 | 261 | 207 | 238 | 198 | 247 | 164 | 108 |
| 1991 | 67  | 118 | 205 | 253 | 296 | 300 | 284 | 282 | 282 | 261 | 237 | 130 |
| 1992 | 133 | 210 | 298 | 334 | 312 | 333 | 311 | 327 |     | 387 | 243 | 140 |

Usted recuerda vagamente que, en cierto momento de su experiencia educativa, desarrolló un modelo autorregresivo y le pregunta a Dorothy si le gustaría que usted evaluara esa posibilidad. Ella decide que ésta sería una buena oportunidad para que usted ocupara su tiempo.

#### Tarea

Analice la significancia de las variables en el modelo de regresión de Dorothy. Desarrolle un modelo de regresión (asegúrese de incluir variables ficticias aditivas para el componente estacional, si fuera necesario) y utilícelo para pronosticar el número de nuevos clientes

para los primeros tres meses de 1993. Compare sus pronósticos con las observaciones reales.

Desarrolle un modelo autorregresivo y genere pronósticos para los primeros tres meses de 1993. ¿Qué modelo (regresión múltiple o autorregresión) cree usted que es el mejor candidato para generar los pronósticos para el resto de 1993?

Escriba un informe a Dorothy para remitirle la información que le ha solicitado acerca del problema de la correlación serial. Incluya un análisis de los resultados de sus esfuerzos para desarrollar un modelo apropiado que permita pronosticar el número de nuevos clientes para el resto de 1993.■

## CASO 8-6 AAA WASHINGTON<sup>16</sup>

En el caso 5.5 se proporcionó una visión general de AAA Washington, cuando a los alumnos se les solicitó que elaboraran una descomposición de series de tiempo para las llamadas que recibe el club para el servicio de emergencia en carretera durante un periodo de cinco años. La descomposición de la serie de tiempo realizada en el caso 5.5 mostró el patrón que Michael DeCoria había observado en el volumen de llamadas para el servicio de emergencia en carretera que probablemente fuera cíclico por naturaleza. A Michael le gustaría ser capaz de predecir el efecto cíclico en el volumen de llamadas del servicio de emergencia en carretera para años futuros.

Otra investigación realizada por el club descubrió diversos factores que repercuten en el volumen de las llamadas del servicio de emergencia en carretera. Entre estos factores están la temperatura diaria promedio y la cantidad de lluvia recibida en un día. La investigación ha mostrado que las llamadas para el servicio de emergencia en carretera se incrementan cuando crece el volumen de lluvia y a medida que se reduce la temperatura promedio diaria. El club también cree que el número total de llamadas del servicio de emergencia en carretera que recibe depende del número de miembros asociados.

<sup>16</sup> Este caso fue proporcionado por Steve Branton, antiguo estudiante y graduado de MBA en la Eastern Washington University.

Michael ha observado que la tendencia cíclica de la serie de tiempos parece atrasarse respecto del ciclo económico general. Asimismo, ha sugerido que el desempleo en el estado de Washington podría ser una medida sustituta para la economía del estado de Washington. Los datos acerca de la temperatura mensual promedio, la lluvia mensual y la tasa de desempleo en esa entidad federativa tendrían que recopilarse y se presentan en la tabla 8.21. Una conversación sostenida con el gerente del centro de llamadas del servicio de emergencia en carretera ha llevado a dos observaciones importantes: (1) parece que los fabricantes de automóviles diseñan vehículos que funcionan mejor a una temperatura de 65° Fahrenheit, y (2) en apariencia, el volumen de llamadas aumenta más rápido cuando la temperatura cae por debajo de 30° que cuando el descenso está por debajo de 60 grados. Esta información sugiere que el efecto de la temperatura en el servicio de emergencia en carretera no es lineal.

En el caso 6.6 se investigaron cuatro modelos de regresión lineal que usaron el número total de llamadas al servicio de emergencia en carretera como variable dependiente, y la tasa de desempleo, la temperatura, la

lluvia y el número de socios del club como cuatro variables independientes. La variable de la temperatura fue transformada al sustraer 65° de los valores de la temperatura mensual promedio. Entonces, se investigó una relación no lineal.

En el caso 7.2 se desarrolló un modelo de regresión múltiple. Se evaluaron variables como la lluvia, el número de socios, la temperatura mensual promedio transformada exponencialmente y la tasa de desempleo retrasada 11 meses.

### Tarea

1. ¿Cuál modelo de regresión es mejor para la predicción? ¿Los signos de coeficientes para las variables independientes son los que usted esperaba? ¿Los coeficientes de las variables independientes son significativamente distintos a cero?
2. ¿La correlación serial es un problema?
3. Prepare un informe para el señor DeCoria en donde usted recomiende un modelo de regresión que considere como el más apropiado para predecir la naturaleza cíclica del volumen de llamadas para el servicio de emergencia en carretera.■

**TABLA 8.21 Datos de AAA Washington para el caso 8.6**

| Año  | Mes        | Llamadas | Tasa   | Temperatura | Lluvia | Socios  |
|------|------------|----------|--------|-------------|--------|---------|
| 1987 | Junio      | —        | 6.9940 | —           | —      | —       |
|      | Julio      | —        | 6.8137 | —           | —      | —       |
|      | Agosto     | —        | 6.3648 | —           | —      | —       |
|      | Septiembre | —        | 6.5435 | —           | —      | —       |
|      | Octubre    | —        | 6.7826 | —           | —      | —       |
|      | Noviembre  | —        | 6.9225 | —           | —      | —       |
|      | Diciembre  | —        | 7.1560 | —           | —      | —       |
| 1988 | Enero      | —        | 7.9143 | —           | —      | —       |
|      | Febrero    | —        | 7.7576 | —           | —      | —       |
|      | Marzo      | —        | 7.0641 | —           | —      | —       |
|      | Abril      | —        | 6.0977 | —           | —      | —       |
|      | Mayo       | 20,002   | 5.7867 | 55.1        | 3.75   | —       |
|      | Junio      | 21,591   | 5.7592 | 59.0        | 1.95   | —       |
|      | Julio      | 22,696   | 5.5718 | 63.8        | 0.89   | —       |
| 1989 | Agosto     | 21,509   | 5.2939 | 63.8        | 0.51   | 384,746 |
|      | Septiembre | 22,123   | 5.4709 | 59.1        | 2.31   | 388,652 |
|      | Octubre    | 21,449   | 5.5049 | 54.6        | 3.12   | 392,241 |
|      | Noviembre  | 23,475   | 5.8630 | 45.4        | 8.42   | 393,115 |
|      | Diciembre  | 23,529   | 6.1349 | 41.0        | 4.44   | 392,631 |
|      | Enero      | 23,327   | 7.5474 | 40.3        | 4.30   | 396,975 |
|      | Febrero    | 24,050   | 7.8157 | 34.3        | 3.18   | 395,186 |
|      | Marzo      | 24,010   | 7.1390 | 43.2        | 6.57   | 397,791 |
|      | Abril      | 19,735   | 6.2637 | 52.5        | 2.39   | 397,764 |
|      | Mayo       | 20,153   | 5.8332 | 55.3        | 2.83   | 399,348 |
|      | Junio      | 19,512   | 5.8077 | 62.4        | 1.30   | 401,949 |

TABLA 8.21 (Continúa)

| <i>Año</i> | <i>Mes</i> | <i>Llamadas</i> | <i>Tasa</i> | <i>Temperatura</i> | <i>Lluvia</i> | <i>Socios</i> |
|------------|------------|-----------------|-------------|--------------------|---------------|---------------|
| 1990       | Julio      | 19,892          | 5.6713      | 62.9               | 0.83          | 404,866       |
|            | Agosto     | 20,326          | 5.4977      | 63.5               | 1.53          | 405,341       |
|            | Septiembre | 19,378          | 5.2989      | 60.9               | 0.32          | 407,479       |
|            | Octubre    | 21,263          | 5.6028      | 51.9               | 3.44          | 405,430       |
|            | Noviembre  | 21,443          | 5.9143      | 46.2               | 7.24          | 412,134       |
|            | Diciembre  | 23,366          | 5.0000      | 41.8               | 4.72          | 415,342       |
|            | Enero      | 23,836          | 6.1917      | 41.8               | 9.55          | 416,255       |
|            | Febrero    | 23,336          | 6.3775      | 38.9               | 5.73          | 423,001       |
|            | Marzo      | 22,003          | 5.7234      | 46.3               | 3.40          | 428,559       |
|            | Abril      | 20,155          | 4.7792      | 51.7               | 2.91          | 431,429       |
|            | Mayo       | 20,070          | 4.5715      | 54.9               | 2.15          | 434,675       |
|            | Junio      | 19,588          | 4.3899      | 59.8               | 3.55          | 435,864       |
| 1991       | Julio      | 20,804          | 4.2559      | 66.7               | 0.59          | 437,969       |
|            | Agosto     | 19,644          | 3.9359      | 66.4               | 1.33          | 440,565       |
|            | Septiembre | 17,424          | 3.9048      | 61.9               | 0.24          | 441,936       |
|            | Octubre    | 20,833          | 4.4294      | 50.4               | 1.17          | 448,595       |
|            | Noviembre  | 22,490          | 5.1523      | 45.8               | 10.66         | 446,291       |
|            | Diciembre  | 24,861          | 5.5102      | 33.9               | 7.93          | 446,455       |
|            | Enero      | 23,441          | 6.8901      | 37.9               | 4.40          | 445,392       |
|            | Febrero    | 19,205          | 7.0308      | 46.9               | 5.42          | 445,787       |
|            | Marzo      | 20,386          | 6.7186      | 43.4               | 4.35          | 445,746       |
|            | Abril      | 19,988          | 6.1280      | 49.1               | 5.69          | 446,430       |
|            | Mayo       | 19,077          | 5.8146      | 54.3               | 2.12          | 450,001       |
|            | Junio      | 19,141          | 5.9480      | 58.2               | 1.61          | 452,303       |
| 1992       | Julio      | 20,883          | 5.9026      | 65.4               | 0.51          | 456,551       |
|            | Agosto     | 20,709          | 5.7227      | 66.0               | 2.80          | 455,747       |
|            | Septiembre | 19,647          | 5.6877      | 60.9               | 0.20          | 456,764       |
|            | Octubre    | 22,013          | 6.2922      | 51.0               | 1.70          | 462,340       |
|            | Noviembre  | 22,375          | 7.0615      | 46.2               | 6.50          | 460,492       |
|            | Diciembre  | 22,727          | 7.4370      | 42.4               | 3.45          | 465,361       |
|            | Enero      | 22,367          | 8.4513      | 43.0               | 7.26          | 465,492       |
|            | Febrero    | 21,155          | 8.7699      | 46.0               | 3.59          | 466,775       |
|            | Marzo      | 21,209          | 8.0728      | 48.9               | 1.47          | 467,168       |
|            | Abril      | 19,286          | 7.2392      | 52.7               | 4.35          | 464,575       |
|            | Mayo       | 19,725          | 7.0461      | 58.3               | 0.60          | 459,019       |
|            | Junio      | 20,276          | 7.0478      | 63.6               | 1.84          | 463,665       |
| 1993       | Julio      | 20,795          | 7.1080      | 64.9               | 1.41          | 463,775       |
|            | Agosto     | 21,126          | 6.7824      | 65.0               | 1.01          | 466,230       |
|            | Septiembre | 20,251          | 6.7691      | 58.4               | 2.16          | —             |
|            | Octubre    | 22,069          | 7.5896      | 53.2               | 2.55          | —             |
|            | Noviembre  | 23,268          | 7.9908      | 44.8               | 6.23          | —             |
|            | Diciembre  | 26,039          | 8.2460      | 37.8               | 4.38          | —             |
|            | Enero      | 26,127          | 9.5301      | 34.9               | 4.08          | —             |
|            | Febrero    | 20,067          | 9.2790      | —                  | —             | —             |
|            | Marzo      | 19,673          | 8.6802      | —                  | —             | —             |
|            | Abril      | 19,142          | 7.7815      | —                  | —             | —             |
|            | Mayo       | —               | 7.4338      | —                  | —             | —             |
|            | Junio      | —               | 7.3701      | —                  | —             | —             |
|            | Julio      | —               | 7.2442      | —                  | —             | —             |

— = Datos no disponibles.

## CASO 8-7 ALOMEGA FOOD STORES

En el ejemplo 1.1 Julie Ruth reunió los datos de ventas mensuales de su empresa, junto con otras variables que ella pensó que podrían estar relacionadas con las ventas. Sus esfuerzos para llevar a cabo los pronósticos mediante estos datos, junto con Minitab, se detallan en los casos que aparecen al final de los capítulos 2, 3 y 5. Su deseo por mejorar la precisión de sus pronósticos sugirió el uso de la regresión múltiple.

A partir de su análisis de descomposición de las ventas, Julie reconoció que sus datos tenían un componente estacional fuerte (vea el caso 5.6). Debido a que creyó que las ventas estaban influidas por las cantidades gastadas en publicidad de periódicos y televisión (vea el caso 2.3), decidió correr una regresión que relacionaba las ventas con la publicidad en periódicos, televisión y 11 variables ficticias 0-1, siendo diciembre el

mes base, para considerar los efectos estacionales mensuales. Algunos resultados obtenidos en Minitab, a partir de la regresión de Julie, se muestran en la tabla 8.22.

Julie quedó muy satisfecha al observar un valor  $R^2$  de aproximadamente 91%, una mejoría significativa sobre el valor cuadrado  $r$  para el modelo de regresión que contiene la variable explicativa única TV (vea los resultados de regresión en el caso 2.3). Además, la varianza del factor inflación (*VIF*, por sus siglas en inglés) para cada una de las variables explicativas fue pequeño y la estadística Durbin-Watson quedó próxima a 2.

Julie se dio cuenta de que un examen de los residuales será necesario antes de que pueda estar completamente satisfecha con sus resultados; además, intentó realizar un análisis completo de residuales antes de pro-

**TABLA 8.22 Resultados en Minitab para el modelo de regresión de Julie Ruth para las ventas de Alomega**

**Regression Analysis: Sales versus Paper, TV,...**

The regression equation is

$$\begin{aligned} \text{Sales} = & 184393 + 0.363 \text{ Paper} + 0.315 \text{ TV} + 200847 \text{ Dum1} + 55491 \text{ Dum2} \\ & + 199556 \text{ Dum3} + 100151 \text{ Dum4} + 190293 \text{ Dum5} + 135441 \text{ Dum6} \\ & + 156609 \text{ Dum7} + 51586 \text{ Dum8} + 183619 \text{ Dum9} + 109096 \text{ Dum10} \\ & + 96206 \text{ Dum11} \end{aligned}$$

| Predictor | Coef    | SE Coef | T    | P     | VIF |
|-----------|---------|---------|------|-------|-----|
| Constant  | 184393  | 23402   | 7.88 | 0.000 |     |
| Paper     | 0.36319 | 0.06854 | 5.30 | 0.000 | 2.1 |
| TV        | 0.31530 | 0.03638 | 8.67 | 0.000 | 1.6 |
| Dum1      | 200847  | 39151   | 5.13 | 0.000 | 2.7 |
| Dum2      | 55491   | 32399   | 1.71 | 0.096 | 1.8 |
| Dum3      | 199556  | 34147   | 5.84 | 0.000 | 2.0 |
| Dum4      | 100151  | 32388   | 3.09 | 0.004 | 1.8 |
| Dum5      | 190293  | 32822   | 5.80 | 0.000 | 1.9 |
| Dum6      | 135441  | 32581   | 4.16 | 0.000 | 1.9 |
| Dum7      | 156609  | 32699   | 4.79 | 0.000 | 1.9 |
| Dum8      | 51586   | 32420   | 1.59 | 0.121 | 1.8 |
| Dum9      | 183619  | 36522   | 5.03 | 0.000 | 2.3 |
| Dum10     | 109096  | 32439   | 3.36 | 0.002 | 1.8 |
| Dum11     | 96206   | 32417   | 2.97 | 0.005 | 1.8 |

S = 45801.4 R-Sq = 90.8% R-Sq(adj) = 87.3%

**Analysis of Variance**

| Source         | DF | SS          | MS          | F     | P     |
|----------------|----|-------------|-------------|-------|-------|
| Regression     | 13 | 7.06412E+11 | 54339402344 | 25.90 | 0.000 |
| Residual Error | 34 | 71324031968 | 2097765646  |       |       |
| Total          | 47 | 7.77736E+11 |             |       |       |

Durbin-Watson statistic = 2.274

poner su modelo de regresión como una herramienta útil para pronosticar las ventas.

A Julie le ha gustado el hecho de que los elementos de predicción “periódicos” y “televisión” quedarán bajo el control de la compañía. Los valores de estas variables podrían elegirse con anticipación para generar pronósticos de ventas futuras.

Julie está casi lista para confrontar a Jackson Tilson, pero aún tiene algunos cabos que atar. ■

## PREGUNTAS

1. Julie ha recopilado datos y otras variables que no estarán incluidas en su modelo de regresión múltiple. ¿Deberían incluirse en su modelo una o más de esas distintas variables? De manera más general, ¿cómo puede asegurarse Julie de que tiene el conjunto “correcto” de variables explicativas?
2. Suponga que no hay más variables explicativas importantes, ¿está usted satisfecho con el modelo de pronósticos de Julie? ¿De qué manera “vendería”

usted el modelo a la administración (y a Jackson Tilson)?

3. ¿De qué manera podría utilizarse el modelo de Julie para determinar gastos futuros por publicidad en periódicos y televisión?
4. ¿Qué condiciones podrían ocasionar que Julie examinara nuevamente su modelo de regresión, o quizás, que busque otro método para pronosticar las ventas?

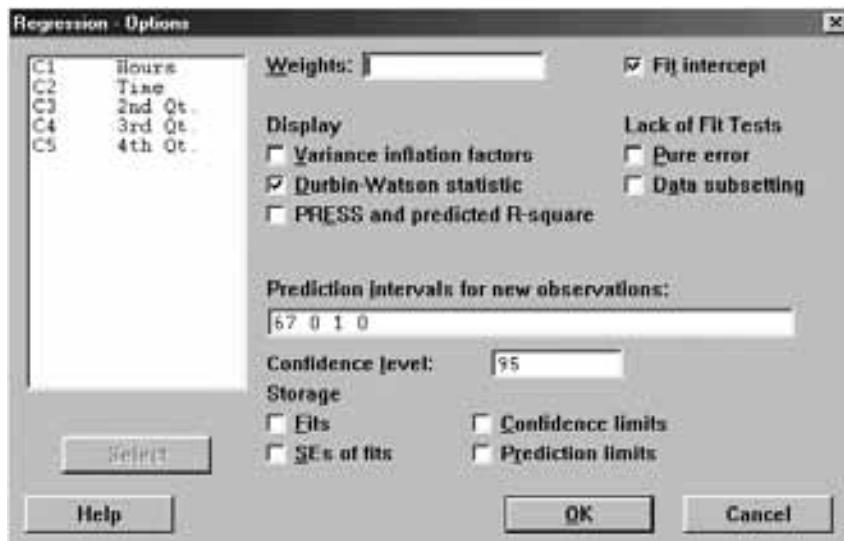
## Aplicaciones en Minitab

**El problema.** En el ejemplo 8.8, James Brown está tratando de pronosticar el consumo de energía eléctrica que efectúan los clientes residenciales para el tercero y cuarto trimestres de 1996 para la Washington Water Power.

### Solución en Minitab

1. Introduzca la variable *Hours* (horas) de la tabla 8.9 a la columna C1. Introduzca el tiempo en la columna C2 (1, 2, 3... 66). Introduzca las variables ficticias como se muestran en la tabla 8.10 en las columnas C3, C4 y C5.
2. Para correr el modelo de regresión estacional, dé clic en los siguientes menús:  
*Stat>Regression>Regression*
3. Aparece el cuadro de diálogo de *Regression* (regresión) similar al que se muestra en la figura 6.21.
  - a) Se selecciona *Hours* como la *Response* (respuesta) o variable dependiente.
  - b) *Time-4th Qt.* (tiempo cuarto trimestre) se introduce como el *Predictor* (elemento de predicción) o variables independientes.
  - c) Dé clic en *Options* (opciones) para que salga el cuadro de diálogo de *Regression-Options*.
4. Aparece el cuadro de *Regression-Options* que se muestra en la figura 8.15.
  - a) Dé clic en el cuadro de diálogo de *Durbin-Watson statistics* (estadística de Durbin-Watson).
  - b) En el espacio de *Prediction intervals for new observations* (intervalos de predicción para nuevas observaciones) escriba las observaciones nuevas para el tercer trimestre: 67 0 1 0.
  - c) Dé clic en OK y después OK en el cuadro de diálogo de *Regression*. Los resultados desplegados en la tabla 8.11 se presentan en la ventana de la sesión.

Observe que en el ejemplo 8.4 Fred Gardner desarrolló un modelo de regresión a través del origen (no había intercepción *Y*). Para correr este tipo de modelo, dé clic y marque el re-cuadro en el cuadro de *Fit intercept* (ajustar intercepción) como se ilustra en la figura 8.15.



**FIGURA 8.15** Regresión en Minitab – Cuadro de diálogo de opciones

## Aplicaciones en Excel

**El problema.** En el ejemplo 8.6 se desarrolló un modelo autorregresivo de primer orden para los datos de ventas de Novak Corporation.

### Solución en Excel

1. Introduzca el archivo Minitab que contiene los datos mostrados en la tabla 8.2 (vea la página 336) y resalte la columna de *Sales* (ventas). Dé clic en los siguientes menús:

Edit>Copy Cells

Ahora introduzca su hoja de cálculo de Excel, resalte A3 y dé clic en los siguientes menús:

Edit>Paste

Los datos de las ventas aparecen en la columna A. Después de incorporar el encabezado de esta columna la primera línea de la hoja de cálculo de parece a la figura 8.16 sin la variable retrasada.

2. Para crear la variable *Lagged one period* (retrasada por un periodo) en ventas, coloque el mouse en A3 y resalte hasta A19. Dé clic en los siguientes menús:

Edit>Copy

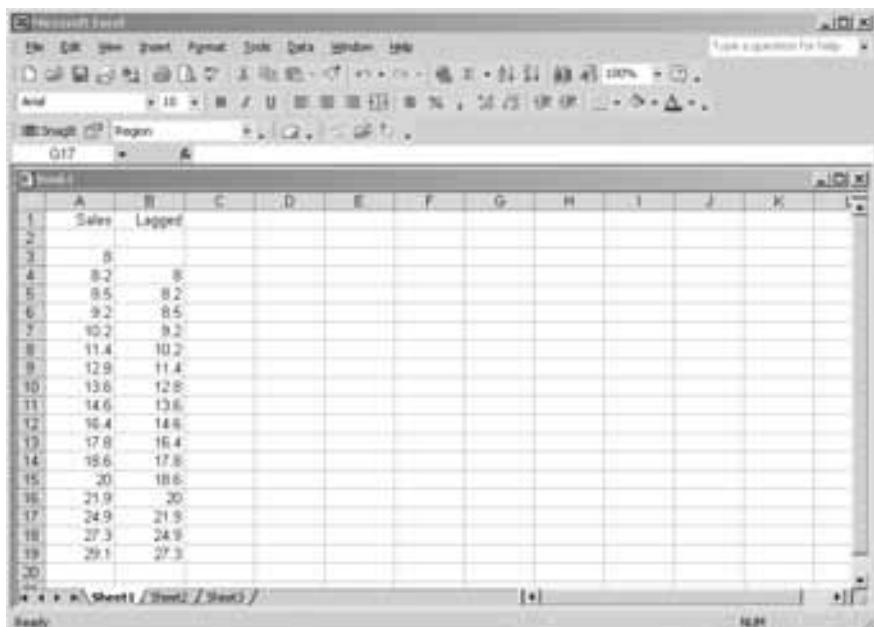
Ahora resalte B4 y dé clic en:

Edit>Paste

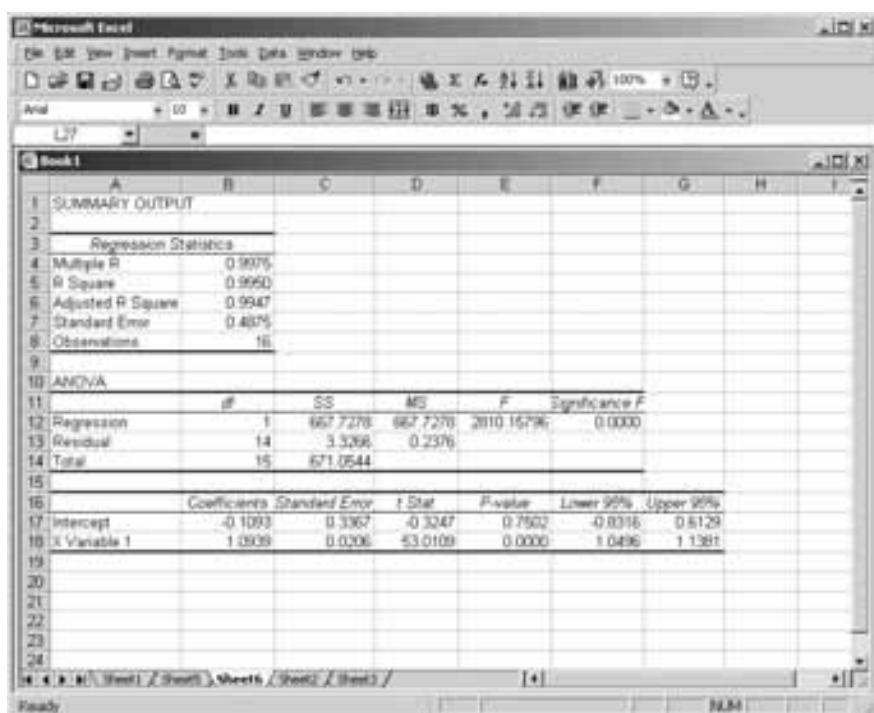
El resultado aparece en la figura 8.16.

3. Dé clic en los siguientes menús para correr el modelo autorregresivo:

Tools>Data Analysis



**FIGURA 8.16** Hoja de cálculo de Excel después de retrasar una variable en ventas por un periodo



**FIGURA 8.17** Resultados de Excel para un modelo autorregresivo

Aparece el cuadro de *Data Analysis* (análisis de datos). Bajo *Analysis Tools* (herramientas para el análisis) seleccione *Regression* (regresión). En la figura 6.25 se mostró el cuadro de diálogo *Regression*.

- a) Introduzca A4:A19 en el *Input Y Range* (rango Y de entrada).
- b) Introduzca B4:B19 en el *Input X Range* (rango X de entrada).
- c) Dé clic en el botón que está junto a *New Worksheet Ply* (en un libro nuevo).
- d) Dé clic en *OK* y el resultado aparece como se muestra en la figura 8.17.

## Referencias

- Diebold, F. X., *Elements of forecasting*, tercera edición, Cincinnati, OH, South-Western, 2004.
- Durbin, J. y G. S. Watson, “Testing for Serial Correlation in Least Squares Regression II”, *Biometrika* 38, 1951: 159-178.
- King J. L. y D. A. Bessler, “A Comparison of Multivariate Forecasting Procedures for Economic Time Series”, *International Journal of Forecasting* 1(1), 1985: 5-24.
- Kleinbaum, D. L. Kupper, K. Muller y A. Nizam, *Applied Regression Analysis and Other Multivariable Methods*, tercera edición, Pacific Grove, CA, Duxbury, 1998.
- Newbold P. y T. Bos, *Introductory Business and Economic Forecasting*, segunda edición, Cincinnati, OH, South-Western, 1994.
- Pindyck, R. S. y D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, cuarta edición, Nueva York, McGraw-Hill, 1998.
- Young, R. M., “Forecasting with an Econometric Model: The Issue of Judgemental Adjustment”, *Journal of Forecasting* 1(2) (1982): 189–204.



## CAPÍTULO

# 9

## LA METODOLOGÍA BOX-JENKINS (ARIMA)

En este texto se han presentado distintos métodos para analizar y pronosticar las series de tiempo. En el capítulo 4 se introdujo el método de suavizamiento y los pronósticos de las series de tiempo se generaron por medio de un mecanismo particular de este método (promedios). En el capítulo 5 se descompusieron las series de tiempo en componentes de tendencia, estacionales e irregulares; y los pronósticos se produjeron al extrapolar la tendencia estimada y la estacionalidad. En el capítulo 8 se consideraron modelos de regresión apropiados para las series de tiempo. Habitualmente, los pronósticos de la variable dependiente  $Y$ , generados a partir de estos modelos, necesitan los pronósticos de los valores futuros de las variables independientes, las  $X$ .

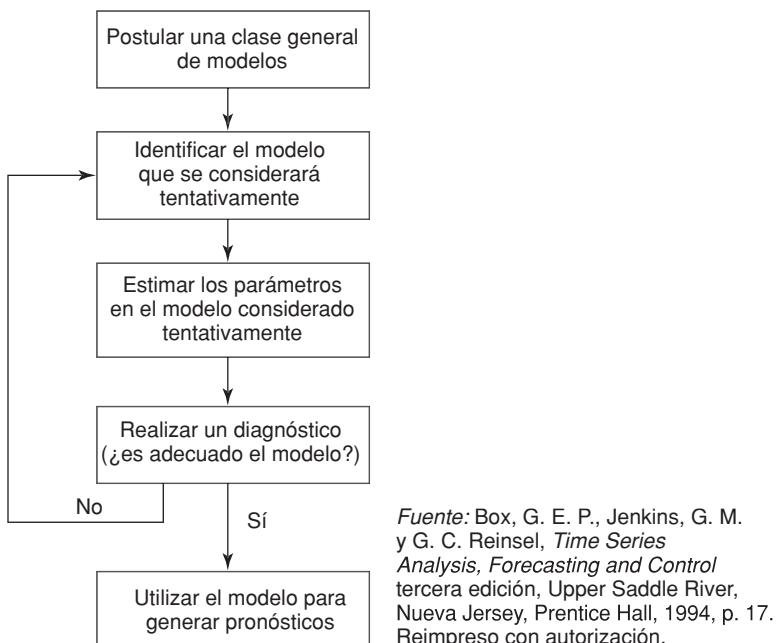
Este capítulo presenta una clase de modelos que puede producir pronósticos precisos con base en la descripción de patrones históricos en los datos. Los modelos de promedio móvil autorregresivo integrado (ARIMA, por sus siglas en inglés) son una clase de modelos lineales que tienen la capacidad de operar sobre series de tiempo *estacionarias o no estacionarias*. Recuerde que los procesos estacionarios varían en torno a un nivel fijo en tanto que los procesos no estacionarios no tienen un nivel promedio constante natural. De hecho, los modelos autorregresivos presentados en el capítulo 8, útiles para modelar las series de tiempo estacionarias, son un subconjunto de los modelos ARIMA.

Los modelos ARIMA no involucran a las variables independientes en su construcción. En cambio, emplean la información que se encuentra en la serie misma para generar los pronósticos. Por ejemplo, un modelo ARIMA para las ventas mensuales proyectaría un patrón histórico de ventas para producir un pronóstico para las ventas del mes siguiente.

Los modelos ARIMA dependen mucho de los patrones de autocorrelación que existen en los datos. Los estadísticos G. E. P. Box y G. M. Jenkins lograron grandes avances en la metodología para identificar, ajustar y verificar los modelos ARIMA adecuados. Por esto, los modelos ARIMA para producir pronósticos suelen ser llamados metodología Box-Jenkins.

### LA METODOLOGÍA BOX-JENKINS

La *metodología Box-Jenkins* para generar pronósticos es distinta de la mayoría de los métodos debido a que no *supone* un patrón particular en los datos históricos de las series que han de pronosticarse. Usa un método iterativo para identificar un modelo posible de una clase general de modelos. Enseguida, el modelo seleccionado se contrasta con los datos históricos para ver si describe con precisión la serie. El modelo se ajusta correctamente si los residuales son pequeños, están distribuidos aleatoriamente y no contienen información útil. Si el modelo especificado no es satisfactorio, el proceso se repite mediante un nuevo modelo diseñado para mejorar el original. Se sigue aplicando este procedimiento iterativo hasta que se encuentra un procedimiento satisfactorio. Hasta este punto se puede usar el modelo para el pronóstico. La figura 9.1 ilustra la estrategia de construcción del modelo Box-Jenkins.



Fuente: Box, G. E. P., Jenkins, G. M. y G. C. Reinsel, *Time Series Analysis, Forecasting and Control*, tercera edición, Upper Saddle River, Nueva Jersey, Prentice Hall, 1994, p. 17. Reimpreso con autorización.

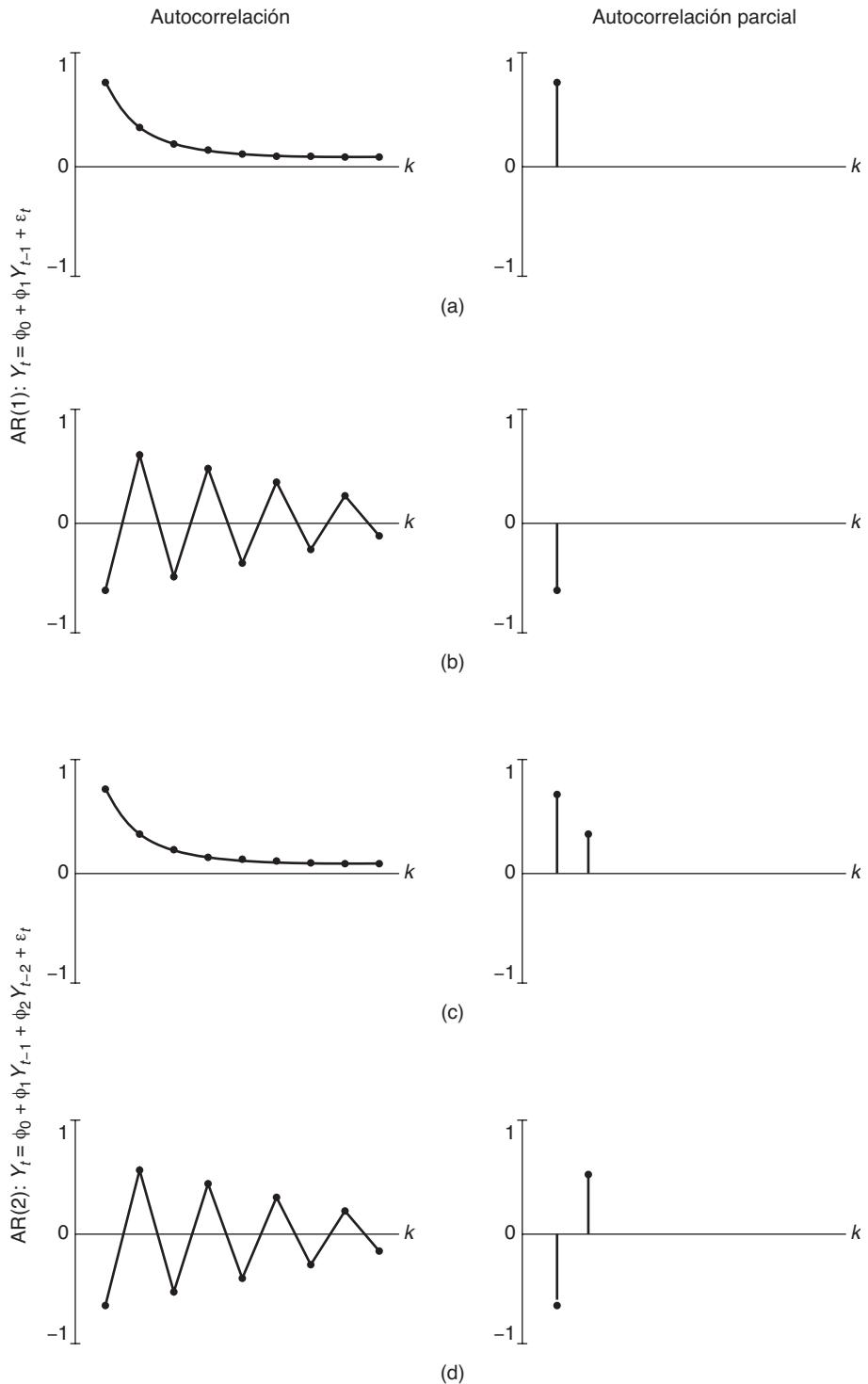
**FIGURA 9.1 Diagrama de flujo de la estrategia de construcción del modelo Box-Jenkins**

La selección inicial de un modelo ARIMA se basa en el examen de una gráfica de la serie de tiempo (para observar su carácter general) y un examen de su autocorrelación para diversos retrasos. De manera particular, el patrón de la autocorrelación de la muestra, calculado a partir de la serie de tiempo, coincide con el patrón de autocorrelación ya conocido que se asocia a un modelo ARIMA específico. Este acoplamiento se realiza para las autocorrelaciones y las autocorrelaciones parciales.<sup>1</sup> Los coeficientes de autocorrelación teóricos para los modelos ARIMA más comunes se muestran en las figuras 9.2, 9.3 y 9.4. La significancia de los patrones mostrados en estas figuras será más evidente conforme avanza el capítulo.

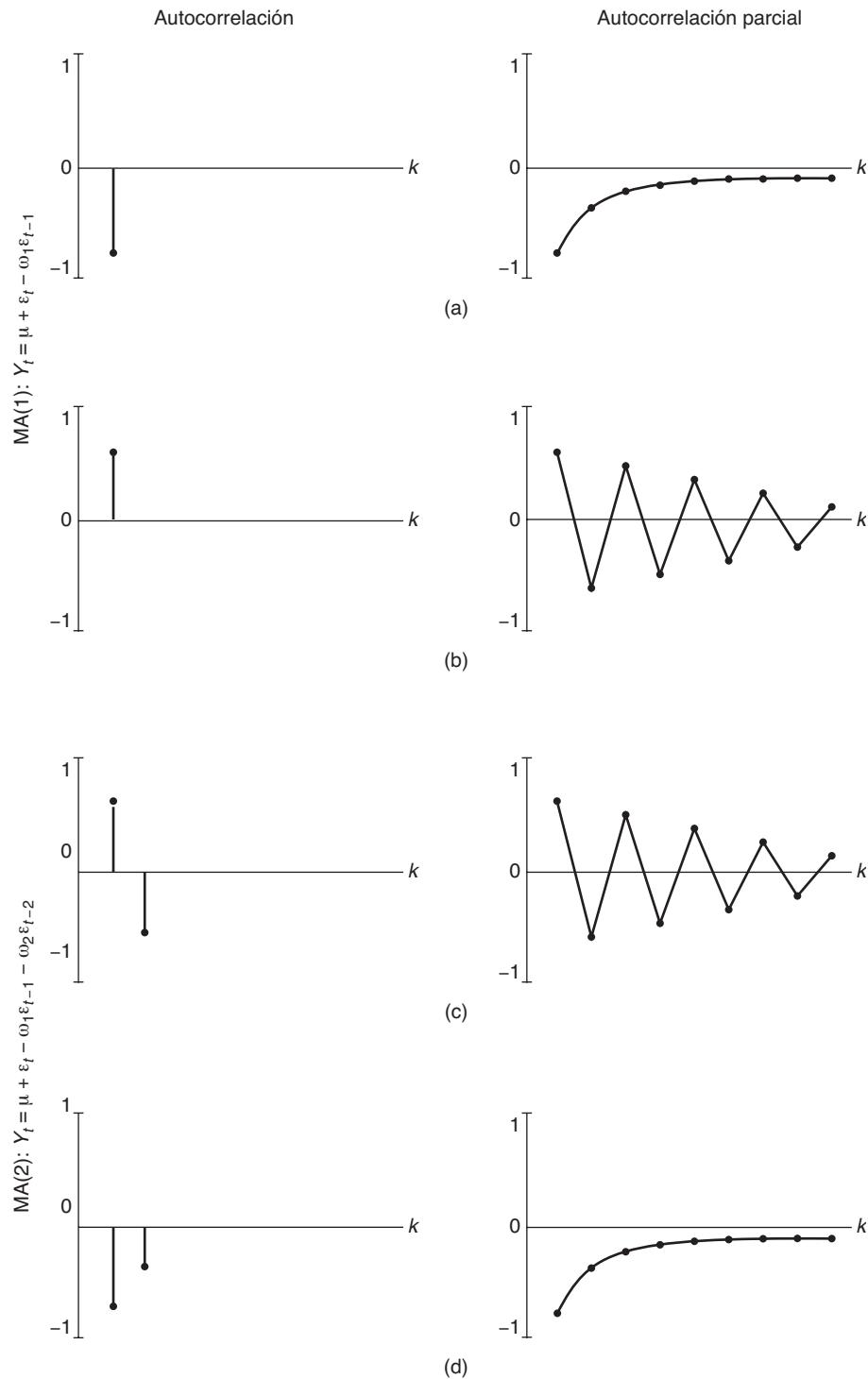
**La metodología Box-Jenkins** se refiere a una serie de procedimientos para identificar, ajustar y verificar los modelos ARIMA con los datos de la serie de tiempo. Los pronósticos proceden directamente de la forma del modelo ajustado.

Al seleccionar un modelo, recuerde que las autocorrelaciones calculadas a partir de los datos no serán exactamente iguales a las autocorrelaciones teóricas asociadas con un modelo ARIMA. Las autocorrelaciones calculadas con base en los datos están sujetas a la variación de la muestra. No obstante, usted debe ser capaz de igualar apropiadamente la mayoría de las series de datos con un modelo ARIMA. Si la selección inicial no es la adecuada, se presentarán insuficiencias en el análisis de los residuales (diagnóstico del modelo) y el modelo original puede verse modificado. Conforme vaya adquiriendo experiencia, la tarea de construir un modelo iterativo será cada vez más fácil.

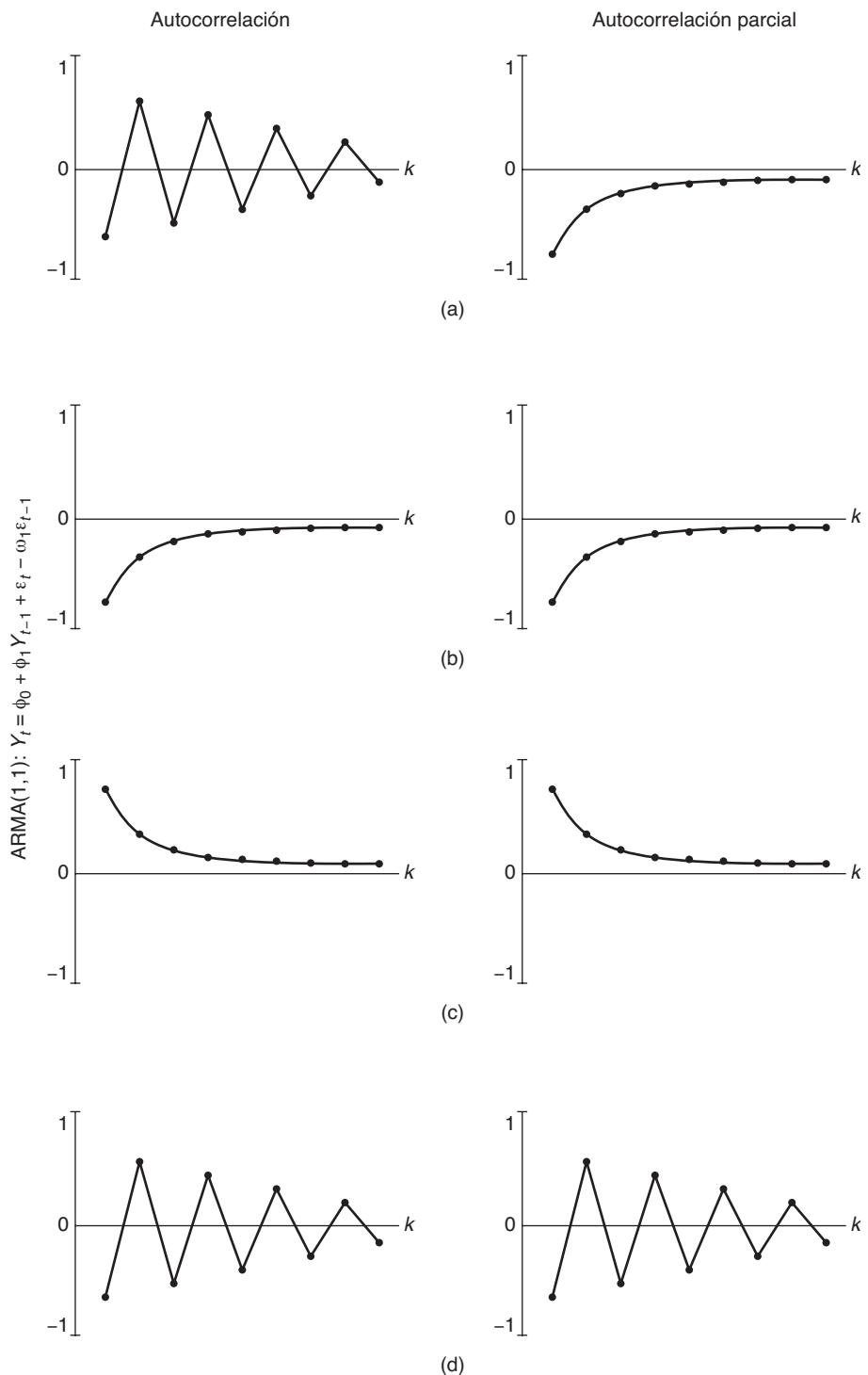
<sup>1</sup> Una autocorrelación parcial en un tiempo de retraso  $k$  es la correlación entre  $Y_t$  y  $Y_{t-k}$ , las respuestas para el periodo  $t$  y  $t-k$ , respectivamente, después de haber ajustado por efectos de los valores intervenidos  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ . Las autocorrelaciones se derivan y se analizan en Newbold y Bos (1994) y Box, Jenkins y Reinsel (1994).



**FIGURA 9.2** Coeficientes de autocorrelación y de autocorrelación parcial de los modelos AR(1) y AR(2)



**FIGURA 9.3** Coeficientes de autocorrelación y de autocorrelación parcial de los modelos MA(1) y MA(2)



**FIGURA 9.4** Coeficientes de autocorrelación y de autocorrelación parcial de un modelo mezclado ARMA(1, 1)

## Modelos autorregresivos

En el capítulo 8 se presentó un modelo autorregresivo de primer orden. Un modelo autorregresivo de cualquier orden  $p$  adopta la siguiente forma

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (9.1)$$

donde

$Y_t$  = variable de la respuesta (dependiente) en el tiempo  $t$

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$  = variable de respuesta en los retrasos  $t-1, t-2, \dots, t-p$ , respectivamente, estas  $Y$  desempeñan la función de variables independientes

$\phi_0, \phi_1, \phi_2, \dots, \phi_p$  = coeficientes que serán estimados<sup>2</sup>

$\varepsilon_t$  = término de error en el tiempo  $t$  que representa los efectos de las variables que no explica el modelo; los supuestos acerca del término de error son las mismas que las del modelo de regresión estándar

El modelo en la ecuación 9.1 aparenta ser un modelo de regresión con valores retrasados en la variable dependiente en las posiciones de la variable independiente; por esta razón se le da el nombre de modelo autorregresivo. Los modelos autorregresivos son apropiados para series de tiempo estacionarias y que tienen un coeficiente de  $\phi_0$  que se relaciona con el nivel constante de la serie. Si los datos varían alrededor de cero o se expresan como desviaciones de la media  $Y_t - \bar{Y}$ , no se requiere el coeficiente  $\phi_0$ .

Las ecuaciones para un modelo AR de orden 1 —AR(1)— y de orden 2 —AR(2)— se muestran en la figura 9.2. La figura 9.2(a) y (b) ilustra la conducta de las funciones autocorrelación teóricas y de una autocorrelación parcial de un modelo AR(1). Observe el comportamiento tan distinto de las funciones de autocorrelación y autocorrelación parcial. Los coeficientes de autocorrelación se aproximan gradualmente al cero, a la vez que los coeficientes de autocorrelación parcial caen a cero después del primer retraso de tiempo. La figura 9.2 (c) y (d) muestra la autocorrelación para un modelo AR(2). De nuevo, los coeficientes de autocorrelación se aproximan a cero y los coeficientes de autocorrelación parcial caen a cero después del segundo tiempo de retraso. Este tipo de patrón generalmente será válido para cualquier modelo AR( $p$ ). Debe recordarse que las funciones de autocorrelación de la muestra serán distintas a las soluciones teóricas debido a las variaciones de la muestra.

### Ejemplo 9.1

Se muestra la realización de pronósticos con un modelo AR(2) al usar la serie de datos de 75 lecturas de procesos que se muestran en la tabla 9.5. (Vea la página 398.) Sólo las últimas cinco observaciones de la tabla 9.5 se utilizarán para construir el pronóstico que se muestra en la tabla 9.1.

Se elige un modelo AR(2) para procesar los datos y mediante el método de mínimos cuadrados con el programa Minitab se obtiene  $\phi_0 = 115.2$ ,  $\phi_1 = -.535$  y  $\phi_2 = .0055$ . Suponga que en el tiempo  $t-1 = 75$  se requiere un pronóstico de las observaciones para el siguiente periodo  $t = 76$ . Debido a que el mejor pronóstico para el término de error es su valor promedio 0, el pronóstico para el periodo  $t = 76$  es

$$\begin{aligned}\hat{Y}_t &= \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2} \\ \hat{Y}_{76} &= 115.2 - .535 Y_{75} + .0055 Y_{74} \\ \hat{Y}_{76} &= 115.2 - .535(72) + .0055(99) = 77.2\end{aligned}$$

Para los modelos autorregresivos, los pronósticos dependen de los valores observados en períodos previos. Particularmente en modelos AR(2), los pronósticos para el siguiente valor dependen de las observaciones en los dos períodos anteriores. Para los modelos AR(3), los pronósticos para el siguiente valor dependen de las observaciones en los tres períodos anteriores, y así sucesivamente.

<sup>2</sup> El coeficiente  $\phi_0$  se relaciona con la media  $\mu$  en el proceso mediante  $\phi_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$ .

**TABLA 9.1 Pronósticos por medio de un modelo autorregresivo AR(2) para el ejemplo 9.1**

| Periodo | Tiempo $t$ | Valores $Y_t$ | Pronósticos $\hat{Y}_t$ | Residuales $e_t$ |
|---------|------------|---------------|-------------------------|------------------|
| $t-5$   | 71         | 90            | 76.4                    | 13.6             |
| $t-4$   | 72         | 78            | 67.5                    | 10.5             |
| $t-3$   | 73         | 87            | 74.0                    | 13.0             |
| $t-2$   | 74         | 99            | 69.1                    | 29.9             |
| $t-1$   | 75         | 72            | 62.7                    | 9.3              |
| $t$     | 76         |               | 77.2                    |                  |

### Modelos de promedio móvil

Un modelo de promedio móvil de orden  $q$  adopta la forma

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \quad (9.2)$$

donde

$Y_t$  = variable de respuesta (dependiente) en un tiempo  $t$

$\mu$  = valor promedio que permanece constante en el proceso

$\omega_1, \omega_2, \dots, \omega_q$  = coeficientes que serán estimados

$\varepsilon_t$  = término de error que representa los efectos de las variables no explicadas por el modelo; los supuestos acerca del término de error son las mismas que los del modelo de regresión estándar

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$  = errores en períodos anteriores al tiempo  $t$ , incorporados en la respuesta  $Y_t$ .

La ecuación 9.2 es similar a la ecuación 9.1, con excepción de que la variable dependiente  $Y$  depende de valores previos de los errores en lugar de la variable por sí misma. Los modelos de promedio móvil (MA) proporcionan pronósticos de  $Y$ , con base en una combinación lineal de un número finito de errores pasados, mientras que los modelos autorregresivos (AR) pronostican a  $Y$  como una función lineal de un número finito de valores anteriores de  $Y$ .

El término de *promedio móvil* para el modelo en la ecuación 9.2 es histórico y no debería confundirse con los procedimientos de promedio móvil que se analizan en el capítulo 4. Aquí, *promedio móvil* se refiere al hecho de que la desviación de respuesta de su media  $Y_t - \mu$  es una combinación lineal de errores actuales y pasados, y conforme avanza el tiempo los errores involucrados en esta combinación lineal también avanzan.

$$Y_t - \mu = \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q}$$

$$Y_{t+1} - \mu = \varepsilon_{t+1} - \omega_1 \varepsilon_t - \omega_2 \varepsilon_{t-1} - \cdots - \omega_q \varepsilon_{t-q+1}$$

Los pesos  $\omega_1, \omega_2, \dots, \omega_q$  no necesariamente se suman a 1 y pueden ser negativos o positivos, a pesar de que cada uno está precedido por un signo de menos en la especificación del modelo.

La figura 9.3 muestra las ecuaciones de un modelo de promedio móvil de primer orden, MA(1), y las de un modelo de promedio móvil de segundo orden, MA(2). Los términos de error pueden sumarse secuencialmente para obtener un modelo MA( $q$ ), en donde  $q$  representa el número de términos de error anteriores que habrán de incluirse en el pronóstico de la siguiente observación. La figura 9.3(a) y (b) también ilustra la conducta de los coeficientes de la autocorrelación teórica y autocorrelación parcial para el modelo

**TABLA 9.2 Pronósticos mediante un modelo de promedio móvil MA(2) para el ejemplo 9.2**

| Periodo     | Tiempo<br><i>t</i> | Valores<br><i>Y<sub>t</sub></i> | Pronósticos<br>$\hat{Y}_t$ | Residuales<br><i>e<sub>t</sub></i> |
|-------------|--------------------|---------------------------------|----------------------------|------------------------------------|
| <i>t</i> -5 | 71                 | 90                              | 76.1                       | 13.9                               |
| <i>t</i> -4 | 72                 | 78                              | 69.1                       | 8.9                                |
| <i>t</i> -3 | 73                 | 87                              | 75.3                       | 11.7                               |
| <i>t</i> -2 | 74                 | 99                              | 72.0                       | 27.0                               |
| <i>t</i> -1 | 75                 | 72                              | 64.3                       | 7.7                                |
| <i>t</i>    | 76                 |                                 | 80.6                       |                                    |

MA(1). Observe cuán afortunado es que las funciones de autocorrelación y autocorrelación parcial de los modelos AR y MA se comporten de manera tan diferente. Los coeficientes de autocorrelación en el modelo MA(1) caen a cero después del primer retraso de tiempo, mientras que los coeficientes de autocorrelación parcial también se aproximan al cero pero lo hacen gradualmente. Además, los coeficientes de autocorrelación del modelo MA(2) equivalen a cero después del segundo retraso de tiempo, mientras que los coeficientes de autocorrelación parcial se aproximan a cero gradualmente (vea la figura 9.3(c) y (d)). De nuevo, debe mencionarse que las funciones de autocorrelación muestrales serán distintas de las funciones teóricas debido a la variación de la muestra.

### Ejemplo 9.2

Las 75 lecturas de procesos que se muestran en la tabla 9.5 pueden utilizarse para ilustrar el pronóstico con un modelo MA(2). La realización del pronóstico se muestra utilizando únicamente las últimas cinco observaciones que se muestran en la tabla 9.2.

Se selecciona el modelo MA(2) y el programa Minitab calcula los estimados de los mínimos cuadrados<sup>3</sup> para los coeficientes del modelo  $\hat{\mu} = 75.4$ ,  $\hat{\omega}_1 = .5667$ , y  $\hat{\omega}_2 = -.3560$ . De nuevo supongamos que en un tiempo  $t - 1 = 75$  se requiere de un pronóstico de las observaciones en un periodo  $t = 76$ . Debido a que en un tiempo  $t - 1$  la mejor alternativa para el término de error en el siguiente periodo es su valor promedio 0, y las mejores alternativas para los errores en los períodos actuales y anteriores son los residuales correspondientes, el pronóstico para el periodo  $t = 76$  es

$$\begin{aligned}\hat{Y}_t &= \hat{\mu} - \hat{\omega}_1 e_{t-1} - \hat{\omega}_2 e_{t-2} \\ \hat{Y}_{76} &= 75.4 - .5667e_{75} + .3560e_{74} \\ \hat{Y}_{76} &= 75.4 - .5667(7.7) + .3560(27) = 80.6\end{aligned}$$

Observe que ambos residuales,  $e_{75}$  y  $e_{74}$ , se sustituyen por los errores  $e_{75}$  y  $e_{74}$  en los cálculos del pronóstico para el periodo 76. Para calcular los pronósticos a partir de los modelos de promedio móvil, se sustituyen los errores correspondientes a períodos que ya ocurrieron por los residuales de esos períodos. El número de residuales involucrados en el pronóstico de la siguiente observación es igual al orden del modelo de promedio móvil.

### Modelos de promedios móviles autorregresivos

Un modelo con términos autorregresivos también puede combinarse con un modelo que tenga términos de promedio móvil para obtener un modelo autorregresivo de promedio móvil “mezclado”. Para representar estos modelos es conveniente usar la notación ARMA(*p, q*)

<sup>3</sup> En los modelos de promedio móvil o los que impliquen términos de promedio móvil, los estimados de mínimos cuadrados para los coeficientes deben obtenerse de manera iterativa utilizando un algoritmo de mínimos cuadrados no lineal. Si se usa un punto de inicio, el algoritmo de mínimos cuadrados no lineal genera estimados de coeficientes mejores y que poseen una suma de errores cuadrados más baja. Los estimados mejoran continuamente hasta que la suma de errores cuadrados no pueda disminuirse más de manera apreciable. Aunque los modelos autorregresivos pueden ser ajustados con paquetes estándar de regresión, los estimados de mínimos cuadrados de los coeficientes en los modelos autorregresivos a menudo se obtienen con el uso de un método no lineal de mínimos cuadrados.

donde  $p$  equivale al orden de la parte autorregresiva y  $q$  es el orden de la parte del promedio móvil. Un modelo ARMA( $p, q$ ) tiene la forma general

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \quad (9.3)$$

Los modelos ARMA( $p, q$ ) pueden describir una amplia variedad de comportamientos para las series de tiempo estacionarias.<sup>4</sup> Los pronósticos generados por un modelo ARMA( $p, q$ ) dependerán de los valores actuales y anteriores de la respuesta  $Y$ , así como de los valores pasados y actuales de los errores (residuales)  $\varepsilon$ .

La figura 9.4 muestra la ecuación de un modelo ARMA(1, 1) y las conductas posibles de una autocorrelación teórica y una autocorrelación parcial. En este caso, ambas autocorrelaciones se desvanecen, ninguna se corta.

### Resumen

Los patrones de autocorrelación y correlación para los procesos de promedio móvil autorregresivos pueden resumirse como se muestra a continuación

|                | <i>Autocorrelaciones</i>                | <i>Autocorrelaciones parciales</i>      |
|----------------|-----------------------------------------|-----------------------------------------|
| MA( $q$ )      | Corte después del orden $q$ del proceso | Desvanecimiento                         |
| AR( $p$ )      | Desvanecimiento                         | Corte después del orden $p$ del proceso |
| ARMA( $p, q$ ) | Desvanecimiento                         | Desvanecimiento                         |

Los números de los términos autorregresivos y de promedio móvil (órdenes  $p$  y  $q$ ) en un modelo ARMA se determinan a partir de los patrones de las autocorrelaciones de la muestra y de las autocorrelaciones parciales, así como de los valores del criterio de selección del modelo que se presenta en una sección posterior de este capítulo. En la práctica, los valores de  $p$  y  $q$  rara vez exceden a 2.

## APLICACIÓN DE UNA ESTRATEGIA PARA LA CONSTRUCCIÓN DE UN MODELO

Como se muestra en la figura 9.1, el método Box-Jenkins utiliza una estrategia iterativa para la construcción de modelos que consiste en seleccionar un modelo inicial (identificación del modelo), mediante la estimación de los coeficientes del modelo (estimación de los parámetros) y el análisis de los residuales (evaluación del modelo). De ser necesario, el modelo inicial se modifica y el proceso se repite hasta que los residuales indican que ya no se requiere modificación alguna. En este punto, el modelo ajustado puede utilizarse para realizar los pronósticos.

A continuación se examinan con cierto detalle los pasos en la estrategia de construcción de modelos.

### Paso 1: identificación del modelo

- El primer paso en la identificación del modelo es determinar si la serie es estacionaria; es decir, si la serie de tiempo aparece variar alrededor de un nivel fijo. Es útil observar una gráfica de la serie junto con una muestra de la función de autocorrelación. Una serie de tiempo no estacionaria se indica si la serie parece crecer o decrecer con relación al tiempo y las autocorrelaciones no pueden desvanecerse con rapidez. La serie de tiempo que se observa en la figura 8.2 no es estacionaria, y el patrón de las autocorrelaciones de la muestra que se indica en la figura 8.3 es típica de una serie no estacionaria (vea la página 330).

<sup>4</sup> Observe que cuando  $q = 0$  el modelo ARMA( $p, 0$ ) se reduce a un modelo autorregresivo puro de orden  $p$ . De manera semejante, cuando  $p = 0$ , el modelo ARMA( $0, q$ ) es un modelo de promedio móvil puro de orden  $q$ .

Si la serie no es estacionaria, con frecuencia puede convertirse en una serie estacionaria al tomar sus diferencias. Es decir, la serie es reemplazada por una serie de diferencias. Entonces, se especifica un modelo ARMA para la serie de las diferencias. En efecto, el analista está modelando los cambios en lugar de los niveles.

Por ejemplo, suponga que la serie original  $Y_t$  por lo general se incrementa con el tiempo, pero las primeras diferencias  $\Delta Y_t = Y_t - Y_{t-1}$  varían alrededor de un nivel fijo. Podría ser apropiado modelar las diferencias estacionarias por medio de un modelo ARMA, de, por ejemplo, un orden  $p = 1$  y  $q = 1$ . En este caso el modelo es<sup>5</sup>

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

o

$$(Y_t - Y_{t-1}) = \phi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

En algunos casos podría ser necesario tomar las diferencias de las diferencias antes de obtener los datos estacionarios. Cuando se toma una diferencia simple dos veces los datos estacionarios son

$$\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

Se toman diferencias hasta que la gráfica de los datos indica que la serie varía alrededor de un nivel fijo y las autocorrelaciones de la muestra desaparecen con rapidez. El número de diferencias requerido para lograr un estado estacionario se denota por la  $d$ .

A los modelos para las series que no son estacionarias se les llama modelos de promedio móvil *integrados* autorregresivos y se denotan como ARIMA  $(p, d, q)$ .<sup>6</sup> En este caso,  $p$  indica el orden de la parte autorregresiva,  $d$  indica el orden de la diferencia y  $q$ , el orden de la parte de promedio móvil. Si la serie original es estacionaria, entonces  $d = 0$  y los modelos ARIMA se reducen a modelos ARMA. En consecuencia, desde este punto, la notación ARIMA  $(p, d, q)$  se utiliza para indicar los modelos tanto para series de tiempo estacionarias ( $d = 0$ ) como para las no estacionarias ( $d > 0$ ).

A pesar de que los modelos ARIMA involucran diferencias, los pronósticos para la serie original siempre pueden calcularse directamente a partir del modelo ajustado.

2. Una vez que se ha obtenido una serie estacionaria, el analista debe identificar la forma del modelo que habrá de utilizar.

La segunda parte del paso 1 se consigue al comparar la autocorrelación y la autocorrelación parcial que se calcularon a partir de los datos para las autocorrelaciones y autocorrelaciones parciales teóricas de los diversos modelos ARIMA. Para ayudar en la selección del modelo, las correlaciones teóricas para los modelos ARIMA más comunes se muestran en las figuras 9.2, 9.3 y 9.4.

Cada modelo ARIMA cuenta con una serie única de autocorrelaciones y autocorrelaciones parciales, y el analista debe tener la capacidad de identificar los valores que se obtienen a partir de la muestra con alguno de los patrones teóricos correspondientes.

Es probable que haya cierta ambigüedad al determinar un modelo ARIMA apropiado a partir de los patrones que provienen de las autocorrelaciones y las autocorrelaciones parciales de la muestra. De esta manera, la selección del modelo inicial deberá considerarse como tentativa. Los análisis pueden realizarse en los pasos 2 y 3 para determinar si el modelo es adecuado. Si no fuera el caso, se deberá intentar con un modelo alterno. Con un poco de práctica, el analista deberá volverse un experto para identificar un modelo adecuado.

---

<sup>5</sup> Es posible que no se requiera el término constante  $\phi_0$  cuando se utiliza un modelo ARIMA para una serie de diferencias.

<sup>6</sup> El término *integrado* significa que las diferencias deberán sumarse (o integrarse) para obtener la serie original.

Recuerde, si las autocorrelaciones de muestra se desvaneцен exponencialmente a cero y las autocorrelaciones parciales de muestra se cortan, el modelo requerirá términos autorregresivos. Si las autocorrelaciones de muestra se cortan y las autocorrelaciones parciales de muestra se desvaneцен, el modelo requerirá términos de promedio móvil. Si las autocorrelaciones de muestra y las autocorrelaciones parciales se desvaneцен, se indican términos tanto autorregresivos como de promedio móvil. Al contar el número de autocorrelaciones de muestra significativas y autocorrelaciones parciales, pueden determinarse las órdenes de MA y AR. Para juzgar su significancia, tanto las autocorrelaciones como las autocorrelaciones parciales de muestra usualmente se comparan con  $\pm 2/\sqrt{n}$  donde  $n$  es el número de observaciones en la serie de tiempo. Estos límites funcionan bien cuando el valor de  $n$  es grande.

Al ser iguales todas las cosas, los modelos más simples se prefieren sobre los modelos más complejos. Esto se conoce como el *principio de parsimonia*. Con una cantidad limitada de datos, es relativamente fácil encontrar un modelo que tenga un gran número de parámetros que ajuste correctamente los datos. Sin embargo, es probable que los pronósticos de tal modelo sean malos porque se está modelando mucha de la variación en los datos debida a un error aleatorio. El objetivo es desarrollar el modelo más simple que proporcione una descripción adecuada de la mayor parte de las características de los datos.

El *principio de parsimonia* se refiere a la preferencia por modelos más simples en lugar de los complejos.

## Paso 2: estimación de modelos

1. Una vez que se ha seleccionado un modelo tentativo, deben estimarse los parámetros para dicho modelo.

Los parámetros en los modelos ARIMA se estiman al minimizar la suma de los cuadrados en los errores de ajuste. En general, estos estimados de los mínimos cuadrados deben obtenerse mediante un procedimiento no lineal de mínimos cuadrados. Un procedimiento no lineal de mínimos cuadrados es, sencillamente, un algoritmo que encuentra el mínimo de la suma de la función de errores cuadrados. Después de que se hayan determinado los estimados de los mínimos cuadrados y sus errores estándar, los valores  $t$  pueden construirse e interpretarse como se hace usualmente. Los parámetros que se juzgan de manera significativa como distintos de cero, se conservan en el modelo ajustado; pero se desechan los parámetros que no son significativos. Por ejemplo, suponga que un ARIMA(1, 0, 1) se ha ajustado a una serie de tiempo de 100 observaciones y la ecuación ajustada es

$$\hat{Y}_t = 33.4 + .25Y_{t-1} - .5\hat{\epsilon}_{t-1}$$

(7.02)      (.17)      (.21)

donde los números entre paréntesis bajo los coeficientes estimados son los errores estándar. Debido a que la proporción  $t$  del coeficiente del término autorregresivo es  $t = .25/.17 = 1.47$  con un valor  $p$  de 0.14 la hipótesis  $H_0 = \phi_1 = 0$  no se rechaza y este término puede borrarse del modelo. Un modelo ARIMA(0, 0, 1)—es decir, un modelo MA(1)—podría ajustarse a los datos.

2. Además, se calcula el *error cuadrado medio de los residuales*, un estimado de la varianza del error  $\epsilon_t$ .

El *error cuadrado medio de los residuales* se define como<sup>7</sup>

$$s^2 = \frac{\sum_{t=1}^n e_t^2}{n-r} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n-r} \quad (9.4)$$

<sup>7</sup> La raíz cuadrada de  $s^2$ ,  $s = \sqrt{s^2}$ , es análoga al error estándar del estimado (vea ecuación 7.4).

donde

$$\begin{aligned} e_t &= Y_t - \hat{Y}_t = \text{el residual en el tiempo } t \\ n &= \text{el número de residuales} \\ r &= \text{el número total de los parámetros estimados} \end{aligned}$$

El error cuadrado medio residual es útil para evaluar el ajuste y comparar distintos modelos. También se utiliza para calcular los límites del error en el pronóstico.

### Paso 3: evaluación del modelo

Antes de usar el modelo para realizar pronósticos, debe asegurarse que es un modelo adecuado. Básicamente, un modelo es adecuado si sus residuales no pueden utilizarse para mejorar los pronósticos. Es decir, los residuales deben ser aleatorios.

1. Muchas de las gráficas de los residuales que son útiles para el análisis de regresión pueden desarrollarse para los residuales de un modelo ARIMA. De manera particular, son útiles un histograma y una gráfica de probabilidad normal (para verificar la normalidad) y una gráfica de secuencia de tiempo (para verificar los datos atípicos).
2. Las autocorrelaciones residuales individuales deberán ser pequeñas y, por lo general, estar dentro de  $\pm 2/\sqrt{n}$  de cero. Las autocorrelaciones residuales significativas en retrasos cortos o estacionales sugieren que el modelo no es adecuado y que se debe elegir un modelo nuevo o modificado.
3. Como un grupo, las autocorrelaciones residuales deberán ser coherentes con aquellas producidas por los errores aleatorios.

Una prueba chi cuadrada ( $\chi^2$ ) que se basa en la estadística de Ljung-Box  $Q$  proporciona una revisión global de la pertinencia del modelo. Esta prueba considera las dimensiones de las autocorrelaciones residuales como un grupo. La estadística de prueba  $Q$  es

$$Q_m = n(n + 2) \sum_{k=1}^m \frac{r_k^2(e)}{n - k} \quad (9.5)$$

la cual se distribuye aproximadamente como una variable aleatoria de chi cuadrada con grados de libertad  $m - r$  en donde  $r$  es el número total de parámetros estimados en el modelo ARIMA. En la ecuación 9.5,

$$\begin{aligned} r_k(e) &= \text{la autocorrelación residual en el retraso } k \\ n &= \text{el número de residuales} \\ k &= \text{el retraso de tiempo} \\ m &= \text{el número de retrasos de tiempo que habrán de ser evaluados} \end{aligned}$$

Si el valor  $p$  asociado con la estadística  $Q$  es pequeño (por ejemplo, un valor  $p$  de  $< .05$ ), se considera que el modelo es inadecuado. El analista deberá considerar un modelo nuevo o modificado y continuar el análisis hasta que se determine un modelo satisfactorio.

El juicio desempeña una gran función en el esfuerzo por construir un modelo. Dos modelos simples que compiten entre sí podrían describir adecuadamente los datos y con base en la naturaleza de los pronósticos se podría hacer una selección. Además, es probable que se ignoren muy pocos residuales grandes si se pueden explicar por circunstancias poco usuales, y el modelo es adecuado para el resto de las observaciones.

### Paso 4: realización de pronósticos con el modelo

1. Despues de que se ha encontrado un modelo adecuado, se pueden llevar a cabo los pronósticos para un periodo, o varios, en el futuro.

También pueden construirse intervalos de predicción con base en los pronósticos. En general, para un nivel de confianza determinado, mientras más largo sea el tiempo guía del pronóstico, mayor será el intervalo de predicción. Esto es sensato porque se espera que la incertidumbre para el pronóstico de un valor lejano sea más grande que, por ejemplo, para la observación próxima. Es tedioso calcular los pronósticos y los intervalos de predicción; por eso es mejor hacerlo mediante la computadora. Los programas de cómputo que ajustan los modelos ARIMA generan pronósticos e intervalos de predicción como lo requiere el analista.

2. A medida que se tienen más datos disponibles, se puede usar el mismo modelo ARIMA para generar pronósticos revisados que procedan de otro origen de tiempo.
3. Si el patrón de la serie parece cambiar con el tiempo, los nuevos datos podrían usarse para volver a estimar los parámetros del modelo o, de ser necesario, desarrollar un modelo completamente nuevo.

Vigilar los errores de pronóstico es una buena idea. Si las magnitudes de los errores más recientes tienden a ser consistentemente mayores que los anteriores, quizás sea la hora de evaluar otra vez el modelo. En este momento podría ser necesaria otra iteración de la estrategia para la construcción de modelos. Lo mismo es válido si los errores recientes de los pronósticos tienden a ser consistentemente positivos (predecir de menos) o negativos (predecir de más).

### Ejemplo 9.3

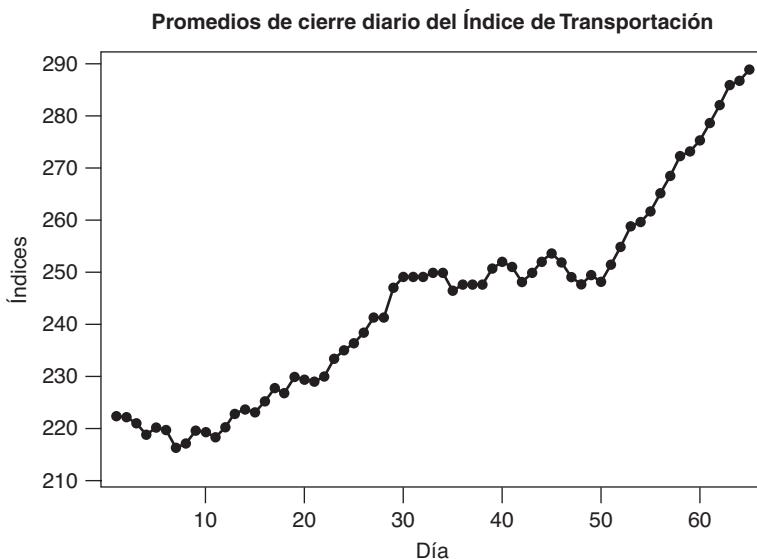
La empresa Cameron Consulting Corporation se especializa en los servicios de inversión de cartera. Lynn Stephens, analista de la compañía, tiene la obligación de desarrollar técnicas más sofisticadas para pronosticar los promedios del Dow Jones. Recientemente Lynn asistió a un taller sobre la metodología Box-Jenkins y decidió poner a prueba esta técnica en el Índice de Transportación. La tabla 9.3 representa 65 promedios de cierre diario del Índice de Transportación durante el verano.

Lynn comenzó el análisis al examinar la gráfica de datos que se muestra en la figura 9.5. En apariencia, allí existe una tendencia ascendente en la serie. Su siguiente paso para identificar un modelo tentativo fue estudiar las autocorrelaciones de la muestra, las cuales se ilustran en la figura 9.6. Cuando Lynn observó que las primeras autocorrelaciones eran persistentemente grandes y se deslizaban hacia cero muy lentamente, reconoció que su observación original era correcta y que esta serie de tiempo no era estacionaria y no variaba alrededor de un nivel fijo.

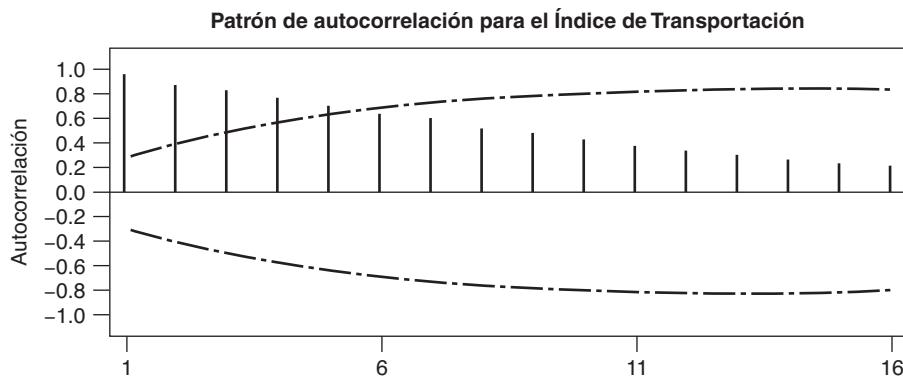
Lynn decidió diferenciar los datos para ver si podía eliminar la tendencia y crear una serie estacionaria. Una gráfica de los datos diferenciados (que no se muestra) parece variar en torno a un nivel fijo. De hecho, la media de la muestra de las diferencias fue 1.035. Las autocorrelaciones

**TABLA 9.3 Promedios de cierre diario del Índice de Transportación para el ejemplo 9.3**

| Periodo | Promedio de cierre | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1       | 222.34             | 15      | 223.07             | 29      | 246.74             | 43      | 249.76             | 57      | 268.21             |
| 2       | 222.24             | 16      | 225.36             | 30      | 248.73             | 44      | 251.66             | 58      | 272.16             |
| 3       | 221.17             | 17      | 227.60             | 31      | 248.83             | 45      | 253.41             | 59      | 272.79             |
| 4       | 218.88             | 18      | 226.82             | 32      | 248.78             | 46      | 252.04             | 60      | 275.03             |
| 5       | 220.05             | 19      | 229.69             | 33      | 249.61             | 47      | 248.78             | 61      | 278.49             |
| 6       | 219.61             | 20      | 229.30             | 34      | 249.90             | 48      | 247.76             | 62      | 281.75             |
| 7       | 216.40             | 21      | 228.96             | 35      | 246.45             | 49      | 249.27             | 63      | 285.70             |
| 8       | 217.33             | 22      | 229.99             | 36      | 247.57             | 50      | 247.95             | 64      | 286.33             |
| 9       | 219.69             | 23      | 233.05             | 37      | 247.76             | 51      | 251.41             | 65      | 288.57             |
| 10      | 219.32             | 24      | 235.00             | 38      | 247.81             | 52      | 254.67             |         |                    |
| 11      | 218.25             | 25      | 236.17             | 39      | 250.68             | 53      | 258.62             |         |                    |
| 12      | 220.30             | 26      | 238.31             | 40      | 251.80             | 54      | 259.25             |         |                    |
| 13      | 222.54             | 27      | 241.14             | 41      | 251.07             | 55      | 261.49             |         |                    |
| 14      | 223.56             | 28      | 241.48             | 42      | 248.05             | 56      | 264.95             |         |                    |



**FIGURA 9.5** Promedios de cierre diario del Índice de Transportación Dow Jones



**FIGURA 9.6** Función de autocorrelación de la muestra para el Índice de Transportación

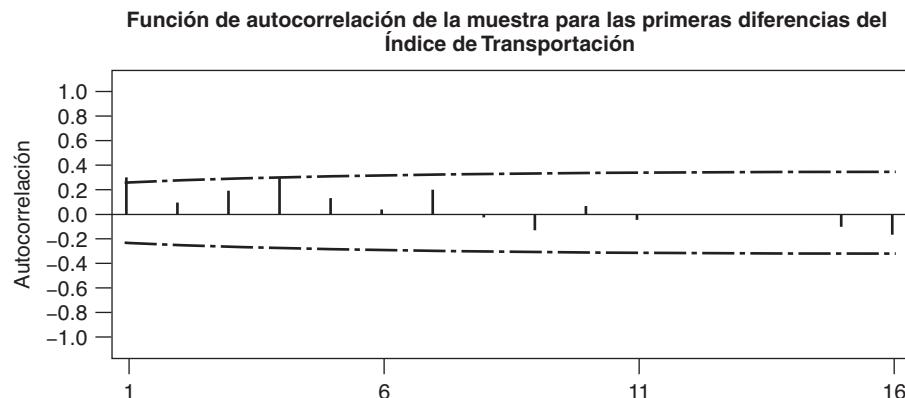
de la muestra para las diferencias se muestran en la figura 9.7. Las autocorrelaciones parciales de la muestra para las diferencias se muestran en la figura 9.8.

Lynn estaba perpleja. Al comparar las autocorrelaciones con sus límites de error, la única autocorrelación significativa fue el retraso 1. De manera semejante, únicamente la autocorrelación parcial del retraso 1 fue significativa. Aparentemente las autocorrelaciones se cortan después del retraso 1, lo cual indica una conducta MA(1). Asimismo, parece que las autocorrelaciones parciales se cortan después del retraso 1, lo cual indica una conducta de tipo AR(1). Al parecer, ninguno de los patrones se desvanece de manera suave en los retrasos cortos. Lynn decidió ajustar los modelos ARIMA (1, 1, 0) y ARIMA (0, 1, 1) al Índice de Transportación. También, determinó incluir un término constante en cada modelo puesto que la serie de diferencias parece variar en torno a un nivel mayor que cero. Si  $Y_t$  denota al Índice de Transportación, entonces la serie diferenciada es  $\Delta Y_t = Y_t - Y_{t-1}$  y los modelos de Lynn son

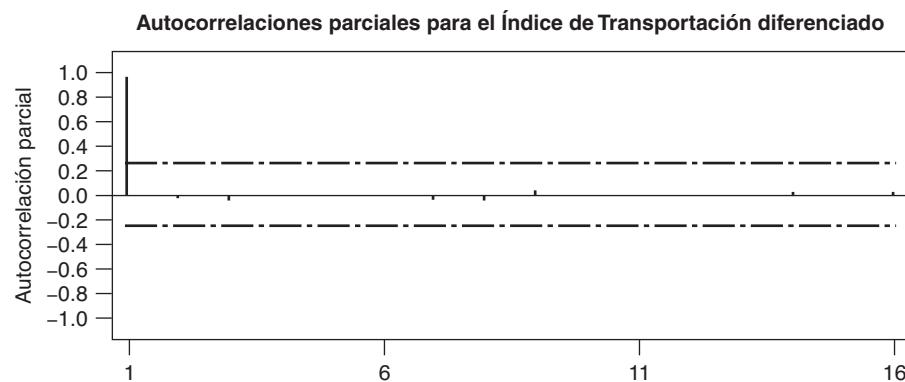
$$\text{ARIMA}(1, 1, 0): \Delta Y_t = \phi_0 + \phi_1 \Delta Y_{t-1} + \varepsilon_t$$

$$\text{ARIMA}(0, 1, 1): \Delta Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

La tabla 9.4 muestra los resultados en Minitab para los modelos de Lynn. Las autocorrelaciones residuales para el ajuste del modelo ARIMA (1, 1, 0) se muestran la figura 9.9.



**FIGURA 9.7** Función de autocorrelación de la muestra para las primeras diferencias del Índice de Transportación



**FIGURA 9.8** Autocorrelaciones parciales de la muestra para las primeras diferencias del Índice de Transportación

Ambos modelos ajustan bastante bien los datos. Los mínimos cuadrados residuales (MS) son

$$\text{ARIMA}(1, 1, 0): \quad s^2 = 3.536$$

$$\text{ARIMA}(0, 1, 1): \quad s^2 = 3.538$$

Lynn también observa que la constante en el modelo ARIMA(0, 1, 1) se estima en  $\hat{\mu} = 1.038$  (en esencia) la media de la muestra de las diferencias.

La figura 9.9 muestra que no hay una autocorrelación residual significativa para el modelo ARIMA (1, 1, 0). A pesar de que no se muestra la función de autocorrelación residual para el modelo ARIMA (0, 1, 1), los resultados son similares. La estadística Ljung-Box  $Q_m$  calculada para los grupos de retrasos  $m = 12, 24, 36$  y  $48$  no es significativa como se indica por los valores  $p$  para cualquiera de los modelos. Lynn decide que cualquier modelo es adecuado. Además, los pronósticos de un paso anticipado proporcionados por los dos modelos son básicamente los mismos.

Para resolver este dilema, Lynn opta por el modelo ARIMA (1, 1, 0) sobre la base de su ajuste ligeramente superior. Ella revisa el pronóstico para el periodo 66 de su modelo de la siguiente manera

$$Y_t - Y_{t-1} = \phi_0 + \phi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t$$

O

$$Y_t = Y_{t-1} + \phi_0 + \phi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t$$

**TABLA 9.4** Resultados en Minitab para el ajuste de los modelos ARIMA(1, 1, 0) y ARIMA(0, 1, 1) para el Índice de Transportación

**ARIMA(1, 1, 0): Model for Transportation Index**

Final Estimates of Parameters

| Type     | Coef   | SE Coef | T    | P     |
|----------|--------|---------|------|-------|
| AR 1     | 0.2844 | 0.1221  | 2.33 | 0.023 |
| Constant | 0.7408 | 0.2351  | 3.15 | 0.003 |

Differencing: 1 regular difference

Number of observations: Original series 65, after differencing 64

Residuals: SS = 219.223 (backforecasts excluded)

MS = 3.536 DF = 62

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag        | 12    | 24    | 36    | 48    |
|------------|-------|-------|-------|-------|
| Chi-Square | 11.8  | 29.1  | 37.1  | 48.1  |
| DF         | 10    | 22    | 34    | 46    |
| P-Value    | 0.297 | 0.141 | 0.328 | 0.389 |

Forecasts from period 65

95 Percent Limits

| Period | Forecast | Lower   | Upper   | Actual |
|--------|----------|---------|---------|--------|
| 66     | 289.948  | 286.262 | 293.634 |        |

**ARIMA(0, 1, 1): Model for Transportation Index**

Final Estimates of Parameters

| Type     | Coef    | SE Coef | T     | P     |
|----------|---------|---------|-------|-------|
| MA 1     | -0.2913 | 0.1226  | -2.38 | 0.021 |
| Constant | 1.0381  | 0.3035  | 3.42  | 0.001 |

Differencing: 1 regular difference

Number of observations: Original series 65, after differencing 64

Residuals: SS = 219.347 (backforecasts excluded)

MS = 3.538 DF = 62

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag        | 12    | 24    | 36    | 48    |
|------------|-------|-------|-------|-------|
| Chi-Square | 11.6  | 32.0  | 41.0  | 51.4  |
| DF         | 10    | 22    | 34    | 46    |
| P-Value    | 0.310 | 0.077 | 0.189 | 0.270 |

Forecasts from period 65

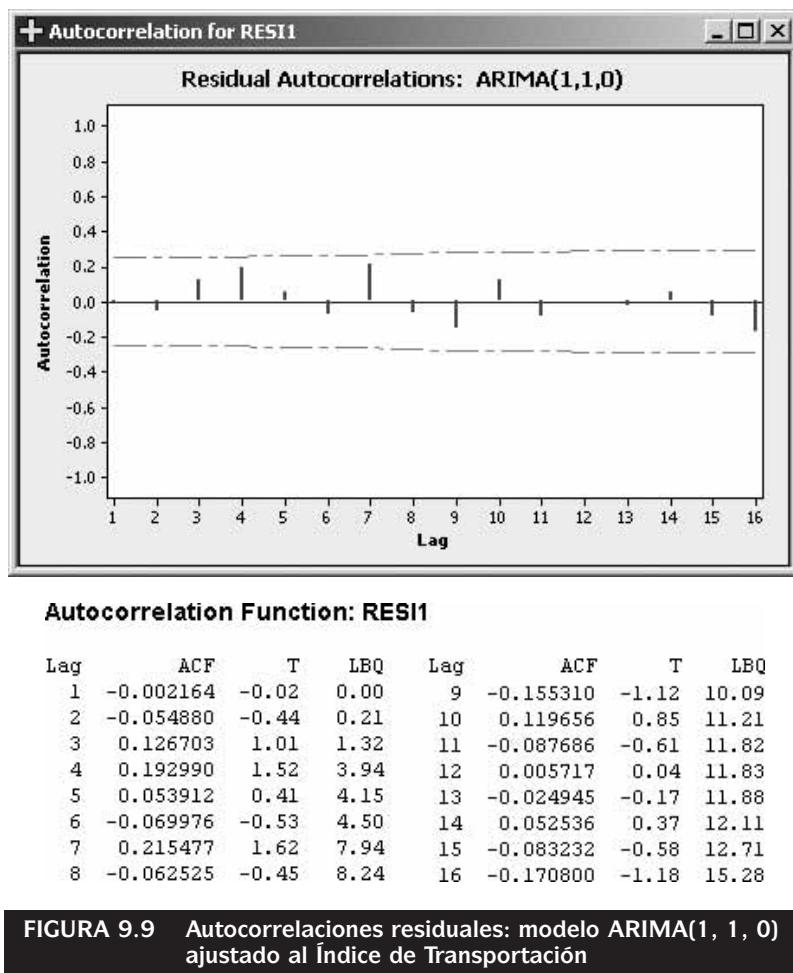
95 Percent Limits

| Period | Forecast | Lower   | Upper   | Actual |
|--------|----------|---------|---------|--------|
| 66     | 290.053  | 286.366 | 293.740 |        |

de manera que  $\phi_0 = .741$  y  $\phi_1 = .284$  la ecuación de pronóstico se transforma en

$$\begin{aligned}\hat{Y}_{66} &= Y_{65} + .741 + .284(Y_{65} - Y_{64}) \\ &= 288.57 + .741 + .284(288.57 - 286.23) = 289.947\end{aligned}$$

El pronóstico tiene correspondencia con el resultado que se muestra en la tabla 9.4. El intervalo de predicción para el valor real del periodo 66 calculado por Minitab es (286.3, 293.6).



**FIGURA 9.9 Autocorrelaciones residuales: modelo ARIMA(1, 1, 0) ajustado al Índice de Transportación**

#### Ejemplo 9.4

Jim White, analista de Atron Corporation, tuvo una serie de tiempo de lecturas de un proceso que necesitaba ser pronosticado. Los datos se muestran en la tabla 9.5. Las lecturas se grafican en la figura 9.11. Él creyó que la metodología Box-Jenkins podría funcionar mejor para estos datos.

Jim comenzó a identificar un modelo tentativo al ver la gráfica de los datos y las autocorrelaciones de la muestra que se ilustran en la figura 9.11. Aparentemente, la serie de tiempo de las lecturas varía en torno a un nivel fijo de alrededor de 80, y las autocorrelaciones se desvaneцен rápidamente hacia cero. Concluyó que la serie de tiempo es estacionaria.

El primer coeficiente de autocorrelación de la muestra ( $-0.53$ ) es significativamente distinto de cero en el nivel de 5%, debido a que permanece fuera del rango.

$$0 \pm 2 \frac{1}{\sqrt{n}} = 0 \pm 2 \frac{1}{\sqrt{75}} = 0 \pm 2(0.115) = 0 \pm .23$$

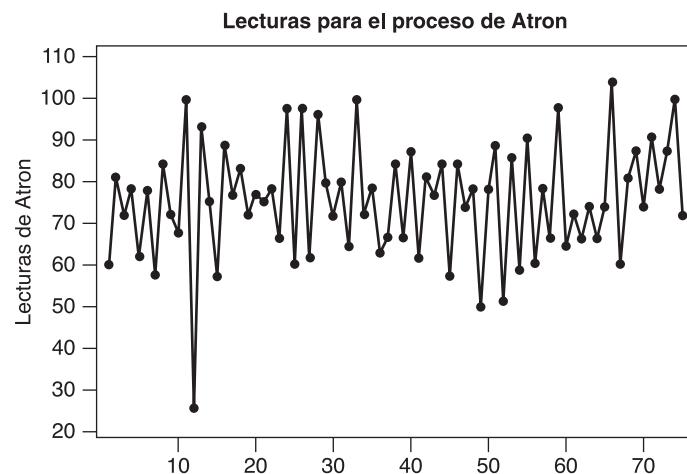
La autocorrelación en el retraso 2 es casi significativa en el nivel de 5% y es de signo opuesto a partir de la autocorrelación del retraso 1  $r_1$ . Las demás autocorrelaciones son pequeñas y correctas dentro de sus límites de error individuales. El patrón de las autocorrelaciones sugiere un modelo AR(1) [vea la figura 9.2(b)] o quizás un modelo MA(2), cualquiera de ellos, si las autocorrelaciones se cortan (equivalentes a cero) después del retraso 2. Jim decidió examinar las autocorrelaciones parciales de la muestra que se ilustran en la figura 9.12.

Asimismo, observó que la primera autocorrelación parcial ( $-0.53$ ) fue significativamente distinta de cero, pero ninguna de las otras autocorrelaciones parciales se acercó a la significancia.<sup>8</sup> Jim

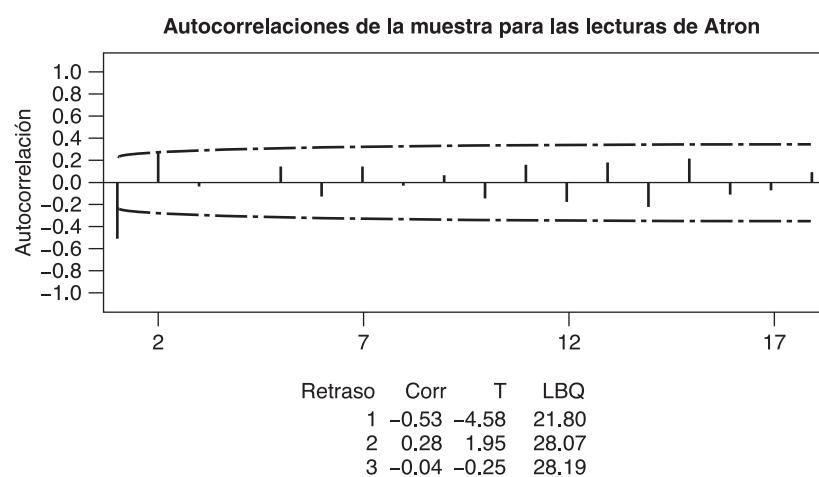
<sup>8</sup> El valor de la primera autocorrelación parcial es igual al valor de la primera autocorrelación (ambas son  $-0.53$ ). Ésta es una característica de un proceso autorregresivo de primer orden.

**TABLA 9.5** Lecturas para el proceso Atron  
para el ejemplo 9.4 (léase por columnas)

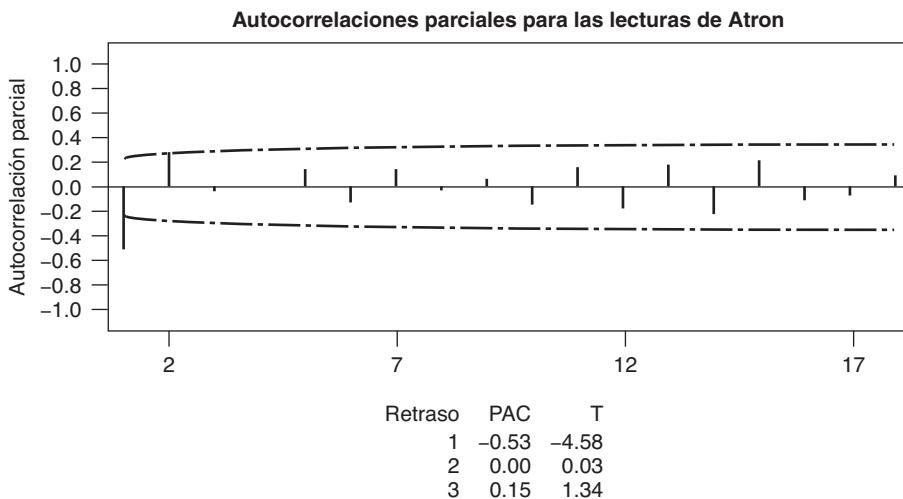
|      |      |      |      |      |      |       |      |
|------|------|------|------|------|------|-------|------|
| 60.0 | 99.0 | 75.0 | 79.5 | 61.5 | 88.5 | 72.0  | 90.0 |
| 81.0 | 25.5 | 78.0 | 64.5 | 81.0 | 51.0 | 66.0  | 78.0 |
| 72.0 | 93.0 | 66.0 | 99.0 | 76.5 | 85.5 | 73.5  | 87.0 |
| 78.0 | 75.0 | 97.5 | 72.0 | 84.0 | 58.5 | 66.0  | 99.0 |
| 61.5 | 57.0 | 60.0 | 78.0 | 57.0 | 90.0 | 73.5  | 72.0 |
| 78.0 | 88.5 | 97.5 | 63.0 | 84.0 | 60.0 | 103.5 |      |
| 57.0 | 76.5 | 61.5 | 66.0 | 73.5 | 78.0 | 60.0  |      |
| 84.0 | 82.5 | 96.0 | 84.0 | 78.0 | 66.0 | 81.0  |      |
| 72.0 | 72.0 | 79.5 | 66.0 | 49.5 | 97.5 | 87.0  |      |
| 67.8 | 76.5 | 72.0 | 87.0 | 78.0 | 64.5 | 73.5  |      |



**FIGURA 9.10** Lecturas para el proceso de Atron



**FIGURA 9.11** Función de autocorrelación de la muestra para las lecturas de Atron



**FIGURA 9.12 Autocorrelaciones parciales de la muestra para las lecturas de Atron**

piensa que las autocorrelaciones y las autocorrelaciones parciales sugieren un modelo AR(1) [o de forma equivalente, en un ARIMA(1, 0, 0)], pero para mayor seguridad, prefiere adaptar un modelo MA(2) [ARIMA (0, 0, 2)]. Si ambos modelos son adecuados, él tomará una decisión por medio del principio de parsimonia.

La tabla 9.6 muestra los resultados obtenidos al utilizar Minitab para ajustar los modelos AR(1) y MA(2) a las lecturas del proceso de Atron. Se incluye un término constante en ambos modelos para permitir que las lecturas varíen en torno a un nivel distinto de cero.<sup>9</sup>

Al parecer, ambos modelos ajustan correctamente los datos. Los coeficientes estimados son significativamente distintos de cero. Los errores cuadráticos medios son similares

$$\text{MA}(2): s^2 = 135.1$$

$$\text{AR}(1): s^2 = 137.9$$

Los pronósticos anticipados de uno o dos pasos obtenidos de ambos modelos son distintos, pero los pronósticos anticipados de tres pasos (periodo 78) son casi iguales. Para obtener un origen de pronóstico fijo, los pronósticos de los procesos estacionarios eventualmente serán iguales al nivel del medio estimado. En este caso, se estima que el nivel medio será de aproximadamente  $\hat{\mu} = 75$  para ambos modelos.

La estadística Ljung-Box  $Q_m$  (Box-Pierce modificada) no es significativa para el conjunto de los retrasos  $m = 12, 24, 36$  y  $48$  para uno u otro modelo. Las autocorrelaciones residuales para el modelo AR(1) se muestran en la figura 9.13. Las autocorrelaciones residuales individuales son pequeñas y correctas dentro de sus límites de error. La función de autocorrelación residual para el modelo MA(2) es similar. Para cada uno de los modelos, no hay razón para creer que los errores no son aleatorios.

Debido a que el modelo AR(1) tiene dos parámetros (incluyendo el término constante) y el modelo MA(2) tiene tres parámetros (incluyendo el término constante), Jim apela al principio de parsimonia y decide utilizar el modelo AR(1) más sencillo para pronosticar las lecturas futuras. La ecuación de pronóstico de AR(1) es<sup>10</sup>

$$\hat{Y}_t = 115.842 + (-.538)Y_{t-1} = 115.842 - .538Y_{t-1}$$

De manera que para el periodo 76,

$$\hat{Y}_{76} = 115.842 - .538Y_{75} = 115.842 - .538(72) = 77.11$$

<sup>9</sup> Si los datos se han expresado como desviaciones a partir de la media de la muestra, no se requerirá un término constante en cualquiera de los modelos.

<sup>10</sup> El término de error  $\varepsilon_t$  no se considera debido a que, para el pronóstico  $\hat{Y}_t$ , la mejor opción de  $\varepsilon_t$  equivale a cero.

**TABLA 9.6 Resultados en Minitab para los modelos MA(2) y AR(1) para las lecturas de Atron**

**ARIMA(0, 0, 2) Model: Atron readings**

Final Estimates of Parameters

| Type     | Coef    | SE Coef | T     | P     |
|----------|---------|---------|-------|-------|
| MA 1     | 0.5667  | 0.1107  | 5.12  | 0.000 |
| MA 2     | -0.3560 | 0.1146  | -3.11 | 0.003 |
| Constant | 75.410  | 1.061   | 71.08 | 0.000 |
| Mean     | 75.410  | 1.061   |       |       |

Number of observations: 75

Residuals: SS = 9724.97 (backforecasts excluded)  
MS = 135.07 DF = 72

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

|            |       |       |       |       |
|------------|-------|-------|-------|-------|
| Lag        | 12    | 24    | 36    | 48    |
| Chi-Square | 7.0   | 23.8  | 31.8  | 46.9  |
| DF         | 9     | 21    | 33    | 45    |
| P-Value    | 0.638 | 0.305 | 0.526 | 0.394 |

Forecasts from period 75

95 Percent Limits

| Period | Forecast | Lower  | Upper   | Actual |
|--------|----------|--------|---------|--------|
| 76     | 80.648   | 57.864 | 103.431 |        |
| 77     | 78.169   | 51.982 | 104.356 |        |
| 78     | 75.410   | 47.996 | 102.825 |        |

**ARIMA(1, 0, 0) Model: Atron readings**

Final Estimates of Parameters

| Type     | Coef    | SE Coef | T     | P     |
|----------|---------|---------|-------|-------|
| AR 1     | -0.5379 | 0.0986  | -5.46 | 0.000 |
| Constant | 115.842 | 1.356   | 85.44 | 0.000 |
| Mean     | 75.3269 | 0.8817  |       |       |

Number of observations: 75

Residuals: SS = 10065.2 (backforecasts excluded)  
MS = 137.9 DF = 73

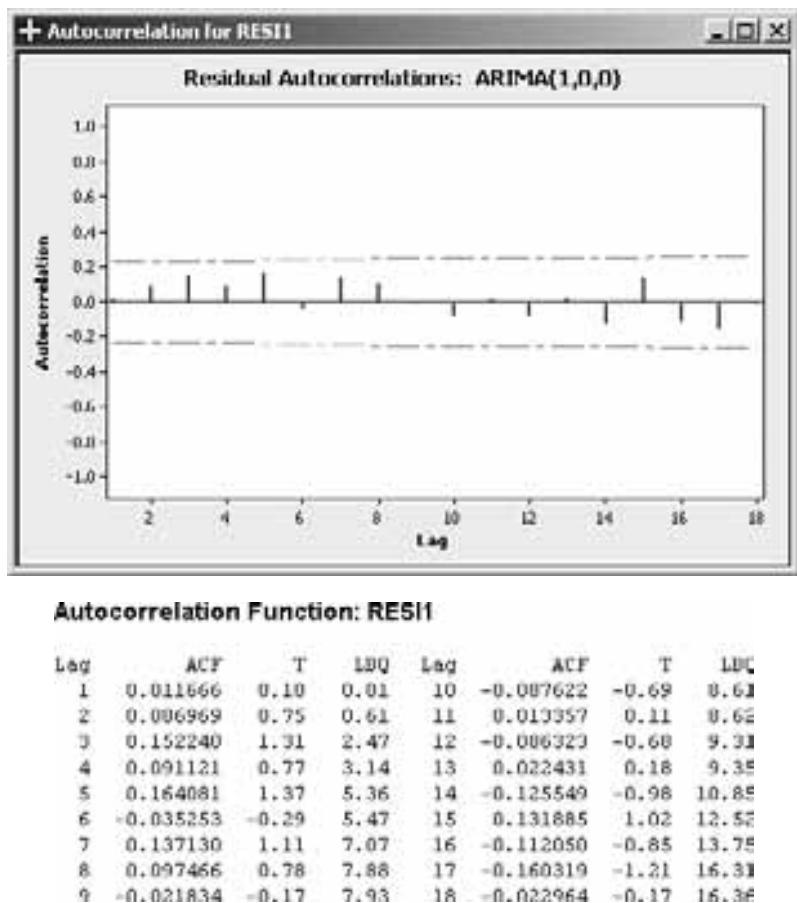
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

|            |       |       |       |       |
|------------|-------|-------|-------|-------|
| Lag        | 12    | 24    | 36    | 48    |
| Chi-Square | 9.3   | 29.9  | 37.3  | 58.3  |
| DF         | 10    | 22    | 34    | 46    |
| P-Value    | 0.503 | 0.121 | 0.319 | 0.105 |

Forecasts from period 75

95 Percent Limits

| Period | Forecast | Lower  | Upper   | Actual |
|--------|----------|--------|---------|--------|
| 76     | 77.116   | 54.097 | 100.136 |        |
| 77     | 74.364   | 48.227 | 100.502 |        |
| 78     | 75.845   | 48.872 | 102.817 |        |



**FIGURA 9.13 Función de autocorrelación para los residuales del modelo AR(1)**

Además, el pronóstico anticipado de dos pasos es

$$\hat{Y}_{77} = 115.842 - .538\hat{Y}_{76} = 115.842 - .538(77.11) = 74.36$$

Jim estuvo satisfecho de ver que estos resultados concordaron con el resultado obtenido en Minitab.

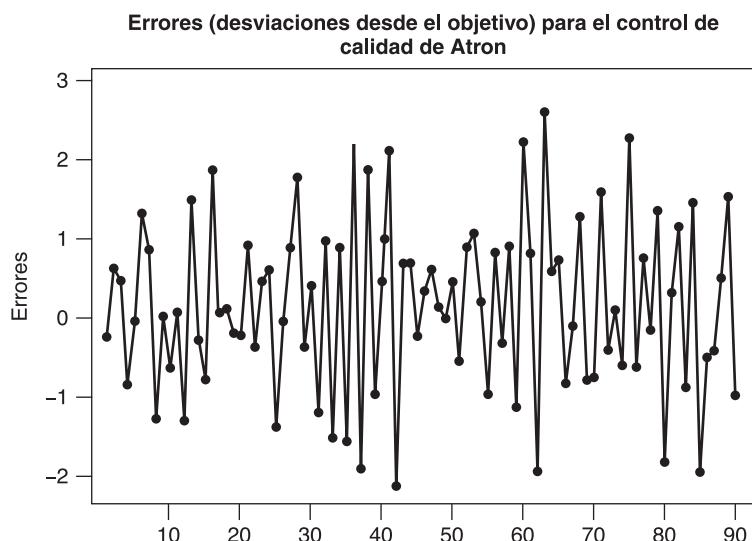
#### Ejemplo 9.5

Jim White estaba contento con los resultados de sus pronósticos para la serie de tiempo de las lecturas que se muestran en la tabla 9.5. Él decidió utilizar la metodología Box-Jenkins para intentar pronosticar los errores (desviaciones desde el objetivo) que resultan del control de calidad de un proceso de manufactura que está bajo su administración. Los datos se proporcionan en la tabla 9.7 y la serie de tiempo de errores se grafica en la figura 9.14.

Jim comenzó el proceso de identificación al examinar la gráfica de la serie de tiempo de errores y evaluar las autocorrelaciones y autocorrelaciones parciales que se muestran en las figuras 9.15 y 9.16. La gráfica de la serie de tiempo y las funciones de autocorrelación indican que la serie es estacionaria. Debido a que existe una autocorrelación significativa de -0.50 en el retraso 1 y las autocorrelaciones en los demás retrasos son pequeñas y correctas dentro de sus límites de error, parece que se cortan las autocorrelaciones de la muestra después del retraso 1. A partir de una autocorrelación parcial significativa en el retraso 1, las tres primeras autocorrelaciones parciales de la muestra son negativas y se deslizan hacia cero. Las conductas de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra se parecen mucho a las cantidades teóricas de un proceso MA(1) [ARIMA(0, 0, 1)] que se ilustra en la figura 9.3 (a). Jim está convencido de que su serie podrá representarse mediante un modelo MA(1).

**TABLA 9.7** Errores para el control de calidad de Atron (léase por columnas)

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| -.23  | -.20  | -1.93 | -.97  | .10   |
| .63   | -.21  | 1.87  | .83   | -.62  |
| .48   | .91   | -.97  | -.33  | 2.27  |
| -.83  | -.36  | .46   | .91   | -.62  |
| -.03  | .48   | 2.12  | -1.13 | .74   |
| 1.31  | .61   | -2.11 | 2.22  | -.16  |
| .86   | -1.38 | .70   | .80   | 1.34  |
| -1.28 | -.04  | .69   | -1.95 | -1.83 |
| 0     | .90   | -.24  | 2.61  | .31   |
| -.63  | 1.79  | .34   | .59   | 1.13  |
| .08   | -.37  | .60   | .71   | -.87  |
| -1.30 | .40   | .15   | -.84  | 1.45  |
| 1.48  | -1.19 | -.02  | -.11  | -1.95 |
| -.28  | .98   | .46   | 1.27  | -.51  |
| -.79  | -1.51 | -.54  | -.80  | -.41  |
| 1.86  | .90   | .89   | -.76  | .49   |
| .07   | -1.56 | 1.07  | 1.58  | 1.54  |
| .09   | 2.18  | .20   | -.38  | -.96  |

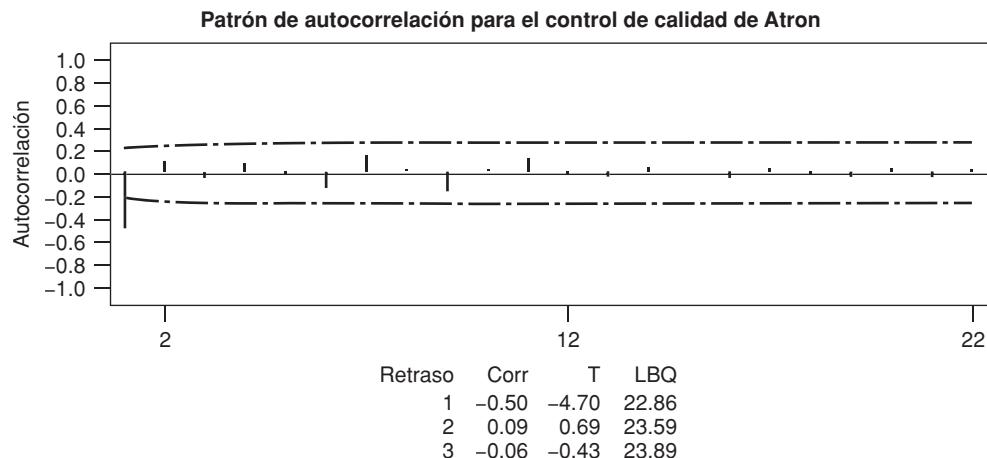


**FIGURA 9.14** Errores (desviaciones desde el objetivo) para el control de calidad de Atron

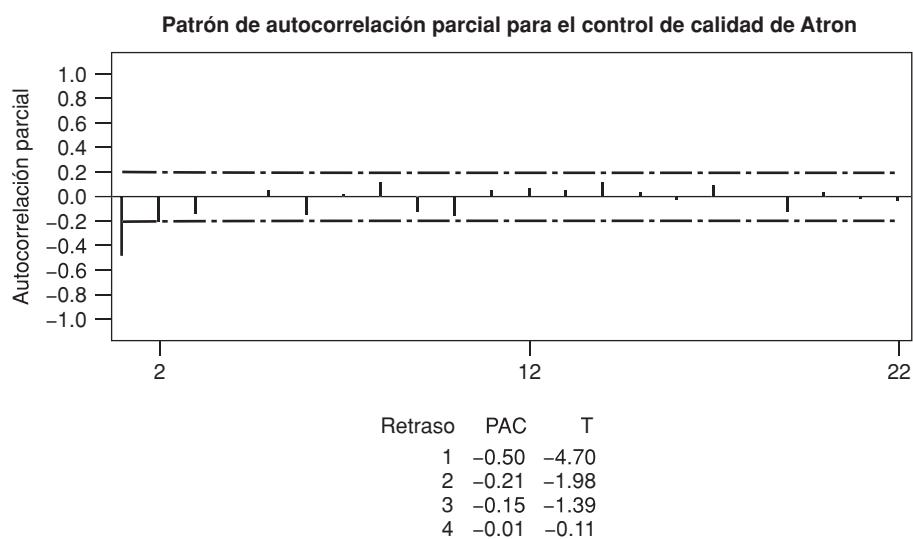
Jim utilizó Minitab para ajustar un modelo MA(1) a sus datos. Los resultados se muestran en la tabla 9.8. Los parámetros en el modelo MA(1) se estiman en  $\hat{\mu} = .1513$  y  $\hat{\omega}_1 = .5875$ . Cada uno es significativamente distinto de cero. Las autocorrelaciones residuales que se muestran en la figura 9.17 y la estadística chi cuadrada de Ljung-Box (Box-Pierce modificada) indican que los errores son aleatorios.

La ecuación del pronóstico de MA(1) es

$$\hat{Y}_t = \hat{\mu} - \hat{\omega}_1 \varepsilon_{t-1} = .1513 - .5875 \varepsilon_{t-1}$$



**FIGURA 9.15 Autocorrelaciones de la muestra para el control de calidad de Atron**



**FIGURA 9.16 Autocorrelaciones parciales de la muestra para el control de calidad de Atron**

donde  $\varepsilon_{t-1}$  se estima por el residual correspondiente  $e_{t-1}$ . Para pronosticar el error (desviaciones del objetivo) para el periodo 91, se requiere del residual para el periodo 90,  $e_{90} = -.4804$ . Jim calcula

$$\hat{Y}_{91} = .1513 - .5875(-.4804) = .4335$$

El pronóstico de Jim para el error del control de calidad en el periodo 92 es simplemente la media estimada de la serie debido a que, en el origen del pronóstico  $t = 90$ , la mejor opción para el término de error en el período 91  $\varepsilon_{91}$  es cero. De esta forma

$$\hat{Y}_{92} = .1513 - .5875(0) = .1513$$

Jim estaba satisfecho al ver que sus pronósticos calculados concordaban con los obtenidos en Minitab.

**TABLA 9.8 Resultados en Minitab para el modelo de MA(1) para el control de calidad de Atron**

**ARIMA(0, 0, 1) Model: Atron Quality Control**

Final Estimates of Parameters

| Type     | Coef    | SE Coef | T    | P     |
|----------|---------|---------|------|-------|
| MA 1     | 0.5875  | 0.0864  | 6.80 | 0.000 |
| Constant | 0.15129 | 0.04022 | 3.76 | 0.000 |
| Mean     | 0.15129 | 0.04022 |      |       |

Number of observations: 90

Residuals: SS = 74.4933 (backforecasts excluded)

MS = 0.8465 DF = 88

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag        | 12    | 24    | 36    | 48    |
|------------|-------|-------|-------|-------|
| Chi-Square | 9.1   | 10.8  | 17.3  | 31.5  |
| DF         | 10    | 22    | 34    | 46    |
| P-Value    | 0.524 | 0.977 | 0.992 | 0.950 |

Forecasts from period 90

| Period | Forecast | 95 Percent Limits |         |        |
|--------|----------|-------------------|---------|--------|
|        |          | Lower             | Upper   | Actual |
| 91     | 0.43350  | -1.37018          | 2.23719 |        |
| 92     | 0.15129  | -1.94064          | 2.24322 |        |

**TABLA 9.9 Errores para el control de calidad de Ed Jones para el ejemplo 9.6 (léase por columnas)**

|       |       |       |       |      |
|-------|-------|-------|-------|------|
| .77   | 1.04  | -2.46 | -.73  | -.23 |
| .33   | 1.02  | -.37  | .10   | 1.05 |
| 2.15  | -2.03 | .80   | -1.47 | -.66 |
| 2.50  | -2.54 | .49   | -.89  | .25  |
| 1.36  | -.23  | .50   | -.53  | -.63 |
| .48   | .49   | .07   | -.20  | .91  |
| 2.05  | -.87  | 1.92  | -.70  | -.21 |
| -1.46 | .61   | 1.00  | -.27  | .24  |
| -1.13 | .20   | 2.16  | .39   | .05  |
| -2.85 | .98   | .04   | -.07  | .85  |
| -2.67 | .78   | 1.91  | .89   | 1.55 |
| -2.71 | .80   | .43   | .37   | .40  |
| -1.30 | .86   | -.32  | -.75  | 1.82 |
| -.88  | 1.72  | -.48  | -1.24 | .81  |
| -.07  | .15   | -.13  | -.62  | .28  |
| -1.47 | -1.15 | -2.26 | -.54  | 1.06 |

**Ejemplo 9.6**

Jim White, analista de Atron Corporation, estaba complacido con los pronósticos de los errores de control de calidad que desarrolló con el empleo de la metodología Box-Jenkins, que se analizó en el ejemplo 9.5. Jim encontró a su viejo amigo Ed Jones en una conferencia y le comentó acerca de su éxito. Ed había estado luchando con una aplicación semejante y decidió aplicar la metodología Box-Jenkins. Los datos de Ed se proporcionan en la tabla 9.9 y la gráfica para la serie de tiempo de los datos se muestra en la figura 9.18.

La gráfica de la serie y las autocorrelaciones de la muestra, que se ilustran la figura 9.19, sugieren que la serie de los errores del control de calidad es estacionaria. Aparentemente, los errores varían en torno a un nivel fijo de cero y las autocorrelaciones se desvanecen rápidamente.

Ed observó que las primeras dos autocorrelaciones eran significativamente distintas de cero y, quizás más importante, que las autocorrelaciones para los primeros retrasos de tiempo caen

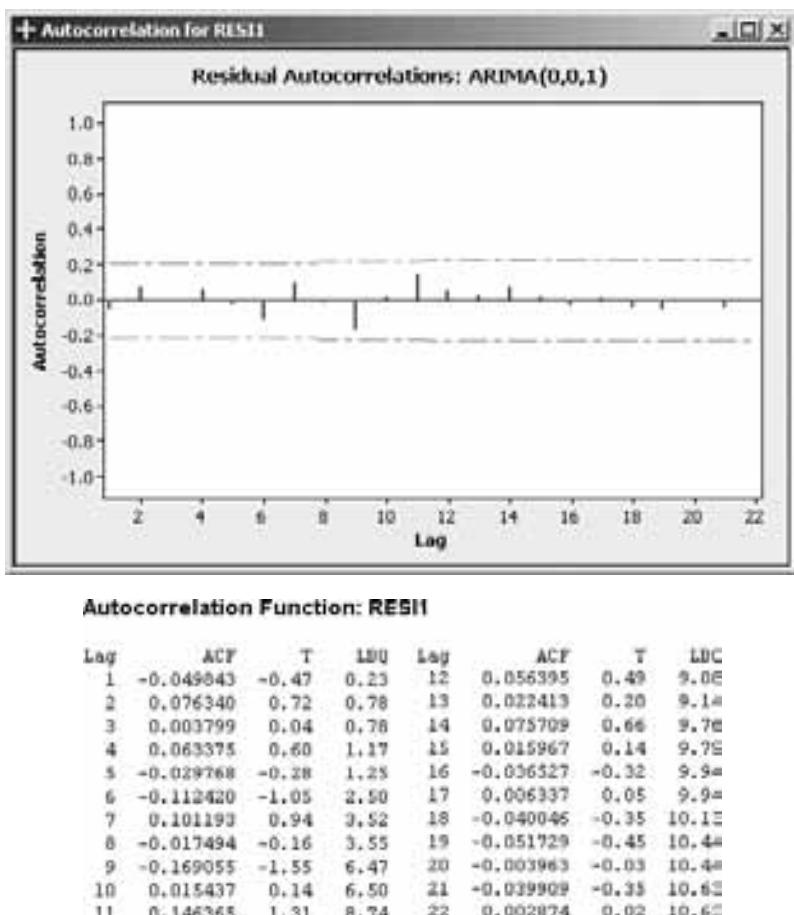


FIGURA 9.17 Función de autocorrelación para los residuales del modelo MA(1)

hacia cero de manera muy semejante a la del patrón teórico de un proceso AR(1) [vea la figura 9.2(a)]. Ed también examinó las autocorrelaciones parciales de la muestra que se presentan en la figura 9.20. Como lo sospechó, las autocorrelaciones parciales parecían cortarse después de una autocorrelación parcial significativa en el retraso 1.<sup>11</sup> Si se consideran en conjunto, los patrones de la autocorrelación de la muestra y la autocorrelación parcial de la muestra fueron coherentes con un proceso AR(1), y Ed se sintió a gusto al ajustar un modelo AR(1) {ARIMA (1, 0, 0)} a su serie de tiempo de los errores (desviaciones del objetivo).

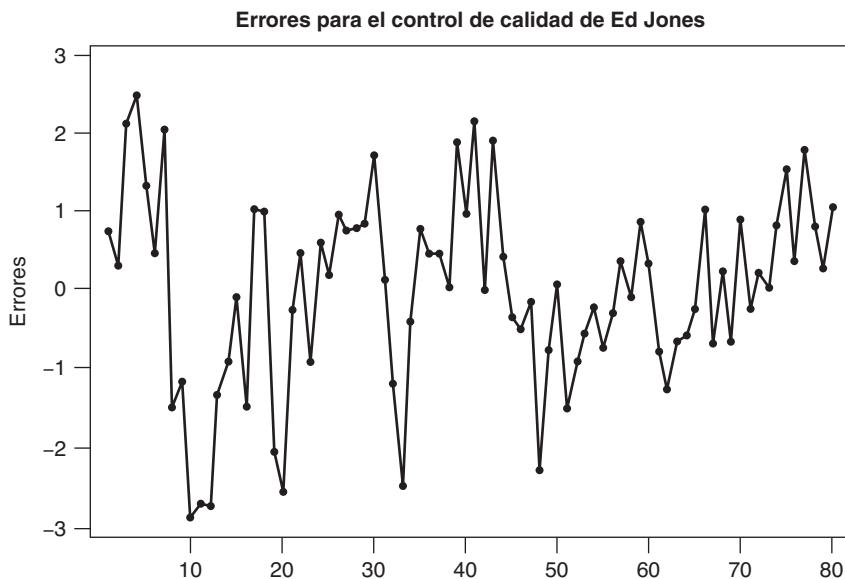
El resultado obtenido en Minitab para el intento inicial de Ed dentro de la metodología Box-Jenkins se proporciona en la tabla 9.10. Debido a que la media de la muestra de la serie de error es extremadamente pequeña (cercana a cero), en comparación con su desviación estándar, Ed no incluyó un término constante en su modelo AR(1).

Se estima que el parámetro en el modelo AR(1) sea  $\hat{\phi}_1 = .501$  y significativamente distinto de cero  $t = 5.11$ . El error cuadrático medio de los residuales es  $s^2 = 1.0998$ . Las gráficas de los residuales (que no se muestran), la estadística Chi cuadrada de Ljung-Box y las autocorrelaciones residuales que se muestran en la figura 9.21, sugieren que el modelo AR(1) es adecuado. No hay razón para dudar de las suposiciones usuales de los términos de error.

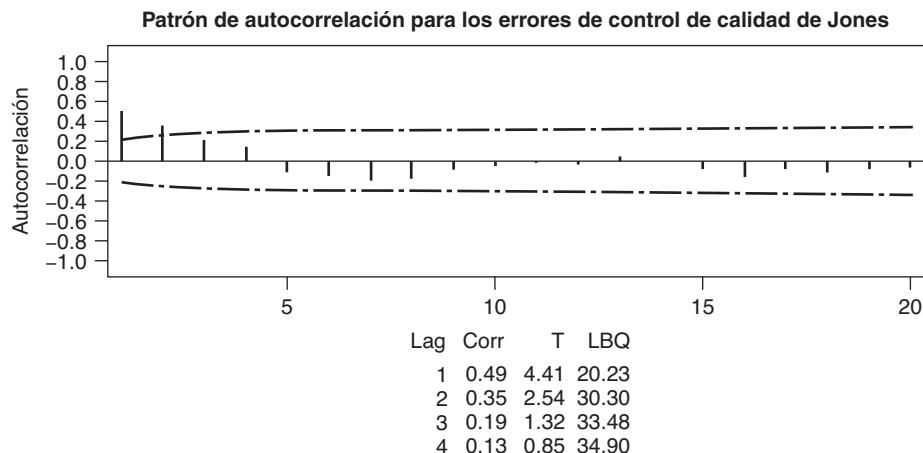
La ecuación AR(1) para realizar pronósticos es

$$\hat{Y}_t = .501 Y_{t-1}$$

<sup>11</sup> Ed atribuyó la autocorrelación parcial significativa en el retraso cinco al error de muestreo, debido a que no pudo encontrar una razón física de por qué los errores del control de calidad de cinco períodos separados debían estar correlacionados.



**FIGURA 9.18** Errores (desviaciones desde el objetivo) para el control de calidad de Ed Jones.



**FIGURA 9.19** Autocorrelaciones de la muestra para el control de calidad de Ed Jones

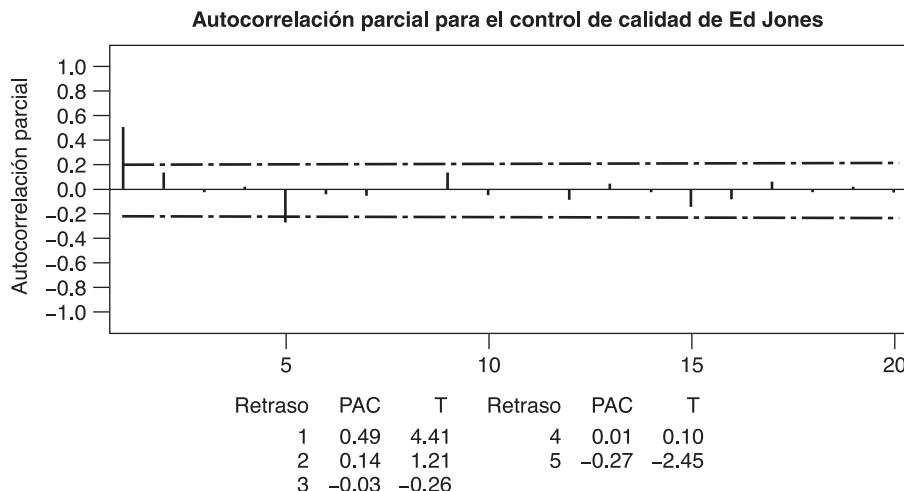
de manera que los pronósticos para los períodos 81 y 82 son

$$\hat{Y}_{81} = .501Y_{80} = .501(1.06) = .531$$

$$\hat{Y}_{82} = .501\hat{Y}_{81} = .501(.531) = .266$$

Los pronósticos que calculó Ed concuerdan con los pronósticos de uno o dos pasos anticipados generados por Minitab.

Debido que ésta fue su primera experiencia con la metodología Box-Jenkins, Ed decidió ajustar un modelo ligeramente más complejo para ver si los resultados confirmarían su elección de un modelo AR(1) para sus datos. Entonces, decidió añadir un parámetro adicional y ajustar



**FIGURA 9.20 Autocorrelaciones parciales de la muestra para el control de calidad de Ed Jones.**

**TABLA 9.10 Resultados de Minitab para el modelo AR(1) para el control de calidad de Ed Jones**

```

ARIMA(1, 0, 0) Model: Ed Jones' Quality Control Errors

Final Estimates of Parameters
Type      Coef    SE Coef      T      P
AR 1    0.5008   0.0980   5.11  0.000

Number of observations: 80
Residuals: SS = 86.8808 (backforecasts excluded)
            MS = 1.0998 DF = 79

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag          12      24      36      48
Chi-Square   10.7    19.5    36.2   44.2
DF           11      23      35      47
P-Value      0.468   0.669   0.410   0.591

Forecasts from period 80
         95 Percent Limits
Period   Forecast      Lower      Upper   Actual
81        0.53088   -1.52498   2.58673
82        0.26588   -2.03340   2.56515

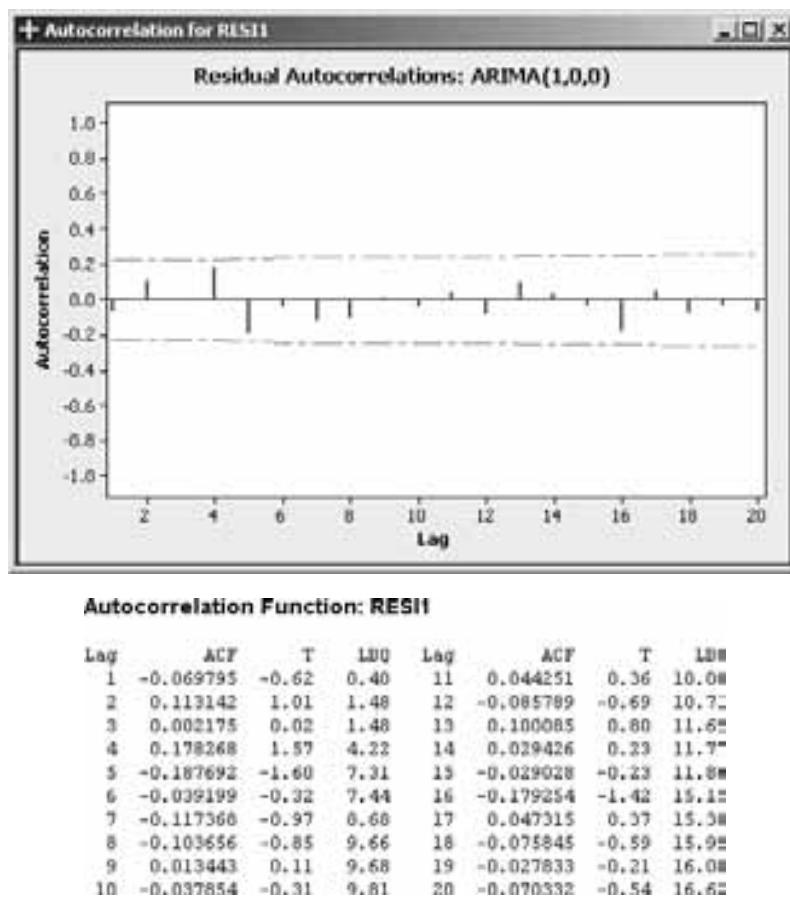
```

su modelo ARMA(1, 1) [ARIMA (1, 0, 1)] para los errores del control de calidad. Pensó que si el modelo AR(1) era correcto, el parámetro de promedio móvil en el modelo ARMA(1, 1) no debería ser significativo.

La tabla 9.11 ilustra los resultados del modelo ARMA(1, 1) ajustado a la serie de Ed, quien estuvo satisfecho de ver que el parámetro MA(1) no era significativamente distinto de cero  $t = 1.04$ , lo cual implica que este término no es necesario. Desde luego, debido a que es un modelo ligeramente más general que el modelo AR(1), aún representa adecuadamente los datos como se indica por  $s^2 = 1.0958$  y la conducta aleatoria de los residuales.

#### Ejemplo 9.7

A Jill Blake, analista de ISC Corporation, se le pidió que desarrollara los pronósticos para los precios de cierre de las acciones de ISC. Por algún tiempo, las acciones languidecieron y su



**FIGURA 9.21 Función de autocorrelación para los residuales del modelo AR(1)**

crecimiento fue mínimo, así que la dirección quería algunas proyecciones para analizarlas con el consejo de administración. Los precios de las acciones de ISC se grafican en la figura 9.22 y se enlistan en la tabla 9.12.

La gráfica de los precios de las acciones sugieren que la serie es estacionaria. Parece que, a partir de la figura 9.22, los precios de las acciones varían en torno a un nivel fijo de aproximadamente 250.

Jill cree que la metodología de Box-Jenkins podría ser útil a su propósito y generar de inmediato las autocorrelaciones y las autocorrelaciones parciales de la muestra para la serie de precios accionarios. Los resultados aparecen en las figuras 9.23 y 9.24.

Jill observa que las autocorrelaciones de la muestra alternan su signo y declinan hacia cero para los retrasos bajos. Las autocorrelaciones parciales de la muestra son semejantes, pero es claro que se cortan (son equivalentes a cero) después del retraso 2. Jill recuerda los patrones posibles para las autocorrelaciones teóricas y las autocorrelaciones parciales para un proceso AR(2) [vea la figura 9.2 (d)] y decide que las correspondientes cantidades de la muestra son coherentes con uno de estos patrones. Ella identifica un modelo AR(2) [ARIMA (2, 0, 0)] para sus datos.

Jill utiliza Minitab para estimar los parámetros en su modelo. Ella incluye un término constante que permite un nivel distinto de cero. Los resultados se muestran en la tabla 9.13.

Los estimados del parámetro son  $\hat{\phi}_0 = 284.9$ ,  $\hat{\phi}_1 = -0.324$  y  $\hat{\phi}_2 = .219$ . El coeficiente estimado  $\hat{\phi}_2$  no es significativo  $t = 1.75$  en el nivel de 5%, pero sí lo es en el nivel de 10%. Jill decide conservar este parámetro en su modelo. El error cuadrático medio de residuales es  $s^2 = 2808$  y las autocorrelaciones residuales que se proporcionan en la figura 9.25 están dentro de sus límites de error estándar. Además, los valores de la estadística Ljung-Box para  $m = 12, 24, 36$  y  $48$  son pequeños. Jill decide que este modelo es adecuado.

Jill utiliza el modelo para generar pronósticos para los períodos 66 y 67, como se indica.

**TABLA 9.11 Resultados de Minitab para el modelo ARIMA(1, 0, 1) para el control de calidad de Ed Jones**

**ARIMA(1, 0, 1) Model: Ed Jones' Errors**

Final Estimates of Parameters

| Type | Coef   | SE Coef | T    | P     |
|------|--------|---------|------|-------|
| AR 1 | 0.6671 | 0.1628  | 4.10 | 0.000 |
| MA 1 | 0.2204 | 0.2121  | 1.04 | 0.302 |

Number of observations: 80

Residuals: SS = 85.4710 (backforecasts excluded)  
MS = 1.0958 DF = 78

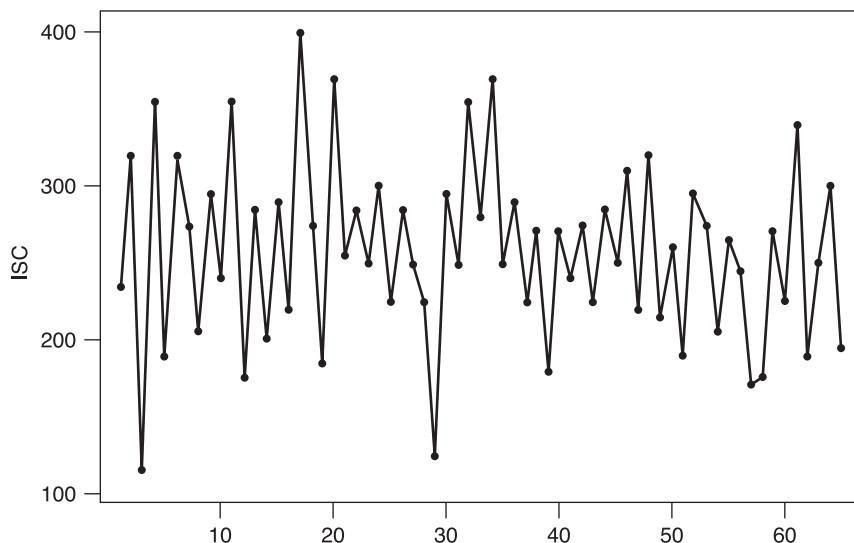
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

| Lag        | 12    | 24    | 36    | 48    |
|------------|-------|-------|-------|-------|
| Chi-Square | 8.8   | 17.5  | 32.2  | 38.7  |
| DF         | 10    | 22    | 34    | 46    |
| P-Value    | 0.547 | 0.737 | 0.555 | 0.770 |

Forecasts from period 80

| Period | Forecast | 95 Percent Limits |         | Actual |
|--------|----------|-------------------|---------|--------|
|        |          | Lower             | Upper   |        |
| 81     | 0.52814  | -1.52399          | 2.58027 |        |
| 82     | 0.35230  | -1.89526          | 2.59985 |        |

**Precios de cierre para las acciones de ISC Corporation**



**FIGURA 9.22 Precios de cierre para las acciones de ISC Corporation**

Desde el origen del pronóstico  $t = 65$ , el pronóstico para el periodo 66 se genera con la ecuación

$$\hat{Y}_t = \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2}$$

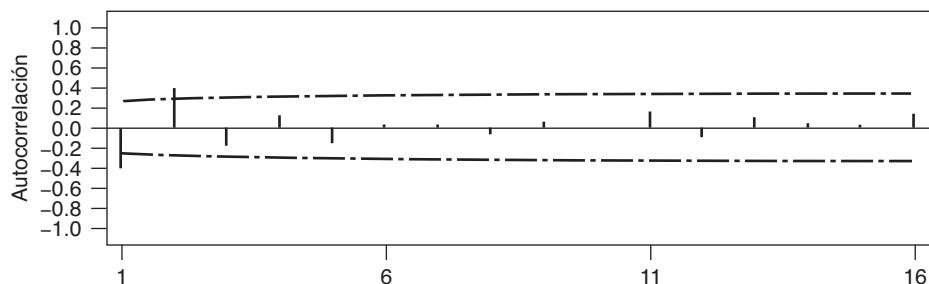
también

$$\begin{aligned}\hat{Y}_{66} &= 284.9 + (-.324)Y_{65} + .219Y_{64} \\ &= 284.9 - .324(195) + .219(300) = 287.4\end{aligned}$$

**TABLA 9.12** Precios de cierre para las acciones de ISC Corporation del ejemplo 9.7 (léase por columnas)

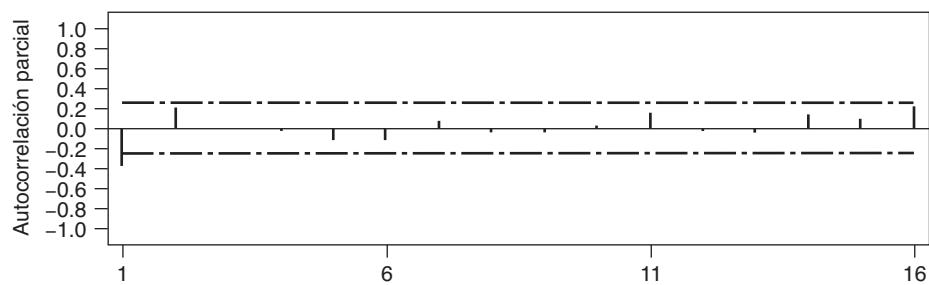
|     |     |     |     |     |
|-----|-----|-----|-----|-----|
| 235 | 200 | 250 | 270 | 275 |
| 320 | 290 | 225 | 240 | 205 |
| 115 | 220 | 125 | 275 | 265 |
| 355 | 400 | 295 | 225 | 245 |
| 190 | 275 | 250 | 285 | 170 |
| 320 | 185 | 355 | 250 | 175 |
| 275 | 370 | 280 | 310 | 270 |
| 205 | 255 | 370 | 220 | 225 |
| 295 | 285 | 250 | 320 | 340 |
| 240 | 250 | 290 | 215 | 190 |
| 355 | 300 | 225 | 260 | 250 |
| 175 | 225 | 270 | 190 | 300 |
| 285 | 285 | 180 | 295 | 195 |

**Patrón de autocorrelación para los precios accionarios de cierre de ISC**



**FIGURA 9.23** Autocorrelaciones de la muestra para los precios accionarios de ISC

**Patrón de autocorrelación parcial para los precios accionarios de cierre de ISC**



**FIGURA 9.24** Autocorrelaciones de la muestra para los precios accionarios de ISC

De manera semejante,

$$\begin{aligned}\hat{Y}_{67} &= 284.9 + (-.324)\hat{Y}_{66} + .219Y_{65} \\ &= 284.9 - .324(287.4) + .219(195) = 234.5\end{aligned}$$

Los resultados concuerdan con los que arrojó Minitab y que se muestran en la tabla 9.13. Para la serie estacionaria, los límites de predicción de 95% son aproximadamente

$$\hat{Y} \pm 2s \quad (9.6)$$

**TABLA 9.13 Resultados de Minitab para el modelo AR(2) para los precios accionarios de ISC**

**ARIMA(2, 0, 0) Model: ISC Closing Stock Prices**

**Final Estimates of Parameters**

| Type     | Coef    | SE Coef | T     | P     |
|----------|---------|---------|-------|-------|
| AR 1     | -0.3243 | 0.1246  | -2.60 | 0.012 |
| AR 2     | 0.2192  | 0.1251  | 1.75  | 0.085 |
| Constant | 284.903 | 6.573   | 43.34 | 0.000 |
| Mean     | 257.828 | 5.949   |       |       |

Number of observations: 65

Residuals: SS = 174093 (backforecasts excluded)

MS = 2808 DF = 62

**Modified Box-Pierce (Ljung-Box) Chi-Square statistic**

| Lag        | 12    | 24    | 36    | 48    |
|------------|-------|-------|-------|-------|
| Chi-Square | 6.3   | 13.3  | 18.2  | 29.1  |
| DF         | 9     | 21    | 33    | 45    |
| P-Value    | 0.707 | 0.899 | 0.983 | 0.969 |

**Forecasts from period 65**

**95 Percent Limits**

| Period | Forecast | Lower   | Upper   | Actual |
|--------|----------|---------|---------|--------|
| 66     | 287.446  | 183.565 | 391.328 |        |
| 67     | 234.450  | 125.244 | 343.656 |        |

donde  $\hat{Y}$  es el pronóstico y  $s$  es la raíz cuadrada del error cuadrado medio. Por ejemplo, los límites de predicción aproximados a 95% para el periodo 66 son

$$287.4 \pm 2\sqrt{2808} = 287.4 \pm 106 \quad \text{o} \quad (181.4, 393.4)$$

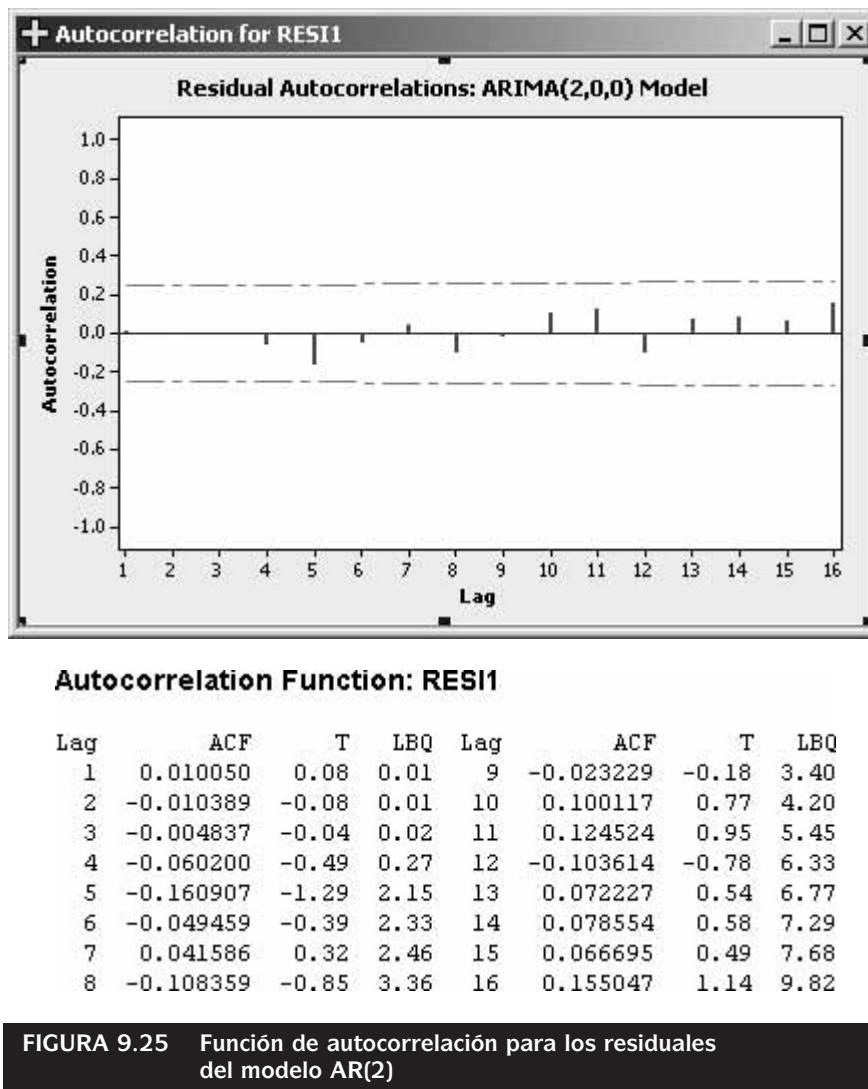
Este intervalo está cerca del intervalo de 95% proporcionado por Minitab para el periodo 66 en la tabla 9.13.

### Comentarios finales

En la modelación ARIMA *no* es una buena práctica incluir parámetros AR y MA para “cubrir todas las posibilidades” sugeridas por las funciones de autocorrelación de la muestra y la autocorrelación parcial de la muestra. Es decir, cuando haya duda comience con un modelo que contenga pocos parámetros, en lugar de muchos. La necesidad de parámetros adicionales será evidente a partir de un examen de las autocorrelaciones residuales y de las autocorrelaciones parciales. Si la conducta de MA es necesaria en apariencia en las autocorrelaciones residuales y las autocorrelaciones parciales, añada un parámetro MA y ajuste el modelo revisado. Si las autocorrelaciones residuales se parecen a las del proceso AR, agregue un término AR y reajuste el modelo.

Los estimados de los mínimos cuadrados de los parámetros autorregresivos y de promedio móvil en los modelos ARIMA, tienden a estar muy correlacionados. Cuando hay más parámetros de los necesarios, esto conduce a una disyuntiva entre demasiados parámetros y los modelos inestables que pueden generar pronósticos deficientes.

Para resumir, se considera una buena práctica comenzar con un pequeño número de parámetros que se justifiquen con claridad y añadir un parámetro cuando sea necesario. Por otra parte, si los parámetros en un modelo ARIMA ajustado no son significativos (según se juzga por sus proporciones  $t$ ), borre un parámetro a la vez y ajuste el modelo. Debido a la elevada correlación que existe entre los parámetros estimados, podría darse el caso de que un parámetro anterior sin significado se vuelva significativo.



### Criterio para la selección de un modelo

Los modelos ARIMA se identifican (seleccionan) al observar una gráfica de la serie y al establecer una correspondencia entre los patrones de autocorrelación de la muestra y la autocorrelación parcial de la muestra con los patrones teóricos conocidos de los procesos ARIMA. Sin embargo, existe cierta subjetividad involucrada en este proceso, y es posible que dos (o más) modelos iniciales pudieran ser coherentes con los patrones de las autocorrelaciones y las autocorrelaciones parciales de la muestra. Además, después de estimarlos y verificarlos, ambos modelos pueden representar los datos de forma adecuada. Si los modelos contienen el mismo número de parámetros, se preferirá el modelo con el error cuadrado medio más pequeño  $s^2$ . (Éste fue el caso de Lynn Stephens en el ejemplo 9.3.) Si los modelos contienen distintos números de parámetros, el principio de parsimonia conduce a la selección del modelo más sencillo. No obstante, es posible que el modelo con más parámetros tenga un error cuadrado medio apreciablemente más pequeño.

Se ha desarrollado una metodología para la selección de los modelos que considera el ajuste del modelo y el número de parámetros. El criterio de la información de Akaike,<sup>12</sup> o

<sup>12</sup> Vea H. Akaike, "A New Look at the Statistical Model Identification", *IEEE Trans. Automatic Control*, AC-19, 1974: 716-723.

*AIC*, es seleccionar el mejor modelo de un grupo de modelos candidatos como aquel que minimiza

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n}r \quad (9.7)$$

donde

$\ln$  = el logaritmo natural

$\hat{\sigma}^2$  = la suma de cuadrado de los residuales dividida entre el número de las observaciones

$n$  = el número de las observaciones (residuales)

$r$  = el número total de los parámetros (más el término constante) en el modelo ARIMA.

El criterio bayesiano de información que desarrolló Schwarz,<sup>13</sup> o *BIC*, selecciona el modelo que minimiza

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n}r \quad (9.8)$$

El segundo término en *AIC* y en *BIC* es un “factor de castigo” por incluir parámetros adicionales en el modelo. Debido a que el criterio *BIC* impone un castigo mayor por el número de parámetros que el criterio *AIC*, el uso de un *BIC* mínimo para la selección del modelo resultará en un modelo cuyo número de parámetros no es mayor que los escogidos por *AIC*. Con frecuencia, los dos criterios producen el mismo resultado.

*AIC* y *BIC* deben verse como procedimientos adicionales que ayudan en la selección del modelo. No deberán emplearse como sustitutos de un examen cuidadoso de las autocorrelaciones de la muestra y las autocorrelaciones parciales.

#### Ejemplo 9.8

En el ejemplo 9.4, Jim White descubrió que los dos modelos ARIMA parecen proporcionar una descripción adecuada de las lecturas del proceso de Atron Corporation. Uno fue el modelo AR(1) con sus parámetros estimados  $r = 2$  (incluyendo un término constante) y  $\hat{\sigma}^2 = 10,065/75 = 134.2$ . El segundo fue un modelo MA(2) con parámetros  $r = 3$  (incluyendo un término constante) y  $\hat{\sigma}^2 = 9,725/75 = 129.7$ . Jim calculó los criterios *AIC* y *BIC* como se muestra continuación:

$$\begin{aligned} \text{AR(1):} \quad AIC &= \ln(134.2) + \frac{2}{75}(2) = 4.8993 + .0533 = 4.9526 \\ &BIC = \ln(134.2) + \frac{\ln(75)}{75}(2) = 4.8993 + .1151 = 5.0144 \\ \text{MA(2):} \quad AIC &= \ln(129.7) + \frac{2}{75}(3) = 4.8652 + .0800 = 4.9452 \\ &BIC = \ln(129.7) + \frac{\ln(75)}{75}(3) = 4.8652 + .1727 = 5.0379 \end{aligned}$$

Tanto *AIC* como *BIC* proporcionan resultados que están en conflicto. *AIC* es más pequeño para el modelo MA(2), mientras que *BIC* es más pequeño para el modelo AR(1). Jim recuerda que *BIC* no selecciona un modelo con más parámetros que *AIC*, debido a que hay un castigo mayor para los parámetros adicionales. Jim estaba contento al ver que su selección inicial para el modelo AR(1), con base en el principio de parsimonia, estaba apoyado por el criterio *BIC*.

---

<sup>13</sup> Vea G. Schwarz, “Estimating the Dimension of a Model”, *Annals of Statistics* 6, 1978, pp. 461-464.

## Modelos para datos estacionales

Los datos estacionales tienen un patrón distintivo que se repite cada año. Las ventas mensuales de Cavanaugh Company, que se muestran en la figura 5.1 (véase la página 160), siempre alcanzan su máximo a finales del otoño (octubre y noviembre). Para los datos mensuales con patrón estacional anual, las observaciones para los mismos meses en distintos años deberán estar correlacionadas. Es decir, determinado enero deberá ser semejante al enero del año siguiente; febrero deberá ser muy semejante al febrero que sigue y así sucesivamente. De esta forma, no sólo las observaciones dentro de un año están relacionadas entre sí (correlacionadas), sino también las relaciones entre los años guardan relación (correlación). Si la longitud del periodo estacional es  $S$ , de forma que  $S = 12$  para los datos mensuales y  $S = 4$  para los datos trimestrales, las autocorrelaciones y autocorrelaciones parciales para los procesos estacionales no equivalen a cero en los retrasos bajos (en una asociación de un año) y de los retrasos que son múltiplos del periodo estacional  $S$  (asociación que se presenta entre un año). Las interpretaciones de las autocorrelaciones y de las autocorrelaciones parciales en los retrasos estacionales son iguales que en la interpretación de las autocorrelaciones y las autocorrelaciones parciales en los retrasos bajos. Los modelos estacionales ARIMA contienen términos autorregresivos regulares y de promedio móvil que explican la correlación en los retrasos bajos y los términos autorregresivos estacionales y de promedio móvil que explican la correlación en los retrasos estacionales. Además, para la serie estacional no estacionaria se necesita una diferencia estacional adicional para especificar el modelo por completo.

Los siguientes ejemplos ilustran el desarrollo de los modelos ARIMA estacionales. Para los modelos estacionales se usa la misma estrategia de construcción de modelos que se aplicó para los no estacionales. Una vez establecido el modelo adecuado, se puede utilizar para pronosticar.

### Ejemplo 9.9

Kathy Fisher ha asumido la responsabilidad de pronosticar las ventas para Keytron Corporation. Kathy ha recopilado los datos correspondientes a 115 meses consecutivos de ventas. Los datos comienzan en enero de 1987 y se muestran en la tabla 9.14.

Kathy examinó la serie de tiempo, graficada en la figura 9.26, y observó el patrón estacional pronunciado con una tendencia general ascendente. Entonces, pensó que la serie no era estacional y decidió desarrollar un modelo ARIMA para sus datos.

Kathy comenzó la identificación del modelo al observar las autocorrelaciones de la muestra que aparecen en la figura 9.27. Kathy vio que las autocorrelaciones en los retrasos cortos parecían cortarse después el retraso 1, a pesar de que había un pequeño margen en el retraso 3. Kathy

**TABLA 9.14 Ventas de Keytron Corporation para el ejemplo 9.9 (léase por columnas)**

|        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|
| 1736.8 | 1627.6 | 1895.4 | 1768.0 | 2202.2 | 1708.2 | 2288.0 |
| 1297.4 | 1575.6 | 1822.6 | 1840.8 | 2449.2 | 806.0  | 2275.0 |
| 559.0  | 1682.2 | 2054.0 | 1804.4 | 2090.4 | 2028.0 | 2581.8 |
| 1455.6 | 1710.8 | 1544.4 | 2007.2 | 2184.0 | 2236.0 | 2540.2 |
| 1526.2 | 1853.8 | 600.6  | 2067.0 | 2267.2 | 2028.0 | 2519.4 |
| 1419.6 | 1788.8 | 1604.2 | 2048.8 | 1705.6 | 2100.8 | 2267.2 |
| 1484.6 | 1822.4 | 1796.6 | 2314.0 | 962.0  | 2327.0 | 2615.6 |
| 1651.0 | 1838.2 | 1822.6 | 2072.2 | 1929.2 | 2225.6 | 2163.2 |
| 1661.4 | 1635.4 | 1835.6 | 1952.6 | 2202.2 | 2321.8 | 899.6  |
| 1851.2 | 618.8  | 1944.8 | 2134.6 | 1903.2 | 2275.0 | 2210.0 |
| 1617.2 | 1593.8 | 2009.8 | 1799.2 | 2337.4 | 2171.0 | 2376.4 |
| 1614.6 | 1898.0 | 2116.4 | 756.6  | 2022.8 | 2431.0 | 2259.4 |
| 1757.6 | 1911.0 | 1994.2 | 1890.2 | 2225.6 | 2165.8 | 2584.4 |
| 1302.6 | 1695.0 | 1895.4 | 2256.8 | 2441.4 | 780.0  |        |
| 572.0  | 1757.6 | 1947.4 | 2111.2 | 2113.8 | 2056.6 |        |
| 1458.6 | 1944.8 | 1770.6 | 2080.0 | 2035.8 | 2340.0 |        |
| 1567.8 | 2108.6 | 626.6  | 2191.8 | 2152.8 | 2033.2 |        |

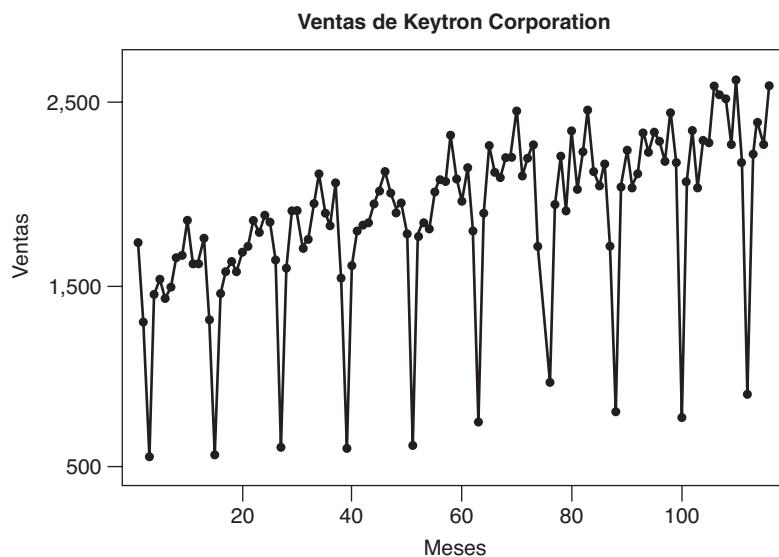
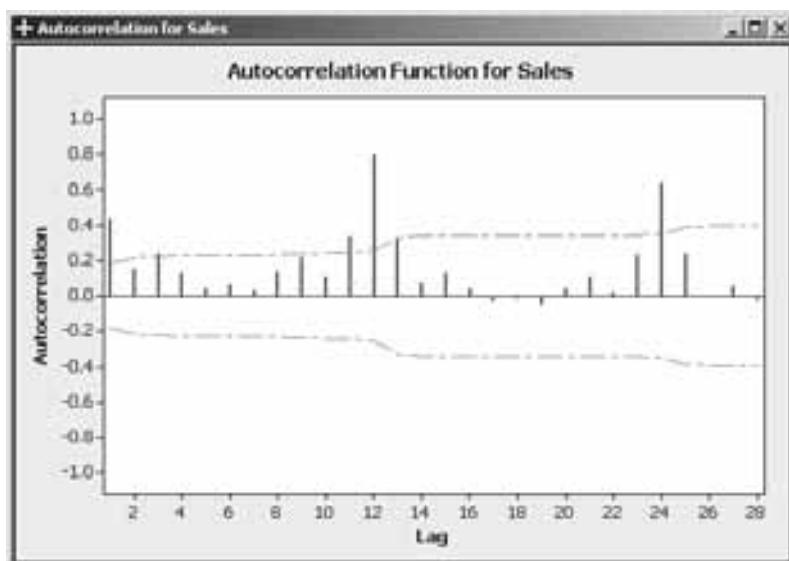


FIGURA 9.26 Ventas de Keytron Corporation



Autocorrelation Function: Sales

| Lag | ACF      | T    | LBQ    | Lag | ACF       | T     | LBQ    |
|-----|----------|------|--------|-----|-----------|-------|--------|
| 1   | 0.431271 | 4.63 | 21.96  | 15  | 0.127275  | 0.74  | 156.73 |
| 2   | 0.153911 | 1.41 | 24.78  | 16  | -0.039851 | 0.25  | 156.98 |
| 3   | 0.237841 | 2.14 | 31.58  | 17  | -0.031030 | -0.18 | 157.12 |
| 4   | 0.126583 | 1.11 | 33.50  | 18  | -0.012252 | -0.07 | 157.14 |
| 5   | 0.041549 | 0.36 | 33.79  | 19  | -0.049556 | -0.29 | 157.46 |
| 6   | 0.081352 | 0.53 | 34.26  | 20  | 0.041444  | 0.24  | 157.73 |
| 7   | 0.028549 | 0.24 | 34.36  | 21  | 0.107932  | 0.62  | 159.39 |
| 8   | 0.136014 | 1.16 | 36.49  | 22  | 0.026250  | 0.15  | 159.48 |
| 9   | 0.212966 | 1.80 | 42.44  | 23  | 0.227137  | 1.30  | 167.04 |
| 10  | 0.105903 | 0.67 | 43.66  | 24  | 0.635500  | 3.59  | 226.75 |
| 11  | 0.330651 | 2.71 | 58.20  | 25  | 0.238127  | 1.22  | 235.23 |
| 12  | 0.791934 | 6.08 | 140.12 | 26  | 0.000420  | 0.00  | 235.23 |
| 13  | 0.324223 | 1.94 | 153.99 | 27  | 0.059577  | 0.26  | 235.62 |
| 14  | 0.067091 | 0.39 | 154.59 | 28  | -0.025580 | -0.13 | 235.72 |

FIGURA 9.27 Autocorrelaciones de la muestra para las ventas de Keytron Corporation

también observó que las autocorrelaciones en los retrasos estacionales 12, 24 y 36 (que no se muestran) eran grandes y no se desvanecían con rapidez. Esto sugirió que la serie no era estacionaria y confirmó la impresión que Kathy tenía acerca de la gráfica de la serie de tiempo. Antes de continuar, Kathy decidió diferenciar la serie en relación con el retraso estacional para ver si podía convertir la serie no estacionaria en una que sí lo fuera.

La diferencia estacional para el periodo  $S = 12$  se define como

$$\Delta_{12} = Y_t - Y_{t-12}$$

Para las ventas de Keytron, la primera diferencia estacional que puede calcularse es

$$Y_{13} - Y_1 = 1757.6 - 1736.8 = 20.8$$

Se calcularon las restantes diferencias estacionales y las series que consisten en diferencias estacionales de ventas se muestran en la figura 9.28.

Las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra para las series diferenciadas se proporcionan en las figuras 9.29 y 9.30, respectivamente. La figura 9.28 indica que los datos diferenciados estacionalmente son estacionarios y parecen variar alrededor de un nivel de aproximadamente 100. Las autocorrelaciones tienen un punto significativo en el retraso 12 (se corta) y las autocorrelaciones parciales de muestra tienen puntos significativos en los retrasos 12 y 24 que se empequeñecen progresivamente (se desvanece). Esta conducta sugiere la presencia de un término MA(1) en el retraso estacional 12.

Kathy identifica un modelo ARIMA(0, 0, 0)(0, 1, 1)<sub>12</sub> para sus datos de ventas. Esta notación significa que este modelo tiene

$p = 0$  términos autorregresivos regulares

$d = 0$  diferencias regulares

$q = 0$  términos de promedio móvil regulares

$P = 0$  términos autorregresivos estacionales

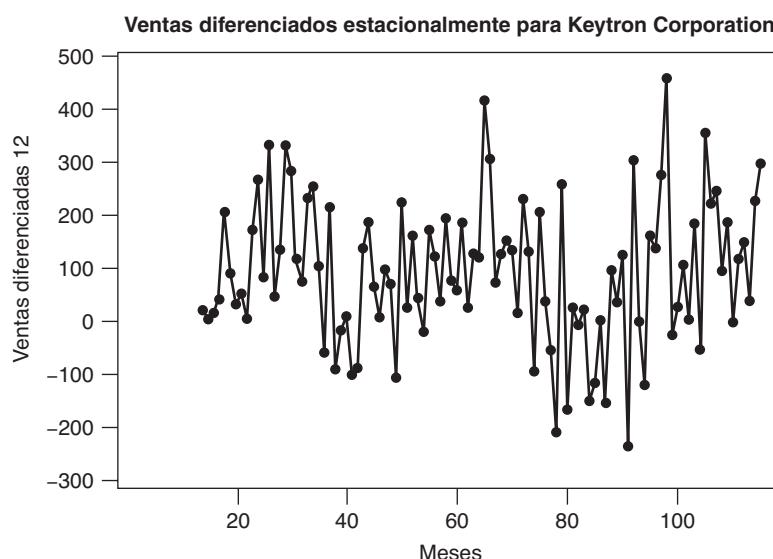
$D = 1$  diferencia estacional en el retraso 12

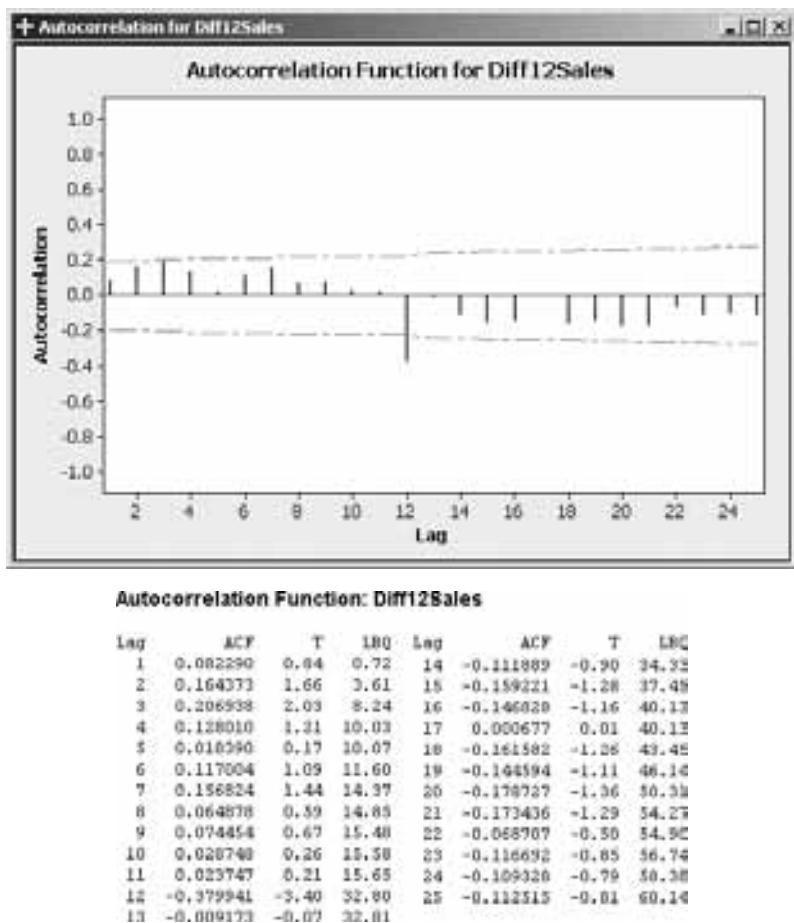
$Q = 1$  término de promedio móvil estacional en el retraso 12

Debido a que la serie diferenciada estacionalmente varía en relación con un nivel distinto de cero, Kathy incluye un término constante en su modelo. El modelo identificado tiene la forma

$$Y_t - Y_{t-12} = \mu + \varepsilon_t - \Omega_1 \varepsilon_{t-12} \quad (9.9)$$

**FIGURA 9.28 Ventas diferenciadas estacionalmente para Keytron Corporation**





**FIGURA 9.29 Autocorrelaciones de la muestra para las ventas diferenciadas estacionalmente de Keytron Corporation**

donde  $\mu$  es el nivel medio del proceso diferenciado estacionalmente y  $\Omega_1$  es el parámetro de promedio móvil estacional.

Kathy utiliza Minitab para estimar los parámetros de su modelo, calcular las autocorrelaciones residuales y generar pronósticos. El resultado se muestra en la tabla 9.15. Las autocorrelaciones residuales se proporcionan en la figura 9.31 y los pronósticos para los siguientes 12 meses se presentan al final de la serie de ventas en la figura 9.32.

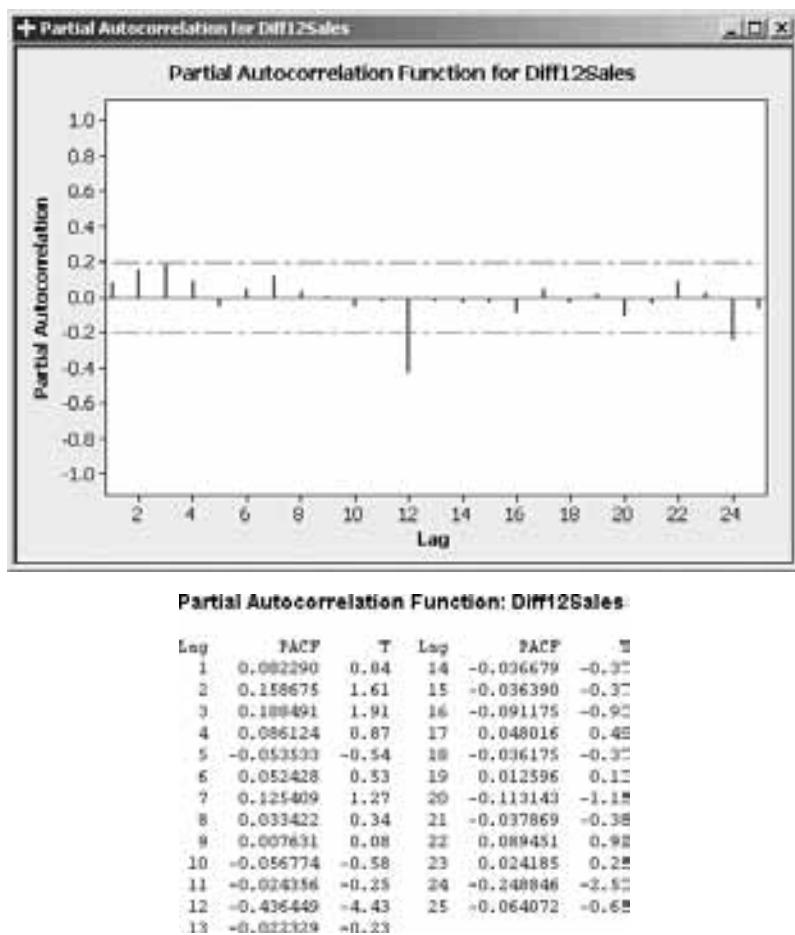
Kathy estaba satisfecha porque su modelo inicial se adaptaba muy bien a los datos. La estadística chi cuadrada Ljung-Box para los grupos de retrasos  $m = 12, 24, 26$  y  $48$ , claramente no era significativa, como se evidencia por sus grandes valores  $p$ . Las autocorrelaciones residuales eran uniformemente pequeñas sin un patrón aparente. Kathy estaba lista para utilizar su modelo ajustado a fin de verificar los pronósticos producidos por Minitab.

Los parámetros estimados fueron  $\hat{m} = 85.457$  y  $\hat{\Omega}_1 = .818$ . Por medio de estos valores, la ecuación 9.9 puede resolverse por  $Y_t$  para proporcionar

$$Y_t = Y_{t-12} + 85.457 + \varepsilon_t - .818\varepsilon_{t-12}$$

A fin de pronosticar las ventas para el periodo 116, Kathy establece  $t = 116$  y observa que en el momento que ella está realizando el pronóstico, su mejor opción de  $\varepsilon_{116}$  (el error para el siguiente periodo) es cero. De esta manera, la ecuación de pronóstico es

$$\hat{Y}_{116} = Y_{104} + 85.457 - .818\varepsilon_{104}$$



**FIGURA 9.30** Autocorrelaciones parciales de la muestra para las ventas diferenciadas estacionalmente de Keytron Corporation

donde  $e_{104} = -72.418$  es el residual (error estimado) para el periodo 104.

$$\hat{Y}_{116} = 2275 + 85.457 - .818(-72.418) = 2419.7$$

De manera similar,

$$\hat{Y}_{117} = Y_{105} + 85.457 - .818e_{105}$$

$$\hat{Y}_{117} = 2581.8 + 85.457 - .818(199.214) = 2504.3$$

Los pronósticos de Kathy concuerdan con los resultados que ofreció Minitab. Ella se convenció de que comprendió cómo utilizar el modelo ajustado para producir pronósticos. Kathy estaba satisfecha con los pronósticos de ventas para los siguientes 12 meses, los cuales se muestran en la figura 9.32. Los pronósticos parecían ser completamente coherentes con la conducta de la serie. Ella pensó que capturaban perfectamente la esencia del patrón estacional y el crecimiento de las ventas.

#### Ejemplo 9.10

En el ejemplo 3.5, Perkin Kendell, analista de Outboard Marine Corporation, utilizó el análisis de autocorrelación para concluir que las ventas trimestrales eran estacionales. Él ha decidido realizar pronósticos para las ventas de 1997 mediante la metodología Box-Jenkins. Los datos se graficaron en la figura 3.14.

**TABLA 9.15 Resultados en Minitab para el modelo ARIMA(0, 0, 0)(0, 1, 1)<sub>12</sub> de las ventas de Keytron Corporation**

```

ARIMA(0, 0, 0)(0, 1, 1)12 Model: Keytron Sales

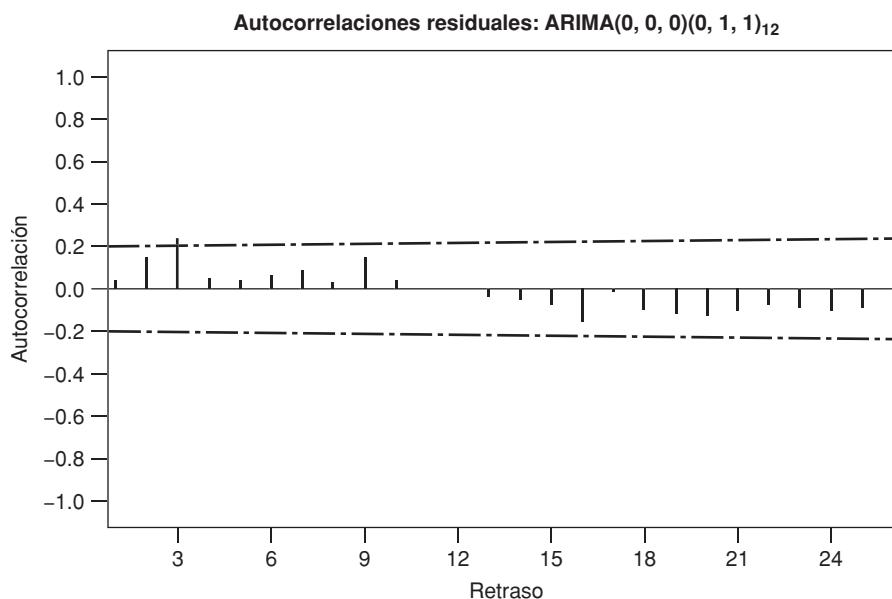
Final Estimates of Parameters
Type      Coef    SE Coef      T      P
SMA 12    0.8180   0.0881    9.28  0.000
Constant  85.457   2.910     29.36  0.000

Differencing: 0 regular, 1 seasonal of order 12
Number of observations: Original series 115, after differencing 103
Residuals: SS = 1172652 (backforecasts excluded)
            MS = 11610 DF = 101

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag          12      24      36      48
Chi-Square   12.5    27.1    34.6    46.4
DF           10      22      34      46
P-Value      0.250   0.209   0.439   0.456

Forecasts from period 115
      95 Percent Limits
Period  Forecast     Lower     Upper Actual
116      2419.69   2208.46   2630.93
117      2504.31   2293.07   2715.54

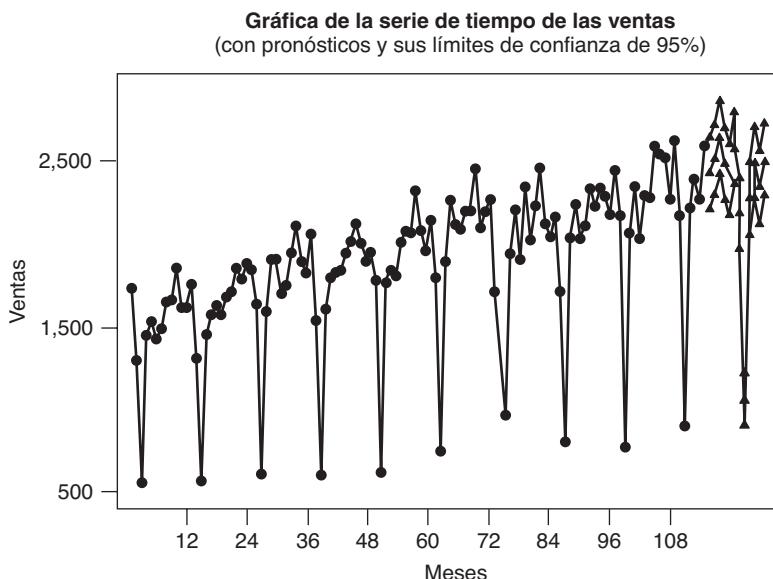
```



**FIGURA 9.31 Resultados en Minitab para el modelo ARIMA(0, 0, 0)(0, 1, 1)<sub>12</sub> de las ventas de Keytron Corporation**

(Vea la página 73). El gráfico de la serie de tiempo muestra un patrón estacional claro con, quizás, una ligera tendencia a la alta.

Perkin comenzó su análisis al ver las autocorrelaciones de la muestra para la serie original y las diversas diferencias de la serie. El amigo de Perkin, un experto en la metodología Box-Jenkins, le dijo que ésta era una buena idea cada vez que se diera la posibilidad de que la serie no fuera estacionaria. Una vez que el modelo más sencillo sugerido por un patrón de autocorrelación



**FIGURA 9.32 Ventas y pronósticos para los siguientes 12 meses para Keytron Corporation**

ha sido identificado tentativamente, este hecho puede confirmarse al examinar las autocorrelaciones parciales de la muestra. Perkin calculó las funciones de autocorrelación mediante Minitab para la siguiente serie:

$Y_t$  = la serie original de ventas

$\Delta Y_t = Y_t - Y_{t-1}$  = la serie de las primeras diferencias (regulares)

$\Delta_4 Y_t = Y_t - Y_{t-4}$  = la serie de las diferencias estacionales de orden  $S = 4$

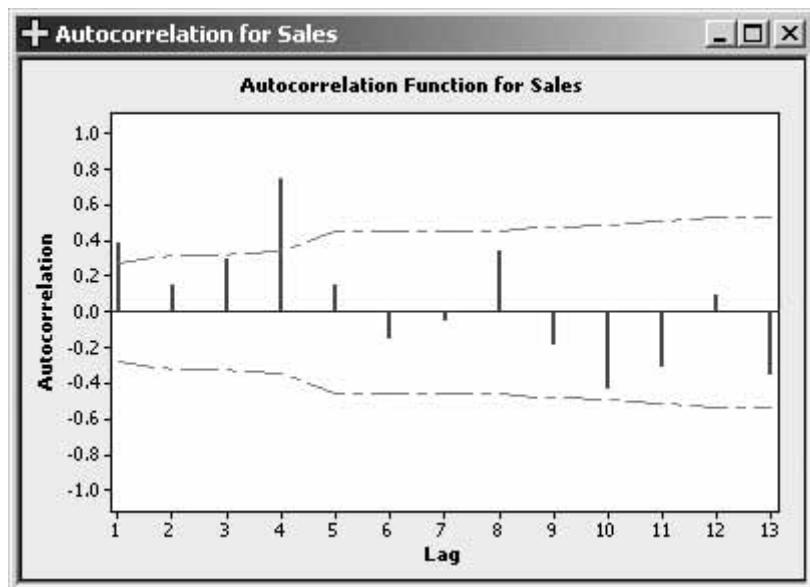
$\left[ \begin{array}{l} \Delta \Delta_4 Y_t = \Delta(\Delta_4 Y_t) = \Delta(Y_t - Y_{t-4}) \\ = Y_t - Y_{t-1} - Y_{t-4} + Y_{t-5} \end{array} \right]$  = la serie que consiste en una diferencia estacional seguida por una diferencia regular

Algunas veces, es necesario tomar una diferencia estacional y una diferencia regular con el fin de que la serie resultante sea estacionaria. Las autocorrelaciones de la muestra de la serie original para la serie diferenciada estacionalmente y para la serie con una diferencia regular y una estacional, se muestran en las figuras 9.33, 9.34 y 9.35.

La función de autocorrelación de la serie original tiene picos grandes (significativos) en los retrasos 1 y 4. No obstante, las autocorrelaciones en los retrasos estacionales 4, 8 y 12 caen hacia cero. Esto puede indicar una serie no estacionaria y la necesidad de una diferencia estacional. De hecho, el intento de Perkin para ajustar un modelo ARIMA a la serie original con un término constante, un término de promedio móvil regular y un término autorregresivo estacional no tuvo éxito. El parámetro autorregresivo estacional estimado resultó cerca de uno (lo que implica la necesidad de una diferencia estacional) y las autocorrelaciones residuales fueron generalmente grandes y no coherentes con los errores aleatorios. Después, Perkin intentó la función de autocorrelación de la serie diferenciada estacionalmente que se observa en la figura 9.34.

Las autocorrelaciones de los datos diferenciados estacionalmente son grandes en los retrasos bajos y declinan lentamente en la forma de un patrón de onda. Perkin decidió que la serie aún no podría ser estacionaria y pensó que podría necesitarse la diferencia regular, además de la diferencia estacional, para lograr hacerla estacionaria. En consecuencia, utilizó Minitab para producir las autocorrelaciones de la muestra para las primeras diferencias de las ventas y las diferencias estacionales que se muestran en la figura 9.35.

Al examinar las autocorrelaciones de la figura 9.35, Perkin observó sólo dos valores significativos en los retrasos 1 y 8. Además, observó que las autocorrelaciones para los primeros dos retrasos alternaban su signo. Perkin pensó que podría ser adecuado un modelo ARIMA con un término autorregresivo regular y, quizás, un término de promedio móvil estacional en el retraso 8. Antes de continuar, Perkin quería confirmar su selección inicial al examinar las autocorrelaciones parciales para la serie de las primeras diferencias y las diferencias estacionales.



**Autocorrelation Function: Sales**

| Lag | ACF       | T     | LBQ   | Lag | ACF       | T     | LBQ   |
|-----|-----------|-------|-------|-----|-----------|-------|-------|
| 1   | 0.392861  | 2.83  | 8.50  | 8   | 0.346975  | 1.51  | 57.72 |
| 2   | 0.153945  | 0.97  | 9.83  | 9   | -0.182601 | -0.76 | 59.90 |
| 3   | 0.293782  | 1.82  | 14.77 | 10  | -0.434729 | -1.80 | 72.53 |
| 4   | 0.743520  | 4.34  | 47.11 | 11  | -0.315031 | -1.23 | 79.33 |
| 5   | 0.150655  | 0.67  | 48.47 | 12  | 0.091203  | 0.35  | 79.91 |
| 6   | -0.153008 | -0.67 | 49.90 | 13  | -0.353274 | -1.34 | 88.90 |
| 7   | -0.046978 | -0.21 | 50.04 |     |           |       |       |

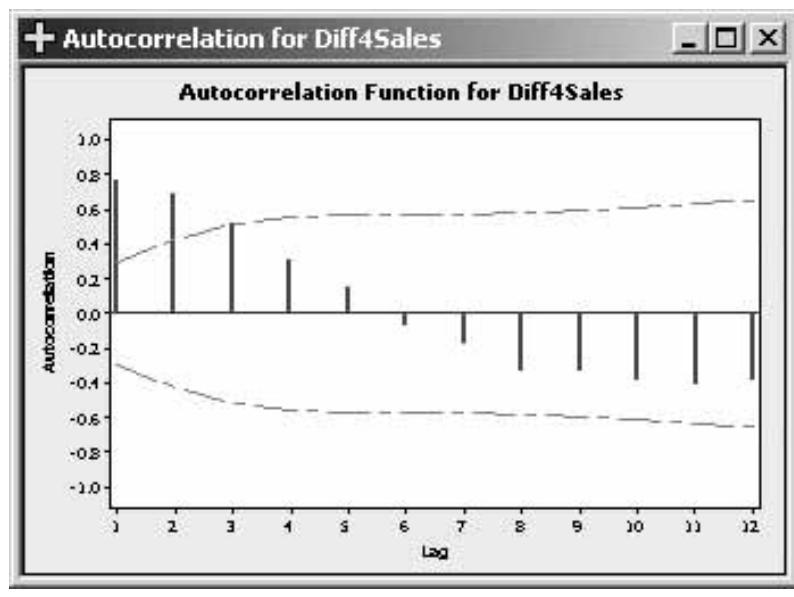
**FIGURA 9.33 Autocorrelaciones de la muestra para las ventas de Outboard Marine Corporation**

Las autocorrelaciones parciales de la muestra se presentaron en la figura 9.36 para las primeras diferencias y para las diferencias estacionales. Las autocorrelaciones parciales parecían cortarse después del primer retraso, en consistencia con la conducta de AR(1) que él observó en las autocorrelaciones y hubo una autocorrelación parcial significativa en el retraso 8.

Perkin no estaba seguro acerca de cómo manejar los parámetros estacionales, de manera que decidió conservarlo de un modo más sencillo y ajustar un modelo ARIMA(1, 1, 0)(0, 1, 0)<sub>4</sub>. Este modelo permite una diferencia regular y otra estacional en el retraso 4, un término autorregresivo regular, pero no permite coeficientes autorregresivos estacionales o de promedio móvil. Una gráfica de la serie después de tomar diferencias mostró que variaba alrededor de cero, de manera que no se incluyó ningún término constante en el modelo. Perkin pensó que si los parámetros estacionales fueran necesarios, la evidencia aparecería en las autocorrelaciones residuales desde su modelo inicial.

Los resultados del modelo ARIMA(1, 1, 0)(0, 1, 0)<sub>4</sub> se presentan en la parte superior de la tabla 9.16 y las autocorrelaciones residuales se presentan al principio de la figura 9.37. El modelo parecía ajustarse bastante bien con un coeficiente estimado  $\hat{\phi}_1 = -.352$  y un error cuadrado medio residual de  $s^2 = 1041.3$ . Sin embargo, un pico significativo ocurrido en el retraso 8 en las autocorrelaciones residuales y la estadística Ljung-Box para los primeros  $m = 12$  retrasos fue  $Q_{12} = 19.7$  con un valor  $p$  de 0.05. Lo último sugiere que como grupo, las primeras 12 autocorrelaciones residuales son mayores de lo que se esperaría para los errores aleatorios.

Perkin decide modificar su modelo inicial e incluir términos de promedio móvil que correspondan a los retrasos estacionales 4 y 8. La evidencia sugiere que únicamente el término en el retraso 8 es necesario. No obstante, el programa Minitab requiere incluir parámetros estacionales



**Autocorrelation Function: Diff4Sales**

| Lag | ACF       | T     | LBQ   | Lag | ACF       | T     | LBQ    |
|-----|-----------|-------|-------|-----|-----------|-------|--------|
| 1   | 0.768185  | 5.32  | 30.13 | 7   | -0.171085 | -0.60 | 77.70  |
| 2   | 0.693331  | 3.25  | 55.21 | 8   | -0.333127 | -1.16 | 84.36  |
| 3   | 0.514793  | 2.01  | 69.35 | 9   | -0.338294 | -1.14 | 91.40  |
| 4   | 0.303188  | 1.10  | 74.36 | 10  | -0.392484 | -1.29 | 101.13 |
| 5   | 0.152404  | 0.54  | 75.66 | 11  | -0.407611 | -1.30 | 111.90 |
| 6   | -0.075474 | -0.26 | 75.98 | 12  | -0.387217 | -1.19 | 121.90 |

**FIGURA 9.34** Autocorrelaciones de la muestra para las ventas diferenciadas estacionalmente de Outboard Marine Corporation

AR o MA en el modelo, así como el múltiplo más elevado del retraso estacional que es necesario. En este caso, el retraso estacional es  $S = 4$  y el parámetro de promedio móvil en el retraso  $2 \times 4 = 8$  es necesario.

Perkin utiliza Minitab para ajustar un modelo ARIMA(1, 1, 0)(0, 1, 2)<sub>4</sub> a sus datos.<sup>14</sup> El resultado se muestra en la tabla 9.16 (parte inferior) y las autocorrelaciones residuales se proporcionan en la figura 9.37 (vea la página 426). Perkin se encuentra satisfecho con los resultados. El error cuadrado medio de los residuales ha sido reducido a  $s^2 = 716.3$  y las autocorrelaciones residuales, conforme se juzgan por los valores individuales y la estadística de Ljung-Box, son coherentes con los errores aleatorios. Las gráficas residuales (que no se muestran) sugieren que las suposiciones acerca del término del otro error son apropiadas.

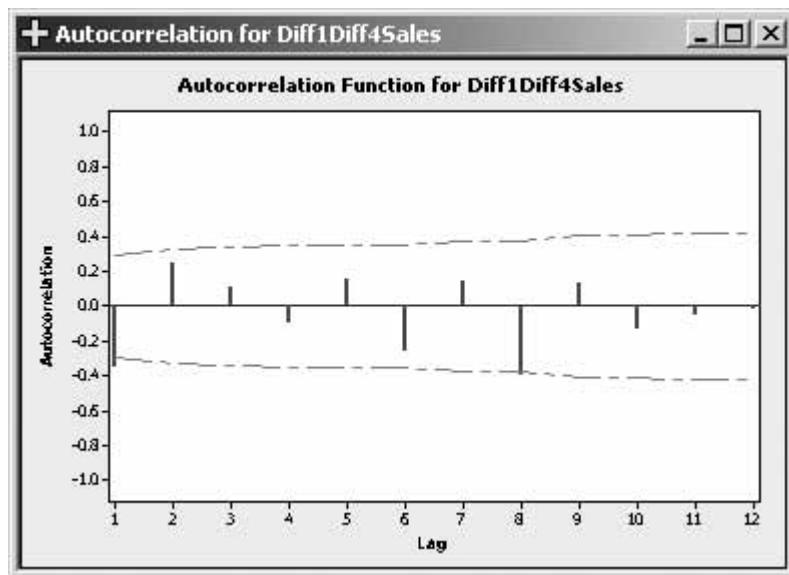
Para obtener experiencia adicional con la metodología Box-Jenkins, Perkin desea verificar los pronósticos de las ventas para los siguientes dos trimestres en relación con los resultados obtenidos en Minitab. Por medio de los estimados de parámetros en lugar de valores verdaderos, el modelo final es

$$W_t = -.351 W_{t-1} + \varepsilon_t - .239 \varepsilon_{t-4} - .672 \varepsilon_{t-8}$$

donde

$$W_t = \Delta \Delta_4 Y_t = Y_t - Y_{t-1} - Y_{t-4} + Y_{t-8}$$

<sup>14</sup> Los comandos de Minitab para correr un modelo ARIMA(1, 1, 0)(0, 1, 2)<sub>4</sub> se presentan el final de este capítulo.



**Autocorrelation Function: Diff1Diff4Sales**

| Lag | ACF       | T     | LBQ   | Lag | ACF       | T     | LBQ   |
|-----|-----------|-------|-------|-----|-----------|-------|-------|
| 1   | -0.347173 | -2.38 | 6.03  | 7   | 0.147388  | 0.80  | 16.47 |
| 2   | 0.248695  | 1.53  | 9.20  | 8   | -0.398618 | -2.14 | 25.86 |
| 3   | 0.106003  | 0.62  | 9.79  | 9   | 0.127499  | 0.63  | 26.84 |
| 4   | -0.093888 | -0.55 | 10.26 | 10  | -0.126568 | -0.62 | 27.84 |
| 5   | 0.158168  | 0.91  | 11.63 | 11  | -0.053319 | -0.26 | 28.02 |
| 6   | -0.252819 | -1.44 | 15.22 | 12  | -0.021388 | -0.10 | 28.05 |

**FIGURA 9.35 Autocorrelaciones de la muestra para las primeras diferencias de las ventas y diferenciadas estacionalmente para Outboard Marine Corporation**

es la serie diferenciada. En el origen del pronóstico  $t - 1$ , la ecuación para realizar pronósticos, escrita en términos de  $Y$ , adopta la forma

$$\hat{Y}_t = Y_{t-1} + Y_{t-4} - Y_{t-5} - .351(Y_{t-1} - Y_{t-2} - Y_{t-5} + Y_{t-6}) - .239e_{t-4} - .672e_{t-8}$$

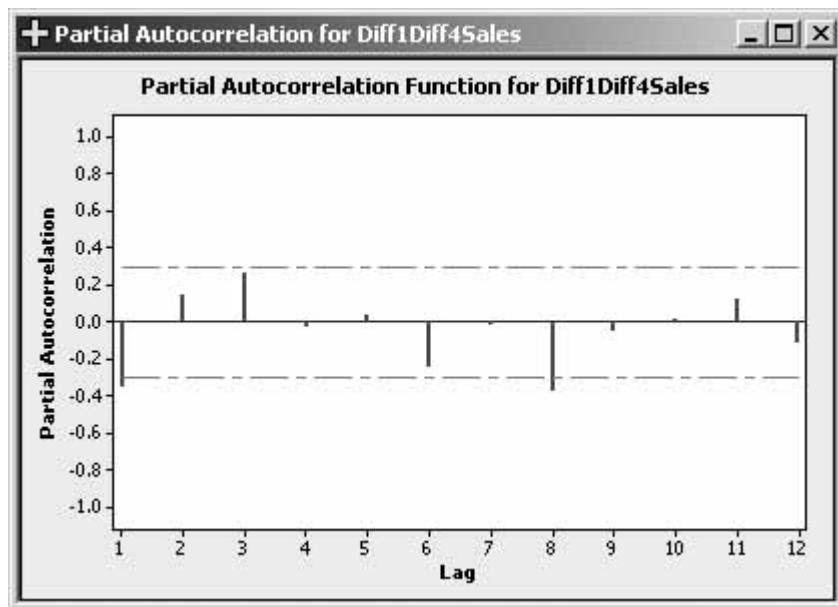
donde  $e_{t-4}$  y  $e_{t-8}$  son los residuales (errores estimados) para los períodos  $t - 4$  y  $t - 8$ , respectivamente.

El pronóstico de las ventas para el periodo 53 es

$$\begin{aligned}\hat{Y}_{53} &= Y_{52} + Y_{49} - Y_{48} - .351(Y_{52} - Y_{51} - Y_{48} + Y_{47}) - .239e_{49} - .672e_{45} \\ \hat{Y}_{53} &= 281.4 + 232.1 - 338.2 - .351(281.4 - 291.0 - 338.2 + 329.6) - .239(-1.726) \\ &\quad - .672(18.898) \\ &= 169.4\end{aligned}$$

y el pronóstico del periodo 54 es

$$\begin{aligned}\hat{Y}_{54} &= \hat{Y}_{53} + Y_{50} - Y_{49} - .351(\hat{Y}_{53} - Y_{52} - Y_{49} + Y_{48}) - .239e_{50} - .672e_{46} \\ \hat{Y}_{54} &= 169.4 + 285.6 - 232.1 - .351(169.4 - 281.4 - 232.1 + 338.2) - .239(-37.040) \\ &\quad - .672(4.781) \\ &= 230.6\end{aligned}$$



**Partial Autocorrelation Function: Diff1Diff4Sales**

| Lag | PACF      | T     | Lag | PACF      | T     |
|-----|-----------|-------|-----|-----------|-------|
| 1   | -0.347173 | -2.38 | 7   | -0.020174 | -0.14 |
| 2   | 0.145731  | 1.00  | 8   | -0.376537 | -2.58 |
| 3   | 0.267607  | 1.83  | 9   | -0.043690 | -0.30 |
| 4   | -0.026767 | -0.18 | 10  | 0.012654  | 0.09  |
| 5   | 0.036594  | 0.25  | 11  | 0.124850  | 0.86  |
| 6   | -0.238699 | -1.64 | 12  | -0.111552 | -0.76 |

**FIGURA 9.36** Autocorrelaciones parciales de la muestra para las primeras diferencias de las ventas y para las ventas diferenciadas estacionalmente para Outboard Marine Corporation

Perkin estaba contento porque sus pronósticos calculados concordaron con los producidos con Minitab. Los pronósticos para los siguientes cuatro trimestres están graficados en la figura 9.38. Los pronósticos parecen razonables. El patrón de los pronósticos se parece más al patrón histórico. Perkin pensó que su esfuerzo por realizar los pronósticos fue un éxito y planeó recomendar la metodología Box-Jenkins a sus colegas.

### Suavizamiento exponencial simple y un modelo ARIMA

Varios modelos ARIMA producen pronósticos que son los mismos o casi los mismos que los métodos de suavizamiento que se presentaron en el capítulo 4. Para ilustrar un caso, considere el modelo ARIMA(0, 1, 1)

$$Y_t - Y_{t-1} = \varepsilon_t - \omega_1 \varepsilon_{t-1} \quad (9.10)$$

Suponga que el pronóstico se origina en  $t$  y se requiere el pronóstico de  $Y_{t+1}$ . Al sustituir  $t + 1$  por  $t$  en la ecuación 9.10 y al resolver  $Y_{t+1}$  se obtiene

$$Y_{t+1} = Y_t + \varepsilon_{t+1} - \omega_1 \varepsilon_t$$

**TABLA 9.16 Resultados en Minitab para el modelo ARIMA(1, 1, 0)(0, 1, 0)<sub>4</sub> (parte superior) y ARIMA(1, 1, 0)(0, 1, 2)<sub>4</sub> (parte inferior) de las ventas de Outboard Marine**

| <b>ARIMA(1, 1, 0)(0, 1, 0)<sub>4</sub> Model: Outboard Marine Sales</b> |          |         |         |        |  |
|-------------------------------------------------------------------------|----------|---------|---------|--------|--|
| Final Estimates of Parameters                                           |          |         |         |        |  |
| Type                                                                    | Coef     | SE Coef | T       | P      |  |
| AR 1                                                                    | -0.3520  | 0.1384  | -2.54   | 0.014  |  |
| Differencing: 1 regular, 1 seasonal of order 4                          |          |         |         |        |  |
| Number of observations: Original series 52, after differencing 47       |          |         |         |        |  |
| Residuals: SS = 47898.3 (backforecasts excluded)                        |          |         |         |        |  |
| MS = 1041.3 DF = 46                                                     |          |         |         |        |  |
| Modified Box-Pierce (Ljung-Box) Chi-Square statistic                    |          |         |         |        |  |
| Lag                                                                     | 12       | 24      | 36      | 48     |  |
| Chi-Square                                                              | 19.7     | 23.7    | 32.1    | *      |  |
| DF                                                                      | 11       | 23      | 35      | *      |  |
| P-Value                                                                 | 0.050    | 0.420   | 0.608   | *      |  |
| <b>ARIMA(1, 1, 0)(0, 1, 2)<sub>4</sub> Model: Outboard Marine Sales</b> |          |         |         |        |  |
| Final Estimates of Parameters                                           |          |         |         |        |  |
| Type                                                                    | Coef     | SE Coef | T       | P      |  |
| AR 1                                                                    | -0.3511  | 0.1423  | -2.47   | 0.018  |  |
| SMA 4                                                                   | 0.2382   | 0.1339  | 1.78    | 0.082  |  |
| SMA 8                                                                   | 0.6730   | 0.1403  | 4.80    | 0.000  |  |
| Differencing: 1 regular, 1 seasonal of order 4                          |          |         |         |        |  |
| Number of observations: Original series 52, after differencing 47       |          |         |         |        |  |
| Residuals: SS = 31518.1 (backforecasts excluded)                        |          |         |         |        |  |
| MS = 716.3 DF = 44                                                      |          |         |         |        |  |
| Modified Box-Pierce (Ljung-Box) Chi-Square statistic                    |          |         |         |        |  |
| Lag                                                                     | 12       | 24      | 36      | 48     |  |
| Chi-Square                                                              | 8.4      | 10.7    | 22.0    | *      |  |
| DF                                                                      | 9        | 21      | 33      | *      |  |
| P-Value                                                                 | 0.493    | 0.969   | 0.927   | *      |  |
| Forecasts from period 52                                                |          |         |         |        |  |
| 95 Percent Limits                                                       |          |         |         |        |  |
| Period                                                                  | Forecast | Lower   | Upper   | Actual |  |
| 53                                                                      | 169.386  | 116.917 | 221.854 |        |  |
| 54                                                                      | 230.533  | 167.987 | 293.080 |        |  |

Debido a que el tiempo  $t$  como la mejor opción de  $\varepsilon_{t+1}$  es cero, y  $\varepsilon_t$  se estima por el residual  $e_t = Y_t - \hat{Y}_t$ , la ecuación del pronóstico es

$$\hat{Y}_{t+1} = Y_t - \omega_1(Y_t - \hat{Y}_t) = (1 - \omega_1)Y_t + \omega_1\hat{Y}_t \quad (9.11)$$

Si se define  $\alpha = 1 - \omega_1$ , la ecuación 9.11 es idéntica a la ecuación 4.13 de suavizamiento exponencial simple

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t$$

Pronosticar mediante el suavizamiento exponencial equivale a generar pronósticos con el modelo ARIMA(0, 1, 1) con un parámetro  $\omega_1 = 1 - \alpha$ . Observe que el modelo ARIMA(0, 1, 1) representa un proceso no estacionario. El suavizamiento exponencial simple deberá funcionar bien para las series de tiempo que puedan describirse de forma adecuada por

medio de un modelo ARIMA(0, 1, 1). Para las series de tiempo que no puedan describirse adecuadamente por medio de este modelo, los pronósticos generados por el suavizamiento exponencial podrían resultar particularmente insatisfactorios.

Para el suavizamiento exponencial simple, el parámetro  $\alpha$  usualmente está restringido al rango  $0 < \alpha < 1$ . El parámetro de promedio móvil  $\omega_1$  en el modelo ARIMA(0, 1, 1) se restringe al rango  $-1 < \omega_1 < 1$  de manera que, hablando de forma estricta, los dos métodos de pronóstico son equivalentes para los valores positivos de los parámetros  $\alpha$  y  $\omega_1$ .

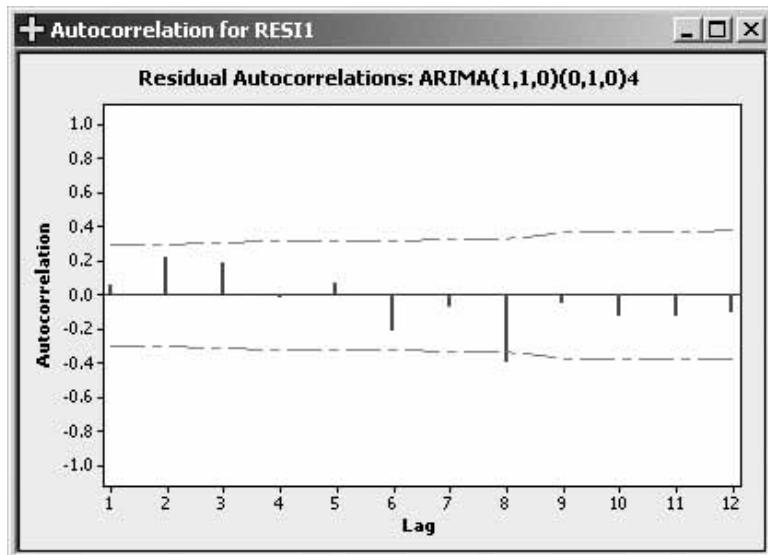
### Ventajas y desventajas de los modelos ARIMA

La metodología Box-Jenkins para el análisis de las series de tiempo es una herramienta poderosa para proporcionar pronósticos precisos de corto alcance. Los modelos ARIMA son muy flexibles y pueden representar un amplio rango de características de las series de tiempo que ocurren en la realidad. Los procedimientos formales para la evaluación de la conveniencia de un modelo están disponibles. Además, los pronósticos y los intervalos de predicción provienen directamente del modelo ajustado.

No obstante, los modelos ARIMA tienen ciertas desventajas.

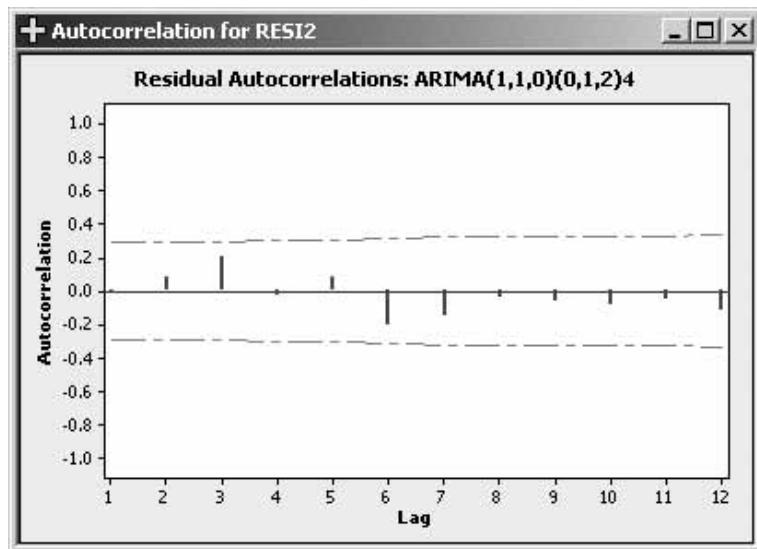
1. Se requiere una cantidad relativamente grande de datos. Debe reconocerse que si los datos son estacionales con, por ejemplo, un periodo estacional de  $S = 12$ , entonces las observaciones mensuales para un año constituyen esencialmente un punto de datos (una sola vista del patrón estacional), no 12. En términos generales, para datos no estacionales y a fin de desarrollar un modelo ARIMA se necesitan 40 observaciones

**FIGURA 9.37 Autocorrelaciones de los residuales del modelo ARIMA(1, 1, 0)(0, 1, 0)<sub>4</sub> y autocorrelaciones de los residuales del modelo ARIMA(1, 1, 0)(0, 1, 2)<sub>4</sub> para las ventas de Outboard Marine**

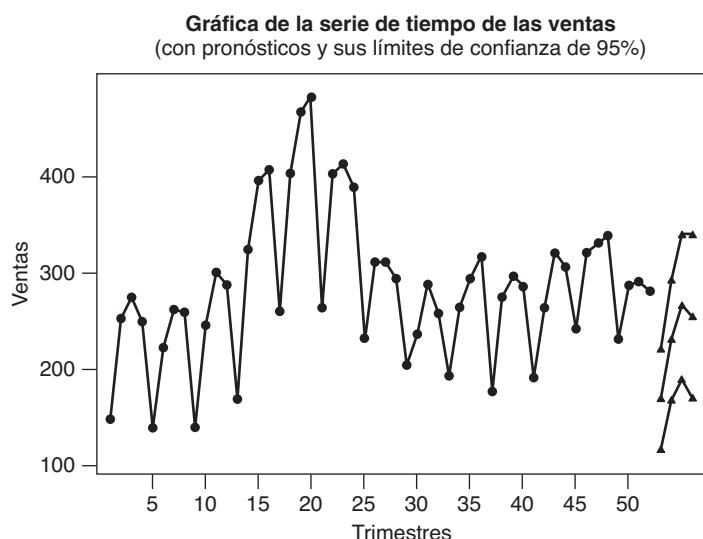


**Autocorrelation Function: RESI1**

| Lag | ACF       | T     | LBQ  | Lag | ACF       | T     | LBQ   |
|-----|-----------|-------|------|-----|-----------|-------|-------|
| 1   | 0.054510  | 0.37  | 0.15 | 7   | -0.070930 | -0.43 | 7.59  |
| 2   | 0.222343  | 1.52  | 2.68 | 8   | -0.393430 | -2.38 | 16.73 |
| 3   | 0.193279  | 1.26  | 4.63 | 9   | -0.043829 | -0.24 | 16.85 |
| 4   | -0.017583 | -0.11 | 4.65 | 10  | -0.126685 | -0.69 | 17.85 |
| 5   | 0.065878  | 0.42  | 4.89 | 11  | -0.129403 | -0.69 | 18.92 |
| 6   | -0.207367 | -1.30 | 7.30 | 12  | -0.107972 | -0.57 | 19.69 |

**Autocorrelation Function: RESI2**

| Lag | ACF       | T     | LBQ  | Lag | ACF       | T     | LBQ  |
|-----|-----------|-------|------|-----|-----------|-------|------|
| 1   | -0.013785 | -0.09 | 0.01 | 7   | -0.152917 | -0.96 | 6.72 |
| 2   | 0.086619  | 0.59  | 0.39 | 8   | -0.042170 | -0.26 | 6.83 |
| 3   | 0.203811  | 1.39  | 2.57 | 9   | -0.060928 | -0.37 | 7.05 |
| 4   | -0.031045 | -0.20 | 2.62 | 10  | -0.080110 | -0.49 | 7.45 |
| 5   | 0.085412  | 0.56  | 3.02 | 11  | -0.044302 | -0.27 | 7.58 |
| 6   | -0.204742 | -1.33 | 5.37 | 12  | -0.113216 | -0.69 | 8.42 |

**FIGURA 9.37 (Continúa)****FIGURA 9.38 Pronósticos para los siguientes cuatro trimestres de un modelo ARIMA(1, 1, 0)(0, 1, 2)<sub>4</sub> para las ventas de Outboard Marine**

- o más. Para construir un modelo ARIMA, en el caso de los datos estacionales, se requieren entre seis y 10 años de datos, según sea la longitud del periodo estacional.
2. No hay formas sencillas para actualizar los parámetros de un modelo ARIMA cuando se incorporan nuevos datos, como sucede con los métodos de suavizamiento. El modelo tiene que readjustarse periódicamente en su totalidad y algunas veces debe desarrollarse un modelo nuevo.
  3. Con frecuencia, la construcción de un modelo ARIMA satisfactorio requiere de una gran inversión de tiempo y otros recursos. Los costos del desarrollo del modelo, tiempo de operación de la computadora y necesidades de almacenamiento pueden ser sustancialmente más elevados para los modelos ARIMA que para las técnicas más tradicionales de pronóstico, como el suavizamiento.

## APLICACIÓN A LA ADMINISTRACIÓN

---

Según Bernstein (1996), con frecuencia los pronósticos constituyen una de las más importantes aportaciones que los administradores han desarrollado para auxiliarse en el proceso de la toma decisiones. Prácticamente, cada decisión operativa importante depende, hasta cierto grado, de un pronóstico. La acumulación de inventarios se relaciona con el pronóstico de la demanda esperada; el departamento de producción tiene que programar las necesidades de empleo y los pedidos de materia prima para los siguientes dos meses; el departamento de finanzas debe preparar los financiamientos de corto plazo por el siguiente trimestre; el departamento de personal debe determinar las necesidades de contratación y despidos. La lista de las aplicaciones de los pronósticos es muy grande.

Los ejecutivos son muy conscientes de la importancia de los pronósticos. Sin duda alguna, gran parte del tiempo se invierte en el estudio de las tendencias en las cuestiones económicas y políticas y la forma en que los sucesos pueden afectar la demanda de los productos o servicios. Un tema de interés es la importancia que los ejecutivos le otorgan a los métodos para realizar pronósticos cuantitativos, en oposición a los pronósticos que se basan estrictamente en sus juicios. La cuestión es especialmente sensible cuando los sucesos pueden tener una repercusión significativa en la demanda. Uno de los problemas de los métodos para realizar pronósticos cuantitativos es que dependen de los datos históricos. Por esta razón, quizás sean menos efectivos al considerar que con frecuencia los resultados implican una demanda considerablemente más elevada o más baja.

Con frecuencia, un tipo de problema que los administradores de negocios enfrentan es la necesidad de preparar pronósticos de corto plazo para un número de artículos diversos. Un ejemplo típico es el gerente que debe programar la producción con base en cierto pronóstico de la demanda para varios cientos de productos distintos en una misma línea de producción. Los métodos de suavizamiento son las técnicas que se utilizan con mayor frecuencia para esta situación.

La mayor ventaja del suavizamiento exponencial la constituyen su bajo costo y sencillez. Es posible que no sea tan preciso como los métodos más sofisticados; por ejemplo, la modelación general ARIMA. Sin embargo, cuando se necesitan pronósticos para sistemas de inventarios que contienen miles de productos, a menudo los métodos de suavizamiento son el único método razonable.

Los pronósticos a partir de una perspectiva de las series de tiempo dependen de la suposición de que el futuro será como el pasado y que los patrones precedentes pueden describirse adecuadamente. Las técnicas de series de tiempo son más útiles para pronosticar las variables que fluctúan en ciertos patrones estables a través del tiempo.

La metodología Box-Jenkins es una herramienta poderosa para proporcionar pronósticos precisos de corto alcance. Los administradores deben estar conscientes de que la construcción de un modelo ARIMA satisfactorio, por medio de la técnica Box-Jenkins, requiere una cantidad considerable de datos históricos y de una elevada inversión del tiempo del analista.

Hay muchas aplicaciones apropiadas de la metodología Box-Jenkins. Los modelos ARIMA se han utilizado como se explica continuación:

- Estimar un cambio en la estructura de precios de la industria telefónica estadounidense.
- Investigar la relación entre la concentración de amonio, flujo lento y temperatura existente en los ríos.
- Pronosticar el volumen accionario anual.
- Pronosticar la cantidad de petróleo activo en los pozos.
- Analizar el número de unidades particulares de vivienda.
- Explorar las observaciones diarias de los avances porcentuales en el número de asuntos negociados.
- Estudiar la competencia existente entre las rutas aéreas y ferroviarias.
- Pronosticar el empleo.
- Examinar un gran número de series de tiempo de energía para una compañía de servicios.
- Investigar el efecto de las promociones de las ventas de los productos de consumo.
- Pronosticar las diversas categorías de aseguramiento de calidad del producto.

## Glosario

**Metodología Box-Jenkins.** Se refiere a una serie de procedimientos para identificar, ajustar y verificar los modelos ARIMA con datos de la serie que tiempo. Los pronósticos proceden directamente de la forma del modelo ajustado.

**Principio de parsimonia.** Este precepto se refiere a la preferencia de modelos más simples sobre los complejos.

## Fórmulas clave

### Modelo ARIMA: modelo autorregresivo de orden $p$

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t \quad (9.1)$$

### Modelo ARIMA: modelo de promedio móvil de orden $q$

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \quad (9.2)$$

### Modelo ARIMA: modelo ARMA ( $p, q$ )

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \quad (9.3)$$

### Error cuadrado medio de los residuales

$$s^2 = \frac{\sum_{t=1}^n e_t^2}{n-r} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n-r} \quad (9.4)$$

### Estadística Ljung-Box $Q$ (Box-Pierce modificada)

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2(e)}{n-k} \quad (9.5)$$

### Intervalos aproximados de predicción para la serie estacionaria

$$\hat{Y} \pm 2s \quad (9.6)$$

### Criterio de información Akaike (AIC)

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n} r \quad (9.7)$$

**Criterio de información (BIC) bayesiano (Schwarz)**

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n} r \quad (9.8)$$

**Modelo de promedio móvil estacional (datos mensuales)**

$$Y_t - Y_{t-12} = \mu + \varepsilon_t - \Omega_1 \varepsilon_{t-12} \quad (9.9)$$

**Modelo ARIMA(0, 1, 1)**

$$Y_t - Y_{t-1} = \varepsilon_t - \omega_1 \varepsilon_{t-1} \quad (9.10)$$

**Ecuación para la generación de pronósticos del modelo ARIMA(0, 1, 1) en forma de suavizamiento exponencial**

$$\hat{Y}_{t+1} = Y_t - \omega_1(Y_t - \hat{Y}_t) = (1 - \omega_1)Y_t + \omega_1 \hat{Y}_t \quad (9.11)$$

**Problemas**

1.
  - a) Para una muestra de 100 observaciones de datos aleatorios, calcule un intervalo de confianza de 95% para el coeficiente de autocorrelaciones en cualquier retraso.
  - b) Si todos los coeficientes de autocorrelación están dentro de sus intervalos individuales de confianza de 95% y no muestran un patrón particular, ¿qué conclusión se puede obtener acerca del proceso?
  - c) Si las primeras autocorrelaciones son positivas, significativamente distintas de cero y los patrones de autocorrelación declinan a cero, ¿qué conclusión puede obtener acerca del proceso?
  - d) Se observa un proceso trimestralmente. Si las autocorrelaciones  $r_4, r_8$  y  $r_{12}$  son significativamente distintas de cero, ¿qué conclusión puede obtenerse?
2. Suponga que el siguiente modelo de una serie de tiempo se ha ajustado a la serie de tiempo y se le considera como un modelo adecuado

$$Y_t = 35 + \varepsilon_t + .25\varepsilon_{t-1} - .30\varepsilon_{t-2}$$

Las primeras cuatro observaciones son  $Y_1 = 32.5$ ,  $Y_2 = 36.6$ ,  $Y_3 = 33.3$  y  $Y_4 = 31.9$ . Se supone que  $\hat{Y}_1 = 35$  y  $\varepsilon_0 = 0$ , calcule los pronósticos para los períodos 5, 6 y 7, si el período 4 es el origen del pronóstico.

3. Un modelo de serie de tiempo ha sido ajustado y verificado con los datos históricos, lo cual produce

$$Y_t = 50 + .45Y_{t-1} + \varepsilon_t$$

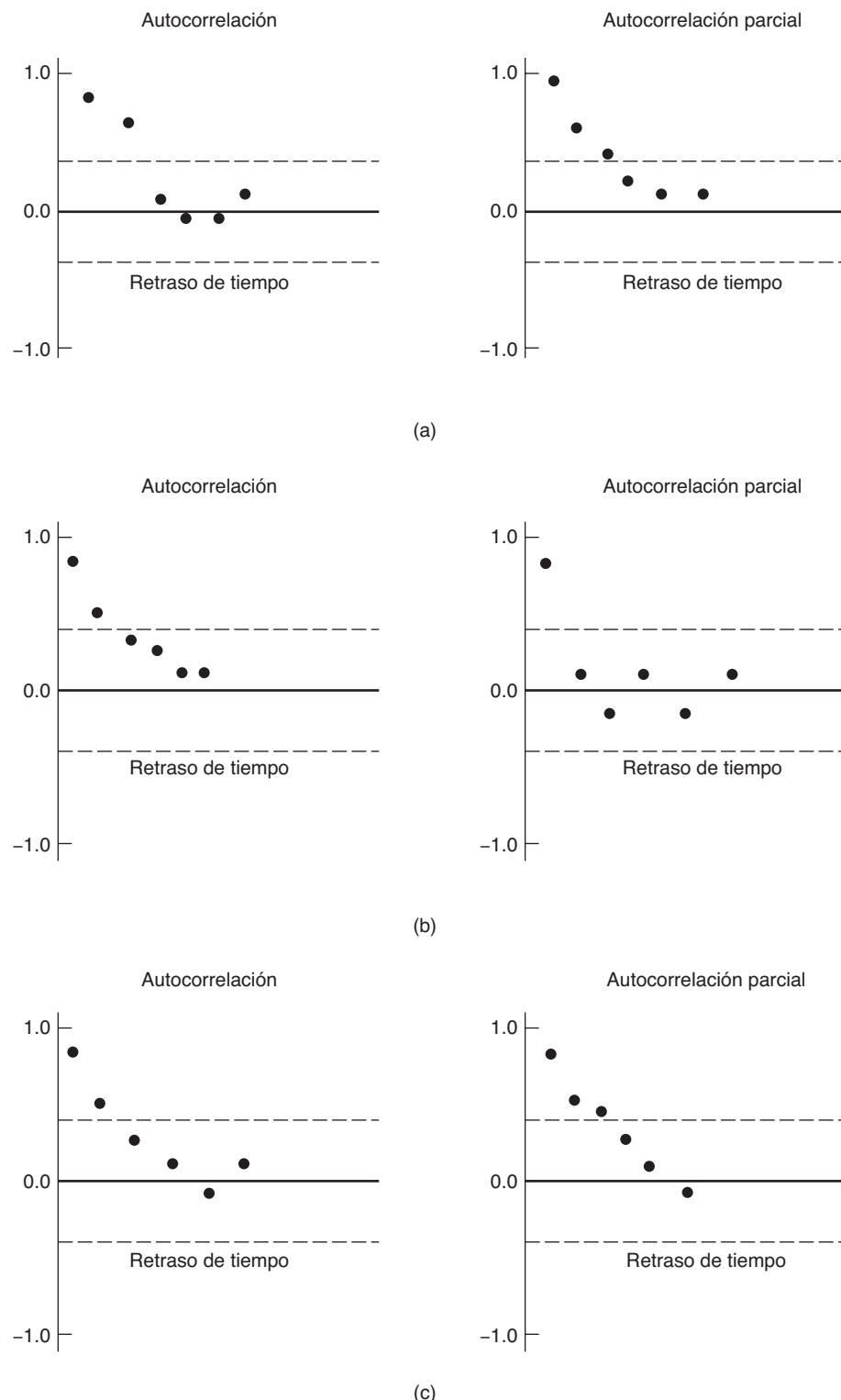
Suponga que en un tiempo  $t = 60$  la observación es  $Y_{60} = 57$ .

- a) Determine los pronósticos para los períodos 61, 62 y 63 de un origen 60.
- b) Suponga que el valor observado de  $Y_{61}$  es 59. Actualice los pronósticos para los períodos 62 y 63.
- c) Suponga que el estimado de la varianza del término del error es  $s^2 = 3.2$ . Calcule un intervalo de predicción de 95% acerca del pronóstico para el período 61.

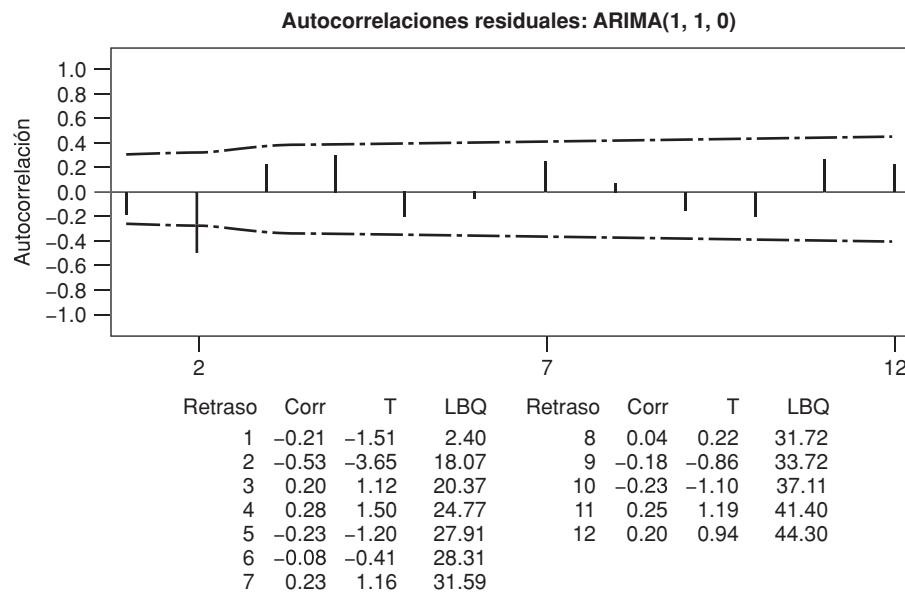
**TABLA P.4**

| <i>Modelo</i> | <i>Autocorrelaciones</i> | <i>Autocorrelaciones parciales</i> |
|---------------|--------------------------|------------------------------------|
| MA            |                          |                                    |
| AR            |                          |                                    |
| ARMA          |                          |                                    |

4. Complete la información faltante en la tabla P.4 que indique si las autocorrelaciones teóricas y las autocorrelaciones parciales se desvanecen o se cortan para estos modelos.
5. Dadas las gráficas de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra, en la figura P.5 a-c, identifique tentativamente un modelo ARIMA para cada par de gráficas.
6. Un modelo ARIMA(1, 1, 0) [modelo AR(1) para las primeras diferencias] se ajusta a las observaciones de una serie tiempo. Las primeras 12 autocorrelaciones residuales se muestran la figura P.6. El modelo fue ajustado con una constante.
  - a) Con base en la inspección de las autocorrelaciones residuales, ¿parece que el modelo es el adecuado? ¿Por qué sí o por qué no?
  - b) Si usted decidiera que el modelo no es el adecuado, indique de qué manera puede modificarse para eliminar cualquier aspecto que lo hace inadecuado.
7. Chips Bakery ha tenido problemas con los pronósticos de la demanda para su pan especial con mucha fibra y quisiera su ayuda. Los datos para la demanda semanal y las autocorrelaciones de los datos originales y diversas diferencias de los datos se muestran en las tablas P.7A-D.
  - a) Inspeccione las gráficas de autocorrelación y sugiera un modelo tentativo para estos datos. ¿Cómo decidiría sobre un modelo? (Nota: no haga más diferencias de las necesarias. Se indica demasiada diferenciación si las autocorrelaciones de la muestra de los retrasos cortos tienden a incrementarse en magnitud.)
  - b) Mediante un programa de cómputo para la modelación ARIMA ajuste y verifique su modelo identificado con los datos en la demanda del pan.
  - c) Escriba una ecuación para pronosticar la demanda del pan de fibra para el periodo 53.
  - d) Con un programa de cómputo para la modelación ARIMA, pronostique la demanda del pan de fibra para los siguientes cuatro periodos desde un origen 52 del pronóstico. Asimismo, construya los intervalos de predicción de 95%.
8. La tabla P.8 contiene una serie tiempo con 126 observaciones. Por medio de un programa de cómputo para la modelación ARIMA, obtenga una gráfica de los datos, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA apropiado y produzca los pronósticos para los siguientes tres periodos a partir del origen del pronóstico  $t = 126$ .
9. La tabla P.9 contiene una serie tiempo de 80 observaciones. Mediante un programa de cómputo para la modelación ARIMA obtenga gráficas de los datos, de las autocorrelaciones de la muestra y de las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA que sea apropiado y genere pronósticos para los siguientes tres periodos a partir del origen de pronóstico  $t = 80$ .
10. La tabla P.10 contiene una serie de tiempo de 80 observaciones. Mediante un programa de cómputo para la modelación ARIMA obtenga gráficas de los datos, de las autocorrelaciones de la muestra y de las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA apropiado y genere pronósticos para los siguientes tres periodos tomando como origen de pronóstico  $t = 80$ .
11. La tabla P.11 contiene una serie de tiempo con 96 observaciones mensuales. Por medio de un programa de cómputo para la modelación ARIMA obtenga gráficas de los



**FIGURA P.5** Autocorrelaciones de la muestra y autocorrelaciones parciales de la muestra



**FIGURA P.6 Autocorrelaciones para los residuales de un modelo ARIMA(1, 1, 0)**

**TABLA P.7A Demanda de las ventas semanales para el pan de fibra (en miles)**

| Semana | Demand | Semana | Demand | Semana | Demand | Semana | Demand |
|--------|--------|--------|--------|--------|--------|--------|--------|
| 1      | 22.46  | 14     | 30.21  | 27     | 39.29  | 40     | 47.31  |
| 2      | 20.27  | 15     | 30.09  | 28     | 39.61  | 41     | 50.08  |
| 3      | 20.97  | 16     | 33.04  | 29     | 41.02  | 42     | 50.25  |
| 4      | 23.68  | 17     | 31.21  | 30     | 42.52  | 43     | 49.00  |
| 5      | 23.25  | 18     | 32.44  | 31     | 40.83  | 44     | 49.97  |
| 6      | 23.48  | 19     | 34.73  | 32     | 42.15  | 45     | 52.52  |
| 7      | 24.81  | 20     | 34.92  | 33     | 43.91  | 46     | 53.39  |
| 8      | 25.44  | 21     | 33.37  | 34     | 45.67  | 47     | 52.37  |
| 9      | 24.88  | 22     | 36.91  | 35     | 44.53  | 48     | 54.06  |
| 10     | 27.38  | 23     | 37.75  | 36     | 45.23  | 49     | 54.88  |
| 11     | 27.74  | 24     | 35.46  | 37     | 46.35  | 50     | 54.82  |
| 12     | 28.96  | 25     | 38.48  | 38     | 46.28  | 51     | 56.23  |
| 13     | 28.48  | 26     | 37.72  | 39     | 46.70  | 52     | 57.54  |

**TABLA P.7B Autocorrelaciones de la muestra para los datos originales**

| Retraso | Autocorrelación | Retraso | Autocorrelación |
|---------|-----------------|---------|-----------------|
| 1       | .94             | 7       | .59             |
| 2       | .88             | 8       | .53             |
| 3       | .82             | 9       | .48             |
| 4       | .77             | 10      | .43             |
| 5       | .71             | 11      | .38             |
| 6       | .65             | 12      | .32             |

**TABLA P.7C** Autocorrelaciones para las primeras diferencias de la serie

| Retraso | Autocorrelación | Retraso | Autocorrelación |
|---------|-----------------|---------|-----------------|
| 1       | -.40            | 7       | .20             |
| 2       | -.29            | 8       | -.03            |
| 3       | .17             | 9       | -.03            |
| 4       | .21             | 10      | -.23            |
| 5       | -.22            | 11      | .21             |
| 6       | -.05            | 12      | .14             |

**TABLA P.7D** Autocorrelaciones para las segundas diferencias de la serie

| Retraso | Autocorrelación | Retraso | Autocorrelación |
|---------|-----------------|---------|-----------------|
| 1       | -.53            | 7       | .16             |
| 2       | -.10            | 8       | -.05            |
| 3       | .11             | 9       | .06             |
| 4       | .18             | 10      | -.23            |
| 5       | -.20            | 11      | .16             |
| 6       | -.04            | 12      | .13             |

**TABLA P.8**

| <i>t</i> | $Y_t$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1        | 60.9  | 27       | 54.2  | 53       | 56.9  | 79       | 51.6  | 105      | 52.9  |
| 2        | 64.2  | 28       | 50.9  | 54       | 58.9  | 80       | 48.2  | 106      | 56.9  |
| 3        | 64.2  | 29       | 52.2  | 55       | 58.2  | 81       | 47.6  | 107      | 51.6  |
| 4        | 59.6  | 30       | 57.6  | 56       | 60.9  | 82       | 50.2  | 108      | 48.2  |
| 5        | 62.2  | 31       | 51.6  | 57       | 51.6  | 83       | 58.2  | 109      | 47.6  |
| 6        | 60.9  | 32       | 46.9  | 58       | 54.9  | 84       | 65.6  | 110      | 50.2  |
| 7        | 54.2  | 33       | 51.6  | 59       | 66.2  | 85       | 53.6  | 111      | 58.2  |
| 8        | 56.9  | 34       | 57.6  | 60       | 57.6  | 86       | 55.6  | 112      | 65.6  |
| 9        | 58.2  | 35       | 60.2  | 61       | 48.9  | 87       | 61.6  | 113      | 53.6  |
| 10       | 56.2  | 36       | 64.2  | 62       | 46.2  | 88       | 57.6  | 114      | 55.6  |
| 11       | 60.9  | 37       | 62.2  | 63       | 50.9  | 89       | 56.2  | 115      | 61.6  |
| 12       | 56.9  | 38       | 53.6  | 64       | 57.6  | 90       | 60.9  | 116      | 57.6  |
| 13       | 55.6  | 39       | 50.9  | 65       | 54.9  | 91       | 57.6  | 117      | 56.2  |
| 14       | 52.2  | 40       | 54.2  | 66       | 51.6  | 92       | 51.6  | 118      | 60.9  |
| 15       | 58.2  | 41       | 56.2  | 67       | 50.2  | 93       | 56.2  | 119      | 57.6  |
| 16       | 62.9  | 42       | 59.6  | 68       | 50.9  | 94       | 52.2  | 120      | 51.6  |
| 17       | 57.6  | 43       | 66.2  | 69       | 56.9  | 95       | 50.2  | 121      | 56.2  |
| 18       | 58.9  | 44       | 57.6  | 70       | 50.2  | 96       | 56.9  | 122      | 52.2  |
| 19       | 59.6  | 45       | 48.9  | 71       | 54.2  | 97       | 56.9  | 123      | 50.2  |
| 20       | 55.6  | 46       | 50.9  | 72       | 58.2  | 98       | 50.2  | 124      | 56.9  |
| 21       | 60.2  | 47       | 60.2  | 73       | 56.9  | 99       | 54.2  | 125      | 56.9  |
| 22       | 62.2  | 48       | 64.2  | 74       | 55.6  | 100      | 58.2  | 126      | 50.2  |
| 23       | 58.2  | 49       | 56.9  | 75       | 64.2  | 101      | 56.9  |          |       |
| 24       | 51.6  | 50       | 56.9  | 76       | 58.2  | 102      | 55.6  |          |       |
| 25       | 50.9  | 51       | 63.6  | 77       | 52.9  | 103      | 64.2  |          |       |
| 26       | 59.6  | 52       | 58.2  | 78       | 56.9  | 104      | 58.2  |          |       |

**TABLA P.9**

| <i>t</i> | $Y_t$ | <i>t</i> | $Y_t$ | <i>t</i> | $Y_t$ | <i>t</i> | $Y_t$ |
|----------|-------|----------|-------|----------|-------|----------|-------|
| 1        | 61    | 21       | 50    | 41       | 59    | 61       | 57    |
| 2        | 50    | 22       | 69    | 42       | 49    | 62       | 56    |
| 3        | 62    | 23       | 53    | 43       | 64    | 63       | 53    |
| 4        | 47    | 24       | 57    | 44       | 55    | 64       | 55    |
| 5        | 64    | 25       | 52    | 45       | 48    | 65       | 55    |
| 6        | 40    | 26       | 66    | 46       | 61    | 66       | 66    |
| 7        | 76    | 27       | 47    | 47       | 47    | 67       | 49    |
| 8        | 38    | 28       | 67    | 48       | 58    | 68       | 57    |
| 9        | 75    | 29       | 51    | 49       | 46    | 69       | 50    |
| 10       | 41    | 30       | 57    | 50       | 58    | 70       | 68    |
| 11       | 74    | 31       | 55    | 51       | 57    | 71       | 42    |
| 12       | 47    | 32       | 64    | 52       | 52    | 72       | 77    |
| 13       | 72    | 33       | 48    | 53       | 62    | 73       | 30    |
| 14       | 47    | 34       | 65    | 54       | 46    | 74       | 88    |
| 15       | 62    | 35       | 52    | 55       | 72    | 75       | 37    |
| 16       | 57    | 36       | 65    | 56       | 37    | 76       | 88    |
| 17       | 56    | 37       | 47    | 57       | 71    | 77       | 32    |
| 18       | 53    | 38       | 68    | 58       | 33    | 78       | 90    |
| 19       | 58    | 39       | 48    | 59       | 71    | 79       | 31    |
| 20       | 61    | 40       | 61    | 60       | 47    | 80       | 85    |

**TABLA P.10**

| <i>t</i> | $Y_t$ | <i>t</i> | $Y_t$ | <i>t</i> | $Y_t$ | <i>t</i> | $Y_t$ |
|----------|-------|----------|-------|----------|-------|----------|-------|
| 1        | 91.5  | 21       | 181.5 | 41       | 125.1 | 61       | 155.6 |
| 2        | 93.0  | 22       | 183.0 | 42       | 119.0 | 62       | 169.3 |
| 3        | 106.8 | 23       | 167.8 | 43       | 117.4 | 63       | 173.8 |
| 4        | 109.8 | 24       | 143.3 | 44       | 115.9 | 64       | 170.8 |
| 5        | 114.4 | 25       | 132.7 | 45       | 94.5  | 65       | 158.6 |
| 6        | 106.8 | 26       | 134.2 | 46       | 90.0  | 66       | 151.0 |
| 7        | 105.2 | 27       | 134.2 | 47       | 135.7 | 67       | 151.0 |
| 8        | 115.9 | 28       | 126.6 | 48       | 114.4 | 68       | 158.6 |
| 9        | 123.5 | 29       | 122.0 | 49       | 106.8 | 69       | 169.3 |
| 10       | 131.2 | 30       | 131.2 | 50       | 91.5  | 70       | 199.8 |
| 11       | 138.8 | 31       | 151.0 | 51       | 96.1  | 71       | 222.7 |
| 12       | 140.3 | 32       | 161.7 | 52       | 106.8 | 72       | 233.3 |
| 13       | 120.5 | 33       | 163.2 | 53       | 108.3 | 73       | 250.1 |
| 14       | 125.1 | 34       | 163.2 | 54       | 93.0  | 74       | 271.4 |
| 15       | 134.2 | 35       | 152.5 | 55       | 93.0  | 75       | 273.0 |
| 16       | 138.8 | 36       | 131.2 | 56       | 120.5 | 76       | 273.0 |
| 17       | 152.5 | 37       | 122.0 | 57       | 131.2 | 77       | 269.9 |
| 18       | 164.7 | 38       | 126.6 | 58       | 129.6 | 78       | 260.8 |
| 19       | 161.7 | 39       | 129.6 | 59       | 143.3 | 79       | 260.8 |
| 20       | 163.2 | 40       | 125.1 | 60       | 151.0 | 80       | 266.9 |

**TABLA P.11**

| <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> |
|----------|----------------------|----------|----------------------|----------|----------------------|----------|----------------------|
| 1        | 97,575               | 26       | 106,650              | 51       | 149,850              | 76       | 166,950              |
| 2        | 97,755               | 27       | 122,550              | 52       | 149,850              | 77       | 181,500              |
| 3        | 105,825              | 28       | 124,650              | 53       | 160,350              | 78       | 173,925              |
| 4        | 104,700              | 29       | 128,925              | 54       | 161,325              | 79       | 174,450              |
| 5        | 110,700              | 30       | 126,038              | 55       | 153,225              | 80       | 179,475              |
| 6        | 108,060              | 31       | 129,300              | 56       | 159,375              | 81       | 180,300              |
| 7        | 111,825              | 32       | 127,050              | 57       | 149,325              | 82       | 177,975              |
| 8        | 114,525              | 33       | 125,025              | 58       | 157,200              | 83       | 174,525              |
| 9        | 108,375              | 34       | 129,675              | 59       | 163,350              | 84       | 220,650              |
| 10       | 117,900              | 35       | 134,250              | 60       | 190,650              | 85       | 152,625              |
| 11       | 117,600              | 36       | 161,475              | 61       | 141,600              | 86       | 143,100              |
| 12       | 142,800              | 37       | 115,350              | 62       | 136,725              | 87       | 164,925              |
| 13       | 109,500              | 38       | 111,750              | 63       | 158,625              | 88       | 167,175              |
| 14       | 103,995              | 39       | 134,025              | 64       | 163,050              | 89       | 181,725              |
| 15       | 119,250              | 40       | 134,325              | 65       | 173,925              | 90       | 174,150              |
| 16       | 121,200              | 41       | 140,175              | 66       | 165,900              | 91       | 174,675              |
| 17       | 123,900              | 42       | 144,300              | 67       | 226,650              | 92       | 179,700              |
| 18       | 124,350              | 43       | 143,475              | 68       | 175,050              | 93       | 180,525              |
| 19       | 125,775              | 44       | 143,700              | 69       | 164,700              | 94       | 178,200              |
| 20       | 126,900              | 45       | 140,325              | 70       | 167,625              | 95       | 174,750              |
| 21       | 121,650              | 46       | 144,375              | 71       | 171,225              | 96       | 220,875              |
| 22       | 126,600              | 47       | 149,175              | 72       | 203,550              |          |                      |
| 23       | 124,350              | 48       | 182,400              | 73       | 152,400              |          |                      |
| 24       | 152,400              | 49       | 132,900              | 74       | 142,875              |          |                      |
| 25       | 108,150              | 50       | 133,950              | 75       | 164,700              |          |                      |

**TABLA P.12**

| <i>Periodo</i> | <i>IBM</i> | <i>Periodo</i> | <i>IBM</i> | <i>Periodo</i> | <i>IBM</i> | <i>Periodo</i> | <i>IBM</i> |
|----------------|------------|----------------|------------|----------------|------------|----------------|------------|
| Ene. 6         | 267        | Abr. 7         | 241        | Jul. 7         | 258        | Oct. 6         | 279        |
| 13             | 267        | 14             | 244        | 14             | 259        | 13             | 287        |
| 20             | 268        | 21             | 254        | 21             | 268        | 20             | 276        |
| 27             | 264        | 28             | 262        | 28             | 276        | 27             | 273        |
| Feb. 3         | 263        | May 5          | 261        | Ago. 4         | 285        | Nov. 3         | 270        |
| 10             | 260        | 12             | 265        | 11             | 288        | 10             | 264        |
| 17             | 256        | 19             | 261        | 18             | 295        | 17             | 261        |
| 24             | 256        | 26             | 261        | 25             | 297        | 24             | 268        |
| Mar. 2         | 252        | Jun. 2         | 257        | Sep. 1         | 292        | Dic. 1         | 270        |
| 10             | 245        | 9              | 268        | 8              | 299        | 8              | 276        |
| 17             | 243        | 16             | 270        | 15             | 294        | 15             | 274        |
| 24             | 240        | 23             | 266        | 22             | 284        | 22             | 284        |
| 31             | 238        | 30             | 259        | 29             | 277        | 29             | 304        |

datos, de las autocorrelaciones de la muestra y de las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA apropiado y genere pronósticos para los siguientes 12 períodos desde un origen de pronóstico  $t = 96$ .

12. Los datos de la tabla P.12 son los precios semanales de las acciones de IBM.
  - a) Por medio de un programa de cómputo para la modelación ARIMA obtenga gráficas de los datos, de las autocorrelaciones de la muestra y de las autocorrelaciones parciales de la muestra. Utilice esta información para identificar de manera tentativa un modelo ARIMA que sea adecuado para esta serie.
  - b) ¿Es estacionaria la serie de IBM? ¿Qué corrección recomendaría usted si la serie no fuera estacionaria?
  - c) Ajuste un modelo ARIMA para la serie de IBM. Interprete el resultado. ¿Son sucesivos los cambios aleatorios?
  - d) Lleve a cabo las evaluaciones de diagnóstico para determinar si el modelo que usted ha ajustado es adecuado.
  - e) Después de que se ha encontrado un modelo satisfactorio, pronostique el precio accionario de IBM para la primera semana de enero del siguiente año. ¿De qué manera difiere su pronóstico del pronóstico simple, el cual dice que el pronóstico de la primera semana de enero es el precio para la última semana de diciembre (precio actual)?
13. Los datos en la tabla P.13 son cotizaciones accionarias de cierre para 150 días de DEF Corporation. Determine un modelo ARIMA apropiado y pronostique el precio accionario para los siguientes cinco días a partir del origen del pronóstico  $t = 145$ . Compare sus pronósticos con los precios reales por medio de *MAPE*. ¿Qué tan precisos son sus pronósticos?

TABLA P.13

| <i>Periodo</i> | <i>DEF</i> | <i>Periodo</i> | <i>DEF</i> | <i>Periodo</i> | <i>DEF</i> |
|----------------|------------|----------------|------------|----------------|------------|
| 1              | 136.0      | 29             | 136.0      | 57             | 134.4      |
| 2              | 132.8      | 30             | 142.4      | 58             | 140.8      |
| 3              | 130.4      | 31             | 140.0      | 59             | 137.6      |
| 4              | 128.8      | 32             | 144.8      | 60             | 132.8      |
| 5              | 136.8      | 33             | 140.0      | 61             | 136.8      |
| 6              | 135.2      | 34             | 139.2      | 62             | 135.2      |
| 7              | 134.4      | 35             | 139.2      | 63             | 132.8      |
| 8              | 139.2      | 36             | 136.8      | 64             | 144.0      |
| 9              | 136.8      | 37             | 140.8      | 65             | 137.6      |
| 10             | 136.0      | 38             | 141.6      | 66             | 138.4      |
| 11             | 133.6      | 39             | 139.2      | 67             | 136.0      |
| 12             | 139.2      | 40             | 142.4      | 68             | 135.2      |
| 13             | 137.6      | 41             | 140.8      | 69             | 138.4      |
| 14             | 139.2      | 42             | 140.0      | 70             | 134.4      |
| 15             | 139.2      | 43             | 132.0      | 71             | 138.4      |
| 16             | 136.0      | 44             | 142.4      | 72             | 139.2      |
| 17             | 138.4      | 45             | 138.4      | 73             | 141.6      |
| 18             | 137.6      | 46             | 138.4      | 74             | 134.4      |
| 19             | 139.2      | 47             | 136.8      | 75             | 135.2      |
| 20             | 134.4      | 48             | 139.2      | 76             | 136.0      |
| 21             | 136.8      | 49             | 135.2      | 77             | 135.2      |
| 22             | 139.2      | 50             | 138.4      | 78             | 136.0      |
| 23             | 139.2      | 51             | 140.8      | 79             | 132.8      |
| 24             | 140.0      | 52             | 135.2      | 80             | 133.6      |
| 25             | 139.2      | 53             | 133.6      | 81             | 134.4      |
| 26             | 140.8      | 54             | 134.4      | 82             | 133.6      |
| 27             | 139.2      | 55             | 134.4      | 83             | 131.2      |
| 28             | 138.4      | 56             | 137.6      | 84             | 132.0      |

**TABLA P.13** *(Continúa)*

| <i>Periodo</i> | <i>DEF</i> | <i>Periodo</i> | <i>DEF</i> | <i>Periodo</i> | <i>DEF</i> |
|----------------|------------|----------------|------------|----------------|------------|
| 85             | 131.2      | 107            | 129.6      | 129            | 137.6      |
| 86             | 132.8      | 108            | 132.8      | 130            | 137.6      |
| 87             | 132.0      | 109            | 135.2      | 131            | 140.0      |
| 88             | 133.6      | 110            | 132.0      | 132            | 135.2      |
| 89             | 131.2      | 111            | 132.8      | 133            | 135.2      |
| 90             | 131.2      | 112            | 132.8      | 134            | 135.2      |
| 91             | 129.6      | 113            | 136.0      | 135            | 136.0      |
| 92             | 131.2      | 114            | 136.8      | 136            | 132.0      |
| 93             | 130.4      | 115            | 136.8      | 137            | 133.6      |
| 94             | 131.2      | 116            | 133.6      | 138            | 134.4      |
| 95             | 136.0      | 117            | 134.4      | 139            | 133.6      |
| 96             | 135.2      | 118            | 130.4      | 140            | 133.6      |
| 97             | 136.8      | 119            | 132.8      | 141            | 132.8      |
| 98             | 136.8      | 120            | 134.4      | 142            | 132.0      |
| 99             | 133.6      | 121            | 135.2      | 143            | 136.0      |
| 100            | 135.2      | 122            | 136.8      | 144            | 133.6      |
| 101            | 136.0      | 123            | 134.4      | 145            | 133.6      |
| 102            | 137.6      | 124            | 136.0      | 146            | 135.2      |
| 103            | 131.2      | 125            | 137.6      | 147            | 139.2      |
| 104            | 136.0      | 126            | 138.4      | 148            | 136.8      |
| 105            | 136.0      | 127            | 137.6      | 149            | 136.0      |
| 106            | 133.6      | 128            | 138.4      | 150            | 134.4      |

**TABLA P.14**

| <i>Periodo</i> | <i>Observación</i> | <i>Periodo</i> | <i>Observación</i> | <i>Periodo</i> | <i>Observación</i> |
|----------------|--------------------|----------------|--------------------|----------------|--------------------|
| 1              | 101                | 25             | 64                 | 49             | 78                 |
| 2              | 84                 | 26             | 58                 | 50             | 83                 |
| 3              | 54                 | 27             | 44                 | 51             | 87                 |
| 4              | 39                 | 28             | 26                 | 52             | 64                 |
| 5              | 26                 | 29             | 24                 | 53             | 44                 |
| 6              | 40                 | 30             | 18                 | 54             | 24                 |
| 7              | 99                 | 31             | 16                 | 55             | 29                 |
| 8              | 148                | 32             | 17                 | 56             | 73                 |
| 9              | 147                | 33             | 21                 | 57             | 138                |
| 10             | 134                | 34             | 28                 | 58             | 154                |
| 11             | 106                | 35             | 30                 | 59             | 119                |
| 12             | 83                 | 36             | 51                 | 60             | 102                |
| 13             | 76                 | 37             | 62                 | 61             | 79                 |
| 14             | 63                 | 38             | 57                 | 62             | 53                 |
| 15             | 57                 | 39             | 46                 | 63             | 40                 |
| 16             | 37                 | 40             | 40                 | 64             | 27                 |
| 17             | 32                 | 41             | 32                 | 65             | 31                 |
| 18             | 22                 | 42             | 23                 | 66             | 56                 |
| 19             | 20                 | 43             | 20                 | 67             | 78                 |
| 20             | 23                 | 44             | 18                 | 68             | 114                |
| 21             | 30                 | 45             | 24                 | 69             | 140                |
| 22             | 50                 | 46             | 33                 | 70             | 112                |
| 23             | 61                 | 47             | 52                 | 71             | 82                 |
| 24             | 59                 | 48             | 66                 | 72             | 80                 |

**TABLA P.14 (Continúa)**

| <i>Periodo</i> | <i>Observación</i> | <i>Periodo</i> | <i>Observación</i> | <i>Periodo</i> | <i>Observación</i> |
|----------------|--------------------|----------------|--------------------|----------------|--------------------|
| 73             | 70                 | 79             | 71                 | 85             | 63                 |
| 74             | 55                 | 80             | 110                | 86             | 46                 |
| 75             | 37                 | 81             | 112                | 87             | 32                 |
| 76             | 23                 | 82             | 93                 | 88             | 23                 |
| 77             | 20                 | 83             | 75                 | 89             | 53                 |
| 78             | 39                 | 84             | 60                 | 90             | 90                 |

14. Los datos en la tabla P.14 son el número de accidentes vehiculares semanales para los años de 1996 a 1997 en el condado de Havana. Determine el modelo ARIMA apropiado y pronostique los accidentes para la semana 91. Comente la precisión de su pronóstico.
15. La tabla P.15 proporciona 120 observaciones mensuales acerca del precio en centavos por silo de maíz en Omaha, Nebraska. Determine el mejor modelo ARIMA para estos datos. Genere pronósticos del precio del maíz para los siguientes 12 meses. Comente el patrón de los pronósticos.

**TABLA P.15**

| <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> |
|----------|----------------------|----------|----------------------|----------|----------------------|
| 1        | 125                  | 30       | 128                  | 59       | 358                  |
| 2        | 126                  | 31       | 130                  | 60       | 359                  |
| 3        | 126                  | 32       | 129                  | 61       | 333                  |
| 4        | 129                  | 33       | 139                  | 62       | 310                  |
| 5        | 132                  | 34       | 131                  | 63       | 293                  |
| 6        | 136                  | 35       | 131                  | 64       | 285                  |
| 7        | 139                  | 36       | 166                  | 65       | 284                  |
| 8        | 137                  | 37       | 159                  | 66       | 283                  |
| 9        | 155                  | 38       | 163                  | 67       | 291                  |
| 10       | 144                  | 39       | 158                  | 68       | 312                  |
| 11       | 142                  | 40       | 165                  | 69       | 288                  |
| 12       | 153                  | 41       | 206                  | 70       | 280                  |
| 13       | 159                  | 42       | 252                  | 71       | 255                  |
| 14       | 160                  | 43       | 215                  | 72       | 258                  |
| 15       | 156                  | 44       | 238                  | 73       | 266                  |
| 16       | 153                  | 45       | 249                  | 74       | 271                  |
| 17       | 151                  | 46       | 250                  | 75       | 272                  |
| 18       | 158                  | 47       | 250                  | 76       | 265                  |
| 19       | 150                  | 48       | 275                  | 77       | 281                  |
| 20       | 129                  | 49       | 292                  | 78       | 294                  |
| 21       | 117                  | 50       | 304                  | 79       | 300                  |
| 22       | 112                  | 51       | 301                  | 80       | 284                  |
| 23       | 109                  | 52       | 265                  | 81       | 282                  |
| 24       | 126                  | 53       | 271                  | 82       | 250                  |
| 25       | 123                  | 54       | 289                  | 83       | 231                  |
| 26       | 121                  | 55       | 319                  | 84       | 247                  |
| 27       | 123                  | 56       | 377                  | 85       | 252                  |
| 28       | 125                  | 57       | 350                  | 86       | 250                  |
| 29       | 128                  | 58       | 378                  | 87       | 253                  |

| <b>TABLA P.15</b> <i>(Continúa)</i> |                      |          |                      |          |                      |
|-------------------------------------|----------------------|----------|----------------------|----------|----------------------|
| <i>t</i>                            | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> | <i>t</i> | <i>Y<sub>t</sub></i> |
| 88                                  | 255                  | 99       | 256                  | 110      | 239                  |
| 89                                  | 236                  | 100      | 253                  | 111      | 251                  |
| 90                                  | 221                  | 101      | 250                  | 112      | 261                  |
| 91                                  | 204                  | 102      | 232                  | 113      | 268                  |
| 92                                  | 187                  | 103      | 221                  | 114      | 304                  |
| 93                                  | 187                  | 104      | 214                  | 115      | 285                  |
| 94                                  | 182                  | 105      | 221                  | 116      | 274                  |
| 95                                  | 210                  | 106      | 225                  | 117      | 281                  |
| 96                                  | 221                  | 107      | 224                  | 118      | 259                  |
| 97                                  | 222                  | 108      | 228                  | 119      | 270                  |
| 98                                  | 236                  | 109      | 233                  | 120      | 251                  |

16. Utilice la metodología Box-Jenkins para modelar y pronosticar las ventas mensuales de Cavanaugh Company que se proporcionan en la tabla P.14 en el problema 14 del capítulo 5. (Sugerencia: considere una transformación logarítmica antes de modelar estos datos.)
17. Utilice la metodología Box-Jenkins para modelar y pronosticar las ventas trimestrales de Disney Company que se proporcionan en la tabla P.16 en el problema 16 del capítulo 5. (Sugerencia: considere una transformación logarítmica antes de modelar estos datos.)
18. Utilice la metodología Box-Jenkins para modelar y pronosticar la demanda mensual de gasolina de Yukong Oil Company que se muestra en la tabla P.17 en el problema 17 del capítulo 5.

## CASOS

---

### CASO 9-1 VENTAS DEL RESTAURANTE

Este caso hace referencia a los datos de las ventas del restaurante que se mencionan en el caso 8.3. Jim Price ha terminado el curso de pronósticos y está ansioso por aplicar la metodología de Box-Jenkins a los datos de las ventas del restaurante. Estos datos, que se muestran en

la tabla 9.17A, comienzan con la semana que termina el domingo 4 de enero de 1981 y continúa hasta el domingo 26 de diciembre de 1982. La tabla 9.17B contiene datos nuevos para la semana que terminó el 2 de enero de 1983 hasta la que terminó el 30 de octubre de 1983. ■

| <b>TABLA 9.17A</b> <i>Ventas del restaurante: datos anteriores para el caso 9.1</i> |                    |                              |                    |                              |                    |
|-------------------------------------------------------------------------------------|--------------------|------------------------------|--------------------|------------------------------|--------------------|
| <i>Semana que termina el</i>                                                        | <i>Ventas (\$)</i> | <i>Semana que termina el</i> | <i>Ventas (\$)</i> | <i>Semana que termina el</i> | <i>Ventas (\$)</i> |
| 1/4/81                                                                              | 1,688              | 3/8/81                       | 3,460              | 5/10/81                      | 4,103              |
| 1/11/81                                                                             | 2,514              | 3/15/81                      | 4,517              | 5/17/81                      | 6,594              |
| 1/18/81                                                                             | 5,843              | 3/22/81                      | 5,188              | 5/24/81                      | 5,742              |
| 1/25/81                                                                             | 4,912              | 3/29/81                      | 5,944              | 5/31/81                      | 3,714              |
| 2/1/81                                                                              | 5,133              | 4/5/81                       | 5,842              | 6/7/81                       | 3,399              |
| 2/8/81                                                                              | 4,563              | 4/12/81                      | 6,589              | 6/14/81                      | 3,376              |
| 2/15/81                                                                             | 5,416              | 4/19/81                      | 5,447              | 6/21/81                      | 3,627              |
| 2/22/81                                                                             | 6,416              | 4/26/81                      | 7,548              | 6/28/81                      | 4,201              |
| 3/1/81                                                                              | 5,879              | 5/3/81                       | 6,403              | 7/5/81                       | 3,515              |

**TABLA 9.17A** *(Continúa)*

| <i>Semana que termina el</i> | <i>Ventas (\$)</i> | <i>Semana que termina el</i> | <i>Ventas (\$)</i> | <i>Semana que termina el</i> | <i>Ventas (\$)</i> |
|------------------------------|--------------------|------------------------------|--------------------|------------------------------|--------------------|
| 7/12/81                      | 3,645              | 1/10/82                      | 1,870              | 7/11/82                      | 3,885              |
| 7/19/81                      | 3,416              | 1/17/82                      | 3,962              | 7/18/82                      | 4,209              |
| 7/26/81                      | 3,565              | 1/24/82                      | 5,973              | 7/25/82                      | 3,614              |
| 8/2/81                       | 2,428              | 1/31/82                      | 5,009              | 8/1/82                       | 3,722              |
| 8/9/81                       | 3,292              | 2/7/82                       | 5,328              | 8/8/82                       | 4,307              |
| 8/16/81                      | 3,460              | 2/14/82                      | 5,014              | 8/15/82                      | 3,322              |
| 8/23/81                      | 6,212              | 2/21/82                      | 4,986              | 8/22/82                      | 5,962              |
| 8/30/81                      | 6,057              | 2/28/82                      | 5,213              | 8/29/82                      | 6,784              |
| 9/6/81                       | 5,739              | 3/7/82                       | 4,807              | 9/5/82                       | 6,069              |
| 9/13/81                      | 5,560              | 3/14/82                      | 3,964              | 9/12/82                      | 5,897              |
| 9/20/81                      | 5,335              | 3/21/82                      | 5,201              | 9/19/82                      | 5,916              |
| 9/27/81                      | 5,305              | 3/28/82                      | 4,863              | 9/26/82                      | 4,998              |
| 10/4/81                      | 5,364              | 4/4/82                       | 5,019              | 10/3/82                      | 5,111              |
| 10/11/81                     | 5,511              | 4/11/82                      | 4,868              | 10/10/82                     | 5,612              |
| 10/18/81                     | 5,698              | 4/18/82                      | 5,777              | 10/17/82                     | 5,906              |
| 10/25/81                     | 5,382              | 4/25/82                      | 6,543              | 10/24/82                     | 6,010              |
| 11/1/81                      | 5,629              | 5/2/82                       | 6,352              | 10/31/82                     | 5,937              |
| 11/8/81                      | 5,617              | 5/9/82                       | 5,837              | 11/7/82                      | 6,004              |
| 11/15/81                     | 5,742              | 5/16/82                      | 7,162              | 11/14/82                     | 5,959              |
| 11/22/81                     | 3,747              | 5/23/82                      | 4,997              | 11/21/82                     | 4,223              |
| 11/29/81                     | 4,159              | 5/30/82                      | 4,063              | 11/28/82                     | 4,679              |
| 12/6/81                      | 4,853              | 6/6/82                       | 3,942              | 12/5/82                      | 5,307              |
| 12/13/81                     | 5,607              | 6/13/82                      | 4,011              | 12/12/82                     | 6,101              |
| 12/20/81                     | 3,946              | 6/20/82                      | 3,999              | 12/19/82                     | 6,896              |
| 12/27/81                     | 1,919              | 6/27/82                      | 4,794              | 12/26/82                     | 2,214              |
| 1/3/82                       | 1,898              | 7/4/82                       | 4,956              |                              |                    |

**TABLA 9.17B** Ventas del restaurante: datos nuevos para el caso 9.1

| <i>Semana que termina el</i> | <i>Ventas (\$)</i> | <i>Semana que termina el</i> | <i>Ventas (\$)</i> | <i>Semana que termina el</i> | <i>Ventas (\$)</i> |
|------------------------------|--------------------|------------------------------|--------------------|------------------------------|--------------------|
| 1/2/83                       | 2,431              | 4/17/83                      | 6,357              | 7/31/83                      | 3,558              |
| 1/9/83                       | 2,796              | 4/24/83                      | 7,273              | 8/7/83                       | 3,791              |
| 1/16/83                      | 4,432              | 5/1/83                       | 8,678              | 8/14/83                      | 3,946              |
| 1/23/83                      | 5,714              | 5/8/83                       | 7,418              | 8/21/83                      | 3,054              |
| 1/30/83                      | 5,324              | 5/15/83                      | 10,388             | 8/28/83                      | 6,893              |
| 2/6/83                       | 5,011              | 5/22/83                      | 4,940              | 9/4/83                       | 8,037              |
| 2/13/83                      | 5,336              | 5/29/83                      | 4,195              | 9/11/83                      | 6,884              |
| 2/20/83                      | 4,999              | 6/5/83                       | 3,895              | 9/18/83                      | 7,143              |
| 2/27/83                      | 5,340              | 6/12/83                      | 3,762              | 9/25/83                      | 8,318              |
| 3/6/83                       | 5,009              | 6/19/83                      | 3,739              | 10/2/83                      | 6,871              |
| 3/13/83                      | 5,590              | 6/26/83                      | 3,975              | 10/9/83                      | 6,514              |
| 3/20/83                      | 3,308              | 7/3/83                       | 4,634              | 10/16/83                     | 6,656              |
| 3/27/83                      | 6,558              | 7/10/83                      | 4,891              | 10/23/83                     | 6,484              |
| 4/3/83                       | 4,991              | 7/17/83                      | 3,463              | 10/30/83                     | 6,125              |
| 4/10/83                      | 6,952              | 7/24/83                      | 3,536              |                              |                    |

## PREGUNTAS

1. ¿Cuál es el modelo Box-Jenkins apropiado para usarse con los datos originales?
2. ¿Cuáles son sus pronósticos para las primeras cuatro semanas de enero de 1983?
3. ¿De qué manera se comparan estos pronósticos con las ventas reales?
4. De qué manera se comprara el modelo Box-Jenkins con los modelos de regresión empleados en el capítulo 8?
5. ¿Utilizaría el mismo modelo Box-Jenkins si los datos nuevos se combinaran con los anteriores?

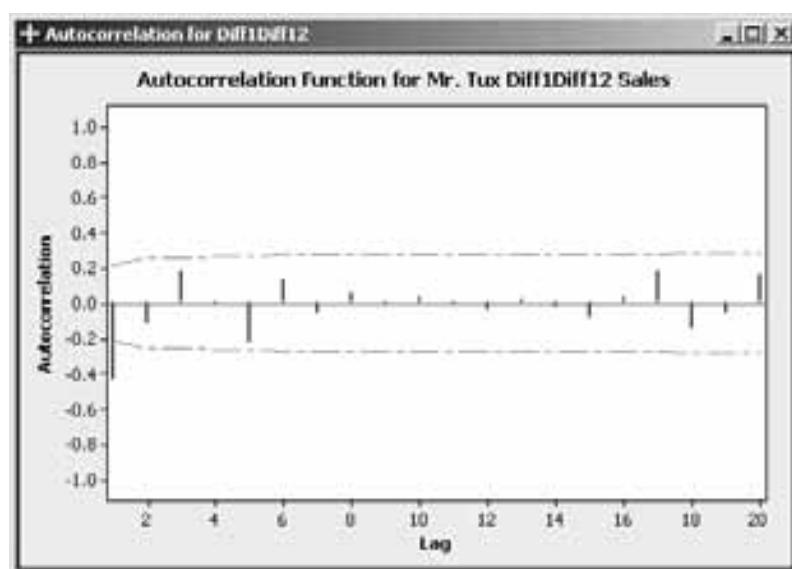
## CASO 9-2 MR. TUX

John Mosby ha decidido poner a prueba el método de Box-Jenkins para pronosticar sus ventas mensuales y comprende que este método es mucho más complejo que los métodos más sencillos que ha venido utilizando. También sabe que es posible obtener pronósticos más precisos gracias a este método más avanzado. Él tiene acceso

al programa de cómputo de Minitab que puede ajustarse a los modelos ARIMA. Debido a que ya tiene los datos recopilados y almacenados en su computadora, opta por esa alternativa.

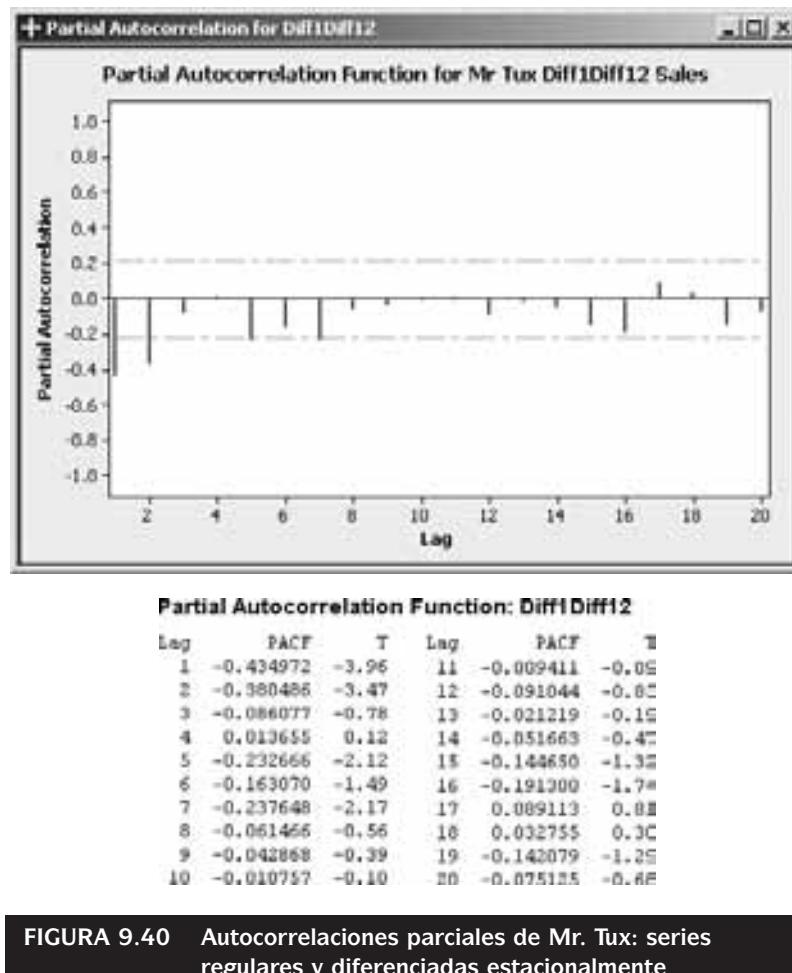
John decide utilizar la serie completa de datos, que consiste en 96 meses de datos de ventas. John sabe que

**FIGURA 9.39 Autocorrelaciones parciales de Mr. Tux: series regulares y diferenciadas estacionalmente**



**Autocorrelation Function: Diff1Diff12**

| Lag | ACF       | T     | LBB   | Lag | ACF       | T     | LBB   |
|-----|-----------|-------|-------|-----|-----------|-------|-------|
| 1   | -0.434972 | -3.96 | 16.20 | 11  | -0.006051 | -0.04 | 26.98 |
| 2   | -0.119297 | -0.93 | 17.52 | 12  | -0.037127 | -0.27 | 27.08 |
| 3   | 0.177448  | 1.36  | 20.29 | 13  | 0.024916  | 0.18  | 27.18 |
| 4   | -0.013968 | -0.10 | 20.31 | 14  | -0.025145 | -0.18 | 27.21 |
| 5   | -0.223625 | -1.66 | 24.04 | 15  | -0.001500 | -0.58 | 27.92 |
| 6   | 0.126474  | 0.92  | 26.30 | 16  | 0.028684  | 0.20  | 28.08 |
| 7   | -0.053425 | -0.38 | 26.57 | 17  | 0.104183  | 1.31  | 31.62 |
| 8   | 0.054957  | 0.39  | 26.85 | 18  | -0.148161 | -1.03 | 34.01 |
| 9   | -0.006380 | -0.05 | 26.85 | 19  | -0.059354 | -0.41 | 34.32 |
| 10  | 0.032094  | 0.23  | 26.95 | 20  | 0.165452  | 1.14  | 37.48 |



**FIGURA 9.40 Autocorrelaciones parciales de Mr. Tux: series regulares y diferenciadas estacionalmente**

un modelo ARIMA puede manejar la estructura estacional en sus series de tiempo de la misma manera en que modela la correlación de un mes a otro.

John calculó la función de la autocorrelación de la muestra para sus datos en el caso 3.2 (vea la figura 3.25 de la página 93) y determinó que sus datos mostraban una tendencia (no estacionaria). Entonces, calculó la función de autocorrelación para las primeras diferencias de sus datos de ventas. El resultado se muestra en la fi-

gura 3.26 de la página 93. De inmediato, John observó los picos significativos en los retrasos 12 y 24, los que indicaban estacionalidad y quizás la necesidad de una diferencia estacional. La función de autocorrelación de la muestra para los datos, después de una diferencia regular (primera) y una diferencia estacional, se presenta en la figura 9.39. Las autocorrelaciones parciales de la muestra para los datos diferenciados se ilustran en la figura 9.40. ■

## PREGUNTAS

1. Comente los problemas, si los hay, de explicar el método de Box-Jenkins al banquero de John y a otros miembros del equipo de administración.
2. Dadas las autocorrelaciones que se presentan en la figura 9.39 y las autocorrelaciones parciales de la figura 9.40, ¿cuáles términos regulares (no estacionales) podría incluir John en un modelo ARIMA para las ventas de Mr. Tux? ¿Qué términos estacionales podrían incluirse en el modelo?
3. Por medio de un programa para la modelación de ARIMA, ajuste y revise el modelo ARIMA empleado para las ventas de Mr. Tux. Genere pronósticos para los siguientes 12 meses.

**CASO 9-3****CONSUMER CREDIT  
COUNSELING**

La operación de Consumer Credit Counseling (CCC) se describió en los capítulos 1 (caso 1.2) y 3 (caso 3.3).

El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de clientes nuevos que habrían de ser atendidos durante el resto de 1993. Marv proporcionó a Dorothy Mercer los datos mensuales de los clientes nuevos atendidos por CCC durante el periodo de enero de 1985 a marzo de 1993 (vea el caso 3.3).

Hasta el momento, Dorothy ha estado muy contenta con su trabajo de pronóstico. No obstante, usted no está del todo satisfecho y decide que ahora es tiempo de impresionarla de verdad. Usted le dice que debería intentar una técnica más poderosa utilizada para realizar

pronósticos, la metodología Box-Jenkins y los modelos ARIMA. Dorothy nunca ha escuchado hablar de Box-Jenkins, pero está dispuesta a emplearla.

**Tarea**

1. Escriba un memorándum a Dorothy para explicarle la metodología Box-Jenkins.
2. Desarrolle un modelo ARIMA por medio de la metodología Box-Jenkins y realice un pronóstico para el número mensual de clientes nuevos para el resto de 1993.
3. Escriba un segundo memorándum dirigido a Dorothy en donde resuma los resultados obtenidos a partir de este análisis. ■

**CASO 9-4****LYDIA E. PINKHAM  
MEDICINE COMPANY<sup>15</sup>**

La Lydia E. Pinkham Medicine Company fue una empresa familiar con ingresos que en gran medida, provenían de la venta del Compuesto de vegetales de Lydia Pinkham. Quizá hoy día los alumnos pudieran utilizar el compuesto para aliviarse del estrés; desafortunadamente, ya no está a la venta. La fotografía de Lydia Pinkham estaba en la etiqueta y el producto estaba dirigido a un mercado femenino. La publicidad del compuesto incluía esta invitación: "Escriba con toda libertad a la señora Pinkham a Lynn, Mass., y asegúrese de recibir el consejo gratuito que le ofrece a todas las mujeres. Este consejo ha traído luz a muchos hogares que, con nerviosismo e irritabilidad, casi se habían colapsado." De hecho, la compañía aseguraba que una empleada contestaba cada una de las cartas. Las mujeres le escribían a la señora Pinkham. Algunos de sus comentarios eran: "Sin el Compuesto de vegetales de Lydia Pinkham ahora yo estaría muerta, o peor aún, loca. Ya me había dado por vencida y como intenté tantas cosas, ya no creía que

algo me fuera de utilidad. Ahora estoy sana y fuerte; de hecho, soy totalmente otra persona." Éste y otros testimonios se reprodujeron en anuncios publicitarios del compuesto.

La naturaleza singular de la compañía (un producto dominante que representaba la mayoría de las ventas de la empresa, ningún personal de ventas y una gran proporción de los ingresos obtenidos por ventas invertidos en publicidad) y la disponibilidad de los datos tanto en ventas como en publicidad, llevó al Comité de Fijación de Precios del National Bureau of Economic Research (NBER, por sus siglas en inglés), en 1943, a recomendar que los datos fueran sujetos a un análisis exhaustivo. La investigación no se llevó a cabo en algunos años.<sup>16</sup> Los analistas han estudiado los datos con modelos informales que incluyen datos de publicidad y variables económicas (similares a las que se presentan en el capítulo 8). Sin embargo, varios investigadores han sugerido que la metodología Box-Jenkins, que sólo

<sup>15</sup> Este caso fue aportado por la doctora Susan C. White, de la George Washington University, en Washington D.C. Para obtener información adicional vea Susan C. White, "Predicting Time Series with Neural Networks versus Statistical Models: The Lydia Pinkham Data". *Proceedings of the 24<sup>th</sup> Annual Conference of the Decision Sciences Institute, Southwestern Region*, 1993: 108-110.

<sup>16</sup> Vea Palda (1964).

usa los datos de ventas, proporciona predicciones comparables e incluso mejores cuando se comparan con las metodologías causales.<sup>17</sup> Los datos de ventas son atractivos para su estudio por dos razones:

1. El producto en sí mismo ha permanecido sin cambio alguno durante el periodo que se considera para los datos, no hay cambios en la serie ocasionados por modificaciones en el producto.
2. No hubo cambio alguno en la fuerza de las ventas durante el lapso en que se consideran los datos y la proporción de las utilidades invertidas en publicidad fue constante. De esta manera, no hay cambios en los datos ocasionados por promociones especiales o por cualquier otro fenómeno de marketing.

De manera típica, los datos reales no son tan “limpios” en términos de continuidad de producto y marketing.

Entonces, la tarea que hay que realizar consiste en determinar cuál modelo Box-Jenkins (ARIMA) es el “mejor” para estos datos. El modelo se desarrollará utilizando los datos de 1907-1948 y evaluando con los datos de 1949-1960, como se muestra en la tabla 9.18.

## IDENTIFICACIÓN DEL MODELO

Se utilizó un programa de cómputo capaz de realizar un modelado ARIMA para examinar los datos de 1907 a 1948; los datos de 1949 a 1960 se emplean para validar la competencia del modelo seleccionado para realizar pronósticos. Las pruebas preliminares sugieren que los datos son estacionarios (es decir, que no existe tendencia aparente), de manera que no se aplica la diferenciación. Después de explorar las autocorrelaciones y las autocorrelaciones parciales se determinó que un modelo AR no era el más adecuado para los datos. [La tabla 9.19 presenta las autocorrelaciones (ACF) y las autocorrelaciones parciales (PACF) para 10 periodos]. Al parecer, las autocorrelaciones y las autocorrelaciones parciales tenían coherencia con las del proceso AR(2). Para verificar el orden  $p$  del componente AR se utilizó el criterio de información de Akaike (*AIC*, vea la ecuación 9.7) con los modelos autorregresivos de órdenes  $p = 1, 2$  y  $3$ . El *AIC* lleva a la selección de un modelo AR(2) para los datos de Lydia Pinkham.

## ESTIMACIÓN DEL MODELO Y VALIDACIÓN DE SU ADECUACIÓN

Con los datos de 1907 a 1948 se usó un programa de computadora para estimar los parámetros (incluyendo

un término constante) del modelo AR(2). El modelo resultante es

$$Y_t = 178.6 + 1.423Y_{t-1} - .521Y_{t-2} + \epsilon_t \quad (.137) \quad (.136)$$

donde los números que se encuentran entre paréntesis junto a los coeficientes autorregresivos son sus desviaciones estándar estimadas. Cada coeficiente autorregresivo es significativamente distinto de cero para cualquier nivel razonable de significancia. Todas las autocorrelaciones residuales fueron pequeñas y cada una estuvo dentro de 95% de sus límites de error. La estadística chi cuadrada de Ljung-Box tenía valores de  $p$  de 0.63, 0.21 y 0.64 para los grupos de retrasos  $m = 12, 24$  y  $36$ , respectivamente. En apariencia, el modelo AR(2) es adecuado para los datos de ventas de Lydia Pinkham.

## Generación de pronósticos con el modelo

El paso final en el análisis del modelo AR(2) para estos datos es pronosticar los valores de 1949 a 1960 con un periodo de anticipación. (Por ejemplo, los datos de 1958 se usaron para calcular el pronóstico de 1959). La ecuación del pronóstico es

$$\hat{Y}_t = 178.6 + 1.423Y_{t-1} - .521Y_{t-2}$$

Los pronósticos con un paso de anticipación y los errores de pronóstico se muestran en la tabla 9.20.

Además de los pronósticos con un paso de anticipación, se calcularon algunas medidas de precisión. Los pronósticos para el modelo AR(2) tienen un error porcentual absoluto medio (MAPE) de 6.9% y una desviación absoluta media (MAD) de 113 mil dólares actuales. Esta cifra se compara, de manera favorable, con las medidas de precisión de los modelos informales desarrollados por otros investigadores.

## Resumen y conclusiones

Se ajustó un modelo utilizando el principio de parsimonia (con un número menor de parámetros) AR(2) a los datos de Lydia Pinkham para los años de 1907 a 1948. Este modelo ha producido pronósticos de un paso anticipado muy precisos para los años de 1949 a 1960. ■

<sup>17</sup> Vea Kyle (1978)

**TABLA 9.18 Datos de Lydia E. Pinkham Medicine**

| <i>Año</i> | <i>Ventas<br/>(miles de dólares<br/>actuales)</i> | <i>Año</i> | <i>Ventas<br/>(miles de dólares<br/>actuales)</i> | <i>Año</i> | <i>Ventas<br/>(miles de dólares<br/>actuales)</i> |
|------------|---------------------------------------------------|------------|---------------------------------------------------|------------|---------------------------------------------------|
| 1907       | 1016                                              | 1925       | 3438                                              | 1943       | 2602                                              |
| 1908       | 921                                               | 1926       | 2917                                              | 1944       | 2518                                              |
| 1909       | 934                                               | 1927       | 2359                                              | 1945       | 2637                                              |
| 1910       | 976                                               | 1928       | 2240                                              | 1946       | 2177                                              |
| 1911       | 930                                               | 1929       | 2196                                              | 1947       | 1920                                              |
| 1912       | 1052                                              | 1930       | 2111                                              | 1948       | 1910                                              |
| 1913       | 1184                                              | 1931       | 1806                                              | 1949       | 1984                                              |
| 1914       | 1089                                              | 1932       | 1644                                              | 1950       | 1787                                              |
| 1915       | 1087                                              | 1933       | 1814                                              | 1951       | 1689                                              |
| 1916       | 1154                                              | 1934       | 1770                                              | 1952       | 1866                                              |
| 1917       | 1330                                              | 1935       | 1518                                              | 1953       | 1896                                              |
| 1918       | 1980                                              | 1936       | 1103                                              | 1954       | 1684                                              |
| 1919       | 2223                                              | 1937       | 1266                                              | 1955       | 1633                                              |
| 1920       | 2203                                              | 1938       | 1473                                              | 1956       | 1657                                              |
| 1921       | 2514                                              | 1939       | 1423                                              | 1957       | 1569                                              |
| 1922       | 2726                                              | 1940       | 1767                                              | 1958       | 1390                                              |
| 1923       | 3185                                              | 1941       | 2161                                              | 1959       | 1397                                              |
| 1924       | 3352                                              | 1942       | 2336                                              | 1960       | 1289                                              |

**TABLA 9.19**

| <i>Retraso k</i> | <i>ACF</i> | <i>PACF</i> |
|------------------|------------|-------------|
| 1                | 0.915      | 0.915       |
| 2                | 0.765      | -0.440      |
| 3                | 0.597      | -0.061      |
| 4                | 0.424      | -0.124      |
| 5                | 0.243      | -0.168      |
| 6                | 0.059      | -0.141      |
| 7                | -0.099     | 0.024       |
| 8                | -0.240     | -0.170      |
| 9                | -0.356     | -0.036      |
| 10               | -0.440     | -0.040      |

**PREGUNTAS**

- Después de haber completado este análisis, la cifra para los ventas en 1961 estuvo disponible: 1426 dólares. ¿Cuál es el pronóstico del modelo para 1961? Si este año fuera añadido a los datos de prueba, ¿cuáles serían los nuevos *MAPE* y *MAD*?
- Existe alguna evidencia de que los datos de Lydia Pinkham serían no estacionarios. Por ejemplo, las autocorrelaciones de la muestra tienden a ser grandes (de manera persistente) en diversos retrasos. Diferencie los datos.
- Construya la gráfica de la serie de tiempo para las diferencias. Por medio de Minitab u otro programa de cómputo ajuste un modelo ARIMA(1, 1, 0) a las ventas anuales de 1907 a 1948. Genere pronósticos de un paso anticipado para los años de 1949 a 1960. ¿Cuál modelo, AR(2) o ARIMA(1, 1, 0), cree usted que es mejor para los datos de Lydia Pinkham?
- Los datos de Lydia Pinkham son interesantes debido a la naturaleza única (invariable) del producto y el marketing para el periodo de 54 años repre-

TABLA 9.20

| <i>Año</i> | <i>Real</i> | <i>Pronóstico</i> | <i>Error</i> |
|------------|-------------|-------------------|--------------|
| 1949       | 1984        | 1896.4            | 87.6         |
| 1950       | 1787        | 2006.7            | -219.7       |
| 1951       | 1689        | 1687.8            | 1.2          |
| 1952       | 1866        | 1651.0            | 215.0        |
| 1953       | 1896        | 1953.9            | -57.9        |
| 1954       | 1684        | 1904.4            | -220.4       |
| 1955       | 1633        | 1587.1            | 45.9         |
| 1956       | 1657        | 1625.0            | 32.0         |
| 1957       | 1569        | 1685.7            | -116.7       |
| 1958       | 1390        | 1548.0            | -158.0       |
| 1959       | 1387        | 1339.1            | -47.9        |
| 1960       | 1289        | 1442.3            | -153.3       |

sentado. ¿Qué factores podrían afectar las ventas anuales para los automóviles y para el cobre durante este mismo periodo? ¿Por qué?

Para mayor información acerca de Lydia Pinkham Medicine Company y Lydia Pinkham vea

Sarah Stage, *Female Complaints: Lydia Pinkham and the Business of Women's Medicine*, Nueva York, Norton, 1979.

## CASO 9-5 LA CIUDAD DE COLLEGE STATION

La ciudad de College Station confía en los ingresos fiscales de las ventas mensuales para financiar sus funciones. Los ingresos fiscales por ventas generan cuentas por aproximadamente 44% del presupuesto general. Es importante poder pronosticar la cantidad de ingresos por impuestos sobre las ventas que la ciudad recibirá cada mes del estado de Texas, de manera que la cantidad de fondos iguale a los gastos anticipados y se deposite en una cuenta de cheques. El exceso de fondos puede colocarse en una cuenta que genere intereses.

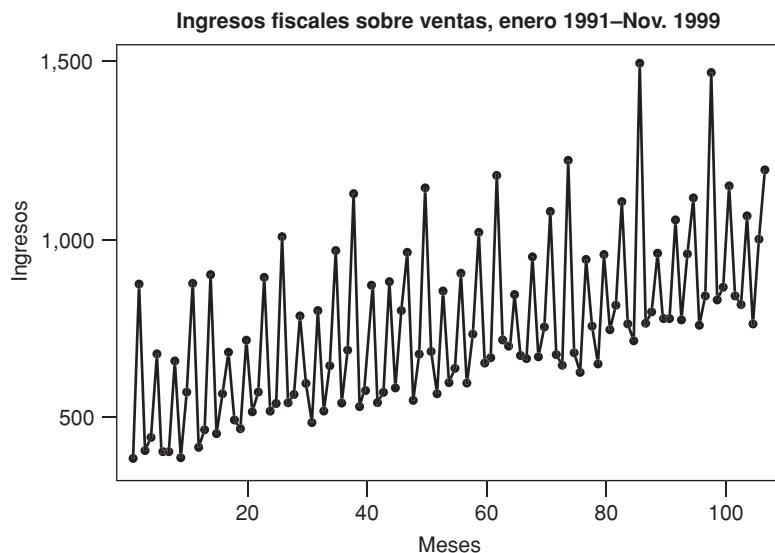
Charles Lemore, un analista de la ciudad, fue asignado para desarrollar un modelo de pronósticos para los ingresos fiscales sobre las ventas. Recientemente, Charles terminó un curso de pronósticos en la universidad local y decide utilizar la metodología Box-Jenkins. Comienza por observar los datos y las autocorrelaciones de la muestra, así como las funciones de la autocorrelación parcial de la muestra.

Los ingresos fiscales mensuales sobre ventas (en miles de dólares), para el periodo de enero de 1991 a noviembre de 1999, están enlistados en la tabla 9.21. En la figura 9.41 aparece una gráfica de la serie de tiempo de los ingresos.

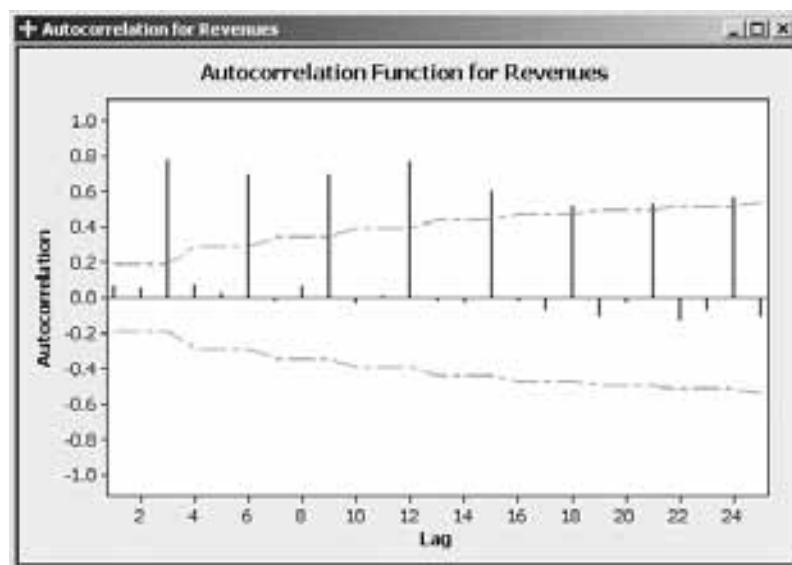
Un componente estacional fuerte es obvio en una gráfica de las series, así como en las autocorrelaciones y las autocorrelaciones parciales de la muestra que se ilustran en las figuras 9.42 y 9.43, respectivamente. Las autocorrelaciones en los múltiplos del retraso 3 sugieren que hay un patrón estacional trimestral dentro del año (el segundo pago mensual de la ciudad dentro de cualquier trimestre es relativamente grande), así como un patrón anual. Además, las series contienen una clara tendencia ascendente.

Charles decide diferenciar la serie de ingresos con respecto al periodo estacional  $S = 12$ . Las autocorrelaciones de la muestra para la serie diferenciada estacionalmente se presentan en la figura 9.44. Aparentemente, una gráfica de la serie diferenciada estacionalmente (que no se presenta) varía en torno a un nivel constante de casi 50.

Charles está listo para ajustar y verificar un modelo ARIMA para la serie de ingresos fiscales sobre ventas, pero necesita de su ayuda. Una vez que cuente con el modelo adecuado, también le gustaría generar pronósticos para los siguientes 12 meses. ■



**FIGURA 9.41** Ingresos fiscales sobre ventas de la ciudad de College Station



**Autocorrelation Function: Revenues**

| Lag | ACF        | T     | LBC    | Lag | ACF       | T     | LBC    |
|-----|------------|-------|--------|-----|-----------|-------|--------|
| 1   | 0.065295   | 0.60  | 0.47   | 14  | -0.028273 | -0.13 | 355.28 |
| 2   | 0.052000   | 0.54  | 0.77   | 15  | 0.398940  | 2.68  | 300.73 |
| 3   | 0.770231   | 9.99  | 68.69  | 16  | -0.034427 | -0.10 | 300.85 |
| 4   | 0.069311   | 0.40  | 69.23  | 17  | -0.073904 | -0.31 | 301.56 |
| 5   | 0.027452   | 0.19  | 69.32  | 18  | 0.515622  | 2.16  | 336.46 |
| 6   | 0.693138   | 4.79  | 124.80 | 19  | -0.109377 | -0.44 | 337.98 |
| 7   | -0.022330  | -0.13 | 124.86 | 20  | -0.034739 | -0.14 | 338.14 |
| 8   | 0.066637   | 0.39  | 125.38 | 21  | 0.527036  | 2.12  | 375.88 |
| 9   | 0.694694   | 4.01  | 182.81 | 22  | -0.126835 | -0.49 | 378.02 |
| 10  | -0.042753  | -0.22 | 183.03 | 23  | -0.075273 | -0.29 | 378.81 |
| 11  | 0.008252   | 0.04  | 183.04 | 24  | 0.571783  | 2.20  | 424.75 |
| 12  | 0.766130   | 3.88  | 255.10 | 25  | -0.111139 | -0.41 | 426.53 |
| 13  | -0.0229266 | -0.12 | 255.19 |     |           |       |        |

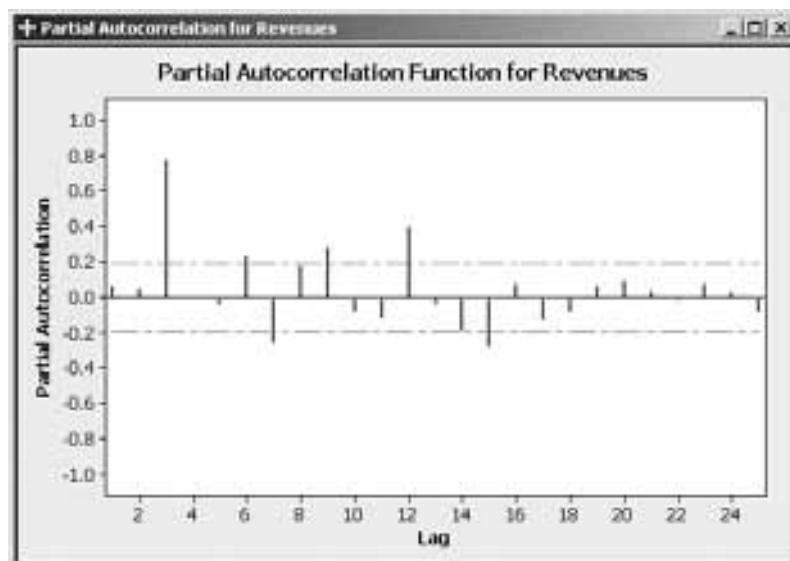
**FIGURA 9.42** Autocorrelaciones de la muestra para los ingresos fiscales sobre ventas

**PREGUNTAS**

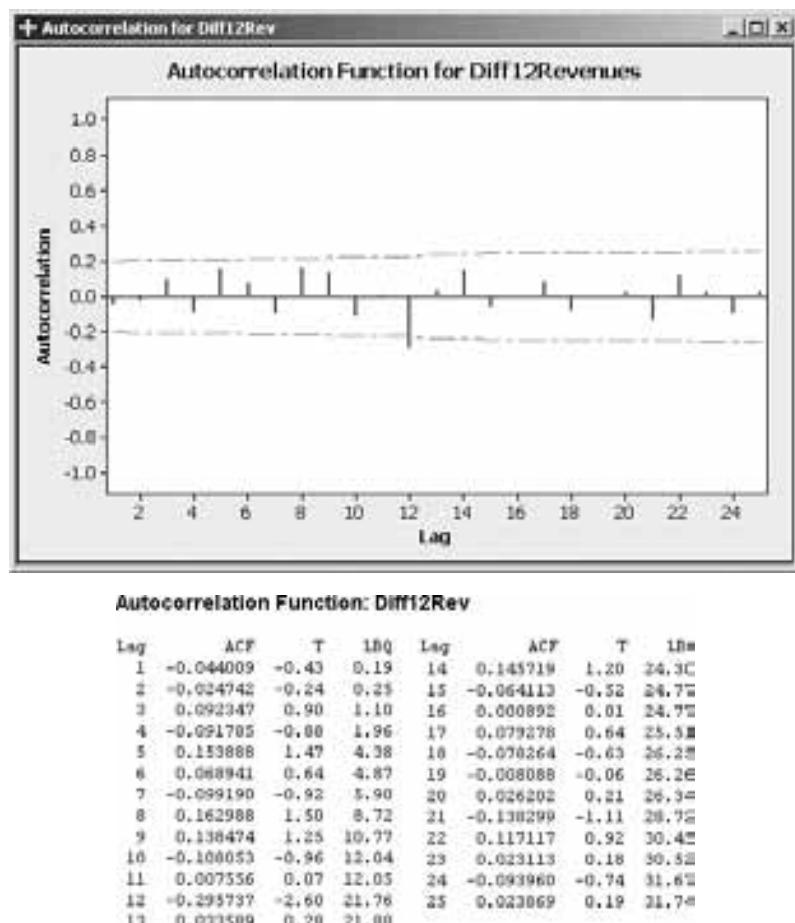
- Desarrolle un modelo ARIMA para los ingresos fiscales sobre ventas mediante la metodología Box-Jenkins
- Por medio de su modelo, genere pronósticos para los ingresos de los siguientes 12 meses. Agréguelos al final de la serie y haga una gráfica con los resultados. ¿Está usted satisfecho con el patrón de los pronósticos?
- Resuma sus resultados en un pequeño reporte.

**TABLA 9.21**

| <i>Año</i> | <i>Ene.</i> | <i>Feb.</i> | <i>Mar.</i> | <i>Abr.</i> | <i>May.</i> | <i>Jun.</i> | <i>Jul.</i> | <i>Ago.</i> | <i>Sept.</i> | <i>Oct.</i> | <i>Nov.</i> | <i>Dic.</i> |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|
| 1991       | 382         | 872         | 408         | 443         | 679         | 406         | 409         | 662         | 388          | 570         | 879         | 417         |
| 1992       | 468         | 895         | 457         | 568         | 685         | 492         | 465         | 713         | 518          | 572         | 891         | 519         |
| 1993       | 544         | 1005        | 543         | 568         | 785         | 596         | 488         | 799         | 518          | 644         | 968         | 544         |
| 1994       | 691         | 1125        | 531         | 576         | 868         | 541         | 571         | 876         | 579          | 801         | 960         | 546         |
| 1995       | 675         | 1139        | 684         | 564         | 851         | 595         | 641         | 902         | 596          | 735         | 1014        | 652         |
| 1996       | 663         | 1174        | 714         | 698         | 842         | 675         | 665         | 948         | 670          | 753         | 1073        | 676         |
| 1997       | 646         | 1219        | 680         | 628         | 943         | 756         | 649         | 951         | 742          | 814         | 1100        | 759         |
| 1998       | 715         | 1487        | 763         | 793         | 955         | 780         | 777         | 1052        | 775          | 957         | 1117        | 757         |
| 1999       | 839         | 1459        | 828         | 862         | 1147        | 840         | 814         | 1064        | 766          | 997         | 1187        |             |

**FIGURA 9.43 Autocorrelaciones parciales de la muestra para los ingresos fiscales sobre ventas****Partial Autocorrelation Function: Revenues**

| Lag | PACF      | T     | Lag | PACF      | T     |
|-----|-----------|-------|-----|-----------|-------|
| 1   | 0.065295  | 0.68  | 14  | -0.109410 | -1.90 |
| 2   | 0.047941  | 0.50  | 15  | -0.285778 | -2.92 |
| 3   | 0.776959  | 0.04  | 16  | 0.072095  | 0.79  |
| 4   | -0.002013 | -0.02 | 17  | -0.130604 | -1.32 |
| 5   | -0.036516 | -0.38 | 18  | -0.082646 | -0.95 |
| 6   | 0.224379  | 2.32  | 19  | 0.065410  | 0.60  |
| 7   | -0.256006 | -2.65 | 20  | 0.091950  | 0.92  |
| 8   | 0.177547  | 1.04  | 21  | 0.028949  | 0.30  |
| 9   | 0.278791  | 2.88  | 22  | -0.017353 | -0.16 |
| 10  | -0.078868 | -0.81 | 23  | 0.072479  | 0.72  |
| 11  | -0.123700 | -1.20 | 24  | 0.022790  | 0.24  |
| 12  | 0.396276  | 4.10  | 25  | -0.079367 | -0.68 |
| 13  | -0.042746 | -0.44 |     |           |       |



**FIGURA 9.44 Autocorrelaciones de la muestra para los ingresos fiscales sobre ventas diferenciadas estacionalmente**

## CASO 9-6 DIVISION DE FINANZAS DE UPS AIR

La División de Finanzas Aéreas es una de las áreas dentro de United Parcel Service Financial Corporation, una subsidiaria de United Parcel Service (UPS), que es su propietario total. La División de Finanzas Aéreas tiene la responsabilidad de proporcionar servicios financieros a la compañía en cuanto a todas las adquisiciones de las aeronaves. Además, esta división también proporciona financiamiento a entidades independientes externas para la compra de aeronaves, en una operación separada y distinta. Históricamente, el segmento financiero que no pertenece a UPS en la División de Finanzas Aéreas ha proporcionado gran rendimiento y

mayores ingresos que el segmento de UPS; sin embargo, es más costoso llevarlo a cabo.

Los pronósticos de financiamiento para el segmento de mercado ajeno a UPS de la División de Finanzas Aéreas, ha sido muy subjetivo y no es confiable como los pronósticos para otros segmentos. La tabla 9.22 enumera 10 años de requerimientos de inversión mensual para el segmento ajeno a UPS de la División de Finanzas Aéreas. Estos datos se obtuvieron de los reportes administrativos del periodo de enero de 1989 a diciembre de 1998, y son las cifras de fin de mes que aparecen en millones de dólares.

Los datos en la tabla 9.22 están graficados en la figura 9.45. La serie de tiempo de los requerimientos de financiamiento tiene un componente estacional sólido junto con una tendencia general ascendente. Las autocorrelaciones y las autocorrelaciones parciales de la muestra se presentan en las figuras 9.46 y 9.47, respectivamente.

Se usará la metodología Box-Jenkins para desarrollar un modelo ARIMA, a fin de pronosticar los requerimientos futuros de financiamiento del segmento ajeno a UPS de la División de Finanzas Aéreas.

Debido al carácter no estacionario de la serie tiempo, se examinan las autocorrelaciones y las autocorrela-

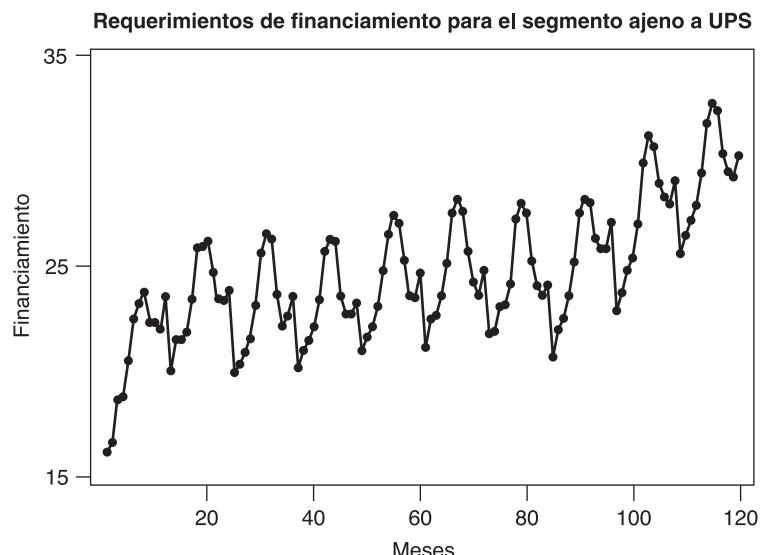
ciones parciales para las variadas diferencias de la serie. La función de autocorrelación para la serie con una diferencia regular y una diferencia estacional del orden  $S = 12$  se muestran la figura 9.48. Es decir, estas autocorrelaciones se calculan para las diferencias

$$\begin{aligned} W_t &= \Delta\Delta_{12}Y_t = \Delta(Y_t - Y_{t-12}) \\ &= Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} \end{aligned}$$

TABLA 9.22

| Año  | Ene. | Feb. | Mar. | Abr. | May. | Jun. | Jul. | Ago. | Sept. | Oct. | Nov. | Dic. |
|------|------|------|------|------|------|------|------|------|-------|------|------|------|
| 1989 | 16.2 | 16.7 | 18.7 | 18.8 | 20.6 | 22.5 | 23.3 | 23.8 | 22.3  | 22.3 | 22.1 | 23.6 |
| 1990 | 20.1 | 21.6 | 21.6 | 21.9 | 23.4 | 25.9 | 26.0 | 26.2 | 24.7  | 23.5 | 23.4 | 23.9 |
| 1991 | 20.0 | 20.4 | 20.9 | 21.6 | 23.2 | 25.6 | 26.6 | 26.3 | 23.7  | 22.2 | 22.7 | 23.6 |
| 1992 | 20.2 | 21.1 | 21.5 | 22.2 | 23.4 | 25.7 | 26.3 | 26.2 | 23.6  | 22.8 | 22.8 | 23.3 |
| 1993 | 21.0 | 21.7 | 22.2 | 23.1 | 24.8 | 26.6 | 27.4 | 27.1 | 25.3  | 23.6 | 23.5 | 24.7 |
| 1994 | 21.2 | 22.5 | 22.7 | 23.6 | 25.1 | 27.6 | 28.2 | 27.7 | 25.7  | 24.3 | 23.7 | 24.9 |
| 1995 | 21.8 | 21.9 | 23.1 | 23.2 | 24.2 | 27.2 | 28.0 | 27.6 | 25.2  | 24.1 | 23.6 | 24.1 |
| 1996 | 20.7 | 22.0 | 22.5 | 23.6 | 25.2 | 27.6 | 28.2 | 28.0 | 26.3  | 25.9 | 25.9 | 27.1 |
| 1997 | 22.9 | 23.8 | 24.8 | 25.4 | 27.0 | 29.9 | 31.2 | 30.7 | 28.9  | 28.3 | 28.0 | 29.1 |
| 1998 | 25.6 | 26.5 | 27.2 | 27.9 | 29.4 | 31.8 | 32.7 | 32.4 | 30.4  | 29.5 | 29.3 | 30.3 |

FIGURA 9.45 Requerimientos de financiamiento para el segmento ajeno a UPS de la División de Finanzas Aéreas



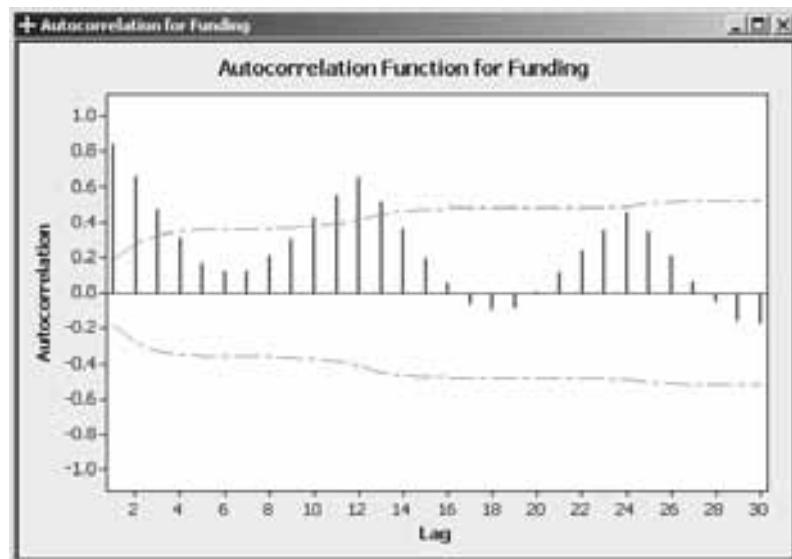
Existe claramente un pico significativo en el retraso 12 en las autocorrelaciones de la muestra que se presentan en la figura 9.48. Esto sugiere un modelo ARIMA para el  $W_t$ 's (el proceso con una diferencia regular y una diferencia estacional) que tiene un término de promedio

móvil en el retraso estacional 12. En consecuencia, un modelo ARIMA(1, 0, 1)(0, 1, 1)<sub>12</sub> podría ser una buena elección inicial para los datos de los requerimientos de financiamiento. ■

## PREGUNTAS

1. Por medio del uso de Minitab o de un software equivalente, ajuste un modelo ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub> a los datos que se presentan en la tabla 9.22. ¿Usted cree que se requiere un término constante en el modelo? Explique su respuesta.
2. ¿El modelo sugerido en la parte 1 es adecuado? Analícelo con referencia a las gráficas residuales, las autocorrelaciones residuales y la estadística chi cuadrada de Ljung-Box. Si el modelo es inadecuado,
- modifique y ajuste el modelo inicial hasta que considere que ha obtenido un modelo satisfactorio.
3. Mediante el modelo que ha desarrollado en la parte 2 genere los pronósticos de los requerimientos de financiamiento para los siguientes 12 meses.
4. Escriba un informe en donde resuma sus descubrimientos. Agregue una gráfica de la serie original y de los pronósticos.

**FIGURA 9.46 Autocorrelaciones de la muestra para los requerimientos de financiamiento**



**Autocorrelation Function: Funding**

| Lag | ACF      | T    | IBC    | Lag | ACF       | T     | IBC    |
|-----|----------|------|--------|-----|-----------|-------|--------|
| 1   | 0.837546 | 9.17 | 86.30  | 16  | 0.051635  | 0.21  | 385.15 |
| 2   | 0.653627 | 4.62 | 139.31 | 17  | -0.070460 | -0.29 | 385.90 |
| 3   | 0.471450 | 2.86 | 167.12 | 18  | -0.101867 | -0.42 | 387.30 |
| 4   | 0.309926 | 1.76 | 179.24 | 19  | -0.090741 | -0.38 | 388.50 |
| 5   | 0.167326 | 0.93 | 182.80 | 20  | 0.006750  | 0.03  | 388.55 |
| 6   | 0.119600 | 0.66 | 184.64 | 21  | 0.114121  | 0.47  | 390.53 |
| 7   | 0.110763 | 0.65 | 186.47 | 22  | 0.233914  | 0.97  | 390.66 |
| 8   | 0.208453 | 1.14 | 192.15 | 23  | 0.353911  | 1.45  | 417.55 |
| 9   | 0.308387 | 1.87 | 204.69 | 24  | 0.457769  | 1.85  | 449.55 |
| 10  | 0.428652 | 2.27 | 229.15 | 25  | 0.340459  | 1.34  | 467.41 |
| 11  | 0.546265 | 2.77 | 269.23 | 26  | 0.102990  | 0.78  | 473.82 |
| 12  | 0.649303 | 3.10 | 326.37 | 27  | 0.061054  | 0.23  | 474.41 |
| 13  | 0.511713 | 2.27 | 362.20 | 28  | -0.016059 | -0.22 | 474.93 |
| 14  | 0.355561 | 1.51 | 379.66 | 29  | -0.156711 | -0.60 | 478.00 |
| 15  | 0.192224 | 0.80 | 384.81 | 30  | -0.175532 | -0.67 | 483.80 |

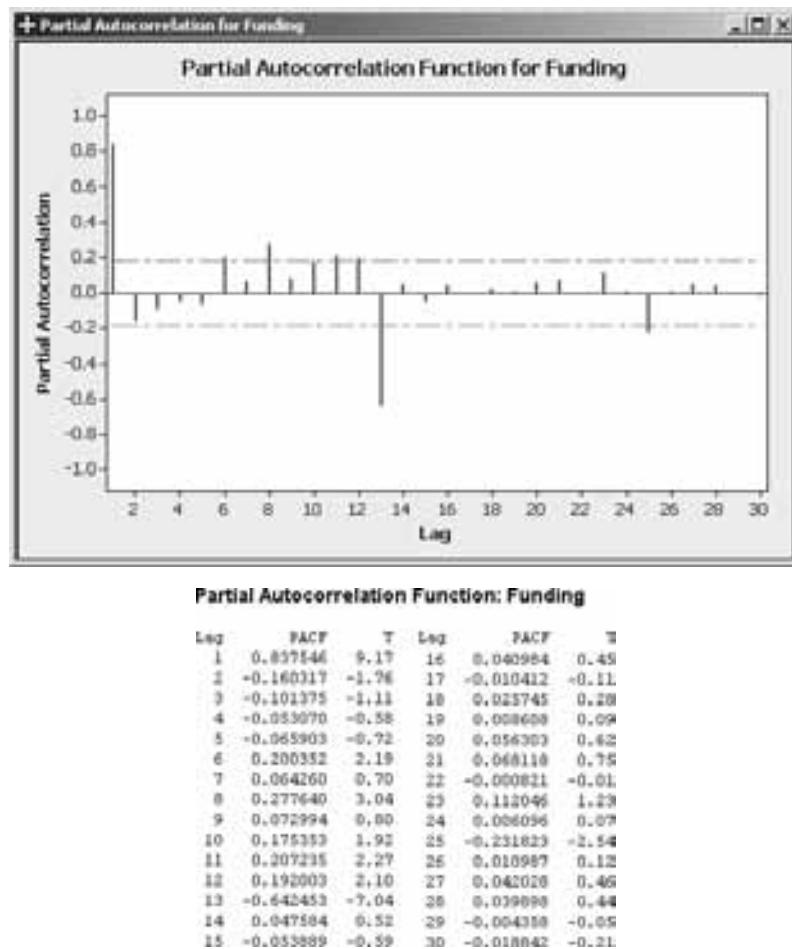


FIGURA 9.47 Autocorrelaciones parciales de la muestra para los requerimientos de financiamiento

## CASO 9-7 AAA WASHINGTON

En el caso 5.5 se proporcionó una apreciación global de AAA Washington, cuando se pidió a los estudiantes que preparasen una descomposición de la serie tiempo de las llamadas de emergencia en carretera que recibió el club durante cinco años. La descomposición de la serie de tiempo que se efectuó en el caso 5.5 mostró que el patrón que Michael DeCoria había observado en el volumen de llamadas del servicio de emergencia en carretera era, probablemente, de naturaleza cíclica. Michael quisiera ser capaz de predecir el efecto cíclico del volumen de llamadas del servicio de emergencia carretera para los próximos años.

En el caso 6.6 se investigaron cuatro modelos de regresión lineal que utilizaban el número total de llamadas

del servicio de emergencia en carretera como la variable dependiente; y las tasas de desempleo, la temperatura, el volumen de lluvia y el número de afiliados al club como las cuatro variables independientes. La variable de temperatura se transformó al sustraer 65 grados de los valores de la temperatura mensual promedio. Entonces, se investigó una relación no lineal.

En el caso 7.2 se desarrolló un modelo de regresión múltiple. Se evaluaron las variables, como la lluvia, el número de afiliados, la temperatura mensual promedio transformada exponencialmente y la tasa de desempleo con retraso de 11 meses.

En el caso 8.6 se revisó el modelo de regresión múltiple desarrollado en el caso 7.2 para una correlación

serial y se creyó que un modelo recomendado a Michael DeCoria era el más apropiado para predecir la naturaleza cíclica del volumen de llamadas del servicio de emergencia carretera.

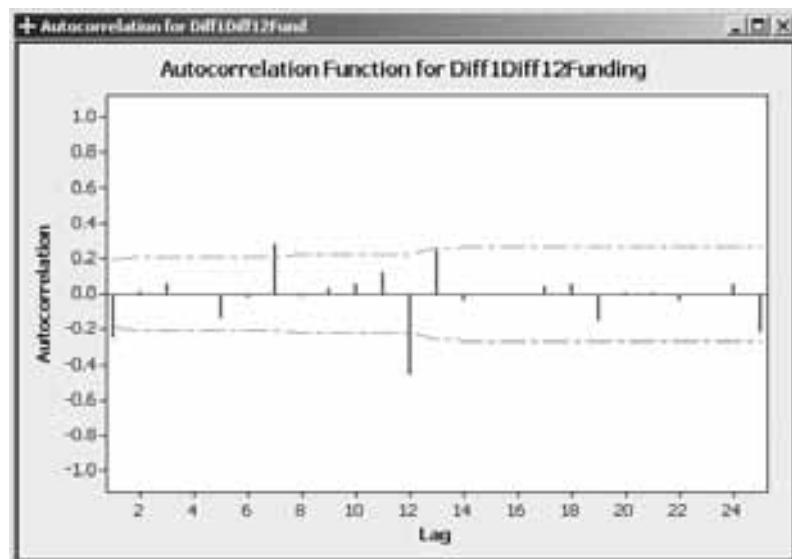
Michael no está satisfecho todavía. El modelo presentado en el caso 8.6 sólo explicaba aproximadamente 71% de la varianza en la variable del volumen de llamadas del servicio en carretera. Michael le pidió a usted que usara la metodología Box-Jenkins para pronosticar los datos del volumen de llamadas de emergencia que se presentan en la tabla 9.23. Usted

decide dividir la serie de datos en dos secciones. Los valores de mayo de 1988 a diciembre de 1992 serán la sección de inicio, o de ajuste, y los primeros cuatro meses de 1993 serán la sección de prueba, o de pronóstico.

### Tarea

Desarrolle un modelo ARIMA para el volumen de los datos de llamadas del servicio de emergencia en carretera. Escriba un memorándum para Michael DeCoria en donde resuma sus descubrimientos. ■

**FIGURA 9.48 Autocorrelaciones de la muestra para los requerimientos de financiamiento diferenciados regular y estacionalmente**



**Autocorrelation Function: Diff1Diff12Fund**

| Lag | ACF       | T     | LBC   | Lag | ACF       | T     | LBC   |
|-----|-----------|-------|-------|-----|-----------|-------|-------|
| 1   | -0.249318 | -2.58 | 6.84  | 14  | -0.040666 | -0.31 | 53.68 |
| 2   | 0.015515  | 0.15  | 6.87  | 15  | -0.000993 | -0.07 | 53.69 |
| 3   | 0.049961  | 0.49  | 7.15  | 16  | -0.000199 | -0.00 | 53.69 |
| 4   | -0.004556 | -0.04 | 7.15  | 17  | 0.035883  | 0.27  | 53.66 |
| 5   | -0.138456 | -1.35 | 9.34  | 18  | 0.052058  | 0.39  | 54.22 |
| 6   | -0.026264 | -0.25 | 9.42  | 19  | -0.152044 | -1.14 | 57.28 |
| 7   | 0.279279  | 2.67  | 18.52 | 20  | 0.004204  | 0.03  | 57.28 |
| 8   | -0.017412 | -0.16 | 18.55 | 21  | 0.009024  | 0.07  | 57.28 |
| 9   | 0.027132  | 0.24  | 18.64 | 22  | -0.040380 | -0.30 | 57.52 |
| 10  | 0.052554  | 0.47  | 19.97 | 23  | -0.005723 | -0.04 | 57.52 |
| 11  | 0.121755  | 1.09  | 20.77 | 24  | 0.052168  | 0.39  | 57.95 |
| 12  | -0.433773 | -4.02 | 46.05 | 25  | -0.221294 | -1.63 | 64.03 |
| 13  | 0.264586  | 1.90  | 53.47 |     |           |       |       |

**TABLA 9.23 Volumen de llamadas del servicio de emergencia carretera para el caso 9.7**

| <i>Año</i> | <i>Mes</i> | <i>Llamadas</i> | <i>Año</i> | <i>Mes</i> | <i>Llamadas</i> |
|------------|------------|-----------------|------------|------------|-----------------|
| 1988       | Mayo       | 20,002          | 1991       | Noviembre  | 22,490          |
|            | Junio      | 21,591          |            | Diciembre  | 24,861          |
|            | Julio      | 22,696          |            | Enero      | 23,441          |
|            | Agosto     | 21,509          |            | Febrero    | 19,205          |
|            | Septiembre | 22,123          |            | Marzo      | 20,386          |
|            | Octubre    | 21,449          |            | Abril      | 19,988          |
|            | Noviembre  | 23,475          |            | Mayo       | 19,077          |
|            | Diciembre  | 23,529          |            | Junio      | 19,141          |
|            | Enero      | 23,327          |            | Julio      | 20,883          |
|            | Febrero    | 24,050          |            | Agosto     | 20,709          |
|            | Marzo      | 24,010          |            | Septiembre | 19,647          |
|            | Abril      | 19,735          |            | Octubre    | 22,013          |
| 1989       | Mayo       | 20,153          | 1992       | Noviembre  | 22,375          |
|            | Junio      | 19,512          |            | Diciembre  | 22,727          |
|            | Julio      | 19,892          |            | Enero      | 22,367          |
|            | Agosto     | 20,326          |            | Febrero    | 21,155          |
|            | Septiembre | 19,378          |            | Marzo      | 21,209          |
|            | Octubre    | 21,263          |            | Abril      | 19,286          |
|            | Noviembre  | 21,443          |            | Mayo       | 19,725          |
|            | Diciembre  | 23,366          |            | Junio      | 20,276          |
|            | Enero      | 23,836          |            | Julio      | 20,795          |
|            | Febrero    | 23,336          |            | Agosto     | 21,126          |
|            | Marzo      | 22,003          |            | Septiembre | 20,251          |
|            | Abril      | 20,155          |            | Octubre    | 22,069          |
| 1990       | Mayo       | 20,070          | 1993       | Noviembre  | 23,268          |
|            | Junio      | 19,588          |            | Diciembre  | 26,039          |
|            | Julio      | 20,804          |            | Enero      | 26,127          |
|            | Agosto     | 19,644          |            | Febrero    | 20,067          |
|            | Septiembre | 17,424          |            | Marzo      | 19,673          |
|            | Octubre    | 20,833          |            | Abril      | 19,142          |

## Aplicaciones en Minitab

**El problema.** En el ejemplo 9.10, Perkin Kendell, analista de Outboard Marine Corporation, desea pronosticar las ventas para 1997.

### Solución en Minitab

- Si los datos están almacenados en un archivo, ábralo con los siguientes menús:

File>Open Worksheet

Si los datos no están en un archivo, intodúzcalos en una hoja de cálculo.

- Para calcular las autocorrelaciones para la variable ventas, dé clic en:

Stat>Time Series>Autocorrelation

- Aparece el cuadro de diálogo de la Función de Autocorrelación que se presenta en la figura 3.28.
  - Dé doble clic en la variable *Sales* (ventas) y ésta aparecerá a la derecha de *Series*.
  - Dé clic en OK y aparecerá la figura 9.33.

4. Para poder diferenciar estacionalmente los datos dé clic en los siguientes menús:  
 Stat>Time Series>Differences
5. Aparece el cuadro de diálogo de *Differences* (diferencias) que se muestra en la figura 3.29.
  - a) Dé doble clic en *Variable Sales* (ventas variables) y ésta aparecerá a la derecha de *Series*.
  - b) Dé clic en el tabulador para almacenar las diferencias de introduzca C2.
  - c) Dé clic en *Lag* (retraso) e introduzca un 4. Dé clic en OK y las cuatro diferencias aparecerán en la columna 2 al principio de la línea 5.
6. Rotule la variable C2 Diff4Sales. Para calcular las autocorrelaciones para esta variable, repita el paso 2 utilizando Diff4Sales como la variable a la derecha de *Series*.
7. Para calcular las primeras diferencias de la variable Diff4Sales repita el paso 5 y almacene las diferencias en C3 utilizando un retraso de 1.
8. Rotule la variable C3 Diff1Diff4Sales. Para calcular las autocorrelaciones para esta variable, repita el paso 2 utilizando Diff1Diff4Sales a la derecha de *Series*.
9. Para calcular las autocorrelaciones parciales para la variable Diff1Diff4Sales dé clic en:  
 Stat>Time Series>Partial Autocorrelation
10. Aparece el cuadro de diálogo de Función de Autocorrelación Parcial, semejante a la figura 3.28.
  - a) Dé doble clic en la variable Diff1Diff4Sales y ésta aparecerá a la derecha de *Series*.
  - b) Dé clic en OK y aparecerá la figura 9.36.
11. Para poder correr el modelo ARIMA(1, 1, 0)(0, 1, 2) de clic en los siguientes menús:  
 Stat>Time Series>ARIMA
12. Aparece el cuadro de diálogo de ARIMA que se muestra la figura 9.49.
  - a) Dé doble clic en la variable *Sales* (ventas) y ésta aparecerá a la derecha de *Series*.
  - b) Dé clic en el cuadro a la izquierda de *Fit seasonal model* (ajustar el modelo estacional) e indique 4 a la derecha de *Period* (periodo).
  - c) Bajo *Nonseasonal* (no estacional) coloque 1 a la derecha de *Autoregressive* (autorregresivo), un 1 a la derecha de *Difference* (diferencia) y un 0 a la derecha de *Lead* (guiar). Dé clic en OK.
  - d) Bajo *Seasonal* (estacional) coloque 0 a la derecha de *Autoregressive* (autorregresivo), un 1 a la derecha de *Difference* (diferencia) y un 2 a la derecha de *Lead* (guiar). Dé clic en OK.
  - e) Debido que los datos han sido diferenciados, utilice el cuadro para incluir a término constante en el modelo.
  - f) Dé clic en *Forecast* (pronóstico) y aparecerá el cuadro de diálogo de ARIMA-Forecast. Para poder pronosticar dos períodos en el futuro coloque 2 a la derecha de *Lead*. Dé clic en OK.
  - g) Dé clic en *Storage* (almacenar) y aparecerá el cuadro de diálogo de ARIMA-Forecast. Dé clic en el cuadro que está junto a *Residuals* (residuales) y después dé clic en OK. Dé clic en OK en el cuadro de diálogo ARIMA y aparecerá la porción inferior de la tabla 9.16.
  - h) Para calcular las autocorrelaciones de los residuales, repita el paso 2 utilizando RES1 como variable a la derecha de *Series*.
13. Para poder desarrollar una gráfica de la serie de tiempo que incluya un pronóstico, dé clic en *Graphs* (gráficas) en el cuadro de diálogo ARIMA y aparecerá el cuadro de diálogo de ARIMA-Graphs que se muestra en la figura 9.50.
  - a) Dé clic en el cuadro que está junto a *Series Plot* (gráfica de la serie).
  - b) Un método alterno para obtener las autocorrelaciones para los residuales consiste en dar clic en el cuadro que está la izquierda de ACF de los residuales
  - c) Dé clic en OK. Dé clic en OK y aparecerá el cuadro de diálogo ARIMA que se muestra en la figura 9.38.

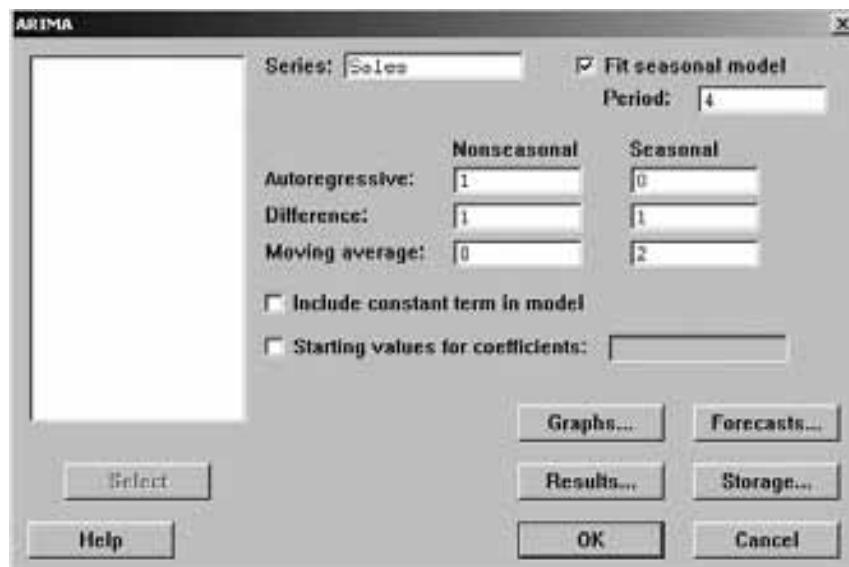


FIGURA 9.46 Cuadro de diálogo de Minitab ARIMA

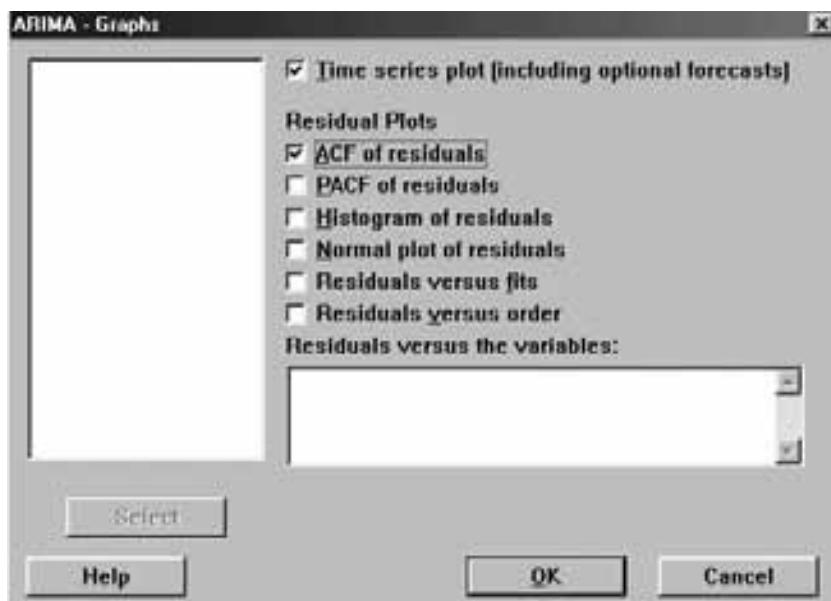


FIGURA 9.50 Cuadro de diálogo de Minitab ARIMA-Graphs

### Aplicaciones en Excel: CB predictor

**El problema.** En el capítulo 3 se utilizó Excel para calcular los coeficientes de autocorrelación y se presentó un correlograma para los datos en la tabla 3.1. (Vea la página 61.) En esta sección se demostrará CB Predictor utilizando los promedios diarios de cierre del Índice de Transportación Dow Jones que se muestra en la tabla 9.3.

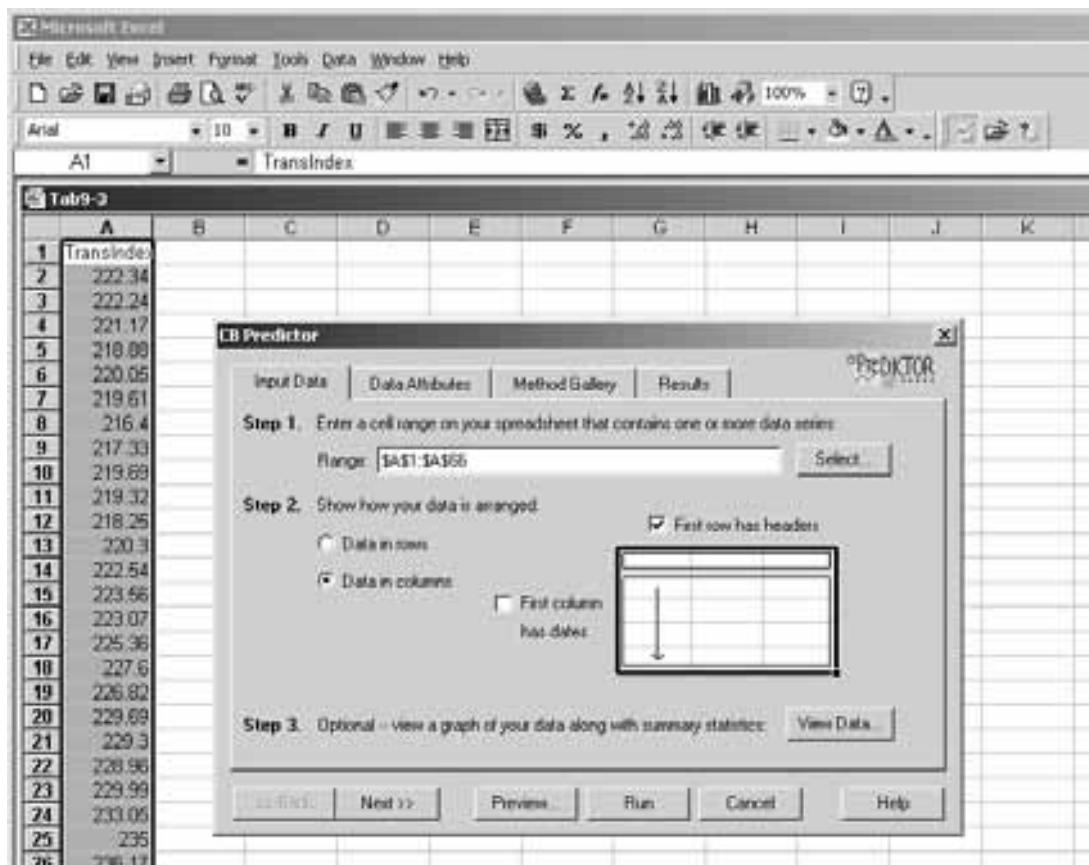


FIGURA 9.51 Cuadro de diálogo para la entrada de datos de CB Predictor

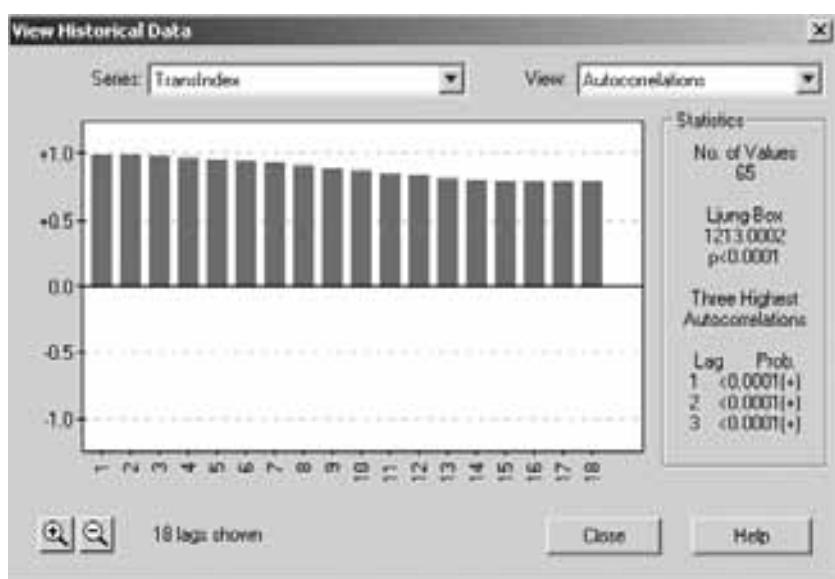


FIGURA 9.52 Pantalla de datos históricos de CB Predictor

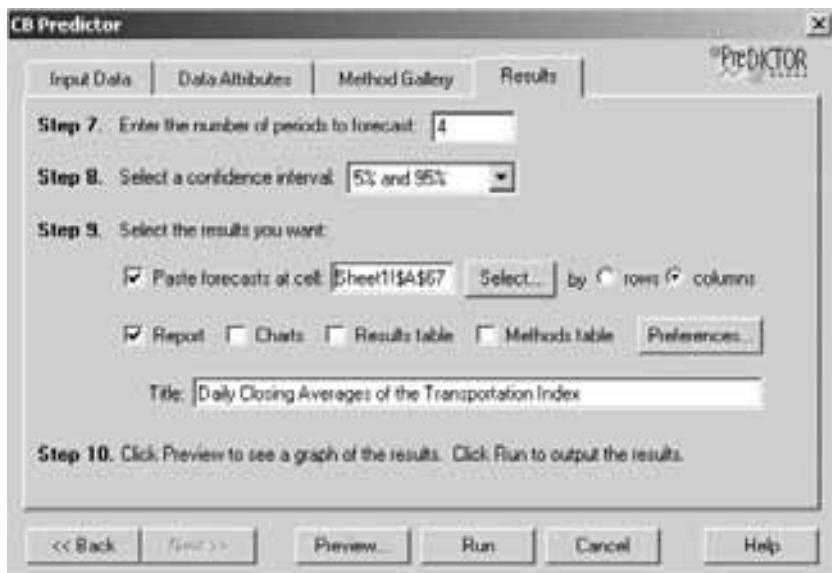


FIGURA 9.53 Pantalla de resultados de CB Predictor

**Solución en CB Predictor**

1. Instale CB Predictor del CD incluido en este texto. Es posible elegir cargar CB Predictor cada vez que abra Excel, de manera que los iconos de CB Predictor aparecerán en la barra de herramientas. Si no, dé clic en el cuadro correspondiente a Tools>Add-Ins  
Aparece el cuadro de diálogo *Add-Ins* (complementos), seleccione CB Predictor y dé clic en OK. Aparecerá CB Predictor en la barra herramientas.
2. Abra el archivo de datos en Excel al dar clic de los siguientes menús  
File>Open  
En el cuadro de diálogo *Open* (abrir) introduzca Tab9-3 a la derecha del nombre del archivo: dé clic en *Open* y aparecerán los datos de la tabla 9.3.
3. Coloque el cursor sobre *Indices* —al cual se ha llamado TransIndex en la figura 9.51— y dé clic en CB Predictor. El cuadro de diálogo de CB Predictor que aparece en la figura 9.52 aparecerá con los datos introducidos en el cuadro de *Range* (rango).
4. En el paso 3 dé clic en *View Data* (ver datos) y seleccione *Autocorrelations* (autocorrelaciones) en el cuadro *View* (ver) que aparece en la pantalla *View Data* (ver datos). Aparece la pantalla que se muestra en la figura 9.52.
5. Se imprimen los valores numéricos de las autocorrelaciones<sup>18</sup> y los valores *p* asociados con las estadísticas *t* para probar la existencia de una autocorrelación significativa en un retraso determinado como parte del resultado *Report* (informe) que ofrece CB Predictor. En consecuencia, es necesario completar el modelo y los pasos para las especificaciones del pronóstico de los pasos en CB Predictor (pasos 4.9). La pantalla de CB Predictor del paso 9 se muestra en la figura 9.53.
6. En el paso 9, seleccione *Report* y dé clic en *Preferences* (preferencias). Cuando la pantalla de preferencias aparezca, dé clic en el cuadro *Autocorrelations* y dé clic en OK.

<sup>18</sup> CB Predictor calcula el coeficiente de autocorrelación en un retraso determinado al usar la definición de Pearson acerca de la correlación, que se proporciona en la ecuación 2.10 con la serie retrasada  $Y_{t-k}$  que adopta la función de las observaciones  $X$  en este ecuación. En consecuencia, los valores numéricos de las autocorrelaciones de CB Predictor será diferente de las autocorrelaciones  $r_k$  proporcionadas por la ecuación 3.1. Sin embargo, las diferencias serán pequeñas para una  $n$  grande.

| Report for Daily Closing Averages of the Transportation Index                                                         |             |                      |
|-----------------------------------------------------------------------------------------------------------------------|-------------|----------------------|
| Created: 3/10/2003 at 1:09:22 PM                                                                                      |             |                      |
| <b>Summary:</b>                                                                                                       |             |                      |
| Number of series: 1<br>Periods to forecast: 4<br>Seasonality: none<br>Error Measure: RMSE                             |             |                      |
| <b>Series: TransIndex</b>                                                                                             |             | <b>Range: A2:A66</b> |
| Method: Double Moving Average<br>Parameters:<br>Periods: 3<br>Error: 2.3581                                           |             |                      |
| Series Statistics:<br>Mean: 244.176<br>Std. Dev.: 19.320<br>Minimum: 216.4<br>Maximum: 288.57<br>Ljung-Box: 1213.0002 |             |                      |
| Autocorrelations:                                                                                                     |             |                      |
| Lag                                                                                                                   | Correlation | Probability          |
| 1                                                                                                                     | 0.9953      | 0.0000               |
| 2                                                                                                                     | 0.9881      | 0.0000               |
| 3                                                                                                                     | 0.9804      | 0.0000               |
| 4                                                                                                                     | 0.9706      | 0.0000               |
| 5                                                                                                                     | 0.9573      | 0.0000               |
| 6                                                                                                                     | 0.9421      | 0.0000               |
| 7                                                                                                                     | 0.9266      | 0.0000               |
| 8                                                                                                                     | 0.9064      | 0.0000               |
| 9                                                                                                                     | 0.8857      | 0.0000               |
| 10                                                                                                                    | 0.8672      | 0.0000               |
| 11                                                                                                                    | 0.8490      | 0.0000               |
| 12                                                                                                                    | 0.8317      | 0.0000               |
| 13                                                                                                                    | 0.8124      | 0.0000               |
| 14                                                                                                                    | 0.7973      | 0.0000               |
| 15                                                                                                                    | 0.7866      | 0.0000               |
| 16                                                                                                                    | 0.7848      | 0.0000               |

**FIGURA 9.54 Resultado del informe de CB Predictor para los datos del Índice de Transportación Dow Jones**

- Después de completar el paso 9, seleccione *Run* (ejecutar) y aparecerá la pantalla *Report* que se muestra en la figura 9.54. Este resultado puede editarse como una hoja de cálculo de Excel. Las autocorrelaciones y sus valores *p* se imprimen bajo la información estadística descriptiva en el informe. También se despliega el valor de la estadística Ljung-Box.

## Referencias

- Akaike, H., "A New Look at the Statistical Model Identification", *IEEE Transactions Automatic Control*, AC-19, 1974: 716-723.
- Bernstein, P., *Against the Gods: The Remarkable Story of Risk*, Nueva York, John Wiley & Sons, 1996.
- Box, G. E. P., G. M. Jenkins y G. C. Reinsel, *Time Series Analysis Forecasting and Control*, tercera edición, Upper Saddle River, NJ, Prentice Hall, 1994.
- DeLurgio, S. A., *Forecasting Principles and Applications*, Nueva York, McGraw-Hill, 1998.

- Diebold, F. X., *Elements of Forecasting*, Cincinnati, OH, South Western, 1998.
- Jenkins, G. y G. McLeod, *Case Studies in Time Series Analysis*, Lancaster, GB, Gwilym Jenkins & Partners Ltd., 1982.
- Kyle, P. W., "Lydia Pinkham Revisited: A Box-Jenkins Approach", *Journal of Advertising Research* 18(2), 1978: 31-39.
- Makridakis, S. S., C. Wheelwright y R. J. Hyndman, *Forecasting Methods and Applications*, tercera edición, Nueva York, John Wiley & Sons, 1998.
- Montgomery, D. S. y L. A. Johnson, *Forecasting and Time Series Analysis*, Nueva York, McGraw-Hill, 1976.
- Newbold, P. y T. Bos, *Introductory Business & Economic Forecasting*, segunda edición, Cincinnati, OH, South-Western, 1994.
- O'Donovan, T. M., *Short Term Forecasting: An Introduction to the Box-Jenkins Approach*, Nueva York, John Wiley & Sons, 1983.
- Pack, D. J., "In Defense of ARIMA Modeling", *International Journal of Forecasting*, 6(2), 1990: 211-218.
- Palda K. S. *The Measurement of Cumulative Advertising Effects*, Upper Saddle River, NJ, Prentice Hall, 1964.
- Pindyck, R. S. y D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, cuarta edición, Nueva York, McGraw-Hill, 1998.
- Quenouille, M. H., "The Joint Distribution of Serial Correlation Coefficients", *Annals of Mathematical Statistics*, 20, 1949: 561-571.
- Schwarz, G., "Estimating the Dimension of a Model", *Annals of Statistics*, 6, 1978: 461-464.



# CAPÍTULO 10

## PRONÓSTICO POR APRECIACIÓN Y AJUSTE DE PRONÓSTICO

Todas las técnicas de pronóstico tratadas en este libro involucran la manipulación de datos históricos para producir los pronósticos, o predicciones, de importantes variables de interés. Los temas analizados en capítulos anteriores trataban complejos análisis de datos y quizás insinuaban que el juicio del analista no estaba involucrado. En realidad, como se enfatizó en el capítulo 1, el criterio es un componente esencial en todas las buenas técnicas de pronóstico. Con el propósito de decidir qué datos son relevantes al problema y para interpretar los resultados del proceso de análisis de datos es necesario el buen juicio, que algunas veces es una parte fundamental del propio análisis.

Este capítulo trata sobre algunos de los principales elementos del pronóstico que se adjuntan o complementan el manejo sistemático de dichos datos históricos. En el capítulo 11 se describen algunas consideraciones sobre el manejo del proceso de pronóstico.

En diversas situaciones de pronóstico, se utiliza sólo el análisis de datos históricos para generar el pronóstico final; la opinión o el criterio del analista no tienen relevancia en el proceso. Este libro se interesa primordialmente en dichas técnicas de pronóstico y, como resultado, en pronósticos de intermedio y corto plazos. Dichos pronósticos son la principal preocupación de los diferentes niveles administrativos en una organización y están asociados con la mayoría de las decisiones básicas que deben tomarse.

Estos procedimientos de pronóstico se basan en el manejo de datos históricos y suponen un pasado y un futuro que son indistinguibles, excepto por las variables específicas identificadas que afectan las probabilidades de resultados futuros. Esta suposición descarta un cambio sustancial en la base tecnológica de la sociedad; sin embargo, desarrollos recientes sugieren que esta conjectura es errónea. Considere, por ejemplo, la introducción y proliferación de computadoras personales de bajo precio y gran velocidad, teléfonos celulares, aparatos personales portátiles de datos, etcétera.

En algunas situaciones de pronóstico, el analista complementa el proceso de análisis de datos después de considerar circunstancias inusuales de la situación o luego de reconocer que la historia pasada no es un pronóstico exacto del futuro. Si los datos históricos son pocos o se consideran parcialmente irrelevantes, el grado de apreciación subjetiva que se aplica al proceso de pronóstico aumenta. En caso extremo, el analista puede opinar que ningún dato histórico es relevante de manera directa en el pronóstico. En estas condiciones, los pronósticos basados únicamente en las opiniones de los “expertos” se usan para formular el pronóstico o escenario para el futuro.

De manera interesante, las investigaciones han mostrado que, con datos históricos a la mano, se reduce la exactitud cuando la apreciación subjetiva modifica los pronósticos producidos por métodos analíticos. Este hallazgo puede atribuirse a algún prejuicio del analista, posiblemente por alguna tendencia a ser demasiado optimista o por menospreciar la

incertidumbre futura. También se ha demostrado que el costo del proceso de pronóstico aumenta si se usa un componente de opinión personal.<sup>1</sup>

Si se desean previsiones o predicciones sobre el futuro y se tienen pocos datos históricos, o ninguno, para ayudar en el proceso de pronóstico, se debe confiar en la apreciación subjetiva. Debido a que tales situaciones son frecuentes, sobre todo en los niveles altos de administración, se han desarrollado técnicas para mejorar la exactitud de dichos pronósticos mediante la utilización óptima de la apreciación ejecutiva disponible. El uso de estas técnicas es digno de consideración porque, habitualmente, los ejecutivos consideran que su juicio es superior a otros métodos para pronosticar el futuro. Como lo afirma Makridakis: “Las personas prefieren hacer pronósticos basados en su apreciación. Creen que su conocimiento del producto, el mercado y los clientes, así como su visión e información interna, les da la habilidad especial de hacer un pronóstico por apreciación.”<sup>2</sup>

A continuación se formulan varias preguntas. Cada una sugiere el uso de la imaginación y una sesión de “lluvia de ideas”, en lugar de basarse completamente en la recopilación y manipulación de datos históricos. Para cada interrogante, una de las técnicas de pronóstico discutidas en este capítulo puede proporcionarle al equipo directivo de una compañía valiosas visiones del futuro ambiente operativo de su empresa.

- ¿Cuál será la distribución de edad en Estados Unidos para el año 2025?
- ¿Hasta qué punto los ciudadanos estadounidenses trabajarán en el hogar dentro de 25 años?
- ¿Qué ciudades serán las más pobladas y tendrán los centros de negocios más importantes dentro de 20 años?
- ¿Hasta qué punto Estados Unidos dependerá de otros países en cuanto a la fabricación de artículos básicos para el consumidor?
- ¿Hasta qué grado las compras en casa mediante la televisión y la computadora, u otros dispositivos electrónicos, serán populares dentro de 20 años?
- ¿Qué tipo de recreaciones preferirán los ciudadanos estadounidenses en el año 2025?
- ¿De cuánto tiempo libre dispondrá el ciudadano promedio estadounidense en el siglo XXI?
- ¿Durante los próximos 25 años se retractará Estados Unidos de sus compromisos alrededor del mundo? Si es así, ¿de qué manera se verán afectados los negocios estadounidenses?

En ocasiones, las técnicas analizadas en este capítulo son llamadas métodos de *pronóstico por apreciación subjetiva*, ya que ésta es el componente primario o único del proceso. A otras se les conoce como métodos de pronósticos tecnológicos, los cuales frecuentemente pronostican los efectos de cambios tecnológicos en el futuro incierto. El resto de este capítulo explica algunos métodos de pronóstico en los que el juicio del analista es el principal ingrediente.<sup>3</sup>

## **EL MÉTODO DELPHI**

---

Cuando se reúne a un grupo de expertos y se les pregunta sobre el futuro, las dinámicas grupales suelen distorsionar el proceso, por lo que se llega a un consenso que no fue considerado cuidadosamente por todos los participantes. El método Delphi, utilizado por primera vez en un proyecto de la RAND Corporation financiado por la fuerza aérea en la década de 1950, trata de excluir las dinámicas de grupo en las deliberaciones de los analistas. En la primera ronda de este método, los expertos responden por escrito a las pregun-

<sup>1</sup> Makridakis (1986), pág. 45.

<sup>2</sup> Makridakis (1986), pág. 63.

<sup>3</sup> Para un resumen organizado de varios métodos de pronóstico, que incluyen pronósticos por apreciación, consulte Georgoff y Murdick (1986).

tas planteadas por el equipo de investigación. Luego, el equipo resume los comentarios de los participantes y se los devuelve por correo.

Entonces, después de leer las reacciones de los demás, los participantes pueden defender sus opiniones originales o modificarlas con base en los razonamientos de los otros.

Este proceso continúa a través de dos o tres rondas hasta que los diversos puntos de vista desarrollados cuidadosamente satisfacen a los investigadores. Entonces, se podría invitar a los participantes a reunirse para compartir y debatir sobre las opiniones. Al concluir este proceso, el equipo investigador debe tener una buena comprensión del futuro y puede empezar a planear la postura de su organización como corresponde.

Como lo señalan Rowe y Wright (1999):

El método Delphi no es un procedimiento diseñado para desafiar los procedimientos estadísticos o los que se basan en modelos, contra los cuales la apreciación humana ha demostrado ser inferior: su intención es ser utilizado en situaciones de apreciación y pronóstico donde los métodos estadísticos basados en modelos no son prácticos o son irrealizables por la carencia de datos históricos, económicos o técnicos..

Cualquier procedimiento Delphi tiene cuatro características básicas: *el anonimato, la repetición, la reacción controlada y la agregación de respuestas del grupo*. El anonimato se logra mediante el uso de cuestionarios que se pueden aplicar a mano, por correo, teléfono o computadora. Cuando los expertos expresan sus opiniones en privado, no sienten las presiones sociales del grupo y se pueden concentrar exclusivamente en los méritos de cada idea, propuesta o situación. Además, con la repetición del cuestionario en varias rondas, los especialistas tienen la oportunidad de cambiar sus opiniones sin perder credibilidad ante los demás integrantes (anónimos) del grupo.

Después de cada repetición del cuestionario se resumen las opiniones individualmente y se informa a cada miembro del grupo sobre la posición de sus colegas anónimos. En consecuencia, la retroalimentación contiene las opiniones y apreciaciones de todos los participantes y no sólo del que más se hace oír. Al terminar la ronda final, la apreciación del grupo se resume (por lo general mediante estadísticas descriptivas, como el promedio o los cuartiles que describen el rango) para que cada juicio reciba la misma consideración.

Muchas veces, la primera ronda del método Delphi no tiene estructura, lo cual deja una gran libertad a los expertos para identificar y comentar los temas que consideran importantes. A continuación, el equipo de monitoreo consolida estos temas en una serie de declaraciones estructuradas (el cuestionario), donde las opiniones y apreciaciones de los panelistas Delphi pueden reunirse de manera cuantitativa en rondas posteriores. A cada experto se le entrega un resumen de los resultados de cada ronda para su futura consideración. Después de la segunda ronda, los miembros tienen la oportunidad de cambiar sus opiniones previas con base en la retroalimentación proporcionada. Es raro que en la mayoría de las aplicaciones de la técnica Delphi las rondas pasen de tres. Los estudios empíricos del método Delphi sugieren que la precisión aumenta con las rondas Delphi y que los panelistas tienden a ser más precisos que en grupos de interacción sin estructura.

#### Ejemplo 10.1

Applied Biosystems suministra a las compañías de ciencias biológicas e instituciones de investigación una gran cantidad de productos y servicios para la investigación de genes y proteínas; estudios sobre cómo interactúan los fármacos con los sistemas del cuerpo y la estructura genética; pruebas sobre alimentos en ambientes contaminados, y la calidad de las identificaciones basadas en el ADN. En los primeros años de la compañía, los sistemas de instrumentos y otros equipos generaban la mayoría de las ganancias por ventas. En años recientes, la rama de consumibles (reactivos, servicios, análisis) ha crecido rápidamente y ahora aporta la mitad de las ganancias de la compañía. Applied Biosystems tiene negocios en Estados Unidos, sobre todo, pero está considerando aumentar su presencia en Europa, Japón y Australia, donde actualmente tiene un porcentaje relativamente pequeño en el mercado de los equipos de biosistemas. Antes de comprometer los gastos de capital asociados a su expansión internacional, la compañía está interesada en desarrollar pronósticos de crecimiento de ventas en Europa, Japón y Australia para los próximos 10 años. Con pocos datos históricos de ventas en estas regiones para guiarse, Ap-

plied Biosystems ha decidido emplear a tres expertos que están familiarizados con las ciencias biológicas y las condiciones económicas internacionales en general.

El experto A es un genetista que tiene mucha experiencia en la industria farmacéutica y ha estudiado el crecimiento de la investigación genética en Europa. El experto B es un renombrado economista con gran experiencia, particularmente en el estudio de la economía de Japón. La experta C es nativa de Australia y ha trabajado en las oficinas centrales de la Organización Mundial de la Salud en Europa, en áreas de tecnología de la salud y farmacéutica. Actualmente trabaja como consultora para el gobierno y la industria de Australia.

Applied Biosystems desea obtener un pronóstico del crecimiento de ventas en las áreas de equipo y consumibles para Europa, Japón y Australia en los próximos 10 años. Para empezar, se le pidió a cada experto que aportara sus estimados de crecimiento de ventas por región para los próximos 10 años de los grupos de consumibles y equipo e instrumentación. A cada uno se le proporcionó el nivel anual de ventas y el estimado de participación de mercado por grupo para Europa, Japón y Australia. Con su presupuesto, Applied Biosystems pudo obtener el compromiso de los tres expertos para dos rondas de reportes. Los expertos anónimos (entre sí), localizados en diferentes partes del mundo, se enlazaban por computadora con el gerente del proyecto de Applied Biosystems.

Los resultados de la primera ronda se resumen en la tabla 10.1.

En la primera ronda, el experto A no observa mucho crecimiento en las ventas para los productos y servicios de Applied Biosystems en Europa en los próximos 10 años. Aunque es probable que la investigación genética tenga un considerable aumento en Europa, la competencia de empresas europeas y foráneas con una fuerte presencia en el mercado biológico puede limitar las oportunidades de Applied Biosystems a pesar de su excelente reputación. Asimismo, el experto A observa un crecimiento relativamente limitado para instrumentación y equipo de Applied Biosystems en Japón, debido a la tendencia de los investigadores japoneses a utilizar el excelente equipo de fabricación japonesa. Sin embargo, él piensa que hay una oportunidad para un crecimiento considerable en las ventas de consumibles en Japón. El experto A no tiene una opinión sobre el crecimiento de las ventas de Applied Biosystems en Australia.

El experto B observa un crecimiento limitado en las ventas de instrumentación y equipo de Applied Biosystems en Europa y Japón, pero considera que hay un importante crecimiento de las oportunidades en Australia; aunque por lo general el mercado australiano es relativamente pequeño. El experto B es un poco más optimista que el experto A sobre las ventas en Europa, pero ambos están de acuerdo en que el crecimiento de las ventas, en especial en consumibles, será pequeño en un periodo de 10 años. El experto B opina que habrá un tremendo crecimiento de las ventas en Australia, tanto para el grupo de instrumentación y equipo como para el de consumibles; y este último es particularmente atractivo. El experto B piensa que Applied Biosystems no tiene gran competencia en Australia y que el compromiso de los australianos de llevar una vida saludable y mantener un ambiente de calidad seguirá siendo fuerte.

La experta C está más insegura que los expertos A y B sobre el crecimiento de ventas de instrumentación y equipo en Europa y sugiere un amplio rango de 0% a 150%. Su pronóstico de crecimiento de ventas en consumibles para Europa, aunque más optimista que los expertos A y B, es aún relativamente pequeño. La experta C no tiene una opinión sobre el crecimiento de ventas de Applied Biosystems en Japón. Es bastante “optimista” sobre el crecimiento potencial de ventas de productos y servicios de Applied Biosystems en su natal Australia. Sus estimados de crecimiento de ventas son más grandes para instrumentación y equipo y consumibles, y de la misma magnitud que los del experto B.

**TABLA 10.1 Resultados del primer ciclo del método Delphi**

|                          | <i>Experto A</i> | <i>Experto B</i> | <i>Experto C</i> |
|--------------------------|------------------|------------------|------------------|
| Instrumentación y equipo |                  |                  |                  |
| –Europa                  | +10% a +40%      | +30% a +70%      | +0% a +150%      |
| Consumibles–Europa       | +5% a +15%       | +10% a +25%      | +20% a +60%      |
| Instrumentación y equipo |                  |                  |                  |
| –Japón                   | +0% a +50%       | -10% a +40%      | —                |
| Consumibles–Japón        | +40% a +100%     | +50% a +200%     | —                |
| Instrumentación y equipo |                  |                  |                  |
| –Australia               | —                | +100% a +200%    | +150% a +300%    |
| Consumibles–Australia    | —                | +200% a +400%    | +200% a +400%    |

A cada experto se le proporcionó la información contenida en la tabla 10.1 y se solicitó una segunda ronda de pronósticos de crecimiento de ventas. Cada especialista tuvo la oportunidad de ajustar sus pronósticos iniciales después de observar los resultados de la primera ronda. Las opiniones del experto A fueron esencialmente iguales. El experto B ajustó el rango de crecimiento de ventas para instrumentación y equipo en el mercado europeo, lo colocó en el rango de +20% a +60%, un ligero ajuste descendente. También ajustó el límite superior de su rango para crecimiento de venta de consumibles en el mercado japonés a +150%. Su pronóstico de crecimiento de ventas para el mercado australiano permaneció sin cambios.

La experta C ajustó los rangos de su pronóstico de crecimiento de ventas para el mercado europeo a

|                                 |             |
|---------------------------------|-------------|
| Instrumentación y equipo—Europa | +10% a +90% |
| Consumibles—Europa              | +15% a +40% |

Sus pronósticos para Australia permanecieron iguales.

Un telefonema final a los panelistas Delphi con los resultados de la segunda ronda no produjo ajustes adicionales. Con los resultados del ejercicio Delphi, Applied Biosystems decidió mantener su actual presencia en Europa haciendo énfasis en sus ventas de instrumentación y equipo. En Japón, la compañía decidió lanzar una importante campaña de marketing para incrementar la venta de los productos de su grupo de consumibles, con la posibilidad de asociarse con una firma japonesa en sus esfuerzos de ventas y distribución. Parecía que Australia era una gran oportunidad, pero en general el mercado es pequeño, así que Applied Biosystems decidió realizar un análisis de costo-beneficio antes de comprometerse con el mercado australiano.

La ventaja del método Delphi es que se les puede pedir a renombrados expertos que consideren el tema de interés y respondan cuidadosamente a los puntos de vista de otros sin la interferencia de una dinámica de grupo. El resultado, si el proceso se maneja con cuidado, puede ser un buen consenso acerca del futuro junto con varios escenarios alternos.<sup>4</sup>

## FORMULACIÓN DE ESCENARIOS

---

La formulación de escenarios involucra la definición de los detalles de un futuro incierto al escribir un “guión” para el entorno de una organización para muchos años en el futuro. La nueva tecnología, la población cambiante y las diversas demandas del consumidor están entre los factores que se consideran y entrelazan en esta especulación para provocar el razonamiento de la administración.

Se describe el escenario más probable junto con otro menos probable, pero posible. Si se considera la postura de la compañía para cada uno de estos posibles entornos futuros, la alta administración está en una mejor posición para reaccionar ante cambios verdaderos en el entorno de los negocios, cuando ocurran, y para reconocer las implicaciones de largo alcance de cambios sutiles que de otra manera podrían pasar inadvertidos. Así, la organización está en una mejor posición para mantener su rentabilidad a largo plazo, en lugar de concentrarse en las ganancias de corto plazo e ignorar el cambiante medio tecnológico en el que opera.

A menudo, el proceso de formular un escenario es seguido por una fase de discusión, algunas veces por un grupo diferente al que desarrolló los escenarios. Entonces, la discusión entre los grupos puede usarse para defender y modificar los puntos de vista, de tal forma que se desarrollan una opinión general sólida y escenarios alternos. Por ejemplo, los escenarios podrían ser generados por el personal de planeación de una compañía y después discutidos con el equipo directivo. Aun cuando posteriormente ninguno de los escenarios demuestre ser totalmente verdadero, este proceso alienta al equipo directivo a pensar ampliamente y lo prepara mejor para reconocer y reaccionar ante los cambios importantes del ambiente.

---

<sup>4</sup> Para una descripción detallada del método Delphi, consulte Parente y Anderson-Parente (1987).

**Ejemplo 10.2**

Una compañía que fabrica cables industriales para teléfonos y televisiones decide llevar a cabo un ejercicio de formulación de escenario antes de su retiro anual de fin de semana. A cada miembro del grupo de retiro se le pide escribir tres escenarios que la compañía podría enfrentar en cinco años: uno del peor de los casos; otro probable, y uno más del mejor de los casos. Después de que las tareas escritas fueron completadas, poco antes del retiro de fin de semana, el presidente y su vicepresidente resumen las contribuciones en los siguientes tres escenarios en los cuales piensan enfocar la discusión del grupo durante el retiro de dos días:

1. El uso de la Internet continúa creciendo rápidamente, pero lentamente se va alejando de los cables en favor del acceso vía satélite. Incluso, el servicio telefónico cada vez utiliza más los medios sin cables, de la misma manera que la televisión lo hace en el hogar. La compañía observa cómo sus ventas y ganancias continúan descendiendo inevitablemente y pronto estará fuera del negocio.
2. Los servicios de Internet y televisión en el hogar continúan creciendo rápidamente, pero se proporcionan mediante varias fuentes. El servicio por satélite se usa ampliamente, pero los cables en tierra continúan siendo una parte básica del servicio de alta tecnología, sobre todo en las ciudades grandes, tanto en aplicaciones industriales como en el hogar. El liderazgo de la compañía en el desarrollo y despliegue de cables produce altas ventas y ganancias.
3. Debido a problemas técnicos y temas de seguridad, el uso del satélite para la Internet y la televisión ha disminuido hasta el punto en que se usa sobre todo en las áreas rurales. El servicio de cable crece rápidamente en las aplicaciones residenciales e industriales y la compañía prospera mientras sus productos bien posicionados y su liderazgo le proporcionan un dominio en la industria.

El presidente y vicepresidente de la compañía piensan tener extensas discusiones sobre cada uno de estos tres escenarios. Asimismo, desean desarrollar estrategias de largo alcance que contengan todas las posibilidades futuras y consideran que si se enfocan en estos tres casos los vigorizarán junto a su equipo de dirección.

## **COMBINACIÓN DE PRONÓSTICOS**

---

Una rama en desarrollo del estudio de los pronósticos involucra la combinación de dos o más métodos de pronóstico para producir los pronósticos finales. Una edición del *International Journal of Forecasting* contenía una sección especial sobre esta nueva técnica. Algunos fragmentos de tres artículos de esta edición ilustran la naturaleza de la combinación de pronósticos en desarrollo:

1. Según Armstrong (1989), la investigación de alrededor de 200 estudios demuestra que la combinación de pronósticos produce ganancias consistentes, aunque modestas, en exactitud. Sin embargo, esta investigación no define bien bajo qué condiciones es más eficaz la combinación ni cómo deben combinarse los métodos en cada situación.
2. Mahoud (1989) afirma que la cantidad de investigación en la mezcla de pronósticos es sustancial. No obstante, es relativamente poco conocido cuándo y cómo los gerentes combinan los pronósticos. Los problemas directivos importantes que requieren el estudio extenso incluyen el ajuste administrativo de pronósticos cuantitativos, el uso de sistemas especializados en la combinación de pronósticos y los análisis de los costos de la combinación de pronósticos.
3. La literatura sobre la combinación de pronósticos se ha acumulado considerablemente al correr de los años. La principal conclusión de esta línea de investigación, según Ciernen (1989), es que la exactitud del pronóstico puede mejorarse sustancialmente por medio de la combinación de pronósticos individuales múltiples. Este documento proporciona una revisión y bibliografía de esa literatura.

**Ejemplo 10.3**

Una compañía que fabrica refacciones para tractores de granja quiere pronosticar cuántas unidades venderá durante los próximos 10 años. Con este pronóstico desarrollará estrategias para permanecer competitiva en su negocio. Específicamente, la empresa está preocupada por la capacidad de su planta. Un pronóstico de negocios futuros le ayudaría mucho a desarrollar sus planes de expansión y a tratar con sus fuentes de apoyo financiero.

TABLA 10.2 Combinación de pronósticos

| Año en el futuro | Pronóstico profesional | Pronóstico de la empresa | Pronóstico final |
|------------------|------------------------|--------------------------|------------------|
| 1                | 328                    | 335                      | 329.8            |
| 2                | 342                    | 340                      | 341.5            |
| 3                | 340                    | 345                      | 341.3            |
| 4                | 348                    | 350                      | 348.5            |
| 5                | 350                    | 352                      | 350.5            |
| 6                | 360                    | 355                      | 358.8            |
| 7                | 366                    | 365                      | 365.8            |
| 8                | 371                    | 370                      | 370.8            |
| 9                | 385                    | 375                      | 382.5            |
| 10               | 390                    | 385                      | 388.8            |

Después de recopilar datos extensivamente y realizar un pronóstico, la firma tiene que decidir cuál de los dos pronósticos debe aceptar. Y aunque no están demasiado lejos para la mayoría de los años, hay algunas diferencias. Una empresa profesional en la materia, que tiene una sólida reputación, generó el primer pronóstico. Los métodos que esta firma usó son desconocidos para la administración de la fábrica, pero le dijeron que el proceso era “matemáticamente sofisticado”. El segundo pronóstico fue el obtenido en un retiro ejecutivo al que asistieron directivos y personal de marketing que estaban familiarizados con las expectativas de sus clientes.

Después de haberlo analizado con ejecutivos de la compañía, el presidente decidió combinar los dos pronósticos. Estas discusiones tendieron a favorecer el pronóstico profesional sobre el interno, por lo que el presidente ponderó el anterior 75% y el último 25%. La tabla 10.2 muestra los dos pronósticos para cada uno de los próximos 10 años en unidades vendidas —en miles— seguidos por el pronóstico combinado. Cada pronóstico final se calculó multiplicando el primer pronóstico por 0.75, el segundo por 0.25 y después sumándolos. Por ejemplo, para el año 1 el pronóstico combinado se da por

$$(0.75)328 + (0.25)335 = 329.75$$

Observe que los pronósticos finales se ubican entre las previsiones profesionales y las internas, pero se acercan más a los pronósticos profesionales porque éstos recibieron el mayor peso.

Uno de los beneficios de la combinación de pronósticos es el de minimizar los efectos de predisposición, es decir, dar un peso indebido a un método de pronóstico en particular. Una estrategia es combinar los diferentes pronósticos simplemente promediando los pronósticos individuales producidos por diferentes métodos. Si  $\hat{Y}_{11}, \hat{Y}_{12}, \dots, \hat{Y}_{1m}$  son pronósticos anticipados un paso producidos por los diferentes métodos  $m$ , entonces el pronóstico anticipado un paso combinado,  $\hat{Y}_{1C}$ , obtenido sencillamente mediante el promedio es

$$\hat{Y}_{1C} = \frac{\hat{Y}_{11} + \hat{Y}_{12} + \dots + \hat{Y}_{1m}}{m} \quad (10.1)$$

Resultados similares aplican para otros pronósticos cuyo tiempo principal puede ser mayor a un periodo de anticipación.

Como en el ejemplo 10.2, los pronósticos pueden combinarse con pesos desiguales. En este caso, cada pronóstico anticipado un paso  $\hat{Y}_{1i}$  recibe el peso  $0 < w_i < 1$ , donde  $\sum w_i = 1$ . Así, con  $m$  pronósticos anticipados un paso producidos por  $m$  diferentes métodos, el pronóstico combinado  $\hat{Y}_{1C}$  es

$$\hat{Y}_{1C} = w_1 \hat{Y}_{11} + w_2 \hat{Y}_{12} + \dots + w_m \hat{Y}_{1m} \quad (10.2)$$

donde

$$\sum_{i=1}^m w_i = 1$$

Aunque el pronóstico principal para varios períodos de anticipación cambie, los pesos permanecen igual. Por ejemplo, el pronóstico anticipado dos pasos,  $\hat{Y}_{2C}$ , se puede calcular como el pronóstico combinado de la ecuación 10.2 con los pronósticos anticipados dos pasos para cada método que reemplaza el pronóstico anticipado un paso. El hecho de que los pesos sumen uno asegura que el pronóstico combinado estará en alguna parte entre los valores más pequeños y los más grandes de los pronósticos individuales.

Hay varios procedimientos para determinar los pesos  $w_i$ . Si está disponible un registro de desempeño pasado para cada método de pronóstico, los pesos pueden tomarse como inversamente proporcionales a la suma del error cuadrático del pronóstico.

Otra manera de obtener los pesos es mediante el uso de los métodos de la regresión. En Newbold y Bos (1994) hay un análisis adicional sobre los procedimientos para la combinación de pronósticos.

En años venideros, es probable que la investigación se dirija a las ventajas de la combinación de pronósticos, junto con las técnicas para lograrlo. El objetivo de dichas combinaciones será desarrollar pronósticos exactos que sean rentables.

## REDES NEURALES Y PRONÓSTICOS

---

Los métodos convencionales de pronóstico, como los que se tratan en este libro, dependen de los datos históricos para desarrollar un modelo y usarlo para proyectar las variables de interés en el futuro. Estos valores proyectados se convierten en los pronósticos que se usan para desarrollar los planes del negocio. En estos modelos se supone que el futuro será exactamente igual que el pasado, excepto por las variables específicamente reconocidas por el modelo.

A veces los modelos convencionales hacen suposiciones —que pueden estar sujetas a verificación o no— sobre la forma de distribución de la población. La estimación del intervalo mediante un modelo de regresión, por ejemplo, supone que la población subyacente sigue una distribución normal.

El campo de la inteligencia artificial, aún en desarrollo, trata de reproducir los procesos del cerebro humano y el sistema nervioso mediante la computadora. Aunque este campo se originó en la biología y la psicología, está avanzando rápidamente hacia otras áreas, incluso en los negocios y la economía. Los principales aspectos de la inteligencia artificial son el procesamiento del idioma, la robótica y las redes neurales artificiales. Este último campo tiene las aplicaciones más comerciales, incluyendo los pronósticos.

En las redes neurales, muchos ejemplos se programan en la computadora, los cuales capturan el rango completo de relaciones del pasado entre todas las variables que podrían afectar el resultado de las variables dependientes. El programa de la red neural asimila estos ejemplos y trata de desarrollar las relaciones subyacentes, “aprendiendo” mientras avanza. Este proceso de aprendizaje, también llamado *entrenamiento*, es análogo a un aprendiz humano que se capacita en el trabajo.

Algunos investigadores de los pronósticos han notado la similitud que hay entre las técnicas de la red neural y los métodos convencionales de pronóstico que intentan encontrar con éxito las variables para predecir la variable dependiente. La ventaja teórica de la red neural como una herramienta de pronóstico es que las relaciones no necesitan especificarse de antemano, pues el método involucra el aprendizaje de las relaciones mediante los ejemplos proporcionados. Asimismo, las redes neurales no requieren ninguna suposición sobre las distribuciones subyacentes de la población y, al contrario de muchos métodos convencionales de pronóstico, pueden operar con datos incompletos.<sup>5</sup>

<sup>5</sup> Aunque las redes neurales no producen pronósticos por apreciación subjetiva en el sentido habitual, los pronósticos de red neural son distintos de los procedimientos que se basan en datos discutidos en los capítulos del 4 al 9, así que están incluidos en este capítulo. Una red neural es un esfuerzo para poner el proceso del pensamiento humano en una computadora. Además, para utilizar una red neural eficazmente se requiere de mucha intuición.

Frecuentemente, una aplicación exitosa de la red neural es llamada *plug-compatible*. Esto significa que el programa de la red neural puede reemplazar rápida y fácilmente al modelo actual, como el análisis de regresión, sin distorsionarlo. Algunas veces se puede lograr un mejoramiento en el desempeño, como pronósticos más exactos, con un impacto mínimo en las operaciones en curso. Las redes neurales son especialmente valiosas cuando no hay datos de entrada o éstos están altamente correlacionados, o cuando los sistemas no son lineales.

A continuación se dan algunos ejemplos del uso exitoso de redes neurales en escenarios prácticos, proporcionados por NeuralWare Inc., un proveedor comercial de software de redes neurales.<sup>6</sup>

- Una planta de Kodak situada en Texas redujo sus costos por tres millones de dólares anuales, manteniendo el rendimiento y la calidad del producto. Comenzó por recolectar datos históricos de operación que fueron usados para entrenar una red neural, a fin de pronosticar la calidad del producto como una función de los diversos parámetros del proceso.
- Uno de los problemas clásicos cuando se interpretan los signos sísmicos para la explotación petrolera es encontrar la primera evidencia de la onda expansiva en las grabaciones del sensor. Los métodos tradicionales para procesar las señales no han tenido mucho éxito en la correcta identificación de la primera evidencia —sólo 30% para un algoritmo—. Algunas compañías petroleras han usado un programa de la red para identificar correctamente la primera onda, de las cuales Amoco tiene un porcentaje de éxito de 95%. Esto ha sido de gran trascendencia en su habilidad para mejorar el procesamiento de datos sísmicos para construir sus depósitos.
- Un fabricante alemán de motores de ventiladores eléctricos para autos utilizó a sus maestros inspectores para determinar si un motor terminado era demasiado ruidoso o hacia sonidos “raros”. Una red neural tardó una semana para aprender del maestro inspector y actualmente la computadora está haciendo la prueba del ruido. La repercusión financiera es un producto de calidad superior con menos quejas de los clientes. Los maestros inspectores ya no realizan este trabajo estresante para concentrarse en otras tareas de mejoramiento de calidad en la fábrica.
- Desde 1989, una compañía de crédito ha usado una red neural para determinar a qué clientes debe otorgar el crédito inmediato.
- Varias compañías han descubierto que las redes neurales se pueden utilizar para desarrollar indicadores útiles para el comercio. Los resultados de los sistemas revelan que los indicadores comerciales de las redes neurales, acoplados con estrategias comerciales adecuadas, pueden producir dinero en el mercado de manera consistente.
- La Fuerza Aérea de Estados Unidos está usando una red neural para pronosticar fallas en los componentes de sus aviones. Se recopiló una variedad de datos sobre cada avión y la red fue entrenada para pronosticar la probabilidad de fracasos específicos en aviones determinados. Estos resultados se usan para abastecer las refacciones y para el mantenimiento preventivo.
- Se ha desarrollado un producto comercial para pronosticar los resultados del hipódromo. El sistema requiere que el usuario introduzca información específica sobre los caballos en una carrera. El 80% de las veces que decide que las probabilidades favorecen la apuesta, gana.

Actualmente, ejemplos similares a éstos son exitosos en la práctica, por lo cual no es difícil imaginar un sistema de redes neurales que pronostique exitosamente el volumen mensual de ventas, la tasa de ausentismo de un empleado o la prima de interés. Dichas redes tienen un gran potencial para hacer un pronóstico exacto, sin las restricciones de tener que especificar las relaciones de antemano y asumir las distribuciones de la población, y su uso continúa ganando aceptación. El caso 10.2, que está al final de este capítulo, es un ejemplo de una aplicación de redes neurales.

---

<sup>6</sup> NeuralWare, Inc., Penn Center West, Building IV, Suite 227, Pittsburgh, PA 15276.

## RESUMEN DEL PRONÓSTICO POR APRECIACIÓN SUBJETIVA

---

Un peligro potencial al utilizar la mayoría de los métodos de pronóstico discutidos en este texto es que involucran la manipulación de datos históricos para generar el pronóstico. Esta práctica es un procedimiento válido si el pronóstico es para el futuro cercano, pero se vuelve muy incierto conforme aumenta el horizonte del pronóstico. Los métodos analizados en este capítulo son valiosas adiciones al arsenal del analista cuando se considera la preocupación particular de la dirección, el pronóstico a largo plazo del entorno de la compañía.

Varios autores han fascinado al público con sus especulaciones sobre las tendencias de largo plazo que se desarrollan actualmente en nuestra sociedad y el futuro que pronostican junto con las habilidades administrativas que se necesitarán para sobrellevarlas. Entre los más provocativos están:

- Bill Gates, *The Road Ahead* (Viking, 1995)
- Rowan Gibson (editor), Alvin Toffler y Heidi Toffler, *Rethinking the Future: Rethinking Business, Principles, Competition, Control and Complexity, Leadership, Markets and the World* (Nicholas Breasley, 1999)
- Di Kamp, *The 21<sup>st</sup> Century Manager: Future-Focused Skills for the Next Millennium* (Kogan Page, 1999).
- Chuck Martin, *Net Future: The 7 Cybertrends That Will Drive Your Business, Create New Wealth, and Define Your Future* (McGraw-Hill, 1998).
- Alvin Toffler, *Future Shock* (Nueva York, Random House, 1970).
- Alvin Toffler, *The Third Wave* (Nueva York, Bantam Books, 1980).
- John Naisbitt, *Megatrends* (Nueva York, Warner Books, 1982).
- John Naisbitt y Patricia Aburdene, *Megatrends 2000* (Nueva York, William Morrow, 1990).
- John Naisbitt (audiocasete), *Eight Asian Megatrends That Are Reshaping the World* (Simon & Schuster Books, 1996).
- John Naisbitt, *Global Paradox* (Avon, 1995).
- Jim Taylor, Watts Wacker y Howard Means, *The 500-Year Delta: What Happens After What Comes Next* (Harperbusiness, 1998).

Las referencias anteriores son lecturas interesantes, aunque algunos de ellos tengan 20 o 30 años de antigüedad, porque estos autores tienen la habilidad única de especular sobre el futuro de maneras provocativas. Un enfoque más formal para la clase de cambios que ellos predicen es conocido como *pronóstico tecnológico*, el cual es campo que se está volviendo muy importante para muchas empresas. Esto se debe a que un estimado 25% de la tecnología existente se reemplaza cada año. Bajo estas circunstancias, incluso una manipulación muy sofisticada de datos históricos para producir pronósticos puede fallar por un amplio margen. Por ejemplo, considere los siguientes campos en vía de desarrollo. Cuando estas tecnologías se desplieguen, su trascendencia será considerable en muchas empresas.

- Inteligencia artificial
- Ingeniería genética
- Bioelectricidad
- Robótica multisensorial
- Láser
- Fibra óptica
- Microondas
- Satélites avanzados

- Energía solar
- Superconductores

PricewaterhouseCoopers edita una publicación interna, *Technology Forecast*, que está disponible para el público en general. Este documento anual analiza importantes tecnologías de información y pronostica sus direcciones en los próximos uno a tres años. Éste es un ejemplo de la intensa preocupación de una compañía sobre los adelantos tecnológicos y sus efectos en la sociedad, así como en su propia empresa.

La tasa del desarrollo tecnológico aumenta la necesidad de los administradores de estar atentos a los cambios. Aun así, los patrones históricos pueden proporcionar por lo menos una pista del futuro en casi todos los negocios. Como se enfatizó en el capítulo 1, si se desea generar pronósticos útiles, se requiere una mezcla prudente de la manipulación de los datos históricos apropiados junto con el uso juicioso de la apreciación personal y la experiencia.

## OTRAS HERRAMIENTAS ÚTILES PARA JUZGAR EL FUTURO

---

Se han desarrollado algunas herramientas para ayudar a las personas que toman las decisiones a considerar la información sobre el futuro y tomar las mejores decisiones posibles. Hay una serie de herramientas frecuentemente discutidas en los libros de texto y artículos periodísticos bajo el encabezado general de “toma de decisiones”, otras veces como “toma de decisiones bajo incertidumbre”. Cuando los gerentes se enfrentan a la tarea de tomar decisiones bajo incertidumbre, su habilidad de pronóstico es un elemento fundamental en el proceso de la toma de decisiones.

Cuando se ha completado el análisis de datos históricos, la persona que toma las decisiones debe hacer una apreciación personal respecto a las alteraciones en el curso de acción de la empresa. En otras palabras, el analista debe entrelazar los resultados del proceso de pronóstico con los procedimientos existentes de la toma de decisión de la empresa. A continuación se discuten brevemente algunos elementos de la teoría de las decisiones que a menudo son pertinentes en esta fase del proceso.<sup>7</sup>

El concepto del *valor esperado* se describió en el capítulo 2. Las personas que toman las decisiones lo usan de manera habitual, explícita o implícitamente. Recuerde que este concepto requiere calcular el valor medio que la variable numérica aleatoria asumirá a lo largo de muchas pruebas. En la tabla 10.3 se muestra una variable aleatoria discreta  $X$  en una distribución de probabilidad; asimismo, se presenta cada posible valor futuro que  $X$  puede asumir junto con la probabilidad de cada uno.

Observe que la suma de las probabilidades en la tabla 10.3 es 1.00, o 100%, lo cual significa que se ha identificado cada posible valor que  $X$  puede asumir. En la tabla 10.3,  $X$  re-

**TABLA 10.3 Una distribución de probabilidades**

| $X$ | $P(X)$ |
|-----|--------|
| 1   | .10    |
| 2   | .20    |
| 3   | .25    |
| 4   | .30    |
| 5   | .15    |
|     | 1.00   |

<sup>7</sup> Para un libro excelente y muy ameno sobre la toma de decisiones, consulte Hammond, Keeney y Raiffa (1999).

presenta el número de nuevos contratos grandes que una empresa firmará durante el próximo año fiscal. La pregunta que responde el valor esperado es: ¿cuántos nuevos contratos pueden esperarse, en promedio, si la distribución de probabilidad de la tabla 10.3 es válida? La ecuación 10.3 se usa para calcular el valor esperado de una distribución de probabilidad como el mostrado en la tabla 10.3.

$$E(X) = \Sigma X[P(X)] \quad (10.3)$$

donde

$E(X)$  = Valor esperado

$X$  = Valores que puede asumir la variable aleatoria

$P(X)$  = Probabilidad de que cada  $X$  ocurra

Se puede calcular el valor esperado de la tabla 10.3 utilizando la ecuación 10.3 como sigue:

$$E(X) = 1(.10) + 2(.20) + 3(.25) + 4(.30) + 5(.15) = 3.2$$

El valor esperado de la distribución de probabilidad mostrada en la tabla 10.3 es 3.2. Si en este ejemplo  $X$  representa el número de nuevos contratos importantes durante el próximo año fiscal, entonces se firmarán en promedio 3.2 nuevos documentos, siempre y cuando las oportunidades para los nuevos compromisos sigan siendo las mismas cada año. Observe que el valor 3.2 no es posible en cualquier año; sólo son posibles los valores enteros 1, 2, 3, 4 y 5. No obstante, el valor 3.2 representa el resultado medio de muchas pruebas. Las personas que toman las decisiones frecuentemente están interesadas en los valores esperados y los usan como sus mejores pronósticos para las variables numéricas fundamentales cuando planean para el futuro incierto.

La teoría de las decisiones dirige los elementos que componen la función de toma de decisiones de líderes de negocios. Con frecuencia, los valores esperados se entrelazan en esta consideración más general de la toma de decisiones. El *diagrama de árbol de la decisión* se usa para ayudar al responsable de las decisiones a visualizar una situación compleja y tomar decisiones racionales. Dicho diagrama se muestra en la figura 10.1.

La figura 10.1 refleja la incertidumbre existente sobre la naturaleza de las ventas futuras e incorpora la decisión de si se debe construir una nueva planta o reparar la antigua. El problema radica en que si la compañía supiera que la demanda será alta, lo mejor sería construir una nueva; por otro lado, si se supiera que la demanda bajaría, las ganancias serían más altas si repararan la planta antigua. Incluso en este simple ejemplo se puede observar el beneficio del diagrama de árbol: éste le permite a la persona que toma las decisiones observar las diferentes opciones disponibles, identificar las incertidumbres fuera del control de la empresa y determinar explícitamente los costos, ganancias y probabilidades de sucesos futuros. En situaciones más complicadas, los beneficios del diagrama de árbol y la teoría de la decisión formal son más evidentes.

Un concepto estadístico diseñado para examinar las probabilidades preliminares con base en la evidencia de muestra es el *teorema de Bayes*. Habitualmente, este concepto se aplica a las situaciones en las que se determinan los estimados de las probabilidades de sucesos del futuro desconocido y se modifican después de recolectar la evidencia de la muestra. Un ejemplo es el concepto de prueba de mercado utilizado por muchos fabricantes de productos de consumo. Una de estas compañías podría calcular la probabilidad de aceptación pública de un nuevo producto como muy alta. Sin embargo, antes de arriesgar los millones de dólares que requiere una campaña nacional, se puede emprender una prueba de mercado en las áreas que la compañía considera buenos mercados representativos. Los resultados de la prueba de mercado son utilizados para modificar los cálculos originales de éxito del producto, y se toma una decisión sobre la introducción

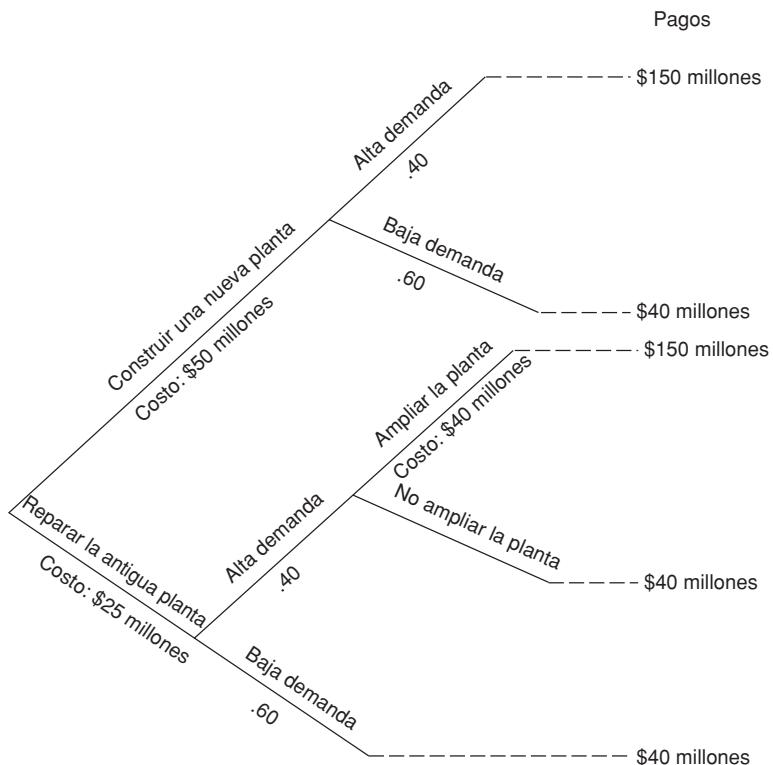


FIGURA 10.1 Diagrama de árbol de la decisión

del producto en el plano nacional. Una versión simplificada del teorema de Bayes se muestra en la ecuación 10.4.<sup>8</sup>

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (10.4)$$

donde

$P(A \setminus B)$  = Probabilidad de que ocurra el evento A dado que ha ocurrido el evento B

$P(B)$  = Probabilidad de que ocurra el evento B

$P(A)$  = Probabilidad de que ocurra el evento A

$P(B \setminus A)$  = Probabilidad de que ocurra el evento B dado que ha ocurrido el evento A

#### Ejemplo 10.4

La figura 10.2 refleja una aplicación específica del teorema de Bayes en una situación de prueba de mercado. La administración de una gran compañía de productos de consumo necesita decidir si debe introducir un nuevo producto en el plano nacional. También estima que su nuevo producto tiene 50% de oportunidad de altas ventas en el mercado nacional; es decir,  $P(H) = 0.50$ . Está considerando el uso de una prueba de mercado para determinar si puede hacer un mejor trabajo al pronosticar las ventas altas o bajas del nuevo producto. La figura 10.2 muestra un árbol de decisión de los resultados de la prueba de mercado.

<sup>8</sup> El enunciado general del teorema de Bayes para los eventos  $n A_1, A_2, \dots, A_n$ , cada uno de los cuales puede ocurrir junto con otro evento B, es:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)} \quad i = 1, 2, \dots, n$$

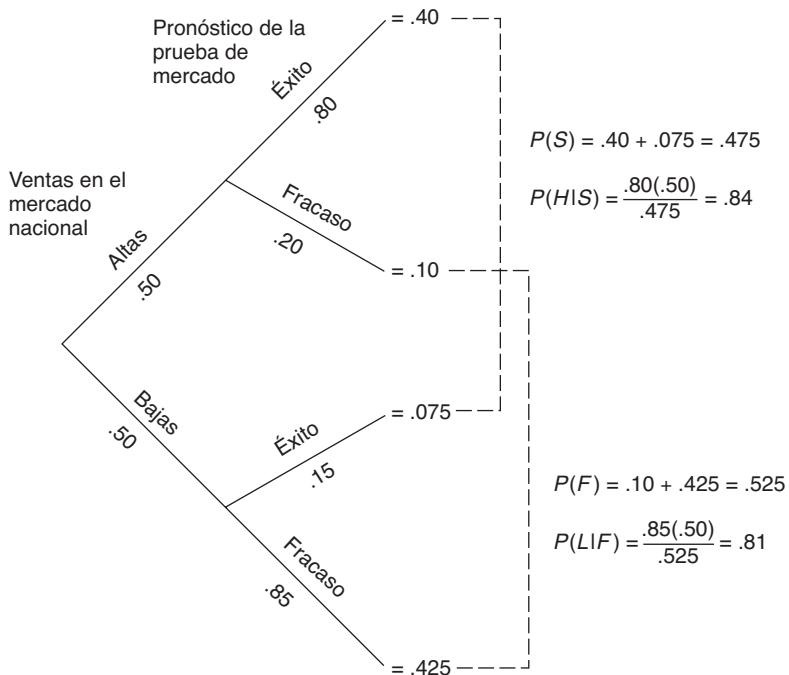


FIGURA 10.2 Decisión de Bayes para la introducción de un nuevo producto

Las experiencias pasadas han mostrado que cuando se introdujo un nuevo producto y se lograron ventas altas, la prueba de mercado tuvo éxito en 80% del tiempo, o  $P(SH) = 0.80$ . Las experiencias pasadas también muestran que cuando se introdujo un nuevo producto y se obtuvieron ventas bajas, las pruebas de mercado fallaron (ventas bajas) 85% del tiempo, o  $P(FL) = 0.85$ .

El teorema de Bayes puede usarse para calcular la probabilidad de éxito del nuevo producto en el mercado nacional bajo diferentes condiciones de prueba. La figura 10.2 muestra el cálculo de las dos probabilidades que nos interesan:

- Si la prueba de mercado pronostica éxito, existe 84% de probabilidad de que el producto tendrá ventas altas en el ámbito nacional. Esto equivale a

$$P(H|S) = 0.84 = 0.80(0.50)/0.475 = P(S|H)P(H)/P(S)$$

- Si la prueba de mercado pronostica fracaso, existe 81% de probabilidad de que el producto tendrá ventas bajas en el pleno nacional, o

$$P(L|F) = .81 = .85(.50)/.525 = P(F|L)P(L)/P(F)$$

En el ejemplo 10.4 una prueba de mercado ayudaría a la administración a decidir si debe, o no, introducir el nuevo producto a nivel nacional. La exactitud de la prueba de mercado es suficiente para cambiar la probabilidad de ventas altas (o bajas) del valor de la prueba preliminar de 50%. Si la prueba de mercado tiene éxito, la oportunidad de ventas altas se incrementa a 84%. Si la prueba de mercado fracasa, la oportunidad de ventas bajas aumenta a 81%. Estas probabilidades se calcularon mediante el teorema de Bayes. La decisión sobre la introducción del producto será mucho más fácil de tomar de lo que lo habría sido sin la prueba de mercado.

Otra herramienta útil para el pronóstico es la simulación en computadora. La *simulación* es una serie de técnicas numéricas y de programación para modelar situaciones sujetas a incertidumbre; también, para efectuar experimentos de muestra en dichos modelos mediante una computadora. Cada vez que se corre una simulación (repetición) ésta pro-

duce un resultado posible (pronóstico) para el problema que está siendo estudiado. Si se corre muchas veces (repeticiones), le da la oportunidad a la persona que toma las decisiones de observar la serie completa de posibles resultados, así como su probabilidad de ocurrencia. Entonces, estos escenarios generados por computadora pueden resumirse y ser usados para tomar la mejor decisión.

Hay muchas razones para usar la simulación en lugar del análisis matemático para recabar la información sobre un ambiente incierto.

- Muchas representaciones realistas de sistemas comerciales verdaderos son demasiado complejas para ser analizadas matemáticamente.
- El interés principal podría ser experimentar con el sistema o encontrar los mejores niveles para las variables que influyen en él o simplemente estudiarlo. Puede resultar imposible experimentar con el sistema real (situaciones de combate) o sumamente caro (el diseño de instalaciones de generación nuclear) o que consume tanto tiempo que sólo es posible realizar una repetición.
- Frecuentemente, la simulación es útil en sí misma porque facilita un mejor entendimiento del sistema.
- Como herramienta, la simulación brinda cierto tipo de credibilidad con la administración. Es relativamente fácil explicar a la administración los esfuerzos involucrados en un estudio de simulación.

Desde una perspectiva comercial, la principal ventaja de un ejercicio de simulación en computadora es que se pueden generar pronósticos de las consecuencias de varias decisiones administrativas antes de que éstas sean tomadas. La simulación se ha usado en una amplia variedad de escenarios, como la planeación de operaciones industriales, la organización de la mano de obra o de las alternativas de inversión, el análisis y precio de sofisticados instrumentos de inversión y los flujos de pasajeros en una terminal aérea.<sup>9</sup> A menudo, los árboles de decisión como el de la figura 10.1 son parte de los experimentos de simulación en computadora.

## Fórmulas clave

### **Pronóstico combinado determinado por el promedio simple**

$$\hat{Y}_{1C} = \frac{\hat{Y}_{11} + \hat{Y}_{12} + \cdots + \hat{Y}_{1m}}{m} \quad (10.1)$$

### **Pronóstico combinado determinado por el promedio ponderado**

$$\hat{Y}_{1C} = w_1 \hat{Y}_{11} + w_2 \hat{Y}_{12} + \cdots + w_m \hat{Y}_{1m} \quad (10.2)$$

### **Fórmula del valor esperado**

$$E(X) = \Sigma X[P(X)] \quad (10.3)$$

### **Teorema de Bayes**

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (10.4)$$

---

<sup>9</sup> Interesantes discusiones sobre la simulación en computadora están disponibles en Shela, Ceric y Tadika-malla (2003), Law y Kelton (2000) y Banks (1998).

TABLA P.1

| Mes | Ventas reales | Pronósticos de Winters | Pronósticos de regresión |
|-----|---------------|------------------------|--------------------------|
| 1   | 5502          | 4586                   | 5263                     |
| 2   | 6535          | 5944                   | 6008                     |
| 3   | 7390          | 6706                   | 6832                     |
| 4   | 4388          | 4530                   | 4886                     |
| 5   | 4521          | 5010                   | 4918                     |
| 6   | 5679          | 6574                   | 5630                     |
| 7   | 8578          | 8462                   | 7963                     |
| 8   | 7088          | 6415                   | 5942                     |
| 9   | 4839          | 4457                   | 5156                     |
| 10  | 4050          | 4638                   | 3819                     |

## Problemas

- Identifique dos situaciones de negocios donde se podría usar el método Delphi para generar los pronósticos.
- Considere las ventas reales mostradas en la tabla P.1, junto con los pronósticos anticipados un paso producidos por el método Winters y por un modelo de regresión.
  - Construya los pronósticos combinados de ventas, producidos por un promedio simple de los pronósticos generados por el método de Winters y el modelo de regresión.
  - Construya los pronósticos combinados de ventas, producidos por un promedio ponderado de los pronósticos de Winters y los pronósticos de regresión con los pesos  $w_1 = 0.8$  y  $w_2 = 1 - w_1 = 0.2$ .
  - Usando las ventas reales, determine el *MAPE* para los pronósticos de Winters y los pronósticos de regresión.
  - Repita el inciso c usando los pronósticos combinados de los incisos a y b. Basándose en la medida *MAPE*, ¿qué serie de pronósticos prefiere usted?

## CASOS

---

### CASO 10-1 RESTAURANTE GOLDEN GARDENS

Sue y Bill Golden han decidido abrir un restaurante en una ciudad del medio oeste de Estados Unidos. Ambos han pasado un año investigando el área y visitando restaurantes de precios entre medianos y altos. Definitivamente creen que hay espacio para otro restaurante y han encontrado un buen sitio a precio razonable.

Además, tienen contacto con varios cocineros de primera clase y creen poder atraer a uno de ellos a su nuevo negocio. Su investigación con banqueros locales los han convencido de que el financiamiento estará disponible pronto, debido a sus propios recursos

financieros y a su experiencia en el negocio de los restaurantes.

Lo único que todavía preocupa a los Golden es la decoración y el tema ambiental de su restaurante. Ya han trabajado con una serie de tres grupos de residentes del área que comen fuera regularmente, pero no han obtenido ningún consenso general en esta materia. También han discutido este asunto entre sí, pero ahora consideran que serían valiosas otras opiniones.

Después de leer sobre algunas de las técnicas utilizadas en los pronósticos por apreciación, creen que alguna de estas técnicas podría ayudarles a decidir la

atmósfera para su nuevo restaurante. Han identificado a varios de sus amigos y socios en otras ciudades que

estarían dispuestos a ayudarles, pero los Golden no están seguros de cómo utilizar sus talentos. ■

#### PREGUNTAS

1. ¿Qué método sugeriría usted a los Golden para utilizar la experiencia de sus amigos a fin de decidir la atmósfera y el motivo de su nuevo restaurante?
2. ¿Existen otros métodos que han pasado inadvertidos mientras trataban de investigar sobre este asunto?

## CASO 10-2 ALOMEGA FOOD STORES

El ejemplo 1.1 del capítulo 1 describió cómo Julie Ruth, la presidenta de Alomega Food Stores, recopiló los datos de las ventas mensuales para su compañía (vea la tabla 10.4) junto con otras variables que ella pensó que se relacionaban con las ventas. Los casos Alomega en los capítulos 2, 3, 5 y 8 describieron los esfuerzos de Julie por usar varios procedimientos en Minitab en un intento por obtener pronósticos significativos de las ventas mensuales.

En el caso 8.7 Julie desarrolló un modelo de regresión múltiple que explicó casi 91% de la varianza en las ventas mensuales. Ella pensó bien sobre este modelo, pero después se sintió especialmente susceptible con los comentarios negativos hechos por Jackson Tilson, su gerente de producción, durante una reunión (vea el final del ejemplo 1.1). Tilson dijo: “He estado tratando de mantener mi boca cerrada durante esta reunión, pero

esto es demasiado. Creo que estamos gastando el tiempo de muchas personas con toda esta colección de datos y estamos jugando con las computadoras. Todo lo que usted tiene que hacer es hablar con el personal de piso y con los gerentes de las tiendas de comestibles para entender lo que está pasando. Yo he visto esto pasar antes, y aquí vamos de nuevo. Algunos de ustedes necesitan apagar sus computadoras, salir de sus elegantes oficinas y hablar con personas reales.” Julie decidió que la política de la oficina dictaba que ella debía prestar atención al consejo de Jackson. Consultó con varias personas, incluso con Tilson, para determinar sus opiniones sobre cómo prever las ventas para enero de 2003. Una gran mayoría pensaba que obtendría un mejor pronóstico si usara las ventas del mes de enero del año anterior. Asimismo, el pronóstico para febrero de 2003 debería basarse en el cuadro de ventas de febrero de 2002.

**TABLA 10.4** Ventas mensuales para 27 establecimientos de Alomega Food Stores, 1999-2002, para el caso 10.2

| Mes  | 1999    | 2000    | 2001    | 2002    |
|------|---------|---------|---------|---------|
| Ene. | 425,075 | 629,404 | 655,748 | 455,136 |
| Feb. | 315,305 | 263,467 | 270,483 | 247,570 |
| Mar. | 432,101 | 468,612 | 429,480 | 732,005 |
| Abr. | 357,191 | 313,221 | 260,458 | 357,107 |
| May. | 347,874 | 444,404 | 528,210 | 453,156 |
| Jun. | 435,529 | 386,986 | 379,856 | 320,103 |
| Jul. | 299,403 | 414,314 | 472,058 | 451,779 |
| Ago. | 296,505 | 253,493 | 254,516 | 249,482 |
| Sep. | 426,701 | 484,365 | 551,354 | 744,583 |
| Oct. | 329,722 | 305,989 | 335,826 | 421,186 |
| Nov. | 281,783 | 315,407 | 320,408 | 397,367 |
| Dic. | 166,391 | 182,784 | 276,901 | 269,096 |

Basándose en esta aportación, Julie desarrolló un modelo informal de pronóstico

$$\hat{Y}_{t+1} = Y_{t-11}$$


---

el cual utilizó el valor mensual del último año para prever el valor mensual de este año. ■

## PREGUNTAS

1. ¿Qué tan exacto es el modelo informal de pronóstico de Julie?
2. ¿Cómo se compara el modelo informal de pronóstico con el modelo de regresión múltiple desarrollado en el caso 8.7?
3. Hasta que Julie pueda experimentar cada uno de estos dos métodos en acción, está considerando la combinación de pronósticos. Piensa que este enfoque contrarrestaría la política de la oficina y aún así le permitiría usar un enfoque más científico. ¿Será ésta una buena idea?
4. ¿Debería Julie usar un enfoque de promedio simple o de promedio ponderado para combinar los pronósticos?

## CASO 10-3 OTRA VISITA A LA LYDIA E. PINKHAM MEDICINE COMPANY<sup>10</sup>

Este caso demuestra una aplicación real del uso de redes neuronales para prever datos de la serie de tiempo. Los autores entendemos que no se les ha proporcionado a los estudiantes el antecedente para entender este caso por completo. Sin embargo, pensamos que sería un gran beneficio para ellos experimentar este caso real.

En el caso 9.4 se presentaron las compañías Lydia E. Pinkham Medicine Company y Lydia Pinkham Vegetable Compound.

Durante el final de los años ochenta y al comienzo de los años noventa del siglo pasado, se hicieron muchos esfuerzos por utilizar las redes para prever los datos de series de tiempo. La mayoría de este trabajo estaba enfocado en el campo de la utilización de energía porque las compañías de este sector necesitan pronósticos exactos de la demanda por hora de su producto. Sin embargo, la investigación se enfocó en las series de tiempo de negocios más tradicionales, como las series micro y macroeconómicas, datos demográficos y datos específicos de las compañías. Virtualmente, todo este trabajo utilizó una red con conexiones hacia adelante (*feed-forward*) entrenada mediante la propagación hacia atrás. Este caso empleará este tipo de red para prever los datos de ventas de Lydia Pinkham. Los pro-

nósticos resultantes se compararán con los pronósticos del modelo AR(2) presentado en el caso 9.4.

La figura 10.3 representa una red neural con conexiones hacia adelante de 2-4-1 —la red utilizada para este estudio—. El primer número en 2-4-1 indica la cantidad de entradas en la red. En este caso, las dos entradas son  $Y_{t-1}$  y  $Y_{t-2}$ . (El uso de los dos períodos anteriores para predecir el período actual es coherente con el modelo AR[2], de esta manera tanto el modelo AR [2] como el modelo de la red neural usan la misma “información” para calcular los pronósticos anticipados un paso.) El 4 indica el número de nodos, o unidades de procesamiento, en la capa oculta. (Se le llama *oculta* porque no está directamente conectada con el “mundo exterior”, como lo están las capas de entrada y salida.) El número de nodos en la capa oculta es en cierto modo escogido arbitrariamente: muy pocos nodos ocultos restringen la habilidad de la red de “ajustar” los datos, y demasiados nodos ocultos causan que la red memorice los datos de entrenamiento (o de estimación). La memorización causa una actuación muy mala sobre la muestra de prueba. En este caso, el número de nodos en la capa oculta es simplemente dos veces el número de entradas. Finalmente, el nodo de salida proporciona el pronóstico anticipado un paso, o  $\hat{Y}_t$ .

<sup>10</sup> Este caso fue proporcionado por Susan C. White, de la George Washington University, Washington, D.C. Para más información, consultar Susan C. White, “Predicting Time Series with Neural Networks versus Statistical Models: The Lydia Pinkham Data”, Actas de la 24 Conferencia Anual del Decision Sciences Institute, Región Suroeste, 1993, pp. 108-110.

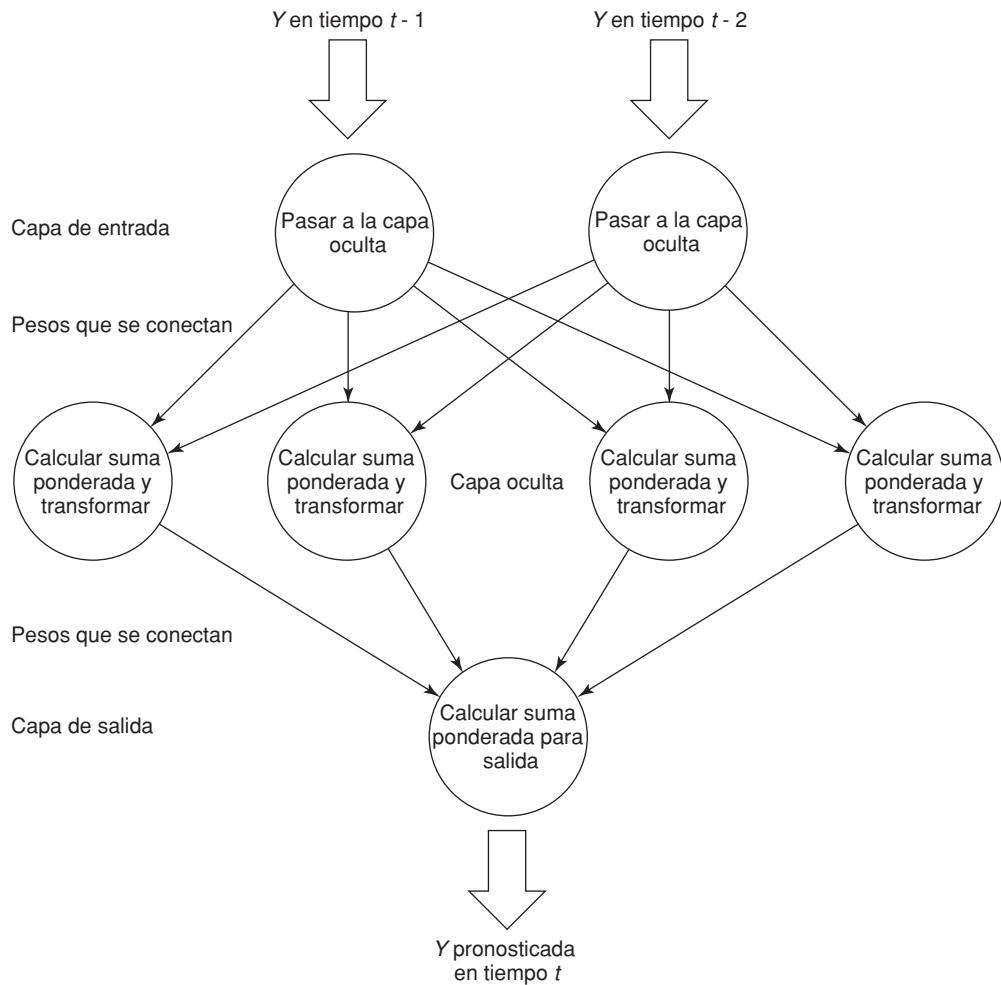


FIGURA 10.3 Red con conexiones hacia delante 2-4-1

La red neural calcula su salida de la siguiente manera: cada uno de los arcos que se conectan entre los nodos en dos capas adyacentes tienen un peso asociado. Cada nodo en la capa oculta calcula una suma ponderada de sus salidas. (Los nodos de la capa de entrada simplemente pasan a las entradas en la capa oculta). Entonces, esta suma ponderada se “transforma” de alguna manera, como la  $Y = 1/(1 + e^{-x})$ , donde  $Y$  es el dato “transformado” y  $X$  es la suma ponderada. La  $Y$  pasa entonces a la capa de salida donde cada nodo calcula de nuevo una suma ponderada. Esta suma ponderada final es la salida de la red. Ésta se entrena ajustando todos los pesos que se conectan de manera iterativa.

### EL MODELO DE LA RED NEURAL

La red neural se entrenó usando BrainMaker,<sup>11</sup> un paquete comercial de simulación de redes neurales en computadora. Para este estudio, el tamaño del paso fue de 0.500 y la tolerancia de entrenamiento de 0.001. (Otros paquetes podrían requerir que el usuario especifique una tasa de aprendizaje y un término de velocidad adquirida.) Como en el caso 9.4, las primeras 42 observaciones se usan para entrenar a la red; las últimas 12 se utilizan para evaluar la actuación de la red al realizar pronósticos anticipados un paso. Un problema al usar las redes neurales para pronosticar datos de series de tiempo es determinar cuánto tiempo se debe entrenar a la red; una red sobreentrenada tie-

<sup>11</sup> BrainMaker, California Scientific Software, 1990.

**TABLA 10.5 Resultados de la red neural para la Lydia E. Pinkham Medicine del caso 10.3**

| <b>Iteraciones</b> | <b>Datos de estimación</b> |             |            | <b>Datos de entrenamiento</b> |             |            |
|--------------------|----------------------------|-------------|------------|-------------------------------|-------------|------------|
|                    | <b>MAD</b>                 | <b>MAPE</b> | <b>MSE</b> | <b>MAD</b>                    | <b>MAPE</b> | <b>MSE</b> |
| 10,000             | 189.7                      | 11.06%      | 51923.0    | 182.2                         | 11.26%      | 43441.1    |
| 20,000             | 189.8                      | 10.80%      | 53926.3    | 169.6                         | 10.18%      | 37737.4    |
| 50,000             | 195.0                      | 11.22%      | 54311.3    | 176.3                         | 10.51%      | 41529.7    |
| 75,000             | 196.0                      | 11.40%      | 54424.2    | 178.1                         | 10.66%      | 42758.1    |
| 100,000            | 196.5                      | 11.53%      | 54725.5    | 181.9                         | 11.02%      | 43985.4    |

de a memorizar los datos de entrenamiento y actuar mal en los datos de prueba. Por lo tanto, algunos investigadores han sugerido simplemente detener el entrenamiento “antes” de que la red haya memorizado los datos. Para determinar el efecto del entrenamiento, se entrenó a la red para 10 000, 20 000, 50 000, 75 000 y 100 000 iteraciones. (Una iteración es la presentación de una observación; esto representa 250, 500, 1250, 1875 y 2500 pasos a través de la serie completa de entrenamiento, respectivamente). Esto le permite al analista evaluar el posible impacto del sobreentrenamiento. (El trabajo se realizó originalmente en una computadora personal con Intel 386SX-20 y el “tiempo” del entrenamiento para cien mil iteraciones fue de aproximadamente 20 minutos).

## RESULTADOS

En la tabla 10.5 se presentan los *MAD*, *MAPE* y *MSE* para los diferentes modelos de redes neurales. No se comparan favorablemente con el modelo AR(2). Actualmente, el autor de este caso está experimentando con un tipo diferente de redes neurales —una red neural con función de base radial— que produce resultados similares a los del modelo AR(2).<sup>12</sup> ■

## PREGUNTAS

1. Encuentre un artículo que describa una aplicación de redes neurales para pronosticar series de tiempo. ¿Qué método usaron los autores y qué tan exitoso fue?
2. Si usted tiene acceso a un programa de la red neural, intente encontrar una red que produzca mejores pronósticos que la red 2-4-1 presentada aquí.
3. ¿Por qué se consideran las redes neurales como una alternativa viable a los otros métodos de pronóstico discutidos en este texto?

## Referencias

- Armstrong, J. S., “Combining Forecasts: The End of the Beginning or the Beginning of the End?”, *International Journal of Forecasting* 5(4), 1989:585-592.
- Balkin, S. D. y J. K. Ord, “Automatic Neural Network Modeling for Univariate Time Series”, *International Journal of Forecasting* 16(4), 2000: 509-515.
- Banks, J., *Handbook of Simulation*, Nueva York, John Wiley & Sons, 1998.
- Bopp, A. E., “On Combining Forecasts: Some Extensions and Results”, *Management Science*, 31(12), diciembre de 1986: 1492-1497.
- Bunn, D. y G. Wright, “Interaction of Judgmental and Statistical Forecasting Methods: Issues and Analysis”, *Management Science*, 37(5), mayo de 1991: 501-516.

<sup>12</sup> Para leer más sobre las redes neurales, consulte Khanna (1990).

- Clemen, R. T., "Combining Forecasts: A Review and Annotated Bibliography", *International Journal of Forecasting*, 5(4), 1989:559-583.
- Denton, J. W., "How Good Are Neural Networks for Causal Forecasting?", *Journal of Business Forecasting*, 14(2), verano de 1995: 17-20.
- Dull, R., "Delphi Forecasting: Market Research for The 1990's", *Marketing News*, agosto de 1988:17.
- Edmundson, R. H., "Decomposition: A Strategy for Judgmental Forecasting", *Journal of Business Forecasting*, 9, verano de 1990: 305-315.
- Georgoff, D. M. y R. G. Murdick, "Manager's Guide to Forecasting", *Harvard Business Review*, 1, enero-febrero de 1986:110-120.
- Hammond, J. S., R. L. Keeney y H. Raiffa, *Smart Choices: A Practical Guide to Making Better Decisions*, Boston, Harvard Business School Press, 1999.
- Jain, C. L., *A Managerial Guide to Judgmental Forecasting*, Nueva York, Graceway, 1987: 101.
- Jain, C. L., "Myths and Realities of Forecasting", *Journal of Business Forecasting*, otoño de 1990: 18-22.
- Khanna, T., *Foundations of Neural Networks*, Reading, MA, Addison-Wesley, 1990.
- Kudlow, L., "The Case for Market-Based Forecasting", *CATO Journal*, 12(1), primavera de 1992: 119.
- Law, A. M. y W. D. Kelton, *Simulation Modeling and Analysis*, tercera edición, Boston, McGraw-Hill, 2000.
- Mahoud, E., "Combining Forecasts: Some Managerial Issues", *International Journal of Forecasting*, 5(4), 1989: 599-600.
- Makridakis, S., "The Art and Science of Forecasting", *International Journal of Forecasting*, 2, 1986: 15-39.
- Newbold, P. y T. Bos, *Introductory Business & Economic Forecasting*, segunda edición, Cincinnati, OH, Suroeste, 1994.
- Parente, F. J. y J. K., Anderson-Parente, "Delphi Inquiry Systems", en *Judgmental Forecasting*, eds. G. Wright y P. Ayton, Nueva York, John Wiley & Sons, 1987: 129-156.
- PricewaterhouseCoopers Global Technology Centre, *Technology Forecast* (versión actual), Menlo Park, CA, PricewaterhouseCoopers Global Technology Centre, 2003.
- Rowe, G. y G. Wright, "The Delphi Technique as a Forecasting Tool: Issues and Analysis", *International Journal of Forecasting*, 15(4), 1999: 353-375.
- Seila, A. F., V. Ceric y P. Tadikamalla, *Applied Simulation Modeling*, Belmont, CA, Brooks/Cole, 2003.
- Walden, M., "How to Evaluate and Improve a Forecasting Process", *Journal of Business Forecasting*, 15(2), verano de 1996: 22.
- Willard, T., "Forecasting: A Key to Business Success", *Futurist*, julio-agosto de 1991: 33-34.



## CAPÍTULO

# 11

## ADMINISTRACIÓN DEL PROCESO DE PRONÓSTICOS

En el capítulo 1 se presentó el tema de la generación de pronósticos con sugerencias acerca de su utilidad dentro de una organización moderna. Las organizaciones modernas deben desarrollar pronósticos para tomar decisiones precisas cuando enfrentan la incertidumbre. Por lo general, este proceso de tomar decisiones adecuadas acerca del futuro incierto (generación de pronósticos) implica un proceso racional para extender los datos históricos y las experiencias hacia el futuro. Ahora es posible revisar las nociones básicas y los puntos importantes presentados en el capítulo 1 con el beneficio de numerosas técnicas para la generación de pronósticos que se han abarcado en los capítulos anteriores.

### EL PROCESO DE PRONÓSTICOS

---

El proceso de pronósticos consiste en dos fases distintas. Una se presenta en un nivel estratégico y la otra en un plano operativo. En la primera, las decisiones incluyen qué se debe decidir, cómo emplear los pronósticos y quién es el responsable de éstos. La fase operativa de la generación de pronósticos consiste en la recopilación de los datos, la realización propia de los pronósticos y la evaluación pertinente de los pronósticos obtenidos. El proceso de los pronósticos es como cualquier otro: si se queda sin supervisión ni evaluación, existe la probabilidad de que se salga de control por completo.

Un punto clave presentado en el capítulo 1 es la importancia de utilizar el buen juicio de la administración junto con técnicas cuantitativas que permitan el desarrollo de buenos pronósticos. Por supuesto, es importante contar con un buen criterio administrativo y ahora puede ilustrarse al emplear varias de las técnicas de pronóstico que se presentaron en los capítulos anteriores. Siempre es necesaria una mezcla razonable de las técnicas cuantitativas con el sentido común si los pronósticos han de ser precisos y comprensibles para los responsables de la toma de decisiones en una empresa.

El análisis de las series de tiempo (descomposición) es un buen ejemplo de una técnica que implica la necesidad de emplear un juicio confiable junto con un análisis de la historia. Por ejemplo, una compañía que usa un programa de análisis de la serie de tiempo con datos mensuales a fin de obtener un análisis histórico de la variación cíclica de la tendencia, una variación estacional y movimientos irregulares, podría recombinar estos tres factores para producir un pronóstico útil de las ventas unitarias. La reorganización apropiada de estos componentes implica un juicio considerable conforme a su conducta futura. Al utilizar el proceso de descomposición es importante decidir cuántos períodos históricos habrán de incluirse en el análisis.

Cuando se realiza el pronóstico, el análisis de regresión requiere de un análisis estadístico, además de la cordura. Si la regresión múltiple se lleva a cabo cuando se emplea la clasificación del desempeño laboral del empleado como la variable dependiente y la puntuación de la prueba de ingreso y la edad como variables explicativas, podría obtenerse un valor de  $R$  cuadrada de 60%. Además, los valores  $t$  para las variables explicativas podrían ser sig-

nificativos junto con el valor  $F$  de regresión. El pronosticador se ve tentado a medir ambas variables explicativas en cada solicitante de trabajo y usarlas para predecir su desempeño laboral. Sin embargo, es necesario considerar tres cuestiones adicionales. En primer lugar, ¿es suficiente la variación explicada de 60% para realizar pronósticos? Quizás un juicio intuitivo acerca del deseo por contratar a una persona sea un método superior, o tal vez se necesite mayor precisión en los pronósticos y se deban considerar otras variables. En segundo lugar, ¿puede considerarse que los futuros aspirantes a un puesto serán exactamente idénticos a los que se muestran en el estudio de regresión? Si de alguna forma sustantiva son distintos, es posible que el modelo de pronóstico no sea válido. Finalmente, ¿se justifica el costo del proceso de pronóstico en términos del beneficio recibido? Para la compañía, la prueba podría ser costosa, especialmente si se le compra a una agencia externa; además, debe justificarse por los beneficios del pronóstico.

La regresión con datos de una serie de tiempo es una práctica común de las organizaciones donde se lleva a cabo el rastreo de importantes medidas del desempeño cada semana, mes o trimestre. Debido a que la existencia de autocorrelaciones es un problema común en tales estudios, la comprensión de esta circunstancia y su solución son vitales si los resultados de tal análisis habrán de ser válidos para el proceso de la toma de decisiones. Desafortunadamente, tal entendimiento no ocurre con frecuencia y esto se ha vuelto un problema cada vez mayor debido a la aparición del software de análisis de regresiones de bajo costo.

Las técnicas de Box-Jenkins que se analizaron en el capítulo 9 ilustran un problema común en la realización de pronósticos, el cual se abordó en el capítulo 1. Con frecuencia, estos procedimientos son métodos superiores de pronóstico que producen errores menores en muchas situaciones complejas. Su desventaja consiste en que se requiere cierto grado de sofisticación por parte del usuario. Si el proceso que genera los pronósticos es totalmente misterioso para el responsable de las decisiones, la administración de la empresa podría desear los pronósticos a pesar de su precisión.

Los aspectos de corto, mediano y largo plazos de las técnicas de pronóstico, conforme se vinculan con los distintos niveles administrativos de una empresa, se pueden ilustrar con un análisis de la serie de tiempo y una realización tecnológica de los pronósticos. Los administradores de primera y segunda líneas de una empresa podrían interesarse en el análisis de la serie tiempo de las ventas unitarias mensuales con los datos recolectados en los últimos cuatro años. Al utilizar el criterio con respecto al futuro del componente del ciclo de tendencia de esta serie, las ventas pueden pronosticarse para el siguiente año fiscal y esta información podrá utilizarse para programar la producción mensual de la empresa. Los gerentes de rango medio podrán utilizar el mismo programa de las series de tiempo para analizar los datos de las ventas unitarias anuales durante los últimos ocho años y pronosticarlas para los cinco años siguientes. En este caso, cualquier componente de largo plazo podría ser ignorado en un intento por planear las necesidades de gastos de capital para la fábrica durante este periodo de cinco años. Asimismo, los gerentes de alto nivel podrían tomar parte en la realización tecnológica de pronósticos mediante el método Delphi, junto con la formulación de escenarios. Su propósito podría ser evaluar la posición actual de la empresa en el mercado y buscar los cambios tecnológicos y sociales que podrían amenazar su nicho de mercado durante los siguientes 20 años, o bien, ofrecerle oportunidades que no son evidentes en las actividades cotidianas.

Las técnicas para el análisis de datos que se presentan en este libro se resumen en la tabla 11.1, la cual proporciona descripciones, muestra aplicaciones, estima niveles de costos e indica si las capacidades de la computadora son necesarias para la aplicación de cada una de las técnicas. Cada una de estas técnicas también hace referencia a los capítulos donde se ha presentado. Los resúmenes, como los que están en la tabla 11.1, deberán observarse como lineamientos y no como declaraciones definitivas que no deben ser cuestionadas.

## MONITOREO DE LOS PRONÓSTICOS

---

La recolección de datos y la selección de una técnica aceptable de pronóstico son los primeros pasos que se siguen en un esfuerzo efectivo y constante para la generación de pronósticos. En esta obra se han descrito diversos pasos de la fase operativa del proceso de pronóstico,

**TABLA 11.1 Modelos de pronóstico**

| <b>Método</b>                                        | <b>Descripción</b>                                                                                                                                                | <b>Aplicaciones</b>                                                                                                                                                                                                                                                            | <b>Computadora</b> |                    |                 |
|------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|--------------------|-----------------|
|                                                      |                                                                                                                                                                   |                                                                                                                                                                                                                                                                                | <b>Costo</b>       | <i>¿Necesario?</i> | <b>Capítulo</b> |
| <b>Modelos causales de pronóstico</b>                |                                                                                                                                                                   |                                                                                                                                                                                                                                                                                |                    |                    |                 |
| Análisis de regresión                                | Pronóstico explicativo: supone una relación de causa y efecto entre la información introducida en el sistema y la información que resulta de éste.                | Pronósticos de corto y mediano plazos de los productos y servicios existentes; estrategias de marketing, producción, contratación de personal y planeación de instalaciones.                                                                                                   | De bajo a medio    | Usualmente         | 6               |
| Regresión múltiple                                   | Pronóstico explicativo: supone una relación de causa y efecto entre más de una información introducida en el sistema y la información que resulta de éste.        | Igual que las anteriores                                                                                                                                                                                                                                                       | De bajo a medio    | Sí                 | 7               |
| <b>Modelos de pronóstico de las series de tiempo</b> |                                                                                                                                                                   |                                                                                                                                                                                                                                                                                |                    |                    |                 |
| Método de descomposición                             | Pronóstico explicativo: supone una relación de causa y efecto entre el tiempo y la información resultante del sistema; el sistema se descompone en sus elementos. | Pronóstico de mediano plazo para una planta nueva y planeación del equipo, financiamiento, desarrollo de productos nuevos y métodos novedosos de ensamblaje; pronóstico de corto plazo para el personal, publicidad, inventario, financiamiento y planeación de la producción. | De bajo a medio    | Sí                 | 5               |
| Promedios móviles                                    | Para eliminar el factor aleatorio en la serie de tiempo, pronóstico basado en los datos de la serie de tiempo suavizados mediante un promedio móvil.              | Pronósticos de corto plazo para operaciones como inventario, programación, control, fijación de precios y coordinación de promociones especiales; se usa para calcular los componentes tanto cíclicos como estacionales del método de descomposición a corto plazo.            | Bajo               | No                 | 4 y 5           |

**TABLA 11.1** *(Continuación)*

| <b>Método</b>             | <b>Descripción</b>                                                                                                                                                                                    | <b>Aplicaciones</b>                                                                                                                                         | <b>Computadora</b> |                    |                 |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|--------------------|-----------------|
|                           |                                                                                                                                                                                                       |                                                                                                                                                             | <b>Costo</b>       | <b>¿Necesario?</b> | <b>Capítulo</b> |
| Suavizamiento exponencial | Similar a los promedios móviles pero con valores ponderados exponencialmente al dar mayor peso a las observaciones recientes.                                                                         | Pronósticos de corto plazo para operaciones como inventario, programación, control, fijación de precios y promociones especiales de temporada.              | Bajo               | Sí                 | 4               |
| Modelos autorregresivos   | Empleado con variables económicas para considerar las relaciones entre las observaciones adyacentes en una serie de tiempo.                                                                           | Pronósticos de corto y mediano plazos para los datos económicos ordenados en una serie de tiempo: precios, inventario, producción, almacenamiento y ventas. | Medio              | Sí                 | 8 y 9           |
| Técnicas Box-Jenkins      | No supone un patrón especial en los datos históricos de la serie que se pronosticará; utiliza un método iterativo para identificar y ajustar un modelo útil a partir de una clase general de modelos. | Igual que las anteriores                                                                                                                                    | Elevado            | Sí                 | 9               |
| Redes neutrales           | Utiliza equipo sofisticado de cómputo para asimilar los datos relevantes y reconocer los patrones por medio del “aprendizaje”, como los hacen las personas.                                           | Su uso está en aumento en una variedad de aplicaciones de pronóstico; está en fase de desarrollo.                                                           | Elevado            | Sí                 | 10              |

con un énfasis especial en aprender las técnicas que comúnmente se usan para generar los pronósticos reales. En la figura 11.1 se resumen los pasos operacionales básicos para el pronóstico.

La recolección y examen de los datos históricos apropiados se han descrito con anterioridad en este libro (capítulo 3), junto con las consideraciones para seleccionar una técnica o modelo para el pronóstico. Como se sugiere en la figura 11.1, por lo general el siguiente paso es pronosticar diversos períodos históricos en donde se conocen los valores reales. Los errores que resultan de esto pueden resolverse de diversas maneras, como se detalló en el capítulo 3, y este proceso continúa hasta que se encuentre una técnica con una razón adecuada de costo-beneficio. Entonces, el modelo se utiliza para pronosticar períodos futuros, de manera que los resultados obtenidos se incorporan al proceso de la toma de decisiones de la empresa.

De vez en cuando es necesario hacer una pausa en el proceso de pronóstico y reconsiderar los procedimientos que se están utilizando. Los pasos usuales son los que se indican a continuación:

1. Se desechan los valores históricos más antiguos de los datos que está utilizando la técnica de pronóstico y se añaden al banco de datos los valores reales más recientes.

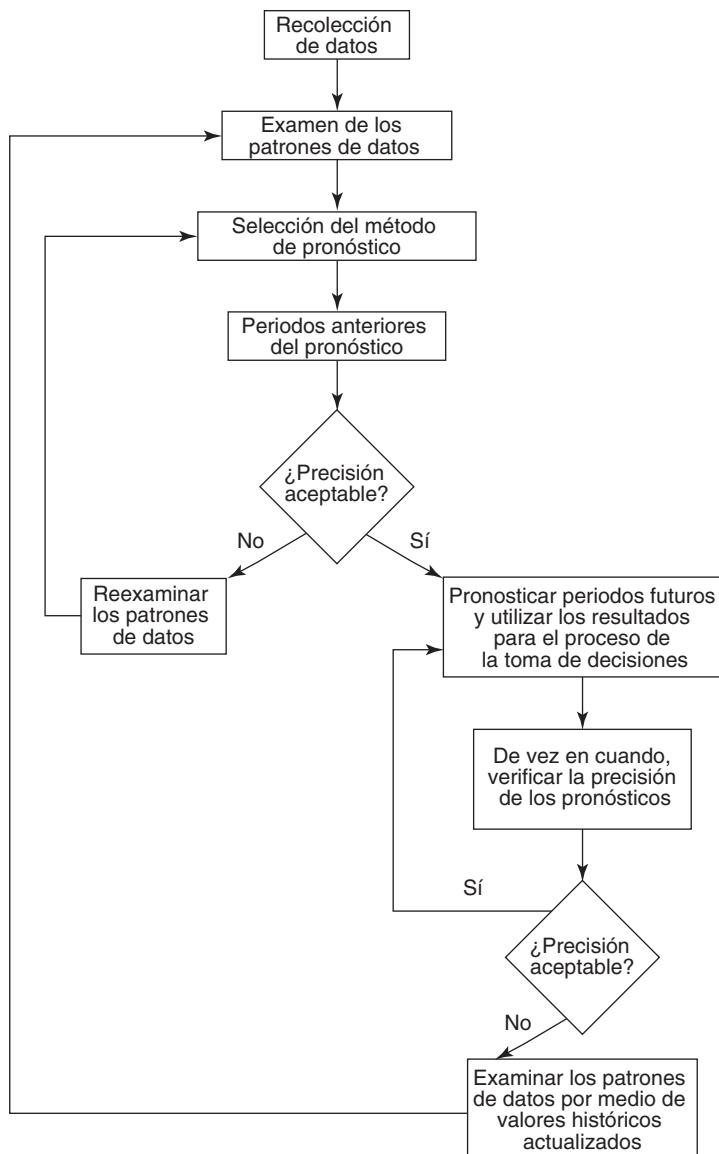


FIGURA 11.1 Fase operacional del proceso de pronóstico

2. Conforme se van actualizando los datos, se vuelven a calcular los parámetros utilizados en el modelo de pronóstico. Por ejemplo, los valores óptimos de las constantes de peso utilizadas en el suavizamiento exponencial podrían cambiar, quizás de manera considerable, cuando se agreguen valores de datos más recientes. Asimismo, los coeficientes en un análisis de regresión pueden cambiar cuando distintos valores de datos se ajusten con una ecuación determinada.
3. El modelo de pronóstico con nuevos parámetros se examina para una precisión adecuada. Si se considera que esta exactitud es suficiente, entonces el modelo se utiliza como antes hasta que se presente el siguiente periodo de actualización. Si la precisión de los pronósticos se considera inadecuada o marginal, se pueden examinar los patrones en los nuevos datos con la posibilidad de seleccionar un nuevo procedimiento de pronóstico. Este proceso continúa hasta que se considere que la exactitud del modelo elegido es la adecuada, de acuerdo con la certeza de los periodos del pronóstico donde los valores reales son conocidos.

El proceso anterior se resume en el diagrama de flujo de la figura 11.1 y constituye el tipo de retroalimentación que habitualmente se encuentra en los diseños de sistemas de todos los tipos.

Algunas veces, los pronósticos se monitorean constantemente utilizando señales de rastreo, concepto que se analizó en el capítulo 4, después del material relacionado con el suavizado exponencial simple. La idea es establecer límites dentro de los cuales se espera que los errores generados por los pronósticos fracasen, si el proceso de pronóstico es adecuado. Mientras los errores estén dentro de los límites aceptables, el proceso de pronóstico deberá continuar. Tan pronto como el error queda fuera del rango aceptable, la administración se concentra en el proceso de pronóstico y se realizan los pasos de actualización y revisión que ya se mencionaron en este libro. Este concepto se ilustra en el ejemplo 11.1.

### Ejemplo 11.1

Sue Bradley es responsable de los pronósticos de la venta mensual en dólares de su compañía. Sue ha seleccionado un modelo de pronóstico que tiene una tasa de error aceptable para sus gerentes. Específicamente, el error estándar de este proceso de pronóstico es de \$935; es decir, la diferencia entre los valores pronosticados y los reales en una venta mensual típicamente es de \$935.

Sue supone que los errores de pronóstico se distribuyen normalmente con una media de cero y una desviación estándar de \$935. Este supuesto lo realiza después de haber examinado la gráfica de los errores pronosticados al encontrar que siguen una curva en forma de campana en torno al cero. Al usar un nivel de confianza de 95%, establece los siguientes límites dentro de los cuales espera que caiga el error pronosticado de cada mes:

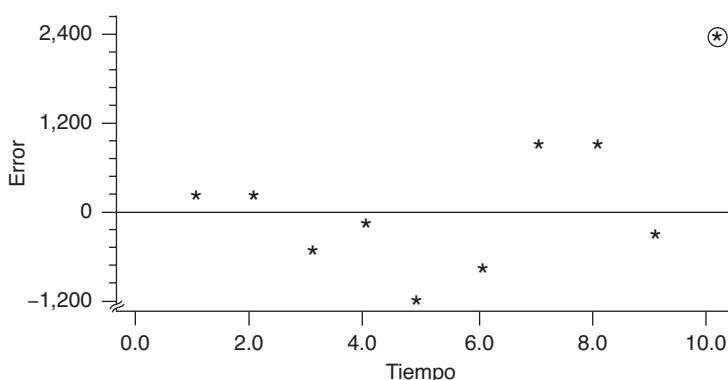
$$0 \pm (1.96)(935)$$

$$0 \pm 1,833$$

De esta manera, Sue espera que los pronósticos mensuales estén a una distancia menor que \$1833 del valor real para el mes, en el nivel de confianza de 95%. Si éste fuera el caso, el procedimiento para la generación de pronósticos continuará sin que le preste atención. Pero si el error fuera mayor a \$1833, ella examinará tanto los parámetros como técnica de pronóstico y podría considerar la aplicación de otra técnica.

Para monitorear los errores de pronóstico de manera más sencilla, Sue diseña una gráfica para rastrearlos. Con el paso de los meses, encuentra dos gráficas que le indican que debe examinar más de cerca sus procedimientos de pronóstico. El primero, que se muestra en la figura 11.2, revela errores de pronóstico que parecen distribuirse de manera aleatoria hasta el periodo más reciente. Este gran error lleva a Sue a estimar nuevamente los parámetros de su modelo de pronóstico después de actualizar su base de datos al añadir valores más recientes y desechar el mismo número de valores más antiguos. Más adelante, una segunda gráfica de error, que se presenta en la figura 11.3, ocasiona que Sue revise de nuevo su proceso de pronóstico. A pesar de que ninguno de los errores ha excedido sus límites establecidos, Sue observa que en apariencia los errores no están distribuidos aleatoriamente. De hecho, los errores crecen en una dirección positiva y es obvio que pronto estarán fuera de control. Sue actualiza su base de datos y, después de examinar con todo cuidado el patrón de datos, selecciona una nueva técnica de pronóstico.

**FIGURA 11.2 Primera gráfica de errores de los primeros datos para el ejemplo 11.1**



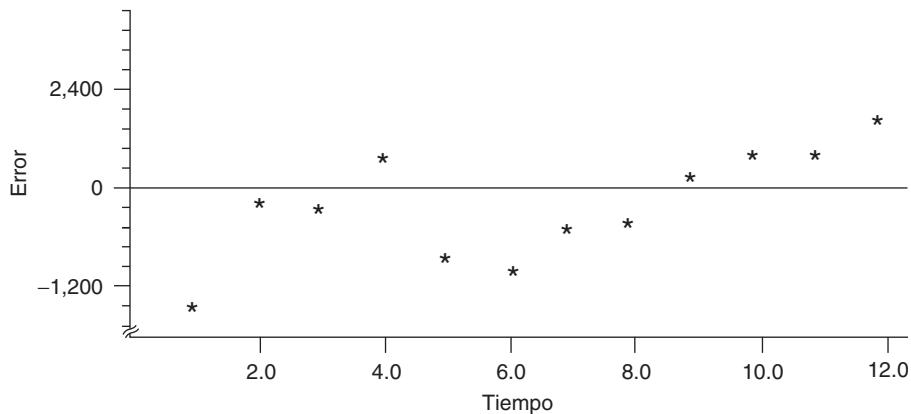


FIGURA 11.3 Segunda gráfica de errores de los datos para el ejemplo 11.1

## REVISIÓN DE LAS ETAPAS DEL PRONÓSTICO

En el capítulo 1 se enfatizó el punto de que el buen juicio deberá ejercitarse constantemente junto con el análisis cuantitativo, para obtener pronósticos útiles y precisos. Se deben plantear diversas preguntas que ya se citaron si se quiere que la administración del proceso de pronóstico tenga una adecuada dirección. Aquí se repiten:

- ¿Por qué se necesita un pronóstico?
- ¿Quién utilizará el pronóstico y cuáles son sus requerimientos específicos?
- ¿Qué nivel de detalle o de acreditación se requiere y cuál es el horizonte adecuado de tiempo?
- ¿Cuáles datos están disponibles?, y ¿éstos serán suficientes para generar el pronóstico necesario?
- ¿Cuál es el costo del pronóstico?
- ¿Qué nivel de precisión puede esperarse del pronóstico?
- ¿El analista entiende con claridad cómo se utilizará el pronóstico en la organización?
- ¿Existe un proceso de retroalimentación para evaluar el pronóstico después de que éste se realice y para ajustar el proceso conforme sea necesario?

Cuando estas preguntas hayan sido dirigidas y contestadas adecuadamente, puede comenzar el proceso real de la generación de pronósticos. El capítulo 1 presenta los pasos del proceso a seguir:

- Recolección de datos
- Reducción o condensación de datos
- Construcción y evaluación del modelo
- Extrapolación del modelo (el pronóstico real)
- Evaluación del pronóstico

A esta lista de pasos podría añadirse otro: Retroalimentación después de que el proceso de pronóstico se pone en marcha para determinar si se ha obtenido la precisión suficiente, así como saber si la administración encuentra útil y efectivo el pronóstico en términos de costos para la toma de decisiones.

A pesar de que la principal preocupación de este libro ha sido la construcción de un modelo o la selección de un método de pronóstico adecuado, los autores esperan que se

haya enfatizado la importancia de la administración del proceso de pronóstico. Las preguntas que se enumeraron anteriormente son importantes en todas las situaciones de pronóstico y deberán dirigirse si se desean resultados útiles. De manera particular, se recomienda el estudio de los casos que se presentan al final de este capítulo. Están diseñados para enfatizar los aspectos gerenciales de la generación de pronósticos.

## RESPONSABILIDAD DEL PROCESO DE PRONÓSTICO

---

La ubicación que tenga el proceso de pronóstico dentro de una empresa varía de acuerdo con el tamaño de ésta, la importancia que se le atribuya a la realización formal de los pronósticos y la naturaleza del estilo de administración del negocio. La responsabilidad del pronóstico recae en algún lugar del continuo, entre un departamento de pronóstico separado y los pronósticos que se efectúan dentro de pequeñas unidades administrativas que no tienen referencia de los demás esfuerzos que se realizan en la empresa.

El personal especializado en la generación de pronósticos es más común en las organizaciones grandes que en las pequeñas. Las empresas grandes pueden permitirse la contratación de expertos para la realización sofisticada de los pronósticos y pueden equipar a su personal con software y herramientas modernas de cómputo. La ventaja de tal esfuerzo centralizado es que esa capacidad está al alcance de todas las unidades de la organización. La desventaja es que, con frecuencia, la coordinación entre el personal encargado de los pronósticos y los gerentes de línea es difícil de lograr. El equipo especializado podría perder más tiempo en negociaciones con los usuarios y explicándoles su función que en realizar pronósticos reales.

En el otro extremo se encuentra la ubicación del proceso de pronóstico dentro de cada unidad de la empresa sin coordinación o cooperación entre ellas. La ventaja de este proceso es que no hay confusión entre quienes hacen los pronósticos y quienes los utilizan. Son las mismas personas. Los pronósticos obtenidos bajo estas condiciones tienden a ser aceptados y usados en la toma de decisiones. La desventaja es que la sofisticación y, por lo tanto, la precisión de los pronósticos son difíciles de alcanzar debido a que la disponibilidad de las herramientas de cómputo y la destreza tecnológica podrían ser desiguales en muchos usuarios. Por lo general, es difícil convencer a los directivos de que adquieran hardware y software apropiados; por ejemplo, cuando saben que ya existen en otras áreas de la empresa.

Además, la falta de coordinación y cooperación puede acarrear problemas cuando una unidad utiliza los pronósticos generados por otro departamento dentro de la compañía. Si los departamentos no son responsables de la calidad de sus propios pronósticos, se pueden dar desviaciones que conducirán a dificultades operativas.

En cierta organización, el departamento de marketing ha sido responsable del proceso de pronóstico. A este departamento se le evalúa por la existencia del producto, medido por el porcentaje del tiempo que un artículo ha estado disponible para que el consumidor lo adquiera. Sin embargo, el departamento de marketing no ha sido sancionado por excesos en el inventario que pudieran generarse como resultado de que los pronósticos influyen sobre la producción. A pesar de que esos pronósticos se basaron en condiciones de mercado, también tuvieron fuerte influencia de las expectativas de ingresos y cuotas de mercado. En consecuencia, el grupo de marketing generó de manera consistente pronósticos optimistas de demanda para asegurar la existencia del producto, la cual permitiría que se ganaran utilidades. Por otro lado, el departamento de planeación de la producción ha sido evaluado por la disponibilidad del producto y el excedente del inventario. Este grupo utilizó métodos estadísticos para neutralizar cualquier incertidumbre y estuvo muy motivado para mantener sus pronósticos precisos y sin desviaciones.

Muchas organizaciones intentan ubicar la responsabilidad de los pronósticos en un punto medio entre los extremos que se acaban de mencionar. Es posible incluir un pequeño grupo de pronosticadores en subunidades dentro de la empresa con el propósito de satisfacer las necesidades de las áreas funcionales. Su labor involucra la coordinación adecuada

con los clientes, así como la generación de pronósticos precisos. Algunas veces, esta responsabilidad de generar pronósticos se combina con otras funciones del equipo especializado, como el apoyo estadístico y de cómputo.

El advenimiento de las computadoras pequeñas de bajo costo y del software para elaborar pronósticos ha restringido la función de los pronósticos dentro de la organización. Ahora es posible que los gerentes adquieran herramientas sofisticadas para generar pronósticos a un precio menor de lo que costaban hasta hace pocos años. No obstante, el conocimiento que se requiere para utilizar adecuadamente estas herramientas no se incluye en los programas de software y hardware; la necesidad de comprender el uso adecuado de las técnicas de pronóstico se ha incrementado conforme la capacidad cibernetica ha pasado del dominio de los “expertos” al de los usuarios dentro de la organización.

## COSTOS DEL PRONÓSTICO

---

El hardware y el software, más el personal especializado, son los costos evidentes de la generación de pronósticos. Pero también existen costos adicionales que no podrían ser tan obvios debido a los gastos de la compañía en el tiempo del personal, así como en el dinero. El tiempo que consumen los empleados —de los distintos departamentos— en la recopilación de datos para el pronóstico, así como en el monitoreo del proceso y la interpretación de los resultados debe considerarse como un costo de los pronósticos. Tal costo debe balancearse con los beneficios recibidos si se han de alcanzar las decisiones racionales respecto de la utilidad de los pronósticos resultantes.

Una alternativa para producir pronósticos de forma interna es emplear consultores para este propósito. Esta práctica es especialmente atractiva si la necesidad de un pronóstico involucra un requerimiento de una sola ocasión, en lugar de uno que sea constante. Además, la necesidad de pronósticos más allá de la capacidad técnica del personal de la compañía sugiere que se empleen consultores profesionales. Tal contratación de ayuda externa para realizar pronósticos permite que la identificación de los costos sea algo sencillo.

## EL PRONÓSTICO Y EL SISTEMA MIS

---

En los últimos años, los sistemas de información gerencial (MIS, por sus siglas en inglés) de las empresas modernas han incrementado su sofisticación y utilización. Su principal beneficio para el proceso de pronóstico se relaciona con su enorme capacidad para recopilar y registrar datos a lo largo de la organización. El pronosticador deberá resistir la tentación de recopilar datos para los modelos de pronóstico que están en uso, si tales datos ya han sido recopilados y registrados en el sistema MIS de la empresa.

En muchas empresas, el rápido desarrollo de las redes de computadoras ha incrementado el acceso a los datos por parte de los usuarios de la red. Además, los programas integradores de software, como el SAP,<sup>1</sup> tienen módulos para pronóstico que hacen más accesible el uso de las bases de datos con el programa.

La confianza en las bases de datos existentes es importante, incluso si los datos disponibles no están precisamente en el formato o en la secuencia de tiempo deseados por el pronosticador. Se deben considerar las modificaciones al modelo de pronóstico o a los datos disponibles antes de abandonar los datos MIS en favor de una nueva recopilación de datos. Desde luego, este consejo supone que la recopilación de datos nuevos involucraría un gasto considerable de tiempo y dinero. Si se obtuvieran con facilidad los datos necesarios para un modelo de pronóstico en un formato correcto, estos datos deberían preferirse sobre los recopilados con anterioridad, los cuales no están en un formato adecuado o ya son obsoletos.

---

<sup>1</sup> SAP, cuyas oficinas centrales están en Walldorf, Alemania, actualmente es la empresa más grande de software interno para las empresas.

Una ventaja adicional que tiene el uso de los datos del MIS o del sistema de redes de la empresa, consiste en que el proceso de pronóstico se vuelve un componente de este sistema. Como tal, se introduce a la distribución y a la red para la toma de decisiones que ya se han establecido dentro del sistema y puede incorporarse más fácilmente a la toma de decisiones de la empresa. Esta situación contrasta con el procedimiento de pronóstico que intenta infiltrar los métodos de la toma de decisiones que ya utilizan los gerentes de la compañía.

## PRONÓSTICOS Y ADMINISTRACIÓN DE LAS VENTAS

---

Hay diversos factores que son importantes si los pronósticos generados dentro de una compañía han de convertirse en aspectos relevantes dentro del proceso de la toma de decisiones. Primero, debe reconocerse que los administradores eficientes están interesados en resultados útiles y prácticos. En general, los pronósticos deben satisfacer las necesidades de esos gerentes; asimismo, deben responder a una serie de preguntas que ya se formularon en este capítulo.

El segundo lugar, para ser útiles, los pronósticos requieren de cierto nivel de precisión. Los buenos administradores no adoptarán un proceso de pronóstico por mucho tiempo, a pesar de su sofisticación, si no se obtienen resultados precisos.

Finalmente, hay que reconocer las intuiciones de los buenos gerentes acerca del costo-beneficio en el proceso de pronóstico. La habilidad para analizar situaciones en términos de los costos pagados, en comparación con los beneficios recibidos, es la piedra angular de un proceso eficaz de administración, y esto lo deberá reconocer el personal encargado de efectuar los pronósticos. Con frecuencia, esta situación crea dificultades entre el pronosticador y el usuario. El pronosticador debe tener en mente que el resultado final del proceso de pronóstico es mejorar la eficiencia del negocio, aún cuando las utilidades están en los niveles más bajos.

## EL PRONÓSTICO Y SU FUTURO

---

Cómo se mencionó en el capítulo 1, los pronósticos siguen ganando importancia en las organizaciones modernas gracias a la progresiva complejidad del mundo empresarial junto con la existencia de equipo de cómputo más poderoso y barato. La continua competencia en las áreas de las computadoras pequeñas y el software es obvia para todos; pero menos evidentes son las tendencias de largo plazo que lentamente cambian la escena empresarial y que ejercen presiones sutiles, aunque poderosas, en las formas en que operan los empresarios.

Considere algunas de las “megatendencias” identificadas por Naisbitt en su libro con ese título (vea las referencias adicionales de este tipo en el capítulo 10). Estas fuerzas tienen una importancia fundamental en el mundo de los negocios y traen a la mente la necesidad de generar pronósticos mediante equipos y software modernos:

- De la industria a la sociedad de la información
- Del plano nacional a la economía mundial
- Del corto al largo plazos
- De la centralización a la descentralización
- De cualquiera a múltiples opciones

La economía actual es de una competencia feroz a escala mundial. Debido a la Internet, una gran cantidad de información está disponible en cualquier momento del día. Los cambios en el mercado son rápidos e imprevistos, y quizás permanentes. Asimismo, se reducen los ciclos de vida de los productos. Llevar a cabo un negocio de manera electrónica (*e-business*) es una forma de vida y las transacciones mediante Internet seguirán creciendo.

Estas tendencias subrayan la importancia de un desarrollo gerencial continuo por parte de los empresarios, quienes deben tratar los asuntos complejos que enfrentarán sus empresas en los años siguientes. De manera particular, enfatizan la necesidad de desarrollar métodos cada vez más sofisticados para tratar con las incertidumbres de los sucesos futuros. En otras palabras, destacan la creciente necesidad de combinar el buen juicio con los métodos avanzados de manipulación de datos en un pronóstico de negocios correcto. Conforme las tendencias de Naisbitt y la dinámica progresiva del medio empresarial continúan desarrollándose, la habilidad de los líderes empresariales para reaccionar de manera rápida y redituable a los sucesos cambiantes siempre será un enfoque importante. La pregunta básica en los negocios “¿qué sucederá a continuación?”, será de gran importancia; esta interrogante es la que concierne al pronóstico en los negocios. A los analistas de pronósticos les esperan nuevos retos y oportunidades.

## Problemas

1. Escriba una respuesta corta para las siguientes afirmaciones:
  - a) Los pronosticadores siempre están mal, entonces, ¿por qué enfatizar la planeación de la demanda?
  - b) Un buen proceso de pronóstico es muy caro.
  - c) Lo que necesitamos hacer es contratar a un “tipo cuantitativo” para que haga nuestros pronósticos.
2. ¿Puede usted pensar en cualquier tendencia empresarial que no se haya presentado en este capítulo y que tenga la posibilidad de influir en la manera en que se efectúen los pronósticos en el futuro? Explique su respuesta.

## CASOS

---

### CASO 11-1 BOUNDARY ELECTRONICS

Boundary Electronics es un gran proveedor de productos electrónicos para uso en el lugar. Entre sus productos con mayores ventas están las videocámaras y los sistemas de televisión satelital. Debido al crecimiento acelerado de la compañía, Guy Preston, presidente de Boundary, está preocupado porque su patrón de ventas podría verse alterado si ocurre algún cambio en las condiciones del mercado.

Al consultar con sus gerentes acerca del futuro de la empresa, Guy ha descubierto dos cosas. Primero, la mayor parte de sus gerentes están demasiado ocupados con los problemas habituales de satisfacer la demanda creciente como para darle mucha atención al futuro de largo plazo. En segundo lugar, sus opiniones varían de forma considerable, pues van de muy optimistas a demasiado pesimistas. Como presidente de la firma, Guy piensa que tiene la obligación de considerar seriamente el ambiente futuro de ésta.

Después de reflexionar en este asunto, Guy planea un retiro sabático para los seis miembros de su equipo directivo. Entonces, renta una sala de juntas en un hotel local y hace arreglos para el almuerzo y la hora del café. Cuando el equipo se reúne el sábado por la mañana, Guy hace una exposición sobre el tema del día y después instruye a cada persona para que prepare una descripción, de una o dos páginas, del ambiente opera-

tivo de la compañía con la perspectiva de los siguientes 20 años, para cada una de las siguientes situaciones:

1. Esencialmente, el ambiente de la compañía continuará siendo como es ahora. Los productos que el mercado demande serán modificaciones de los actuales, y no intervendrá ninguna tecnología novedosa.
2. La actual línea de productos será obsoleta debido a los cambios tecnológicos más importantes. Para satisfacer las demandas que genera el tiempo libre de la población, habrán de crearse nuevos productos.
3. Entre estos dos extremos, ¿cuál es el escenario más probable para el ambiente operativo de la empresa?

Los integrantes del equipo disponen de una hora para desarrollar los escenarios de cada situación. Durante ese tiempo, Guy medita sobre lo que ocurrirá el resto del día. En su cavilación, espera que los gerentes desarrollen algunas ideas provocativas y que las discusiones subsecuentes demuestren ser animadas e interesantes. Además de recibir ideas para su uso personal, Guy desea que el ejercicio del día ayude a sus gerentes a ver más allá de los problemas y oportunidades de la empresa; también espera que logren un enfoque de largo plazo con respecto a la compañía. ■

## PREGUNTAS

1. ¿Qué proceso piensa usted que debería utilizar Guy una vez que la hora de las actividades de escritura haya terminado?
2. ¿Existe otro método que pudiera utilizar Guy para cumplir con sus objetivos?
3. ¿Cree usted que Guy alcanzará sus metas con la junta del sábado?

## CASO 11-2 BUSBY ASSOCIATES

Jill Tilson apenas se había graduado en una escuela de negocios universitaria cuando obtuvo un empleo en Busby Associates, un gran exportador de equipo agrícola. Durante el proceso de contratación, el presidente de Busby encontró en el currículum de Jill un curso de pronósticos, por lo cual determinó que Jill comenzara su trabajo en

un proyecto de pronósticos que había sido analizado muchas veces por el equipo directivo de la empresa.

El presidente de Busby creyó que había una fuerte correspondencia entre las ventas de exportación de la compañía y las cifras de las exportaciones nacionales. Éstas podían obtenerse de fuentes gubernamentales, de

**TABLA 11.2 Datos trimestrales de la serie de tiempo para el caso 11.2**

| <i>Periodo</i> | <i>Variable</i> |          |                    |          |          |          |          |
|----------------|-----------------|----------|--------------------|----------|----------|----------|----------|
|                | <i>1</i>        | <i>2</i> | <i>3</i>           | <i>4</i> | <i>5</i> | <i>6</i> | <i>7</i> |
| 1987           |                 |          |                    |          |          |          |          |
| 1              | 18.2            | 128.3    | 306.2              | 110.0    | —        | —        | —        |
| 2              | 19.8            | 45.8     | 311.6              | 109.7    | 1.6      | 8.79     | 18.2     |
| 3              | 20.9            | 66.1     | 320.7              | 109.9    | 1.1      | 5.56     | 19.8     |
| 4              | 22.1            | 129.7    | 324.2              | 109.7    | 1.2      | 5.74     | 20.9     |
| 1988           |                 |          |                    |          |          |          |          |
| 1              | 24.0            | 136.4    | 331.0              | 109.4    | 1.9      | 8.60     | 22.1     |
| 2              | 26.0            | 140.7    | 337.3              | 110.5    | 2.0      | 8.33     | 24.0     |
| 3              | 27.7            | 156.9    | 342.6              | 110.6    | 1.7      | 6.54     | 26.0     |
| 4              | 29.7            | 148.5    | 352.6              | 110.9    | 2.0      | 7.22     | 27.7     |
| 1989           |                 |          |                    |          |          |          |          |
| 1              | 33.6            | 189.8    | 351.5              | 113.4    | 3.9      | 13.13    | 29.7     |
| 2              | 35.0            | 168.9    | 357.6              | 112.4    | 1.4      | 4.17     | 33.6     |
| 3              | 35.0            | 154.5    | 365.2              | 111.9    | 0.0      | 0.00     | 35.0     |
| 4              | 38.0            | 174.1    | 366.3              | 111.0    | 3.0      | 8.57     | 35.0     |
| 1990           |                 |          |                    |          |          |          |          |
| 1              | 40.7            | 191.3    | 369.1              | 111.9    | 2.7      | 7.11     | 38.0     |
| 2              | 42.0            | 201.2    | 370.0 <sup>a</sup> | 112.1    | 1.3      | 3.19     | 40.7     |

<sup>a</sup> Estimado

Variables fundamentales:

1: Bienes de consumo, exportaciones, miles de millones de dólares

2: Ahorros personales brutos, miles de millones de dólares

3: Ingresos nacionales por comercio al mayoreo, miles de millones de dólares

4: Índices de precios ponderados en dólares constantes para las compras de defensa nacional, equipo militar, 1982 = 100

5: Cambio en la variable dependiente del periodo anterior

6: Cambio porcentual en la variable dependiente del periodo anterior

7: Variable dependiente retrasada un periodo

Fuente de las variables de la 1 a la 4: Survey of Current Business, 70, núm. 7 (julio de 1990), U.S. Department of Commerce

manera que el proyecto de Jill consistía en pronosticar una variable de exportación representativa. Si este esfuerzo fuera exitoso, el presidente de Busby y su empresa contarían con una herramienta poderosa para efectuar pronósticos sobre sus ventas de exportación.

Jill encontró la edición más reciente de *Survey of Current Business* en una biblioteca local y registró las cifras trimestrales de las exportaciones de bienes de consumo en miles de millones de dólares. Ella creyó que la variable representaba correctamente el total de las exportaciones nacionales. En anticipación a la posibilidad de realizar pronósticos mediante un análisis de regresión, también registró los valores de otras variables que podrían correlacionarse adecuadamente con esta variable dependiente. De este modo, terminó con valores para cuatro variables que correspondían a 14 trimestres.

Después calculó tres variables adicionales a partir de la variable dependiente: cambio en  $Y$ , cambio porcentual en  $Y$  y esta  $Y$  retrasada un periodo. Así que cuando comenzó a reflexionar acerca de las diversas formas de pronosticar su variable, había recopilado los datos que se muestran en la tabla 11.2.

Jill introdujo sus datos en un paquete de cómputo que efectuaba análisis de regresión y calculaba la matriz de correlación para sus siete variables. Después de examinar esta matriz, seleccionó tres regresiones con una sola variable explicativa y seis regresiones con dos variables explicativas. Después corrió estas regresiones y seleccionó la que consideró mejor: una regresión simple que utilizaba como variable explicativa al retraso de  $Y$  en un periodo, con los siguientes resultados:

$$r\text{-cuadrada} = .98$$

$$t = 25.9$$

$$F = 671.6$$

$$\text{Durbin-Watson} = 2.98$$

La tabla de Durbin-Watson que Jill empleó comenzaba con  $n = 15$  y el tamaño de la muestra era únicamente 13. De esta forma, interpoló los valores relevantes  $DW$  y estimó en 1.35 el límite superior para la prueba de autocorrelación. Debido a que su estadística  $DW$  correspondía a 2.18, concluyó que no había autocorrelación en su regresión. Creyó que había encontrado una buena variable explicativa ( $Y$  retrasada un periodo).

Jill descubrió que su muestra era muy pequeña: 13 trimestres. Entonces, regresó a *Survey of Current Business* para recopilar más datos, aunque se decepcionó al encontrar que en los años que le interesaba analizar la definición de su variable dependiente había cambiado, lo cual producía una serie de tiempo inconsistente. Es decir, la serie subió a un punto intermedio a través del periodo que estaba estudiando.

Jill le planteó este problema a su jefe y coincidieron en que las exportaciones totales de mercancía podrían utilizarse como la variable dependiente en lugar de las exportaciones de bienes de consumo. Jill descubrió que esta variable estaba presente en varias ediciones de *Survey of Current Business* y que era posible recolectar datos correspondientes a varios años. Entonces, reunió los datos que se muestran en la tabla 11.3, los retrasó un periodo y de nuevo corrió un análisis de regresión utilizando la  $Y$  retrasada un periodo como una variable explicativa.

De nuevo encontró buenas estadísticas en su listado de regresión, excepto por la estadística Durbin-Watson. Este valor era de 0.96; entonces, llegó a la conclusión de que la autocorrelación estaba allí porque este valor está debajo de del valor más bajo de la tabla para la prueba Durbin-Watson. Jill intentó corridas adicionales al agregar el número del periodo y el cambio en  $Y$  como variables explicativas adicionales. Sin

**TABLA 11.3** Datos trimestrales de la serie de tiempo: exportaciones totales de mercancía para el caso 11.2 (miles de millones de dólares)

| Año  | Trimestre | $Y$   | $Y$ retrasada un periodo |
|------|-----------|-------|--------------------------|
| 1984 | 1         | 219.3 | —                        |
|      | 2         | 223.1 | 219.3                    |
|      | 3         | 225.9 | 223.1                    |
|      | 4         | 228.0 | 225.9                    |
| 1985 | 1         | 225.0 | 228.0                    |
|      | 2         | 221.6 | 225.0                    |
|      | 3         | 218.0 | 221.6                    |
|      | 4         | 218.6 | 218.0                    |
| 1986 | 1         | 220.7 | 218.6                    |
|      | 2         | 221.4 | 220.7                    |
|      | 3         | 225.7 | 221.4                    |
|      | 4         | 230.4 | 225.7                    |
| 1987 | 1         | 234.5 | 230.4                    |
|      | 2         | 246.6 | 234.5                    |
|      | 3         | 261.6 | 246.6                    |
|      | 4         | 281.3 | 261.6                    |
| 1988 | 1         | 306.7 | 281.3                    |
|      | 2         | 319.2 | 306.7                    |
|      | 3         | 327.9 | 319.2                    |
|      | 4         | 342.8 | 327.9                    |
| 1989 | 1         | 360.6 | 342.8                    |
|      | 2         | 373.2 | 360.6                    |
|      | 3         | 367.3 | 373.2                    |
|      | 4         | 378.7 | 367.3                    |
| 1990 | 1         | 394.2 | 378.7                    |
|      | 2         | 394.4 | 394.2                    |

Fuente: *Survey of Current Business*, diversas ediciones.

embargo, no pudo encontrar una estadística Durbin-Watson que fuera tan elevada como para concluir que se había eliminado la autocorrelación. Entonces, decidió buscar otras técnicas para pronosticar su nueva variable dependiente: las exportaciones totales de la mercancía. Usó los datos de la serie de tiempo que se muestra en la columna  $Y$  de la tabla 11.3.

Entre el software disponible en Busby se encontraba uno ya anticuado para la generación de pronósticos llamado Sibyl/Runner. Jill preparó sus datos para utilizarlos con este programa y comenzó a trabajar a su manera mediante diversas rutinas de pronóstico. Entre éstas utilizó un procedimiento para graficar datos sin tratar. La gráfica trimestral de las exportaciones totales para el periodo de 1984 al segundo trimestre de 1990 se muestra en la figura 11.4.

Luego de estudiar la figura 11.4., Jill decidió usar sólo los últimos 16 puntos de datos en su esfuerzo por efectuar el pronóstico. Ella dedujo que, comenzando con el periodo 9, la serie había mostrado un incremento relativamente constante, mientras que antes de ese periodo mostraba un aumento y un declive. El programa Sibyl/Runner le advirtió que estaba utilizando pocos puntos de datos (16), pero ella continuó sin hacer caso de la recomendación.

Comenzó su análisis al pedir al programa la rutina Sibyl utilizando únicamente los últimos 16 puntos de datos. El primer análisis producido por el programa indicó que había cierto patrón en los datos, pero que no

había estacionalidad. El programa utilizó diversos métodos para pronosticar los valores de la muestra e imprimir una tabla que mostraba los errores del pronóstico como porcentajes para cada método y periodo.

Después de estudiar estos errores del pronóstico, Jill seleccionó tres técnicas de pronóstico para análisis posteriores. Estas técnicas tenían el porcentaje más bajo de los errores de pronóstico y eran las siguientes:

Suavizamiento exponencial simple.

Suavizamiento exponencial lineal de Holt, que toma en cuenta la tendencia en los datos.

Suavizamiento exponencial cuadrática, que permite una línea de tendencia curva a través de los datos.

Entonces, Jill realizó pronósticos para sus datos mediante cada una de las técnicas, con los resultados que se muestran en la tabla 11.4.

**TABLA 11.4**

| Técnica                | Error cuadrático medio | Constante de suavizamiento óptima |
|------------------------|------------------------|-----------------------------------|
| Exponencial simple     | 185.5                  | .999                              |
| Lineal de Holt         | 81.0                   | .722                              |
| Exponencial cuadrática | 87.7                   | .485                              |

**FIGURA 11.4 Gráfica de los valores trimestrales de datos: exportaciones totales, del primer trimestre de 1984 al segundo trimestre de 1990**



Jill observó que la constante de suavizamiento óptima por medio de un suavizamiento exponencial simple fue casi de 1.00 (0.999). Aparentemente, para poder rastrear los datos de una manera óptima, el programa utilizó básicamente cada valor de datos para predecir el siguiente. Éste es el equivalente a utilizar un método informal simple para generar pronósticos; es decir, un modelo que establece que las diferencias consecutivas son aleatorias.

El suavizamiento exponencial lineal de Holt produjo el error cuadrático medio más bajo, de manera que Jill pensó que podría utilizar este método. Examinó más a fondo el reporte que el programa de cómputo emitió para este método y descubrió que el error porcentual absoluto medio (MAPE) era de 2.39 y el error porcentual medio (MPE), o desviación, era de 0.34. Asimismo, consideró estos valores como satisfactorios y decidió utilizar el suavizamiento exponencial lineal de Holt para pronosticar los valores futuros.

Conforme al método seleccionado, solicitó pronósticos para los siguientes cuatro períodos más allá del final de sus datos. Los valores del pronóstico son:

| <i>Periodo</i> | <i>Pronóstico</i> |
|----------------|-------------------|
| 17             | 401.27            |
| 18             | 407.30            |
| 19             | 413.34            |
| 20             | 419.38            |

Jill observó que, con cada trimestre que pasaba, debía existir un nuevo valor real de las exportaciones totales y que podrían actualizarse los pronósticos para los períodos futuros.

Jill se reunió con su jefe para comentar sus resultados. Dijo que creía tener una buena manera para pronosticar la variable nacional, las exportaciones totales, mediante un suavizamiento exponencial con ajustes a la tendencia. El jefe le pidió que explicara tal método, y así lo hizo. Su siguiente tarea fue utilizar los datos reales para verificar la corazonada del presidente de Busby con respecto a que las exportaciones de la empresa estaban bien correlacionadas con las exportaciones nacionales. Si ella pudiera establecer esto, Busby tendría un buen método de pronóstico para sus exportaciones y podría utilizar los pronósticos para planear sus operaciones futuras.

## PREPREGUNTAS

1. Jill no consideró combinar los pronósticos generados por los tres métodos que analizó. ¿Cómo haría esto? ¿Cuáles serían las ventajas y las desventajas de tal acción?
2. Las constantes de suavizamiento óptimas utilizadas en el suavizamiento exponencial lineal de Holt corresponden a  $\alpha = .722$  y  $\beta = .722$ . Conforme aparezcan nuevos datos en los próximos trimestres, posiblemente Jill deberá correr nuevamente sus datos para ver si han cambiado estos valores. ¿Con qué frecuencia cree usted que ella debería realizar esto?
3. Es posible que la selección del método de pronóstico cambie a otra técnica conforme se añaden datos trimestrales a la base de datos. ¿Debería Jill correr otra vez su análisis entero con cierta frecuencia para verificar esto? ¿Con qué frecuencia debería repetirse esta acción?
4. Un colega de Jill sugirió que utilizara la metodología ARIMA Box-Jenkins. ¿Qué le recomendaría a Jill si ella decide utilizar la metodología Box-Jenkins?

## CASO 11-3 CONSUMER CREDIT COUNSELING

En los capítulos 1 (caso 1.2) y 3 (caso 3.3) se describieron las operaciones de Consumer Credit Counseling (CCC).

El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de clientes nuevos que se tendrían por el resto de 1993. Marv proporcionó a Do-

rothy Mercer los datos mensuales para los números de nuevos clientes que se habían sido atendidos por CCC en el periodo de enero de 1985 a marzo de 1993 (véase el caso 3.3).

Con ayuda del lector, Dorothy ha intentado diversas formas para pronosticar la variable más importante. Estos esfuerzos se presentan en los casos 4.3, 5.3,

6.5, 8.5 y 9.3. Luego de completar estos intentos de generar pronósticos, Dorothy decide que es tiempo de hacer un resumen y de proponerse lograr un método de pronóstico que funcione para el resto del año.

### Tarea

Suponga que Dorothy le pide su ayuda para solucionar este problema de pronóstico. Escriba un informe que recomiende un curso de la acción. Considere que Marv

debe desarrollar pronósticos precisos para el número de clientes vistos y que los pueda utilizar en la toma cotidiana de decisiones de la empresa. Especifique lo que usted les recomienda a Dorothy y Marv. Recuerde tomar en cuenta las cuestiones que se comentaron en este capítulo, como el costo. ■

## CASO 11-4 MR. TUX

El propietario de varias tiendas Mr. Tux, John Mosby, ha intentado varias maneras para pronosticar su variable más importante, el importe de las ventas mensuales. Sus esfuerzos se resumen al final de diversos capítulos de este libro. Luego de completar los intentos para generar pronósticos, John decide que es tiempo de resumir sus esfuerzos y de alcanzar un método para pronosticar el futuro. Se da cuenta de que debe actualizar tanto sus datos como su método de pronóstico en algún momento futuro, pero actualmente necesita seleccionar una forma de pronosticar las ventas de los próximos meses.

Para comenzar, John resume los resultados de los métodos que hasta este momento ha puesto a prueba.

- Caso 2.2: El uso del promedio anual para pronosticar las ventas futuras por año podría funcionar; pero, debido a que John identificó una tendencia ascendente, necesita una manera efectiva para extender estos promedios al futuro. De igual forma, a John le preocupa mucho el efecto estacional debido a que sabe que sus ventas varían considerablemente cada mes. Sus esfuerzos no han rendido frutos al utilizar promedios anuales.
- Caso 3.2: Mediante el uso que le ha dado al programa Minitab, John estableció que tanto la tendencia como el efecto estacional existían en estos datos. No obstante, él sabía que estos elementos estaban ahí antes de que comenzara, y estaba satisfecho porque su software los había establecido de una manera estadística. El programa también indicó que diversos coeficientes de autocorrelación estaban fuera de los límites de error de la muestra, lo cual indicaba a John que tanto los efectos de tendencia como los estacionales necesitaban ser reflejados en su modelo final de pronósticos.
- Caso 4.2: Cuando John utilizó el suavizamiento exponencial, incluyendo los métodos que consideraban los factores de tendencia y estacionales, las medidas del error resultante no fueron satisfactorias.
- Caso 5.2: Finalmente, John obtuvo algunos resultados alentadores utilizando el método de descomposición para construir una línea de tendencia, índices estacionales y un componente cíclico para sus datos. Entonces, fue capaz de mostrar a su banquero los índices estacionales y llegar a los acuerdos deseados para los pagos de su préstamo. También generó pronósticos para los siguientes meses al ensamblar nuevamente sus componentes estimados. No obstante, John estaba un poco inquieto por el amplio rango de sus proyecciones.
- Caso 6.4: El análisis de regresión simple fue la siguiente técnica que probó John, con el periodo como la variable independiente o explicativa. Entonces, dedujo que esta variable sería parte del factor de tendencia que estaba en sus datos. Este método no consideró el factor de estacionalidad en las ventas y el valor  $r$  cuadrado de 56.3% no fue satisfactorio.
- Caso 8.4: John llevó a la práctica una regresión múltiple utilizando el número de períodos para reflejar una tendencia y una serie de variables ficticias para considerar el efecto estacional (meses). Su valor de 88% para  $R$  cuadrada fue una mejoría considerable con respecto a la regresión simple. Sin embargo, el error de pronóstico para los últimos 12 meses de sus datos, conforme se midió por el error porcentual absoluto medio (*MAPE*) de 21% era inaceptable. Con esta clase de error, John decidió no utilizar una regresión múltiple. También probó un modelo autorregresivo estacional que dio como resultado un valor para  $R$  cuadrada de 90.6%. John estaba muy satisfecho con este resultado.
- Caso 9.2: A John le molestaba no comprender del todo la metodología Box-Jenkins ARIMA desde

Entonces, observó que estas medidas, como el promedio de los errores y el promedio de los errores porcentuales, resultaban de predecir los valores pasados de esta variable. Pero como eran demasiado altos, él no quería utilizar estas técnicas para predecir el futuro desconocido.

el comienzo. Sin embargo, reconoció que su modelo autorregresivo estacional era un modelo ARIMA particular y deseaba saber si podría mejorarlo. Asimismo, sabía que tendría que explicar cualquiera de los pronósticos que surgieran a los inversionistas y banqueros en su intento por obtener más capital para la expansión, de manera que deseó un método que fuera preciso y comprensible.

Al reflexionar sobre estos factores, John advirtió que el tiempo pasaba y que pronto debía generar pronósticos de los ingresos mensuales. Tenía poco tiempo para intentar hacer modificaciones en los métodos que usó y podría pensar en la combinación de dos o más métodos. Sin embargo, no tenía tiempo para adquirir software nuevo ni para probar métodos distin-

tos. Mientras se preguntaba qué debía hacer, leyó una de sus frases favoritas que estaba sobre la pared de su oficina: "Hagamos algo, incluso si no es del todo correcto."

### Tarea

Suponga que usted ha sido contratado para ayudar a John Mosby con su problema de pronósticos. Escríbale un informe que resuma sus esfuerzos a la fecha y recomiéndele acciones a seguir. Considere que, muy pronto, John debe desarrollar pronósticos para las ventas mensuales que sean muy precisos y que él pueda utilizar en las negociaciones con los inversionistas. Sea muy específico en lo que usted recomienda al propietario de Mr. Tux. ■

## CASO 11-5 ALOMEGA FOOD STORES

El ejemplo 1.1 del capítulo 1 describió cómo Julie Ruth, presidenta de Alomega Food Stores, recopiló datos de las ventas mensuales para su compañía junto con otras variables que podrían estar relacionadas con las ventas. Los casos de Alomega que se presentaron en los capítulos 2, 3, 5, 8, 9 y 10 describieron sus esfuerzos para utilizar diversos procedimientos Minitab en un intento para realizar pronósticos significativos de las ventas mensuales.

Julie sabía que su personal técnico estaba utilizando los mismos datos para generar buenos pronósticos, pero no sabía de qué manera estaban resultando. Además, quería obtener un buen método de pronóstico por sí misma. Sabía que al ser la primera presidenta de

Alomega, había superado a diversos candidatos para ese cargo y que podría existir cierto resentimiento entre su equipo directivo. Debido a los comentarios negativos que expresó en una junta reciente Jackson Tilson, gerente de producción, ella estaba especialmente sensible (véase el final del ejemplo 1.1).

Al revisar sus esfuerzos, Julie decidió desechar el análisis de regresión simple que había empleado, como se reseñó en el caso 2.3. En ese momento, debía elegir entre el análisis por descomposición descrito en el caso 5.6, la regresión múltiple descrita en el caso 8.7 o una combinación de métodos, como se describió en el caso 10.2. ■

### PREGUNTAS

1. Suponga que usted ha sido contratado por Alomega Food Stores y se le asigna la tarea de ayudar a Julie a desarrollar un método de pronóstico eficaz para las ventas mensuales. Después de revisar los esfuerzos hechos por Julie en la aplicación de los métodos, ¿cuál de éstos le recomendaría?
2. Con base en su respuesta a la pregunta 1, escriba un informe detallado a Julie para explicarle las razones de su elección; también, para indicarle hasta qué grado es eficaz este método de pronóstico.
3. Además de seleccionar entre los métodos que Julie probó, ¿qué otros procedimientos de pronóstico le recomendaría usted? Utilice Minitab o cualquier otro software y compárelos con su selección de la pregunta 1.
4. ¿Debería Julie combinar los pronósticos?

### Referencias

Armstrong, J. S. (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA, Kluwer Academic Publishers, 2001.

Naisbitt, J., *Megatrends*, Nueva York, Warner Books, 1982.



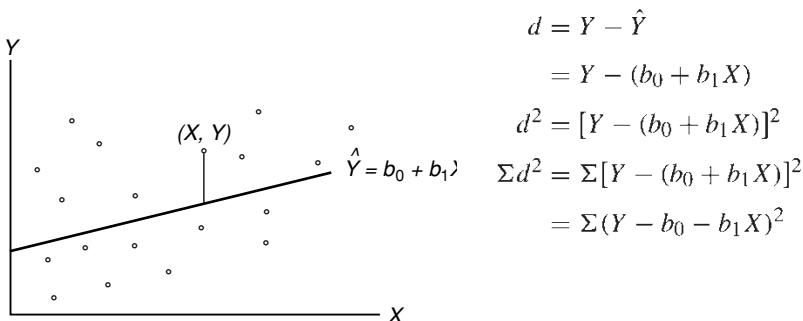
## Apéndice A

# Deducción de fórmulas

### Deducción de fórmulas de correlación

$$\begin{aligned}
 r &= \frac{\sum Z_X Z_Y}{n-1} = \sum \frac{[(X - \bar{X})/S_X][(Y - \bar{Y})/S_Y]}{n-1} \\
 &= \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum X^2 - n(\Sigma X/n)^2} \sqrt{\sum Y^2 - n(\Sigma Y/n)^2}} \\
 &= \frac{n \sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{n \sum X^2 - (\Sigma X)^2} \sqrt{n \sum Y^2 - (\Sigma Y)^2}} \\
 &= \frac{n[\sum XY - (\Sigma X \Sigma Y)/n]}{\sqrt{n \sum X^2 - (\Sigma X)^2} \sqrt{n \sum Y^2 - (\Sigma Y)^2}}
 \end{aligned}$$

### Deducción de los estimadores de mínimos cuadrados



### Derivadas parciales

$$\begin{aligned}
 \frac{\delta \Sigma}{\delta b_1} &= 2 \sum (Y - b_0 - b_1 X)(-X) \\
 &= 2 \sum (-XY + b_0 X + b_1 X^2)
 \end{aligned}
 \quad
 \begin{aligned}
 \frac{\delta \Sigma}{\delta b_0} &= 2 \sum (Y - b_0 - b_1 X)(-1) \\
 &= 2 \sum (-Y + b_0 + b_1 X)
 \end{aligned}$$

**Para obtener mínimos las derivadas parciales se igualan a cero.**

$$\begin{array}{ll} \frac{\delta \Sigma}{\delta b_1} = 0 \Rightarrow 2\Sigma(-XY + b_0X + b_1X^2) = 0 & \frac{\delta \Sigma}{\delta b_0} = 0 \Rightarrow 2\Sigma(-Y + b_0 + b_1X) = 0 \\ \Sigma(-XY + b_0X + b_1X^2) = 0 & \Sigma(-Y + b_0 + b_1X) = 0 \\ -\Sigma XY + b_0\Sigma X + b_1\Sigma X^2 = 0 & -\Sigma Y + nb_0 + b_1\Sigma X = 0 \end{array}$$

**Encuentre un  $b_0$  y  $b_1$  a fin de que  $\Sigma d^2$  sea un mínimo**

$$\begin{array}{ll} b_0\Sigma X + b_1\Sigma X^2 = \Sigma XY & \times n \\ nb_0 + b_1\Sigma X = \Sigma Y & \times \Sigma X \end{array}$$

Con

$$\begin{array}{c} nb_0\Sigma X + nb_1\Sigma X^2 = n\Sigma XY \\ nb_0\Sigma X + b_1(\Sigma X)^2 = \Sigma X\Sigma Y \\ \hline nb_1\Sigma X^2 - b_1(\Sigma X)^2 = n\Sigma XY - \Sigma X\Sigma Y \end{array} \quad \text{réstale}$$

y resuelva  $b_1$  para obtener la fórmula de la pendiente

$$\begin{aligned} b_1[n\Sigma X^2 - (\Sigma X)^2] &= n\Sigma XY - \Sigma X\Sigma Y \\ b_1 &= \frac{n\Sigma XY - \Sigma X\Sigma Y}{n\Sigma X^2 - (\Sigma X)^2} \end{aligned}$$

Resuelva

$$nb_0 + b_1\Sigma X = \Sigma Y$$

para obtener la fórmula de la intercepción con el eje  $Y$

$$\begin{aligned} nb_0 &= \Sigma Y - b_1\Sigma X \\ b_0 &= \frac{\Sigma Y}{n} - b_1 \frac{\Sigma X}{n} = \bar{Y} - b_1 \bar{X} \end{aligned}$$

## Apéndice B

# Datos para el caso 7.1

|       |      |      |      |       |       |       |       |      |      |      |       |       |       |
|-------|------|------|------|-------|-------|-------|-------|------|------|------|-------|-------|-------|
| 14.75 | 0.00 | 0.00 | 2.01 | 12.57 | 10.00 | 20.50 | 16.13 | 0.00 | 0.00 | 2.80 | 13.39 | 10.00 | 16.50 |
| 14.50 | 0.00 | 0.00 | 2.53 | 12.57 | 10.00 | 20.00 | 15.75 | 1.00 | 0.00 | 4.00 | 13.35 | 30.00 | 16.00 |
| 14.13 | 1.00 | 0.00 | 2.10 | 12.57 | 10.00 | 20.00 | 16.13 | 0.00 | 0.00 | 2.81 | 13.50 | 10.00 | 15.80 |
| 14.63 | 0.00 | 1.00 | 4.13 | 12.14 | 30.00 | 20.00 | 16.25 | 1.00 | 0.00 | 3.38 | 13.50 | 30.00 | 15.80 |
| 14.00 | 1.00 | 0.00 | 2.10 | 12.57 | 10.00 | 20.00 | 16.00 | 0.00 | 0.00 | 2.57 | 13.50 | 10.00 | 15.80 |
| 13.38 | 0.00 | 1.00 | 3.97 | 12.57 | 10.00 | 20.00 | 15.88 | 0.00 | 1.00 | 3.96 | 13.50 | 30.00 | 15.80 |
| 14.57 | 0.00 | 1.00 | 3.27 | 12.14 | 30.00 | 20.00 | 16.50 | 1.00 | 0.00 | 2.67 | 13.50 | 30.00 | 15.80 |
| 13.88 | 1.00 | 0.00 | 3.50 | 13.19 | 10.00 | 19.50 | 16.38 | 1.00 | 0.00 | 3.05 | 13.50 | 30.00 | 15.80 |
| 15.38 | 0.00 | 0.00 | 2.85 | 13.19 | 10.00 | 19.50 | 12.50 | 1.00 | 0.00 | 2.36 | 10.60 | 30.00 | 15.30 |
| 15.63 | 0.00 | 0.00 | 1.81 | 13.12 | 10.00 | 18.50 | 12.25 | 1.00 | 0.00 | 2.54 | 10.60 | 30.00 | 15.30 |
| 15.88 | 1.00 | 0.00 | 2.72 | 12.69 | 30.00 | 18.50 | 14.25 | 1.00 | 0.00 | 2.20 | 12.13 | 30.00 | 15.30 |
| 15.00 | 1.00 | 0.00 | 2.43 | 13.12 | 10.00 | 18.00 | 15.00 | 1.00 | 0.00 | 3.03 | 12.13 | 30.00 | 15.80 |
| 16.13 | 0.00 | 0.00 | 3.27 | 12.69 | 30.00 | 18.00 | 15.25 | 1.00 | 0.00 | 3.24 | 12.13 | 30.00 | 16.50 |
| 15.25 | 0.00 | 1.00 | 3.13 | 12.69 | 30.00 | 17.50 | 16.00 | 0.00 | 0.00 | 1.95 | 12.34 | 30.00 | 17.80 |
| 16.00 | 0.00 | 0.00 | 2.55 | 13.68 | 10.00 | 17.00 | 14.88 | 1.00 | 0.00 | 2.86 | 12.34 | 30.00 | 17.80 |
| 16.25 | 0.00 | 0.00 | 2.08 | 13.68 | 10.00 | 17.50 | 14.75 | 1.00 | 0.00 | 2.64 | 12.34 | 30.00 | 19.00 |
| 17.38 | 0.00 | 0.00 | 2.12 | 13.20 | 30.00 | 17.50 | 15.50 | 1.00 | 0.00 | 2.23 | 11.40 | 30.00 | 20.00 |
| 16.35 | 1.00 | 0.00 | 3.40 | 14.10 | 10.00 | 19.00 | 13.75 | 1.00 | 0.00 | 2.24 | 11.40 | 30.00 | 19.50 |
| 17.00 | 1.00 | 0.00 | 2.63 | 13.60 | 30.00 | 19.00 | 11.30 | 1.00 | 0.00 | 3.24 | 11.36 | 30.00 | 17.50 |
| 16.00 | 0.00 | 1.00 | 2.61 | 14.10 | 10.00 | 19.50 | 12.38 | 1.00 | 0.00 | 1.95 | 11.36 | 30.00 | 17.50 |
| 16.63 | 1.00 | 0.00 | 2.06 | 14.10 | 10.00 | 19.50 | 12.15 | 1.00 | 0.00 | 2.32 | 11.36 | 30.00 | 14.50 |
| 16.38 | 0.00 | 0.00 | 2.08 | 14.10 | 10.00 | 20.00 | 11.75 | 1.00 | 0.00 | 2.45 | 9.81  | 30.00 | 13.00 |
| 16.75 | 1.00 | 0.00 | 2.09 | 13.60 | 30.00 | 20.00 | 12.38 | 1.00 | 0.00 | 1.88 | 9.81  | 30.00 | 13.00 |
| 15.13 | 0.00 | 1.00 | 4.29 | 12.69 | 30.00 | 20.00 | 12.63 | 0.00 | 0.00 | 1.76 | 9.81  | 30.00 | 13.00 |
| 16.00 | 1.00 | 0.00 | 2.50 | 12.96 | 30.00 | 20.00 | 11.13 | 1.00 | 0.00 | 1.99 | 9.81  | 30.00 | 12.50 |
| 14.50 | 0.00 | 1.00 | 3.32 | 13.47 | 10.00 | 20.00 | 11.38 | 0.00 | 0.00 | 2.20 | 9.78  | 10.00 | 12.50 |
| 16.25 | 0.00 | 0.00 | 2.95 | 12.96 | 30.00 | 20.00 | 11.88 | 1.00 | 0.00 | 2.14 | 9.81  | 30.00 | 12.00 |
| 16.88 | 0.00 | 0.00 | 1.85 | 14.28 | 10.00 | 20.50 | 11.75 | 1.00 | 0.00 | 2.61 | 9.81  | 30.00 | 12.00 |
| 17.38 | 0.00 | 0.00 | 1.55 | 13.59 | 30.00 | 20.50 | 13.63 | 0.00 | 0.00 | 1.84 | 10.24 | 30.00 | 11.00 |
| 16.00 | 0.00 | 1.00 | 3.33 | 14.28 | 10.00 | 20.50 | 13.88 | 0.00 | 0.00 | 1.62 | 11.00 | 30.00 | 11.00 |
| 16.75 | 1.00 | 0.00 | 2.77 | 14.94 | 10.00 | 20.50 | 13.00 | 1.00 | 0.00 | 3.56 | 11.00 | 30.00 | 11.00 |
| 17.13 | 0.00 | 0.00 | 2.18 | 14.94 | 10.00 | 20.50 | 12.00 | 1.00 | 0.00 | 2.65 | 11.10 | 10.00 | 11.00 |
| 17.50 | 0.00 | 1.00 | 4.21 | 14.67 | 30.00 | 20.50 | 13.13 | 1.00 | 0.00 | 2.65 | 11.00 | 30.00 | 11.00 |
| 17.00 | 1.00 | 0.00 | 2.66 | 15.32 | 10.00 | 19.50 | 14.27 | 0.00 | 0.00 | 1.80 | 11.34 | 30.00 | 12.30 |
| 16.75 | 0.00 | 1.00 | 3.58 | 15.32 | 10.00 | 19.50 | 14.63 | 0.00 | 0.00 | 1.69 | 11.34 | 30.00 | 12.30 |
| 17.20 | 0.00 | 1.00 | 2.96 | 15.32 | 10.00 | 19.50 | 15.25 | 0.00 | 0.00 | 1.88 | 11.34 | 30.00 | 12.20 |
| 18.75 | 0.00 | 0.00 | 1.93 | 15.32 | 10.00 | 19.50 | 14.25 | 1.00 | 0.00 | 2.77 | 11.34 | 30.00 | 12.30 |
| 17.50 | 0.00 | 1.00 | 2.57 | 14.68 | 30.00 | 19.00 | 13.52 | 1.00 | 0.00 | 2.22 | 11.75 | 10.00 | 13.50 |
| 17.50 | 0.00 | 0.00 | 3.18 | 15.15 | 10.00 | 18.00 | 14.63 | 1.00 | 0.00 | 2.42 | 11.59 | 30.00 | 13.50 |
| 18.00 | 0.00 | 0.00 | 1.93 | 15.15 | 10.00 | 18.00 | 14.75 | 0.00 | 0.00 | 1.77 | 11.39 | 30.00 | 13.50 |
| 15.63 | 0.00 | 0.00 | 2.20 | 13.39 | 10.00 | 17.00 | 14.00 | 0.00 | 0.00 | 2.22 | 11.75 | 10.00 | 13.50 |
| 14.75 | 1.00 | 0.00 | 2.21 | 13.39 | 10.00 | 17.00 | 14.50 | 0.00 | 0.00 | 2.99 | 11.59 | 30.00 | 13.50 |
| 15.25 | 0.00 | 1.00 | 3.24 | 13.35 | 30.00 | 16.50 | 14.25 | 0.00 | 0.00 | 2.22 | 11.75 | 10.00 | 13.50 |
| 15.75 | 1.00 | 0.00 | 2.35 | 13.35 | 30.00 | 16.50 | 14.63 | 0.00 | 0.00 | 1.93 | 11.75 | 10.00 | 14.50 |
| 15.25 | 1.00 | 0.00 | 2.11 | 13.39 | 10.00 | 16.50 | 13.30 | 1.00 | 0.00 | 3.35 | 12.68 | 10.00 | 15.50 |
| 15.75 | 1.00 | 0.00 | 2.80 | 13.35 | 30.00 | 16.50 | 14.50 | 0.00 | 0.00 | 2.21 | 12.68 | 10.00 | 17.00 |
| 15.63 | 0.00 | 0.00 | 1.95 | 13.39 | 10.00 | 16.50 |       |      |      |      |       |       |       |



## Apéndice C

# Tablas

TABLA C-1 Términos individuales de la distribución binomial

TABLA C-2 Tabla de las áreas para la distribución de probabilidad normal estándar

TABLA C-3 Valores críticos de la estadística *t de Student*

TABLA C-4 Valores críticos de la distribución chi cuadrada

TABLA C-5 Tabla de la distribución *F*

TABLA C-6 Límites en la prueba Durbin-Watson

**TABLA C-1 Términos individuales de la distribución binomial**

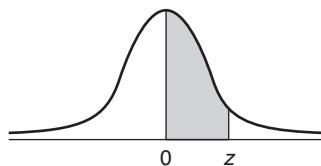
| <i>n</i> | <i>x</i> | <i>p</i> |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|----------|----------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|          |          | .05      | .10   | .15   | .20   | .25   | .30   | .35   | .40   | .45   | .50   | .55   | .60   | .65   | .70   | .75   | .80   | .85   | .90   | .95   |
| 1        | 0        | .9500    | .9000 | .8500 | .8000 | .7500 | .7000 | .6500 | .6000 | .5500 | .5000 | .4500 | .4000 | .3500 | .3000 | .2500 | .2000 | .1500 | .1000 | .0500 |
|          | 1        | .0500    | .1000 | .1500 | .2000 | .2500 | .3000 | .3500 | .4000 | .4500 | .5000 | .5500 | .6000 | .6500 | .7000 | .7500 | .8000 | .8500 | .9000 | .9500 |
| 2        | 0        | .9025    | .8100 | .7225 | .6400 | .5625 | .4900 | .4225 | .3600 | .3025 | .2500 | .2025 | .1600 | .1225 | .0900 | .0625 | .0400 | .0225 | .0100 | .0025 |
|          | 1        | .0950    | .1800 | .2550 | .3200 | .3750 | .4200 | .4550 | .4800 | .4950 | .5000 | .4950 | .4800 | .4550 | .4200 | .3750 | .3200 | .2550 | .1800 | .0950 |
| 3        | 0        | .8574    | .7290 | .6141 | .5120 | .4219 | .3430 | .2746 | .2160 | .1664 | .1250 | .0911 | .0640 | .0429 | .0270 | .0156 | .0080 | .0034 | .0010 | .0001 |
|          | 1        | .1354    | .2430 | .3251 | .3840 | .4219 | .4410 | .4436 | .4320 | .4084 | .3750 | .3341 | .2880 | .2389 | .1890 | .1406 | .0960 | .0574 | .0270 | .0071 |
| 4        | 0        | .0071    | .0270 | .0574 | .0960 | .1406 | .1890 | .2389 | .2880 | .3341 | .3750 | .4084 | .4320 | .4436 | .4410 | .4219 | .3840 | .3251 | .2430 | .1354 |
|          | 1        | .0001    | .0010 | .0034 | .0080 | .0156 | .0270 | .0429 | .0640 | .0911 | .1250 | .1664 | .2160 | .2746 | .3430 | .4116 | .5120 | .6141 | .7290 | .8574 |
| 5        | 0        | .0005    | .0015 | .0026 | .0046 | .0075 | .0115 | .0153 | .0196 | .0241 | .0285 | .0330 | .0385 | .0441 | .0496 | .0551 | .0606 | .0661 | .0715 | .0769 |
|          | 1        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |
| 6        | 0        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |
|          | 1        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |
| 7        | 0        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |
|          | 1        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |
| 8        | 0        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |
|          | 1        | .0000    | .0001 | .0005 | .0016 | .0039 | .0081 | .0150 | .0256 | .0410 | .0625 | .0915 | .1296 | .1785 | .2401 | .3164 | .4096 | .5220 | .6561 | .8145 |

**TABLA C-1** (*Continúa*)

| n  | x     | p     |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|    |       | .05   | .10   | .15   | .20   | .25   | .30   | .35   | .40   | .45   | .50   | .55   | .60   | .65   | .70   | .75   | .80   | .85   | .90   | .95   |
| 2  | .0515 | .1488 | .2376 | .2936 | .3115 | .2965 | .2587 | .2090 | .1569 | .1094 | .0703 | .0413 | .0217 | .0100 | .0038 | .0011 | .0002 | .0000 | .0000 |       |
| 3  | .0054 | .0331 | .0839 | .1468 | .2076 | .2541 | .2786 | .2787 | .2568 | .2188 | .1719 | .1239 | .0808 | .0467 | .0231 | .0092 | .0026 | .0004 | .0000 |       |
| 4  | .0004 | .0046 | .0185 | .0459 | .0865 | .1361 | .1875 | .2322 | .2627 | .2734 | .2627 | .2322 | .1875 | .1361 | .0865 | .0459 | .0185 | .0046 | .0004 |       |
| 5  | .0000 | .0004 | .0026 | .0092 | .0231 | .0467 | .0808 | .1239 | .1719 | .2188 | .2568 | .2787 | .2786 | .2541 | .2076 | .1468 | .0839 | .0331 | .0054 |       |
| 6  | .0000 | .0000 | .0002 | .0011 | .0038 | .0100 | .0217 | .0413 | .0703 | .1094 | .1569 | .2090 | .2587 | .2965 | .3115 | .2936 | .2376 | .1488 | .0515 |       |
| 7  | .0000 | .0000 | .0000 | .0001 | .0004 | .0012 | .0033 | .0079 | .0164 | .0312 | .0548 | .0896 | .1373 | .1977 | .2670 | .3355 | .3847 | .3826 | .2793 |       |
| 8  | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0002 | .0007 | .0017 | .0039 | .0084 | .0168 | .0319 | .0576 | .1001 | .1678 | .2725 | .4305 | .6634 |       |
| 9  | 0     | .6302 | .3874 | .2316 | .1342 | .0751 | .0404 | .0207 | .0101 | .0046 | .0020 | .0008 | .0003 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 |       |
|    | 1     | .2986 | .3874 | .3679 | .3020 | .2253 | .1556 | .1004 | .0605 | .0339 | .0176 | .0083 | .0035 | .0013 | .0004 | .0001 | .0000 | .0000 | .0000 |       |
|    | 2     | .0629 | .1722 | .2597 | .3020 | .3003 | .2668 | .2162 | .1612 | .1110 | .0703 | .0407 | .0212 | .0098 | .0039 | .0012 | .0003 | .0000 | .0000 |       |
|    | 3     | .0077 | .0446 | .1069 | .1762 | .2336 | .2668 | .2716 | .2508 | .2119 | .1641 | .1160 | .0743 | .0424 | .0210 | .0087 | .0028 | .0006 | .0001 | .0000 |
|    | 4     | .0006 | .0074 | .0283 | .0661 | .1168 | .1715 | .2194 | .2508 | .2600 | .2461 | .2128 | .1672 | .1181 | .0735 | .0389 | .0165 | .0050 | .0008 | .0000 |
|    | 5     | .0000 | .0008 | .0050 | .0165 | .0389 | .0735 | .1181 | .1672 | .2128 | .2461 | .2600 | .2508 | .2194 | .1715 | .1168 | .0661 | .0283 | .0074 | .0006 |
|    | 6     | .0000 | .0001 | .0006 | .0028 | .0087 | .0210 | .0424 | .0743 | .1160 | .1641 | .2119 | .2508 | .2716 | .2668 | .2336 | .1762 | .1069 | .0446 | .0077 |
|    | 7     | .0000 | .0000 | .0000 | .0003 | .0012 | .0039 | .0098 | .0212 | .0407 | .0703 | .1110 | .1612 | .2162 | .2668 | .3003 | .3020 | .2597 | .1722 | .0629 |
|    | 8     | .0000 | .0000 | .0000 | .0000 | .0001 | .0004 | .0013 | .0035 | .0083 | .0176 | .0339 | .0605 | .1004 | .1556 | .2253 | .3020 | .3679 | .3874 | .2986 |
|    | 9     | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0003 | .0008 | .0020 | .0046 | .0101 | .0207 | .0404 | .0751 | .1342 | .2316 | .3874 | .6302 |
| 10 | 0     | .5987 | .3487 | .1969 | .1074 | .0563 | .0282 | .0135 | .0060 | .0025 | .0010 | .0003 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |       |
|    | 1     | .3151 | .3874 | .3474 | .2684 | .1877 | .1211 | .0725 | .0403 | .0207 | .0098 | .0042 | .0016 | .0005 | .0001 | .0000 | .0000 | .0000 | .0000 |       |
|    | 2     | .0746 | .1937 | .2759 | .3020 | .2816 | .2335 | .1757 | .1209 | .0763 | .0439 | .0229 | .0106 | .0043 | .0014 | .0004 | .0001 | .0000 | .0000 |       |
|    | 3     | .0105 | .0574 | .1298 | .2013 | .2503 | .2668 | .2522 | .2150 | .1665 | .1172 | .0746 | .0425 | .0212 | .0090 | .0031 | .0008 | .0001 | .0000 | .0000 |
|    | 4     | .0010 | .0112 | .0401 | .0881 | .1460 | .2001 | .2377 | .2508 | .2384 | .2051 | .1596 | .1115 | .0689 | .0368 | .0162 | .0055 | .0012 | .0001 | .0000 |
|    | 5     | .0001 | .0015 | .0085 | .0264 | .0584 | .1029 | .1536 | .2007 | .2340 | .2461 | .2340 | .2007 | .1536 | .1029 | .0584 | .0264 | .0085 | .0015 | .0001 |
|    | 6     | .0000 | .0001 | .0012 | .0055 | .0162 | .0368 | .0689 | .1115 | .1596 | .2051 | .2384 | .2508 | .2377 | .2001 | .1460 | .0881 | .0401 | .0112 | .0010 |
|    | 7     | .0000 | .0000 | .0001 | .0008 | .0031 | .0090 | .0212 | .0425 | .0746 | .1172 | .1665 | .2150 | .2522 | .2668 | .2503 | .2013 | .1298 | .0574 | .0105 |
|    | 8     | .0000 | .0000 | .0000 | .0001 | .0004 | .0014 | .0043 | .0106 | .0229 | .0439 | .0763 | .1209 | .1757 | .2335 | .2816 | .3020 | .2759 | .1937 | .0746 |
|    | 9     | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0005 | .0016 | .0042 | .0098 | .0207 | .0403 | .0725 | .1211 | .1877 | .2684 | .3474 | .3874 | .3151 |
|    | 10    | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0003 | .0010 | .0025 | .0060 | .0135 | .0282 | .0563 | .1074 | .1969 | .3487 | .5987 |
| 11 | 0     | .5688 | .3138 | .1673 | .0859 | .0422 | .0198 | .0088 | .0036 | .0014 | .0005 | .0002 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |       |
|    | 1     | .3293 | .3835 | .3248 | .2362 | .1549 | .0932 | .0518 | .0266 | .0125 | .0054 | .0021 | .0007 | .0002 | .0000 | .0000 | .0000 | .0000 | .0000 |       |
|    | 2     | .0867 | .2131 | .2866 | .2953 | .2581 | .1998 | .1395 | .0887 | .0513 | .0269 | .0126 | .0052 | .0018 | .0005 | .0001 | .0000 | .0000 | .0000 |       |
|    | 3     | .0137 | .0710 | .1517 | .2215 | .2581 | .2568 | .2254 | .1774 | .1259 | .0806 | .0462 | .0234 | .0102 | .0037 | .0011 | .0002 | .0000 | .0000 |       |
|    | 4     | .0014 | .0158 | .0536 | .1107 | .1721 | .2201 | .2428 | .2365 | .2060 | .1611 | .1128 | .0701 | .0379 | .0173 | .0064 | .0017 | .0003 | .0000 |       |
|    | 5     | .0001 | .0025 | .0132 | .0388 | .0803 | .1321 | .1830 | .2207 | .2360 | .2256 | .1931 | .1471 | .0985 | .0566 | .0268 | .0097 | .0023 | .0003 | .0000 |
|    | 6     | .0000 | .0003 | .0023 | .0097 | .0268 | .0566 | .0985 | .1471 | .1931 | .2256 | .2360 | .2270 | .1830 | .1321 | .0803 | .0388 | .0132 | .0025 | .0001 |
|    | 7     | .0000 | .0000 | .0003 | .0017 | .0064 | .0173 | .0379 | .0701 | .1128 | .1611 | .2060 | .2365 | .2428 | .2201 | .1721 | .1107 | .0536 | .0158 | .0014 |
|    | 8     | .0000 | .0000 | .0000 | .0002 | .0011 | .0037 | .0102 | .0234 | .0462 | .0806 | .1259 | .1774 | .2254 | .2568 | .2581 | .2215 | .1517 | .0710 | .0137 |
|    | 9     | .0000 | .0000 | .0000 | .0001 | .0005 | .0018 | .0052 | .0126 | .0269 | .0513 | .0887 | .1395 | .1998 | .2581 | .2953 | .2866 | .2131 | .0867 |       |
|    | 10    | .0000 | .0000 | .0000 | .0000 | .0000 | .0002 | .0007 | .0021 | .0054 | .0125 | .0266 | .0518 | .0932 | .1549 | .2362 | .3248 | .3835 | .3293 |       |
|    | 11    | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0002 | .0005 | .0014 | .0036 | .0088 | .0198 | .0422 | .0859 | .1673 | .3138 | .5688 |       |
| 12 | 0     | .5404 | .2824 | .1422 | .0687 | .0317 | .0138 | .0057 | .0022 | .0008 | .0002 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 |       |
|    | 1     | .3413 | .3766 | .3012 | .2062 | .1267 | .0712 | .0368 | .0174 | .0075 | .0029 | .0010 | .0003 | .0001 | .0000 | .0000 | .0000 | .0000 | .0000 |       |
|    | 2     | .0988 | .2301 | .2924 | .2835 | .2323 | .1678 | .1088 | .0639 | .0339 | .0161 | .0068 | .0025 | .0008 | .0002 | .0000 | .0000 | .0000 | .0000 |       |
|    | 3     | .0173 | .0852 | .1720 | .2362 | .2581 | .2397 | .1954 | .1419 | .0923 | .0537 | .0277 | .0125 | .0048 | .0015 | .0004 | .0001 | .0000 | .0000 |       |
|    | 4     | .0021 | .0213 | .0683 | .1329 | .1936 | .2311 | .2367 | .2128 | .1700 | .1208 | .0762 | .0420 | .0199 | .0078 | .0024 | .0005 | .0001 | .0000 |       |
|    | 5     | .0002 | .0038 | .0193 | .0532 | .1032 | .1585 | .2039 | .2270 | .2225 | .1934 | .1489 | .1009 | .0591 | .0291 | .0115 | .0033 | .0006 | .0000 |       |
|    | 6     | .0000 | .0005 | .0040 | .0155 | .0401 | .0792 | .1281 | .1766 | .2124 | .2256 | .2124 | .1766 | .1281 | .0792 | .0401 | .0155 | .0040 | .0005 |       |
|    | 7     | .0000 | .0000 | .0006 | .0033 | .0115 | .0291 | .0591 | .1009 | .1489 | .1934 | .2225 | .2270 | .2039 | .1585 | .1032 | .0532 | .0193 | .0038 | .0002 |
|    | 8     | .0000 | .0000 | .0001 | .0005 | .0024 | .0078 | .0199 | .0420 | .0762 | .1208 | .1700 | .2128 | .2367 | .2311 | .1936 | .1329 | .0683 | .0213 | .0021 |
|    | 9     | .0000 | .0000 | .0000 | .0001 | .0004 | .0015 | .0048 | .0125 | .0277 | .0537 | .0923 | .1419 | .1954 | .2397 | .2581 | .2362 | .1720 | .0852 | .0173 |
|    | 10    | .0000 | .0000 | .0000 | .0000 | .0002 | .0008 | .0025 | .0068 | .0161 | .0339 | .0639 | .1088 | .1678 | .2323 | .2835 | .2924 | .2301 | .0988 |       |
|    | 11    | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0003 | .0010 | .0029 | .0075 | .0174 | .0368 | .0712 | .1267 | .2062 | .3012 | .3766 | .3413 |       |
|    | 12    | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0000 | .0001 | .0002 | .0008 | .0022 | .0057 | .0138 | .0317 | .0687 | .1422 | .2824 | .5404 |       |

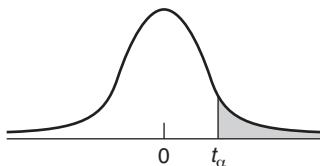
Fuente: Tabla A, páginas 464-466 en *Business Statistics: Concepts and Applications*, de William J. Stevenson. Derechos de autor © 1978 de William J. Stevenson. Reimpreso con autorización de Harper & Row, Editores.

**TABLA C-2 Tabla de las áreas para la distribución de probabilidad normal estándar**



Para  $z = 1.93$ , el área sombreada es 0.4732 de un área total de 1.

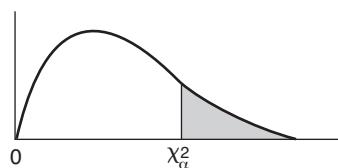
**TABLA C-3 Valores críticos de la distribución *t* de Student**



| <i>df</i> | $t_{.100}^2$ | $t_{.050}^2$ | $t_{.025}^2$ | $t_{.010}^2$ | $t_{.005}^2$ |
|-----------|--------------|--------------|--------------|--------------|--------------|
| 1         | 3.078        | 6.314        | 12.706       | 31.821       | 63.657       |
| 2         | 1.886        | 2.920        | 4.303        | 6.965        | 9.925        |
| 3         | 1.638        | 2.353        | 3.182        | 4.541        | 5.841        |
| 4         | 1.533        | 2.132        | 2.776        | 3.747        | 4.604        |
| 5         | 1.476        | 2.015        | 2.571        | 3.365        | 4.032        |
| 6         | 1.440        | 1.943        | 2.447        | 3.143        | 3.707        |
| 7         | 1.415        | 1.895        | 2.365        | 2.998        | 3.499        |
| 8         | 1.397        | 1.860        | 2.306        | 2.896        | 3.355        |
| 9         | 1.383        | 1.833        | 2.262        | 2.821        | 3.250        |
| 10        | 1.372        | 1.812        | 2.228        | 2.764        | 3.169        |
| 11        | 1.363        | 1.796        | 2.201        | 2.718        | 3.106        |
| 12        | 1.356        | 1.782        | 2.179        | 2.681        | 3.055        |
| 13        | 1.350        | 1.771        | 2.160        | 2.650        | 3.012        |
| 14        | 1.345        | 1.761        | 2.145        | 2.624        | 2.977        |
| 15        | 1.341        | 1.753        | 2.131        | 2.602        | 2.947        |
| 16        | 1.337        | 1.746        | 2.120        | 2.583        | 2.921        |
| 17        | 1.333        | 1.740        | 2.110        | 2.567        | 2.898        |
| 18        | 1.330        | 1.734        | 2.101        | 2.552        | 2.878        |
| 19        | 1.328        | 1.729        | 2.093        | 2.539        | 2.861        |
| 20        | 1.325        | 1.725        | 2.086        | 2.528        | 2.845        |
| 21        | 1.323        | 1.721        | 2.080        | 2.518        | 2.831        |
| 22        | 1.321        | 1.717        | 2.074        | 2.508        | 2.819        |
| 23        | 1.319        | 1.714        | 2.069        | 2.500        | 2.807        |
| 24        | 1.318        | 1.711        | 2.064        | 2.492        | 2.797        |
| 25        | 1.316        | 1.708        | 2.060        | 2.485        | 2.787        |
| 26        | 1.315        | 1.706        | 2.056        | 2.479        | 2.779        |
| 27        | 1.314        | 1.703        | 2.052        | 2.473        | 2.771        |
| 28        | 1.313        | 1.701        | 2.048        | 2.467        | 2.763        |
| 29        | 1.311        | 1.699        | 2.045        | 2.462        | 2.756        |
| inf.      | 1.282        | 1.645        | 1.960        | 2.326        | 2.576        |

Fuente: "Table of Percentage Points of the t-Distribution." Cálculos realizados por Maxine Merrington, *Biometrika*, vol. 32 (1941), p. 300. Reproducido con permiso del profesor D. V. Lindley.

TABLA C-4 Valores críticos de la distribución chi cuadrada

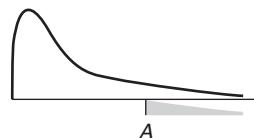


| <i>df</i> | $\chi^2_{.995}$ | $\chi^2_{.990}$ | $\chi^2_{.975}$ | $\chi^2_{.950}$ | $\chi^2_{.900}$ |
|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1         | 0.0000393       | 0.0001571       | 0.0009821       | 0.0039321       | 0.0157908       |
| 2         | 0.0100251       | 0.0201007       | 0.0506356       | 0.102587        | 0.210720        |
| 3         | 0.0717212       | 0.114832        | 0.215795        | 0.351846        | 0.584375        |
| 4         | 0.206990        | 0.297110        | 0.484419        | 0.710721        | 1.063623        |
| 5         | 0.411740        | 0.554300        | 0.831211        | 1.145476        | 1.61031         |
| 6         | 0.675727        | 0.872085        | 1.237347        | 1.63539         | 2.20413         |
| 7         | 0.989265        | 1.239043        | 1.68987         | 2.16735         | 2.83311         |
| 8         | 1.344419        | 1.646482        | 2.17973         | 2.73264         | 3.48954         |
| 9         | 1.734926        | 2.087912        | 2.70039         | 3.32511         | 4.168216        |
| 10        | 2.15585         | 2.55821         | 3.24697         | 3.94030         | 4.86518         |
| 11        | 2.60321         | 3.05347         | 3.81575         | 4.57481         | 5.57779         |
| 12        | 3.07382         | 3.57056         | 4.40379         | 5.22603         | 6.30380         |
| 13        | 3.56503         | 4.10691         | 5.00874         | 5.89186         | 7.04150         |
| 14        | 4.07468         | 4.66043         | 5.62872         | 6.57063         | 7.78953         |
| 15        | 4.60094         | 5.22935         | 6.26214         | 7.26094         | 8.54675         |
| 16        | 5.14224         | 5.81221         | 6.90766         | 7.96164         | 9.31223         |
| 17        | 5.69724         | 6.40776         | 7.56418         | 8.67176         | 10.0852         |
| 18        | 6.26481         | 7.01491         | 8.23075         | 9.39046         | 10.8649         |
| 19        | 6.84398         | 7.63273         | 8.90655         | 10.1170         | 11.6509         |
| 20        | 7.43386         | 8.26040         | 9.59083         | 10.8508         | 12.4426         |
| 21        | 8.03366         | 8.89720         | 10.28293        | 11.5913         | 13.2396         |
| 22        | 8.64272         | 9.54249         | 10.9823         | 12.3380         | 14.0415         |
| 23        | 9.26042         | 10.19567        | 11.6885         | 13.0905         | 14.8479         |
| 24        | 9.88623         | 10.8564         | 12.4011         | 13.8484         | 15.6587         |
| 25        | 10.5197         | 11.5240         | 13.1197         | 14.6114         | 16.4734         |
| 26        | 11.1603         | 12.1981         | 13.8439         | 15.3791         | 17.2919         |
| 27        | 11.8076         | 12.8786         | 14.5733         | 16.1513         | 18.1138         |
| 28        | 12.4613         | 13.5648         | 15.3079         | 16.9279         | 18.9302         |
| 29        | 13.1211         | 14.2565         | 16.0471         | 17.7083         | 19.7677         |
| 30        | 13.7867         | 14.9535         | 16.7908         | 18.4926         | 20.5992         |
| 40        | 20.7065         | 22.1643         | 24.4331         | 26.5093         | 29.0505         |
| 50        | 27.9907         | 29.7067         | 32.3574         | 34.7642         | 37.6886         |
| 60        | 35.5347         | 37.4848         | 40.4817         | 43.1879         | 46.4589         |
| 70        | 43.2752         | 45.4418         | 48.7576         | 51.7393         | 55.3290         |
| 80        | 51.1720         | 53.5400         | 57.1532         | 60.3915         | 64.2778         |
| 90        | 59.1963         | 61.7541         | 65.6466         | 69.1260         | 73.2912         |
| 100       | 67.3276         | 70.0648         | 74.2219         | 77.9295         | 82.3581         |

**TABLA C-4** *(Continúa)*

| <i>df</i> | $\chi^2_{.100}$ | $\chi^2_{.050}$ | $\chi^2_{.025}$ | $\chi^2_{.010}$ | $\chi^2_{.005}$ |
|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 1         | 2.70554         | 3.84146         | 5.02389         | 6.63490         | 7.87944         |
| 2         | 4.60517         | 5.99147         | 7.37776         | 9.21034         | 10.5966         |
| 3         | 6.25139         | 7.81473         | 9.34840         | 11.3449         | 12.8381         |
| 4         | 7.77944         | 9.48773         | 11.1433         | 13.2767         | 14.8602         |
| 5         | 9.23635         | 11.0705         | 12.8325         | 15.0863         | 16.7496         |
| 6         | 10.6446         | 12.5916         | 14.4494         | 16.8119         | 18.5476         |
| 7         | 12.0170         | 14.0671         | 16.0128         | 18.4753         | 20.2777         |
| 8         | 13.3616         | 15.5073         | 17.5346         | 20.0902         | 21.9550         |
| 9         | 14.6837         | 16.9190         | 19.0228         | 21.6660         | 23.5893         |
| 10        | 15.9871         | 18.3070         | 20.4831         | 23.2093         | 25.1882         |
| 11        | 17.2750         | 19.6751         | 21.9200         | 24.7250         | 26.7569         |
| 12        | 18.5494         | 21.0261         | 23.3367         | 26.2170         | 28.2995         |
| 13        | 19.8119         | 22.3621         | 24.7356         | 27.6883         | 29.8194         |
| 14        | 21.0642         | 23.6848         | 26.1190         | 29.1413         | 31.3193         |
| 15        | 22.3072         | 24.9958         | 27.4884         | 30.5779         | 32.8013         |
| 16        | 23.5418         | 26.2962         | 28.8454         | 31.9999         | 34.2672         |
| 17        | 24.7690         | 27.5871         | 30.1910         | 33.4087         | 35.7185         |
| 18        | 25.9894         | 28.8693         | 31.5264         | 34.8053         | 37.1564         |
| 19        | 27.2036         | 30.1435         | 32.8523         | 36.1908         | 38.5822         |
| 20        | 28.4120         | 31.4104         | 34.1696         | 37.5662         | 39.9968         |
| 21        | 29.6151         | 32.6705         | 35.4789         | 38.9321         | 41.4010         |
| 22        | 30.8133         | 33.9244         | 36.7807         | 40.2894         | 42.7956         |
| 23        | 32.0069         | 35.1725         | 38.0757         | 41.6384         | 44.1813         |
| 24        | 33.1963         | 36.4151         | 39.3641         | 42.9798         | 45.5585         |
| 25        | 34.3816         | 37.6525         | 40.6465         | 44.3141         | 46.9278         |
| 26        | 35.5631         | 38.8852         | 41.9232         | 45.6417         | 48.2899         |
| 27        | 36.7412         | 40.1133         | 43.1944         | 46.9630         | 49.6449         |
| 28        | 37.9159         | 41.3372         | 44.4607         | 48.2782         | 50.9933         |
| 29        | 39.0875         | 42.5569         | 45.7222         | 49.5879         | 52.3356         |
| 30        | 40.2560         | 43.7729         | 46.9792         | 50.8922         | 53.6720         |
| 40        | 51.8050         | 55.7585         | 59.3417         | 63.6907         | 66.7659         |
| 50        | 63.1671         | 67.5048         | 71.4202         | 76.1539         | 79.4900         |
| 60        | 74.3970         | 79.0819         | 83.2976         | 88.3794         | 91.9517         |
| 70        | 85.5271         | 90.5312         | 95.0231         | 100.425         | 104.215         |
| 80        | 96.5782         | 101.879         | 106.629         | 112.329         | 116.321         |
| 90        | 107.565         | 113.145         | 118.136         | 124.116         | 128.299         |
| 100       | 118.498         | 124.342         | 129.561         | 135.807         | 140.169         |

Fuente: "Tables of the Percentage Points of the  $\chi^2$ -Distribution", *Biometrika*, vol. 32 (1941), pp. 188-189, de Catherine M. Thompson. Reproducida con permiso del profesor D. V. Lindley.

TABLA C-5 Tabla de la distribución  $F$ 

Por ejemplo, el valor de escala de  $F$  para  $\delta_1 = 3$ ,  $\delta_2 = 10$  correspondiente al área 0.01 en la cola derecha es 6.55.

El valor  $F$  correspondiente al área 0.05 en la cola derecha se muestra en tipografía normal.

**El valor  $F$  correspondiente al área 0.01 en la cola derecha se muestra en tipografía negrita.**

| $\delta_2$ ,<br>Grados de<br>libertad del<br>denominador | $\delta_1$ , Grados de libertad del numerador |              |              |              |              |              |              |              |              |              |
|----------------------------------------------------------|-----------------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                                                          | 1                                             | 2            | 3            | 4            | 5            | 6            | 7            | 8            | 9            | 10           |
| 1                                                        | 161                                           | 200          | 216          | 225          | 230          | 234          | 237          | 239          | 241          | 242          |
|                                                          | <b>4,052</b>                                  | <b>4,999</b> | <b>5,403</b> | <b>5,625</b> | <b>5,764</b> | <b>5,859</b> | <b>5,928</b> | <b>5,981</b> | <b>6,022</b> | <b>6,056</b> |
| 2                                                        | 18.51                                         | 19.00        | 19.16        | 19.25        | 19.30        | 19.33        | 19.36        | 19.37        | 19.38        | 19.39        |
|                                                          | <b>98.49</b>                                  | <b>99.00</b> | <b>99.17</b> | <b>99.25</b> | <b>99.30</b> | <b>99.33</b> | <b>99.36</b> | <b>99.37</b> | <b>99.39</b> | <b>99.40</b> |
| 3                                                        | 10.13                                         | 9.55         | 9.28         | 9.12         | 9.01         | 8.94         | 8.88         | 8.84         | 8.81         | 8.78         |
|                                                          | <b>34.12</b>                                  | <b>30.82</b> | <b>29.46</b> | <b>28.71</b> | <b>28.24</b> | <b>27.91</b> | <b>27.67</b> | <b>27.49</b> | <b>27.34</b> | <b>27.23</b> |
| 4                                                        | 7.71                                          | 6.94         | 6.59         | 6.39         | 6.26         | 6.16         | 6.09         | 6.04         | 6.00         | 5.96         |
|                                                          | <b>21.20</b>                                  | <b>18.00</b> | <b>16.69</b> | <b>15.98</b> | <b>15.52</b> | <b>15.21</b> | <b>14.98</b> | <b>14.80</b> | <b>14.66</b> | <b>14.54</b> |
| 5                                                        | 6.61                                          | 5.79         | 5.41         | 5.19         | 5.05         | 4.95         | 4.88         | 4.82         | 4.78         | 4.74         |
|                                                          | <b>16.26</b>                                  | <b>13.27</b> | <b>12.06</b> | <b>11.39</b> | <b>10.97</b> | <b>10.67</b> | <b>10.45</b> | <b>10.29</b> | <b>10.15</b> | <b>10.05</b> |
| 6                                                        | 5.99                                          | 5.14         | 4.76         | 4.53         | 4.39         | 4.28         | 4.21         | 4.15         | 4.10         | 4.06         |
|                                                          | <b>13.74</b>                                  | <b>10.92</b> | <b>9.78</b>  | <b>9.15</b>  | <b>8.75</b>  | <b>8.47</b>  | <b>8.26</b>  | <b>8.10</b>  | <b>7.98</b>  | <b>7.87</b>  |
| 7                                                        | 5.59                                          | 4.74         | 4.35         | 4.12         | 3.97         | 3.87         | 3.79         | 3.73         | 3.68         | 3.63         |
|                                                          | <b>12.25</b>                                  | <b>9.55</b>  | <b>8.45</b>  | <b>7.85</b>  | <b>7.46</b>  | <b>7.19</b>  | <b>7.00</b>  | <b>6.84</b>  | <b>6.71</b>  | <b>6.62</b>  |
| 8                                                        | 5.32                                          | 4.46         | 4.07         | 3.84         | 3.69         | 3.58         | 3.50         | 3.44         | 3.39         | 3.34         |
|                                                          | <b>11.26</b>                                  | <b>8.65</b>  | <b>7.59</b>  | <b>7.01</b>  | <b>6.63</b>  | <b>6.37</b>  | <b>6.19</b>  | <b>6.03</b>  | <b>5.91</b>  | <b>5.82</b>  |
| 9                                                        | 5.12                                          | 4.26         | 3.86         | 3.63         | 3.48         | 3.37         | 3.29         | 3.23         | 3.18         | 3.13         |
|                                                          | <b>10.56</b>                                  | <b>8.02</b>  | <b>6.99</b>  | <b>6.42</b>  | <b>6.06</b>  | <b>5.80</b>  | <b>5.62</b>  | <b>5.47</b>  | <b>5.35</b>  | <b>5.26</b>  |
| 10                                                       | 4.96                                          | 4.10         | 3.71         | 3.48         | 3.33         | 3.22         | 3.14         | 3.07         | 3.02         | 2.97         |
|                                                          | <b>10.04</b>                                  | <b>7.56</b>  | <b>6.55</b>  | <b>5.99</b>  | <b>5.64</b>  | <b>5.39</b>  | <b>5.21</b>  | <b>5.06</b>  | <b>4.95</b>  | <b>4.85</b>  |
| 11                                                       | 4.84                                          | 3.98         | 3.59         | 3.36         | 3.20         | 3.09         | 3.01         | 2.95         | 2.90         | 2.86         |
|                                                          | <b>9.65</b>                                   | <b>7.20</b>  | <b>6.22</b>  | <b>5.67</b>  | <b>5.32</b>  | <b>5.07</b>  | <b>4.88</b>  | <b>4.74</b>  | <b>4.63</b>  | <b>4.54</b>  |
| 12                                                       | 4.75                                          | 3.88         | 3.49         | 3.26         | 3.11         | 3.00         | 2.92         | 2.85         | 2.80         | 2.76         |
|                                                          | <b>9.33</b>                                   | <b>6.93</b>  | <b>5.95</b>  | <b>5.41</b>  | <b>5.06</b>  | <b>4.82</b>  | <b>4.65</b>  | <b>4.50</b>  | <b>4.39</b>  | <b>4.30</b>  |
| 13                                                       | 4.67                                          | 3.80         | 3.41         | 3.18         | 3.02         | 2.92         | 2.84         | 2.77         | 2.72         | 2.67         |
|                                                          | <b>9.07</b>                                   | <b>6.70</b>  | <b>5.74</b>  | <b>5.20</b>  | <b>4.86</b>  | <b>4.62</b>  | <b>4.44</b>  | <b>4.30</b>  | <b>4.19</b>  | <b>4.10</b>  |
| 14                                                       | 4.60                                          | 3.74         | 3.34         | 3.11         | 2.96         | 2.85         | 2.77         | 2.70         | 2.65         | 2.60         |
|                                                          | <b>8.86</b>                                   | <b>6.51</b>  | <b>5.56</b>  | <b>5.03</b>  | <b>4.69</b>  | <b>4.46</b>  | <b>4.28</b>  | <b>4.14</b>  | <b>4.03</b>  | <b>3.94</b>  |
| 15                                                       | 4.54                                          | 3.68         | 3.29         | 3.06         | 2.90         | 2.79         | 2.70         | 2.64         | 2.59         | 2.55         |
|                                                          | <b>8.68</b>                                   | <b>6.36</b>  | <b>5.42</b>  | <b>4.89</b>  | <b>4.56</b>  | <b>4.32</b>  | <b>4.14</b>  | <b>4.00</b>  | <b>3.89</b>  | <b>3.80</b>  |
| 16                                                       | 4.49                                          | 3.36         | 3.24         | 3.01         | 2.85         | 2.74         | 2.66         | 2.59         | 2.54         | 2.49         |
|                                                          | <b>8.53</b>                                   | <b>6.23</b>  | <b>5.29</b>  | <b>4.77</b>  | <b>4.44</b>  | <b>4.20</b>  | <b>4.03</b>  | <b>3.89</b>  | <b>3.78</b>  | <b>3.69</b>  |
| 17                                                       | 4.45                                          | 3.59         | 3.20         | 2.96         | 2.81         | 2.70         | 2.62         | 2.55         | 2.50         | 2.45         |
|                                                          | <b>8.40</b>                                   | <b>6.11</b>  | <b>5.18</b>  | <b>4.67</b>  | <b>4.34</b>  | <b>4.10</b>  | <b>3.93</b>  | <b>3.79</b>  | <b>3.68</b>  | <b>3.59</b>  |
| 18                                                       | 4.41                                          | 3.55         | 3.16         | 2.93         | 2.77         | 2.66         | 2.58         | 2.51         | 2.46         | 2.41         |
|                                                          | <b>8.28</b>                                   | <b>6.01</b>  | <b>5.09</b>  | <b>4.58</b>  | <b>4.25</b>  | <b>4.01</b>  | <b>3.85</b>  | <b>3.71</b>  | <b>3.60</b>  | <b>3.51</b>  |
| 19                                                       | 4.38                                          | 3.52         | 3.13         | 2.90         | 2.74         | 2.63         | 2.55         | 2.48         | 2.43         | 2.38         |
|                                                          | <b>8.18</b>                                   | <b>5.93</b>  | <b>5.01</b>  | <b>4.50</b>  | <b>4.17</b>  | <b>3.94</b>  | <b>3.77</b>  | <b>3.63</b>  | <b>3.52</b>  | <b>3.43</b>  |

**TABLA C-5** *(Continúa)*

| $\delta_2$<br><i>Grados de<br/>libertad del<br/>denominador</i> | $\delta_1$ , <i>Grados de libertad del numerador</i> |             |             |             |             |             |             |             |             |             |
|-----------------------------------------------------------------|------------------------------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                                                                 | 1                                                    | 2           | 3           | 4           | 5           | 6           | 7           | 8           | 9           | 10          |
| 20                                                              | 4.35                                                 | 3.49        | 3.10        | 2.87        | 2.71        | 2.60        | 2.52        | 2.45        | 2.40        | 2.35        |
|                                                                 | <b>8.10</b>                                          | <b>5.85</b> | <b>4.94</b> | <b>4.43</b> | <b>4.10</b> | <b>3.87</b> | <b>3.71</b> | <b>3.56</b> | <b>3.45</b> | <b>3.37</b> |
| 21                                                              | 4.32                                                 | 3.47        | 3.07        | 2.84        | 2.68        | 2.57        | 2.49        | 2.42        | 2.37        | 2.32        |
|                                                                 | <b>8.02</b>                                          | <b>5.78</b> | <b>4.87</b> | <b>4.37</b> | <b>4.04</b> | <b>3.81</b> | <b>3.65</b> | <b>3.51</b> | <b>3.40</b> | <b>3.31</b> |
| 22                                                              | 4.30                                                 | 3.44        | 3.05        | 2.82        | 2.66        | 2.55        | 2.47        | 2.40        | 2.35        | 2.30        |
|                                                                 | <b>7.94</b>                                          | <b>5.72</b> | <b>4.82</b> | <b>4.31</b> | <b>3.99</b> | <b>3.76</b> | <b>3.59</b> | <b>3.45</b> | <b>3.35</b> | <b>3.26</b> |
| 23                                                              | 4.28                                                 | 3.42        | 3.03        | 2.80        | 2.64        | 2.53        | 2.45        | 2.38        | 2.32        | 2.28        |
|                                                                 | <b>7.88</b>                                          | <b>5.66</b> | <b>4.76</b> | <b>4.26</b> | <b>3.94</b> | <b>3.71</b> | <b>3.54</b> | <b>3.41</b> | <b>3.30</b> | <b>3.21</b> |
| 24                                                              | 4.26                                                 | 3.40        | 3.01        | 2.78        | 2.62        | 2.51        | 2.43        | 2.36        | 2.30        | 2.26        |
|                                                                 | <b>7.82</b>                                          | <b>5.61</b> | <b>4.72</b> | <b>4.22</b> | <b>3.90</b> | <b>3.67</b> | <b>3.50</b> | <b>3.36</b> | <b>3.25</b> | <b>3.17</b> |
| 25                                                              | 4.24                                                 | 3.38        | 2.99        | 2.76        | 2.60        | 2.49        | 2.41        | 2.34        | 2.28        | 2.24        |
|                                                                 | <b>7.77</b>                                          | <b>5.57</b> | <b>4.68</b> | <b>4.18</b> | <b>3.86</b> | <b>3.63</b> | <b>3.46</b> | <b>3.32</b> | <b>3.21</b> | <b>3.13</b> |

Fuente: Abreviada con permiso de *Statistical Methods*, séptima edición, de George W. Snedecor y William C. Cochran. Derechos de autor © 1980 de la Iowa State University Press, Ames, Iowa.

**TABLA C-6** *Límites en la prueba Durbin-Watson*

| <i>n</i> | <i>Nivel de significancia <math>\alpha = 0.05</math></i> |                      |                      |                      |                      |                      |                      |                      |                      |                      |
|----------|----------------------------------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|          | <i>k = 1</i>                                             |                      | <i>k = 2</i>         |                      | <i>k = 3</i>         |                      | <i>k = 4</i>         |                      | <i>k = 5</i>         |                      |
|          | <i>d<sub>L</sub></i>                                     | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> |
| 15       | 1.08                                                     | 1.36                 | 0.95                 | 1.54                 | 0.82                 | 1.75                 | 0.69                 | 1.97                 | 0.56                 | 2.21                 |
| 16       | 1.10                                                     | 1.37                 | 0.98                 | 1.54                 | 0.86                 | 1.73                 | 0.74                 | 1.93                 | 0.62                 | 2.15                 |
| 17       | 1.13                                                     | 1.38                 | 1.02                 | 1.54                 | 0.90                 | 1.71                 | 0.78                 | 1.90                 | 0.67                 | 2.10                 |
| 18       | 1.16                                                     | 1.39                 | 1.05                 | 1.53                 | 0.93                 | 1.69                 | 0.82                 | 1.87                 | 0.71                 | 2.06                 |
| 19       | 1.18                                                     | 1.40                 | 1.08                 | 1.53                 | 0.97                 | 1.68                 | 0.86                 | 1.85                 | 0.75                 | 2.02                 |
| 20       | 1.20                                                     | 1.41                 | 1.10                 | 1.54                 | 1.00                 | 1.68                 | 0.90                 | 1.83                 | 0.79                 | 1.99                 |
| 21       | 1.22                                                     | 1.42                 | 1.13                 | 1.54                 | 1.03                 | 1.67                 | 0.93                 | 1.81                 | 0.83                 | 1.96                 |
| 22       | 1.24                                                     | 1.43                 | 1.15                 | 1.54                 | 1.05                 | 1.66                 | 0.96                 | 1.80                 | 0.86                 | 1.94                 |
| 23       | 1.26                                                     | 1.44                 | 1.17                 | 1.54                 | 1.08                 | 1.66                 | 0.99                 | 1.79                 | 0.90                 | 1.92                 |
| 24       | 1.27                                                     | 1.45                 | 1.19                 | 1.55                 | 1.10                 | 1.66                 | 1.01                 | 1.78                 | 0.93                 | 1.90                 |
| 25       | 1.29                                                     | 1.45                 | 1.21                 | 1.55                 | 1.12                 | 1.66                 | 1.04                 | 1.77                 | 0.95                 | 1.89                 |
| 26       | 1.30                                                     | 1.46                 | 1.22                 | 1.55                 | 1.14                 | 1.65                 | 1.06                 | 1.76                 | 0.98                 | 1.88                 |
| 27       | 1.32                                                     | 1.47                 | 1.24                 | 1.56                 | 1.16                 | 1.65                 | 1.08                 | 1.76                 | 1.01                 | 1.86                 |
| 28       | 1.33                                                     | 1.48                 | 1.26                 | 1.56                 | 1.18                 | 1.65                 | 1.10                 | 1.75                 | 1.03                 | 1.85                 |
| 29       | 1.34                                                     | 1.48                 | 1.27                 | 1.56                 | 1.20                 | 1.65                 | 1.12                 | 1.74                 | 1.05                 | 1.84                 |
| 30       | 1.35                                                     | 1.49                 | 1.28                 | 1.57                 | 1.21                 | 1.65                 | 1.14                 | 1.74                 | 1.07                 | 1.83                 |
| 31       | 1.36                                                     | 1.50                 | 1.30                 | 1.57                 | 1.23                 | 1.65                 | 1.16                 | 1.74                 | 1.09                 | 1.83                 |
| 32       | 1.37                                                     | 1.50                 | 1.31                 | 1.57                 | 1.24                 | 1.65                 | 1.18                 | 1.73                 | 1.11                 | 1.82                 |
| 33       | 1.38                                                     | 1.51                 | 1.32                 | 1.58                 | 1.26                 | 1.65                 | 1.19                 | 1.73                 | 1.13                 | 1.81                 |
| 34       | 1.39                                                     | 1.51                 | 1.33                 | 1.58                 | 1.27                 | 1.65                 | 1.21                 | 1.73                 | 1.15                 | 1.81                 |
| 35       | 1.40                                                     | 1.52                 | 1.34                 | 1.58                 | 1.28                 | 1.65                 | 1.22                 | 1.73                 | 1.16                 | 1.80                 |
| 36       | 1.41                                                     | 1.52                 | 1.35                 | 1.59                 | 1.29                 | 1.65                 | 1.24                 | 1.73                 | 1.18                 | 1.80                 |
| 37       | 1.42                                                     | 1.53                 | 1.36                 | 1.59                 | 1.31                 | 1.66                 | 1.25                 | 1.72                 | 1.19                 | 1.80                 |

TABLA C-6 (Continúa)

| <i>n</i>                                      | <i>Nivel de significancia</i> $\alpha = 0.05$ |                      |                      |                      |                      |                      |                      |                      |                      |                      |
|-----------------------------------------------|-----------------------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|                                               | <i>k = 1</i>                                  |                      | <i>k = 2</i>         |                      | <i>k = 3</i>         |                      | <i>k = 4</i>         |                      | <i>k = 5</i>         |                      |
|                                               | <i>d<sub>L</sub></i>                          | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> |
| 38                                            | 1.43                                          | 1.54                 | 1.37                 | 1.59                 | 1.32                 | 1.66                 | 1.26                 | 1.72                 | 1.21                 | 1.79                 |
| 39                                            | 1.43                                          | 1.54                 | 1.38                 | 1.60                 | 1.33                 | 1.66                 | 1.27                 | 1.72                 | 1.22                 | 1.79                 |
| 40                                            | 1.44                                          | 1.54                 | 1.39                 | 1.60                 | 1.34                 | 1.66                 | 1.29                 | 1.72                 | 1.23                 | 1.79                 |
| 45                                            | 1.48                                          | 1.57                 | 1.43                 | 1.62                 | 1.38                 | 1.67                 | 1.34                 | 1.72                 | 1.29                 | 1.78                 |
| 50                                            | 1.50                                          | 1.59                 | 1.46                 | 1.63                 | 1.42                 | 1.67                 | 1.38                 | 1.72                 | 1.34                 | 1.77                 |
| 55                                            | 1.53                                          | 1.60                 | 1.49                 | 1.64                 | 1.45                 | 1.68                 | 1.41                 | 1.72                 | 1.38                 | 1.77                 |
| 60                                            | 1.55                                          | 1.62                 | 1.51                 | 1.65                 | 1.48                 | 1.69                 | 1.44                 | 1.73                 | 1.41                 | 1.77                 |
| 65                                            | 1.57                                          | 1.63                 | 1.54                 | 1.66                 | 1.50                 | 1.70                 | 1.47                 | 1.73                 | 1.44                 | 1.77                 |
| 70                                            | 1.58                                          | 1.64                 | 1.55                 | 1.67                 | 1.52                 | 1.70                 | 1.49                 | 1.74                 | 1.46                 | 1.77                 |
| 75                                            | 1.60                                          | 1.65                 | 1.57                 | 1.68                 | 1.54                 | 1.71                 | 1.51                 | 1.74                 | 1.49                 | 1.77                 |
| 80                                            | 1.61                                          | 1.66                 | 1.59                 | 1.69                 | 1.56                 | 1.72                 | 1.53                 | 1.74                 | 1.51                 | 1.77                 |
| 85                                            | 1.62                                          | 1.67                 | 1.60                 | 1.70                 | 1.57                 | 1.72                 | 1.55                 | 1.75                 | 1.52                 | 1.77                 |
| 90                                            | 1.63                                          | 1.68                 | 1.61                 | 1.70                 | 1.59                 | 1.73                 | 1.57                 | 1.75                 | 1.54                 | 1.78                 |
| 95                                            | 1.64                                          | 1.69                 | 1.62                 | 1.71                 | 1.60                 | 1.73                 | 1.58                 | 1.75                 | 1.56                 | 1.78                 |
| 100                                           | 1.65                                          | 1.69                 | 1.63                 | 1.72                 | 1.61                 | 1.74                 | 1.59                 | 1.76                 | 1.57                 | 1.78                 |
| <i>Nivel de significancia</i> $\alpha = 0.01$ |                                               |                      |                      |                      |                      |                      |                      |                      |                      |                      |
| <i>n</i>                                      | <i>k = 1</i>                                  |                      | <i>k = 2</i>         |                      | <i>k = 3</i>         |                      | <i>k = 4</i>         |                      | <i>k = 5</i>         |                      |
|                                               | <i>d<sub>L</sub></i>                          | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> | <i>d<sub>L</sub></i> | <i>d<sub>U</sub></i> |
| 15                                            | 0.81                                          | 1.07                 | 0.70                 | 1.25                 | 0.59                 | 1.46                 | 0.49                 | 1.70                 | 0.39                 | 1.96                 |
| 16                                            | 0.84                                          | 1.09                 | 0.74                 | 1.25                 | 0.63                 | 1.44                 | 0.53                 | 1.66                 | 0.44                 | 1.90                 |
| 17                                            | 0.87                                          | 1.10                 | 0.77                 | 1.25                 | 0.67                 | 1.43                 | 0.57                 | 1.63                 | 0.48                 | 1.85                 |
| 18                                            | 0.90                                          | 1.12                 | 0.80                 | 1.26                 | 0.71                 | 1.42                 | 0.61                 | 1.60                 | 0.52                 | 1.80                 |
| 19                                            | 0.93                                          | 1.13                 | 0.83                 | 1.26                 | 0.74                 | 1.41                 | 0.65                 | 1.58                 | 0.56                 | 1.77                 |
| 20                                            | 0.95                                          | 1.15                 | 0.86                 | 1.27                 | 0.77                 | 1.41                 | 0.68                 | 1.57                 | 0.60                 | 1.74                 |
| 21                                            | 0.97                                          | 1.16                 | 0.89                 | 1.27                 | 0.80                 | 1.41                 | 0.72                 | 1.55                 | 0.63                 | 1.71                 |
| 22                                            | 1.00                                          | 1.17                 | 0.91                 | 1.28                 | 0.83                 | 1.40                 | 0.75                 | 1.54                 | 0.66                 | 1.69                 |
| 23                                            | 1.02                                          | 1.19                 | 0.94                 | 1.29                 | 0.86                 | 1.40                 | 0.77                 | 1.53                 | 0.70                 | 1.67                 |
| 24                                            | 1.04                                          | 1.20                 | 0.96                 | 1.30                 | 0.88                 | 1.41                 | 0.80                 | 1.53                 | 0.72                 | 1.66                 |
| 25                                            | 1.05                                          | 1.21                 | 0.98                 | 1.30                 | 0.90                 | 1.41                 | 0.83                 | 1.52                 | 0.75                 | 1.65                 |
| 26                                            | 1.07                                          | 1.22                 | 1.00                 | 1.31                 | 0.93                 | 1.41                 | 0.85                 | 1.52                 | 0.78                 | 1.64                 |
| 27                                            | 1.09                                          | 1.23                 | 1.02                 | 1.32                 | 0.95                 | 1.41                 | 0.88                 | 1.51                 | 0.81                 | 1.63                 |
| 28                                            | 1.10                                          | 1.24                 | 1.04                 | 1.32                 | 0.97                 | 1.41                 | 0.90                 | 1.51                 | 0.83                 | 1.62                 |
| 29                                            | 1.12                                          | 1.25                 | 1.05                 | 1.33                 | 0.99                 | 1.42                 | 0.92                 | 1.51                 | 0.85                 | 1.61                 |
| 30                                            | 1.13                                          | 1.26                 | 1.07                 | 1.34                 | 1.01                 | 1.42                 | 0.94                 | 1.51                 | 0.88                 | 1.61                 |
| 31                                            | 1.15                                          | 1.27                 | 1.08                 | 1.34                 | 1.02                 | 1.42                 | 0.96                 | 1.51                 | 0.90                 | 1.60                 |
| 32                                            | 1.16                                          | 1.28                 | 1.10                 | 1.35                 | 1.04                 | 1.43                 | 0.98                 | 1.51                 | 0.92                 | 1.60                 |
| 33                                            | 1.17                                          | 1.29                 | 1.11                 | 1.36                 | 1.05                 | 1.43                 | 1.00                 | 1.51                 | 0.94                 | 1.59                 |
| 34                                            | 1.18                                          | 1.30                 | 1.13                 | 1.36                 | 1.07                 | 1.43                 | 1.01                 | 1.51                 | 0.95                 | 1.59                 |
| 35                                            | 1.19                                          | 1.31                 | 1.14                 | 1.37                 | 1.08                 | 1.44                 | 1.03                 | 1.51                 | 0.97                 | 1.59                 |
| 36                                            | 1.21                                          | 1.32                 | 1.15                 | 1.38                 | 1.10                 | 1.44                 | 1.04                 | 1.51                 | 0.99                 | 1.59                 |
| 37                                            | 1.22                                          | 1.32                 | 1.16                 | 1.38                 | 1.11                 | 1.45                 | 1.06                 | 1.51                 | 1.00                 | 1.59                 |
| 38                                            | 1.23                                          | 1.33                 | 1.18                 | 1.39                 | 1.12                 | 1.45                 | 1.07                 | 1.52                 | 1.02                 | 1.58                 |
| 39                                            | 1.24                                          | 1.34                 | 1.19                 | 1.39                 | 1.14                 | 1.45                 | 1.09                 | 1.52                 | 1.03                 | 1.58                 |
| 40                                            | 1.25                                          | 1.34                 | 1.20                 | 1.40                 | 1.15                 | 1.46                 | 1.10                 | 1.52                 | 1.05                 | 1.58                 |

**TABLA C-6 (Continúa)**

| n   | Nivel de significancia $\alpha = 0.01$ |                |                |                |                |                |                |                |                |                |
|-----|----------------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|     | n = 1                                  |                | n = 2          |                | n = 3          |                | n = 4          |                | n = 5          |                |
|     | d <sub>L</sub>                         | d <sub>U</sub> | d <sub>L</sub> | d <sub>U</sub> | d <sub>L</sub> | d <sub>U</sub> | d <sub>L</sub> | d <sub>U</sub> | d <sub>L</sub> | d <sub>U</sub> |
| 45  | 1.29                                   | 1.38           | 1.24           | 1.42           | 1.20           | 1.48           | 1.16           | 1.53           | 1.11           | 1.58           |
| 50  | 1.32                                   | 1.40           | 1.28           | 1.45           | 1.24           | 1.49           | 1.20           | 1.54           | 1.16           | 1.59           |
| 55  | 1.36                                   | 1.43           | 1.32           | 1.47           | 1.28           | 1.51           | 1.25           | 1.55           | 1.21           | 1.59           |
| 60  | 1.38                                   | 1.45           | 1.35           | 1.48           | 1.32           | 1.52           | 1.28           | 1.56           | 1.25           | 1.60           |
| 65  | 1.41                                   | 1.47           | 1.38           | 1.50           | 1.35           | 1.53           | 1.31           | 1.57           | 1.28           | 1.61           |
| 70  | 1.43                                   | 1.49           | 1.40           | 1.52           | 1.37           | 1.55           | 1.34           | 1.58           | 1.31           | 1.61           |
| 75  | 1.45                                   | 1.50           | 1.42           | 1.53           | 1.39           | 1.56           | 1.37           | 1.59           | 1.34           | 1.62           |
| 80  | 1.47                                   | 1.52           | 1.44           | 1.54           | 1.42           | 1.57           | 1.39           | 1.60           | 1.36           | 1.62           |
| 85  | 1.48                                   | 1.53           | 1.46           | 1.55           | 1.43           | 1.58           | 1.41           | 1.60           | 1.39           | 1.63           |
| 90  | 1.50                                   | 1.54           | 1.47           | 1.56           | 1.45           | 1.59           | 1.43           | 1.61           | 1.41           | 1.64           |
| 95  | 1.51                                   | 1.55           | 1.49           | 1.57           | 1.47           | 1.60           | 1.45           | 1.62           | 1.42           | 1.64           |
| 100 | 1.52                                   | 1.56           | 1.50           | 1.58           | 1.48           | 1.60           | 1.46           | 1.63           | 1.44           | 1.65           |

k = Número de variables independientes.

Fuente: Reimpresa, con permiso, de J. Durbin y G. S. Watson, "Testing for Serial Correlation in Least Squares Regresión – II." *Biométrika*, vol. 38 (1951), pp. 159-178.

## Apéndice D

# Series de datos y bases de datos

Este apéndice contiene 18 datos de series de tiempo y tres bases de datos de regresión múltiple. La primera base de datos involucra siete variables para 25 campamentos, la segunda tiene nueve variables financieras medidas en 266 corporaciones estadounidenses en un año reciente,<sup>1</sup> y la tercera base de datos tiene 13 variables relacionadas con la compensación de un director ejecutivo medidas en 50 compañías estadounidenses en un año reciente.

Las series de datos para la mayoría de los ejemplos de texto, problemas y casos están disponibles en el CD incluido en este libro y en la Internet. Cada uno contiene tres versiones (Minitab, Excel y todos los demás programas). Para tener acceso a estas series de datos en la Internet, vaya al sitio de Prentice Hall [www.prenhall.com/Hanke](http://www.prenhall.com/Hanke) y siga las instrucciones de descarga.

### *Series de datos*

| Año  | Población total | Poder adquisitivo del dólar, precios al consumidor | Índice del salario mínimo federal, dólares actuales |
|------|-----------------|----------------------------------------------------|-----------------------------------------------------|
| 1960 | 180,671         | 3.373                                              | 1.00                                                |
| 1961 | 183,691         | 3.340                                              | 1.15                                                |
| 1962 | 186,538         | 3.304                                              | 1.15                                                |
| 1963 | 189,242         | 3.265                                              | 1.25                                                |
| 1964 | 191,889         | 3.220                                              | 1.25                                                |
| 1965 | 194,303         | 3.166                                              | 1.25                                                |
| 1966 | 196,560         | 3.080                                              | 1.25                                                |
| 1967 | 198,712         | 2.993                                              | 1.40                                                |
| 1968 | 200,706         | 2.873                                              | 1.60                                                |
| 1969 | 202,677         | 2.726                                              | 1.60                                                |
| 1970 | 205,052         | 2.574                                              | 1.60                                                |
| 1971 | 207,661         | 2.466                                              | 1.60                                                |
| 1972 | 209,896         | 2.391                                              | 1.60                                                |
| 1973 | 211,909         | 2.251                                              | 1.60                                                |
| 1974 | 213,854         | 2.029                                              | 2.00                                                |
| 1975 | 215,973         | 1.859                                              | 2.10                                                |
| 1976 | 218,035         | 1.757                                              | 2.30                                                |
| 1977 | 220,239         | 1.649                                              | 2.30                                                |
| 1978 | 222,585         | 1.532                                              | 2.65                                                |
| 1979 | 225,055         | 1.380                                              | 2.90                                                |
| 1980 | 227,726         | 1.215                                              | 3.10                                                |
| 1981 | 229,966         | 1.098                                              | 3.35                                                |
| 1982 | 232,188         | 1.035                                              | 3.35                                                |
| 1983 | 234,307         | 1.003                                              | 3.35                                                |

### *Series de datos (cont.)*

| Año  | Población total           | Poder adquisitivo del dólar, precios al consumidor | Índice del salario mínimo federal, dólares actuales |
|------|---------------------------|----------------------------------------------------|-----------------------------------------------------|
| 1984 | 236,348                   | 0.961                                              | 3.35                                                |
| 1985 | 238,466                   | 0.928                                              | 3.35                                                |
| 1986 | 240,651                   | 0.913                                              | 3.35                                                |
| 1987 | 242,804                   | 0.880                                              | 3.35                                                |
| 1988 | 245,021                   | 0.846                                              | 3.35                                                |
| 1989 | 247,342                   | 0.807                                              | 3.35                                                |
| 1990 | 249,948                   | 0.766                                              | 3.80                                                |
| 1991 | 252,639                   | 0.734                                              | 4.25                                                |
| 1992 | 255,374                   | 0.713                                              | 4.25                                                |
| 1993 | 258,083                   | 0.692                                              | 4.25                                                |
| 1994 | 260,599                   | 0.675                                              | 4.25                                                |
| 1995 | 263,044                   | 0.656                                              | 4.25                                                |
| 1996 | 265,463                   | 0.638                                              | 4.75                                                |
| 1997 | 268,008                   | 0.623                                              | 5.15                                                |
| 1998 | 270,561                   | 0.600                                              |                                                     |
| Año  | Ingresos de Eastman Kodak | Depósitos del mercado monetario                    | Suscriptores de televisión por cable                |
| 1975 | 5.0                       |                                                    |                                                     |
| 1976 | 5.4                       |                                                    |                                                     |
| 1977 | 6.0                       |                                                    |                                                     |
| 1978 | 7.0                       |                                                    |                                                     |
| 1979 | 8.0                       |                                                    |                                                     |
| 1980 | 9.7                       | 400                                                |                                                     |
| 1981 | 10.3                      | 344                                                |                                                     |
| 1982 | 10.8                      | 400                                                | 21,000                                              |
| 1983 | 10.2                      | 685                                                | 25,000                                              |
| 1984 | 10.6                      | 705                                                | 30,000                                              |
| 1985 | 10.6                      | 815                                                | 32,000                                              |
| 1986 | 11.5                      | 941                                                | 37,500                                              |
| 1987 | 13.3                      | 937                                                | 41,100                                              |
| 1988 | 17.0                      | 926                                                | 44,000                                              |
| 1989 | 18.4                      | 894                                                | 47,500                                              |
| 1990 | 18.9                      | 924                                                | 50,000                                              |
| 1991 | 19.4                      | 1,045                                              | 51,000                                              |
| 1992 | 20.2                      | 1,187                                              | 53,000                                              |
| 1993 | 16.3                      | 1,219                                              | 55,000                                              |
| 1994 | 13.7                      | 1,150                                              | 57,000                                              |
| 1995 | 15.3                      | 1,135                                              | 58,000                                              |
| 1996 | 16.2                      | 1,272                                              | 60,280                                              |
| 1997 |                           | 1,397                                              | 64,050                                              |
| 1998 |                           |                                                    | 64,170                                              |

<sup>1</sup> Estamos en deuda con la doctora Lynn Stephens, de la Eastern Washington University, por habernos proporcionado los valores de estos datos.

*Aparatos domésticos grandes, embarques industriales, congeladores (miles)*

| <i>Año</i> | <i>1983</i> | <i>1984</i> | <i>1985</i> | <i>1986</i> | <i>1987</i> | <i>1988</i> | <i>1989</i> | <i>1990</i> | <i>1991</i> | <i>1992</i> | <i>1993</i> |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Enero      | 100         | 109         | 110         | 90          | 84          | 91          | 89          | 98          | 95          | 90          | 78          |
| Febrero    | 97          | 100         | 73          | 77          | 82          | 89          | 88          | 79          | 75          | 91          | 121         |
| Marzo      | 115         | 94          | 88          | 80          | 105         | 82          | 85          | 87          | 100         | 128         | 130         |
| Abril      | 111         | 97          | 90          | 99          | 97          | 108         | 89          | 90          | 114         | 120         | 119         |
| Mayo       | 113         | 118         | 112         | 106         | 106         | 112         | 106         | 103         | 109         | 132         | 112         |
| Junio      | 136         | 134         | 136         | 140         | 136         | 126         | 110         | 126         | 142         | 176         | 158         |
| Julio      | 148         | 153         | 149         | 159         | 137         | 156         | 125         | 155         | 154         | 205         | 167         |
| Agosto     | 134         | 133         | 131         | 122         | 138         | 152         | 139         | 136         | 138         | 166         | 156         |
| Septiembre | 117         | 106         | 108         | 110         | 109         | 128         | 121         | 123         | 132         | 135         | 137         |
| Octubre    | 92          | 87          | 100         | 94          | 87          | 110         | 93          | 101         | 118         | 160         |             |
| Noviembre  | 78          | 80          | 66          | 72          | 93          | 99          | 84          | 97          | 106         | 138         |             |
| Diciembre  | 82          | 70          | 75          | 72          | 86          | 94          | 90          | 100         | 125         | 136         |             |

*Producción de cerveza (millones de barriles)*

| <i>Año</i> | <i>1983</i> | <i>1984</i> | <i>1985</i> | <i>1986</i> | <i>1987</i> | <i>1988</i> | <i>1989</i> | <i>1990</i> | <i>1991</i> | <i>1992</i> | <i>1993</i> |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Enero      | 14.77       | 14.15       | 15.50       | 15.71       | 15.60       | 15.80       | 15.88       | 16.46       | 16.27       | 15.65       | 15.36       |
| Febrero    | 14.53       | 14.75       | 14.55       | 15.21       | 15.63       | 15.85       | 15.29       | 15.74       | 15.17       | 16.10       | 15.78       |
| Marzo      | 16.78       | 17.72       | 16.76       | 16.51       | 17.66       | 17.12       | 17.57       | 17.97       | 16.08       | 18.06       | 17.41       |
| Abril      | 18.42       | 16.81       | 17.97       | 17.99       | 17.42       | 17.73       | 17.30       | 17.48       | 17.23       | 18.00       | 17.44       |
| Mayo       | 18.17       | 18.74       | 18.86       | 18.67       | 17.44       | 18.31       | 18.41       | 18.10       | 18.90       | 18.89       | 18.87       |
| Junio      | 18.47       | 18.47       | 18.23       | 18.65       | 18.58       | 18.58       | 18.82       | 18.58       | 19.16       | 18.95       | 18.96       |
| Julio      | 18.50       | 19.12       | 18.59       | 18.33       | 18.09       | 18.17       | 18.28       | 18.25       | 19.88       | 18.34       | 18.51       |
| Agosto     | 18.27       | 17.59       | 17.71       | 17.06       | 16.81       | 17.72       | 18.88       | 18.96       | 18.63       | 17.55       |             |
| Septiembre | 15.71       | 14.58       | 14.54       | 15.26       | 15.82       | 15.78       | 15.28       | 16.08       | 16.11       | 15.66       |             |
| Octubre    | 15.41       | 15.14       | 14.36       | 15.62       | 15.50       | 15.61       | 15.82       | 16.62       | 16.65       | 16.15       |             |
| Noviembre  | 13.62       | 13.06       | 13.12       | 13.53       | 13.81       | 14.02       | 14.78       | 15.44       | 14.47       | 14.43       |             |
| Diciembre  | 12.46       | 12.89       | 13.13       | 13.97       | 13.69       | 13.22       | 13.45       | 13.97       | 13.64       | 14.32       |             |

*Recetas nuevas (miles)*

| <i>Año</i> | <i>1994</i> | <i>1995</i> | <i>1996</i> | <i>1997</i> | <i>1998</i> | <i>1999</i> | <i>2000</i> |
|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Enero      | 154         | 200         | 223         | 346         | 518         | 613         | 628         |
| Febrero    | 96          | 118         | 104         | 261         | 404         | 392         | 308         |
| Marzo      | 73          | 90          | 107         | 224         | 300         | 273         | 324         |
| Abril      | 49          | 79          | 85          | 141         | 210         | 322         | 248         |
| Mayo       | 36          | 78          | 75          | 148         | 196         | 189         | 272         |
| Junio      | 59          | 91          | 99          | 145         | 186         | 257         | 634         |
| Julio      | 95          | 167         | 135         | 223         | 247         | 324         | 299         |
| Agosto     | 169         | 169         | 121         | 272         | 343         | 404         | 424         |
| Septiembre | 210         | 289         | 335         | 445         | 464         | 677         | 548         |
| Octubre    | 278         | 347         | 460         | 560         | 680         | 858         | 372         |
| Noviembre  | 298         | 375         | 488         | 612         | 711         | 895         | 876         |
| Diciembre  | 245         | 203         | 326         | 467         | 610         | 664         | 676         |

**Seattle, Washington, número de usuarios por día del autobús público de la ruta de Pike Street para 146 días**

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 350 | 339 | 351 | 364 | 369 | 331 | 331 | 340 | 346 | 341 | 357 | 398 | 381 | 367 | 383 | 375 |
| 353 | 361 | 375 | 371 | 373 | 366 | 382 | 429 | 406 | 403 | 429 | 425 | 427 | 409 | 402 | 409 |
| 419 | 404 | 429 | 463 | 428 | 449 | 444 | 467 | 474 | 463 | 432 | 453 | 462 | 456 | 474 | 514 |
| 489 | 475 | 492 | 525 | 527 | 533 | 527 | 522 | 526 | 513 | 564 | 599 | 572 | 587 | 599 | 601 |
| 611 | 620 | 579 | 582 | 592 | 581 | 630 | 663 | 638 | 631 | 645 | 682 | 601 | 595 | 521 | 521 |
| 516 | 496 | 538 | 575 | 537 | 534 | 542 | 538 | 547 | 540 | 526 | 548 | 555 | 545 | 594 | 643 |
| 625 | 616 | 640 | 625 | 637 | 634 | 621 | 641 | 654 | 649 | 662 | 699 | 672 | 704 | 700 | 711 |
| 715 | 718 | 652 | 664 | 695 | 704 | 733 | 772 | 716 | 712 | 732 | 755 | 761 | 748 | 748 | 750 |
| 744 | 731 | 782 | 810 | 777 | 816 | 840 | 868 | 872 | 811 | 810 | 762 | 634 | 626 | 649 | 697 |
| 657 | 549 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

Los datos de la tabla se leen por renglones.

**Estadísticas de ocupación mensual del Motel Nine**

| Año        | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 |
|------------|------|------|------|------|------|------|------|------|------|------|
| Enero      | 563  | 635  | 647  | 676  | 748  | 795  | 843  | 778  | 895  | 875  |
| Febrero    | 599  | 639  | 658  | 748  | 773  | 788  | 847  | 856  | 856  | 993  |
| Marzo      | 669  | 712  | 713  | 811  | 814  | 890  | 942  | 939  | 893  | 977  |
| Abril      | 598  | 622  | 688  | 729  | 767  | 797  | 804  | 813  | 875  | 969  |
| Mayo       | 580  | 621  | 724  | 701  | 729  | 751  | 840  | 783  | 835  | 872  |
| Junio      | 668  | 676  | 707  | 790  | 749  | 821  | 872  | 828  | 935  | 1006 |
| Julio      | 499  | 501  | 629  | 594  | 681  | 692  | 656  | 657  | 833  | 832  |
| Agosto     | 215  | 220  | 238  | 231  | 241  | 291  | 370  | 310  | 300  | 346  |
| Septiembre | 556  | 561  | 613  | 617  | 680  | 727  | 742  | 780  | 791  | 850  |
| Octubre    | 587  | 603  | 730  | 691  | 708  | 868  | 847  | 860  | 900  | 914  |
| Noviembre  | 546  | 626  | 735  | 701  | 694  | 812  | 732  | 780  | 782  | 869  |
| Diciembre  | 571  | 606  | 652  | 706  | 772  | 800  | 899  | 808  | 880  | 994  |

**Producción de pulpa de madera  
(miles de toneladas cortas)**

| Año        | 1992 | 1993 | 1994 | 1995 | 1996 |
|------------|------|------|------|------|------|
| Enero      | 5276 | 5466 | 5480 | 5579 | 5681 |
| Febrero    | 4919 | 4821 | 4988 | 5033 | 5254 |
| Marzo      | 5227 | 5307 | 5272 | 5319 | 5525 |
| Abril      | 5003 | 5193 | 5254 | 5285 | 5489 |
| Mayo       | 5004 | 5087 | 5156 | 5188 | 5320 |
| Junio      | 4949 | 5102 | 5175 | 5197 | 5417 |
| Julio      | 5219 | 5399 | 5490 | 5604 | 5711 |
| Agosto     | 5265 | 5301 | 5366 | 5357 | 5419 |
| Septiembre | 5007 | 5063 | 5200 | 5133 | 5318 |
| Octubre    | 5079 | 5295 | 5410 | 5483 | 5409 |
| Noviembre  | 4974 | 5021 | 5123 | 5186 | 5310 |
| Diciembre  | 5282 | 4991 | 5428 | 5472 | 5458 |

| <i>Año</i> | <i>Índice de precios<br/>de educación<br/>superior<br/>1983 = 100</i> | <i>Número de<br/>granjas<br/>estadounidenses<br/>(miles)</i> |
|------------|-----------------------------------------------------------------------|--------------------------------------------------------------|
| 1970       | 39.5                                                                  |                                                              |
|            | 42.1                                                                  |                                                              |
|            | 44.3                                                                  |                                                              |
|            | 46.7                                                                  |                                                              |
|            | 49.9                                                                  |                                                              |
| 1975       | 54.3                                                                  | 2521                                                         |
|            | 57.9                                                                  | 2497                                                         |
|            | 61.7                                                                  | 2456                                                         |
|            | 65.8                                                                  | 2436                                                         |
|            | 70.6                                                                  | 2437                                                         |
| 1980       | 77.5                                                                  | 2440                                                         |
|            | 85.9                                                                  | 2440                                                         |
|            | 94.0                                                                  | 2407                                                         |
|            | 100.0                                                                 | 2379                                                         |
|            | 104.7                                                                 | 2334                                                         |
| 1985       | 110.5                                                                 | 2293                                                         |
|            | 115.6                                                                 | 2250                                                         |
|            | 120.3                                                                 | 2213                                                         |
|            | 125.8                                                                 | 2197                                                         |
|            | 133.1                                                                 | 2171                                                         |
| 1990       | 140.8                                                                 | 2140                                                         |
|            | 148.3                                                                 | 2105                                                         |
|            | 153.1                                                                 | 2094                                                         |
|            | 158.2                                                                 | 2040                                                         |

| <i>Año</i> | <i>Promedio del precio<br/>por pozo, gas natural<br/>(centavos por millón<br/>de pies cúbicos)</i> | <i>Precio promedio anual de las<br/>acciones ordinarias de la<br/>empresa eléctrica de servicio<br/>público dado por Moody's</i> |
|------------|----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| 1972       | 18.6                                                                                               | 80.20                                                                                                                            |
|            | 21.6                                                                                               | 71.21                                                                                                                            |
|            | 30.4                                                                                               | 48.26                                                                                                                            |
|            | 44.5                                                                                               | 51.25                                                                                                                            |
|            | 58.0                                                                                               | 60.10                                                                                                                            |
| 1975       | 79.0                                                                                               | 67.55                                                                                                                            |
|            | 90.5                                                                                               | 63.54                                                                                                                            |
|            | 117.8                                                                                              | 60.28                                                                                                                            |
|            | 158.8                                                                                              | 54.80                                                                                                                            |
|            | 198.2                                                                                              | 55.41                                                                                                                            |
| 1980       | 245.7                                                                                              | 63.56                                                                                                                            |
|            | 259.3                                                                                              | 74.04                                                                                                                            |
|            | 265.5                                                                                              | 71.16                                                                                                                            |
|            | 248.0                                                                                              | 87.24                                                                                                                            |
|            | 194.2                                                                                              | 111.11                                                                                                                           |
| 1985       | 166.5                                                                                              | 105.90                                                                                                                           |
|            | 169.0                                                                                              | 97.99                                                                                                                            |
|            | 169.0                                                                                              | 110.45                                                                                                                           |
|            | 172.0                                                                                              | 112.61                                                                                                                           |
|            | 164.0                                                                                              | 126.97                                                                                                                           |
| 1990       | 174.0                                                                                              |                                                                                                                                  |
|            | 203.0                                                                                              |                                                                                                                                  |

*Miles de cargas de un vagón de flete, productos forestales*

| <b>Año</b> | <b>1992</b> | <b>1993</b> | <b>1994</b> | <b>1995</b> | <b>1996</b> |
|------------|-------------|-------------|-------------|-------------|-------------|
| Enero      | 17.6        | 16.9        | 16.1        | 15.6        | 13.2        |
| Febrero    | 19.0        | 18.5        | 15.8        | 16.7        | 14.5        |
| Marzo      | 19.1        | 18.6        | 16.0        | 17.2        | 14.7        |
| Abril      | 18.7        | 17.9        | 16.6        | 16.7        | 14.1        |
| Mayo       | 19.1        | 18.0        | 16.2        | 16.3        | 15.2        |
| Junio      | 19.9        | 17.4        | 15.8        | 16.4        | 15.3        |
| Julio      | 18.3        | 16.4        | 16.6        | 13.9        | 14.9        |
| Agosto     | 19.1        | 16.5        | 15.8        | 15.9        | 14.0        |
| Septiembre | 18.6        | 16.7        | 15.6        | 15.5        | 15.3        |
| Octubre    | 18.8        | 17.0        | 15.6        | 15.2        | 14.4        |
| Noviembre  | 16.7        | 12.4        | 14.4        | 13.9        | 14.2        |
| Diciembre  | 17.3        | 16.9        | 14.3        | 14.0        | 15.0        |

*Precio industrial/tasa de ganancias*

| <b>Año</b> | <b>Trimestre</b> | <b>P/E</b> | <b>Año</b> | <b>Trimestre</b> | <b>P/E</b> |
|------------|------------------|------------|------------|------------------|------------|
| 1986       | 1                | 17.56      | 1990       | 1                | 15.80      |
|            | 2                | 18.09      |            | 2                | 17.00      |
|            | 3                | 16.47      |            | 3                | 14.80      |
|            | 4                | 18.73      |            | 4                | 15.90      |
| 1987       | 1                | 22.16      | 1991       | 1                | 18.80      |
|            | 2                | 21.80      |            | 2                | 20.40      |
|            | 3                | 20.82      |            | 3                | 23.70      |
|            | 4                | 14.09      |            | 4                | 29.20      |
| 1988       | 1                | 13.61      | 1992       | 1                | 28.40      |
|            | 2                | 13.29      |            | 2                | 26.80      |
|            | 3                | 12.50      |            | 3                | 26.00      |
|            | 4                | 12.35      |            | 4                | 26.00      |
| 1989       | 1                | 12.46      |            |                  |            |
|            | 2                | 13.30      |            |                  |            |
|            | 3                | 14.73      |            |                  |            |
|            | 4                | 15.40      |            |                  |            |

**DATOS DEL CAMPAMENTO**

En Spokane, un grupo de empresarios planea desarrollar algunos campamentos en el estado. Uno de sus problemas es decidir qué cantidad deben cobrar por año. De acuerdo con sus observaciones, la cuota depende de algunas variables, por ejemplo, si se instalará una alberca y el tamaño del campamento. Para ser objetivos, han recolectado la siguiente información de 25 campamentos en Washington (de la Rand McNally Campground Guide for Washington) y completaron un análisis por computadora. Las variables analizadas fueron:

$Y$  = Cuota diaria (FEE)

$X_1$  = Número de acres (ACRES)

$X_2$  = Número de campamentos (SITES)

$X_3$  = Si habría o no inodoros con cisterna (TOILET)

$X_4$  = Si se instalaría o no una alberca (POOL)

$X_5$  = Si se contaría o no con conexiones eléctricas (HOOKUP)

$X_6$  = Número de complejos recreativos adicionales (ADD)

| <i>Campamento</i> | <i>Y</i> | <i>X<sub>1</sub></i> | <i>X<sub>2</sub></i> | <i>X<sub>3</sub></i> | <i>X<sub>4</sub></i> | <i>X<sub>5</sub></i> | <i>X<sub>6</sub></i> |
|-------------------|----------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| 1                 | 7.00     | 40                   | 32                   | 0                    | 0                    | 0                    | 2                    |
| 2                 | 8.50     | 20                   | 47                   | 1                    | 0                    | 1                    | 2                    |
| 3                 | 9.00     | 45                   | 18                   | 1                    | 1                    | 1                    | 1                    |
| 4                 | 8.00     | 110                  | 32                   | 1                    | 0                    | 1                    | 3                    |
| 5                 | 8.00     | 30                   | 54                   | 1                    | 0                    | 1                    | 2                    |
| 6                 | 7.00     | 50                   | 30                   | 1                    | 0                    | 1                    | 3                    |
| 7                 | 7.75     | 35                   | 30                   | 1                    | 0                    | 1                    | 2                    |
| 8                 | 8.00     | 18                   | 40                   | 1                    | 0                    | 1                    | 1                    |
| 9                 | 8.50     | 23                   | 60                   | 1                    | 1                    | 1                    | 1                    |
| 10                | 8.50     | 9                    | 60                   | 1                    | 0                    | 1                    | 3                    |
| 11                | 9.00     | 52                   | 50                   | 1                    | 1                    | 1                    | 2                    |
| 12                | 7.00     | 25                   | 21                   | 0                    | 0                    | 1                    | 1                    |
| 13                | 9.00     | 250                  | 30                   | 1                    | 0                    | 1                    | 2                    |
| 14                | 8.50     | 140                  | 70                   | 1                    | 1                    | 1                    | 2                    |
| 15                | 9.00     | 120                  | 80                   | 1                    | 1                    | 1                    | 1                    |
| 16                | 7.50     | 60                   | 50                   | 1                    | 1                    | 1                    | 2                    |
| 17                | 8.50     | 120                  | 35                   | 1                    | 0                    | 1                    | 2                    |
| 18                | 9.00     | 173                  | 25                   | 1                    | 1                    | 1                    | 2                    |
| 19                | 8.00     | 100                  | 75                   | 1                    | 0                    | 1                    | 2                    |
| 20                | 9.50     | 134                  | 35                   | 1                    | 1                    | 1                    | 1                    |
| 21                | 7.50     | 114                  | 120                  | 0                    | 1                    | 1                    | 2                    |
| 22                | 7.50     | 2                    | 17                   | 0                    | 0                    | 1                    | 2                    |
| 23                | 7.50     | 32                   | 15                   | 0                    | 1                    | 0                    | 3                    |
| 24                | 9.00     | 25                   | 30                   | 1                    | 1                    | 1                    | 2                    |
| 25                | 7.50     | 66                   | 100                  | 1                    | 0                    | 1                    | 2                    |

**Base de datos**

| <i>Número<br/>de la<br/>compañía (en millones)</i> | <i>Ventas<br/>(en millones)</i> | <i>Empleados<br/>(en miles)</i> | <i>Gastos de<br/>capital<br/>(en millones)</i> | <i>Gastos<br/>intangibles<br/>(en millones)</i> | <i>Costo de los<br/>productos<br/>vendidos<br/>(en millones)</i> | <i>Mano de<br/>obra y gastos<br/>relacionados<br/>(en millones)</i> | <i>Gastos de<br/>publicidad<br/>(en millones)</i> | <i>Gastos de<br/>investigación<br/>y desarrollo<br/>(en millones)</i> |
|----------------------------------------------------|---------------------------------|---------------------------------|------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|
| 1                                                  | 3,221.8008                      | 42.0000                         | 147.9000                                       | 30.6000                                         | 2,285.2007                                                       | 599.7998                                                            | 118.3000                                          | 28.0000                                                               |
| 2                                                  | 1,690.6001                      | 20.9050                         | 93.0000                                        | 29.1000                                         | 1,057.2002                                                       | 343.2000                                                            | 114.9000                                          | 8.9000                                                                |
| 3                                                  | 2,197.2764                      | 39.0000                         | 66.8670                                        | 55.8600                                         | 1,387.0679                                                       | 661.3997                                                            | 95.5680                                           | 11.1820                                                               |
| 4                                                  | 2,357.8206                      | 23.3000                         | 59.5560                                        | 69.6080                                         | 1,743.7952                                                       | 25.6320                                                             | 51.9170                                           | 8.5000                                                                |
| 5                                                  | 8,129.0000                      | 35.0000                         | 297.0000                                       | 29.0000                                         | 7,423.0000                                                       | 1,178.0000                                                          | 12.8000                                           | 9.2530                                                                |
| 6                                                  | 11,851.0000                     | 23.0000                         | 394.0000                                       | 20.0000                                         | 10,942.0000                                                      | 2,556.0000                                                          | 11.6530                                           | 14.6000                                                               |
| 7                                                  | 323.8606                        | 3.9000                          | 2.5900                                         | 4.2880                                          | 233.5300                                                         | 22.8350                                                             | 3.5290                                            | 30.7320                                                               |
| 8                                                  | 660.4856                        | 8.3780                          | 10.9840                                        | 3.3720                                          | 582.2649                                                         | 25.6250                                                             | 44.9990                                           | 64.8730                                                               |
| 9                                                  | 4,351.1601                      | 50.9120                         | 102.7080                                       | 217.0920                                        | 4,156.8671                                                       | 12.8360                                                             | 66.2640                                           | 8.7790                                                                |
| 10                                                 | 985.8357                        | 5.5000                          | 16.6010                                        | 29.5900                                         | 874.1287                                                         | 19.5000                                                             | 112.3860                                          | 18.3650                                                               |
| 11                                                 | 3,802.5581                      | 39.6000                         | 206.1020                                       | 157.3520                                        | 2,997.2703                                                       | 518.0000                                                            | 139.7290                                          | 16.4130                                                               |
| 12                                                 | 2,576.0464                      | 22.6000                         | 50.6690                                        | 47.0790                                         | 1,885.9053                                                       | 349.4910                                                            | 48.8170                                           | 9.5000                                                                |
| 13                                                 | 106.0160                        | 28.0000                         | 1.3120                                         | 42.0000                                         | 84.6590                                                          | 35.5550                                                             | 22.9370                                           | 8.7330                                                                |
| 14                                                 | 5,669.8945                      | 46.8810                         | 103.0000                                       | 31.1000                                         | 4,424.3007                                                       | 785.0000                                                            | 141.3000                                          | 18.5000                                                               |
| 15                                                 | 319.6570                        | 2.8940                          | 4.5770                                         | 2.2090                                          | 246.6980                                                         | 42.8370                                                             | 87.0000                                           | 1.1000                                                                |
| 16                                                 | 511.7217                        | 10.1000                         | 19.5600                                        | 27.0000                                         | 286.2288                                                         | 48.9990                                                             | 1.8700                                            | 23.6520                                                               |
| 17                                                 | 884.6189                        | 22.8010                         | 58.0940                                        | 33.0000                                         | 467.4436                                                         | 36.5000                                                             | 16.0350                                           | 29.6320                                                               |
| 18                                                 | 166.3750                        | 2.3000                          | 3.9510                                         | 5.2890                                          | 111.0310                                                         | 31.0000                                                             | 4.0230                                            | 38.5420                                                               |
| 19                                                 | 59.1310                         | 18.0000                         | 1.1400                                         | 14.5000                                         | 43.7430                                                          | 26.3210                                                             | 90.3250                                           | 56.9820                                                               |
| 20                                                 | 136.6970                        | 3.1000                          | 2.0090                                         | 18.4930                                         | 105.3300                                                         | 15.8880                                                             | 46.3000                                           | 8.6330                                                                |
| 21                                                 | 767.8799                        | 8.1000                          | 37.4250                                        | 18.0560                                         | 519.3948                                                         | 112.1350                                                            | 21.8470                                           | 2.7860                                                                |
| 22                                                 | 61.3280                         | 1.1390                          | 1.3880                                         | 26.3250                                         | 35.2020                                                          | 17.3140                                                             | 2.4270                                            | 88.5230                                                               |
| 23                                                 | 445.6387                        | 5.8000                          | 18.9780                                        | 12.6000                                         | 213.2880                                                         | 12.1000                                                             | 62.8060                                           | 1.4600                                                                |

**Base de datos (Cont.)**

| <b>Número<br/>de la<br/>compañía</b> | <b>Ventas<br/>(en millones)</b> | <b>Empleados<br/>(en miles)</b> | <b>Gastos de<br/>capital<br/>(en millones)</b> | <b>Gastos<br/>intangibles<br/>(en millones)</b> | <b>Costo de los<br/>productos<br/>vendidos<br/>(en millones)</b> | <b>Mano de<br/>obra y gastos<br/>relacionados<br/>(en millones)</b> | <b>Gastos de<br/>publicidad<br/>(en millones)</b> | <b>Gastos de<br/>investigación<br/>y desarrollo<br/>(en millones)</b> |
|--------------------------------------|---------------------------------|---------------------------------|------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|
| 24                                   | 2,259.6316                      | 16.0270                         | 228.7270                                       | 27.3350                                         | 1,696.3772                                                       | 421.8057                                                            | 116.5990                                          | 9.6000                                                                |
| 25                                   | 624.8040                        | 8.7000                          | 86.4030                                        | 2.8080                                          | 408.4707                                                         | 168.0200                                                            | 33.4700                                           | 9.4440                                                                |
| 26                                   | 329.9578                        | 4.0000                          | 14.9460                                        | 8.3710                                          | 225.0410                                                         | 20.9850                                                             | 12.9790                                           | 32.0000                                                               |
| 27                                   | 308.7327                        | 2.1070                          | 14.8080                                        | 43.5920                                         | 239.1300                                                         | 36.5000                                                             | 18.1220                                           | 1.8510                                                                |
| 28                                   | 598.9507                        | 5.0000                          | 39.7150                                        | 27.8920                                         | 481.9436                                                         | 45.0000                                                             | 39.8230                                           | .7500                                                                 |
| 29                                   | 172.7920                        | 1.5760                          | 1.6590                                         | 23.5420                                         | 118.7090                                                         | 48.2000                                                             | 7.9090                                            | 26.3330                                                               |
| 30                                   | 910.8406                        | 7.0000                          | 14.4610                                        | 5.5880                                          | 677.2527                                                         | 7.0000                                                              | 58.2130                                           | 1.8000                                                                |
| 31                                   | 142.1830                        | 1.6000                          | 5.5880                                         | 72.5190                                         | 126.9660                                                         | 1.6000                                                              | 2.7310                                            | 57.2710                                                               |
| 32                                   | 425.0828                        | 6.8330                          | 72.5190                                        | 31.8030                                         | 256.2837                                                         | 6.8330                                                              | 12.1440                                           | 44.1550                                                               |
| 33                                   | 4,337.9140                      | 36.1000                         | 306.0220                                       | 101.4290                                        | 2,344.1631                                                       | 36.1000                                                             | 270.2576                                          | 16.1100                                                               |
| 34                                   | 271.0076                        | 2.0780                          | 27.1230                                        | 6.5030                                          | 134.3790                                                         | 35.7730                                                             | 20.2540                                           | 87.4830                                                               |
| 35                                   | 209.4520                        | 2.9980                          | 14.4690                                        | 14.6060                                         | 176.4890                                                         | 2.0780                                                              | 1.8970                                            | 714.9990                                                              |
| 36                                   | 62.4180                         | 3.8000                          | 3.7390                                         | 7.6680                                          | 34.4700                                                          | 2.9980                                                              | 44.0500                                           | 121.3300                                                              |
| 37                                   | 4,300.0000                      | 95.5000                         | 412.2886                                       | 157.6030                                        | 2,108.5503                                                       | 5.1000                                                              | 257.6807                                          | 11.6440                                                               |
| 38                                   | 390.6829                        | 5.1000                          | 30.8480                                        | 10.8550                                         | 225.1080                                                         | 6.3000                                                              | 18.3780                                           | 33.4770                                                               |
| 39                                   | 270.0127                        | 6.3000                          | 40.0340                                        | 22.4540                                         | 189.8000                                                         | 2.0000                                                              | 4.9080                                            | 43.7430                                                               |
| 40                                   | 97.9660                         | 2.0000                          | 6.9940                                         | 5.2500                                          | 64.5920                                                          | 31.9700                                                             | 2.5900                                            | 18.9700                                                               |
| 41                                   | 66.4090                         | 12.5261                         | 3.7570                                         | 1.0090                                          | 57.2310                                                          | 33.2000                                                             | 59.1300                                           | 14.9460                                                               |
| 42                                   | 56.5550                         | 3.9000                          | 1.6240                                         | 6.9940                                          | 44.0550                                                          | 53.5000                                                             | 19.5600                                           | 1.6590                                                                |
| 43                                   | 3,267.9551                      | 31.9790                         | 502.0398                                       | 45.6140                                         | 2,517.7566                                                       | 754.8977                                                            | 3.9510                                            | 57.7210                                                               |
| 44                                   | 2,745.7439                      | 43.9680                         | 251.0340                                       | 16.1110                                         | 1,638.7969                                                       | 45.0000                                                             | 161.2000                                          | 108.1480                                                              |
| 45                                   | 2,609.0000                      | 33.2000                         | 248.0001                                       | 10.0000                                         | 1,874.0000                                                       | 564.0000                                                            | 18.0000                                           | 83.0000                                                               |
| 46                                   | 1,677.6016                      | 11.6440                         | 284.6089                                       | 87.4830                                         | 1,185.9717                                                       | 24.4530                                                             | 6.4840                                            | 36.1310                                                               |
| 47                                   | 6,887.6210                      | 53.5000                         | 1,075.1719                                     | 84.0390                                         | 4,721.9570                                                       | 1,375.7996                                                          | 44.0700                                           | 231.4690                                                              |
| 48                                   | 10,584.1990                     | 132.1400                        | 714.2002                                       | 22.6000                                         | 7,353.5000                                                       | 3,204.2688                                                          | 93.4000                                           | 377.1001                                                              |
| 49                                   | 2,912.7644                      | 45.8540                         | 195.2680                                       | 45.6430                                         | 2,189.5293                                                       | 879.6548                                                            | 14.9460                                           | 66.0560                                                               |
| 50                                   | 4,309.5820                      | 66.8000                         | 275.3079                                       | 67.3120                                         | 2,913.9036                                                       | 993.3997                                                            | 1.6590                                            | 40.5470                                                               |
| 51                                   | 1,946.4766                      | 24.4530                         | 121.3300                                       | 6.2920                                          | 1,403.4976                                                       | 546.0508                                                            | 35.2020                                           | 40.0810                                                               |
| 52                                   | 9,254.1171                      | 151.2000                        | 1,431.0906                                     | 121.3300                                        | 6,187.7851                                                       | 2,125.2012                                                          | 95.9510                                           | 334.8057                                                              |
| 53                                   | 5,018.6914                      | 62.8510                         | 479.8997                                       | 1.6240                                          | 3,478.0989                                                       | 1,318.0999                                                          | 9.2530                                            | 144.3000                                                              |
| 54                                   | 1,510.7798                      | 15.3000                         | 207.9320                                       | 63.5190                                         | 1,157.2117                                                       | 13.9700                                                             | 27.6660                                           | 39.7150                                                               |
| 55                                   | 1,560.0750                      | 22.7000                         | 162.5190                                       | 61.9380                                         | 1,188.9126                                                       | 18.4340                                                             | 19.3190                                           | 24.7010                                                               |
| 56                                   | 2,794.0000                      | 37.4000                         | 256.0999                                       | 7.3000                                          | 1,928.4988                                                       | 780.7996                                                            | 18.3650                                           | 70.1000                                                               |
| 57                                   | 921.3689                        | 13.9700                         | 61.9380                                        | 18.4340                                         | 597.7000                                                         | 45.1640                                                             | 19.2020                                           | 22.6500                                                               |
| 58                                   | 1,253.5430                      | 13.0580                         | 66.4310                                        | 13.9700                                         | 806.6758                                                         | 236.5000                                                            | 32.0000                                           | 48.6510                                                               |
| 59                                   | 1,328.1138                      | 13.1160                         | 201.1960                                       | 31.2730                                         | 851.8938                                                         | 1.1550                                                              | 31.2730                                           | 33.5620                                                               |
| 60                                   | 1,314.6299                      | 27.3460                         | 36.9330                                        | 43.0750                                         | 569.7327                                                         | 6.4690                                                              | 174.4610                                          | 42.1160                                                               |
| 61                                   | 7,869.6914                      | 113.3710                        | 687.7998                                       | 90.2000                                         | 5,580.5976                                                       | 1,931.5005                                                          | 76.5000                                           | 155.9000                                                              |
| 62                                   | 73.0550                         | 7.8240                          | 26.5680                                        | 20.6650                                         | 38.9980                                                          | 22.8990                                                             | 43.0750                                           | 99.8430                                                               |
| 63                                   | 108.5090                        | 87.4350                         | 5.6630                                         | 37.3860                                         | 77.1740                                                          | 36.9990                                                             | 90.2000                                           | 1.6500                                                                |
| 64                                   | 1,422.4507                      | 16.5000                         | 100.4700                                       | 69.8820                                         | 1,060.5420                                                       | 305.7000                                                            | 6.3970                                            | 25.4520                                                               |
| 65                                   | 87.4350                         | 7.6550                          | 8.5150                                         | 15.3750                                         | 51.3970                                                          | 11.3940                                                             | 69.8820                                           | 2.7200                                                                |
| 66                                   | 7.8240                          | 9.5280                          | 26.6950                                        | 7.7640                                          | 6.7860                                                           | 20.5720                                                             | 4.2100                                            | 52.1780                                                               |
| 67                                   | 868.7107                        | 15.3400                         | 42.4040                                        | 1.2120                                          | 686.0518                                                         | 200.4850                                                            | 10.4000                                           | 22.7240                                                               |
| 68                                   | 137.3950                        | 2.8750                          | 14.1080                                        | 9.7470                                          | 112.2350                                                         | 30.7620                                                             | 83.1580                                           | 1.9000                                                                |
| 69                                   | 753.8848                        | 6.5480                          | 24.2870                                        | 4.2120                                          | 596.5076                                                         | 13.4000                                                             | 88.8250                                           | 6.4200                                                                |
| 70                                   | 1,445.0166                      | 27.0030                         | 84.1490                                        | 99.9080                                         | 786.8777                                                         | 1.9360                                                              | 39.8650                                           | 76.1870                                                               |
| 71                                   | 3,062.6316                      | 49.6190                         | 67.6310                                        | 83.1580                                         | 1,446.5227                                                       | 668.9910                                                            | 243.0450                                          | 74.5240                                                               |
| 72                                   | 2,450.4285                      | 32.6000                         | 81.9220                                        | 88.8250                                         | 906.9639                                                         | 6.7120                                                              | 423.2698                                          | 90.5730                                                               |
| 73                                   | 141.2580                        | 1.3040                          | 4.5050                                         | 6.7300                                          | 95.1540                                                          | 3.7000                                                              | 9.9040                                            | 9.7580                                                                |

**Base de datos (Cont.)**

| <b>Número<br/>de la<br/>compañía</b> | <b>Ventas<br/>(en millones)</b> | <b>Empleados<br/>(en miles)</b> | <b>Gastos<br/>de capital<br/>(en millones)</b> | <b>Gastos<br/>intangibles<br/>(en millones)</b> | <b>Costo de los<br/>productos<br/>vendidos<br/>(en millones)</b> | <b>Mano de<br/>obra y gastos<br/>relacionados<br/>(en millones)</b> | <b>Gastos de<br/>publicidad<br/>(en millones)</b> | <b>Gastos de<br/>investigación<br/>y desarrollo<br/>(en millones)</b> |
|--------------------------------------|---------------------------------|---------------------------------|------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|
| 74                                   | 6.8030                          | 5.1000                          | 9.5230                                         | 1.4590                                          | 2.3980                                                           | 12.2490                                                             | .7230                                             | 11.9490                                                               |
| 75                                   | 1,852.0896                      | 25.4000                         | 89.5500                                        | 57.7900                                         | 672.7947                                                         | 4.5070                                                              | 28.4910                                           | 148.0770                                                              |
| 76                                   | 365.7217                        | 4.9030                          | 17.0620                                        | 16.7160                                         | 217.5420                                                         | 3.4720                                                              | 6.7300                                            | 11.8950                                                               |
| 77                                   | 1,981.4397                      | 28.7000                         | 155.8530                                       | 141.2700                                        | 668.7720                                                         | 634.0596                                                            | 55.2940                                           | 161.3500                                                              |
| 78                                   | 2,362.1326                      | 40.7000                         | 110.1000                                       | 99.8430                                         | 1,055.4187                                                       | 11.3940                                                             | 75.7000                                           | 113.1280                                                              |
| 79                                   | 357.0696                        | 5.5500                          | 12.6430                                        | 52.1780                                         | 141.2700                                                         | 2.1330                                                              | 36.8860                                           | 18.9510                                                               |
| 80                                   | 220.3790                        | 3.7000                          | 10.7860                                        | 9.7580                                          | 67.1220                                                          | 20.5720                                                             | 7.1610                                            | 6.2610                                                                |
| 81                                   | 1,082.4927                      | 17.9000                         | 51.3360                                        | 52.1780                                         | 310.7820                                                         | 315.8997                                                            | 114.9660                                          | 65.6910                                                               |
| 82                                   | 848.3799                        | 17.1000                         | 41.2990                                        | 11.9490                                         | 386.0066                                                         | 16.0000                                                             | 40.6150                                           | 61.6940                                                               |
| 83                                   | 1,112.0386                      | 16.5890                         | 74.5790                                        | 44.6610                                         | 378.7710                                                         | 7.3000                                                              | 91.2150                                           | 77.3130                                                               |
| 84                                   | 1,515.8816                      | 37.0000                         | 108.0460                                       | 52.3290                                         | 758.5320                                                         | 469.9229                                                            | 74.5950                                           | 61.8300                                                               |
| 85                                   | 1,328.5508                      | 19.9200                         | 44.6810                                        | 6.2850                                          | 566.2200                                                         | 323.7090                                                            | 36.9560                                           | 115.5890                                                              |
| 86                                   | 2,878.4956                      | 58.0000                         | 182.2670                                       | 348.1426                                        | 1,247.2339                                                       | 1.1500                                                              | 391.6277                                          | 85.3970                                                               |
| 87                                   | 4,312.0507                      | 56.6000                         | 169.2950                                       | 66.9970                                         | 2,672.3262                                                       | 6.4600                                                              | 260.3870                                          | 37.6540                                                               |
| 88                                   | 54.3250                         | 37.3860                         | 1.0660                                         | 2.8130                                          | 26.5960                                                          | 4.7670                                                              | 0.7520                                            | 44.6610                                                               |
| 89                                   | 122.9470                        | 57.1720                         | 13.7480                                        | 7.5620                                          | 94.6720                                                          | 17.6580                                                             | 1.4590                                            | 3.8670                                                                |
| 90                                   | 2,014.7056                      | 31.0000                         | 74.7910                                        | 0.0000                                          | 700.4778                                                         | 503.6768                                                            | 45.0900                                           | 21.1460                                                               |
| 91                                   | 969.8328                        | 18.5170                         | 40.8340                                        | 54.2710                                         | 448.5286                                                         | 9.4450                                                              | 91.2690                                           | 8.5670                                                                |
| 92                                   | 45.3670                         | 8.3500                          | 1.6430                                         | 7.0670                                          | 15.7310                                                          | 2.1230                                                              | 5.1820                                            | 52.3290                                                               |
| 93                                   | 255.1320                        | 3.3000                          | 10.6420                                        | 20.2520                                         | 131.6750                                                         | 12.2220                                                             | 42.5670                                           | 6.2850                                                                |
| 94                                   | 1,710.4700                      | 31.7000                         | 91.5640                                        | 54.7540                                         | 752.5889                                                         | 530.2456                                                            | 239.9010                                          | 42.0600                                                               |
| 95                                   | 365.8809                        | 3.4800                          | 20.0140                                        | 6.7300                                          | 177.5500                                                         | 25.8740                                                             | 16.7100                                           | 23.7910                                                               |
| 96                                   | 33.2650                         | 2.0870                          | 1.5120                                         | 4.4840                                          | 19.7100                                                          | 19.7100                                                             | 1.1550                                            | 2.8890                                                                |
| 97                                   | 53.7460                         | .5250                           | 2.0870                                         | 42.2810                                         | 16.1820                                                          | 16.1800                                                             | 7.6770                                            | 19.7100                                                               |
| 98                                   | 52.8760                         | 1.1420                          | 2.4190                                         | 1.2160                                          | 27.1500                                                          | 27.1500                                                             | 6.4690                                            | 16.1820                                                               |
| 99                                   | 9.6630                          | 2.4190                          | 12.7460                                        | 7.9670                                          | 5.6960                                                           | 5.6950                                                              | .4570                                             | 27.1500                                                               |
| 100                                  | 1,451.6687                      | 29.0000                         | 86.6820                                        | 97.2690                                         | 505.8267                                                         | 36.1200                                                             | 137.7250                                          | 30.7620                                                               |
| 101                                  | 321.3638                        | 4.9110                          | 13.1180                                        | 11.0840                                         | 268.0159                                                         | 57.2600                                                             | 1.1110                                            | 13.4000                                                               |
| 102                                  | 156.4580                        | 2.3500                          | 4.5670                                         | 3.8620                                          | 114.1930                                                         | 6.4800                                                              | 4.7670                                            | 5.6960                                                                |
| 103                                  | 52.1870                         | .8650                           | 1.5100                                         | 20.6490                                         | 36.5130                                                          | 59.3250                                                             | 18.0150                                           | 1.9360                                                                |
| 104                                  | 447.2100                        | 7.7670                          | 12.7460                                        | 41.7940                                         | 280.3218                                                         | 26.8120                                                             | 9.4400                                            | 505.8267                                                              |
| 105                                  | 86.8170                         | 1.1000                          | 1.2810                                         | 19.3850                                         | 57.2600                                                          | 26.6950                                                             | 2.1230                                            | 1.3420                                                                |
| 106                                  | 1,132.3499                      | 18.0150                         | 16.8570                                        | 1.6970                                          | 785.0718                                                         | 36.9240                                                             | 25.8740                                           | 6.7120                                                                |
| 107                                  | 217.4120                        | 3.2000                          | 4.4840                                         | 10.5440                                         | 142.6020                                                         | 57.5790                                                             | 3.2520                                            | 3.7000                                                                |
| 108                                  | 7.7640                          | 86.6820                         | 1.2810                                         | 7.2210                                          | 6.4800                                                           | 9.5280                                                              | 20.8580                                           | 268.0159                                                              |
| 109                                  | 1,581.8760                      | 20.8580                         | 142.2810                                       | 5.8820                                          | 1,280.1670                                                       | 359.0999                                                            | 1.1000                                            | 12.2490                                                               |
| 110                                  | 201.4650                        | 1.1000                          | 7.9670                                         | 1.3370                                          | 169.2630                                                         | 57.5700                                                             | 7.4000                                            | 114.1930                                                              |
| 111                                  | 198.9010                        | .9110                           | 9.7470                                         | .4290                                           | 164.1940                                                         | 73.9670                                                             | 1.8400                                            | 36.5130                                                               |
| 112                                  | 1,497.0076                      | 7.4000                          | 131.9400                                       | 6.0210                                          | 1,098.2969                                                       | 99.4080                                                             | 5.1000                                            | 280.3280                                                              |
| 113                                  | 153.2290                        | 1.8400                          | 11.0840                                        | 3.4390                                          | 59.2350                                                          | 9.2800                                                              | 8.3500                                            | 4.5070                                                                |
| 114                                  | 367.9246                        | 5.1000                          | 20.6490                                        | 11.2110                                         | 230.1690                                                         | 73.9670                                                             | 1.2110                                            | 11.3940                                                               |
| 115                                  | 494.4136                        | 8.3500                          | 19.3850                                        | 3.1490                                          | 342.9849                                                         | 6.4690                                                              | 3.8030                                            | 2.1000                                                                |
| 116                                  | 52.4550                         | 1.2120                          | 7.2210                                         | 7.0620                                          | 26.8120                                                          | 4.7670                                                              | 4.0510                                            | 57.2600                                                               |
| 117                                  | 37.3860                         | .8200                           | 1.3370                                         | 44.3730                                         | 26.6950                                                          | 9.4400                                                              | 5.6000                                            | 785.0718                                                              |
| 118                                  | 57.7120                         | 13.1190                         | 3.4390                                         | .7160                                           | 36.9240                                                          | 2.1230                                                              | 7.5620                                            | .9510                                                                 |
| 119                                  | 586.4766                        | 3.8030                          | 44.3730                                        | 34.2780                                         | 391.3706                                                         | 25.8740                                                             | 2.8100                                            | 20.5720                                                               |
| 120                                  | 476.2078                        | 4.0510                          | 34.2780                                        | 30.2360                                         | 244.7830                                                         | 99.9080                                                             | 5.8820                                            | 16.0000                                                               |
| 121                                  | 15.3570                         | 4.5671                          | 16.8570                                        | 53.2830                                         | 9.5280                                                           | 29.0000                                                             | 6.0200                                            | 142.6020                                                              |
| 122                                  | 393.6016                        | 5.6000                          | 30.2360                                        | 2.8890                                          | 265.3079                                                         | 9.2800                                                              | 11.2110                                           | 7.3000                                                                |
| 123                                  | 4,701.1210                      | 7.5620                          | 353.2830                                       | 48.6920                                         | 3,707.6846                                                       | 4.9110                                                              | 3.1490                                            | 6.4800                                                                |

**Base de datos (Cont.)**

| <b>Número<br/>de la<br/>compañía</b> | <b>Ventas<br/>(en millones)</b> | <b>Empleados<br/>(en miles)</b> | <b>Gastos<br/>de capital<br/>(en millones)</b> | <b>Gastos<br/>intangibles<br/>(en millones)</b> | <b>Costo de los<br/>productos<br/>vendidos<br/>(en millones)</b> | <b>Mano de<br/>obra y gastos<br/>relacionados<br/>(en millones)</b> | <b>Gastos de<br/>publicidad<br/>(en millones)</b> | <b>Gastos de<br/>investigación<br/>y desarrollo<br/>(en millones)</b> |
|--------------------------------------|---------------------------------|---------------------------------|------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|
| 124                                  | 1,167.8340                      | 2.8100                          | 48.6920                                        | 8.4580                                          | 1,017.6038                                                       | 2.3500                                                              | 7.0620                                            | 59.2350                                                               |
| 125                                  | 12,298.3980                     | 50.7000                         | 1,221.8008                                     | 10.4000                                         | 9,285.7109                                                       | 1,016.5000                                                          | 13.1160                                           | 64.6000                                                               |
| 126                                  | 439.4727                        | 1.9020                          | 65.1100                                        | 39.8650                                         | 263.8108                                                         | 51.1480                                                             | 27.3460                                           | 31.2730                                                               |
| 127                                  | 29,127.0030                     | 108.7000                        | 1,897.0005                                     | 9.9040                                          | 20,032.0000                                                      | 78.7700                                                             | 16.5000                                           | 86.0000                                                               |
| 128                                  | 1,993.6624                      | 8.0000                          | 43.4190                                        | 45.7820                                         | 1,755.5662                                                       | 3.5730                                                              | 31.1370                                           | 43.0750                                                               |
| 129                                  | 4,660.8945                      | 18.1000                         | 636.1238                                       | 28.4900                                         | 3,675.6895                                                       | 440.7996                                                            | 3.4000                                            | 11.6000                                                               |
| 130                                  | 976.4578                        | 8.8280                          | 14.8590                                        | 55.2940                                         | 879.3516                                                         | 91.8000                                                             | 15.3440                                           | 90.2000                                                               |
| 131                                  | 3,834.9324                      | 6.6610                          | 316.7156                                       | 68.2690                                         | 3,557.4734                                                       | 7.4050                                                              | 2.8250                                            | 69.8820                                                               |
| 132                                  | 9,535.7382                      | 42.7800                         | 1,107.3838                                     | 75.7000                                         | 7,075.1875                                                       | 971.0000                                                            | 6.5480                                            | 29.7730                                                               |
| 133                                  | 657.7776                        | 1.2640                          | 56.1460                                        | 36.8860                                         | 565.0176                                                         | 14.4700                                                             | 27.0030                                           | 4.2120                                                                |
| 134                                  | 100.4570                        | 43.0750                         | 44.0680                                        | 7.1610                                          | 72.7830                                                          | 22.0310                                                             | 49.6110                                           | 83.1580                                                               |
| 135                                  | 60,334.5110                     | 130.0000                        | 4,186.9296                                     | 40.6150                                         | 45,999.0070                                                      | 3,405.0000                                                          | 32.6000                                           | 290.0000                                                              |
| 136                                  | 2,150.0000                      | 90.2110                         | 311.7000                                       | 91.2150                                         | 1,460.7996                                                       | 57.4030                                                             | 1.3040                                            | 25.1000                                                               |
| 137                                  | 18,069.0000                     | 58.3000                         | 1,680.0000                                     | 74.5900                                         | 13,442.0000                                                      | 1,345.0000                                                          | 25.4000                                           | 88.8250                                                               |
| 138                                  | 109.7380                        | 69.8870                         | 32.2560                                        | 36.9560                                         | 97.0130                                                          | 2.5200                                                              | 4.9030                                            | 6.7300                                                                |
| 139                                  | 592.7710                        | 3.2520                          | 123.7680                                       | 3.8770                                          | 420.3206                                                         | 67.3300                                                             | 28.7000                                           | 1.4590                                                                |
| 140                                  | 4,642.3945                      | 14.3280                         | 353.5999                                       | 33.5620                                         | 4,085.0989                                                       | 324.0000                                                            | 40.7000                                           | 25.0000                                                               |
| 141                                  | 2,072.4412                      | 11.1480                         | 270.1846                                       | 42.1160                                         | 1,640.8118                                                       | 1.2400                                                              | 5.5500                                            | 4.9810                                                                |
| 142                                  | 4,509.3828                      | 13.3540                         | 502.2720                                       | 1.6500                                          | 2,838.0845                                                       | 236.4540                                                            | 2.0370                                            | 12.8000                                                               |
| 143                                  | 34,736.0030                     | 207.7000                        | 1,760.7100                                     | 2.7200                                          | 26,053.9060                                                      | 20.9400                                                             | 3.7000                                            | 16.7160                                                               |
| 144                                  | 1,191.0337                      | 4.2070                          | 255.6150                                       | 1.9000                                          | 865.6477                                                         | 82.6730                                                             | .2670                                             | 99.8430                                                               |
| 145                                  | 312.7300                        | 4.2120                          | 76.5000                                        | 6.4200                                          | 452.4130                                                         | 17.0050                                                             | 17.9000                                           | 52.1780                                                               |
| 146                                  | 1,553.1077                      | 9.1500                          | 343.9539                                       | 23.6410                                         | 988.8760                                                         | 185.6600                                                            | 12.5840                                           | 9.7580                                                                |
| 147                                  | 6,997.7734                      | 30.0080                         | 956.1719                                       | 11.2330                                         | 4,886.8125                                                       | 720.5000                                                            | 17.1000                                           | 58.2460                                                               |
| 148                                  | 513.1880                        | 5.1420                          | 41.9800                                        | 41.9800                                         | 375.3599                                                         | 25.0200                                                             | 11.3330                                           | 11.9490                                                               |
| 149                                  | 28,085.0030                     | 94.8000                         | 2,913.0000                                     | 32.5600                                         | 20,632.0000                                                      | 2,344.0000                                                          | 89.0000                                           | 231.0000                                                              |
| 150                                  | 11,062.8980                     | 34.9740                         | 1,774.3904                                     | 43.0250                                         | 8,259.7656                                                       | 1,051.0000                                                          | 16.5890                                           | 114.0000                                                              |
| 151                                  | 23,232.4060                     | 37.5750                         | 1,049.6729                                     | 90.2110                                         | 19,964.6050                                                      | 994.0000                                                            | 37.0000                                           | 89.7370                                                               |
| 152                                  | 14,961.5000                     | 47.0110                         | 1,744.0364                                     | 69.8870                                         | 10,046.0000                                                      | 1,126.7310                                                          | 19.9200                                           | 80.3250                                                               |
| 153                                  | 5,197.7070                      | 24.1450                         | 762.2510                                       | 4.2120                                          | 3,336.7566                                                       | 431.9976                                                            | 7.7130                                            | 15.0520                                                               |
| 154                                  | 7,428.2343                      | 33.7210                         | 601.1216                                       | 6.7310                                          | 5,714.3085                                                       | 9.7320                                                              | 58.0000                                           | 21.0000                                                               |
| 155                                  | 28,607.5030                     | 67.8410                         | 1,344.3777                                     | 10.4000                                         | 24,787.6050                                                      | 1,572.7996                                                          | 56.6000                                           | 52.0000                                                               |
| 156                                  | 87.6100                         | 6.7310                          | 12.7120                                        | 39.8650                                         | 74.5510                                                          | 31.5580                                                             | 31.0000                                           | 44.6610                                                               |
| 157                                  | 1,165.6736                      | 3.5310                          | 26.6780                                        | 9.9040                                          | 1,035.7129                                                       | 6.6000                                                              | 18.5170                                           | 2.4490                                                                |
| 158                                  | 567.3650                        | 1.5420                          | 97.4910                                        | 28.9400                                         | 480.5110                                                         | 23.5230                                                             | 3.300                                             | 52.3290                                                               |
| 159                                  | 5,954.9414                      | 16.2970                         | 732.0000                                       | 55.9240                                         | 4,540.4609                                                       | 444.8997                                                            | 31.7000                                           | 18.5000                                                               |
| 160                                  | 368.0940                        | 2.3150                          | 15.0860                                        | 2.7160                                          | 319.4939                                                         | 10.6050                                                             | 3.4800                                            | 6.2850                                                                |
| 161                                  | 751.7327                        | 6.2550                          | 51.1060                                        | 13.5380                                         | 606.8318                                                         | 3.5230                                                              | 6.8000                                            | 9.9000                                                                |
| 162                                  | 895.4087                        | 10.9000                         | 145.5140                                       | 9.3840                                          | 681.9656                                                         | 26.3250                                                             | 39.0000                                           | 30.6000                                                               |
| 163                                  | 1,063.2908                      | 16.1790                         | 51.1480                                        | 25.7670                                         | 746.2820                                                         | 12.6000                                                             | 16.6980                                           | 14.6320                                                               |
| 164                                  | 1,306.0867                      | 19.3970                         | 78.7700                                        | 2.7490                                          | 1,021.4856                                                       | 435.2998                                                            | 23.3000                                           | 13.2830                                                               |
| 165                                  | 140.4440                        | 1.9190                          | 3.5730                                         | 55.8600                                         | 122.3210                                                         | 27.3350                                                             | 35.0000                                           | 29.1000                                                               |
| 166                                  | 4,357.2812                      | 52.1400                         | 110.4470                                       | 12.0830                                         | 3,540.9612                                                       | 1,235.0000                                                          | 3.9000                                            | 55.8600                                                               |
| 167                                  | 263.9048                        | 3.7000                          | 7.4050                                         | 27.2080                                         | 203.3440                                                         | 2.8080                                                              | 8.3780                                            | 3.2500                                                                |
| 168                                  | 6,184.8945                      | 94.5000                         | 398.2000                                       | 69.6080                                         | 5,224.0000                                                       | 2,550.0000                                                          | 50.9120                                           | 37.1000                                                               |
| 169                                  | 257.6509                        | 3.3640                          | 14.4730                                        | 7.5700                                          | 190.4190                                                         | 8.3710                                                              | 5.5000                                            | 69.6080                                                               |
| 170                                  | 50.5150                         | 52.5350                         | 29.1000                                        | 29.0000                                         | 18.0560                                                          | 43.5920                                                             | 39.6000                                           | 29.0000                                                               |
| 171                                  | 419.6470                        | 4.3020                          | 22.0310                                        | 20.0000                                         | 341.5906                                                         | 135.6000                                                            | 22.6000                                           | 20.0000                                                               |
| 172                                  | 1,227.4490                      | 20.0000                         | 57.4030                                        | 4.2880                                          | 999.7520                                                         | 27.8920                                                             | 28.0000                                           | 9.0000                                                                |
| 173                                  | 779.3450                        | 8.8000                          | 22.0670                                        | 3.3700                                          | 678.4258                                                         | 229.1270                                                            | 46.8810                                           | 4.2880                                                                |

**Base de datos (Cont.)**

| <b>Número<br/>de la<br/>compañía</b> | <b>Ventas<br/>(en millones)</b> | <b>Empleados<br/>(en miles)</b> | <b>Gastos<br/>de capital<br/>(en millones)</b> | <b>Gastos<br/>intangibles<br/>(en millones)</b> | <b>Costo de los<br/>productos<br/>vendidos<br/>(en millones)</b> | <b>Mano de<br/>obra y gastos<br/>relacionados<br/>(en millones)</b> | <b>Gastos de<br/>publicidad<br/>(en millones)</b> | <b>Gastos de<br/>investigación<br/>y desarrollo<br/>(en millones)</b> |
|--------------------------------------|---------------------------------|---------------------------------|------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|
| 174                                  | 72.1760                         | 1.3000                          | 2.5210                                         | 29.5900                                         | 50.9650                                                          | 24.8290                                                             | 2.8940                                            | 3.3700                                                                |
| 175                                  | 3,248.0076                      | 36.0620                         | 263.6167                                       | 19.4460                                         | 2,710.3455                                                       | 974.3379                                                            | 10.1000                                           | 29.5900                                                               |
| 176                                  | 921.1270                        | 12.6590                         | 67.3340                                        | 10.5250                                         | 771.0059                                                         | 23.5420                                                             | 22.8010                                           | 157.3520                                                              |
| 177                                  | 711.9827                        | 12.5120                         | 133.3850                                       | 45.0790                                         | 653.8069                                                         | 351.4700                                                            | 2.3000                                            | 45.0790                                                               |
| 178                                  | 72.4110                         | 1.0250                          | 1.2400                                         | 42.0000                                         | 60.0820                                                          | 93.0000                                                             | 18.0000                                           | 42.0000                                                               |
| 179                                  | 297.5686                        | 4.1520                          | 20.9420                                        | .8990                                           | 248.7160                                                         | 123.1000                                                            | 3.1000                                            | 31.1000                                                               |
| 180                                  | 677.8489                        | 6.0700                          | 17.0050                                        | 31.1000                                         | 613.3047                                                         | 169.2570                                                            | 8.1000                                            | 2.2090                                                                |
| 181                                  | 582.6238                        | 1.4000                          | 25.0290                                        | 2.2090                                          | 474.3450                                                         | 66.8670                                                             | 1.1390                                            | 27.0000                                                               |
| 182                                  | 3,750.4109                      | 38.1700                         | 120.8280                                       | 27.0000                                         | 3,240.7886                                                       | 1,132.6216                                                          | 5.8000                                            | 33.0000                                                               |
| 183                                  | 88.8070                         | 2.3330                          | 9.7320                                         | 7.2110                                          | 66.6540                                                          | 59.5500                                                             | 16.0270                                           | 5.2890                                                                |
| 184                                  | 306.9397                        | 2.8000                          | 31.5880                                        | 33.0000                                         | 220.4980                                                         | 2.5900                                                              | 8.7000                                            | 14.5000                                                               |
| 185                                  | 331.7366                        | 5.2000                          | 6.6000                                         | 11.0250                                         | 295.3848                                                         | 10.9840                                                             | 4.0000                                            | 18.4930                                                               |
| 186                                  | 546.9500                        | 8.9000                          | 23.5230                                        | 5.2890                                          | 439.0479                                                         | 16.0010                                                             | 2.1000                                            | 18.0560                                                               |
| 187                                  | 7.5910                          | 30.6000                         | 7.8900                                         | 14.5000                                         | 5.0480                                                           | 50.6690                                                             | 5.0000                                            | 26.3250                                                               |
| 188                                  | 3,479.4573                      | 41.3940                         | 170.3720                                       | 18.4930                                         | 3,100.5391                                                       | 1,177.5999                                                          | 1.5760                                            | 15.0500                                                               |
| 189                                  | 485.6138                        | 6.6580                          | 58.6750                                        | 3.5250                                          | 335.3318                                                         | 42.0000                                                             | 93.0000                                           | 1.4320                                                                |
| 190                                  | 123.2280                        | 2.0450                          | 10.6050                                        | 5.2550                                          | 96.6630                                                          | 20.9000                                                             | 66.8670                                           | 12.6000                                                               |
| 191                                  | 488.2327                        | 4.6500                          | 20.4800                                        | 1.1111                                          | 402.8457                                                         | 402.8400                                                            | 77.0101                                           | 22.2426                                                               |
| 192                                  | 100.7820                        | 1.7030                          | 2.4430                                         | 1.6800                                          | 88.7960                                                          | 4.0000                                                              | 21.0000                                           | 28.3032                                                               |
| 193                                  | 165.7970                        | 4.7660                          | 3.2790                                         | 88.0003                                         | 120.1080                                                         | 2.0000                                                              | 4.0008                                            | 18.2022                                                               |
| 194                                  | 274.8440                        | 3.5500                          | 21.7900                                        | 2.9530                                          | 213.1860                                                         | 3.0000                                                              | 3.7521                                            | 24.2628                                                               |
| 195                                  | 11,049.5000                     | 166.8480                        | 667.7998                                       | 55.5000                                         | 9,955.3984                                                       | 4,485.1953                                                          | 22.0007                                           | 52.5000                                                               |
| 196                                  | 1,154.8477                      | 14.4190                         | 32.2360                                        | 4.0800                                          | 1,037.4727                                                       | 424.4556                                                            | 21.1234                                           | 30.3234                                                               |
| 197                                  | 578.7107                        | 11.4920                         | 26.3000                                        | 8.0141                                          | 433.8230                                                         | 1.0111                                                              | 12.3456                                           | 5.3300                                                                |
| 198                                  | 124.5440                        | 1.8000                          | 4.6280                                         | 1.9850                                          | 101.5300                                                         | 23.6630                                                             | 78.9101                                           | 36.3840                                                               |
| 199                                  | 3,711.2029                      | 63.4000                         | 303.3838                                       | 4.5720                                          | 2,729.9280                                                       | 22.0222                                                             | 91.0111                                           | 33.0000                                                               |
| 200                                  | 124.8600                        | 2.0000                          | 5.2240                                         | 2.3200                                          | 79.7770                                                          | 51.0000                                                             | 21.3141                                           | 2.6500                                                                |
| 201                                  | 2,466.0000                      | 26.8650                         | 161.7000                                       | 2.0202                                          | 2,028.7996                                                       | 18.4021                                                             | 3.2000                                            | 14.9000                                                               |
| 202                                  | 2,829.2991                      | 36.2000                         | 156.8000                                       | 27.1000                                         | 2,261.0000                                                       | 930.2000                                                            | 51.1617                                           | 25.2000                                                               |
| 203                                  | 814.8196                        | 14.8000                         | 48.5520                                        | 16.0111                                         | 622.9507                                                         | 204.9000                                                            | 18.1920                                           | 1.4150                                                                |
| 204                                  | 4,051.7996                      | 46.0000                         | 349.5999                                       | 2.6000                                          | 3,036.5999                                                       | 1,215.2996                                                          | 21.2223                                           | 56.6000                                                               |
| 205                                  | 67.0390                         | 28.0000                         | 3.5010                                         | 2.5170                                          | 54.9070                                                          | 66.5620                                                             | 24.2526                                           | 42.4446                                                               |
| 206                                  | 240.5670                        | 4.0000                          | 5.5670                                         | 1.3220                                          | 184.1350                                                         | 61.6900                                                             | 2.5860                                            | 3.0470                                                                |
| 207                                  | 45.2140                         | 2.0000                          | 1.4110                                         | 18.1010                                         | 38.0970                                                          | 62.3201                                                             | 27.2829                                           | 48.5052                                                               |
| 208                                  | 69.9520                         | 81.0000                         | 33.3333                                        | 8.0033                                          | 65.4570                                                          | 52.3302                                                             | 30.3132                                           | 54.5658                                                               |
| 209                                  | 54.5490                         | 1.1270                          | 1.7720                                         | 17.7200                                         | 42.5990                                                          | 42.4444                                                             | 33.3435                                           | 60.6264                                                               |
| 210                                  | 317.4480                        | 5.7840                          | 12.6650                                        | 11.0330                                         | 254.1990                                                         | 80.1010                                                             | 36.3738                                           | 66.6870                                                               |
| 211                                  | 847.9927                        | 24.0000                         | 85.0240                                        | 19.7930                                         | 664.9578                                                         | 34.1021                                                             | 39.4041                                           | 10.4000                                                               |
| 212                                  | 467.9546                        | 4.8450                          | 13.1650                                        | 2.3810                                          | 400.5806                                                         | 4.0999                                                              | 42.4344                                           | 1.0011                                                                |
| 213                                  | 126.6750                        | 14.0007                         | 7.7490                                         | 14.1441                                         | 109.6830                                                         | 50.6410                                                             | 45.4647                                           | 1.0022                                                                |
| 214                                  | 85.7290                         | 49.0000                         | 2.1610                                         | 49.4949                                         | 72.8400                                                          | 9.9901                                                              | 48.4950                                           | 1.0033                                                                |
| 215                                  | 680.7666                        | 8.2200                          | 19.2340                                        | 77.7878                                         | 578.8528                                                         | 9.8175                                                              | 51.5253                                           | 1.3090                                                                |
| 216                                  | 211.3230                        | 1.5670                          | 4.8350                                         | 15.6180                                         | 171.4130                                                         | 65.0000                                                             | 54.5556                                           | 1.8201                                                                |
| 217                                  | 254.3030                        | 3.1000                          | 2.7620                                         | 2.3570                                          | 205.8410                                                         | 42.4381                                                             | 57.5859                                           | 2.0880                                                                |
| 218                                  | 1,396.8108                      | 29.4160                         | 79.9820                                        | 28.2626                                         | 1,000.2886                                                       | 3.8107                                                              | 16.1580                                           | 3.4510                                                                |
| 219                                  | 3,981.0000                      | 52.9000                         | 188.3000                                       | 70.3000                                         | 3,120.5999                                                       | 1,085.7996                                                          | 75.8000                                           | 37.5000                                                               |
| 220                                  | 3,943.0990                      | 56.5320                         | 259.5000                                       | 49.9000                                         | 3,352.3008                                                       | 1,275.7002                                                          | 60.6162                                           | 42.3000                                                               |

**Base de datos** (Cont.)

| <i>Número<br/>de la<br/>compañía</i> | <i>Ventas<br/>(en millones)</i> | <i>Empleados<br/>(en miles)</i> | <i>Gastos<br/>de capital<br/>(en millones)</i> | <i>Gastos<br/>intangibles<br/>(en millones)</i> | <i>Costo de los<br/>productos<br/>vendidos<br/>(en millones)</i> | <i>Mano de<br/>obra y gastos<br/>relacionados<br/>(en millones)</i> | <i>Gastos de<br/>publicidad<br/>(en millones)</i> | <i>Gastos de<br/>investigación<br/>y desarrollo<br/>(en millones)</i> |
|--------------------------------------|---------------------------------|---------------------------------|------------------------------------------------|-------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------|-----------------------------------------------------------------------|
| 221                                  | 1,260.2349                      | 17.2880                         | 103.0320                                       | 11.4810                                         | 1,055.9436                                                       | 12.0000                                                             | 63.6465                                           | 2.1133                                                                |
| 222                                  | 973.2527                        | 9.8850                          | 25.4530                                        | 5.5580                                          | 848.7227                                                         | 4.0877                                                              | 66.6768                                           | 3.3210                                                                |
| 223                                  | 19.9060                         | 18.0002                         | 5.6666                                         | 1.4100                                          | 16.5170                                                          | 3.3876                                                              | 69.7071                                           | 4.2242                                                                |
| 224                                  | 66.8260                         | 1.3200                          | 6.1110                                         | 88.1388                                         | 48.9480                                                          | 4.5222                                                              | 72.7374                                           | 5.6210                                                                |
| 225                                  | 178.7460                        | 2.1980                          | 5.5430                                         | 138.0000                                        | 138.5690                                                         | 43.4350                                                             | 75.7677                                           | 6.2155                                                                |
| 226                                  | 26.7510                         | 1.0560                          | 8.8888                                         | 211.0113                                        | 17.9930                                                          | 18.1111                                                             | 78.7980                                           | 7.2102                                                                |
| 227                                  | 20.5750                         | 43.1111                         | 7.7777                                         | 82.1003                                         | 13.9720                                                          | 14.2222                                                             | 81.8283                                           | 8.9712                                                                |
| 228                                  | 51.5960                         | 18.5216                         | 1.6940                                         | 1.1620                                          | 38.8190                                                          | 88.9922                                                             | 81.0077                                           | 24.2601                                                               |
| 229                                  | 106.1150                        | 2.6000                          | 4.6850                                         | 9.9210                                          | 64.0500                                                          | 12.4243                                                             | 77.0222                                           | 23.2810                                                               |
| 230                                  | 8.5160                          | 14.2421                         | 12.0818                                        | 12.1402                                         | 5.9500                                                           | 7.8246                                                              | 22.4443                                           | 24.8588                                                               |
| 231                                  | 308.8916                        | 5.7000                          | 15.8370                                        | 13.1402                                         | 144.7340                                                         | 42.4444                                                             | 47.8877                                           | 2.7060                                                                |
| 232                                  | 753.8069                        | 16.8750                         | 37.4620                                        | 3.6210                                          | 491.1160                                                         | 210.0050                                                            | 16.4370                                           | 4.9340                                                                |
| 233                                  | 41.2960                         | 1.1080                          | 2.5820                                         | 12.1213                                         | 28.1320                                                          | 81.8118                                                             | 12.5456                                           | 24.5293                                                               |
| 234                                  | 145.6840                        | 3.4190                          | 13.3250                                        | 1.0087                                          | 105.1630                                                         | 51.7100                                                             | 51.8196                                           | 1.8480                                                                |
| 235                                  | 51.3130                         | 1.0000                          | 1.5700                                         | 8.0025                                          | 35.9730                                                          | 43.4400                                                             | 21.4231                                           | 59.6085                                                               |
| 236                                  | 21.4070                         | 12.5358                         | 18.7842                                        | 5.5554                                          | 12.9550                                                          | 12.8888                                                             | 37.8286                                           | 64.8821                                                               |
| 237                                  | 585.6597                        | 8.2000                          | 56.0530                                        | 80.9960                                         | 359.8350                                                         | 77.9999                                                             | 13.6920                                           | 8.9610                                                                |
| 238                                  | 516.7239                        | 10.3000                         | 17.9320                                        | 9.3610                                          | 376.4170                                                         | 1.1007                                                              | 5.6670                                            | 5.6000                                                                |
| 239                                  | 316.8147                        | 7.0000                          | 3.9360                                         | 12.1314                                         | 267.2456                                                         | 2.0008                                                              | 86.8686                                           | 76.7686                                                               |
| 240                                  | 509.7000                        | 10.0000                         | 27.0360                                        | 15.1617                                         | 375.3457                                                         | 179.9240                                                            | 85.8686                                           | 3.6080                                                                |
| 241                                  | 341.3887                        | 7.1270                          | 7.1570                                         | 8.1819                                          | 287.6907                                                         | 9.0007                                                              | 86.8888                                           | 86.7795                                                               |
| 242                                  | 33.0660                         | 1.0872                          | 1.9540                                         | 9.2021                                          | 24.0720                                                          | 12.7210                                                             | 83.1111                                           | 95.9594                                                               |
| 243                                  | 200.5920                        | 4.0000                          | 5.3320                                         | 20.0290                                         | 153.5480                                                         | 7.6660                                                              | 82.2222                                           | 94.9388                                                               |
| 244                                  | 184.5810                        | 4.0500                          | 7.2780                                         | 10.3570                                         | 142.7160                                                         | 8.7770                                                              | 22.6665                                           | 1.0790                                                                |
| 245                                  | 217.7520                        | 4.0880                          | 7.3840                                         | 10.1146                                         | 179.1020                                                         | 78.3910                                                             | 44.6621                                           | 89.9012                                                               |
| 246                                  | 386.8118                        | 7.4040                          | 18.4880                                        | 47.1213                                         | 302.5586                                                         | 2.9990                                                              | 18.1716                                           | 3.8620                                                                |
| 247                                  | 69.1530                         | 12.1212                         | 1.6190                                         | 48.1415                                         | 54.4310                                                          | 11.3410                                                             | 15.1413                                           | 13.8125                                                               |
| 248                                  | 81.4650                         | 1.6220                          | 4.1890                                         | 16.4950                                         | 70.5080                                                          | 4.4555                                                              | 12.1110                                           | 47.8552                                                               |
| 249                                  | 329.5518                        | 6.0040                          | 12.2520                                        | 8.0540                                          | 269.6377                                                         | 12.1417                                                             | 9.8765                                            | 51.9185                                                               |
| 250                                  | 36.3870                         | 133.0000                        | 12.7246                                        | 51.5355                                         | 27.7690                                                          | 21.8283                                                             | 4.3210                                            | 54.3321                                                               |
| 251                                  | 344.7937                        | 7.5000                          | 24.7400                                        | 57.5982                                         | 205.0610                                                         | 92.9395                                                             | 8.1234                                            | 4.8200                                                                |
| 252                                  | 22.8030                         | 84.1000                         | 2.1060                                         | 83.4952                                         | 10.6830                                                          | 96.9899                                                             | 5.6788                                            | 43.8388                                                               |
| 253                                  | 196.3030                        | 5.4660                          | 5.9730                                         | 99.9242                                         | 142.1520                                                         | 97.9294                                                             | 12.4582                                           | 2.2710                                                                |
| 254                                  | 31.5660                         | 13.7124                         | 8.1264                                         | 10.1115                                         | 22.3750                                                          | 95.0092                                                             | 14.5220                                           | 66.7274                                                               |
| 255                                  | 108.8580                        | 1.7000                          | 1.2870                                         | 92.4445                                         | 45.9130                                                          | 92.6666                                                             | 1.4330                                            | 53.5422                                                               |
| 256                                  | 83.6260                         | 1.2320                          | 4.1220                                         | 55.6677                                         | 45.0950                                                          | 92.5555                                                             | 13.5620                                           | 22.5673                                                               |
| 257                                  | 390.8726                        | 6.1660                          | 17.3310                                        | 40.5880                                         | 296.8577                                                         | 58.2130                                                             | 18.0000                                           | 10.0000                                                               |
| 258                                  | 363.9839                        | 7.0160                          | 11.2700                                        | 11.5610                                         | 234.6320                                                         | 2.7310                                                              | 6.4860                                            | 86.0000                                                               |
| 259                                  | 52.2620                         | .4420                           | 5.1030                                         | 1.1500                                          | 43.5110                                                          | 12.1440                                                             | 44.0700                                           | 16.0000                                                               |
| 260                                  | 228.6110                        | 5.6500                          | 1.8370                                         | 41.5600                                         | 161.4700                                                         | 20.5400                                                             | 14.9460                                           | 3.0730                                                                |
| 261                                  | 60.8250                         | 1.5000                          | 1.4910                                         | 45.3100                                         | 41.6820                                                          | 1.8970                                                              | 1.6590                                            | 30.7300                                                               |
| 262                                  | 16.6890                         | 40.5000                         | 57.6000                                        | 9.8540                                          | 9.8450                                                           | 18.3700                                                             | 35.2020                                           | 63.5300                                                               |
| 263                                  | 39.8290                         | 62.1000                         | 3.9900                                         | 1.5610                                          | 32.6580                                                          | 4.9080                                                              | 9.2530                                            | 30.0000                                                               |
| 264                                  | 28.9020                         | 93.4000                         | 1.1040                                         | 36.5000                                         | 23.1410                                                          | 2.5900                                                              | 18.3650                                           | 10.0000                                                               |
| 265                                  | 8.7410                          | 27.0000                         | 55.6000                                        | 32.1400                                         | 6.3700                                                           | 59.3100                                                             | 27.6600                                           | 56.6660                                                               |
| 266                                  | 61.9446                         | 7.0000                          | 35.0470                                        | 43.2000                                         | 432.3777                                                         | 160.6660                                                            | 19.2020                                           | 6.3530                                                                |

*Datos de la compensación del director ejecutivo (CEO)*

| <i>Bonificación</i><br><i>(miles de</i><br><i>de dólares)</i> | <i>Otras compen-</i><br><i>saciones</i><br><i>(miles de</i><br><i>de dólares)</i> | <i>Total de las compen-</i><br><i>saciones</i><br><i>(miles de</i><br><i>de dólares)</i> | <i>Educa-</i><br><i>Experiencia</i><br><i>lidad profesional</i><br><i>lariad riencia</i> | <i>Valoración en el</i><br><i>mercado</i><br><i>(Millones de</i><br><i>dólares)</i> | <i>Ganancias de la</i><br><i>empresa</i><br><i>(Millones de</i><br><i>dólares)</i> | <i>Ventas de la empresa</i><br><i>(Millones de</i><br><i>dólares)</i> |
|---------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| 173                                                           | 275                                                                               | 5                                                                                        | 453                                                                                      | 64                                                                                  | 3                                                                                  | 40                                                                    |
| 1,441                                                         | 429                                                                               | 78                                                                                       | 1,948                                                                                    | 55                                                                                  | 1                                                                                  | 23                                                                    |
| 1,646                                                         | 0                                                                                 | 89                                                                                       | 1,735                                                                                    | 47                                                                                  | 2                                                                                  | 7                                                                     |
| 294                                                           | 325                                                                               | 24                                                                                       | 643                                                                                      | 65                                                                                  | 1                                                                                  | 3                                                                     |
| 1,254                                                         | 105                                                                               | 102                                                                                      | 1,461                                                                                    | 63                                                                                  | 1                                                                                  | 6                                                                     |
| 325                                                           | 25                                                                                | 7                                                                                        | 357                                                                                      | 54                                                                                  | 2                                                                                  | 5                                                                     |
| 658                                                           | 0                                                                                 | 11                                                                                       | 669                                                                                      | 61                                                                                  | 2                                                                                  | 5                                                                     |
| 1,723                                                         | 289                                                                               | 82                                                                                       | 2,094                                                                                    | 63                                                                                  | 1                                                                                  | 3                                                                     |
| 504                                                           | 69                                                                                | 24                                                                                       | 597                                                                                      | 57                                                                                  | 2                                                                                  | 5                                                                     |
| 822                                                           | 38                                                                                | 29                                                                                       | 889                                                                                      | 56                                                                                  | 2                                                                                  | 1                                                                     |
| 374                                                           | 129                                                                               | 11                                                                                       | 514                                                                                      | 57                                                                                  | 2                                                                                  | 4                                                                     |
| 447                                                           | 11                                                                                | 8                                                                                        | 466                                                                                      | 48                                                                                  | 2                                                                                  | 1                                                                     |
| 2,781                                                         | 0                                                                                 | 52                                                                                       | 2,833                                                                                    | 50                                                                                  | 1                                                                                  | 3                                                                     |
| 128                                                           | 282                                                                               | 17                                                                                       | 427                                                                                      | 54                                                                                  | 1                                                                                  | 8                                                                     |
| 1,782                                                         | 0                                                                                 | 74                                                                                       | 1,856                                                                                    | 60                                                                                  | 2                                                                                  | 7                                                                     |
| 1,137                                                         | 423                                                                               | 92                                                                                       | 1,652                                                                                    | 60                                                                                  | 2                                                                                  | 1                                                                     |
| 761                                                           | 20                                                                                | 1                                                                                        | 782                                                                                      | 49                                                                                  | 2                                                                                  | 9                                                                     |
| 505                                                           | 0                                                                                 | 108                                                                                      | 613                                                                                      | 56                                                                                  | 2                                                                                  | 9                                                                     |
| 976                                                           | 448                                                                               | 64                                                                                       | 1,488                                                                                    | 58                                                                                  | 1                                                                                  | 3                                                                     |
| 434                                                           | 12                                                                                | 1                                                                                        | 447                                                                                      | 50                                                                                  | 2                                                                                  | 9                                                                     |
| 1,010                                                         | 687                                                                               | 55                                                                                       | 1,752                                                                                    | 63                                                                                  | 1                                                                                  | 7                                                                     |
| 956                                                           | 1,452                                                                             | 89                                                                                       | 2,497                                                                                    | 64                                                                                  | 0                                                                                  | 1                                                                     |
| 700                                                           | 37                                                                                | 31                                                                                       | 768                                                                                      | 60                                                                                  | 2                                                                                  | 9                                                                     |
| 1,813                                                         | 489                                                                               | 40                                                                                       | 2,342                                                                                    | 71                                                                                  | 1                                                                                  | 4                                                                     |
| 3,396                                                         | 0                                                                                 | 13                                                                                       | 3,409                                                                                    | 64                                                                                  | 0                                                                                  | 9                                                                     |

|       |       |     |       |       |    |    |    |     |    |      |      |      |      |      |      |      |      |         |       |      |        |        |
|-------|-------|-----|-------|-------|----|----|----|-----|----|------|------|------|------|------|------|------|------|---------|-------|------|--------|--------|
| 2,108 | 38    | 98  | 38    | 2,244 | 64 | 2  | 2  | 2   | 59 | 601  | 59   | 2    | 2    | 2    | 5    | 5    | 4    | 0.04    | 30    | 0.1  | 0.12   | -85    |
| 597   | 0     | 4   | 4     | 1,554 | 61 | 1  | 1  | 8   | 41 | 41   | 17   | 11   | 11   | 11   | 17   | 17   | 17   | 17      | 30,6  | 2,23 | 82     | 1,911  |
| 616   | 862   | 76  | 4     | 462   | 61 | 2  | 2  | 5   | 25 | 25   | 11   | 11   | 11   | 11   | 11   | 11   | 11   | 11      | 16,8  | 1,03 | 27     | 1,435  |
| 237   | 221   | 4   | 0     | 571   | 0  | 16 | 16 | 587 | 55 | 2    | 4    | 5    | 5    | 5    | 5    | 5    | 5    | 1,6     | 0,17  | -76  | 1,314  |        |
| 571   | 269   | 391 | 28    | 688   | 54 | 2  | 2  | 9   | 28 | 28   | 28   | 28   | 28   | 28   | 28   | 28   | 28   | 1,689,0 | 34,04 | 317  | 2,301  |        |
| 721   | 101   | 71  | 893   | 60    | 2  | 2  | 5  | 36  | 15 | 15   | 15   | 15   | 15   | 15   | 15   | 15   | 15   | 2,0     | 0,04  | 417  | 3,277  |        |
| 328   | 238   | 34  | 600   | 60    | 0  | 7  | 7  | 42  | 1  | 85,6 | 85,6 | 85,6 | 85,6 | 85,6 | 85,6 | 85,6 | 85,6 | 85,6    | 17,66 | 43   | 4,444  |        |
| 538   | 25    | 7   | 570   | 60    | 1  | 1  | 4  | 3   | 3  | 3    | 3    | 3    | 3    | 3    | 3    | 3    | 3    | 0,2     | 0,21  | 49   | 1,214  |        |
| 741   | 104   | 9   | 854   | 62    | 1  | 1  | 7  | 30  | 30 | 30   | 30   | 30   | 30   | 30   | 30   | 30   | 30   | 2,6     | 0,17  | 81   | 804    |        |
| 607   | 380   | 47  | 1,034 | 51    | 1  | 1  | 6  | 23  | 23 | 23   | 23   | 23   | 23   | 23   | 23   | 23   | 23   | 7,0     | 0,83  | 82   | 669    |        |
| 1,044 | 107   | 36  | 1,187 | 55    | 2  | 2  | 8  | 2   | 2  | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 1    | 3,1     | 1,21  | 10   | 578    |        |
| 2,409 | 1,487 | 143 | 4,039 | 55    | 1  | 1  | 7  | 32  | 32 | 32   | 32   | 32   | 32   | 32   | 32   | 32   | 32   | 35,2    | 0,29  | 715  | 1,214  |        |
| 287   | 198   | 32  | 517   | 59    | 1  | 1  | 8  | 37  | 37 | 37   | 37   | 37   | 37   | 37   | 37   | 37   | 37   | 181,0   | 6,70  | 136  | 3,180  |        |
| 567   | 15    | 34  | 616   | 51    | 2  | 2  | 1  | 16  | 16 | 16   | 16   | 16   | 16   | 16   | 16   | 16   | 16   | 0,3     | 0,01  | 237  | 2,754  |        |
| 682   | 0     | 2   | 684   | 62    | 2  | 1  | 1  | 36  | 36 | 36   | 36   | 36   | 36   | 36   | 36   | 36   | 36   | 2       | 0,8   | 0,01 | -1,086 |        |
| 1,226 | 174   | 56  | 1,456 | 52    | 1  | 4  | 4  | 30  | 30 | 30   | 30   | 30   | 30   | 30   | 30   | 30   | 30   | 10      | 6,0   | 0,17 | 98     |        |
| 952   | 80    | 56  | 1,088 | 45    | 2  | 2  | 7  | 11  | 11 | 11   | 11   | 11   | 11   | 11   | 11   | 11   | 11   | 11      | 2,6   | 0,17 | 48     | 4,439  |
| 432   | 0     | 12  | 444   | 50    | 2  | 2  | 7  | 25  | 25 | 25   | 25   | 25   | 25   | 25   | 25   | 25   | 25   | 3       | 0,1   | 0,01 | -50    | 415    |
| 1,085 | 440   | 97  | 1,622 | 57    | 1  | 8  | 8  | 29  | 29 | 29   | 29   | 29   | 29   | 29   | 29   | 29   | 29   | 5       | 9,7   | 0,19 | 347    | 1,569  |
| 1,009 | 117   | 0   | 1,126 | 64    | 1  | 3  | 3  | 28  | 28 | 28   | 28   | 28   | 28   | 28   | 28   | 28   | 28   | 14      | 3,4   | 0,21 | 63     | 9,886  |
| 1,711 | 182   | 134 | 2,027 | 62    | 2  | 2  | 2  | 25  | 25 | 25   | 25   | 25   | 25   | 25   | 25   | 25   | 25   | 5       | 3,4   | 0,04 | 806    | 2,545  |
| 408   | 183   | 9   | 600   | 52    | 2  | 5  | 7  | 12  | 12 | 12   | 12   | 12   | 12   | 12   | 12   | 12   | 12   | 24      | 4,2   | 0,26 | 10     | 8,379  |
| 543   | 13    | 2   | 558   | 62    | 2  | 5  | 5  | 34  | 34 | 34   | 34   | 34   | 34   | 34   | 34   | 34   | 34   | 12      | 0,2   | 0,01 | 265    | 21,351 |
| 278   | 209   | 74  | 561   | 61    | 2  | 6  | 6  | 24  | 24 | 24   | 24   | 24   | 24   | 24   | 24   | 24   | 24   | 15,4    | 0,95  | 52   | 2,359  |        |
|       |       |     |       |       |    |    |    |     |    |      |      |      |      |      |      |      |      |         |       |      | 695    |        |

\*0 = no tiene título universitario; 1 = estudiante universitario; 2 = título universitario



# Índice analítico

## A

AAA Washington, 200-201, 260-262, 311-313, 372-374  
Ace Manufacturing, 257-258  
Administradores y pronósticos (*véase* Pronóstico en los negocios, Aplicaciones a la administración)  
AIC (Criterio de Información de Akaike), 412-413  
Ajuste de curva-S, 78  
estacional, 18, 159, 171-172, 184  
Alcom Electronics, 50-51  
Alfa (*véase* Nivel de significancia)  
Alomega Food Stores, 94-95, 201-203, 375-376, 479-480  
Análisis en una serie de tiempo, 177  
de regresión, 32-41, 76, 77, 78, 168-170, 185-186, 211-245, 487  
aplicación a la administración del, 242-243, 298-299  
autocorrelación y, 230, 245, 324, 470  
coeficiente de correlación y, 225  
coeficiente de determinación en, 222-226, 229-243  
correlación serial y, 220  
datos de series de tiempo para el, 328-331  
de series de tiempo, 19, 58-60, 76, 77, 78, 327-355, 414-426  
definición de la recta de regresión, 212  
descomposición de la varianza y, 220-223  
ecuación para la predicción de un punto, 217  
error estándar de la estimación para, 216-217  
error estándar de la regresión, coeficiente para, 226, 245  
error estándar del pronóstico para, 218  
estadística *F*, 228-229  
estadística *t* para, 226, 244  
intervalo de predicción, 218, 224  
método de mínimos cuadrados, 221

modelo de regresión en línea recta, 216  
procedimiento de todas las regresiones posibles en, 319-321  
pronóstico de *Y*, 217-220  
prueba de hipótesis en, 226-229, recta de regresión ajustada, 44, 213, 243  
recta de regresión de la población, 196  
recta de regresión muestral en, 221  
regresión lineal simple contra regresión lineal múltiple, 242-243  
regresión lineal simple, 211-267  
regresión por pasos y, 292, 295  
residuales en, 229-233  
suma de los cuadrados de la regresión, 221, 244  
suma de los cuadrados del error, 222, 223, 224  
suma del total de los cuadrados, 222, 223, 224  
supuestos en, 220, 229, 327, 470  
transformación de variable y, 229, 233-237  
variables explicativas en, 234-237  
Anderson-Pareto, 467  
ANOVA (Análisis de Varianza), 221  
Aplicación(es) a la administración, 83-84, 130-131, 181-182, 298-299, 353, 428-429 del análisis de regresión simple, 242-243 de la estadística a las ciencias (SAS), 7  
ARIMA, modelo (*Autoregressive Integrated Moving Average*), 75, 381, 388-389, 429  
análisis estacional, 414-424  
aplicación a la administración, 428-429  
características de, 381  
criterio de Información de Akaike (AIC), 412-413  
Bayesiano (BIC) (Schwarz), 413  
diferencias y, 390  
estimación del, 391-392  
formulación y aplicación del, 389-411  
identificación del, 389-391

verificación del, mejores prácticas para la selección de parámetro, 411, 392-411  
Armstrong, 468  
Asociación Estadounidense de Economía (American Economic Association), 8  
de Estadística (American Statistical Association), 8  
Autocorrelación, 60-63, 84, 327-346, 353  
coeficiente de, 64, 85  
correlación serial de primer orden y, 328-331  
correlogramas y, 22, 63  
de k-ésimo orden, 60-61, 84  
definición de, 60, 84  
diferencias generalizadas y, 335, 342-345, 354  
error estándar del coeficiente de, 64, 85  
fórmula del coeficiente de, 64, 85  
fórmulas para, 60, 85  
método de mínimos cuadrados no lineales y, 342-345  
parcial, 383-385  
prueba Durbin-Watson para, 331-334, 353  
regresión con datos de series de tiempo y, 327-346  
segundo orden y, 61  
soluciones para, 334-346  
técnica Box-Jenkins y, 382-386

## B

B&E DataLinks, 8  
Basura, entra, sale basura (GIGO, *Garbage In, Garbage Out*), 57  
Beisbol de fantasía, 313-319 (A), 319-321 (B)  
Bell y Hillmer, 172  
Bernstein, Peter, 3, 428  
Beta (*véase* Prueba de hipótesis)  
BIC (Criterio de Información Bayesiano), 413  
Boundary Electronics, 495-496  
Buen ajuste, 35, 37  
Burns, Arthur, F., 176  
Busby Associates, 496-499

Business Week, 298  
Butcher Products, Inc., 256-257

## C

Causalidad contra correlación, 37  
Censo II, 179-181  
Chi cuadrada, 64  
Ciclo de vida del producto, 164  
Ciclo-tendencia, 158, 161  
City of College Station, 447-448  
Cleman, Robert, 468  
Coeficiente  
de autocorrelación, 64, 85  
de correlación, 35-36, 37, 38, 43, 44,  
225, 277  
fórmula para, 36, 44  
de determinación (para regresión  
lineal simple), 224-226, 229,  
243  
contra coeficiente de correlación,  
225  
fórmula para, 224, 245  
relaciones con estadística F, 245  
de regresión parcial o neta, 273-274,  
300  
Colinealidad, 269-270, 285-288, 300  
Combinación de pronósticos, 468-470  
beneficios de la, 469, 470  
por medio del promedio  
ponderado, 477  
simple, 477  
Componente  
cíclico, 59, 76-77, 84  
irregular, 184  
estacional de las series, 69, 84, 158,  
167-172  
análisis de series de tiempo y, 59-60,  
104, 158-159  
promedio móvil y, 109  
suavizamiento exponencial ajustado  
para, 158-177  
y técnica Box-Jenkins, 414-424  
irregular (en análisis de series de  
tiempo), 159, 172  
Computadoras  
acumulación de datos y, 57  
cambio tecnológico, 1, 2, 77, 477, 493  
consideraciones de costo, 77, 493  
organización de la función de pronós-  
tico y, 77, 493  
Confiabilidad (de los datos), 57  
Consejo de la Reserva Federal, 179  
Consultores, pronóstico hecho por, 493  
Consumer Credit Counseling, 94, 141,  
197, 259-260, 371-372, 444,  
499-500  
Continuo micro-macro, 3  
Control, 118-120, 131  
Correlación, 32, 34, 42, 211  
aplicación a la administración de la, 42  
causalidad contra, 37

coeficiente de, 35-36, 38, 43, 44, 225,  
277  
diagramas de dispersión y, 32-35  
matriz de, 270  
relación lineal y, 37  
serial  
de primer orden, 328, 354  
modelos autorregresivos y, 345-346  
prueba Durbin-Watson para,  
331-334  
soluciones para, 334-346  
Correlograma, 22, 63, 84  
Costos del pronóstico, 7, 77, 464, 486, 493  
Cristal Ball (CB) Predictor, 7, 8, 12,  
151-155, 457-460  
Criterio  
de información  
de Akaike (AIC, Akaike  
Information Criterion),  
412-413, 412  
Bayesiano (BIC, Bayesian  
Information Criterion), 413  
de mínimos cuadrados, 212 (véase  
también Método de mínimos  
cuadrados)  
Cuantitativo contra cualitativo, 2, 4, 6-7,  
77, 131, 485  
Cuartiles, 18  
Curvas de crecimiento, 78, 166, 237-242

## D

Datos  
acumulación de, 5-6, 57  
aleatorios, 60, 61  
atípicos, 19  
cíclicos, 59, 76-77  
confiabilidad de los, 57  
en bases de datos existentes,  
493-494  
consistencia de los, 57  
de corte transversal (mismo marco  
temporal), 19, 57, 58, 84  
estacionales, 59-60, 76, 158-159  
estacionarios, 69-71, 75-76  
horizontales, 58  
no estacionarios, 67, 103  
oportunos, 57  
precisión de los, 57  
relevancia de los, 57  
de series de tiempo, 58-60, 84, 157  
tendencia en los, 59, 67, 69, 84  
tipos de, 57  
utilidad de los, 53  
Deducción  
de fórmulas de correlación, 503  
de los estimadores de mínimos  
cuadrados, 503  
Deflación de precios, 183, 185  
Descomposición  
clásica, 76, 77, 158-177  
de la variabilidad, 220-221

en el análisis de series de tiempo,  
158-177  
Census II, método para, 179-181  
clásica, 76, 77, 158-177  
problemas con, 158  
Desviación  
cuadrática media (MSD), 164  
estándar, 16  
media absoluta, (MAD), 78, 79, 164,  
181, 185

Diagrama (s)  
de árbol, 475  
de dispersión, 21, 32-35, 43, 58  
Diferencias, construcción de, 67-69  
Distribución  
binomial, 23-24, 42, 43, 44, 507-508  
continua, 23  
chi cuadrada, 511-512  
de probabilidad, 22  
muestral, 23-28, 42, 43  
normal, 24-25, 41-42, 43, 509  
*t*, 27, 510

## E

*Early Warning Signals for the  
Economy* (Moore, Shiskin),  
177

Ecuación  
alternativa para la pendiente de la  
línea de regresión, 38, 44  
de regresión ajustada, 214, 244  
*Eight Asian Megatrends That Are  
Reshaping the World* (Naisbitt),  
472  
Ejemplos en computadora (véase Excel;  
Minitab)  
El mercado de bonos, 310-311  
Eliminación de la estacionalidad de los  
datos, 167-172  
Error (es)  
cuadrático medio (MSE), 78, 79, 81, 85,  
221, 223  
estándar  
de la estimación, 216-217, 244  
de la función de autocorrelación, 85  
del pronóstico, 244  
permitido, 28  
porcentual  
absoluto medio (MAPE), 78, 79, 81,  
85  
medio (MPE), 79, 81, 85  
pronóstico de, 79-81, 85, 102, 118-121,  
215, 216-217, 297-298, 392  
Estadística (s)  
Box-Pierce, 64  
Descriptiva (s), 15-22, 42  
estimación de, 28-29, 42  
*F*, 228-229, 245  
regresión múltiple, para, 277  
relación con el coeficiente de deter-  
minación, 245

*Q*, 64-66, 82, 85  
*t*, 27, 32, 44, 244  
Excel, ejemplos de,  
análisis de regresión, 324-326  
múltiple, 265-266  
ARIMA, 457-460  
regresiones de una serie de tiempo,  
377-379

**F**

*Federal Reserve Bulletin*, 157  
Fildes, R., 77  
Filtración adaptativa, 78  
Fórmula binomial, 23, 44  
Formulación  
de escenarios, 467-468  
del modelo, 6, 389-391  
*Fortune*, 19, 298, 500  
*Future Shock* (Toffler), 472  
*500-Year Delta: What Happens After What Comes Next* (Taylor), 472

**G**

Gates, Bill, 472  
Georgoff y Murdick, 464  
Gibson, Rowan, 472  
GIGO (entra basura, sale basura), 57  
*Global Paradox* (Naisbitt), 472  
Golden Gardens Restaurant, 478-479  
Grados de libertad, 16-17, 27-28, 43  
Gráfica  
de caja, 19-20  
de cuadrados normales, 42  
de puntos, 19

**H**

Heteroscedasticidad, 220, 346-349  
Hibon, M., 77n  
Hipótesis nula, 29-30  
Histograma, 20-21, 229  
Holt, método de, 83, 121-126, 127, 132  
Horizonte de tiempo, 77, 78

**I**

Implementación del modelo, 6  
Indicadores  
coincidentes, 177  
principales, 78, 176  
retrasados, 177  
Índice  
de Actividad para el condado de Spokane, 363-366  
de precios, 182-183  
Interferencia estadística, 15  
*International Journal of Forecasting*, 468  
Intersección con eje Y, 37, 44, 213, 244

Intervalo  
de confianza, (intervalo estimado), 28, 29, 43, 44  
en regresión, 217-220  
fórmula para la media en, 29, 44  
de predicción de 95% de confianza para muestras grandes, 244  
para la predicción, 218

**J**

Juicio (subjetivo) personal o intuitivo (*véase* Pronóstico en los negocios, aspectos del juicio sobre)

**K**

Kamp, Di, 472

**L**

Ljung (y Box), 64, 82, 85, 109, 124

**M**

M3-IJF, competencia, 77  
MA (*véase* Promedio móvil)  
MAD (*véase* Desviación absoluta media)  
Mahoud, E., 468  
Makridakis, Spyros, 7, 77, 179, 464, 468  
Martin, Chuck, 472  
Matriz de correlación, 270  
Media (estadística), 15 (*véase también* Promedio)  
Medidas de tendencia central (*véase* Media, Mediana), 17  
Megatrends 2000 (Naisbitt y Aburdene), 472

Método(s)  
de mínimos cuadrados, 37-38, 44, 161  
análisis de regresión con, 213-214  
fórmula para, 243  
Delphi, 464-467  
características clave del, 465  
ventaja del, 467  
informales, 75, 78, 101-105, 132  
para seleccionar la ecuación de regresión múltiple, 288  
de pronóstico por promedios, 15, 39, 75, 101, 105-113, 129, 130, 131  
de suavizamiento exponencial, 76, 78  
ajustado a la tendencia (Holt), 83, 121-126, 127, 132  
ajustado a la tendencia y a la variación estacional (Winters), 126-129, 133  
aplicación a la administración, 130-131  
control, 118-121  
de pronóstico a corto plazo, 129  
(*véase también* Pronóstico, método de)

Microcomputadoras (*véase* Computadoras)  
Minitab, ejemplos  
análisis de regresión, 39-40, 262-265  
ARIMA, 453-457  
autocorrelación y diferencias, 95-97  
descomposición de series de tiempo, 203-206

estadística descriptiva, 53-55  
promedio móvil, 148-150  
regresión con series de tiempo, 376-377  
regresión por pasos, 321-324  
valoración de la normalidad, 41-42

Mitchell, Wesley, 176

Modelo  
aleatorio, 85  
autorregresivos (AR), 345-346, 353  
aplicaciones a la administración, 428-429

autocorrelación, 382-386  
técnica Box-Jenkins y, 386  
integrados de promedio móvil (*véase* ARIMA)

de componentes aditivos, 159, 168, 171, 179, 184  
de componentes no observados, 159  
de crecimiento de Gompertz, 78, 166  
de regresión en línea recta, 216  
de ruido blanco, 65  
de suavizamiento de Holt, 83, 121-126, 127, 132

de Winters, 126-129, 133  
económicos, 76, 78, 352-353  
logístico Pearl-Reed, 166  
multiplicativo, 159, 168

*Monthly Labor Review*, 157

Moore, Geoffrey H., 177

Morgenthau, Hery, 176

MSE (*véase* Error cuadrático medio)

Muestra, 15, 36

Multicolinealidad, 269-270

Murphy Brothers Furniture, 90-92, 141-142, 197-200

**N**

Naisbitt, John  
*Eight Asian Megatrends that are Reshaping the World* (cinta de audio), 472

*Global Paradox*, 472

*Megatrends 2000* (con Patricia

Aburdene), 472

*Megatrends*, 472

NBER, 176, 177

Negocios (*véase también* Pronóstico en los negocios)

indicadores del ciclo de, 176, 177, 183

métodos de pronóstico en los,

autorregresión (*véase* Modelos de promedios móviles autorregresivos integrados) (ARIMA),

criterio de selección del modelo, 412-413  
de suavizamiento para (*véase* Métodos de suavizamiento exponencial)  
exponentiales para, 78, 83, 101, 105, 113, 129-130, 131  
formulación y evaluación del modelo, 6  
informales para, 78, 101-105, 132  
modelos estacionales, 414-424  
resumen de, 78, 487-488  
tipos de, 3-4  
*Net Future: The 7 Cybertrends that Hill Drive Your Business, Create New Wealth, and Define Your Future* (Martin), 472  
NeuralWare, Inc., 471  
Nivel  
    de confianza, 28  
    de significancia, 30, 31  
Normalidad, valoración de, 41  
Número índice (para variación estacional), 168, 183

**O**

Oficina Nacional de Investigación Económica (NBER, National Bureau of Economic Research), 176, 177

**P**

*p*, valor (*p-value*), 31, 42, 43  
Parente, 467  
Pendiente, en relación lineal, 37  
*Plug-compatible*, 471  
Población (estadística), 15, 36  
Políticas de pronóstico, 10  
Precios  
    índice de, 182-183  
    deflación de, 183, 185  
Predicción,  
Pricewaterhouse Coopers, 473  
Primera diferencia, 338, 354  
Principio de parsimonia, 391, 429  
Probabilidad de la significancia, 31  
Procedimiento  
    de regresión por pasos, 292, 295  
    de todas las regresiones posibles, 319-321  
Promedio  
    móvil (MA), modelos, 75, 76, 78, 107-113, 131, 387-388  
        doble, 109-113, 132  
        simple, 101, 132  
    simple, 105-107, 130, 131  
Pronóstico(s)  
    a corto plazo, 77, 167, 172, 177, 486  
    a largo plazo, 77, 181-182, 486  
    econométrico, 352-353

en los negocios (*véase* Pronóstico en los negocios)  
intermedios, 77, 486  
macro, 3, 4  
micro, 4  
necesidad de, 1, 2, 7, 71, 78, 130-131, 157, 353, 428, 486  
por apreciación. (*véase* Pronóstico en los negocios, aspectos del juicio sobre)  
sobre la macroeconomía, 4, 59, 172  
tecnológicos, 463, 473  
Pronóstico en los negocios,  
    aceptación por parte de la administración, 9-10, 77  
    administración de, 6-7, 71, 75, 485-495  
    los sistemas de información (MIS) y, 493-494  
    análisis de series de tiempo y, 19, 58-60, 78  
    aplicaciones a la administración de los, 83-84, 130-131, 181-182, 242-243, 298-299, 353, 428-429  
    aspectos de juicio sobre los, 1, 4, 6, 8, 9-10, 37, 101, 105, 109, 157-158, 219, 229, 238-239, 294-295, 327, 411, 463-477, 485  
    combinación de pronósticos, 469, 470  
    correlación y, 32, 34, 42, 211  
    en el corto plazo, 77, 167, 172, 177, 486  
    costo del, 7, 77, 464, 486, 493  
    cuantitativo contra cualitativo, 2, 4, 6-7, 77, 131, 485  
    datos para los, 5-6, 57-84  
    evaluación de, 6  
    formulación  
        de escenarios en, 467-468  
        del problema y recolección de datos en, 5-6  
    futuro de, 494-495  
    historia de los, 1  
    horizontes de tiempo en, 77, 78  
    intuición en (*véase* Pronóstico en los negocios, aspectos de juicio sobre)  
    juicio personal en (*véase* Pronóstico en los negocios, aspectos de juicio sobre)  
    largo plazo, 77, 181-182, 486  
    manipulación y limpieza de datos y, 6  
    mediano plazo, 77, 486  
    medición del error en, 79-81, 85, 102, 118-121, 216-217, 297-298, 392  
    método  
        Delphi, 464-467  
        de promedio para, 15, 39, 75, 101, 105-113, 129, 130, 131  
    necesidad de los, 1, 2, 7, 71, 78, 130-131, 157, 353, 428, 486  
    pasos en, 5-6, 491-492  
    políticas de, 10  
    precisión de las técnicas de, 81-83  
procedimiento de la curva de crecimiento, 76, 78, 166, 237-242  
proceso de, 5-6, 485-495  
    pronóstico sobre la macroeconomía, 4  
redes neurales y, 470-471  
reglas para, 9  
regresión múltiple y, 269-326  
residuales en, 79, 84, 85, 213, 229-233, 392, 410  
responsabilidad en, 492-493  
selección y evaluación de métodos para, 4-5, 71, 75-79, 101, 118, 121, 131, 167, 172, 243-274, 485-486, 487-488  
subjetivo (*véase* Pronóstico en los negocios, Aspectos de juicio sobre)  
tecnológico, 463, 473  
teoría de decisiones y, 474-477  
valor esperado y, 22, 43, 473-474, 477  
variables independientes contra variables dependientes, 211, 229, 269-270. (*véase también* Autocorrelación; Colinealidad; Correlación serial)  
y análisis de regresión, 32, 42, 78, 211-243, 269-299, 327-348, 485-486, 487  
Proyección del ingreso a cinco años para Downtown Radiology, 142-148  
Prueba  
    Anderson-Darling, 41, 42  
    de hipótesis, 29-32, 41-42, 226-229  
    Durbin-Watson, 333-334, 353, 354, 514  
Publicaciones sobre el cambio tecnológico, 472

**Q**

Quenouille, M. H., 64

**R**

Rand Corporation y el método Delphi, 464-467  
Rango, 18  
    inter-cuartil, 18  
Razón a método de promedio móvil (lineal simple), 168-170  
Recta  
    de mínimos cuadrados, 213  
    de regresión adaptada, 213, 244  
    de regresión, 212-216  
Redes neurales (y pronósticos), 470-471  
Regresión lineal simple, 211-245  
Regresión múltiple, 76, 77, 78, 269-326, 300, 487  
    aplicación a la administración, 298-299  
    coeficiente  
        de correlación, 277, 301

de determinación (ajustado), 278, 300, 301  
 de regresión parcial o neta en, 273, 274, 300  
 contra regresión simple, 271  
 definición de, 269  
 ecuación de, selección de la mejor, 288-295  
 error estándar de la estimación en, 275-276, 300  
 estadística  
   *F* para, 277-300  
   *t* para, 279, 301  
 factor de expansión de la varianza, 285, 301  
 función de, de la población, 271, 300  
 función de, 399  
 grados de libertad en, 277, 300  
 interferencia para modelos de, 274-279  
 interpretación de, 295-297  
 matriz de correlación y, 270  
 modelos para, 271-273  
 multicolinealidad y, 269-270, 285-288  
 procedimiento de la  
   de todas las regresiones posibles en, 319-321  
   por pasos, 292, 295, 300  
 pronóstico econométrico y, 298-299  
 residuales y, 295-297  
 selección de ecuación en, 288-295  
 sobreajuste en, 297-298, 300  
 tabla ANOVA, 276  
 variables  
   explicativas en, 269-270, 278-280  
   ficticias en, 281-285, 300  
 Relación  
   curva de, 34  
   lineal, 35-37, 38, 167, 184, 211  
   negativa  
     imperfecta, 34  
     perfecta, 34  
     positiva perfecta, 33-34  
   X - Y, 214. (*véase también* Análisis de regresión)  
 Residuales, 79, 84, 85, 213, 229-233, 392, 410  
*Rethinking the Future: Rethinking Business, Principles, Competition, Control and Complexity, Leadership, Markets and the World* (Toffler y Toffler, edit. Gibson), 429  
*Road Ahead, The* (Gates), 472

**S**

Señal de control, 490-491  
 Series  
   aleatorias, 63  
   de tiempo estacionarias, 67, 75, 84  
 Shiskin, Julius, 177, 179  
 Simulación por computadora, 476-477  
*Small Engine Doctor*, 191-193  
 Sobreajuste, 279-298  
 Software para el pronóstico, 7-8  
*Solar Alternative Company*, 139-140  
*Statistical Abstract of the United States*, 157  
 Suavizamiento exponencial  
   cuadrático, 78, 164, 184  
   lineal, 78  
   simple, 114, 132  
 Supuesto(s)  
   de semejanza pasado-futuro, 5  
   en el análisis de regresión, 220, 229, 327, 470  
*Survey of Current Business, The*, 90, 176

**T**

Taylor, Jim, 472  
*Technology Forecast*, 473  
 Técnica(s)  
   Box-Jenkins, 1, 75, 381-429  
   cuantitativas (*véase* Pronóstico en los negocios, aspectos del juicio sobre)  
 Tendencia central, medidas de, 18, 42  
 Teorema  
   de Bayes, 474-476, 477  
   del límite central, 26-27  
 Teoría de la decisión, 474-477  
*The Lydia E. Pinkham Medicine Company*, 444-447, 480-482  
*Third Wave, The*, (Toffler), 472  
 Tiger Transport, 254- 256  
 Transformación de variables, 229, 233-237  
*21st Century Manager: Future-Focused Skills for the Next Millennium* (Kamp), 472

**U**

UPS, Air Finance Division, 448-452

**V**

Valor  
   esperado, 22, 43, 473-474, 477  
   de variable aleatoria, 22  
   probable (valor-p), 31, 42, 43  
 Variable(s)  
   aleatoria, 22  
   continua, 22, 43  
   discreta, 22, 43  
   dependiente, 211, 269  
   endógena, 352  
   exógena, 352  
   explicativa, 234-237, 269-270, 327  
   ficticia, 209, 281- 285, 300, 349  
   independiente, 211, 229-230, 443  
   indicadoras (*véase* Variables ficticias)  
   retrasadas, 345 (*véase también* Modelos autorregresivos)  
 Variaciones cíclicas, 158, 172, 177, 184  
 Varianza, 16  
 Ventas de un restaurante, 367-369

**W**

*Wall Street Journal*, 298  
*What Comes Next* (Taylor), 472  
 Winters, método de, 126-129, 133  
 World Wide Web, 8, 157, 494

**X**

X-12 ARIMA, 179-181

**Y**

Y, pronóstico de, 217-220

**Z**

Z como número de varianzas, 24, 25, 36, 509

## ACUERDO DE LICENCIA Y GARANTÍA LIMITADA.

**LEA ESTA LICENCIA CON CUIDADO ANTES DE USAR ESTE PAQUETE.** AL USARLO, USTED ACEPTA LOS TÉRMINOS Y CONDICIONES DE ESTA LICENCIA. SI USTED NO ESTÁ DE ACUERDO, NO USE EL PAQUETE. DEVUELVA RÁPIDAMENTE EL PAQUETE CERRADO Y SIN USAR ASÍ COMO TODOS LOS ARTÍCULOS QUE LO ACOMPAÑAN AL LUGAR DONDE LO OBTUVO. **ESTOS TÉRMINOS SE APLICAN A TODO EL SOFTWARE AUTORIZADO EN EL DISCO SALVO QUE LOS TÉRMINOS PARA EL USO DE CUALQUIER PROGRAMA COMPARTIDO (SHAREWARE) O PROGRAMA GRATUITO (FREEWARE) EN LOS DISCOS SE EXPONGAN EN LA LICENCIA ELECTRÓNICA LOCALIZADA EN EL DISCO.**

**1. ACUERDO DE LICENCIA Y PROPIEDAD:** Prentice-Hall, Inc. ("nosotros" o la "compañía") otorga a usted una licencia de uso de los programas de computadora y los datos incluidos (el "software"), en consideración a su compra o adopción de los libros de texto y/o otros materiales de la compañía que lo acompañan, y su acuerdo con estos términos, el software no le es vendido a usted. Nos reservamos cualquier derecho no concedido a usted. Usted es dueño sólo del disco o discos pero nosotros y/o los licenciatarios tenemos la propiedad del software como tal. La licencia permite a los individuos que han comprado el libro de texto que acompaña al software de la compañía, a usar y mostrar su copia del software en una sola computadora (es decir una sola UCP) en una única ubicación solamente para uso académico, en tanto cumpla con los términos de este Acuerdo. Usted puede hacer una copia de respaldo de seguridad, o transferir su copia a otra UCP, con la condición de que el software sea utilizable en una sola computadora. Esta licencia permite (únicamente) a instructores en instituciones educativas que han adoptado el libro de texto de la compañía que acompaña al software, a instalar, usar y mostrar la copia adjunta del software en computadoras individuales en el laboratorio de cómputo diseñado para ser utilizado por cualquier estudiante de un curso que requiera el libro de texto de la compañía que acompaña al software, y sólo mientras tal libro de texto sea un texto requerido para tal curso, en un solo campus, cursarial o posición geográfica de una institución educativa, únicamente para uso académico, mientras usted cumpla con los términos de este Acuerdo.

No obstante lo anterior, el acuerdo anterior de licencia no se aplicará a aquella parte del software cuyo título actual es Crystal Ball Professional 2000. El uso de Crystal Ball Professional 2000 es determinado por el Acuerdo Electrónico de Licencia de Usuario Final de Crystal Ball Professional 2000 que se incluye en este disco y que debe aceptarse antes de ser utilizado. Además, no obstante lo anterior, dichos términos acerca del **uso del laboratorio de cómputo** no se aplicarán a aquella parte del software actualmente titulado Crystal Ball Professional 2000. El uso de Crystal Ball Professional 2000 en un laboratorio de cómputo está expresamente prohibido por esta licencia. Para detalles de Crystal Ball Professional 2000, incluyendo licencias de laboratorio para tal producto, por favor lea el Acuerdo Electrónico de Licencia de Usuario Final para Crystal Ball Professional 2000 que se incluye en este disco. Si alguna institución académica desea obtener una Licencia de Laboratorio, por favor póngase en contacto con Decisioneering en el siguiente número: 800-289-2550.

**2. RESTRICCIONES:** Usted no está autorizado a transferir o distribuir el software o la documentación a nadie más. Usted no está autorizado a copiar la documentación o el software. Usted no está autorizado para utilizar ingeniería inversa, desmontar, descompilar, modificar, adaptar, traducir, o crear trabajos derivados basados en el software o la documentación. Usted puede ser legalmente responsable de cualquier copia o infracción de los derechos de reproducción que sea causada por su falta de cumplimiento de los términos de estas restricciones.

**3. TERMINACIÓN:** Esta licencia es efectiva hasta que se termine. Esta licencia terminará automáticamente sin previo aviso de la compañía si usted no cumple con alguna provisión o limitación de esta licencia. En caso de terminación, usted destruirá la documentación y todas las copias del software. Todas las provisiones de este Acuerdo en cuanto a limitación y descargo de responsabilidad de las garantías, limitación de responsabilidad, remedios o daños, y nuestros derechos de propiedad serán válidos después de la terminación.

**4. GARANTÍA LIMITADA Y EXENCIÓN DE RESPONSABILIDAD DE LA GARANTÍA:** La compañía garantiza que durante un periodo de 60 días posteriores a la fecha de compra de este software (o de la compra o adopción del libro de texto que lo acompaña), el software, cuando es instalado correctamente y usado de acuerdo con la documentación, funcionará en conformidad sustancial con la descripción del software que se indica en la documentación, y que, por un periodo de 30 días, el disco (o discos) en que se entrega el software, estará libre de defectos en materiales y mano de obra cuando se use de manera normal. La compañía no le garantiza que el software satisfaga sus exigencias o que la operación del software sea ininterrumpida o libre de errores. La única solución y la única obligación de la compañía conforme a estas garantías limitadas es, en la opinión de la compañía, la devolución del disco para un reembolso de cualquier cantidad pagada por éste o el reemplazo del disco. **ESTA GARANTÍA LIMITADA ES LA ÚNICA GARANTÍA PROPORCIONADA POR LA COMPAÑÍA Y SUS LICENCIADORES, Y LA COMPAÑÍA Y SUS LICENCIADORES RECHAZAN CUALQUIER OTRA GARANTÍA, EXPRESA O IMPLÍCITA, INCLUIDA LA GARANTÍA ILIMITADA, LAS GARANTÍAS IMPLÍCITAS DE COMERCIALIDAD Y APTITUD PARA UN OBJETIVO PARTICULAR. LA COMPAÑÍA NO GARANTIZA, O HACE REPRESENTACIÓN ALGUNA EN CUANTO A LA EXACTITUD, FIABILIDAD, ACTUALIDAD, USO O RESULTADOS DEL USO, DEL SOFTWARE.**

**5. LIMITACIÓN DE SOLUCIONES Y DAÑOS:** BAJO NINGUNA CIRCUNSTANCIA, LA COMPAÑÍA O SUS EMPLEADOS, AGENTES, LICENCIADORES, O CONTRATISTAS SERÁN RESPONSABLES DE DAÑO SECUNDARIO, INDIRECTO, ESPECIAL O RESULTANTE ALGUNO, QUE SURJAN POR CAUSA DE O EN RELACIÓN A ESTA LICENCIA O EL SOFTWARE, INCLUYENDO LA PÉRDIDA DE USO, LA PÉRDIDA DE DATOS, LA PÉRDIDA DE INGRESOS O GANANCIAS U OTRAS PÉRDIDAS SOSTENIDAS COMO CONSECUENCIA DE LA HERIDA DE ALGUNA PERSONA, O PÉRDIDA DE, O DAÑO A PROPIEDAD, O RECLAMACIONES DE TERCEROS, INCLUSO SI LA COMPAÑÍA O UN REPRESENTANTE AUTORIZADO DE LA COMPAÑÍA HAN SIDO INFORMADOS DE LA POSIBILIDAD DE TALES DAÑOS. BAJO NINGUNA CIRCUNSTANCIA, LA COMPAÑÍA SERÁ RESPONSABLE POR DAÑOS QUE ASCIENDAN A UN MONTO SUPERIOR A LA CANTIDAD REALMENTE PAGADA POR USTED, SI FUERA EL CASO, POR EL SOFTWARE O EL LIBRO DE TEXTO QUE LO ACOMPAÑA. ALGUNAS JURISDICCIONES NO PERMITEN LA LIMITACIÓN DE LA RESPONSABILIDAD EN CIERTAS CIRCUNSTANCIAS, ES POSIBLE QUE LAS LIMITACIONES MENCIONADAS ANTERIORMENTE NO SIEMPRE SE PUEDAN APLICAR.

**6. GENERAL:** ESTE ACUERDO SERÁ INTERPRETADO DE ACUERDO CON LAS LEYES DE LOS ESTADOS UNIDOS DE AMÉRICA Y EL ESTADO DE NUEVA YORK, APLICABLE A CONTRATOS HECHOS EN NUEVA YORK, Y BENEFICIARÁ A LA COMPAÑÍA, SUS AFILIADOS Y CESIONARIOS. Este Acuerdo es la declaración completa y exclusiva del acuerdo entre usted y la compañía y reemplaza toda oferta, acuerdo previo oral o escrito, y cualquier otra comunicación entre usted y la compañía o cualquiera de sus representantes acerca de la materia. Si usted es un usuario del gobierno estadounidense, este software se autoriza "con derechos restringidos" como se expone en los subpárrafos (a)-(d) de la cláusula de Derechos Restringidos para Computadoras Comerciales (Commercial Computer-Restricted Rights) en FAR 52.227-19 o en los subpárrafos (c)(1)(ii) de la cláusula de los Derechos para los Datos Técnicos y el Software de Computadora (Rights in Technical Data and Computer Software) en DFARS 252-227-7013, y cláusulas similares, como corresponda.

Si usted tiene cualquier pregunta acerca de este acuerdo o si usted desea ponerse en contacto con la compañía por alguna razón, por favor hágalo por escrito a:

Director, Media Production  
Pearson Education  
1 Lake Street  
Upper Saddle River, NJ 07458

Este software está protegido por los derechos de reproducción y usted tiene autorización para usarlo y mostrarlo únicamente para su uso privado, a menos que el software le fuera expresamente vendido para utilizar en una red de cómputo. Usted no tiene autorización para transferir o distribuir el software a alguna otra persona o institución por medio alguno, lo que incluye su distribución por Internet. Usted no tiene autorización para copiar el software o la documentación excepto para realizar una copia de seguridad. Usted no tiene autorización para utilizar ingeniería inversa, desmontar, descompilar, modificar, adaptar, traducir o crear trabajos derivados del software o la documentación. Usted puede ser legalmente responsable de cualquier copia o transferencia ilegal.