

PRONÓSTICOS EN LOS NEGOCIOS

NOVENA EDICIÓN

PEARSON

JOHN E. HANKE
DEAN W. WICHERN

Novena
edición

PRONÓSTICOS EN LOS NEGOCIOS

John E. Hanke

Profesor emérito de Eastern Washington University

Dean W. Wichern

Profesor emérito de Texas A&M University

Traducción

Antonio P. Enríquez Brito

Traductor especialista en estadística

Revisión técnica

Ofelia Vizcaíno Díaz

*Instituto Tecnológico y de Estudios Superiores de Monterrey,
campus Ciudad de México*

José Antonio Rosas Ferrer

*Instituto Tecnológico y de Estudios Superiores de Monterrey,
campus Monterrey*

Prentice Hall

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

HANKE, JOHN E.

Pronósticos en los negocios. 9a. edición

PEARSON EDUCACIÓN, México, 2010

ISBN: 978-607-442-700-4

Formato: 20 × 25.5 cm

Páginas: 576

Authorized translation from the English language editions, entitled *Business forecasting 9th ed.* by John E. Hanke and Dean W. Wichern published by Pearson Education, Inc., publishing as Prentice Hall, Copyright © 2009. All rights reserved.
ISBN 9780132301206

Traducción autorizada de la edición en idioma inglés, *Business forecasting 9a ed. por John E. Hanke y Dean W. Wichern* publicada por Pearson Education, Inc., publicada como Prentice Hall, Copyright © 2009. Todos los derechos reservados.

Esta edición en español es la única autorizada.

Edición en español

Editor: Rubén Fuerte Rivera
e-mail: ruben.fuerte@pearsoned.com
Editor de desarrollo: Felipe Hernández Carrasco
Supervisor de producción: José D. Hernández Garduño

Edición en inglés

AVP/Executive Editor: Mark Pfaltzgraff
AVP/Editor-in-Chief: Eric Svendsen
Manager, Product Development: Ashley Santora
Permissions Manager: Charles Morris
Editorial Project Manager: Susie Abraham
Editorial Assistant: Vanessa Bain
Marketing Manager: Anne Howard
Marketing Assistant: Susan Osterlitz
Senior Managing Editor: Judy Leale
Production Project Manager: Debbie Ryan

Sr. Operations Supervisor: Arnold Vila
Operations Specialist: Michelle Klein
Cover Design: Jayne Conte
Designer: Kiwi Design
Cover Illustration/Photo: Getty Images, Inc.
Composition Full-Service Project Management: GGS Book Services
Printer/Binder: Courier/Westford
Typeface: Times Ten-Roman, 10pt.

NOVENA EDICIÓN, 2010

D.R. © 2010 por Pearson Educación de México, S.A. de C.V.
Atlacomulco 500-5° piso
Industrial Atoto
53519, Naucalpan de Juárez, Edo. de México
E-mail: editorial.universidades@pearsoned.com

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

Prentice Hall es una marca registrada de Pearson Educación de México, S.A. de C.V.

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 VERSIÓN IMPRESA 978-607-442-700-4

ISBN VERSIÓN E-BOOK 978-607-442-701-1

ISBN VERSIÓN E-CHAPTER 978-607-442-702-8

Impreso en México. *Printed in Mexico.*

1 2 3 4 5 6 7 8 9 0 - 11 10 09 08

Prentice Hall
es una marca de

PEARSON

Dedicado a la memoria de Harry
(quien, en realidad, no quería leerlo);

a Gerry
(quien no necesita leerlo);

a Donna y Dorothy
(quienes probablemente no lo leerán);

a Katrina, Michael y Andrew
(quienes tal vez lo lean);

a Kevin
(quien dice que ya lo leyó);

y a todos los estudiantes
(a quienes les convendría leerlo).

Resumen de contenido

Prefacio xv

CAPÍTULO 1	Introducción a los pronósticos en los negocios	1
CAPÍTULO 2	Repaso de conceptos estadísticos básicos	15
CAPÍTULO 3	Exploración de patrones de datos e introducción a las técnicas de pronósticos	61
CAPÍTULO 4	Métodos de promedios móviles y de suavización	107
CAPÍTULO 5	Series de tiempo y sus componentes	165
CAPÍTULO 6	Regresión lineal simple	221
CAPÍTULO 7	Análisis de regresión múltiple	281
CAPÍTULO 8	Regresión con datos de series de tiempo	339
CAPÍTULO 9	La metodología Box-Jenkins (ARIMA)	399
CAPÍTULO 10	Pronósticos de juicio y ajuste de pronóstico	481
CAPÍTULO 11	Administración del proceso de pronósticos	503
APÉNDICE A	Datos del caso 7-1	521
APÉNDICE B	Tablas	523
APÉNDICE C	Conjunto de datos y bases de datos	533

Índice 547

Contenido

Prefacio xv

CAPÍTULO 1 Introducción a los pronósticos en los negocios 1

La historia de los pronósticos en los negocios	1
¿Es necesario el pronóstico en los negocios?	2
Tipos de pronósticos	2
Consideraciones macroeconómicas del pronóstico	3
Selección de un método de pronósticos	4
Etapas del pronóstico	4
Administración del proceso de elaboración del pronóstico	6
Software de pronósticos	6
Información en línea	7
Ejemplos de elaboración de pronósticos	7
Resumen	9
Caso 1-1: Mr. Tux	10
Caso 1-2: Consumer Credit Counseling	10
Aplicaciones de Minitab	11
Aplicaciones de Excel	12
Referencias	12

CAPÍTULO 2 Repaso de conceptos estadísticos básicos 15

Descripción de datos con resúmenes numéricos	15
Presentación de información numérica	19
Distribuciones de probabilidad	22
Distribuciones muestrales	26
Inferencia de una muestra	29
<i>Estimación</i>	29
<i>Prueba de hipótesis</i>	30
<i>Valor p</i>	32

Análisis de correlación	34
<i>Diagramas de dispersión</i>	34
<i>Coefficiente de correlación</i>	37
Ajuste de una línea recta	39
Evaluación de la normalidad	42
Aplicación en la administración	44
Glosario	44
Fórmulas clave	45
Problemas	46
Caso 2-1: Alcam Electronics	53
Caso 2-2: Mr. Tux	54
Caso 2-3: Alomega Food Stores	56
Aplicaciones de Minitab	56
Aplicaciones de Excel	58
Referencias	60

CAPÍTULO 3 Exploración de patrones de datos e introducción a las técnicas de pronósticos 61

Estudio de patrones de datos en las series de tiempo	62
Exploración de patrones de datos con análisis de autocorrelación	64
<i>¿Los datos son aleatorios?</i>	69
<i>¿Los datos muestran una tendencia?</i>	72
<i>¿Los datos son estacionales?</i>	76
Selección de una técnica de pronósticos	76
<i>Técnicas de pronósticos para datos estacionarios</i>	78
<i>Técnicas de pronósticos para datos con una tendencia</i>	78
<i>Técnicas de pronósticos para datos estacionales</i>	79
<i>Técnicas de pronósticos para series cíclicas</i>	79
<i>Otros factores a considerar en la selección de una técnica de pronóstico</i>	79
<i>Evaluación empírica de métodos para pronosticar</i>	81
Medición del error de pronóstico	81
Determinación de una técnica adecuada de pronóstico	84
Aplicación en la administración	86
Glosario	87
Fórmulas clave	87
Problemas	88
Caso 3-1A: Murphy Brothers Furniture	94
Caso 3-1B: Murphy Brothers Furniture	96
Caso 3-2: Mr. Tux	97
Caso 3-3: Consumer Credit Counseling	98
Caso 3-4: Alomega Food Stores	99

Caso 3-5: Surtido Cookies	100
Aplicaciones de Minitab	101
Aplicaciones de Excel	103
Referencias	105

CAPÍTULO 4 Métodos de promedios móviles y de suavización 107

Modelos informales	108
Métodos de pronósticos basados en promedios	111
<i>Promedios simples</i>	111
<i>Promedios móviles</i>	113
<i>Promedios móviles dobles</i>	116
Métodos de suavización exponencial	119
<i>Suavización exponencial ajustada a la tendencia: Método de Holt</i>	126
<i>Suavización exponencial ajustada a la tendencia y a la variación estacional: Método de Winters</i>	130
Aplicación a la administración	135
Glosario	136
Fórmulas clave	136
Problemas	138
Caso 4-1: The Solar Alternative Company	145
Caso 4-2: Mr. Tux	147
Caso 4-3: Consumer Credit Counseling	148
Caso 4-4: Murphy Brothers Furniture	148
Caso 4-5: Proyección de cinco años de ingresos para Downtown radiology	149
Caso 4-6: Minorista en la Web	154
Caso 4-7: Southwest Medical Center	158
Caso 4-8: Surtido Cookies	159
Aplicaciones de Minitab	159
Aplicaciones de Excel	161
Referencias	163

CAPÍTULO 5 Series de tiempo y sus componentes 165

Descomposición	166
<i>Tendencia</i>	168
<i>Curvas de tendencia no lineal</i>	171
<i>Pronóstico de la tendencia</i>	174
<i>Estacionalidad</i>	175
<i>Datos ajustados a la estacionalidad</i>	179
<i>Variaciones cíclicas e irregulares</i>	180
<i>Ejemplo de resumen</i>	180
<i>Indicadores de negocios</i>	184
Pronóstico de una serie de tiempo estacional	185

Método de descomposición de censo II	187
Aplicación a la administración	189
Apéndice: Índice de precios	190
Glosario	192
Fórmulas clave	192
Problemas	193
Caso 5-1: The Small Engine Doctor	201
Caso 5-2: Mr. Tux	202
Caso 5-3: Consumer Credit Counseling	206
Caso 5-4: Murphy Brothers Furniture	207
Caso 5-5: AAA Washington	210
Caso 5-6: Alomega Food Stores	212
Caso 5-7: Surtido Cookies	213
Caso 5-8: Southwest Medical Center	214
Aplicaciones de Minitab	214
Aplicaciones de Excel	217
Referencias	219

CAPÍTULO 6 Regresión lineal simple 221

Línea de regresión	222
Error estándar de la estimación	226
Pronóstico de Y	227
Descomposición de la varianza	230
Coeficiente de determinación	234
Prueba de hipótesis	236
Análisis de residuos	239
Resultado de computadora	241
Transformaciones de variables	243
Curvas de crecimiento	246
Aplicación a la administración	250
Glosario	252
Fórmulas clave	253
Problemas	254
Caso 6-1: Tiger Transport	266
Caso 6-2: Butcher Products, Inc.	268
Caso 6-3: Ace Manufacturing	269
Caso 6-4: Mr. Tux	270
Caso 6-5: Consumer Credit Counseling	270
Caso 6-6: AAA Washington	271
Aplicaciones de Minitab	274
Aplicaciones de Excel	277
Referencias	279

CAPÍTULO 7	Análisis de regresión múltiple	281
	Diversas variables explicativas	281
	Matriz de correlación	282
	Modelo de regresión múltiple	283
	<i>Modelo estadístico para la regresión múltiple</i>	283
	Interpretación de los coeficientes de regresión	285
	Inferencias en modelos de regresión múltiple	286
	<i>Error estándar de la estimación</i>	287
	<i>Significancia de la regresión</i>	288
	<i>Variables explicativas individuales</i>	290
	<i>Pronóstico de una respuesta futura</i>	291
	Resultado de la computadora	292
	Variables ficticias	293
	Multicolinealidad	297
	Selección de la “mejor” ecuación de regresión	300
	<i>Todas las regresiones posibles</i>	302
	<i>Regresión por pasos</i>	304
	<i>Notas finales sobre la regresión por pasos</i>	306
	Diagnósticos de regresión y análisis residual	307
	Advertencias del pronóstico	309
	<i>Sobreajuste de la prueba</i>	309
	<i>Regresiones útiles, valores grandes calculados de F</i>	310
	Aplicación a la administración	310
	Glosario	312
	Fórmulas clave	312
	Problemas	313
	Caso 7-1: El mercado de bonos	324
	Caso 7-2: AAA Washington	328
	Caso 7-3: Béisbol de fantasía (A)	330
	Caso 7-4: Béisbol de fantasía (B)	334
	Aplicaciones de Minitab	336
	Aplicaciones de Excel	337
	Referencias	338
CAPÍTULO 8	Regresión con datos de series de tiempo	339
	Series de tiempo y el problema de autocorrelación	339
	Autocorrelación y la prueba de Durbin-Watson	343
	Soluciones a los problemas de autocorrelación	347
	<i>Error de especificación del modelo (omisión de una variable)</i>	348
	<i>Regresión con diferencias</i>	350
	<i>Errores autocorrelacionados y diferencias generalizadas</i>	354
	<i>Modelos autorregresivos</i>	357
	<i>Resumen</i>	358

Datos de series de tiempo y el problema de heteroscedasticidad	358
Uso de la regresión para pronosticar datos estacionales	361
Pronósticos econométricos	364
Series de tiempo cointegradas	365
Aplicación a la administración	367
Glosario	367
Fórmulas clave	367
Problemas	369
Caso 8-1: Compañía de su elección	378
Caso 8-2: Índice de actividad de negocios para el condado de Spokane	379
Caso 8-3: Ventas de un restaurante	383
Caso 8-4: Mr. Tux	385
Caso 8-5: Consumer Credit Counseling	388
Caso 8-6: AAA Washington	389
Caso 8-7: Alomega Food Stores	392
Caso 8-8: Surtido Cookies	393
Caso 8-9: Southwest Medical Center	394
Aplicaciones de Minitab	395
Aplicaciones de Excel	396
Referencias	398

CAPÍTULO 9 La metodología Box-Jenkins (ARIMA) 399

Metodología Box-Jenkins	399
<i>Modelos autorregresivos</i>	404
<i>Modelos de promedios móviles</i>	405
<i>Modelos de promedio móvil autorregresivos</i>	407
<i>Resumen</i>	407
Estrategia de implementación de la construcción del modelo	407
<i>Paso 1: Identificación del modelo</i>	407
<i>Paso 2: Estimación del modelo</i>	409
<i>Paso 3: Verificación del modelo</i>	410
<i>Paso 4: Elaboración de pronósticos con el modelo</i>	411
<i>Advertencias sobre la construcción de modelos</i>	430
<i>Criterios de selección del modelo</i>	431
<i>Modelos ARIMA para datos estacionales</i>	432
<i>Suavización exponencial simple y el modelo ARIMA</i>	442
<i>Ventajas y desventajas de los modelos ARIMA</i>	443
Aplicación a la administración	444
Glosario	445
Fórmulas clave	445
Problemas	446

Caso 9-1: Ventas en un restaurante	457
Caso 9-2: Mr. Tux	459
Caso 9-3: Consumer Credit Counseling	460
Caso 9-4: The Lydia E. Pinkham Medicine Company	461
Caso 9-5: Ciudad College Station	463
Caso 9-6: UPS Air Finance Division	466
Caso 9-7: AAA Washington	469
Caso 9-8: Web Retailer	471
Caso 9-9: Surtido Cookies	474
Caso 9-10: Southwest Medical Center	476
Aplicaciones de Minitab	478
Referencias	480

CAPÍTULO 10 Pronósticos de juicio y ajuste de pronóstico 481

Pronósticos de juicio	483
<i>El método Delphi</i>	483
<i>Formulación de escenarios</i>	485
<i>Combinación de pronósticos</i>	486
<i>Los pronósticos y las redes neurales</i>	488
<i>Resumen de pronósticos de juicio</i>	490
Otras herramientas útiles para hacer juicios acerca del futuro	491
Fórmulas clave	496
Problemas	496
Caso 10-1: Restaurante Golden Gardens	497
Caso 10-2: Alomega Food Stores	497
Caso 10-3: The Lydia E. Pinkham Medicine Company	498
Referencias	501

CAPÍTULO 11 Administración del proceso de pronósticos 503

El proceso del pronóstico	503
Monitoreo de pronósticos	504
Revisión de los pasos del pronóstico	509
Responsabilidad al elaborar el pronóstico	510
Costos de los pronósticos	511
Sistemas de información para administrar y pronosticar	511
Importancia de la gerencia en la administración de los pronósticos	512
El futuro de los pronósticos	512
Problemas	513
Caso 11-1: Boundary Electronics	513
Caso 11-2: Busby Associates	514
Caso 11-3: Consumer Credit Counseling	517

Caso 11-4: Mr. Tux	518
Caso 11-5: Alomega Food Stores	519
Caso 11-6: Southwest Medical Center	520
Referencias	520

APÉNDICE A Datos del caso 7-1 521

APÉNDICE B Tablas 523

<i>Tabla B-1</i>	<i>Términos individuales de la distribución binomial</i>	523
<i>Tabla B-2</i>	<i>Áreas de distribución de probabilidad normal estándar</i>	525
<i>Tabla B-3</i>	<i>Valores críticos de t</i>	526
<i>Tabla B-4</i>	<i>Valores críticos de chi cuadrada</i>	527
<i>Tabla B-5</i>	<i>Distribución F</i>	529
<i>Tabla B-6</i>	<i>Límites de la prueba de Durbin-Watson</i>	530

APÉNDICE C Conjunto de datos y bases de datos 533

Índice 547

Prefacio

El propósito de la novena edición de *Pronósticos en los negocios* es el mismo de las ediciones anteriores: presentar las técnicas estadísticas básicas que son útiles en la elaboración de pronósticos de negocios y planes a largo plazo. El libro está escrito en un estilo sencillo y directo e incluye un gran número de ejemplos prácticos de negocios. Sesenta y cinco casos se encuentran distribuidos al final de los capítulos para mostrar al estudiante cómo se vinculan los conceptos teóricos con las aplicaciones en el mundo real. Se hace énfasis en la aplicación de técnicas para la toma de decisiones de la dirección de las empresas. Se supone que los estudiantes tienen conocimientos básicos de estadística y que están familiarizados con el uso de la computadora para tener acceso a paquetes de software de procesamiento de palabras y hojas de cálculo.

ORGANIZACIÓN

Se han revisado todos los capítulos para mejorar la claridad de la escritura e incrementar la efectividad de la enseñanza y el aprendizaje. El contenido se ha organizado en seis secciones.

La primera sección (capítulos 1 y 2) contiene los antecedentes. Una descripción de la naturaleza de los pronósticos y un rápido repaso de conceptos estadísticos básicos preparan al alumno para conocer las técnicas de pronósticos que inician en la segunda sección.

La segunda sección (capítulo 3) hace énfasis en la exploración de patrones de datos y en la selección de una técnica de pronóstico adecuada. La tercera sección (capítulos 4 y 5) cubre las técnicas de promedios y suavización, así como una introducción a la descomposición de las series de tiempo en términos de los componentes subyacentes. La cuarta sección (capítulos 6 y 7) hace énfasis en las técnicas causales de pronóstico, tales como el análisis de la correlación, la regresión y la regresión múltiple.

La quinta sección (capítulos 8 y 9) considera las técnicas usadas para pronosticar con los datos de las series de tiempo. El libro concluye con una sección final (capítulos 10 y 11) sobre pronósticos de juicio y ajustes de pronósticos, junto con una discusión sobre la administración y el seguimiento del proceso de elaboración de pronósticos.

CAMBIOS EN LA NOVENA EDICIÓN

La novena edición se ha modificado en gran parte. Si bien se ha conservado el tono de ediciones anteriores, se ha puesto un énfasis adicional en los descubrimientos teóricos y empíricos más recientes. Se actualizó el material y se reorganizó el libro por completo agregando nuevos problemas, ejemplos, conjuntos de datos y casos de estudio.

Las siguientes características también son nuevas o se mejoraron en esta edición:

- Se agregaron 12 nuevos casos de estudio.
- Se agregaron 32 nuevos problemas.
- Se actualizaron los conjuntos de datos.

- En el capítulo 8 se agregó la sección de Series de tiempo cointegradas.
- Al final de la mayoría de los capítulos se presenta un ejemplo de cómo usar Minitab 15 en la elaboración de pronósticos.
- Al final de la mayoría de los capítulos se muestra un ejemplo de cómo usar Excel 2003 para resolver un problema de elaboración de pronósticos.
- En el sitio www.pearsoneducacion.net/hanke se encuentra disponible una versión del Manual para el profesor.
- En Internet se encuentran disponibles conjuntos de datos para Minitab, Excel y otros programas.

EL PAPEL DE LA COMPUTADORA

En las primeras ocho ediciones, se reconoció a la computadora como una poderosa herramienta para pronosticar. En la actualidad, la computadora es aún más importante, ya que se dispone de poderoso software para pronosticar y se ha facilitado considerablemente el acceso a los datos gracias a las capacidades de las redes e Internet.

Un estudio de investigación nacional de todas las instituciones afiliadas al AACSB dirigida por los autores para determinar qué podía hacer el personal docente para mejorar la enseñanza de los pronósticos utilizando la computadora indicó lo siguiente: **1.** La mayor parte del personal docente del área de pronósticos (94.2%) intenta brindar a los estudiantes experiencia en el uso de la computadora, y **2.** en la investigación se mencionaron varios paquetes estadísticos y paquetes específicos de computación para la elaboración de pronósticos. Los paquetes mencionados con más frecuencia fueron Minitab, SAS y hojas de cálculo.

Los autores han intentado varios enfoques diferentes para ayudar al profesorado y a los estudiantes a utilizar la computadora en la elaboración de pronósticos. En ese sentido, la presente edición incluye lo siguiente:

1. Instrucciones de Minitab presentadas al final de la mayoría de los capítulos.
2. Instrucciones de Excel presentadas al final de la mayoría de los capítulos.
3. Tres colecciones de datos disponibles en Internet (Minitab, Excel, otros programas). Cada colección contiene datos de los ejemplos y problemas del texto. Cada colección contiene también varias series de datos nuevos. Para tener acceso a los datos en Internet, entre al sitio Web www.pearsoneducacion.net/hanke.
4. A lo largo del texto están distribuidos varios ejemplos de diferentes resultados de la computadora.

RECONOCIMIENTOS

Los autores están en deuda con muchos profesores alrededor del mundo que han usado las primeras ocho ediciones de este libro, quienes hicieron valiosas sugerencias para mejorarlo. Un agradecimiento especial para Frank Forest (Marquette University), William Darrow (Townsend State University), Susan Winters (Northwestern State University), Shik Chun Young (Eastern Washington University), Mark Craze, Judy Johnson, Steve Brandon y Dorothy Mercer por facilitar los casos.

Partes de este texto, especialmente algunos conjuntos de datos, son adaptaciones de los que aparecen en la segunda edición de *Understanding Business Statistics*, de Hanke y Reitsch, publicado por Richard D. Irwin, Inc., a quien aquí le damos el crédito correspondiente.

También agradecemos a los revisores John Liechty, de la Universidad de Michigan; John Tamura, de la Universidad de Washington; y Ted Taukahaara, de la Universidad de St. Mary por sus comentarios constructivos sobre ediciones anteriores de este libro. Otros revisores merece-

dores de nuestro reconocimiento son Perry Sadorski, de la Universidad de York; Shady Kholdy, de la Universidad Politécnica de Pomona en el Estado de California; Michael Niemira, de la Universidad de Nueva York; Fred Zufryden, de la Universidad del Sur de California; y Haizheg Li, del Instituto Tecnológico de Georgia.

Finalmente, tenemos una gran deuda con las computadoras de escritorio y el software de procesamiento de palabras. La tecnología moderna ha hecho considerablemente más fácil la escritura de libros de texto. Sin embargo, nosotros, no las computadoras, somos responsables de cualquier error.

INTRODUCCIÓN A LOS PRONÓSTICOS EN LOS NEGOCIOS

Este libro examina los métodos usados para predecir la naturaleza incierta de las tendencias en los negocios, en un esfuerzo por ayudar a los administradores y gerentes a realizar mejores planes y a tomar mejores decisiones. Tales esfuerzos con frecuencia requieren el estudio de datos históricos y la manipulación de esos datos, en la búsqueda de patrones que se extrapolen efectivamente para realizar pronósticos.

En este libro, continuamente recordamos a los lectores que se debe hacer un buen juicio de los resultados numéricos, si queremos obtener pronósticos confiables. Los ejemplos en este capítulo, así como los casos de estudio al final de éste y del resto de los capítulos del texto, resaltan este punto. Hay más análisis acerca del papel del juicio en este capítulo y en el capítulo final.

LA HISTORIA DE LOS PRONÓSTICOS EN LOS NEGOCIOS

En un libro sobre la historia del riesgo, su autor, Peter Bernstein (1996), señaló que el desarrollo de los pronósticos en los negocios durante el siglo xvii fue una gran innovación. Él escribió:

—el pronóstico —largamente desacreditado como una pérdida de tiempo en el mejor de los casos, y como un pecado en el peor— se convirtió en una necesidad absoluta en el curso del siglo xvii, para algunos empresarios audaces que estaban dispuestos tomar el riesgo de dar forma al futuro, de acuerdo con su propio diseño. (p. 95)

Durante los siguientes 300 años, hubo avances significativos en los métodos para pronosticar basados en datos, muchos de los cuales ocurrieron en el siglo xx. Los métodos de análisis de regresión, descomposición, suavizamiento y promedios móviles autorregresivos son ejemplos de procedimientos para realizar pronósticos, basados en datos que se analizan en este libro. Tales procedimientos han demostrado ser altamente eficientes y aparecen rutinariamente en los menús del software comercial de pronósticos en los negocios.

Con el desarrollo de los métodos basados en datos, los enfoques del rol del buen juicio y la crítica para realizar pronósticos han crecido significativamente durante los últimos 25 años. Sin datos históricos, el juicio humano sería el único modo de hacer predicciones acerca del futuro. En los casos donde los datos están disponibles, el buen juicio debería utilizarse para revisar, y quizá modificar, el pronóstico elaborado mediante procedimientos cuantitativos. Si bien la mayor parte de este libro está dedicada a los métodos de elaboración de pronósticos basados en datos, se incluye un análisis sobre métodos críticos de pronósticos en el capítulo 10.

Con la difusión de poderosas computadoras personales y la disponibilidad de paquetes de software avanzado, se generan fácilmente los pronósticos de los valores futuros para variables de interés. Sin embargo, esta facilidad en el cálculo no sustituye el razonamiento claro. La falta de visión gerencial y el uso inadecuado de las técnicas para pronosticar quizá originen decisiones costosas.

Nuevos procedimientos de elaboración de pronósticos continúan desarrollándose conforme se acelera la necesidad de pronósticos más exactos.¹ Se pone especial atención al proceso de elaboración de pronósticos en organizaciones que tienen necesidad de coordinar objetivos, métodos, evaluación e interpretación.

¿ES NECESARIO EL PRONÓSTICO EN LOS NEGOCIOS?

A pesar de las imprecisiones inherentes al intentar predecir el futuro, los pronósticos necesariamente guían el establecimiento de políticas y la planeación. ¿Cómo puede la Reserva Federal ajustar en forma realista las tasas de interés, sin alguna noción del crecimiento económico futuro ni de las presiones inflacionarias? ¿Cómo un gerente de operaciones establecería realísticamente programas de producción sin algún estimado de las ventas futuras? ¿Cómo podría una compañía determinar la plantilla laboral de sus *call centers*, sin alguna suposición de la demanda futura de sus servicios? ¿Cómo un banco haría planes realistas sin algún pronóstico de futuros depósitos y saldos en préstamos? Todos requieren los pronósticos. La necesidad de pronósticos está en todas las líneas funcionales, así como en todos tipos de organizaciones. Los pronósticos son absolutamente necesarios para avanzar en el ambiente de negocios actual, siempre cambiante y altamente interactivo.

Este libro analiza varias formas de realizar pronósticos, que dependen de métodos lógicos de manipulación de datos que se han generado en eventos históricos. Pero creemos que el pronosticador eficaz debe ser capaz de establecer una hábil mezcla de pronóstico cuantitativo y buen juicio, evitando los extremos de confiar totalmente en uno o en otro. En un extremo, encontramos al ejecutivo que, por ignorancia y miedo a las computadoras y a las técnicas cuantitativas, confía solamente en su intuición y sensibilidad. En el otro extremo, está el pronosticador diestro en las más avanzadas técnicas actuales de manipulación de datos, pero que parece incapaz o está indispuerto para relacionar el proceso de elaboración del pronóstico, con las necesidades de la organización y de sus tomadores de decisiones. Nosotros vemos las técnicas cuantitativas para pronosticar, estudiadas en la mayor parte de este libro, como el único punto de arranque en el pronóstico efectivo de resultados importantes para la organización: análisis, juicio, sentido común y experiencia en el negocio deben aplicarse para brindar apoyo al proceso a través del cual estas importantes técnicas han generado sus resultados.

Otro pasaje de Bernstein (1996) resume efectivamente el rol del pronóstico en las organizaciones.

Usted no planearía enviar bienes a través del océano, ensamblar mercancía para vender ni solicitar dinero en préstamo, sin primero tratar de determinar qué depara el futuro a su empresa. Asegurarse de que los materiales que usted solicitó se entreguen a tiempo, saber que los artículos que usted planea vender se produzcan conforme a lo programado, y ver sus instalaciones reales de ventas deben planearse antes del momento en que los clientes aparecen y ponen su dinero en el mostrador. El ejecutivo de negocios exitoso es, en primer lugar, un pronosticador: la compra, la producción, el marketing, la fijación de precios y la organización vienen todos después. (pp. 21-22)

TIPOS DE PRONÓSTICOS

Cuando los gerentes se enfrentan con la necesidad de tomar decisiones en una atmósfera de incertidumbre, ¿qué tipos de pronósticos tienen disponibles? Los procedimientos para realizar pronósticos se clasifican, en primer lugar, como de largo plazo y de corto plazo. Los pronósticos a largo plazo necesariamente señalan el curso general de una organización para un tiempo de

¹Un repaso reciente acerca del estado actual de la elaboración de pronósticos está disponible en un número especial del *Internacional Journal of Forecasting*, editado por R. J. Hyndman y J. K. Ord (2006).

funcionamiento largo; de modo que se convierten en un tema de particular relevancia para la alta dirección. Los pronósticos a corto plazo se requieren para diseñar estrategias inmediatas, y las gerencias media y de primera línea los emplean para satisfacer las necesidades del futuro inmediato.

Los pronósticos también se clasifican en términos de su posición en un continuo micro-macro, es decir, en términos del grado en que implican pequeños detalles o grandes valores de resumen. Por ejemplo, un gerente de planta estaría interesado en pronosticar el número de trabajadores necesarios para los meses siguientes (un micro pronóstico); mientras que el gobierno federal pronostica el número total de gente con empleo en todo el país (un macro pronóstico). Otra vez, diferentes niveles de administración en una organización tienden a enfocarse en diferentes niveles del continuo micro-macro. La alta gerencia estaría interesada en el pronóstico de ventas de la toda compañía, por ejemplo; en tanto que los vendedores individuales estarían mucho más interesados en pronosticar sus propios volúmenes de ventas.

Los procedimientos para la elaboración de pronósticos también se clasifican de acuerdo con la tendencia a ser más cualitativos o más cuantitativos. En un extremo, una técnica puramente cualitativa es aquella que evidentemente no requiere manipulación de datos alguna, tan sólo se usa el “juicio” del pronosticador. Aun aquí, por supuesto, el “juicio” del pronosticador quizá sea realmente el resultado de la manipulación mental de ciertos datos históricos. En el otro extremo, las técnicas puramente cuantitativas no necesitan el juicio como insumo, pues se trata de procedimientos mecánicos para producir resultados cuantitativos. Por supuesto, algunos procedimientos cuantitativos requieren una manipulación de datos mucho más compleja que otros. Este libro se enfoca en las técnicas de pronóstico cuantitativas, porque la comprensión profunda de estos procedimientos tan útiles es necesaria en la administración eficiente de las organizaciones modernas. Sin embargo, destacamos una vez más que el buen juicio y el sentido común deberían utilizarse junto con los procedimientos mecánicos y de manipulación de datos. Sólo de este modo se obtendrá un pronóstico inteligente.

Finalmente, los pronósticos podrían clasificarse de acuerdo con la naturaleza del producto obtenido. Uno debe decidir si el pronóstico será un número individual (*pronóstico puntual*), un intervalo de números dentro del cual se espera que esté el valor futuro (un pronóstico por intervalo) o la distribución de probabilidad total del valor futuro (un *pronóstico de densidad*). Puesto que “shocks” imprevisibles pueden afectar los valores futuros (el futuro nunca es exactamente como el pasado), ocurrirán errores de pronóstico diferentes de cero, incluso en aquellos pronósticos muy buenos. De modo que hay cierta incertidumbre asociada con un pronóstico puntual específico. La incertidumbre que rodea los pronósticos puntuales sugiere la inutilidad de los pronósticos por intervalo. Sin embargo, si los pronósticos tan sólo son el resultado del juicio, los pronósticos puntuales normalmente son el único recurso. En situaciones de juicio, resulta extremadamente difícil describir con precisión la incertidumbre asociada con el pronóstico.

CONSIDERACIONES MACROECONÓMICAS DEL PRONÓSTICO

Usualmente concebimos los pronósticos en términos de la predicción de variables importantes para una compañía individual o quizá para un componente de la empresa. Ventas mensuales de la compañía, ventas unitarias de una de las tiendas de ésta, y horas de ausentismo por empleado por mes en una fábrica son buenos ejemplos.

En contraste, hay un interés creciente en la predicción de variables importantes para toda la economía de un país. Se ha desarrollado mucho trabajo en la evaluación de métodos para hacer esta clase de pronósticos económicos generales, llamados *pronósticos macroeconómicos*. Ejemplos de interés para el gobierno federal de Estados Unidos, y de muchos otros países, son la tasa de desempleo, el producto interno bruto y la tasa de interés de referencia. La política económica se basa, en parte, en las proyecciones de importantes indicadores económicos como éstos. Por tal motivo, existe un gran interés en mejorar los métodos para realizar pronósticos, que se enfoquen en mediciones generales del funcionamiento económico de un país.

Una de las principales dificultades en la elaboración de pronósticos precisos de la actividad económica general es el cambio relevante e inesperado de un factor económico clave. Cambios importantes en los precios del petróleo, inflación súbita, y cambios políticos significativos en el gobierno de un país son ejemplos de cambios de un factor clave que puede afectar la economía global.

La posibilidad de estos cambios relevantes en la escena económica motivó una pregunta fundamental en la elaboración de pronósticos macroeconómicos: ¿Los pronósticos generados por el modelo de elaboración de pronósticos deberían modificarse usando el buen juicio del pronosticador? Los trabajos actuales sobre la metodología en la realización de pronósticos a menudo implican dicha pregunta.

Se siguen efectuando trabajos teóricos y prácticos sobre la elaboración de pronósticos macroeconómicos. Considerando la importancia de la realización de pronósticos económicos precisos para la formulación de la política económica, para este y otros países, se esperaría una mayor atención futura para esta clase de pronósticos. Una buena referencia de introducción a los pronósticos macroeconómicos es Pindyck y Rubinfeld (1998).

SELECCIÓN DE UN MÉTODO DE PRONÓSTICOS

El análisis precedente sugiere que se consideren varios factores en la selección de un método para pronosticar. Se debe considerar el nivel de detalle. ¿Son necesarios los pronósticos de detalles específicos (un micro pronóstico)? ¿O es necesario el estatus futuro de algo general, o factor de resumen (un macro pronóstico)? ¿Es necesario el pronóstico para algún momento en el futuro cercano (un pronóstico a corto plazo) o de un momento en el futuro lejano (un pronóstico a largo plazo)? ¿Hasta dónde son adecuados los métodos cualitativos (juicio) y cuantitativos (manipulación de datos)? Y finalmente, ¿qué forma debería tomar el pronóstico (pronóstico puntual, por intervalo o de densidad)?

La consideración primordial en la selección de un método para pronosticar es que los resultados deben facilitar el proceso de toma de decisiones a los gerentes de la organización. Rara vez un método funciona para todos los casos. Diferentes productos (por ejemplo, nuevo contra existente), metas (por ejemplo, predicción sencilla contra la necesidad de controlar una importante directriz del negocio con valores futuros) y restricciones (por ejemplo, costo, habilidad requerida y lo prioritario) deben considerarse en la selección de un método de pronósticos. Con la disponibilidad del software actual para pronosticar, es mejor pensar en los métodos de pronósticos como herramientas genéricas que pueden aplicarse simultáneamente. En una situación dada es posible intentar varios métodos. La metodología que produce los pronósticos más exactos en un caso quizá no sea la mejor metodología en otra situación. Sin embargo, el(los) método(s) seleccionado(s) debería(n) producir un pronóstico que sea preciso, oportuno y comprensible para la gerencia, de manera que el pronóstico ayude a tomar mejores decisiones.

La selección de una técnica para pronosticar se analiza en el capítulo 3 y se resume en la tabla 3-6. Las metodologías para elaborar pronósticos que estudiamos en este libro se resumen en la tabla 11-1. Dichas tablas, junto con el análisis adicional disponible en Chase (1997), ayudan al pronosticador a seleccionar un conjunto inicial de procedimientos para pronosticar.

ETAPAS DEL PRONÓSTICO

Todos los procedimientos formales para pronosticar requieren extender las experiencias del pasado hacia el futuro. Así, implican la suposición de que las condiciones que generaron los datos y las relaciones pasados son iguales a las condiciones futuras.

Un departamento de recursos humanos contrata empleados, en parte, sobre la base de la puntuación en el examen de ingreso porque, en el pasado, esa puntuación parecía ser un predictor importante del nivel de desempeño en el trabajo. Siempre que se mantenga esta relación, el pronóstico del futuro desempeño en el trabajo —y, por lo tanto, las decisiones de contratación— puede mejorar usando las calificaciones del examen. Si, por alguna razón, la asociación

entre la puntuación del examen y el desempeño en el trabajo cambia, entonces el pronóstico del nivel de desempeño en el trabajo a partir de las puntuaciones del examen que usan el modelo histórico originará pronósticos imprecisos y, potencialmente, decisiones de contratación deficientes. Esto es lo que hace difícil pronosticar. El futuro no siempre es como el pasado. Cuando sí lo es, los métodos cuantitativos de elaboración de pronósticos funcionan bien. Cuando no lo es, llegan a producirse pronósticos imprecisos. Sin embargo, generalmente es mejor tener algún pronóstico construido razonablemente, que no pronosticar.

El reconocimiento de que las técnicas de pronósticos funcionan con datos generados por eventos históricos conduce a la identificación de los siguientes cinco pasos en el proceso de pronosticar:

1. Formulación del problema y recopilación de datos
2. Manipulación y limpieza de datos
3. Construcción y evaluación del modelo
4. Implementación 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 solo paso porque están íntimamente relacionadas. El problema determina los datos apropiados. Si se está considerando una metodología cuantitativa para pronosticar, los datos pertinentes deben estar disponibles y ser correctos. Con frecuencia, el acceso y la recopilación de los datos adecuados constituyen una tarea desafiante y consumidora de tiempo. Si los datos apropiados no están disponibles, quizá tenga que volver a definirse el problema o tal vez se tenga que emplear una metodología no cuantitativa para pronosticar. Los problemas de recopilación y control de calidad de los datos surgen siempre que sea necesario obtener datos pertinentes para elaborar un pronóstico de negocios.

El paso 2, manipulación y limpieza de datos, a menudo es necesario. Es posible tener demasiados datos o muy pocos, en el proceso para realizar pronósticos. Algunos datos quizá no sean pertinentes. Tal vez a algunos datos les falten valores que deban estimarse. En ocasiones ciertos datos tienen que expresarse en unidades diferentes de las originales. Algunos datos deben volverse a procesar (por ejemplo, recolectarse de varias fuentes y sumarse). Otros datos pueden ser adecuados, pero sólo para ciertos periodos históricos (por ejemplo, en la elaboración del pronóstico de ventas de automóviles compactos, uno quizá disponga únicamente de los datos de venta de autos desde el embargo petrolero en la década de 1970, en vez de datos de venta durante los últimos 60 años). Normalmente se requiere algún esfuerzo para obtener datos en la forma requerida, para usar ciertos procedimientos para pronosticar.

El paso 3, construcción y evaluación del modelo, incluye ajustar los datos recolectados a un modelo de pronóstico que sea adecuado, en términos de minimizar errores en el pronóstico. Cuanto más sencillo sea el modelo, mejor será en términos de la aceptación del proceso de pronósticos por parte de los gerentes que deben tomar las decisiones de la empresa. Con frecuencia, se debe alcanzar un equilibrio entre un enfoque para pronosticar complejo que ofrezca un poco más de precisión, y un enfoque sencillo que se entienda fácilmente y gane el apoyo de quienes toman las decisiones de la compañía —y sea activamente usado por éstos. Evidentemente, el buen juicio interviene en este proceso de selección. Puesto que este libro analiza diversos modelos de pronósticos y su aplicación, la capacidad del lector para ejercitar el buen juicio, en la selección y el uso de modelos para pronosticar adecuados, se incrementará después de estudiar este material.

El paso 4, implementación del modelo, es la generación del modelo real una vez que se hayan recopilado y depurado los datos apropiados, y se haya seleccionado el modelo de pronóstico adecuado. Los datos de periodos históricos más recientes se mantienen como respaldo y más tarde se usan para verificar la exactitud del proceso.

El paso 5, evaluación del pronóstico, implica la comparación de los valores del pronóstico con valores históricos reales. Después de que se termina la implementación del modelo, se rea-

lizan los pronósticos para los periodos históricos más recientes, donde se conocen los valores de los datos, pero se mantienen como respaldo de los datos que se analizan. Estos pronósticos se comparan después con los valores históricos conocidos y se analizan cualesquiera errores en el pronóstico. Algunos procedimientos para pronosticar suman los valores absolutos de los errores y reportan esta suma; o dividen esta suma entre el número de intentos para producir el pronóstico, y conocer el error promedio del pronóstico. Otros procedimientos brindan la suma de los errores al cuadrado, lo cual se compara luego con cifras similares de métodos alternativos para pronosticar. Algunos procedimientos también rastrean y reportan la magnitud de los términos del error, durante el periodo del pronóstico. El examen de los patrones de error a menudo lleva al analista a modificar el modelo para pronosticar. Métodos específicos de medición de los errores en los pronósticos se examinan cerca del final del capítulo 3.

ADMINISTRACIÓN DEL PROCESO DE ELABORACIÓN DEL PRONÓSTICO

En este capítulo, el análisis sirve para resaltar nuestra creencia de que la capacidad gerencial y el sentido común deben intervenir en el proceso de elaboración del pronóstico. El pronosticador debería concebirse como un asesor del gerente, en vez de un monitor de un dispositivo automático para la toma de decisiones. Por desgracia, en la práctica se da el último caso, especialmente con la imagen que tiene la computadora. Otra vez, las técnicas cuantitativas de elaboración del pronóstico deben verse como realmente son; a saber, herramientas que el gerente va a utilizar para tomar mejores decisiones. De acuerdo con Makridakis (1986):

La utilidad del pronóstico mejoraría si la gerencia adoptara una actitud más realista. El pronóstico no debería verse como un sustituto del vaticinio, sino como el mejor camino para identificar y extrapolar patrones o relaciones establecidos para la elaboración de pronósticos. Si se acepta dicha actitud, los errores del pronóstico deben considerarse inevitables y se tienen que investigar las circunstancias que los causan. (p. 33)

Entonces, siempre deberían surgir algunas preguntas fundamentales, si el proceso de elaboración del pronóstico se administra adecuadamente:

- ¿Por qué es necesario el pronóstico?
- ¿Quién usará el pronóstico y cuáles son los requerimientos específicos?
- ¿Qué nivel de detalle o generalización se requiere, y cuál es el horizonte de tiempo adecuado?
- ¿Qué datos están disponibles y serán suficientes los datos para generar el pronóstico que se necesita?
- ¿Cuánto costará el pronóstico?
- ¿Qué tan exacto se espera que sea el pronóstico?
- ¿el pronóstico estará a tiempo para ayudar en el proceso de toma de decisiones?
- ¿El pronosticador entiende claramente cómo se usará el pronóstico en la organización?
- ¿Está disponible un proceso de retroalimentación para evaluar el pronóstico y ajustar el proceso como corresponde?

SOFTWARE DE PRONÓSTICOS

Actualmente, hay un gran número de paquetes de software diseñados específicamente para ofrecer al usuario diversos métodos para elaborar pronósticos. Dos tipos de paquetes computacionales son, en primera instancia, de interés para los pronosticadores: **1.** paquetes estadísticos generales que incluyen análisis de regresión, análisis de series de tiempo y otras técnicas usadas frecuentemente por quienes hacen pronósticos; y **2.** paquetes para elaboración de pronósticos diseñados específicamente para aplicaciones en pronósticos. Adicionalmente, algu-

nas herramientas de pronósticos están disponibles en los sistemas de planeación de recursos de la empresa (*enterprise resource planning*) (PRE).

Capacidades gráficas, interfaces con hoja de cálculo y fuentes de datos externas, métodos numéricos y estadísticos confiables, así como algoritmos automáticos sencillos para la selección y configuración de modelos para la elaboración de pronósticos, son ahora características comunes del software para pronóstico en los negocios. Sin embargo, aun cuando el desarrollo y conocimiento del software para pronósticos han crecido considerablemente durante los últimos años, la mayoría de las compañías usan simplemente hojas de cálculo (quizá con la función de complementos, *add-ins*) para generar pronósticos y desarrollar planes de negocios.

Ejemplos de paquetes de software específicos con *herramientas* para elaboración del pronóstico son Minitab, SAS y SPSS. Hay muchos programas complementarios que brindan herramientas para pronosticar en un ambiente de hoja de cálculo. Por ejemplo, *add-in* de Analysis ToolPak (Herramientas, análisis de datos) de Microsoft Excel ofrece algunas capacidades de análisis de regresión y de suavizamiento. Existen otros complementos más integrales que proporcionan un rango (casi) completo de capacidades para la elaboración de pronósticos.²

Algunas veces, particularmente en el ambiente de hojas de cálculo, están disponibles pronósticos “automáticos”. Es decir, el software selecciona el modelo o procedimiento que sea mejor para elaborar el pronóstico e, inmediatamente, genera los pronósticos. No obstante, debemos estar concientes de que esta ventaja tiene un precio. Los procedimientos automáticos producen números, pero raras veces brindan al pronosticador la comprensión real de la naturaleza y la calidad de los pronósticos. La generación de pronósticos significativos requiere de la intervención humana: un continuo intercambio entre el conocimiento del problema y los procedimientos (software) para la elaboración de pronósticos.

Muchas de las técnicas de este libro se ilustrarán con Minitab 15 y Microsoft Excel 2003 (con *add-in* de Análisis ToolPak). Se seleccionó Minitab 15 por su uso fácil y su amplia disponibilidad. Si bien Excel es limitado en su funcionalidad para la elaboración de pronósticos, a menudo es la herramienta seleccionada para proyecciones de cálculos.

INFORMACIÓN EN LÍNEA

Hay información disponible de interés para los pronosticadores en la Web. Quizás el mejor modo de aprender lo que está disponible en el ciberespacio sea pasar algún tiempo buscando todo lo que le interese a usted, usando un navegador como Netscape o Microsoft Internet Explorer.

Cualquier lista de sitios Web para los pronosticadores tal vez esté obsoleta para el tiempo en que aparezca esta edición; sin embargo, hay dos sitios Web que probablemente estén vigentes por algún tiempo. B&E DataLinks, disponible en www.econ-datalinks.org, es un sitio Web administrado por la Business and Economic Statistics Section of the American Statistical Association. Este sitio Web contiene muchos vínculos con fuentes de datos económicos y financieros que son de interés para los pronosticadores. El segundo sitio, Resources for Economists on the Internet, patrocinado por la American Economic Association, y disponible en rfe.org, contiene un extenso conjunto de vínculos con fuentes de datos, publicaciones, organizaciones profesionales, etcétera.

EJEMPLOS DE ELABORACIÓN DE PRONÓSTICOS

Los análisis en este capítulo enfatizan que la elaboración de pronósticos requiere un alto grado de juicio, junto con la manipulación matemática de los datos recopilados. Los siguientes ejemplos muestran la clase de pensamiento que precede a un esfuerzo de elaboración del pronóstico en una empresa real. Note que los valores de datos que producirán pronósticos útiles, incluso si

²Al momento de escribir esta edición, el Institute for Forecasting Education efectuó revisiones del software para pronósticos en su sitio Web. Se pueden consultar tales revisiones en www.forecastingeducation.com/forecastingsoftwareviews.asp.

existen, quizá no sean evidentes al inicio del proceso, y tal vez no se identifiquen conforme el proceso evoluciona. En otras palabras, los esfuerzos iniciales podrían terminar siendo inútiles y entonces se requeriría otro enfoque.

No se muestran los resultados de los esfuerzos en la elaboración de pronósticos para los dos ejemplos que se analizan aquí, ya que requieren temas que se estudian a lo largo del texto. Busque las técnicas descritas en capítulos posteriores y que se aplican a esos datos. Por el momento, esperamos que estos ejemplos ilustren el trabajo en la elaboración de pronósticos que los gerentes enfrentan en la realidad.

Ejemplo 1.1 Alomega Food Stores

Alomega Food Stores es un proveedor minorista de alimentos, con 27 tiendas en un estado del oeste medio. La compañía ha contratado varias clases de publicidad y, hasta hace poco, nunca había estudiado el efecto que su dinero en publicidad tienen sobre las ventas, aun cuando algunos datos se han recopilado y almacenado durante tres años.

Los ejecutivos de Alomega decidieron empezar a examinar sus campañas publicitarias, en relación con los volúmenes de ventas mensuales. Esperaban que después de algunos meses de recolectar datos, tuvieran la posibilidad de descubrir las relaciones que ayudarían en la determinación de los gastos futuros en publicidad.

El departamento de contabilidad empezó la generación de sus registros históricos, anotando los volúmenes de ventas mensuales junto con el dinero gastado en publicidad, tanto en anuncios de periódicos, como en spots televisivos. También registraron los valores de las ventas y la publicidad que se habían retrasado por uno y dos meses. Esto se hizo porque algunas personas del comité ejecutivo pensaron que quizá las ventas dependían de los gastos en publicidad en los meses previos, en vez del mes en que ocurrieron las ventas.

Asimismo, los ejecutivos creían que las ventas tenían un efecto estacional. Por tal razón, se usó una variable ficticia (*dummy*) o categórica para identificar cada mes. Adicionalmente, ellos querían conocer cualquier tendencia en el volumen de ventas.

Por último, los ejecutivos creían que el dinero que Alomega gastaba en publicidad podría tener un efecto en el presupuesto de publicidad de sus principales competidores, durante el mes siguiente. Para cada mes sucesivo, se decidió que la publicidad de los competidores se clasificaría como **1**, una cantidad pequeña, **2**, una cantidad moderada o **3**, una cantidad grande.

Después de pocos meses de recolección de datos y análisis de los registros pasados, el departamento de contabilidad terminó una muestra de 48 datos mensuales, usando las siguientes variables:

- Dinero en ventas
- Dinero en publicidad en periódicos
- Dinero en publicidad televisiva
- Código del mes, donde enero = 1, febrero = 2, hasta diciembre = 12
- Serie de 11 variables ficticias para indicar el mes
- Publicidad en el periódico atrasada un mes
- Publicidad en el periódico atrasada dos meses
- Publicidad televisiva atrasada un mes
- Publicidad televisiva atrasada dos meses
- Número de mes de 1 a 48
- Código 1, 2 o 3 para indicar las campañas publicitarias de los competidores el siguiente mes

Los gerentes de Alomega, especialmente Julie Ruth, la presidenta de la compañía, ahora quiere saber todo lo que pueda a partir de los datos que se han recolectado. Además del conocimiento acerca de los efectos de la publicidad sobre los volúmenes de ventas y la publicidad de los competidores, Julie se pregunta acerca de cualquier tendencia y efecto estacional en las ventas. Sin embargo, el gerente de producción de la compañía, Jackson Tilson, no comparte su entusiasmo. Al final de la junta de planeación del pronóstico, él hizo la siguiente declaración: “He estado tratando de mantener mi boca cerrada durante esta junta, pero es realmente demasiado. Creo que estamos perdiendo mucho tiempo de la gente con toda esta recolección de datos, y haciendo nimiedades con las computadoras. Todo lo que ustedes tienen que hacer es hablar con nuestra gente de piso y con los gerentes de la tienda de abarrotes, para entender qué está pasando. He visto que esto ya ha pasado antes y aquí vamos de nuevo. Alguien de ustedes necesita apagar sus computadoras, salir de sus confortables oficinas y hablar con algunas personas reales”.

Ejemplo 1.2 Minorista grande con operaciones en la Web

Uno de los objetivos de un minorista grande con operaciones en Internet es ser el centro de consumo más grande del mundo. La compañía reconoce que la capacidad para establecer y conservar relaciones de largo plazo con los clientes, y motivarlos a repetir visitas y compras, depende, en parte, de la fortaleza de las operaciones de servicio al cliente. En lo referente al servicio que no puede manejarse usando las características de sitios Web, los representantes de servicio al cliente están localizados en centros de contacto disponibles las 24 horas del día, para atender llamadas de voz y correos electrónicos.

Debido al crecimiento de sus ventas y a su estacionalidad (el volumen del servicio es relativamente bajo en verano y alto cerca del fin de año), es un reto para la compañía dotar del personal adecuado a sus centros de contacto. El problema de planeación implica tomar decisiones acerca de la contratación y la capacitación de los centros administrados internamente, así como de la distribución del trabajo para los proveedores externos, con base en el volumen de llamadas de voz y correos electrónicos. El manejo de cada tipo de contacto debe satisfacer un nivel de servicio determinado cada semana.

Para hacer el problema aún más difícil, el tiempo de duración de cada llamada de voz y correo electrónico se ve afectado por varios atributos del contacto, incluyendo el tipo de producto, de cliente y de compra. Dichos atributos se usan para clasificar los contactos en categorías: en este caso, una categoría “principal” y siete categorías “especiales”. Se necesita un conjunto de habilidades específicas para resolver las diferentes clases de asuntos que surgen en las diferentes categorías. Puesto que la contratación y la capacitación requieren seis semanas, se hacen necesarios pronósticos de contactos de servicio, para tener el número requerido de representantes de servicio disponibles las 24 horas del día, los 7 días a la semana durante todo el año.

Pat Niebuhr y su equipo son los responsables de desarrollar un plan global de dotación de personal para los centros de contacto. Su reto inicial es pronosticar los contactos para las categorías principal y especiales. Pat debe trabajar con pronósticos mensuales de los pedidos totales (los cuales, a la vez, se derivan de los pronósticos de ingresos mensuales) y los contactos por pedido (CPP), números proporcionados por el departamento de finanzas. Pat sabe que los contactos están dados por:

$$\text{Contactos} = \text{pedidos} \times \text{CPP}$$

Para efectos de la planeación de la dotación de personal, Pat debe tener pronósticos de contactos sobre una base semanal. Por fortuna, existe una historia de los pedidos reales, contactos reales, contactos reales por pedido y otra información relevante, en algunos casos, registrada por día de la semana. Esta historia está organizada en una hoja de cálculo. Pat considera usar esta información histórica para desarrollar los pronósticos necesarios.

Resumen

La finalidad de un pronóstico es reducir el nivel de incertidumbre con que deben realizarse los juicios de la gerencia. Tal propósito sugiere dos reglas fundamentales, las cuales debe seguir el proceso de elaboración de pronósticos:

1. El pronóstico debe ser técnicamente correcto y generar pronósticos lo suficientemente precisos para satisfacer las necesidades de la empresa.
2. El procedimiento de elaboración de pronósticos y sus resultados tienen que presentarse convincentemente a la gerencia, de modo que los pronósticos se utilicen en el proceso de toma de decisiones para beneficio de la compañía; los resultados también deben justificarse desde el punto de vista del costo-beneficio.

Los pronosticadores a menudo dan mucha atención a la primera regla y gastan menos esfuerzo en la segunda. Si los pronósticos bien preparados y de costo efectivo van a beneficiar a la empresa, deben usarlos quienes tienen la autoridad para tomar decisiones. Esto motiva la cuestión de lo que podría llamarse las “políticas” del pronóstico. Gastos sustanciales y frecuentemente mayores, así como la asignación de recursos dentro de una empresa, con frecuencia se basan en el punto de vista de la gerencia acerca del curso de eventos futuros. Puesto que el movimiento de recursos y poder dentro de una organización por lo general se basa en la percepción del futuro (pronósticos), no es sorprendente encontrar cierta cantidad de intriga política alrededor del proceso de la elaboración de pronósticos. La necesidad de vender efectivamente los pronósticos a la gerencia es, por lo menos, tan importante como la necesidad de desarrollar los pronósticos mismos.

El resto de este libro analiza varios modelos y procedimientos para realizar pronósticos. Primero revisaremos los conceptos estadísticos básicos, y daremos una introducción a la correlación y al análisis de regresión. Luego, dedicamos un capítulo a los métodos de recolección de datos y exploración de conjuntos de datos para establecer patrones. Muchos métodos específicos de elaboración de pronósticos se estudian con detalle en los siguientes capítulos; los dos capítulos finales se dedican a los aspectos de aplicación del juicio a los pronósticos y la administración del proceso de realizar pronósticos.

CASOS

CASO 1-1 MR. TUX

John Mosby es propietario de varias tiendas de alquiler Mr. Tux, la mayoría de ellas en el área de Spokane, Washington.³ Su tienda de Spokane también confecciona camisas para esmoquin, las cuales distribuye a tiendas de alquiler a lo largo del país. Debido a que la actividad de alquiler varía de una temporada a otra por los bailes colegiales, reuniones y otras actividades, John sabe que su negocio es estacional. A él le gustaría medir este efecto estacional, tanto para ayudarse en la administración de su negocio, como para su uso en la negociación de la deuda de un préstamo con su banquero.

Aun de mayor interés para John es encontrar la forma de pronosticar sus ventas mensuales. Su negocio continúa creciendo, lo cual, a la vez, requerirá de más capital y deudas a largo plazo. Él tiene fuentes de financia-

miento de dos tipos: inversionistas y banqueros; no obstante, ambos están interesados en contar con un modo concreto de pronosticar sus ventas futuras. Aún cuando confían en John, su dicho de que el futuro de su negocio “luce grandioso” los deja inquietos.

Como primer paso en la construcción de un modelo de pronóstico, John da instrucciones a uno de sus empleados, McKennah Lane, para recolectar mensualmente los datos de ventas durante varios años pasados. En los capítulos que siguen, se usarán varias técnicas para pronosticar tales datos de ventas para Mr. Tux. En el capítulo 11, se resumen dichas técnicas, así como los intentos de John Mosby para seleccionar la técnica de elaboración de pronósticos que mejor se adapte a sus necesidades.

CASO 1-2 CONSUMER CREDIT COUNSELING

Consumer Credit Counseling (CCC), una empresa privada sin fines de lucro, se fundó en 1982.⁴ El objetivo de CCC es proporcionar a los consumidores ayuda en la planeación y el seguimiento de presupuestos, así como asistencia en la negociación con acreedores para liquidar deudas con morosidad y en capacitación sobre la administración del dinero.

El asesoramiento financiero privado se ofrece sin costo a las familias y a los individuos que atraviesan por dificultades financieras, o que quieren mejorar sus

habilidades para administrar el dinero. Los programas educativos de administración del dinero se dan a escuelas, grupos comunitarios y negocios. Se cuenta con un programa de administración de la deuda como una alternativa a la bancarrota. A través de este programa, CCC negocia con los acreedores en representación del cliente para acuerdos de pagos especiales. El cliente hace un pago fijo a CCC que luego se desembolsa a los acreedores.

³Agradecemos a John Mosby, el propietario de las tiendas de alquiler Mr. Tux, por su ayuda en la elaboración de este caso.

⁴Agradecemos a Marv Harnishfeger, director ejecutivo de Consumer Credit Counseling de Spokane, y a Dorothy Mercer, presidenta de su consejo de administración, por su ayuda en la preparación de este caso. Dorothy es una ex alumna de MBA de JH, que se ha mantenido consistentemente en contacto con nosotros en el uso de métodos cuantitativos en el mundo de negocios real.

CCC tiene una mezcla de plantilla laboral remunerada y voluntaria; de hecho, los voluntarios superan en número al personal remunerado en proporción de tres a uno. Siete empleados pagados brindan administración, apoyo de oficina y cerca de la mitad de necesidades de asesoría de CCC. Veintiún consejeros voluntarios completan la otra mitad de las necesidades de asesoría del servicio. CCC depende de en primera instancia de fondos corporativos para financiar operaciones y servicios. The Fair Share Funding Program permite a los acreedores, que reciben pagos de un cliente endeudado, administrar programas para donar al servicio una porción de los fondos devueltos a ellos a través de dichos programas.

La mayoría del apoyo corporativo viene de una empresa local de servicio público, que proporciona financiamiento para una posición de asesor de tiempo completo, así como espacio de oficina para asesoramiento en todas las oficinas.

Adicionalmente, los cobros a los clientes son una fuente de recursos. Los clientes que participan en la administración de la deuda pagan una cuota mensual de \$15 para ayudar a cubrir los costos administrativos de este programa. (Las cuotas se reducen o se condonan a los clientes que no pueden solventarlos).

Este antecedente se usará en los capítulos que siguen cuando CCC enfrente problemas difíciles relacionados con el pronóstico de variables importantes.

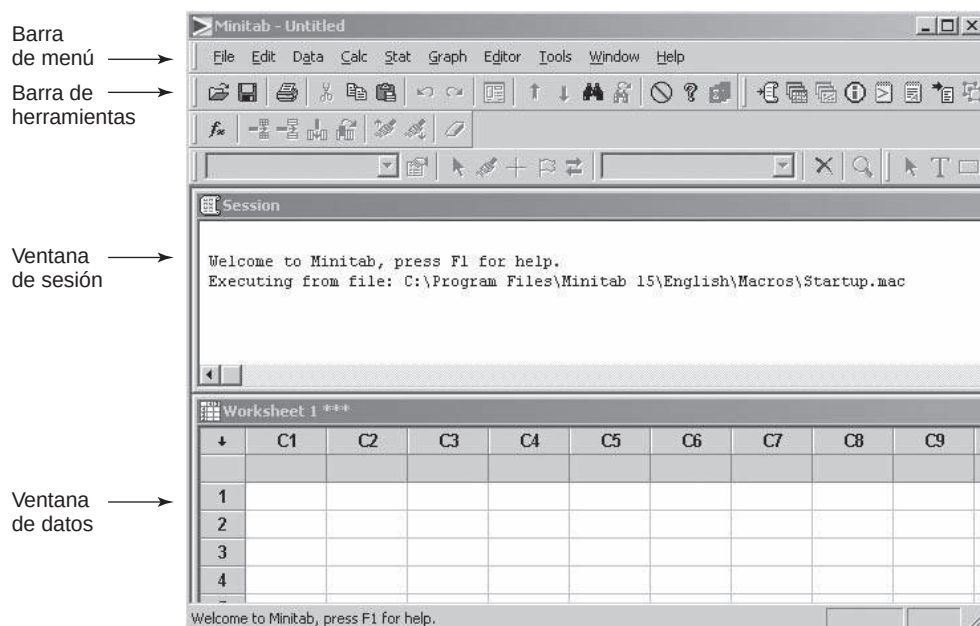
Aplicaciones de Minitab

Minitab es un programa estadístico avanzado que se mejora en cada versión. Lo descrito aquí es la versión 15.

La figura 1-1 le muestra a usted cuatro aspectos importantes de Minitab. La barra del menú es donde el usuario selecciona los comandos. Por ejemplo, al hacer clic en Stat aparece un menú desplegable que contiene todas las técnicas estadísticas disponibles. La barra de herramientas despliega botones de las funciones usadas comúnmente. Note que estos botones cambian dependiendo de cuál ventana de Minitab esté abierta. Hay dos ventanas separadas en la pantalla de Minitab: la ventana de datos, donde usted introduce, edita y observa la columna de datos de cada hoja de cálculo; y la ventana de sesión, que despliega el texto de salida o producto, como tablas de estadísticas.

En los capítulos siguientes se darán instrucciones específicas para facilitarle la introducción de datos en la hoja de cálculo de Minitab y para activar los procedimientos de elaboración de pronósticos, para producir los pronósticos necesarios.

FIGURA 1-1 Pantalla básica de Minitab



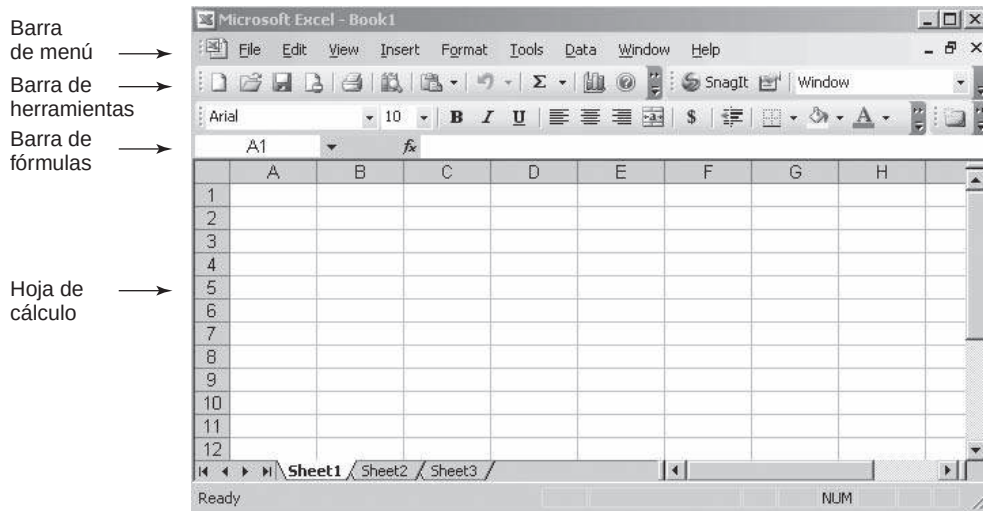


FIGURA 1-2 Pantalla básica de Excel

Aplicaciones de Excel

Excel es un programa de hoja de cálculo de uso común, que se usa frecuentemente para la elaboración de pronósticos. La figura 1.2 muestra la pantalla abierta de la versión 2003. Los datos se introducen en las filas y columnas de la hoja de cálculo y, luego, se dan los comandos para ejecutar varias operaciones con los datos introducidos.

Por ejemplo, los salarios anuales de varios empleados se pueden introducir en la columna A y el promedio de esos valores es calculado por Excel. Como otro ejemplo, se colocan las edades de los empleados en la columna B y se examina la relación entre la edad y el salario. Los capítulos que siguen le mostrarán a usted cómo usar Excel para resolver estos y otros problemas en la elaboración de pronósticos.

Existen algunas funciones estadísticas disponibles en Excel, que quizá no estén en los menús desplegables en su pantalla. Para activar estas funciones, haga clic en lo siguiente:

Tools>Add-Ins o Herramientas>Complementos

Aparece la ventana de diálogo Add-Ins. Seleccione Analysis ToolPak y dé clic en OK. La función disponible bajo ToolPak se utilizará en capítulos posteriores.

Se recomienda ampliamente que un *add-in* o complemento de Excel se use para ayudar con los diversos cálculos estadísticos, requeridos por las técnicas de elaboración de pronósticos estudiadas en este libro de texto.

Referencias

- Bernstein, P. L. *Against the Gods: The Remarkable Store of Risk*. Nueva York: Wiley, 1996.
- Carlberg, C. "Use Excel's Forecasting to Get Terrific Projections". *Denver Business Journal* 47 (18)(1996): 2B.

- Chase, C. W., Jr. "Selecting the Appropriate Forecasting Method". *Journal of Business Forecasting* 15 (otoño de 1997): 2.
- Diebold, F. X. *Elements of Forecasting*, 3a. ed. Cincinnati, Ohio: South-Western, 2004.

- Georgoff, D. M. y R. G. Mardick. "Manager's Guide for Forecasting". *Harvard Business Review* 1 (1986): 110-120.
- Hogarth, R. M., and S. Makridakis. "Forecasting and Planning: An Evaluation." *Management Science* 27 (2) (1981): 115-138.
- Hyndman, R. J., and J. K. Ord, eds. "Special Issue: Twenty Five Years of Forecasting." *International Journal of Forecasting* 22 (2006): 413-636.
- Levenbach, H., and J. P. Cleary. *Forecasting Practice and Process for Demand Management*. Belmont, Calif.: Thomson Brooks/Cole, 2006.
- Makridakis, S. "The Art and Science of Forecasting." *International Journal of Forecasting* 2 (1986): 15-39.
- Newbold, P., and T. Bos. *Introductory Business and Economic Forecasting*, 2nd ed. Cincinnati, Ohio: South-Western, 1994.
- Ord, J. K., and S. Lowe. "Automatic Forecasting." *American Statistician* 50 (1996): 88-94.
- Perry, S. "Applied Business Forecasting." *Management Accounting* 72 (3) (1994): 40.
- Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts*, 4th ed. New York: McGraw-Hill, 1998.
- Wright, G., and P. Ayton, eds. *Judgemental Forecasting*. New York: Wiley, 1987.

CAPÍTULO

2

REPASO DE CONCEPTOS ESTADÍSTICOS BÁSICOS

La mayoría de las técnicas para la elaboración de pronósticos se basan en conceptos estadísticos fundamentales que se estudian en libros de texto de estadística en los negocios y en cursos introductorios de estadística. Este capítulo examina algunos de los conceptos básicos que servirán como base para la mayoría del material en el resto del libro.

La mayoría de los procedimientos estadísticos hacen deducciones sobre datos de interés, llamados *población*, después de la selección y la medición de un subgrupo de tales datos, llamado *muestra*. La selección cuidadosa de una muestra representativa y el uso de una muestra lo suficientemente grande son componentes importantes de un proceso estadístico deductivo, que tenga un grado de riesgo aceptablemente bajo.

En la elaboración de pronósticos, generalmente trabajamos con datos históricos en un intento por predecir el futuro incierto. Por tal motivo, nos concentraremos en examinar los datos de la muestra, en manipular estos datos de alguna forma y en usar los resultados para efectuar pronósticos.

DESCRIPCIÓN DE DATOS CON RESÚMENES NUMÉRICOS

El objetivo de los procedimientos estadísticos descriptivos es detallar brevemente un conjunto grande de mediciones, con unos cuantos valores clave resumidos. El resumen más común se obtiene promediando los valores. En estadística, el proceso de promediar generalmente se realiza calculando la *media*, la cual implica la suma de todos los valores y la división entre el número de éstos. Aunque en cierto sentido la media indica la “parte de en medio”, este cálculo es sensible a unas cuantas mediciones que sean mucho más grandes o mucho más pequeñas que el resto. Estas mediciones inusuales se denominan *valores atípicos*. El ejemplo 2.1 presenta una ilustración de este fenómeno.

La media de la muestra ($\bar{X}$ o X -testada) se calcula usando la ecuación 2.1.

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

donde

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

$\sum X$ = la suma de todos los valores de la muestra

n = el tamaño de la muestra

Para simplificar los cálculos en este texto, se usa un tipo de notación abreviada. En la notación simplificada para la suma de todos los valores de X (véase la ecuación 2.1), se entiende que la sumatoria va desde 1 hasta n . Un sistema de notación más formal para este procedimiento es

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

donde el subíndice i varía de su valor inicial de 1 a n , en incrementos de 1. Puesto que casi todas las sumas van de 1 a n , omitiremos los índices inicial y final, y usaremos la notación más sencilla, excepto donde sea necesaria una notación más completa para fines de claridad.

Además de la tendencia central de un conjunto de valores, obtenidos al calcular la media, también es interesante conocer el rango (intervalo) a través del cual se dispersan los valores alrededor de la media. La *desviación estándar* es una unidad para medir la distancia desde la media. La fórmula para la desviación estándar está dada por

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

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

Muchos procedimientos estadísticos utilizan la *varianza* de la muestra. La varianza de un conjunto de medidas es la desviación estándar al cuadrado. Entonces, 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 Cálculo de media, varianza y desviación estándar

Considere el siguiente conjunto de edades de personas:

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

Para esta muestra, $n = 10$ y

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

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

$$S^2 = \frac{\sum (X - \bar{X})^2}{n - 1} = \frac{(23 - 40.2)^2 + \cdots + (35 - 40.2)^2}{10 - 1} = \frac{1,339.6}{9} = 148.84$$

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

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 12.2 años.

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
		$\Sigma(X - \bar{X})^2 = 1,339.60$
		$S^2 = \frac{1,339.60}{10 - 1} = 148.84$

Para indicar la sensibilidad de la media a una medición atípica, suponga que sustituimos la edad 60 con una edad mucho mayor, digamos 90. Sustituyendo esta edad, la media sería

$$\bar{X} = \frac{\Sigma X}{n} = \frac{23 + \cdots + 25 + 90 + 55 + \cdots + 35}{10} = \frac{432}{10} = 43.2$$

Incrementar el tamaño de una sola de las diez edades originales da como resultado un incremento de 3 años en la media de la edad.

El término *grados de libertad* se usa para indicar el número de datos que son libres entre sí, en el sentido de que no pueden calcularse uno a partir del otro y, por lo tanto, llevan piezas de información únicas. Por ejemplo, suponga que se hacen los tres enunciados siguientes:

Estoy pensando en el número 5. Estoy pensando en el número 7. La suma de los dos números en que estoy pensando es 12.

A primera vista, se presentan tres piezas de información. Sin embargo, si se conocen cualesquiera dos de estos enunciados, el otro se podría inferir. Diríamos entonces que sólo hay dos piezas únicas de información en los tres enunciados o, usando el término estadístico, sólo hay dos grados de libertad porque únicamente dos de los valores tienen libertad para variar; no así el tercero.

Los *grados de libertad* se refieren al número de datos que son independientes uno de otro y que tienen piezas de información únicas

En el ejemplo que se presenta en la tabla 2-1, las edades de 10 personas constituyen una muestra con 10 grados de libertad. Cualquier edad pudo haberse incluido en la muestra y, por lo tanto, cada una de las edades tiene libertad para variar. Cuando se calcula la media, se usan las 10 edades para contar con una edad media igual a 40.2 años.

El cálculo de la desviación estándar de la muestra difiere. Cuando se calcula la desviación estándar de la muestra, se usa un estimado de la media de la población (la media de la muestra $\bar{X}$). Si en los cálculos se emplea la media de la muestra como un estimado de la media de la población, por lo general se obtendrá una desviación estándar más pequeña que la desviación estándar de la población. Sin embargo, este problema se corrige dividiendo el valor $\sum(X - \bar{X})^2$ entre los grados de libertad adecuados. Una vez que se calcula la media de la muestra, sólo se requieren nueve de las desviaciones $X - \bar{X}$, en este ejemplo, para calcular la desviación estándar de la muestra. Dadas nueve de las desviaciones, se determina la última desviación, puesto que $\sum(X - \bar{X})$ debe ser igual a cero. Por lo tanto, decimos 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 muestral para estimar un parámetro de la población, se pierde un grado de libertad.

Se pueden definir estadísticos de resumen para poblaciones. Para distinguir entre parámetros poblacionales de estadísticos muestrales, usamos notaciones diferentes. La tabla 2-2 presenta los símbolos usados tanto para parámetros de la población como para estadísticos muestrales.

La media y la desviación estándar son las medidas más comúnmente usadas para describir datos muestrales de forma breve y significativa. Sin embargo, están disponibles otras medidas de resumen descriptivas. Por ejemplo, la *mediana* se usa a menudo para indicar el valor de la parte media o central en un conjunto de datos. La mediana es el valor que divide por la mitad un conjunto de mediciones ordenadas. Es decir, la mitad de las mediciones son menores que la mediana y la mitad son mayores. Puesto que la mediana simplemente divide un conjunto de mediciones por la mitad, es invariable si, por ejemplo, se incrementa la medición más grande. Decimos que la mediana es *resistente* a los efectos de valores atípicos.

Algunas veces, el *rango* de datos se presenta como una medida de dispersión burda. El rango es la diferencia entre los valores mayor y menor. Por ejemplo, el rango de edades en la tabla 2-1 es 37 (60-23).

Después de que los valores numéricos se ordenan de menor a mayor, los *cuartiles* dividen el conjunto de datos en cuatro partes iguales. Como hemos visto, la mediana divide el conjunto ordenado en dos partes iguales y, algunas veces, se llama el *segundo cuartil* y se denota con Q_2 . El *primer cuartil* (Q_1) divide la mitad inferior del conjunto en dos partes iguales, en tanto que el *tercer cuartil* (Q_3) divide la mitad superior en dos partes iguales. El conjunto 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 $Q_3 = 51.25$.

Finalmente, el *rango intercuartil* es otro indicador de la variabilidad de un conjunto de datos. Es simplemente la diferencia entre el tercer cuartil y el primer cuartil ($Q_3 - Q_1$), es decir, el rango para el 50% central de los valores de los datos. Para los datos de edades, el rango intercuartil es 21 ($51.25 - 30.25$).

Minitab y Excel sirven para calcular la mayoría de los estadísticos descriptivos presentados hasta aquí. La figura 2-1 muestra el resultado de Minitab para los datos de edades de la tabla 2-1. Las instrucciones para el cálculo de estadísticos descriptivos usando Minitab y Excel se presentan en las secciones de aplicaciones de Minitab y Excel, al final de este capítulo.

TABLA 2-2 Notación para parámetros de la población y de la muestra

<i>Estadístico</i>	<i>Símbolo de población</i>	<i>Símbolo de muestra</i>
Media	μ	$\bar{X}$
Varianza	σ^2	S^2
Desviación estándar	σ	S

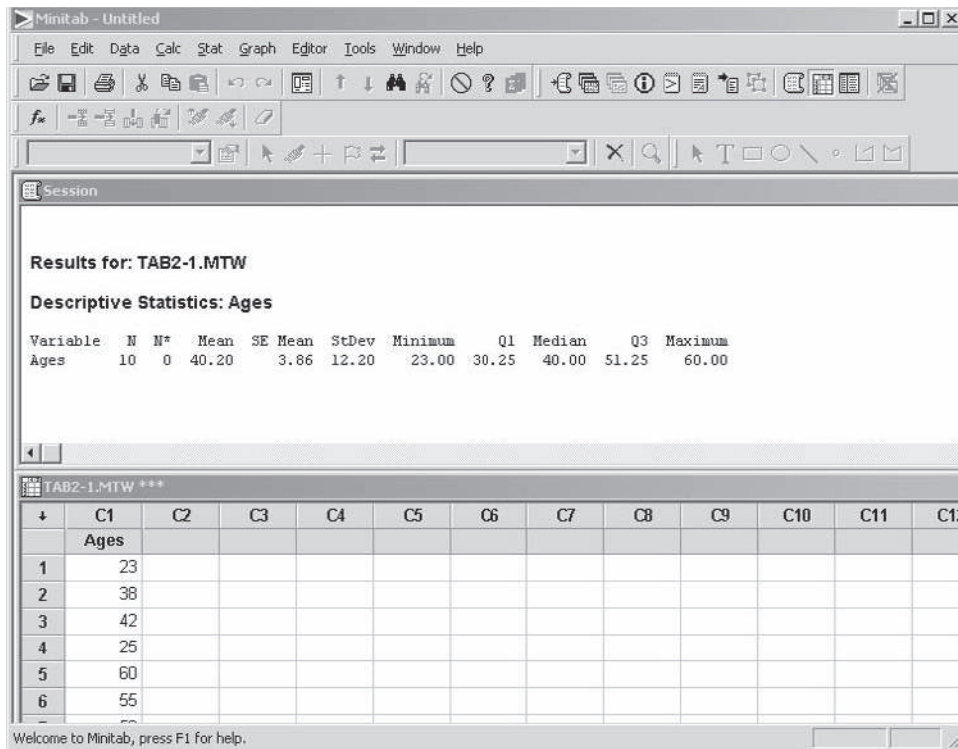


FIGURA 2-1 Resultado de Minitab para estadísticos descriptivos

PRESENTACIÓN DE INFORMACIÓN NUMÉRICA

En la elaboración de pronósticos, trabajamos con dos tipos de datos: *datos transversales*, los cuales consisten en observaciones desde el mismo marco de referencia temporal; y datos de *series de tiempo*, que consisten en una secuencia de observaciones en el tiempo. La tabla 2-3 muestra un ejemplo de datos transversales: el ingreso neto como un porcentaje del capital invertido para una muestra de 209 compañías en una encuesta específica de Fortune 500. Otro ejemplo son los ingresos anuales de ejecutivos corporativos en 2006 y los precios de venta de casas en cierta ciudad en un día determinado. Algunos ejemplos de datos de series de tiempo son el número de visitantes mensuales al Grand Coulee Dam Visitor Center durante varios años, las ventas mensuales de Sears Corporation durante los últimos diez años, el precio diario de la acción de IBM para 2006 y la producción anual de trigo en Estados Unidos para el periodo 2000-2005.

Una de las cuestiones más importantes que se deben hacer primero cuando se explora una variable es visualizar los datos en tablas y gráficas. Las características básicas de los datos, incluyendo observaciones inusuales y patrones únicos, se perciben más fácil visualmente. A veces, incluso, las gráficas sugieren explicaciones posibles para alguna variación en los datos.

Un *diagrama de puntos* es una de las formas más simples de visualizar gráficamente los datos. En la figura 2-2 el eje horizontal presenta el rango de valores para el ingreso neto como un porcentaje del capital invertido de la tabla 2-3. Cada observación se representa usando un punto colocado encima del eje. Los diagramas de puntos muestran los detalles de los datos y también le permiten a un pronosticador comparar dos o más conjuntos de datos.

Un *diagrama de caja* es útil para mostrar las características de distribución de los datos. La figura 2-3 ilustra un diagrama de caja del ingreso neto como un porcentaje de los datos del capital invertido de la tabla 2-3. En la mediana se traza una línea a lo largo de la caja. Esta línea

TABLA 2-3 Ingreso neto como un porcentaje del capital invertido para una muestra de 209 compañías en una 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	

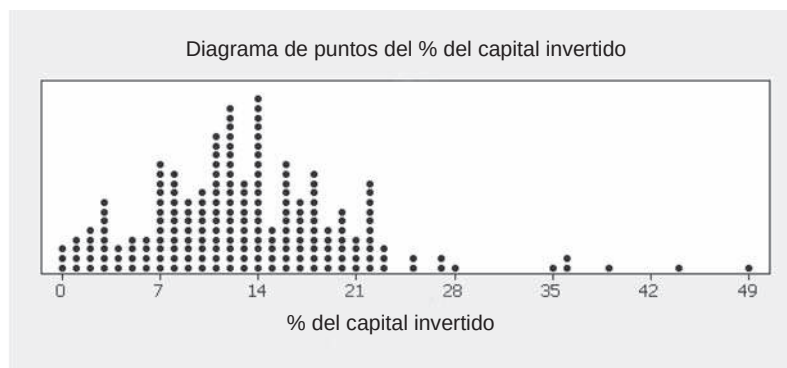


FIGURA 2-2 Diagrama de puntos para el ingreso neto como un porcentaje del capital invertido para una muestra de 209 compañías en una encuesta de Fortune 500

divide los datos en dos secciones iguales. El borde inferior del cuadro es el primer cuartil (Q_1), y el borde superior es el tercer cuartil (Q_3). Se construyen límites adicionales usando el rango intercuartil ($Q_3 - Q_1$). El límite inferior se localiza en $Q_1 - 1.5(Q_3 - Q_1)$, y el límite superior se localiza en $Q_3 + 1.5(Q_3 - Q_1)$. Los valores atípicos están identificados como aquellos puntos fuera de los límites inferior y superior, y están representados por asteriscos. En la figura 2-3, el primer cuartil es 8, la mediana es 13, el tercer cuartil es 17, y el rango intercuartil es

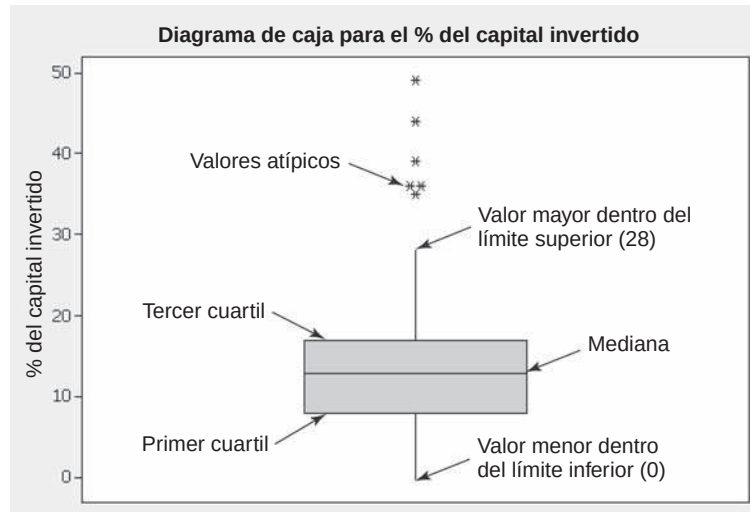


FIGURA 2-3 Diagrama de caja para el ingreso neto como un porcentaje del capital invertido para una muestra de 209 compañías en una encuesta de Fortune 500

9 ($17 - 8$). El límite inferior es -5.5 ($8 - 1.5 \times 9$) y el límite superior es 30.5 ($17 + 1.5 \times 9$). Note que el valor más bajo dentro del límite inferior, -5.5 , es 0; en tanto 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 los valores atípicos.

El *histograma* condensa los datos, agrupando valores similares en clases. Se puede construir un histograma colocando la variable de interés en el eje horizontal, y la frecuencia, frecuencia relativa o frecuencia porcentual, en el eje vertical. Al observar un histograma como el de la figura 2-4, usted puede indicar la proporción del área total, arriba de un intervalo del eje

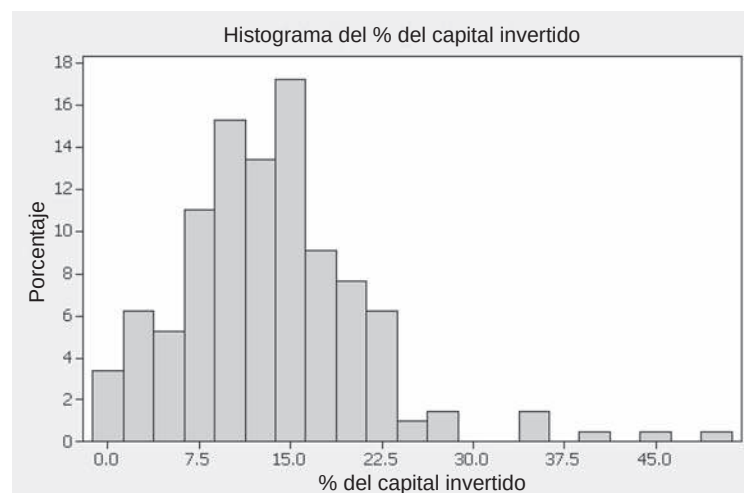


FIGURA 2-4 Histograma del ingreso neto como un porcentaje del capital invertido para una muestra de 209 compañías en una encuesta de Fortune 500

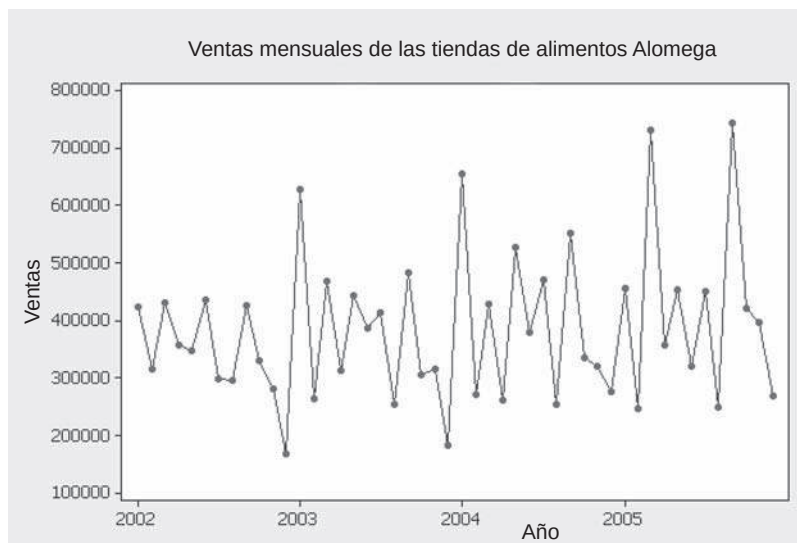


FIGURA 2-5 Ventas mensuales de las tiendas de alimentos Alomega (de la tabla 3-11)

horizontal. Un total de 23 de las 209 compañías, o el 11.0%, tienen ingreso neto como porcentaje del capital invertido en el rango entre 6.25% y 8.75%. Por lo tanto, el cuarto rectángulo de la figura 2-4 contiene .110 del área total representada por las barras verticales sombreadas.

Aunque los histogramas ofrecen buenas descripciones visuales de conjuntos de datos, especialmente de conjuntos de datos muy grandes, no nos permiten identificar observaciones individuales. En contraste, en un diagrama de puntos, está visible cada uno de los valores originales.

Se utilizan *diagramas de dispersión* para visualizar la relación entre dos variables. Tales diagramas se analizarán más adelante, en el capítulo de la sección sobre análisis de correlación.

Para datos cronológicos, la forma gráfica más frecuentemente empleada es un *diagrama de series de tiempo*, en el cual los datos se grafican a través del tiempo. La figura 2-5 ilustra el diagrama de una serie de tiempo con las ventas anuales en las tiendas de alimentos Alomega, que analizamos en el ejemplo 1.1. Un diagrama de series de tiempo revela la variabilidad de los datos y el momento en que ocurren picos y valles. También muestra el tamaño relativo de los picos y los valles, en comparación con el resto de la serie.

Un paso importante en la selección de una técnica de pronósticos adecuada consiste en identificar los patrones de datos que existen dentro de una serie de tiempo. Una vez que se identifican los patrones de datos, se pueden utilizar los métodos de pronóstico más adecuados para tales patrones. Se identifican cuatro tipos de patrones de datos para series de tiempo: horizontal, de tendencia, cíclico y estacional. Cada tipo se estudiará con profundidad en el capítulo 3.

Una herramienta utilizada con frecuencia para identificar patrones de datos en las series de tiempo es el *correlograma* o la *función de autocorrelación*, que es una herramienta gráfica para presentar las correlaciones entre diversos retrasos de una serie de tiempo. Los correlogramas también se examinarán en el capítulo 3.

DISTRIBUCIONES DE PROBABILIDAD

Variable aleatoria es el nombre que se le da a una cantidad que puede tomar valores diferentes de una prueba a otra en un experimento, donde el resultado exacto es un suceso fortuito. Si sólo son posibles ciertos valores específicos, la variable aleatoria se llama *variable discreta*. Ejemplos son el número de habitaciones en una casa, el número de gente que llega a pagar a las cajas de un supermercado en cierta hora, y el número de unidades defectuosas en un lote de partes elec-

trónicas. Si es posible cualquier valor de la variable aleatoria dentro de cierto rango, se llama *variable continua*. Ejemplos de este tipo de variable son el peso de un individuo, la longitud de una parte manufacturada, y el tiempo entre llegadas de automóviles a una caseta de peaje.

Una *variable aleatoria discreta* puede tener valores tan sólo de un conjunto predeterminado. A menudo tales resultados se presentan como números enteros. Una *variable aleatoria continua* puede tomar cualquier valor dentro de un rango específico. Tales resultados se representan numéricamente con intervalos de valores.

La *distribución de probabilidad* de una variable aleatoria discreta enlista todos los valores posibles que puede tomar la variable, junto con la probabilidad de que ocurra cada uno. El *valor esperado* de una variable aleatoria es el valor medio (promedio) que adquiere la variable en muchos intentos realizados. El valor esperado, $E(X)$, para una distribución de probabilidad discreta se obtiene al multiplicar cada valor posible de X por su probabilidad, $P(X)$, y después se suman los productos. La ecuación 2.4 presenta 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 varios ensayos u observaciones.

Ejemplo 2.2 Una distribución de probabilidad discreta

El número de días sin ventas de un vendedor, durante un mes, se describe mediante la distribución de probabilidad mostrada en la tabla 2-4. Estos valores se basan en experiencias del vendedor y se usarán para pronosticar la actividad futura de ventas. La columna X lista todos los valores (días sin ventas) que son posibles, mientras que la columna $P(X)$ lista las probabilidades correspondientes. Observe que puesto que se listan todos los valores posibles de X , las probabilidades suman 1.00, o 100%. Esto es válido para todas las distribuciones de probabilidad, ignorando los errores por redondeo.

Para la distribución de probabilidad dada en la tabla 2-4, el valor esperado es

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

De modo que si se observara a este vendedor durante un número de meses muy grande y se registrara el número de días sin venta, la media sería 3.35 días sin venta, si la actividad futura está pronosticada correctamente por los datos históricos, en los cuales se basa la distribución de probabilidad.

TABLA 2-4 Distribución de probabilidad discreta para el ejemplo 2.2

X	$P(X)$
1	.10
2	.20
3	.25
4	.15
5	.30

Advierta que la media cae cerca de la parte central de los valores X . Es atraída hacia el extremo superior de la distribución de probabilidad debido a la probabilidad relativamente alta asociada con $X = 5$.

Para una distribución continua, la probabilidad de obtener un valor específico se aproxima a cero. Por ejemplo, la probabilidad de que alguien pese 150 libras podría considerarse igual a cero, ya que esto significaría que tal peso sea exactamente 150.0000 —sin importar la precisión de la escala que se use. En el caso de la distribución continua, las probabilidades se asignan a intervalos o rangos de valores. Por ejemplo, podría calcularse la probabilidad de que el peso de un individuo caiga en el intervalo de 149 a 151 libras.

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

1. Hay n ensayos idénticos, cada uno de los cuales da como resultado uno de dos resultados posibles —digamos, éxito y fracaso.
2. La probabilidad de éxito para cada resultado se mantiene igual de un ensayo a otro.
3. Los ensayos son independientes (no están relacionados).

Nos interesa encontrar la probabilidad X de éxitos que ocurren en n ensayos, donde arbitrariamente el éxito se define como uno de los dos resultados posibles. Los diversos valores de X , junto con sus probabilidades, forman la distribución binomial. Estas probabilidades se calculan con 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

$\binom{n}{X}$ = el número de combinaciones de n elementos tomados X a la vez

p = la probabilidad de éxito en cada ensayo

X = el número específico de éxitos que nos interesa

n = el número de ensayos

La *distribución binomial* es una distribución de probabilidad discreta, que describe la probabilidad de X éxitos en n ensayos independientes de un experimento binomial.

Una forma más sencilla de encontrar probabilidades binomiales que usar la ecuación 2.5 es remitirse a una tabla de distribución binomial como la tabla B-1, que se encuentra en el Apéndice B. Los bloques que representan n agrupan las probabilidades. Cada bloque tiene una columna titulada con p y filas indicadas con X .¹

¹Minitab y Excel también calculan probabilidades binomiales.

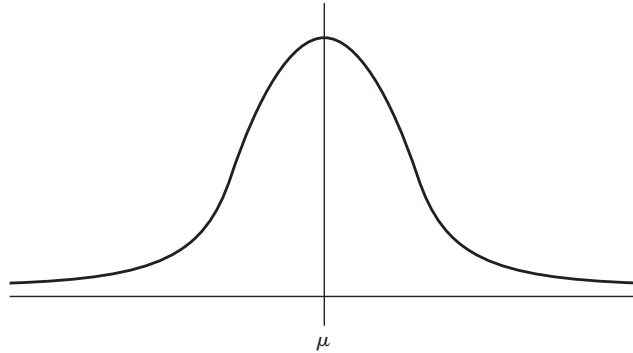


FIGURA 2-6 Una distribución normal

Ejemplo 2.3 Una probabilidad binomial

Suponga que ocho artículos se extraen aleatoriamente de una línea de producción que, se sabe, produce partes defectuosas el 5% de las veces. ¿Cuál es la probabilidad de obtener exactamente cero defectuosos? La respuesta de la tabla B-1 es 0.6634. (Aquí, $n = 8$, $p = .05$ y $X = 0$).

La distribución continua importante, ya que muchas poblaciones de números se pueden aproximar mediante ella, es la *distribución normal*. Para identificar una población normal específica, es necesario conocer la media y la desviación estándar. Una curva normal es simétrica y tiene forma de campana, como se ilustra en la figura 2-6. Esta distribución representa muchas variables de la vida real que se miden en una escala continua.

Las probabilidades de valores extraídos de una distribución normal, que caen en varios intervalos, se obtienen convirtiendo primero todos los intervalos en unidades de desviación estándar, llamadas Z como número de varianzas.² El resultado Z , como el número de varianzas de cualquier valor de X , es el número de desviaciones estándar desde el valor central de la curva (μ) hasta ese valor. Entonces, la fórmula es

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

donde

X = el valor de interés

μ = la media

σ = desviación estándar

Después de que se haya calculado Z como número de varianzas, se puede consultar la tabla de la curva normal, para encontrar el área bajo la curva entre el centro de la curva original (μ) y el valor de interés (X).³

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

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

³Se pueden calcular las probabilidades normales usando Minitab o Excel.

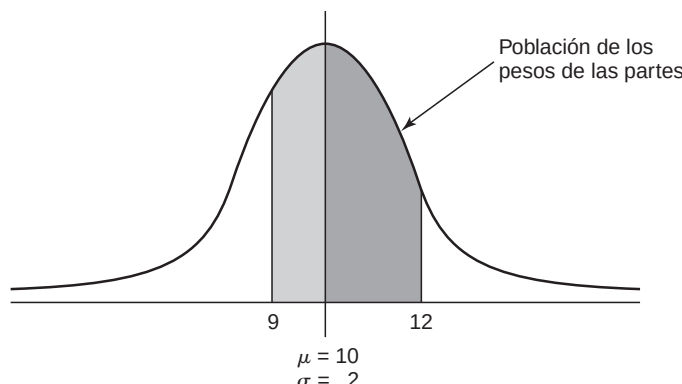


FIGURA 2-7 Áreas de la curva normal para el ejemplo 2.4

Ejemplo 2.4 Cálculo de una probabilidad normal

Los pesos de una población de partes fabricadas con cierta máquina, están distribuidos normalmente con una media de 10 libras y una desviación estándar de 2 libras. ¿Cuál es la probabilidad de que una pieza seleccionada al azar de la máquina pese entre 9 y 12 libras? La curva normal con el área sombreada adecuada se ilustra en la figura 2-7.

Puesto que a menudo las tablas de la curva normal se diseñan para dar áreas desde el centro de la curva hasta algún punto, deben obtenerse dos áreas separadas: una a cada lado de la media. Luego se suman estas áreas. Dicho proceso indicará la probabilidad de que un valor caiga en este intervalo. Los dos valores de Z son

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

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

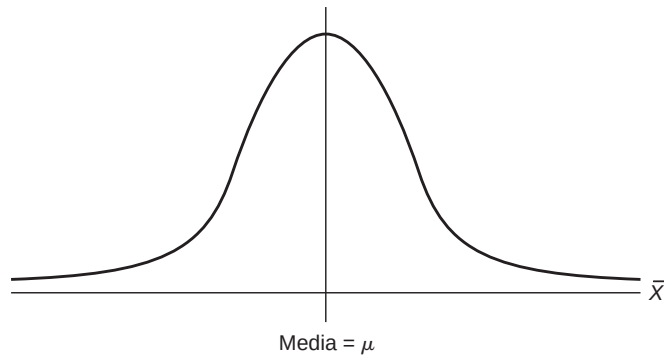
Por lo tanto, el área bajo la curva normal entre 9 y 12 de la figura 2-7 es la misma que el área bajo la curva normal de Z (véase la nota al pie de página 2) entre $-.5$ y 1 . Para la tabla normal de este libro, se ignora el signo negativo en el primer valor de Z , ya que el área de la curva normal para Z entre $-.5$ y 0 es la misma que el área bajo la curva entre 0 y $.5$. Estos dos valores de Z se localizan en la tabla B-2, una tabla de curva normal, darán como resultado las siguientes dos áreas, las cuales se suman después:

$$\begin{array}{rcl}
 Z_1 = -.50 & : & .1915 \\
 Z_2 = 1.00 & : & .3413 \\
 & & \hline
 & & .5328
 \end{array}$$

Se concluye que hay aproximadamente un 53% de probabilidad de que una parte de esta población, extraída al azar, pesará entre 9 y 12 libras.

DISTRIBUCIONES MUESTRALES

En la mayoría de aplicaciones estadísticas, se extrae una muestra al azar de la población en estudio, se calcula un estadístico a partir de los datos de la muestra y se infieren conclusiones acerca de la población sobre la base de dicha muestra. Una *distribución muestral* es la distribución de todos los valores posibles del estadístico muestral que se pueden obtener de la

FIGURA 2-8 Distribución muestral de $\bar{X}$

población, para un tamaño de muestra dado. Por ejemplo, de una población se toma una muestra aleatoria de 100 personas, se pesan y luego se calcula su media (promedio) de peso. Esta media de la muestra ($\bar{X}$) puede considerarse como si fuera extraída de la distribución de todas las medias muestrales posibles, con un tamaño de 100, que podrían tomarse de la población. Más generalmente, cada estadístico muestral que se calcule a partir de los datos de la muestra puede considerarse como si se hubiera tomado de una distribución muestral.

Una *distribución muestral* es el arreglo de todos los valores posibles de un estadístico muestral, que puede extraerse de una población, para un tamaño de muestra dado.

El *teorema central del límite* indica que, conforme 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 μ , la media de la población; y la desviación estándar es $\sigma/\sqrt{n}$, la desviación estándar de la población, dividida entre la raíz cuadrada del tamaño de la muestra. La cantidad $\sigma/\sqrt{n}$ se conoce como el *error estándar* de la media de la muestra. Así, la distribución muestral de la media de la muestra tenderá hacia la normalidad, sin importar la forma de la distribución de la población de la cual se extrajeron las muestras. La figura 2-8 demuestra cómo se observaría tal distribución muestral.

El teorema central del límite es de importancia fundamental en estadística, ya que permite a los analistas calcular la probabilidad de varios resultados de muestras, usando el conocimiento de las probabilidades de la curva normal.

Ejemplo 2.5 Cálculo de una probabilidad usando la distribución muestral de $\bar{X}$

¿Cuál es la probabilidad de que la media de una muestra aleatoria de 100 pesos —extraída de una población— estará dentro de 2 libras del peso medio de la población verdadera, 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 $\sigma/\sqrt{n} = 15/\sqrt{100} = 1.5$, que da como resultado un valor Z de $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n}) = 2/1.5 = 1.33$ y una área de la tabla de la curva normal de .4082. Al duplicar esta área para contabilizar ambos lados de la media, da como resultado una área total de .8164. Las probabilidades son del 82% de que la media de la muestra estará dentro de 2 libras de la media verdadera, dado el tamaño de la muestra de 100 y la variabilidad estimada de la población, $\sigma = 15$. Como se verá, esta destreza al calcular las probabilidades de los resultados de una muestra facilitará a un analista hacer inferencias acerca de la población de la muestra, con aplicabilidad directa a la realización de pronósticos.

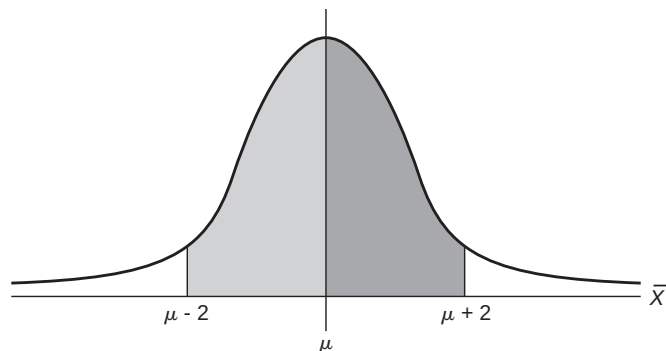


FIGURA 2-9 Áreas de distribución muestral para el ejemplo 2.5

Para tamaños de muestra pequeños, quizá las medias muestrales no estén normalmente distribuidas. Sin embargo, si la población de la cual se selecciona la muestra es esencialmente normal y la desviación estándar de la población se estima con 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 *distribución t*) con $n - 1$ grados de libertad (que se abrevian gl). La distribución t está centrada en 0 y tiene una forma de campana, pero con colas más gruesas que en una distribución normal. Sin embargo, conforme se incrementa el número de grados de libertad (y de forma equivalente, el tamaño de la muestra n), la distribución t parece cada vez más una distribución normal con una media de 0 y una desviación estándar de 1.

La tabla B-3 (en el Apéndice B) presenta valores tomados de la distribución t . Observe que sólo se requiere especificar un valor antes de remitirse a la tabla; a saber, los grados de libertad. Una vez que se conocen los grados de libertad, se pueden obtener los valores t que excluyen 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 en cada lado del 0 incluirá el 95% del área bajo la curva y excluirá el 5% de ella, es decir, 2.5% en cada cola.

Ejemplo 2.6 Uso de la distribución t

El gerente de una compañía de tarjetas de crédito afirma que el tiempo medio para resolver una queja de los clientes es de 30 días. Para probar tal aseveración, se recopilaron datos de una muestra de 15 quejas, las cuales tuvieron un tiempo medio para resolverse de 35.9 días y una desviación estándar de 10.2 días.

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

Si la afirmación del gerente es correcta, la media de la población es $\mu = 30$ y 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 afirmación del gerente será sospechosa, si el valor de esta variable es demasiado grande o demasiado pequeño, porque esto ocurre cuando el valor de $\bar{X}$ está lejos de μ . Consultando la tabla B-3 para 14 grados de libertad, vemos que $t_{.025} = 2.145$. Antes de obtener la muestra, la probabilidad de que t excederá 2.145 es .025. Asimismo, la probabilidad de que t será menor que -2.145 es de .025. Sumando estas dos probabilidades iguales, la probabilidad es $.025 + .025 = .5$ de que t será más grande

que 2.145 o menor que -2.145 . Ahora estamos listos para determinar si el valor de $\bar{X}$, 35.9, está lejos de μ , 30. El valor observado de t es

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

el cual excede $t_{.025} = 2.145$. En vez de aceptar la explicación de que ocurrió un evento con escasa probabilidad, cuestionamos la afirmación de que la media sea de 30 días. Un valor más grande para la media sería más consistente con los datos.

INFERENCIA DE UNA MUESTRA

Estimación

Cuando no es viable o incluso es imposible medir la población total de interés, se selecciona una muestra de la población y se utiliza para aprender acerca de la población. Este aprendizaje (llamado *inferencia*) por lo común toma dos formas. La primera de estas se denomina *estimación*, donde los resultados de la muestra se usan para estimar una característica desconocida de la población. Si bien la estimación es el término estadístico común para dicha tarea, también podría llamarse pronóstico en muchas situaciones de negocios, porque la muestra consiste en observaciones históricas recopiladas y se requiere un valor (estimado o pronosticado) para una observación futura. La segunda forma de aprendizaje acerca de la población a partir de una muestra se llama *prueba de hipótesis* y se estudiará en la siguiente sección.

Una *estimación puntual* (un pronóstico) de una cantidad de la población, como la media poblacional, es un solo valor calculado a partir de los datos de la muestra, y estima el valor de la cantidad desconocida de una población. En estadística, una cantidad de la población se conoce como *parámetro* de ésta. Una estimación puntual es la “mejor suposición” de un parámetro de la población calculada a partir de la muestra. Con frecuencia, la mejor suposición de un parámetro de la población lo proporciona la cantidad correspondiente de la muestra. Por ejemplo, la mejor suposición 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 la población y los estadísticos muestrales que dieron estimaciones puntuales de aquéllos.

Un *intervalo de estimación* es un intervalo dentro del cual es probable que esté el parámetro de la población de interés. Se obtiene construyendo un intervalo alrededor de la estimación puntual:

$$\text{Estimación puntual} = \pm \text{múltiplo} \times (\text{desviación estándar estimada de la estimación puntual}) \quad (2.8)$$

donde el “múltiplo” es a menudo un punto porcentual de la distribución normal o un punto porcentual de la distribución t . Se prefiere un intervalo de estimación por encima de la estimación puntual, ya que aquélla toma en cuenta la incertidumbre asociada con una estimación.

Idealmente, nos gustaría obtener una muestra y luego determinar un intervalo que, por ejemplo, contuviera en definitiva la media de la población, μ . Sin embargo, debido a la variabilidad de una muestra a otra, tal objetivo es inalcanzable. En vez de eso, especificamos una alta probabilidad —digamos, .95 o .99— de que un intervalo propuesto cubrirá el valor verdadero del parámetro de la población de interés. Como esta probabilidad pertenece al intervalo antes de que se observe la muestra, se denomina el *nivel de confianza*. Un intervalo obtenido mediante un procedimiento que satisfaga el requisito de probabilidad se llama *intervalo de confianza*. La cantidad a la derecha del signo $\pm$ en la ecuación 2.8 se conoce a menudo como el *error permitido*, y su magnitud dependerá directamente del nivel de confianza (a través de un múltiplo) y la información en la muestra (a través de la desviación estándar estimada de la estimación puntual).

Una *estimación puntual* es un solo valor estimado de un parámetro de la población. Un *intervalo de estimación* es un intervalo dentro del cual es probable que esté el parámetro de la población.

Ejemplo 2.7 Cálculo de estimaciones puntual y de intervalos de estimación

Se elige una muestra aleatoria de 500 compradores en el Centro Comercial Northgate, para determinar la distancia promedio que recorrieron para llegar al centro comercial. 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, para todos los compradores del centro comercial, es de 23.5 millas, la media de la muestra.

La desviación estándar estimada de $\bar{X}$ es $S/\sqrt{n}$ y para muestras aleatorias grandes (digamos, $n > 30$), el múltiplo en la construcción de un intervalo de confianza (véase la ecuación 2.8) para la media de la población, μ , es Z , un punto porcentual normal estándar. Por lo tanto, un intervalo de estimación para μ es

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

Si establecemos el nivel de confianza en 95%, entonces $Z = 1.96$, el punto más alto .025 (2.5%) de un distribución normal estándar (véase nota al pie 2). Usando la ecuación 2.9 con nuestros resultados de la muestra, se produce el intervalo

$$\begin{aligned} 23.5 \pm 1.96 \frac{10.4}{\sqrt{500}} \\ 23.5 \pm 1.96(.465) \\ 23.5 \pm .91 : (22.6, 24.4) \end{aligned}$$

Así, se dice, con 95% de confianza, que el número medio de millas recorridas para llegar al centro comercial, por la población de compradores, está en algún punto entre 22.6 millas y 24.4 millas. El 95% de confianza se refiere al hecho de que si, digamos, se seleccionaran 100 muestras de tamaño de 500, se calculara la media y la desviación estándar de cada muestra y se construyeran intervalos de estimación, 95 de 100 de los intervalos contendrían, en efecto, la media real de la población. Por lo tanto, es altamente probable que el intervalo específico construido arriba contenga la media de la población en millas viajadas al centro comercial.

Si el tamaño de la muestra hubiera sido pequeño, quizá $n = 20$, entonces, un intervalo de confianza de 95% para μ podría construirse con la ecuación 2.9, usando un punto porcentual t con 19 gl en vez de Z . Puesto que el punto más alto .025 (2.5%) de una distribución t con 19 gl es $t_{.025} = 2.093$ y $S/\sqrt{n} = 10.4/\sqrt{20} = 2.326$, el intervalo de confianza de 95% de la muestra pequeña será considerablemente más amplio, que el intervalo con la muestra grande calculado arriba. Esto es razonable porque hay más incertidumbre asociada con una muestra pequeña. Note que para muestras pequeñas, también hacemos la suposición adicional de que las millas viajadas están distribuidas aproximadamente normal.

Prueba de hipótesis

En muchas situaciones de inferencia, incluyendo la elaboración de pronósticos, el interés está en probar alguna clase de afirmación acerca de la población, en vez de la estimación o el pronóstico de uno de sus parámetros. Este procedimiento se llama *prueba de hipótesis* y es la segunda forma principal de examinar evidencia de la muestra. La prueba de hipótesis implica los siguientes pasos:

- Paso 1.** Formule la hipótesis que se va a probar (llamada hipótesis nula, representada con H_0) y la definición de la hipótesis alternativa (la que se acepta si H_0 se rechaza, representada con H_1).
- Paso 2.** Obtenga una muestra aleatoria de elementos de la población, mídalos y calcule el estadístico de prueba adecuado de la muestra.

TABLA 2-5 Resultados de una prueba de hipótesis

		Acción	
		No rechace H_0	Rechace H_0
Condición natural	H_0 verdadera	Decisión correcta	Error tipo I: probabilidad α
	H_0 Falsa	Error tipo II: probabilidad β	Decisión correcta

Paso 3. Suponga que la hipótesis nula es verdadera, y determine la distribución de muestral, del estadístico de prueba.

Paso 4. Calcule la probabilidad de que un valor del estadístico de prueba, por lo menos tan grande como el observado, se hubiera extraído de esta distribución muestral.

Paso 5. Si esta probabilidad es alta, no rechace la hipótesis nula; si esta probabilidad es baja, se desacredita la hipótesis nula y puede rechazarse con poco margen de error.

Cuando se siguen estos pasos, es posible que haya dos tipos de error, como se indica en la tabla 2-5. Se espera que se tome la decisión correcta con respecto a la hipótesis nula, después de examinar la evidencia de la muestra; no obstante, siempre hay una posibilidad de rechazar una H_0 verdadera y fallar en rechazar una H_0 falsa. Las probabilidades de dichos eventos se conocen como alfa (β) y beta (σ), respectivamente. Alfa también se conoce como *nivel de significancia* de la prueba.

Ejemplo 2.8 Construcción de una prueba de hipótesis

El supervisor de un taller cree que disminuyó el peso medio de las piezas producidas por cierta máquina. Se desea probar la hipótesis de que el peso medio de las piezas producidas por la máquina es todavía de 50 libras, es decir, el peso medio de las piezas en años pasados. Se toma una muestra aleatoria de 100 piezas. Se supone que la desviación estándar de los pesos de las piezas es de 5 libras, sin importar el peso medio, ya que este valor ha permanecido constante en estudios anteriores de las piezas. Usaremos la media de la muestra, como nuestro estadístico de prueba y tomaremos como hipótesis nula el *statu quo*; es decir, el peso medio de la pieza, μ , es de 50 libras. Por lo tanto, si la hipótesis nula H_0 es verdadera, el teorema central del límite indica que la distribución muestral de $\bar{X}$ es, en este caso, normal con media $\mu = 50$ y desviación estándar $\sigma/\sqrt{n} = 5/\sqrt{100} = .5$. Un valor de la media de la muestra, $\bar{X}$, cercano o mayor que 50 apoya la H_0 ; un valor de $\bar{X}$ muy por debajo de 50 es evidencia contra la H_0 , puesto que tal evento es improbable si H_0 es verdadera. Entonces, una regla de decisión sensible es aquella que rechaza H_0 para un valor de $\bar{X}$ muy por debajo de 50. Sin embargo, con esta regla, es probable que cometamos un error tipo I, es decir, que rechacemos una H_0 verdadera. Podemos controlar la probabilidad de rechazar una H_0 verdadera cuando seleccionamos α , el nivel de significancia. Suponga que hacemos a α pequeña —digamos .05 (5%). Entonces, rechazaremos H_0 si el valor observado de nuestro estadístico de prueba, $\bar{X}$, cae más allá del punto más bajo 0.05 (5%) de su distribución muestral de la H_0 . La prueba se muestra en la figura 2-10.

La regla de decisión para esta prueba es como sigue:

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

Suponga que la media de la muestra resulta ser de 49.6 libras. Entonces, podríamos no rechazar H_0 al nivel de 5%. Por otro lado, si el peso medio de la muestra resultara ser de 48.6 libras, rechazaríamos H_0 . Note que para $\bar{X} = 48.6$, la probabilidad de obtener un valor en este extremo (muy por debajo de 50), si H_0 es verdadera, es $P(\bar{X} < 48.6) = P(Z < \frac{48.6 - 50}{.5}) = P(Z < -28) = P(Z > 2.8) = (.5000 - .4974) = .0026$ lo cual es, desde luego, un evento improbable.

En la prueba de hipótesis, siempre existe alguna ambigüedad asociada con la selección de las hipótesis nula y alternativa. En general, la hipótesis nula es el *estatus quo* (condición actual) o hipótesis “sin cambio”. La hipótesis alternativa es la hipótesis de investigación o de cambio.

$H_0: \mu = 50$ libras, (la hipótesis nula)

$H_1: \mu < 50$ libras (la hipótesis alternativa)

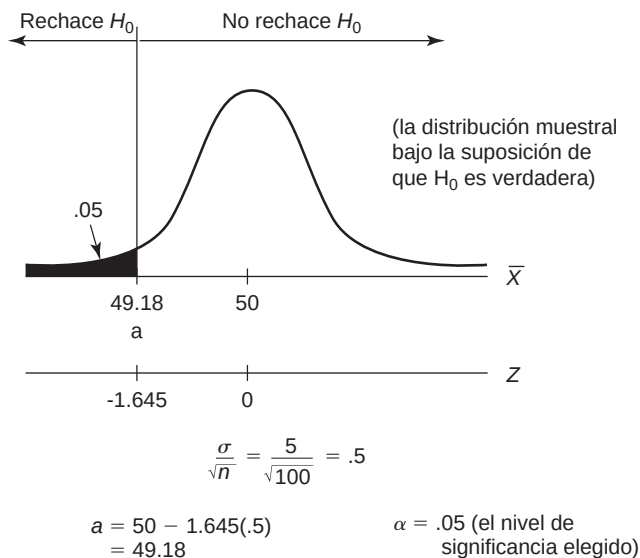


FIGURA 2-10 Prueba de hipótesis para el ejemplo 2.8

Para descubrir si lo que hemos hecho marcó una diferencia o resultó en un cambio del *estatus quo* o de los procedimientos actuales, formulamos la diferencia o el cambio como la hipótesis alternativa. El objetivo es ver si podemos rechazar la hipótesis nula, con una probabilidad pequeña de un error tipo I, a favor de la alternativa.

La prueba de hipótesis que se describe en el ejemplo 2-8 se conoce como prueba *unilateral*. Se sigue esta terminología porque, en este ejemplo, la hipótesis alternativa especifica valores de los parámetros de la población, a un lado del valor especificado en la hipótesis nula. Las pruebas para las cuales la hipótesis alternativa especifica valores tanto más grandes como más pequeños, que el valor con la hipótesis nula, se llaman pruebas *bilaterales*. En este caso, valores del estadístico de prueba mucho más grandes o mucho más pequeños que el valor del parámetro de la población, con H_0 , conducen al rechazo.

Valor p

En el ejemplo 2-8, calculamos la probabilidad de obtener un valor de la media de la muestra tan extremo como $\bar{X} = 48.6$ si la hipótesis nula $H_0: \mu = 50$ es verdadera. La probabilidad encontrada fue de .0026. Esta probabilidad se llama *valor de prueba* o simplemente *valor p* de la prueba. En vez de seleccionar un nivel de significancia, ahora es práctica común calcular e informar el valor p para la prueba; de hecho, los paquetes de software estadístico en general reportan los valores p asociados con estadísticos de prueba. En efecto, los valores p y su uso en pruebas de hipótesis se analizan en los pasos 4 y 5 de nuestro procedimiento de prueba de hipótesis, al inicio de esta sección. Un valor p pequeño significa un fuerte rechazo de la hipótesis nula. El valor p puede considerarse como evidencia a favor (un valor p grande) o en contra (un valor p pequeño) de la hipótesis nula. Los puntos límite comunes para valores p pequeños son .05 y .01.

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

El valor p o *probabilidad de significancia* es la probabilidad de obtener, al menos, un resultado de la muestra tan extremo como el realmente observado, si H_0 fuera verdadera. De forma equivalente, el valor p puede considerarse como la α más pequeña, para la cual el estadístico de prueba observado conduce al rechazo de H_0 .

Ejemplo 2.9 Cálculo de un valor p

Suponga que se desea probar la hipótesis de que la puntuación media de estudiantes en un examen nacional es de 500, contra la hipótesis alternativa de que es menor de 500. Se toma una muestra aleatoria de 15 estudiantes de la población que produce una puntuación media de la muestra de $\bar{X} = 475$. La desviación estándar de la población se estima con la desviación estándar de la muestra, $S = 35$. Suponga que la población de las puntuaciones de los exámenes se distribuye normalmente.

Queremos saber si los puntajes promedio de la prueba son menores que su valor histórico de 500. Tomaremos una decisión con base en la evidencia de la muestra y, en particular, del valor de la media de la muestra. Las hipótesis en competencia son

$$H_0: \mu = 500$$

$$H_1: \mu < 500$$

Si H_0 fuera verdadera, esperaríamos ver un valor del promedio de la muestra cercano a 500 o mayor. Si H_0 fuera falsa (H_1 es verdadera), esperaríamos ver un valor del promedio de la muestra bastante menor de 500. Puesto que el tamaño de la muestra es pequeño, la población es normal y la desviación estándar de la población se estima con la desviación estándar de la muestra, un estadístico de prueba adecuado es el estadístico t dado en la ecuación 2.7, con $n - 1 = 15 - 1 = 14$ gl. Un valor negativo grande del estadístico t es evidencia contra H_0 (que indica que el promedio de la muestra es mucho menor que el promedio hipotético de la población de 500). Primero, realicemos la prueba para un nivel de significancia de $\alpha = .05$. La prueba (regla de decisión) se ilustra en la figura 2-11.

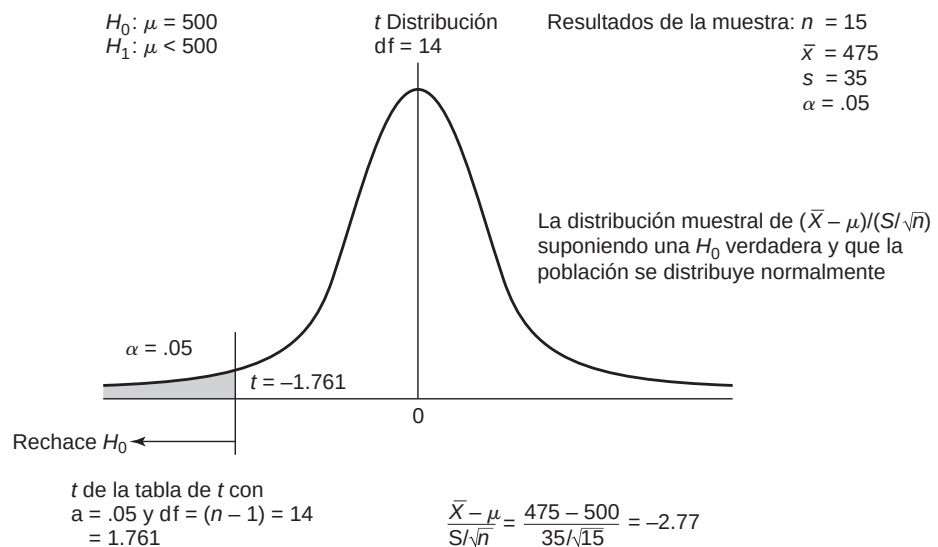


FIGURA 2-11 Prueba de hipótesis para el ejemplo 2.9

Puesto que la regla de decisión es como sigue:

Si observó que $t < -1.76$, rechace H_0 ; $\mu = 500$

y el estadístico t observado es:

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

rechazaríamos la hipótesis nula y concluiríamos que la puntuación media de la prueba de estudiantes en el examen nacional es menor de 500. El *valor p* en este caso es $P(t < -2.77)$. Usando la tabla B-3 con 14 gl y la simetría de la distribución t , esto probablemente esté entre .010 y .005. Nuestro estadístico de prueba observado es muy improbable si H_0 es verdadera. El *valor p* nos indica que deberíamos rechazar la hipótesis nula.

ANÁLISIS DE CORRELACIÓN

En la construcción de modelos estadísticos para realizar pronósticos, a menudo es útil examinar la relación entre dos variables. Dos técnicas, el análisis de correlación y el análisis de regresión, se estudian aquí. Además, en capítulos posteriores se consideran casos especiales de correlación y de regresión. Este énfasis en la correlación y en la regresión se justifica en vista del uso generalizado de dichas técnicas, en todas las clases de aplicaciones para la elaboración de pronósticos.

Diagramas de dispersión

Un estudio de la relación entre variables inicia con el caso más sencillo: el de la relación existente entre dos variables. Suponga que se toman dos mediciones en cada uno de varios objetos. Un analista quiere determinar si una de estas variables medibles, llamada Y , tiende a aumentar o a disminuir cuando la otra variable, llamada X , cambia. Por ejemplo, suponga que se miden tanto la edad como el ingreso de varios individuos, como se indica en la tabla 2-6. ¿Qué se puede decir acerca de la relación entre X y Y ?

En la tabla 2-6, se observa que Y y X tienen una relación evidente. Conforme X aumenta, Y tiende a incrementarse. De esta muestra de cinco individuos, parece que cuanto más vieja se vuelve una persona, más dinero ganará esa persona. Por supuesto, es peligroso llegar a conclusiones sobre la base de una muestra pequeña, un tema que se estudiará más tarde. Incluso con estas observaciones, parece existir una relación evidente entre Y y X .

Estos cinco puntos de datos pueden graficarse sobre una escala bidimensional, con valores de X a lo largo del eje horizontal y valores de Y a lo largo del eje vertical. Una gráfica como ésta se llama *diagrama de dispersión* y se ilustra en la figura 2-12.

TABLA 2-6 Mediciones de ingreso y edad

<i>Persona</i>	<i>Ingreso (en miles) Y</i>	<i>Edad (años) X</i>
1	27.8	22
2	28.5	23
3	30.0	26
4	35.0	27
5	36.4	35

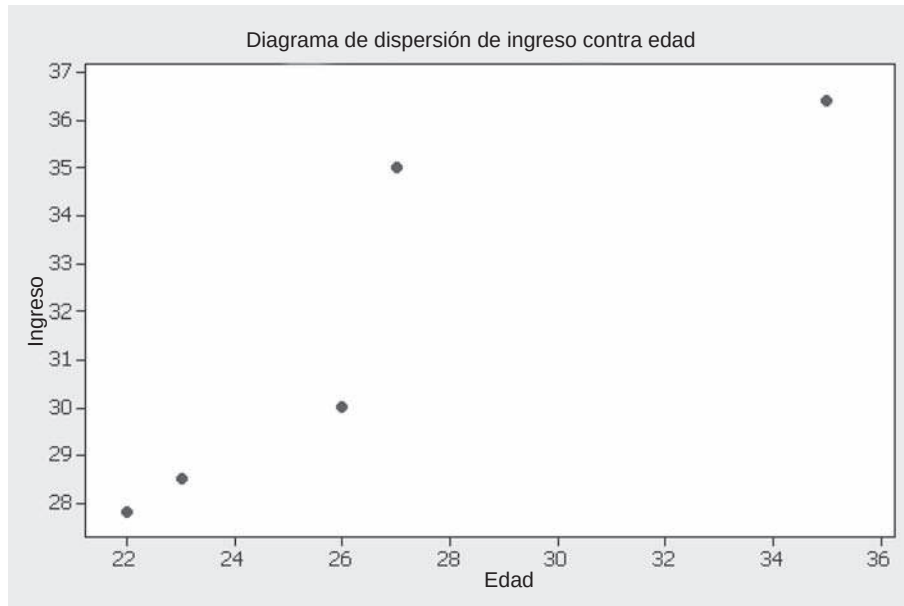


FIGURA 2-12 Diagrama de dispersión para edad e ingreso

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

El diagrama de dispersión ayuda a ilustrar lo que la intuición sugiere cuando se observan a primera vista los datos; a saber: la evidencia de una relación lineal entre Y y X . Esta relación se llama una relación lineal *positiva*, porque conforme X se incrementa, lo hace también Y .

En otras situaciones donde intervienen dos variables, quizá surjan diferentes patrones en el diagrama de dispersión. Considere los diagramas de la figura 2-13.

La figura 2-13(a) sugiere lo que se llama una *relación lineal positiva perfecta*. Conforme X se incrementa, Y aumenta también, y de una forma perfectamente predecible. Es decir, los puntos de datos X y Y parecen formar una línea recta. La figura 2-13(b) sugiere una *relación lineal negativa perfecta*. Conforme X se incrementa, Y disminuye de un modo perfectamente predecible.

Las figuras 2-13(c) y (d) ilustran una *relación lineal imperfecta, positiva y negativa*, respectivamente. Conforme X se incrementa en estos diagramas de dispersión, Y se incrementa (parte c) o disminuye (parte d), aunque no lo hace de un modo perfectamente predecible. Así, Y podría estar ligeramente más arriba o más abajo de “lo esperado”. Es decir, los puntos X - Y no forman una línea recta.

Los diagramas de dispersión en las figuras 2-13(a) a d) ilustran *relaciones lineales*. La relación X - Y , sea perfecta o imperfecta, puede resumirse con una línea recta. En cambio, en la figura 2-13(e) aparece una *relación curva*.

Finalmente, la figura 2-13(f) sugiere que no hay relación de ninguna clase entre las variables X y Y . Conforme X se incrementa, Y no parece aumentar ni disminuir de algún modo predecible. Sobre la base de la evidencia de la muestra que se indica en la figura 2-13(f), podría concluirse que en el mundo que contiene todos los puntos de datos X - Y , no existe relación alguna, ni lineales ni de otra clase, entre las variables X y Y .

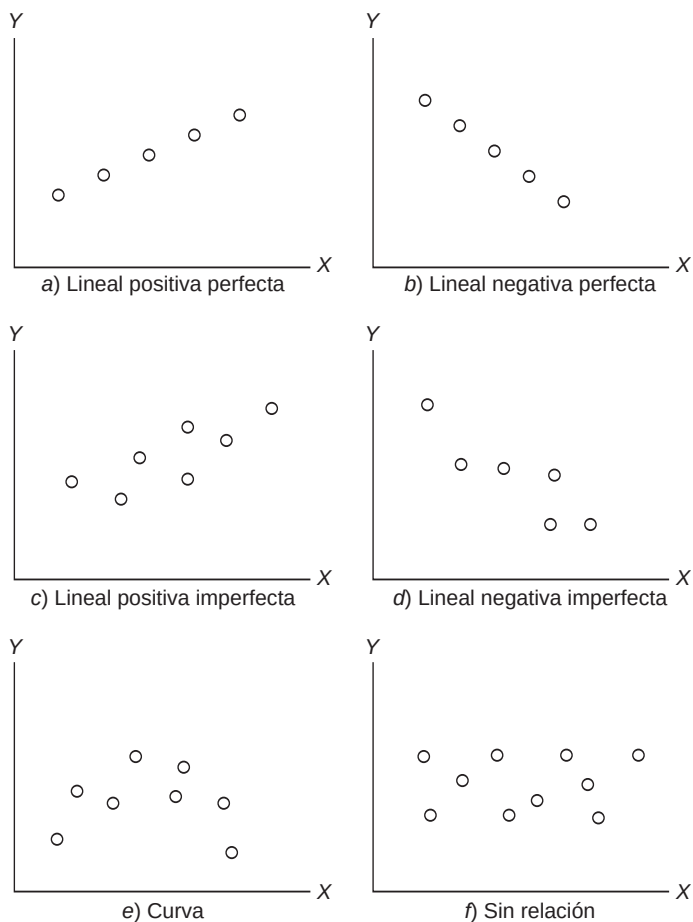


FIGURA 2-13 Patrones para puntos de datos X - Y

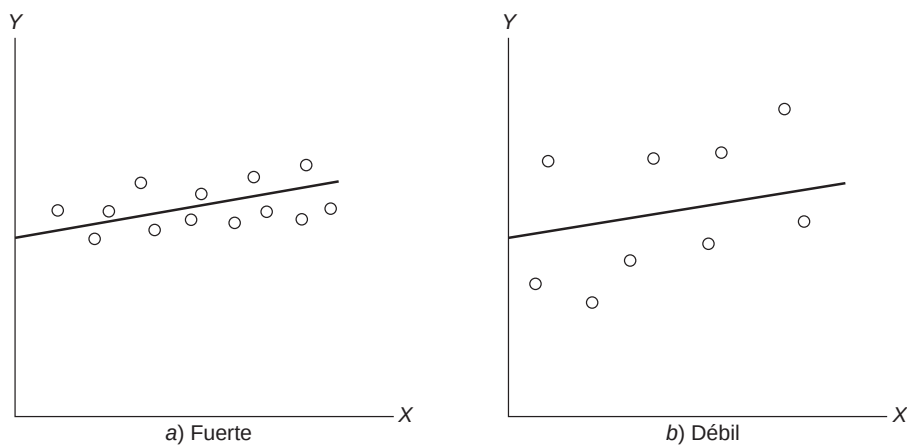


FIGURA 2-14 Asociación lineal fuerte y débil para puntos de datos X - Y

Considere ahora los dos diagramas de dispersión en la figura 2-14. Ambos diagramas de dispersión sugieren relaciones lineales positivas imperfectas entre Y y X . La figura 2-14(a) ilustra una relación fuerte porque los puntos de datos están bastante cerca de la línea recta que pasa a través de ellos. En la figura 2-14(b), los puntos de datos están más lejos de la recta que pasa a través de ellos, sugiriendo así una relación lineal más débil. Más adelante en este capítulo, mostraremos cómo medir la fortaleza de la relación que existe entre dos variables.

Como sugieren los dos diagramas de dispersión de la figura 2-14, a menudo es deseable resumir la relación entre dos variables ajustando una recta a través de los puntos de datos. Usted aprenderá cómo se hace esto en breve; por el momento, piense que una línea recta puede ajustarse a los puntos de un diagrama de dispersión, de modo que resulte un “buen ajuste”. Una pregunta de interés para la elaboración del pronóstico es ¿qué tan rápido se sube o baja la línea recta?

La respuesta a esta cuestión requiere del cálculo de la pendiente de la línea. La pendiente de cualquier recta se define como el cambio de Y asociado con una unidad de incremento en X .

Para resumir, cuando se estudia una relación entre dos variables, uno debe saber primero si la relación es lineal (ilustrada con una línea recta) o no lineal. Si es lineal, uno necesita saber si la relación es positiva o negativa, y qué tan abruptamente sube o desciende la recta que ajusta los puntos de datos. Por último, uno necesita conocer la fortaleza de la relación, es decir, qué tan cerca están los puntos de datos de la línea que mejor se ajusta a ellos.

Coeficiente de correlación

La fortaleza de la relación lineal que existe entre dos variables se mide usando la *correlación* que existe entre ellas. El *coeficiente de correlación* mide la fortaleza de la siguiente manera. 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 . De esta manera, el coeficiente de correlación varía de -1 a 1 inclusive, dependiendo de la cantidad de asociación entre las dos variables que se estén midiendo.

El *coeficiente de correlación* mide el grado en que dos variables están linealmente relacionadas entre sí.

El diagrama de dispersión en la figura 2-13(a) ilustra una situación que produce un coeficiente de correlación de 1 . El diagrama de dispersión en la figura 2-13(b) tiene un coeficiente de correlación de -1 . Las figuras 2-13(e) y (f) ilustran dos variables que no están linealmente relacionadas. Los coeficientes de correlación para estas relaciones son iguales a 0 ; es decir, se presenta una relación no lineal.

Los pronosticadores se interesan tanto en la población como en los datos muestreados. En la *población* que contiene todos los puntos de datos X - Y de interés, hay un coeficiente de correlación cuyo símbolo es ρ , la letra griega ro. Si se extrae una muestra aleatoria de estos puntos de datos X - Y , el coeficiente de correlación para tales datos de muestra se denota con la letra r .

Frecuentemente, X - Y se miden en unidades diferentes, como libras y dólares, unidades vendidas y dólares en ventas, o tasa de desempleo y dólares del PIB. A pesar de tales unidades diferentes para X - Y , es todavía importante medir el grado en que X y Y están relacionadas. Estas mediciones se realizan convirtiendo primero las variables X y Y a unidades estándar o valores de Z .

Después de que las mediciones 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 de interés porque la media de estos valores es el coeficiente de correlación. El cálculo del coeficiente de correlación (esencialmente) como el producto cruzado medio de Z producirá el valor correcto; pero en la mayoría de los casos, el coeficiente de correlación se calcula directamente a partir de los valores X - Y . La ecuación 2-10 muestra cómo se calcula el coeficiente de correlación de la muestra, r , a partir de Z como número de varianza y de las mediciones 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 como la 2-7 facilita el cálculo de r .

Ejemplo 2.10 Cálculo del coeficiente de correlación de una muestra

Si se estudia la relación entre la edad y el ingreso, sería de interés conocer el valor de r para estos datos. Los cálculos requeridos se presentan en la tabla 2-7. Los totales de la tabla 2-7 se sustituyen en la ecuación 2.10, con lo cual se obtiene:

$$\begin{aligned} r &= \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \\ &= \frac{5(4,266.1) - (133)(157.7)}{\sqrt{5(3,643) - (133)^2} \sqrt{5(5,035.05) - (157.7)^2}} \\ &= \frac{21,330.5 - 20,974.1}{\sqrt{526} \sqrt{305.96}} = \frac{356.4}{(22.93)(17.49)} = .89 \end{aligned}$$

Se observa que el coeficiente de correlación de la muestra confirma lo que se ilustró en la figura 2-12. El valor de r es positivo, lo cual sugiere una relación lineal positiva entre la edad y el ingreso. También, en una escala de 0 a 1, el valor de r es bastante alto (0.89). Este resultado sugiere una relación lineal fuerte en vez de una débil. La pregunta restante es si la combinación del tamaño de la muestra y el coeficiente de correlación es lo suficientemente fuerte para hacer afirmaciones significativas acerca de la población, de la cual se extrajeron los valores de los datos.

Ahora deberían hacerse dos observaciones importantes acerca de la correlación. Primero, siempre debe tenerse en mente que se mide la *correlación*, no la *causalidad*. Puede ser perfectamente válido decir que dos variables están relacionadas sobre la base de un coeficiente de

TABLA 2-7 Cálculos para la correlación entre edad e ingreso del ejemplo 2.10

Persona	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

correlación alto. Quizá sea válido o no decir que una variable es causante del cambio en la otra; esto es una cuestión para el juicio del analista. Por ejemplo, puede ser verdad que el volumen de ventas de una tienda local, en una área escasamente poblada, está altamente correlacionado con el precio de mercado promedio de las acciones en Nueva York. Después de examinar una muestra grande de estas dos variables, podría concluirse que existe esa correlación alta. Probablemente no sea verdad que una de tales variables esté causando el cambio en la otra. De hecho, los cambios de ambas variables quizá sean causados por un tercer factor, el estado general de la economía. Los políticos, los anunciantes y otros a menudo cometen el error de suponer una causa con base en la correlación.

Segundo, advierta que el coeficiente de correlación mide una relación *lineal* entre dos variables. En el caso en que el coeficiente de correlación es bajo, se concluiría que las dos variables no están estrechamente relacionadas de forma lineal. Tal vez estén estrechamente relacionadas de un modo no lineal o curvado. Así, un coeficiente de correlación bajo no significa que las dos variables no estén relacionadas, solo que parece que no existe una relación lineal o en línea recta.

AJUSTE DE UNA LÍNEA RECTA

Como se mencionó antes, con frecuencia es de interés ajustar una línea recta a través de un conjunto de puntos de datos X - Y , de manera tal que resulte un “buen ajuste”. Como se mostrará en los capítulos posteriores, esa línea bien ajustada puede usarse para pronosticar valores desconocidos de Y , la variable de interés.

En la práctica, se usa un procedimiento matemático bien definido para calcular tanto la intersección con el eje Y (ordenada al origen), como la pendiente de la línea recta mejor ajustada. El enfoque más común para determinar esa línea mejor ajustada se llama *método de los mínimos cuadrados*. Con este procedimiento, se construye una línea que minimiza la suma de los cuadrados de las distancias entre los puntos de datos y la recta, medidos en la dirección (Y) vertical.

El método de los mínimos cuadrados selecciona los valores de la pendiente y de la intersección con el eje Y que minimizan la suma de errores al cuadrado (distancias), SSE , entre los valores de Y y la recta.

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

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

Usando el cálculo, se pueden derivar expresiones algebraicas específicas para los valores de los mínimos cuadrados. Particularmente:

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

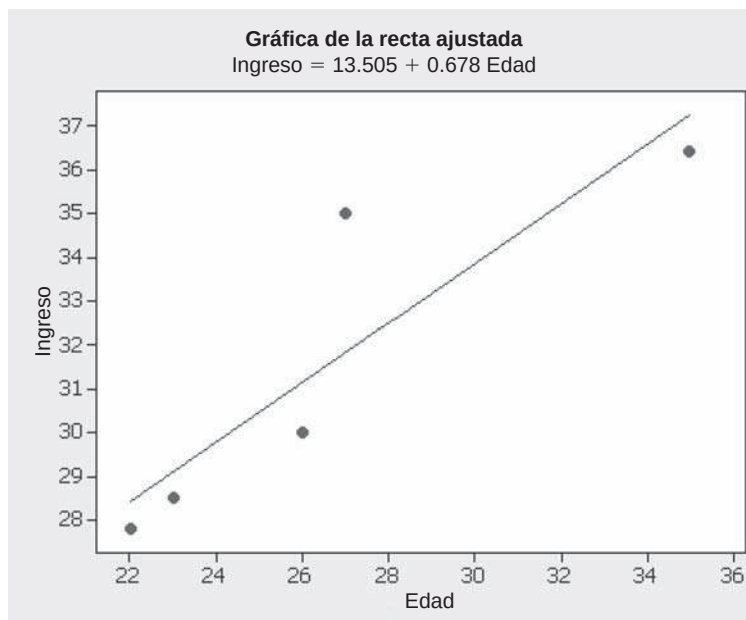
$$b_0 = \bar{Y} - b_1 \bar{X} = \frac{\sum Y}{n} - \frac{b_1 \sum X}{n} \quad (2.13)$$

donde

b_1 = la pendiente de la recta

b_0 = la intersección con el eje Y

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

**FIGURA 2-15** Recta ajustada del ejemplo 2-11**Ejemplo 2.11 Cálculo de una recta ajustada usando el método de los mínimos cuadrados**

El ejemplo 2-10 sugiere una relación lineal positiva fuerte entre edad e ingreso. Sustituyendo los totales de la tabla 2-7 en las ecuaciones 2.12 y 2.13, la ecuación de una recta que mejor se ajuste a los puntos se calcula como:

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

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

La recta que mejor se ajusta a los datos, $\hat{Y} = 13.505 + .678 X$, se ilustra en la figura 2-15.

La ecuación calculada en el ejemplo 2-11, junto con otros valores que pueden calcularse a partir de los datos de la muestra, es de enorme utilidad para los gerentes en el pronóstico de valores futuros de una variable importante, y en la evaluación anticipada de qué tan precisos serían tales pronósticos. En un capítulo posterior, usted aprenderá cómo extraer gran parte de información a partir de los datos de la muestra, y cómo usarla para hacer pronósticos con el análisis de regresión.

El coeficiente de la pendiente de cuadrados mínimos está relacionado con el coeficiente de correlación de la muestra. Específicamente:

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

Como resultado, b_1 y r son proporcionales entre sí y tienen el mismo signo.

Ejemplo 2.12 Demostración de la relación entre el coeficiente de correlación y el coeficiente de la pendiente

Usando los resultados de la tabla 2-7 y los ejemplos 2-10 y 2-11, verificamos numéricamente la ecuación 2.14. Usando la ecuación 2.14, tenemos

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

Este valor, considerando el error por redondeo, coincide con el valor del coeficiente de la pendiente calculado directamente en el ejemplo 2.11, usando la ecuación 2.12.

Ejemplo 2.13

Suponga que el director general de una empresa constructora grande sospecha que los gastos estimados de los proyectos de construcción de su compañía no están muy cercanos de los gastos reales al final del proyecto. Los datos que se presentan en la tabla 2-8 se recopilaron de unos cuantos proyectos anteriores, para analizar la relación entre los costos estimados y los costos reales. Puesto que tiene a su disposición Minitab, se analizan los datos usando este programa.

La figura 2-16 ilustra que la correlación entre los costos de construcción estimados y reales, con base en los datos de la muestra, es $r = .912$. El director general de la compañía está sorprendido por saber que éste sea alto.

La figura 2-17 presenta los datos graficados como un diagrama de dispersión y la recta que mejor se ajusta a estos datos: $\hat{Y} = .683 + .922X$. El director general puede ahora pronosticar el costo de construcción real (Y), después de que se preparó el estimado para el proyecto (X).

TABLA 2-8 Costos estimados y reales del proyecto de construcción del ejemplo 2.13

<i>Renglón</i>	<i>Actual</i>	<i>Estimado</i>
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

Correlations: ActualCost, EstimatedCost

Pearson correlation of ActualCost and EstimatedCost = 0.912

Regression Analysis: ActualCost versus EstimatedCost

The regression equation is

ActualCost = 0.683 + 0.9223 EstimatedCost

FIGURA 2-16 Resultado de Minitab para el ejemplo 2.13

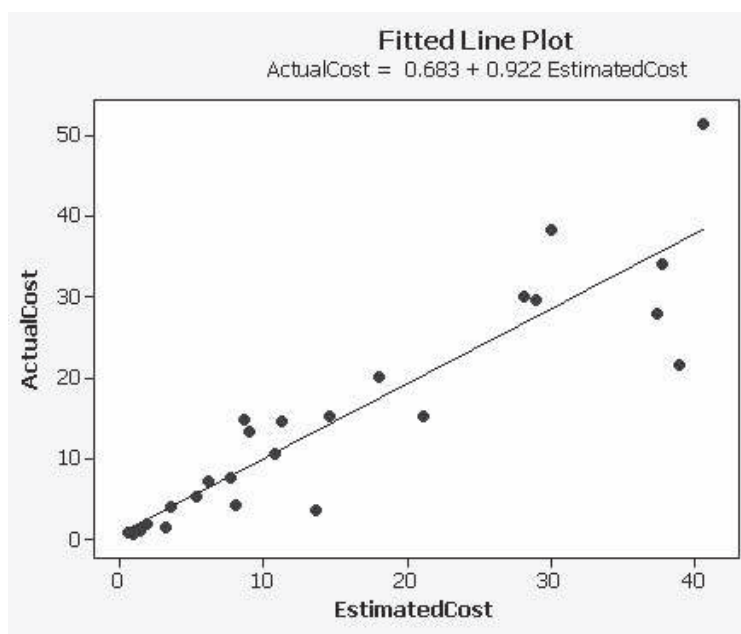


FIGURA 2-17 Gráfica de la línea ajustada del ejemplo 2.13

EVALUACIÓN DE LA NORMALIDAD

Muchas técnicas estadísticas, incluyendo algunas de las que se usan en la elaboración de pronósticos, requieren la suposición de que un conjunto de datos sigue una distribución normal. Por tal motivo, se han desarrollado técnicas estadísticas que prueban la hipótesis de que un conjunto de datos muestrales fue extraído de una población normalmente distribuida.

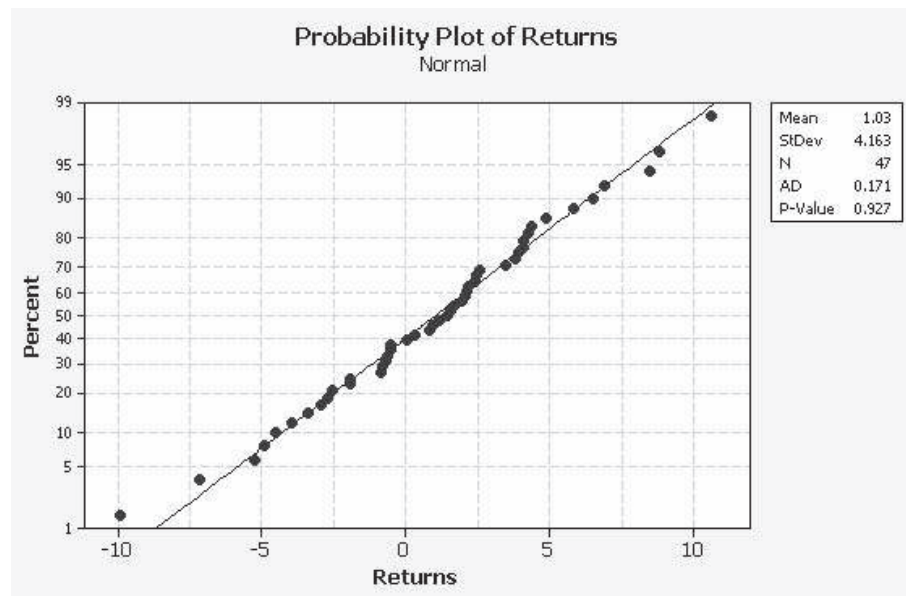
Considere las tasas de rendimiento mensual del índice bursátil Standard & Poor's 500, mostradas en la tabla 2-9. ¿Se puede suponer que estos datos siguen una distribución normal? La respuesta, obtenida con Minitab, está contenida en la figura 2-18.

La recta trazada en la figura 2-18 muestra qué puntos de una curva normal perfecta se verían como si se hubieran graficado usando esta escala especial.⁴ Como se indica, los datos de la tabla 2-9 están muy cerca de esta línea, lo cual sugiere un buen ajuste entre los datos de S&P y una distribución normal.

⁴Están disponibles otras gráficas de probabilidad normal. Una de ellas, llamada *gráfica de puntuaciones normal*, se usa frecuentemente. En todas estas gráficas, normalmente se indica si los datos graficados están cerca de una recta.

TABLA 2-9 Tasas mensuales de rendimiento (%) del índice Standard & Poor's 500

	Año			
	1	2	3	4
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
Mayo	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
Sept.	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 de probabilidad normal para tasas de rendimiento mensuales del índice Standard & Poor's 500**

La prueba de normalidad preestablecida de Minitab es la prueba Anderson-Darling, cuyos resultados se ilustran en la figura 2-18. Los detalles de esta prueba no nos interesan por el momento; sin embargo, note el *valor p* (indicado como P-Value en la parte superior derecha de la figura) de .927. En este caso, el *valor p* grande sugiere que los datos de la muestra de S&P son consistentes con la hipótesis nula (en efecto, no es inusual si la hipótesis nula es verdadera). El rechazo de la hipótesis nula de la normalidad daría como resultado casi con seguridad un error. Por lo tanto, la hipótesis nula no debería rechazarse y la suposición de que los datos de S&P siguen una distribución normal puede hacerse con seguridad.

APLICACIÓN EN LA ADMINISTRACIÓN

Muchos de los conceptos en este capítulo de repaso pueden considerarse como los antecedentes necesarios para la comprensión de técnicas más avanzadas para elaborar pronósticos, que se encuentran a lo largo del resto del libro. Sin embargo, los conceptos en este capítulo también tienen valor por sí mismos en muchas aplicaciones estadísticas. Aunque algunas de dichas aplicaciones quizá no caigan lógicamente dentro del título de “pronósticos”, implican el uso de datos recopilados para contestar preguntas acerca de incertidumbres en la operación de negocios, en especial de los resultados futuros inciertos.

Los procedimientos estadísticos descriptivos mencionados al principio del capítulo son ampliamente usados siempre que grandes cantidades de datos deban describirse de forma convincente, de modo que puedan utilizarse en el proceso de toma de decisiones. Sería casi imposible pensar en una sola área que implique mediciones numéricas, en las cuales los conjuntos de datos no estén resumidos rutinariamente, usando estadísticos descriptivos. Este hecho se aplica específicamente a la media, que por lo general se conoce como el “promedio”, y —en menor medida— a la desviación estándar. Los promedios son cosa común y se han utilizado por muchos años para tener medidas centrales de conjuntos ordenados de datos. El énfasis reciente en la calidad requiere una comprensión de la variación y, por lo tanto, medidas de dispersión, tales como la desviación estándar, están apareciendo con cada vez mayor frecuencia en la práctica de negocios.

Las distribuciones binomial y normal son buenos ejemplos de distribuciones teóricas que sirven como modelos para muchas situaciones de la vida real. Como tales, se han utilizado abundantemente en aplicaciones, incluyendo la elaboración de pronósticos.

La estimación y la prueba de hipótesis son el principal fundamento de la inferencia estadística básica. La elaboración de pronósticos, o estimación, acerca de los valores de interés de población, a partir de una muestra aleatoria, se emplea comúnmente siempre que las restricciones de tiempo y de costos descartan un examen de todos los elementos en consideración. El muestreo está muy extendido en auditorías. La prueba de hipótesis está relacionada con dos enunciados en competencia, acerca de parámetros de la población, de modo que se emplea a menudo para contestar preguntas como éstas:

- ¿Cómo se compara la media de ventas de este periodo con la del periodo anterior?
- ¿Cómo se compara la razón activos/pasivos media de empresas exitosas con la de empresas malogradas?
- ¿La producción media por hora de este proceso de producción es menor de lo que había sido?

La correlación se utiliza ampliamente para examinar las relaciones entre pares de variables numéricas. Como veremos en capítulos posteriores, estas relaciones son de gran importancia en la elaboración de pronósticos, porque el pronóstico de una variable a menudo requiere los valores de las variables relevantes relacionadas. Tanto los análisis de regresión simple como múltiple son técnicas de pronósticos que caen dentro de esta categoría.

Glosario

Distribución binomial. La distribución binomial es una distribución de probabilidad discreta, que describe la probabilidad de X éxitos en n ensayos independientes de un experimento binomial.

Variable aleatoria continua. Una variable aleatoria continua puede tomar cualquier valor dentro de un rango específico. Estos resultados se representan numéricamente con intervalos de valores.

Coefficiente de correlación. Mide el grado en que dos variables están linealmente relacionadas entre sí.

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

Variable aleatoria discreta. Tipo de variable que sólo puede tomar ciertos valores de un conjunto predeterminado. Estos resultados a menudo se representan numéricamente con enteros.

Valor esperado. El valor medio de la variable a lo largo de muchos ensayos u observaciones.

Intervalo de estimación. Intervalo numérico dentro del cual es probable que se encuentre el parámetro de la población.

Método de los mínimos cuadrados. Se usa para calcular la ecuación de una línea recta que minimiza la suma de los cuadrados de las distancias entre los puntos de datos X - Y y la línea, medidos en la dirección vertical (Y).

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

Estimación puntual. Estimación de un parámetro poblacional mediante un valor único.

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.

Distribución muestral. Es el arreglo de todos los valores posibles de un estadístico muestral, que puede obtenerse a partir de una población para un tamaño de muestra determinado.

Diagrama de dispersión. Un diagrama de dispersión es una gráfica de puntos de datos X - Y sobre una gráfica bidimensional.

Fórmulas clave

Media de la muestra

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

Desviación estándar de la muestra

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

Varianza de la muestra

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

Valor esperado

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

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

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

Estadístico 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)

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

Intervalo para la media de la población (muestra pequeña, población normal)

$$\bar{X} \pm t \frac{S}{\sqrt{n}}$$

Coefficiente de correlación

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

Método de mínimos cuadrados

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

Ecuación de la pendiente de regresión ajustada

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

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

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

Ecuación alternativa de la pendiente de la ecuación de regresión ajustada

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

Problemas

1. Dick Hoover, propietario de Modern Office Equipment, está preocupado por los costos de fletes (de envío) y por los costos de oficina en que incurren pedidos pequeños. En un esfuerzo por reducir gastos en esta área, decide aplicar una

política de descuentos en pedidos de más de \$40, con la esperanza de que esto motive a los clientes a consolidar varios pedidos pequeños en pedidos más grandes. Los siguientes datos presentan las cantidades por transacción para una muestra de 28 clientes.

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

- Calcule la media de la muestra.
 - Calcule la desviación estándar de la muestra.
 - Calcule la varianza de la muestra
 - Si la política tiene éxito, ¿la media de la distribución aumentará, disminuirá o no se verá afectada?
 - Si la política tiene éxito, ¿la desviación estándar de la distribución aumentará, disminuirá o no se verá afectada?
 - Dados los datos anteriores, pronostique la cantidad del siguiente pedido de parte de los clientes.
2. Sandy James cree que los precios de las viviendas se han estabilizado en los recientes meses. Para convencer a su jefe, ella tiene pensado comparar los precios actuales con los precios del año pasado. Ella reúne 12 precios de viviendas de anuncios clasificados:
- 125,900 253,000 207,500 146,950 121,450 135,450
175,000 200,000 210,950 166,700 185,000 191,950
- Luego, calcula la media y la desviación estándar de los precios que encontró. ¿Cuáles son estos dos valores?
3. Una compañía constructora grande planea idear una manera útil de visualizar las ganancias típicas de trabajos obtenidos por licitación. Debido a que los trabajos varían significativamente en tamaño y en la cantidad final de subasta exitosa, la compañía ha decidido 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 y también 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

- Calcule un estimado de las ganancias porcentuales medias para la población de trabajos al incluir todos los trabajos potenciales.
- Construya un intervalo de confianza de 95% para las ganancias porcentuales medias de la población de trabajos, usando un argumento de muestra grande.
- Construya un intervalo de confianza de 95%, para las ganancias porcentuales medias para la población de trabajos, suponiendo que 30 es un tamaño de muestra pequeño. ¿Cuál es la suposición adicional que usted necesita hacer en este caso?
- Compare los dos intervalos en los incisos b) y c). Explique por qué un tamaño de muestra de 30 a menudo se toma como límite entre una muestra grande y una pequeña.

4. De los datos en una muestra grande de transacciones de ventas, el propietario de un negocio pequeño informa que un intervalo de confianza de 95% para la utilidad media por transacción, μ , es (23.41, 102.59). Use estos datos para determinar
 - a) Una estimación puntual (mejor conjetura) de la media, μ , y su margen de error de 95%.
 - b) Un intervalo de confianza de 90% para la media, μ .
5. Queremos predecir si el número medio de días de ausencia por año se ha incrementado para una muestra grande de trabajadores. Se sabía que hace un año la media era de 12.1. Una muestra reciente de 100 empleados revela una media muestral de 13.5, con una desviación estándar de la muestra de 1.7 días. Pruebe con nivel de significancia de .05, para determinar si la media de la población se ha incrementado, o si la diferencia entre 13.5 y 12.1 simplemente representa un error de muestreo.
6. New Horizons Airlines quiere predecir el número medio de asientos vacíos por vuelo a Alemania para el siguiente año. Para realizar tal pronóstico, se seleccionan al azar los registros de 49 vuelos de los archivos del año pasado, y se anota el número de asientos desocupados para cada vuelo. La media y la desviación estándar de la muestra son de 8.1 asientos y 5.7 asientos, respectivamente. Desarrolle una estimación puntual y por intervalo de 95% del número medio de asientos desocupados por vuelo, durante el año pasado. Pronostique la media de asientos vacíos por vuelo a Alemania para el siguiente año. Analice la exactitud de su pronóstico.
7. Por un periodo de varios años, un dentífrico ha recibido una puntuación media de 5.9, en una escala de 7 puntos, en cuanto a la satisfacción general del cliente con el producto. Debido a un cambio no anunciado en el producto, existe la preocupación de que quizás haya cambiado la satisfacción del cliente. Suponga que las puntuaciones de satisfacción para una muestra de 60 clientes tienen una media de 5.60 y una desviación estándar de .87. ¿Indican estos datos que la puntuación media de satisfacción es diferente de 5.9? Pruebe con $\alpha = .05$. ¿Cuál es el valor p de la prueba?
8. La gerente de una tienda de yogurt congelado afirma que una porción de tamaño mediano contiene un promedio de más de 4 onzas de producto. A partir de una muestra aleatoria de 14 raciones, se obtiene una media de 4.31 onzas y una desviación estándar de .52 onzas. Pruebe, con $\alpha = .05$, la afirmación de la gerente. Calcule el valor p para la prueba. Suponga que la distribución de peso por ración es normal.
9. Con base en experiencias pasadas, la California Power Company predice que la media del uso residencial de electricidad por hogar será de 700 kwh para el próximo enero. En enero, la compañía selecciona una muestra aleatoria simple de 50 hogares, y calcula una media y una desviación estándar de 715 y 50, respectivamente. Utilice un nivel de significancia de 0.05, para determinar si el pronóstico de California Power es razonable. Calcule e interprete el valor p de esta prueba.
10. Expertos demógrafos indican que el tamaño de la familia ha disminuido en los años recientes. Hace 10 años, el tamaño promedio de la familia era de 2.9. Considere la población de 200 tamaños de familia mostrados en la tabla P-10. Seleccione al azar una muestra de 30 tamaños de familia y pruebe la hipótesis de que el tamaño promedio de la familia no ha cambiado en los últimos 10 años.
11. A James Dobbins, supervisor de mantenimiento en Atlanta Transit Authority, le gustaría determinar si hay una relación positiva entre el costo de mantenimiento anual de un autobús y su antigüedad. Si existe tal relación, James cree que puede predecir mejor el presupuesto anual de mantenimiento de los autobuses. Él reúne los datos que se presentan en la tabla P-11.
 - a) Grafique un diagrama de dispersión.
 - b) ¿Qué tipo de relación hay entre estas dos variables?
 - c) Calcule el coeficiente de correlación.

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 mantenimiento (\$)</i>	<i>Antigüedad</i>
	<i>Y</i>	<i>(años) 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 Y</i>	<i>Espacio en estantes 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

12. Ana Sheehan es la gerente de la cadena de supermercados Spendwise. A ella le gustaría predecir las ventas (ejemplares por semana) de libros de pasta blanda (rústica), con base en la cantidad de espacio de exhibición en los anaqueles (en pies). Anna reúne datos para una muestra de 11 semanas, como se presenta en la tabla P-12.
- Grafique un diagrama de dispersión.
 - ¿Qué tipo de relación hay entre estas dos variables?
 - Calcule el coeficiente de correlación. Utilice el método de 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 en estantes (es decir, $X = 5.2$).
13. Considere la población de 200 observaciones semanales presentadas en la tabla P-13. La variable independiente X es la temperatura semanal promedio en Spokane, Washington. La variable dependiente Y es el número de acciones de Sunshine Mining Stock, con las que se hicieron operaciones en la bolsa de valores de Spokane en una semana. Elija datos al azar para 16 semanas y calcule el coeficiente de correlación. (*Sugerencia:* Asegúrese de que su muestra se obtiene aleatoriamente de la población). Luego determine la recta de los mínimos cuadrados y pronostique Y para una temperatura promedio semanal de 63.
14. Un inversionista inmobiliario recolecta los siguientes datos de una muestra aleatoria de apartamentos en el lado oeste de College Station, Texas.
- Grafique los datos como un diagrama de dispersión donde Y = renta y X = tamaño.
 - Determine la ecuación de la recta ajustada que relacione la renta con el tamaño.
 - ¿Cuál es el incremento estimado en renta, para un pie adicional de espacio?
 - Pronostique la renta mensual para un departamento de 750 pies cuadrados.

Renta (\$ mensuales)	Tamaño (pies ²)	Renta (\$ mensuales)	Tamaño (pies ²)
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>	<i>OBS.</i>	<i>Y</i>	<i>X</i>	<i>OBS.</i>	<i>Y</i>	<i>X</i>	<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

15. Abbot & Sons necesita pronosticar la edad media μ de sus trabajadores contratados por hora. Se obtiene una muestra aleatoria de los archivos personales, con los resultados de abajo. Prepare una estimación puntual y un intervalo de confianza de 98%, para la edad media de los trabajadores. Pruebe la hipótesis $H_0: \mu = 44$ contra $H_1: \mu \neq 44$ al nivel de 2%. ¿Los resultados de la prueba de hipótesis son consistentes con el intervalo de confianza para μ ? ¿Usted esperaría que fueran?

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

16. En cada una de las siguientes situaciones, establezca una hipótesis nula adecuada, H_0 , y una hipótesis alternativa, H_1 . Identifique el parámetro que utilizará para establecer la hipótesis.
- Datos de Census Bureau indican que el ingreso familiar medio, en el área que atiende un centro comercial, es de \$63,700 anuales. Una empresa de investigación de mercados encuesta a los compradores del centro comercial, para averiguar si el ingreso familiar medio de los clientes es mayor que el de la población general.
 - El año pasado, al departamento de bomberos local le tomó un promedio de 4.3 minutos dar respuesta a las llamadas. ¿Los datos de este año indican un promedio diferente en el tiempo de respuesta?
 - Se anuncia que el área media de varios miles de apartamentos en un nuevo desarrollo es de 1,300 pies cuadrados. Un grupo de arrendatarios cree que los apartamentos son más pequeños que lo que dice la publicidad. Ellos contrataron a un ingeniero para medir una muestra de apartamentos para probar sus sospechas.
17. Un inversionista con un portafolio de inversiones sustancial demandó a su agente bursátil y a su empresa de corretaje, a causa de que la falta de diversificación en su portafolio le ocasionó un desempeño deficiente. Las tasas de rendimiento para los 39 meses que la cuenta fue administrada por tal agente produjeron estos estadísticos de resumen: $\bar{r} = -1.10\%$, $S = 5.99\%$. Considere los 39 rendimientos mensuales como una muestra aleatoria de la población de rendimientos que la correduría generaría, si administrara la cuenta para siempre. Usando los resultados de la muestra, construya un intervalo de confianza de 95%, para la media de rendimientos mensuales del mercado. Permita que S&P 500 represente al mercado y suponga que el rendimiento medio de S&P 500 para el mismo periodo es de 0.94%. ¿Éste un valor realista para la media de la población de cuentas del cliente? Explique.
18. La tabla P-18 proporciona sueldos mensuales en dólares (SUELDO) y tiempo de servicio (TDS) en meses, en un momento específico, para 16 mujeres que tienen trabajos de servicio al cliente en bancos de Texas.

TABLA P-18

<i>Empleado</i>	<i>SUELDO</i>	<i>TDS</i>
1	389	94
2	377	60
3	315	45
4	316	39
5	324	20
6	403	76
7	390	106
8	441	62
9	418	68
10	413	55
11	396	27
12	353	58
13	483	139
14	437	98
15	341	70
16	476	60

- Grafique los datos de la tabla P-18 como un diagrama de dispersión, con SUELDOS a lo largo del eje vertical (Y) y TDS en el eje horizontal (X).
- Calcule el coeficiente de correlación de la muestra, r . Usando el signo y la magnitud de r , describa la naturaleza de la asociación lineal entre SUELDOS y TDS. ¿Puede pensar en otras variables que afectarían los sueldos semanalmente, además del tiempo de servicio?
- Calcule la recta ajustada que pueda usarse para predecir SUELDOS a partir de TDS. Si el TDS es de 80 meses, ¿cuál será el SUELDO pronosticado?

CASOS

CASO 2-1 ALCAM ELECTRONICS

Jarrick Tilby se tituló recientemente en administración de negocios en una pequeña universidad y comenzó a trabajar a Alcam Electronics, un fabricante de varios componentes electrónicos para la industria. Después de unas semanas en el trabajo, la gerente y propietaria de Alcam, McKennah Labrum, lo llamó a su oficina y le pidió investigar una cuestión respecto a cierto transistor fabricado por Alcam, porque una compañía de TV grande estaba interesada en una compra significativa.

McKennah quería pronosticar el tiempo de vida promedio de este tipo de transistor, un asunto de gran preocupación para la compañía de TV. Las unidades que actualmente están en existencia podrían ser representativas de aquellas que se producirían durante la vigencia del nuevo contrato si es que éste fuera aceptado.

Jarrick decidió tomar una muestra aleatoria de los transistores en cuestión y formuló un plan para llevar a cabo esa tarea. Numeró los contenedores de almacenamiento de los transistores, les puso números aleatorios y muestreó todos los transistores en cada contenedor seleccionado para la muestra. Puesto que cada contenedor incluía aproximadamente de 20 transistores, seleccionó 10 números al azar, lo cual le dio un tamaño de muestra final de 205 transistores. Debido a que había seleccionado 10 de 55 contenedores, creía que tenía una buena muestra representativa y podría usar los resultados de esta muestra para generalizar a la población completa de transistores en el inventario, así como a las unidades aún habrían de ser fabricadas con el mismo proceso.

Entonces, Jerrick consideró la cuestión del tiempo de vida promedio de las unidades. Debido a que el tiempo de vida de la unidad puede extenderse varios años, se dio cuenta de que ninguna de las unidades muestreadas podría probarse, si se deseaba una res-

puesta oportuna. Por lo tanto, decidió contactar a varios usuarios de dichos componentes para determinar si había registros acerca de la vida de los componentes. Por fortuna, encontró tres compañías que han usado el transistor en el pasado y que tenían registros limitados acerca del tiempo de vida del componente. En total, recibió datos sobre 38 transistores cuyos tiempos de falla eran conocidos. Puesto que estos transistores se fabricaron usando el proceso actual, Jarrick pensó que los resultados de esta muestra podrían utilizarse para hacer inferencias acerca de las unidades en el inventario y aquellas que se fabricarían.

Los resultados de los cálculos que Jarrick ejecutó con su muestra de datos del tiempo de vida son:

$$n = 38$$

$$\text{Tiempo de vida promedio } \bar{X} = 4,805 \text{ horas}$$

$$\text{Desviación estándar de tiempos de vida } S = 675 \text{ horas}$$

Después de encontrar que el tiempo de vida promedio de la muestra era tan sólo de 4,805 horas, Jarrick estaba preocupado porque sabía que otro proveedor de componentes estaba garantizando un tiempo de vida promedio de 5,000 horas. Aunque su promedio de la muestra estaba un poco por debajo de 5,000 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 Alcam fuera inferior a la del otro proveedor. Decidió probar la hipótesis de que el tiempo de vida promedio de todos los transistores era de 5,000 horas contra la alternativa de que fuera menor. En seguida se presentan los cálculos que realizó usando $\alpha = .01$:

$$H_0: \mu = 5,000$$

$$H_1: \mu < 5,000$$

Si $S = 675$, entonces el punto de la regla de decisión es

$$5,000 - 2.33 \frac{675}{\sqrt{38}} = 4,744.9$$

y la regla de decisión es como sigue:

$$\text{Si } \bar{X} < 4,744.9, \text{ rechace } H_0$$

Puesto que la media de la muestra (4,805) no estaba por debajo del punto de la regla de decisión para el rechazo (4,744.9), Jarrick no rechazó la hipótesis de que el tiempo de vida medio de todos los componentes era igual a 5,000 horas. Pensó que esto serían

buenas noticias para McKennah Labrum e incluyó un resumen de sus hallazgos en su informe final. Pocos días después, le entregó a McKennah su informe verbal y escrito. Ella lo llamó a su oficina para felicitarlo por su buen trabajo y compartir con él una inquietud que tenía en relación con sus hallazgos. Ella dijo: “Estoy preocupada por el nivel de significancia tan bajo de tu prueba de hipótesis. Tan sólo tomaste una probabilidad de 1% de rechazar la hipótesis nula si fuera verdadera. Esto me suena muy conservador. Estoy alarmada de firmar el contrato y luego entrarme de que nuestro nivel de calidad no satisface la especificación deseada de 5,000 horas”.

PREGUNTA

1. ¿Cómo respondería ante el 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 de ventas mensual (véase caso 1-1). Como primer paso, John recolecta datos de ventas mensuales de los años 1998 a 2005, como se expone en la tabla 2-10.

Después, John calcula el valor promedio de ventas mensuales para cada año (es decir, suma los 12 valores

para 1998 y los divide entre 12). John también calcula la desviación estándar para los 12 valores mensuales de cada año. Los resultados se presentan en la tabla 2-11. John también decide construir una gráfica de series de tiempo, que se ilustra en la figura 2-19. Grafica los valores medios de ventas mensuales sobre el eje Y y el tiempo sobre el eje X .

TABLA 2-10		Datos de ventas mensuales de Mr. Tux						
	1998	1999	2000	2001	2002	2003	2004	2005
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
Mayo	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
Sept.	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

TABLA 2-11 Valores mensuales promedio de ventas de Mr. Tux		
Año	Promedio (media) (\$)	Desviación estándar (\$)
1998	22,013	13,165
1999	44,603	35,290
2000	64,841	47,217
2001	103,610	57,197
2002	141,381	70,625
2003	169,432	63,376
2004	213,866	96,387
2005	247,231	99,153

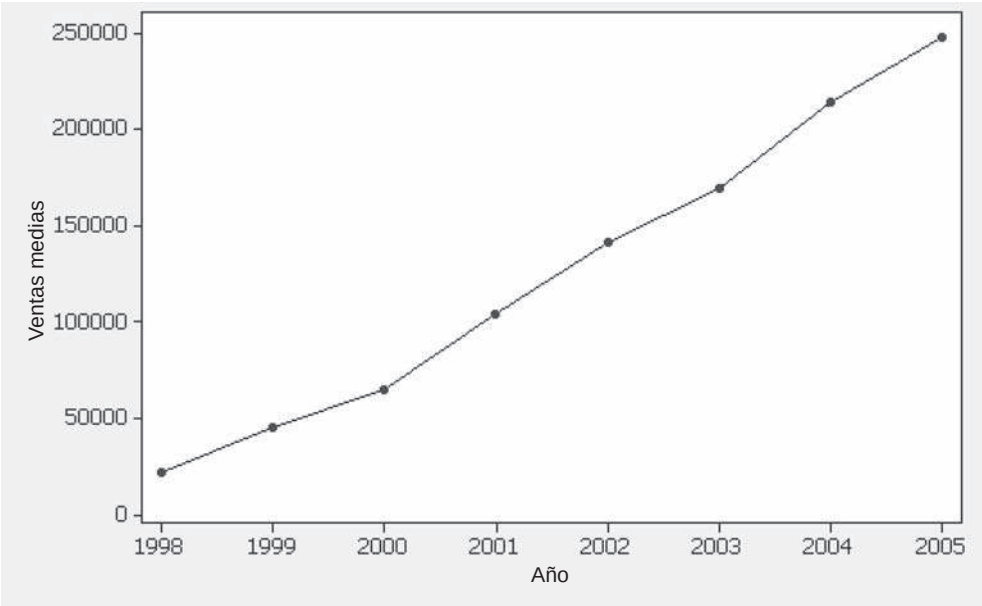


FIGURA 2-19 Ventas mensuales medias de Mr. Tux

PREGUNTAS

1. ¿Qué ideas de pronósticos le vienen a la mente cuando estudia los valores medios de ventas mensuales para los años de esos datos?

2. Suponga que John traza una recta a mano a través de su diagrama de dispersión, de modo que “se ajuste bien” y, luego, extiende esta línea hacia el futuro, usando puntos a lo largo de la línea como sus pronósticos mensuales. ¿Qué tan exactos piensa usted que serían tales pronósticos? Use los valores de la desviación estándar que calculó John,
- al contestar esta pregunta. Con base en su análisis, ¿animaría a John para continuar buscando un método de pronóstico más exacto? John tiene la última versión de Minitab en su computadora. ¿Cree usted que él debería utilizar la función de análisis de regresión de Minitab, para calcular una línea de mínimos cuadrados? Si lo hizo, ¿qué variable *X* debería utilizar para pronosticar las ventas mensuales (*Y*)?

CASO 2-3 ALOMEGA FOOD STORES

En el ejemplo 1.1, la presidenta de Alomega, Julie Ruth, recolectó datos de las operaciones de su compañía. Obtuvo datos de varios meses de ventas, junto con posibles variables explicativas (repase esta situación en el ejemplo 1.1). Mientras su equipo de análisis estaba trabajando con los datos, en un intento por pronosticar las ventas mensuales, se impacientó y se preguntó cuál de las variables de predicción era mejor para este propósito.

Debido a que ella tenía un programa estadístico en su computadora de escritorio, decidió echar una mirada a los datos por sí misma. Primero, encontró los coeficientes de correlación entre la variable de ventas mensuales y varias de las variables explicativas potenciales. Específicamente, estaba interesada en las correlaciones entre las ventas mensuales y el importe de la publicidad en periódicos, el importe mensual de la publicidad en TV, el de la publicidad de uno o dos meses antes, la publicidad en televisión uno o dos meses antes y las calificaciones de la publicidad de sus competidores. Los valores r (coeficientes de correlación) fueron los siguientes:

Coeficiente de correlación r entre
la variable ventas y

Importe mensual de publicidad en periódico	.45
Importe mensual de publicidad en TV	.60
Importe de publicidad en periódico un mes antes	-.32
Importe de publicidad en periódico dos meses antes	.21
Importe de publicidad en TV un mes antes	-.06
Importe de publicidad en TV dos meses antes	.03
Calificación de la publicidad de los competidores	-.18

Julie no estaba sorprendida de encontrar que la correlación más alta era entre las ventas mensuales y los importes de publicidad en TV ($r = .60$); pero esperaba una correlación más fuerte. Decidió usar una función de regresión para calcular la ecuación de la recta

de mínimos cuadrados, usando las ventas como la variable dependiente, y el importe de publicidad en TV como la variable explicativa. Los resultados de esta corrida fueron

$$\begin{aligned}\text{Ventas} &= 341,663 + .336 (\text{publicidad mensual en TV}) \\ r\text{-cuadrada} &= .36(36\%) \quad \text{valor } p = .000\end{aligned}$$

Julie buscó en su libro de texto de estadística universitaria, para interpretar los resultados para r -cuadrada y $\text{valor } p$ que imprimió. Después de la lectura, ella recordó que r -cuadrada (que es el cuadrado del coeficiente de correlación r) mide el porcentaje de la variabilidad en ventas, que puede explicarse con la variabilidad en los dólares mensuales de publicidad en TV (esto se estudiará en el capítulo 6). El $\text{valor } p$ también indica que el coeficiente de la pendiente (.336) es significativo; es decir, puede rechazarse la hipótesis de que es cero en la población de la cual se extrajo la muestra sin que exista posibilidad de error.

Julie concluyó que la ecuación de regresión que encontró era importante y podría utilizarse para el pronóstico de ventas mensuales, si se conoce el presupuesto asignado a publicidad en TV. Puesto que los gastos de publicidad en TV están bajo el control de la compañía, ella sintió que tenía una buena forma de pronosticar las ventas futuras. En una breve conversación con el gerente de su departamento de administración de datos, Roger Jackson, ella mencionó sus hallazgos. Él contestó: “Sí, nosotros encontramos eso también. Pero nos dimos cuenta de que la publicidad en TV explica sólo aproximadamente un tercio de la variabilidad en las ventas. Muy bien, 36%. Realmente no creemos que sea suficiente, y estamos tratando de usar otras variables juntas para intentar obtener ese valor r -cuadrado más elevado. Además, creemos que estamos en un método que consistentemente hará un mejor trabajo que el análisis de regresión”.

PREGUNTAS

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

predichos por la línea recta. ¿Cómo podría examinar usted los residuos para decidir si la representación de la línea recta de Julie es adecuada?

Aplicaciones de Minitab

El problema. En el ejemplo 2.1, se analizó un conjunto de edades usando estadística descriptiva.

Soluciones de Minitab

1. Introduzca el nombre de la variable Ages (Edades) debajo de C1.
2. Introduzca los datos en la columna C1.
3. Haga clic en los menús mostrados abajo.

Stat>Basic Statistics>Graphical Summary

4. Aparece la ventana de diálogo Graphical Summary:
 - a) Haga clic en Variables y seleccione Ages (C1), como se ilustra en la figura 2-20.
 - b) Haga clic en OK en la ventana de diálogo Graphical Summary y aparecerá el resumen mostrado en la figura 2-21.
 - c) Para imprimir esta gráfica, haga clic en los siguientes menús.

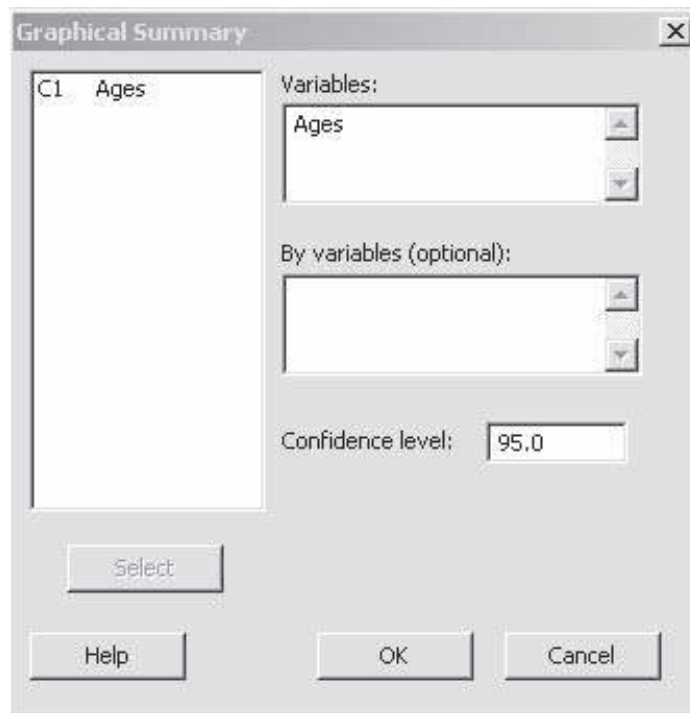
File>Print Graph

5. Haga clic en los menús mostrados a continuación:

Stat>Basic Statistics>Display Descriptive Statistics

6. Aparece la ventana de diálogo Display Descriptive Statistics.
 - a) Haga clic en Variables y seleccione Ages (C1).
 - b) Haga clic en OK y aparece el resumen mostrado en la figura 2-1 en la ventana Session. (Para obtener estadísticas descriptivas adicionales, dé clic en Statistics y seleccione medidas descriptivas adicionales, antes del clic en OK).

FIGURA 2-20 Pantalla de Minitab de la ventana de diálogo Graphical Summary



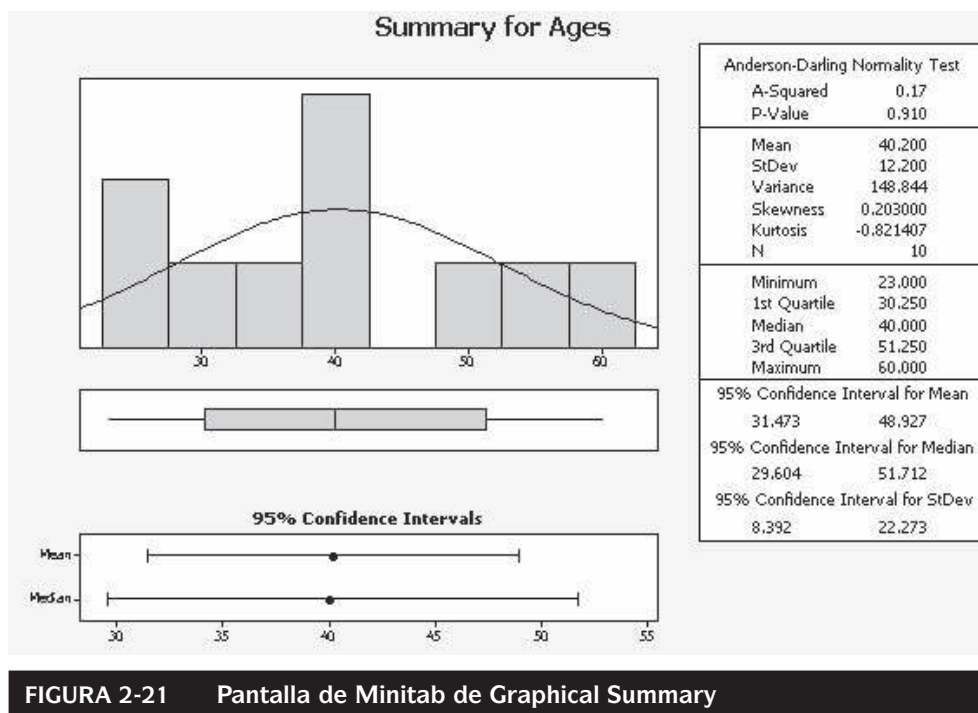


FIGURA 2-21 Pantalla de Minitab de Graphical Summary

- c) Para imprimir la ventana de la sesión que contiene un resumen de los estadísticos descriptivos, haga clic en:

File>Print Session Window

El resultado se ilustra en la figura 2-1.

7. Los siguientes comandos le permitirán a usted desarrollar un histograma, un diagrama de puntos y un diagrama de caja:

Graph>Histogram

Graph>Dotplot

Graph>Boxplot

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

Aplicaciones de Excel

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

Soluciones de Excel

1. Abra el programa Excel y aparecerá la pantalla de hoja de cálculo de la figura 1-2. Mueva el cursor a la celda A1 resaltada en la esquina superior izquierda de la hoja de cálculo.

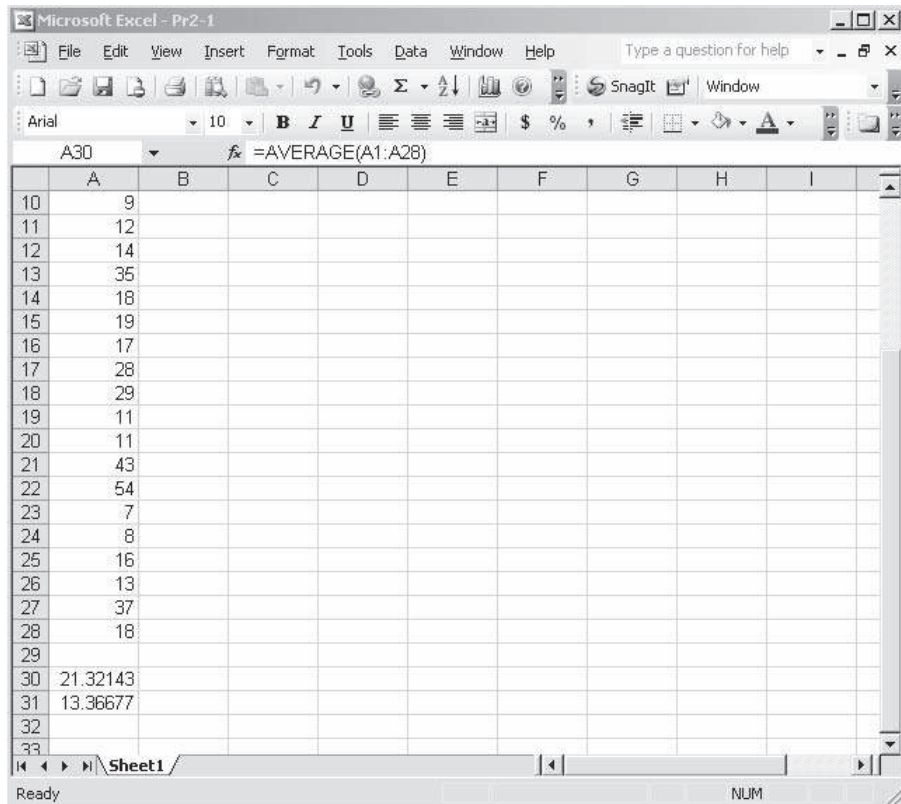


FIGURA 2-22 Hoja de cálculo de Excel para el problema 1

2. Introduzca el primer valor, 10, seguido por la tecla return (enter), luego introduzca el siguiente valor de los datos, y así sucesivamente.
3. Después de que se hayan tecleado los 28 valores de datos en la columna A, se coloca el cursor en la celda donde se desean colocar los resultados del primer cálculo, A30.
4. El promedio de los datos de las celdas A1 a A28 se calcula tecleando la fórmula en A30. Para introducir una fórmula, el signo = (igual) debe ir antes de ella. La fórmula es `= Average(A1:A28)`. *Nota:* Se muestra A30 a la izquierda de la barra de la fórmula e `= Average(A1:A28)` arriba a la derecha de la hoja de cálculo.
5. El mismo enfoque se utiliza para calcular la desviación estándar. La fórmula `= Stdev(A1:A28)` se introduce en la celda A31. Los resultados se muestran en la figura 2-22.

También se pueden calcular el promedio y la desviación estándar usando la función Insert o la herramienta Data Analysis. Estos enfoques se estudiarán en capítulos posteriores.

Las instrucciones de cómo usar Excel para correr los análisis de correlación y de regresión se presentan al final del capítulo 6.

Si usted usa Minitab o Excel, se le sugiere que intente diferentes conjuntos de datos y rutinas estadísticas, para familiarizarse con estos poderosos programas. La habilidad que obtenga será muy útil cuando aprenda acerca de los procedimientos para elaboración de pronósticos en este texto.

Referencias

Anderson, D. R., D. J. Sweeney y T. A. Williams.
Essentials of Modern Business Statistics, 3a. ed.
Belmont, Calif.: Thomson/South-Western, 2007.

Keller, G. *Statistics for Management and Economics*, 7a
ed. Belmont, Calif.: Thomson/South-Western, 2005.

Moore, D. S., G. P. McCabe, W. M. Duckworth y S. L.
Sclove, *The Practice of Business Statistics*. Nueva
York: Freeman, 2003.

EXPLORACIÓN DE PATRONES DE DATOS E INTRODUCCIÓN A LAS TÉCNICAS DE PRONÓSTICOS

Una de las partes más difícil y que más tiempo consume en la elaboración de pronósticos es la recopilación de datos válidos y confiables. El personal de procesamiento de datos usa comúnmente la expresión “basura entra, basura sale”. Esta expresión también se aplica a la elaboración de pronósticos. Un pronóstico no puede ser tan preciso como los datos en que se basa. El modelo de pronóstico más elaborado fallará si se aplica a datos poco confiables.

El poder y la capacidad de las computadoras modernas han traído consigo la acumulación de una cantidad increíble de datos en casi todas las disciplinas. La difícil tarea a la que se enfrenta la mayoría de los pronosticadores es cómo obtener los datos adecuados que les ayuden a resolver sus problemas propios de toma de decisiones.

Para determinar qué datos serán útiles, se aplican cuatro criterios:

1. Los datos deben ser fidedignos y precisos. Se debe tener mucho cuidado en que los datos se obtengan de una fuente confiable, poniendo especial atención en la exactitud.
2. Los datos deberían ser relevantes. Los datos tienen que ser representativos de las circunstancias para las cuales se están usando.
3. Los datos tienen que ser consistentes. Cuando cambian las definiciones relacionadas con la recopilación de datos, se tienen que hacer los ajustes necesarios para conservar la consistencia en los patrones históricos. Esto quizá sea un problema, por ejemplo, cuando las dependencias gubernamentales cambian la miscelánea o la “canasta básica”, que se emplea al determinar el índice del costo de la vida. Hace varios años, las computadoras personales no formaban parte de la combinación de productos que compraban los consumidores; ahora sí lo son.
4. Los datos deberían ser oportunos. Los datos recopilados, resumidos y publicados oportunamente tendrán el mayor valor para el pronosticador. Puede haber muy pocos datos (una historia insuficiente sobre la cual apoyar resultados futuros) o demasiados datos (datos de periodos históricos irrelevantes lejanos en el pasado).

Generalmente, son dos los tipos de datos de interés para el pronosticador. El primer tipo son los datos recopilados en un periodo único, ya sea una hora, un día, una semana, un mes, o un trimestre. El segundo tipo son las observaciones de datos realizadas a través del tiempo. Cuando todas las observaciones se hacen durante el mismo periodo, las llamamos datos de corte *transversal*. El propósito es examinar esos datos y luego extrapolar o extender las relaciones identificadas a una población en general. Por ejemplo, la extracción de una muestra aleatoria de archivos del personal, para conocer la situación de los empleados de una compañía. La recopilación de datos sobre la antigüedad y el costo del mantenimiento de nueve autobuses de Spokane Transit Authority es otro. Un diagrama de dispersión, como el de la figura 3-1, nos ayuda a visualizar las relaciones y sugiere que la antigüedad podría ayudar en la elaboración del pronóstico del presupuesto anual de mantenimiento.

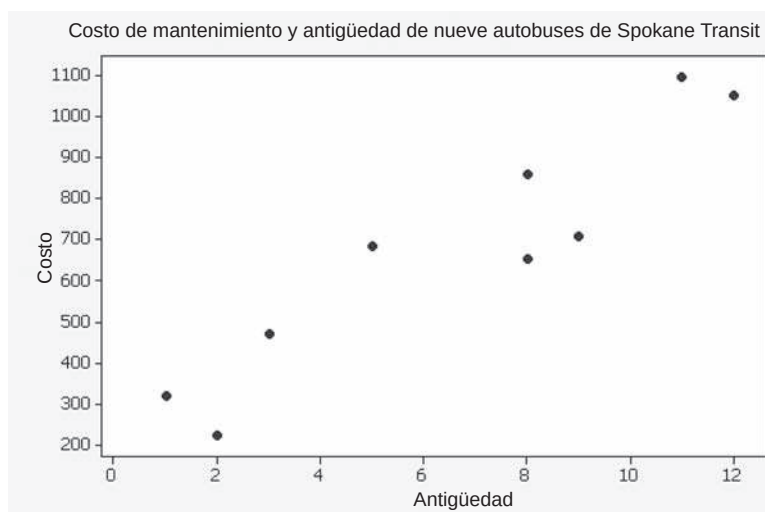


FIGURA 3-1 Diagrama de dispersión de la antigüedad y el costo de mantenimiento para nueve autobuses de Spokane Transit Authority

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

Cualquier variable integrada con datos recopilados, registrados u observados durante incrementos de tiempo sucesivos se llama *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 durante incrementos sucesivos de tiempo.

ESTUDIO DE PATRONES DE DATOS EN LAS SERIES DE TIEMPO

Uno de los pasos más importantes en la selección de un método para pronosticar adecuado con datos de una serie de tiempo es considerar los diferentes tipos de patrones de datos. Existen cuatro tipos generales: horizontal, tendencias, estacionales y cíclicos.

Cuando los datos recopilados en el transcurso del tiempo fluctúan alrededor de un nivel o una media constantes, hay un patrón *horizontal*. Se dice que este tipo de series es *estacionario* en su media. Se considera que las ventas mensuales de un producto alimenticio que no se incrementan, ni disminuyen, consistentemente durante un largo periodo tienen un patrón horizontal.

Cuando los datos crecen o descienden en varios periodos, existe un patrón *de tendencia*. La figura 3-2 muestra el crecimiento (tendencia) a largo plazo de una serie de tiempo (costos de viviendas) con datos anuales. Para ilustrar el crecimiento se dibuja una recta de tendencia lineal. Si bien la variable costo de la vivienda no se ha incrementado cada año, el cambio de la variable ha sido generalmente hacia arriba entre los periodos 1 a 20. Algunos ejemplos de las fuerzas básicas que afectan y ayudan a explicar la tendencia de una serie son el crecimiento de la población, la inflación de los precios, el avance tecnológico, las preferencias del consumidor y los incrementos en la productividad.

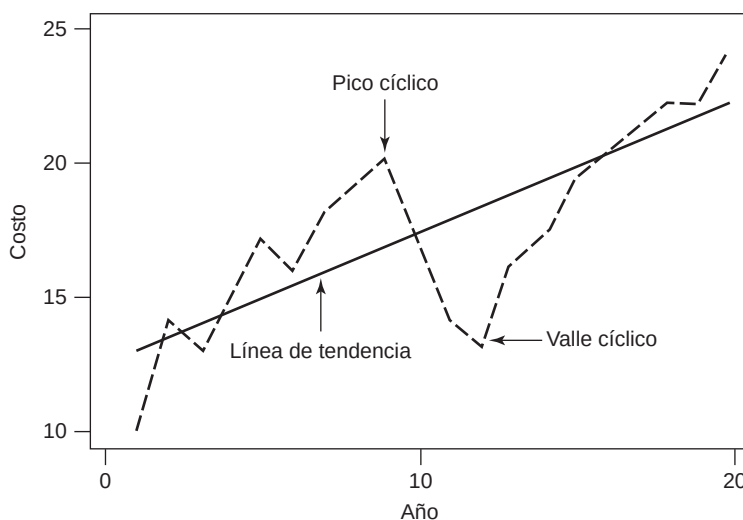


FIGURA 3-2 Tendencia y componentes cíclicos de una serie de tiempo anual, como el costo de la vivienda

Muchas variables macroeconómicas, como el producto interno bruto (PIB), el empleo y la producción industrial de un país presentan un comportamiento con tendencia. La figura 3-10 (véase p. 74) contiene otro ejemplo de una serie de tiempo con una tendencia preponderante. Esta figura muestra el crecimiento de ingresos operativos para Sears, de 1995 a 2004.

La *tendencia* es el componente de largo plazo que representa el crecimiento o el descenso en la serie de tiempo, durante un periodo extenso.

Cuando las observaciones indican aumentos y caídas que no tienen un periodo fijo, existe un patrón cíclico. El *componente cíclico* es la fluctuación con forma de onda alrededor de la tendencia y, por lo común, se ve afectada por las condiciones económicas generales. Un componente cíclico, si existe, típicamente presenta un ciclo durante varios años. Las fluctuaciones cíclicas a menudo están influidas por cambios en las expansiones y contracciones económicas, mejor conocidas como el ciclo de negocios. La figura 3-2 también muestra una serie de tiempo con un componente cíclico. El pico cíclico en el periodo 9 ilustra una expansión económica; y el valle cíclico en el periodo 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 tras año. Para las series mensuales, el componente estacional mide la variabilidad de la serie cada enero, cada febrero, y así sucesivamente. Para una serie trimestral, hay cuatro elementos estacionales, uno por cada trimestre. La figura 3-3 indica que el consumo de electricidad de clientes residenciales de la compañía Washington Water Power es mayor en el primer trimestre (los

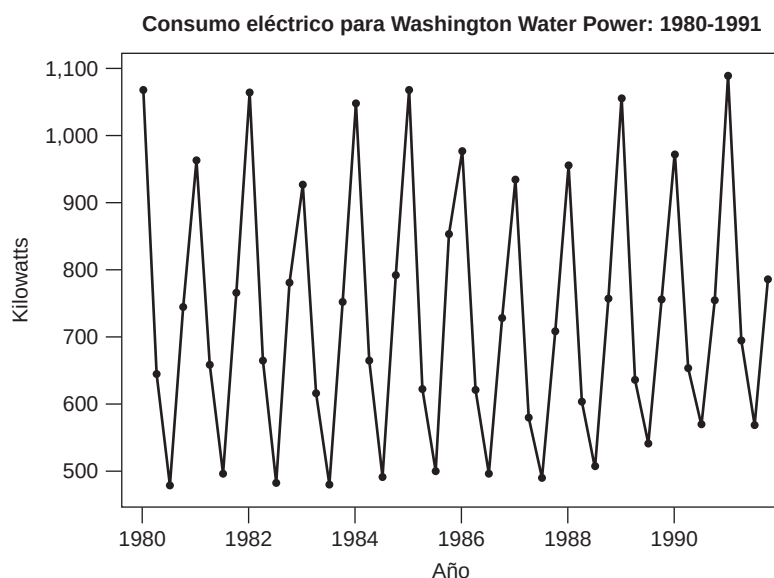


FIGURA 3-3 Consumo eléctrico para la compañía Washington Water Power, 1980-1991

meses de invierno) de cada año. La figura 3-14 (véase p. 77) indica que las ventas trimestrales de Coastal Marine son usualmente bajas en el primer trimestre de cada año. La variación estacional puede representar condiciones climáticas, horarios de escuela, días feriados o duración de los meses calendario.

El componente estacional es un patrón de cambio 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 periodos a menudo están relacionadas o correlacionadas. Esta correlación se mide usando el coeficiente de autocorrelación.

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

Los patrones de datos que incluyen componentes como tendencia y estacionalidad pueden estudiarse usando autocorrelaciones. Los patrones se identifican examinando los coeficientes de autocorrelación de una variable en diferentes retrasos de tiempo.

El concepto de autocorrelación se ilustra con los datos presentados en la tabla 3-1. Observe que los valores de las variables Y_{t-1} y Y_{t-2} son valores reales de Y , que se han retrasado uno y dos periodos, respectivamente. Los valores para marzo, los cuales se muestran en el renglón para el periodo 3, son ventas de marzo, $Y_t = 125$; ventas de febrero, $Y_{t-1} = 130$; y ventas de enero, $Y_{t-2} = 123$.

TABLA 3-1 Datos de la VCR para el ejemplo 3.1

Tiempo t	Mes	Datos originales Y_t	Y atrasada un periodo Y_{t-1}	Y atrasada dos periodos Y_{t-2}
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	Octubree	157	147	146
11	Noviembre	150	157	147
12	Diciembre	160	150	157

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

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

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 k periodos anteriores o durante un periodo $t - k$

Ejemplo 3.1

Harry Vernon, ha recopilado datos del número de VCR vendidos el año pasado por Vernon's Music Store. Los datos se presentan en la tabla 3-1. La tabla 3-2 muestra las operaciones que llevan al cálculo del coeficiente de autocorrelación para un retraso de 1. La figura 3-4 contiene un diagrama de dispersión de los pares de observaciones (Y_t , Y_{t-1}). Al revisar el diagrama de dispersión, queda claro que la correlación de retraso 1 será positiva.

El coeficiente de autocorrelación de retraso 1 (r_1), es decir, la autocorrelación entre Y_t y Y_{t-1} , se calcula usando los totales de la tabla 3.2 y la ecuación 3.1. De modo que:

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

Como se sugiere en el diagrama de la figura 3-4, en esta serie de tiempo hay una autocorrelación positiva para el retraso 1. La correlación entre Y_t y Y_{t-1} o la autocorrelación para el retraso de 1 es .572. Esto significa que las ventas mensuales sucesivas de VCR están correlacionadas de algún modo unas con otras. Esta información daría a Harry perspectivas valiosas sobre su serie de tiempo, que le ayudaría a prepararse para usar un método avanzado para la elaboración del pronóstico y quizás advertirle acerca del uso de un análisis de regresión con tales datos. Todas estas ideas se analizarán en capítulos posteriores.

TABLA 3-2 Cálculo del coeficiente de autocorrelación del retraso 1, para los datos de 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

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

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

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 con la ecuación 3.1.

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

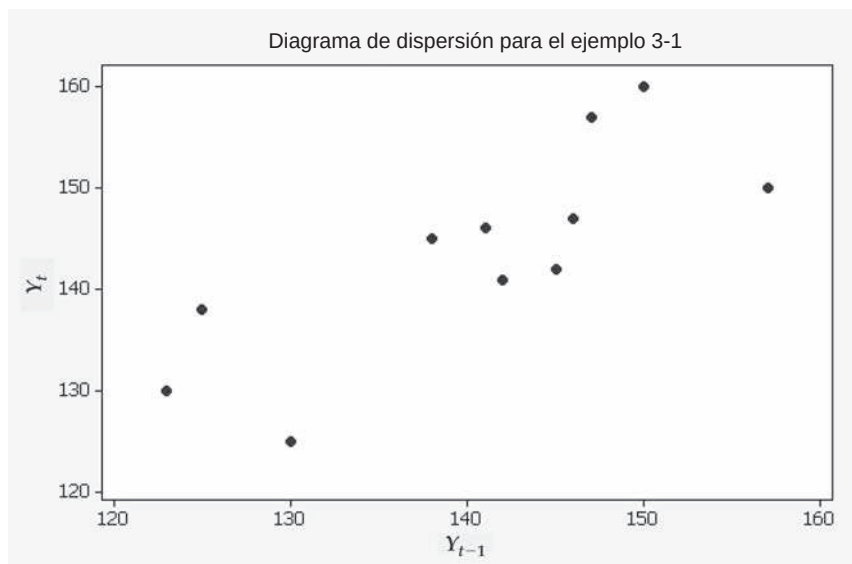


FIGURA 3-4 Diagrama de dispersión de los datos de Vernon's Music Store para el ejemplo 3-1

Parece que existe una autocorrelación moderada en esta serie de tiempo, para dos periodos de retraso. La correlación entre Y_t y Y_{t-2} , o la autocorrelación para un retraso 2, es de .463. Observe que el coeficiente de autocorrelación para el retraso de 2 (.463) es menor que el coeficiente de autocorrelación para el retraso de 1 (.572). Generalmente, conforme aumenta el número de retrasos (k), disminuyen las magnitudes de los coeficientes de autocorrelación.

La figura 3-5 ilustra una gráfica de autocorrelaciones con retrasos de tiempo, para los datos de Harry Vernon usados en el ejemplo 3-1. La escala horizontal en la parte inferior de la gráfica presenta cada retraso de tiempo de interés: 1, 2, 3, etc. La escala vertical de la izquierda indica el posible rango del 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 por arriba de un retraso de tiempo 1 muestra un coeficiente de autocorrelación de .57, o $r_1 = .57$. La línea vertical que se extiende hacia arriba de un retraso de tiempo 2 muestra un coeficiente de autocorrelación .46, o $r_2 = .46$. Los estadísticos de las líneas punteadas, los estadísticos T y LBQ (Ljung-Box Q) desplegados en la ventana de sesión de Minitab se examinarán en los ejemplos 3.2 y 3.3. Los patrones en un correlograma se usan para analizar las características clave de los datos, un concepto que se estudia en la siguiente sección. El software Minitab (para instrucciones específicas, véase la sección de aplicaciones de Minitab al final del capítulo) puede usarse para calcular autocorrelaciones y desarrollar correlogramas.

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

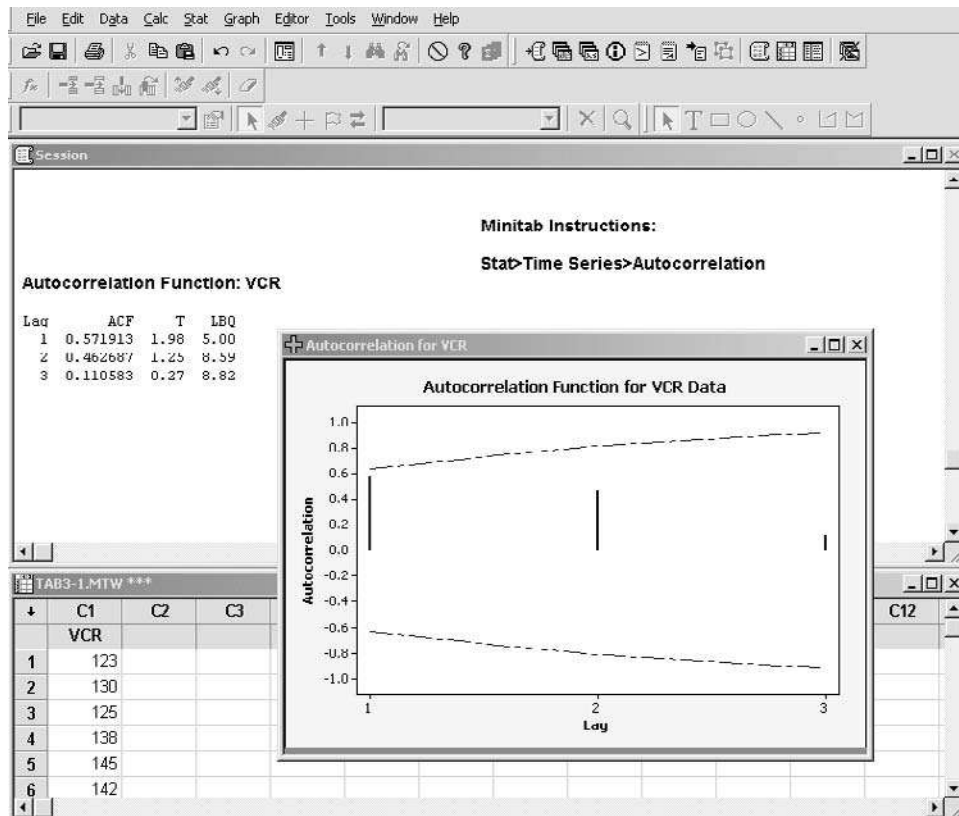


FIGURA 3-5 Correlograma o función de autocorrelación de los datos usados para el ejemplo 3.1

Con una pantalla como la de la figura 3-5, se estudian los patrones de datos incluyendo la tendencia y la estacionalidad. Los coeficientes de autocorrelación para diferentes retrasos de tiempo de una variable pueden usarse para contestar las siguientes preguntas acerca de una serie de tiempo:

1. ¿Los datos son aleatorios?
2. ¿Los datos muestran una tendencia (son no estacionarios)?
3. ¿Los datos son estacionarios?
4. ¿Los datos son estacionales?

Si una serie es aleatoria, las autocorrelaciones entre Y_t y Y_{t-k} para cualquier retraso de tiempo k son cercanas a cero. Los valores sucesivos de una serie de tiempo no están relacionados entre sí.

Si una serie muestra una tendencia, las observaciones sucesivas están altamente correlacionadas y es típico que los coeficientes de correlación sean significativamente diferentes de cero, para los primeros retrasos de tiempo, y de forma gradual tienden a cero conforme se incrementa el número de retrasos. El coeficiente de autocorrelación para un retraso de tiempo 1 a menudo será 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 presentará un coeficiente de autocorrelación significativo en el retraso de tiempo estacional o en los múltiplos del retraso estacional. El retraso estacional se considera de 4 para datos trimestrales y 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 una media de cero y una desviación estándar aproximada de $1/\sqrt{n}$. Sabiendo esto, el analista compara los coeficientes de autocorrelación de la muestra, con esta distribución teórica de la muestra y determinar si, para retrasos de tiempo dados, provienen de una población cuya media sea cero.

En realidad, algunos paquetes de software usan una fórmula ligeramente diferente, como se indica en la ecuación 3.2, para calcular las desviaciones estándar (o errores estándar) de los coeficientes de autocorrelación. Esta fórmula supone que cualquier autocorrelación antes del retraso de tiempo k es diferente de cero, y que cualquier autocorrelación anterior al retraso 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 (desviación estándar estimada) de la autocorrelación en retraso k

r_i = la autocorrelación en el retraso i

k = retraso de tiempo

n = el número de observaciones en la serie de tiempo

Este cálculo se demostrará en el ejemplo 3.2. Si la serie es verdaderamente aleatoria, casi todos los coeficientes de autocorrelación de la muestra deberían estar dentro de un rango especificado por cero, más o menos cierto número de errores estándar. En un nivel de confianza específico, una serie puede considerarse aleatoria, si cada uno de los coeficientes de autocorrelación calculados se encuentra dentro del intervalo alrededor de 0, definido por $0 \pm t \times SE(r_k)$, donde el multiplicador t es un punto porcentual adecuado de una distribución t .

Si bien es útil cada prueba de r_k para saber si, en lo individual, es significativamente diferente de 0, también es recomendable examinar un conjunto de r_k consecutivos, como un grupo. Podemos usar una prueba de conjunto para ver si, por ejemplo, de los primeros 10 valores de r_k son significativamente diferentes de un conjunto donde los 10 valores son cero.

Una prueba de baúl común es el estadístico modificado Q de Ljung-Box (ecuación 3.3). Esta prueba se aplica usualmente a los residuos de un modelo de pronóstico. Si las autocorrelaciones se calculan mediante un proceso aleatorio (o de ruido blanco), el estadístico Q tiene una distribución chi cuadrada con m grados de libertad (el número de retrasos de tiempo por probar). Sin embargo, para los residuos de un modelo de pronóstico, el estadístico Q tiene una distribución chi cuadrada, con m grados de libertad menos el número de parámetros estimados en el modelo. El valor del estadístico Q puede compararse con la tabla de chi cuadrada (tabla B-4), para determinar si es mayor de lo que esperaríamos que fuera, con la hipótesis nula de que todas las autocorrelaciones en el conjunto son cero. Alternativamente, el valor p generado por el estadístico de prueba Q puede ser calculado e interpretado. El estadístico Q está dado por la ecuación 3.3 y se demuestra 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 = retraso de tiempo

m = número de retrasos de tiempo que se van a probar

r_k = la función de autocorrelación muestral de los residuos atrasados k periodos

¿Los datos son aleatorios?

Un modelo aleatorio simple, a menudo llamado *modelo de ruido blanco*, está representado por la ecuación 3.4. La observación Y_t se compone de dos partes: c , el nivel general; y ε_t , el componente de error aleatorio. Es importante advertir que se supone que el componente ε_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 consistentes con este modelo? Este punto 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 significativamente diferente de cero, para el correlograma de la figura 3-5. Las hipótesis nula y alternativa para la prueba de significancia del coeficiente de autocorrelación de la población del retraso 1 son:

$$H_0: \rho_1 = 0$$

$$H_1: \rho_1 \neq 0$$

Si la hipótesis nula es verdadera, el estadístico 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 $df = n - 1$. Donde, $n - 1 = 12 - 1 = 11$, entonces, para un nivel de significancia de 5%, la regla de decisión es como sigue:

Si $t < -2.2$, o bien, $t > 2.2$ rechace H_0 y concluya que la autocorrelación del retraso 1 es significativamente diferente de 0.

Los valores críticos ± 2.2 son los puntos .025 superior e inferior, de una distribución t con 11 grados de libertad. El error estándar de r_1 es $SE(r_1) = \sqrt{1/12} = \sqrt{.083} = .289$, y el valor del estadístico de prueba se convierte en

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

Con la anterior regla de decisión, no puede rechazarse $H_0: \rho_1 = 0$ ya que $-2.2 < 1.98 < 2.2$. Observe que el valor de nuestro estadístico de prueba, $t = 1.98$, es el mismo que la cantidad del renglón de retraso 1, debajo de la T en el resultado de Minitab en la figura 3-5. Los valores T en la figura de Minitab son simplemente los valores del estadístico de prueba al realizar la prueba para la autocorrelación igual a 0 en los diferentes retrasos.

Para la prueba de autocorrelación cero en el retraso de tiempo 2, consideramos

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

Y el estadístico de prueba

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

Usando la ecuación 3.2,

$$SE(r_2) = \sqrt{\frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{n}} = \sqrt{\frac{1 + 2 \sum_{i=1}^{2-1} r_i^2}{n}} = \sqrt{\frac{1 + 2(.572)^2}{12}} = \sqrt{\frac{1.6544}{12}} = \sqrt{.138} = .371$$

y

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

Este resultado coincide con el valor T para el retraso 2 en el resultado de Minitab de la figura 3-5.

Usando la anterior regla de decisión, $H_0: \rho_1 = 0$ no puede rechazarse con un nivel de 0.05, ya que $-2.2 < 1.25 < 2.2$. Una forma alternativa para verificar la autocorrelación significativa es construir, digamos, límites de confianza de 95% centrados en cero. Estos límites para los retrasos de tiempo 1 y 2 son como sigue:

$$\begin{aligned} \text{retraso 1: } 0 \pm t_{.025} \times SE(r_1) & \quad o \text{ bien} \quad 0 \pm 2.2(.289) \rightarrow (-.636, .636) \\ \text{retraso 2: } 0 \pm t_{.025} \times SE(r_2) & \quad o \text{ bien} \quad 0 \pm 2.2(.371) \rightarrow (-.816, .816) \end{aligned}$$

Está indicada una autocorrelación significativamente diferente de 0, siempre que un valor de r_k caiga fuera de los límites de confianza correspondientes. En la figura 3-5 se presentan los límites de confianza de 95%, representados con líneas punteadas, en la gráfica de la función de autocorrelación.

Ejemplo 3.3

Se usó Minitab para generar la serie de tiempo de 40 números de tres dígitos pseudoaleatorios presentados en la tabla 3-3. La figura 3-6 muestra una gráfica de la serie de tiempo con estos datos. Puesto que tales datos son aleatorios (independientes entre sí y todos de la misma población), las autocorrelaciones para todos los retrasos de tiempo, teóricamente, deberían ser iguales a cero. Por supuesto, los

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

40 valores de la tabla 3-3 son sólo un conjunto, de 40 elementos, de un gran número de muestras posibles de tamaño 40. Cada muestra producirá autocorrelaciones diferentes. La mayoría de estas muestras producirán coeficientes de autocorrelación muestrales que estén cercanos a cero. No obstante, es posible que una muestra produzca un coeficiente de autocorrelación que sea significativamente diferente de cero, tan sólo por casualidad.

En seguida, usando Minitab, se construye la función de autocorrelación representada en la figura 3-7. Observe que las dos líneas punteadas indican los límites de confianza de 95%. Se examinan diez retrasos de tiempo y todos los coeficientes de autocorrelación individuales están dentro de estos límites. No hay razón para dudar de que cada una de las autocorrelaciones para los primeros 10 retrasos sea cero. Sin embargo, aun cuando las autocorrelaciones muestrales individuales no sean significativamente diferentes de cero, ¿las magnitudes de los primeros $10r_k$, consideradas como grupo, son mayores de lo que uno esperaría con la hipótesis de que no hay autocorrelación en ningún retraso? Esta pregunta la contesta el estadístico Q de Ljung-Box (LBQ en Minitab).

Si no hay autocorrelación en cualquiera de los retrasos, el estadístico Q tiene una distribución chi cuadrada con, en este caso, con $df = 10$. Como resultado, un valor grande para Q (un extremo de la distribución chi cuadrada) es evidencia contra la hipótesis nula. A partir la figura 3-7 vemos que el valor de Q (LBQ) para 10 retrasos de tiempo es 7.75. En la tabla B-4 observamos que el punto .05 más alto de una distribución chi cuadrada con 10 grados de libertad es 18.31. Puesto que $7.75 < 18.31$, no se puede rechazar la hipótesis nula con el nivel de significancia de 5%. Estos datos no están correlacionados en ningún retraso de tiempo, lo cual es consistente con el modelo de la ecuación 3.4.

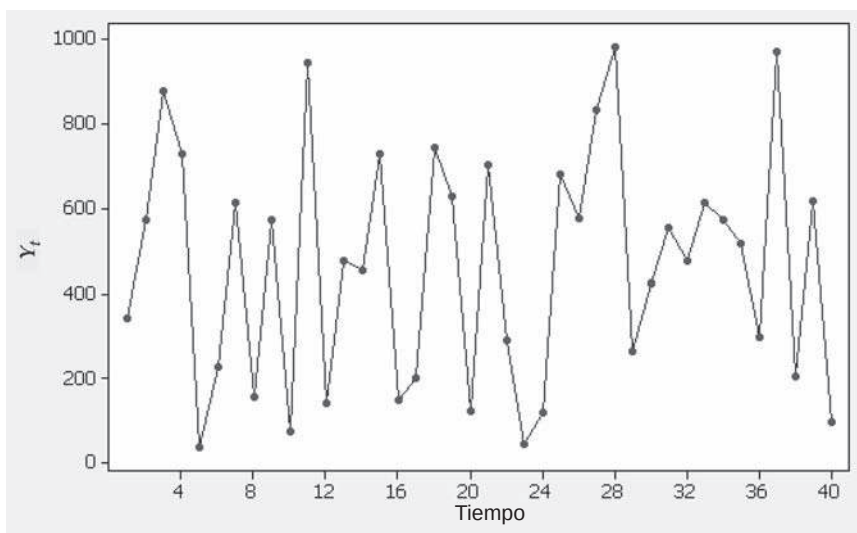
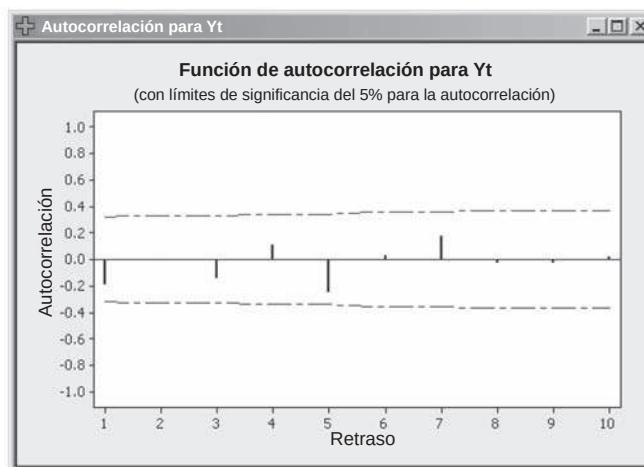


FIGURA 3-6 Gráfica de serie de tiempo de 40 números aleatorios usados para el ejemplo 3.3

Función de autocorrelación: Yt

Retraso	ACF	T	LBQ
1	-0.191171	-1.21	1.57
2	-0.006293	-0.04	1.58
3	-0.145224	-0.89	2.53
4	0.104475	0.63	3.04
5	-0.253756	-1.50	6.13
6	0.028640	0.16	6.17
7	0.169071	0.95	7.63
8	-0.027390	-0.15	7.67
9	-0.032551	-0.18	7.73
10	0.021546	0.12	7.75

**FIGURA 3-7** Función de autocorrelación para los datos usados en el ejemplo 3.3

¿Los datos muestran una tendencia?

Si una serie muestra una tendencia, hay una relación significativa entre los valores sucesivos de la serie de tiempo. Los coeficientes de autocorrelación son usualmente grandes para varios de los primeros retrasos de tiempo y luego, conforme se incrementa el número de retrasos, caen gradualmente hacia cero.

Una serie de tiempo *estacionaria* es aquella cuyas propiedades estadísticas básicas, como la media y la varianza, permanecen constantes en el tiempo. Por lo tanto, se dice que una serie que varía alrededor de un nivel fijo (sin crecimiento ni decrecimiento) con el paso del tiempo es estacionaria. Se dice también que una serie que contiene una tendencia es *no estacionaria*. Los coeficientes de autocorrelación de una serie estacionaria decrecen hacia cero bastante rápidamente, por lo común después del segundo o tercer retraso de tiempo. Por otro lado, las autocorrelaciones muestrales de serie no estacionaria se permanecen muy grandes durante varios periodos. A menudo, para analizar la serie no estacionaria, se elimina la tendencia antes de aplicar cualquier modelo. Los procedimientos que se estudian en el capítulo 9 utilizan dicho enfoque.

Con frecuencia se utiliza un método llamado *diferencias* para eliminar la tendencia de una serie no estacionaria. Los datos de las VCR incluidos originalmente en la tabla 3-1 se presentan otra vez en la columna A de la figura 3-8. Los valores de Y_t , retrasados un periodo Y_{t-1} , se presentan en la columna B. Las diferencias de $Y_t - Y_{t-1}$ (columna A - columna B) se presentan en la columna C. Por ejemplo, el primer valor de las diferencias es $Y_2 - Y_1 = 130 - 123 = 7$. Observe el crecimiento o la tendencia de los datos de las VCR presentados en la gráfica A de la figura 3-9. Ahora observe el patrón estacionario de los datos diferenciados en la gráfica B de la figura 3-9. La diferencia de los datos elimina la tendencia.

Ejemplo 3.4

A Maggie Trymane, una analista de Sears, se le asignó la tarea de pronosticar los ingresos operativos para 2005. Ella reúne los datos de los años 1955 a 2004, que se muestran en la tabla 3-4. Los datos se grafican como una serie de tiempo en la figura 3-10. Observe que, si bien los ingresos operativos de Sears estuvieron decreciendo durante el periodo 2000 a 2004, la tendencia general durante el periodo completo 1955 a 2004 fue hacia arriba. Primero, Maggie calcula un intervalo de confianza de 95% para los coeficientes de autocorrelación para un retraso de tiempo 1, usando $0 \pm Z_{0.025}(1/\sqrt{n})$ donde, para muestras grandes, la distribución normal estándar correspondiente a .025 se ha sustituido a la distribución t correspondiente:

$$0 \pm 1.96(\sqrt{1/50})$$

$$0 \pm .277$$

Después, Maggie corre los datos en Minitab y obtiene la función de autocorrelación que se presenta en la figura 3-11. Durante la revisión, se observa que las autocorrelaciones para los primeros cua-

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Arial 10 B I U

C3 =A3-B3

	A	B	C	D	E	F	G	H	I	J
1	Yt	Yt-1	Differences							
2	123									
3	130	123	7							
4	125	130	-5							
5	138	125	13							
6	145	138	7							
7	142	145	-3							
8	141	142	-1							
9	146	141	5							
10	147	146	1							
11	157	147	10							
12	150	157	-7							
13	160	150	10							
14										
15										
16										
17										
18										
19										
20										

FIGURA 3-8 Resultados de Excel de la diferenciación de los datos de las VCR para el ejemplo 3.1

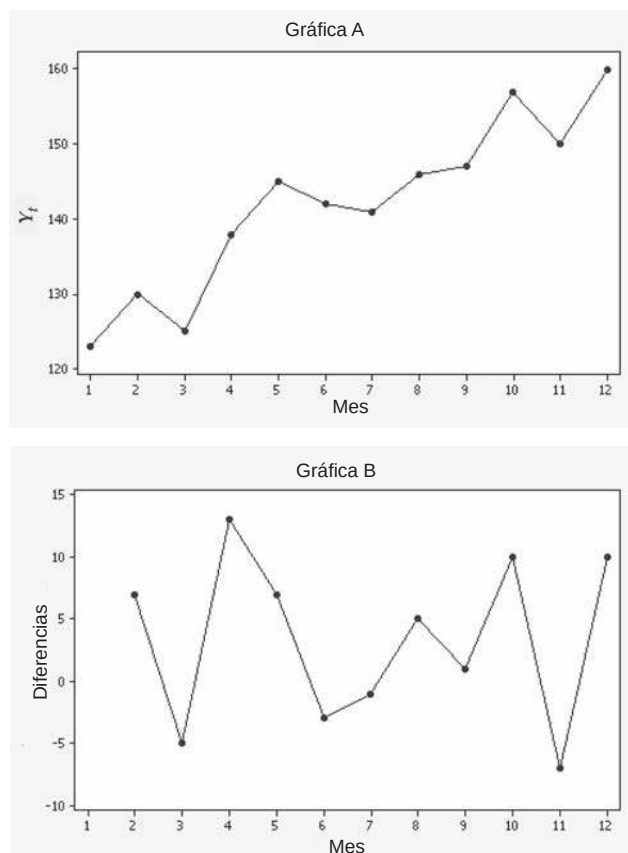


FIGURA 3-9 Gráficas de series de tiempo de los datos de las VCR y los datos diferenciados de las VCR para el ejemplo 3.1

TABLA 3-4 Ingresos operativos anuales de Sears, 1955 a 2004, para el ejemplo 3.4

Año	Y_t	Año	Y_t	Año	Y_t	Año	Y_t	Año	Y_t
1955	3,307	1967	7,296	1979	17,514	1991	57,242	2003	23,253
1956	3,556	1968	8,178	1980	25,195	1992	52,345	2004	19,701
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	2001	36,151		
1966	6,769	1978	17,946	1990	55,972	2002	30,762		

Fuente: Industry Surveys, varios años.

**FIGURA 3-10 Gráficas de series de tiempo de los ingresos operativos de Sears para el ejemplo 3.4**

tro retrasos de tiempo son significativamente diferentes de cero (.96, .92, .87 y .81) y que luego los valores tienden gradualmente hacia cero. Finalmente, Maggie mira el estadístico Q para 10 retrasos de tiempo. El LBQ es 300.56, el cual es mayor que el valor 18.3 de chi cuadrada (el punto .05 superior de una distribución chi cuadrada con 10 grados de libertad). Este resultado indica que las autocorrelaciones para los primeros 10 retrasos como grupo son significativamente diferentes de cero. Ella decide que los datos están altamente autocorrelacionados y que muestran una tendencia en su comportamiento.

Maggie sospecha que las series pueden diferenciarse para eliminar la tendencia y crear una serie estacionaria. Luego diferencia los datos (véase la sección de aplicaciones de Minitab al final del capítulo), y los resultados se presentan en la figura 3-12. Las series diferenciadas no presentan evidencia de alguna tendencia, en tanto que la función de autocorrelación de la figura 3-13 parece apoyar esta conclusión. Examinando la figura 3-13, Maggie observa que el coeficiente de autocorrelación en el retraso de tiempo 3, .32, es significativamente diferente de cero (probado al nivel de significancia .05). Las autocorrelaciones de los retrasos diferentes del retraso 3 son pequeñas, mientras que la del estadístico LBQ para un retraso 10 también es relativamente pequeña, de manera que hay una pequeña evidencia que sugiere que los datos diferenciados están autocorrelacionados. Todavía Maggie se pregunta si existe algún patrón en estos datos, que pueda modelarse con una de las técnicas más avanzadas como las que se estudian en el capítulo 9.

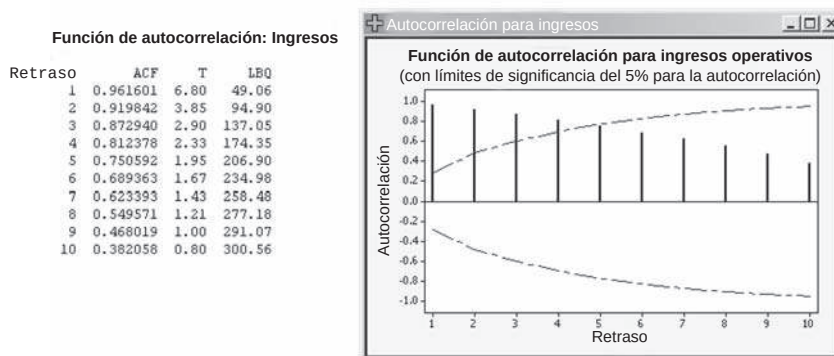


FIGURA 3-11 Función de autocorrelación de ingresos operativos de Sears para el ejemplo 3.4

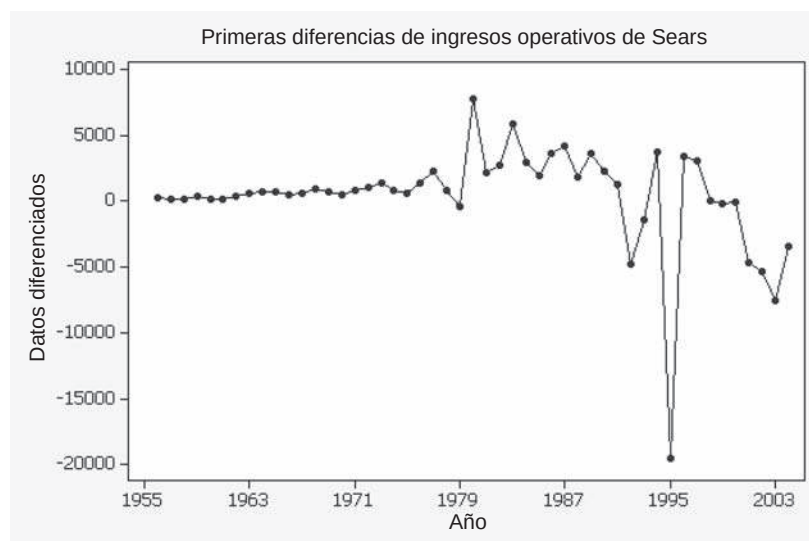


FIGURA 3-12 Gráfica de serie de tiempo de las primeras diferencias de los ingresos operativos de Sears para el ejemplo 3.4

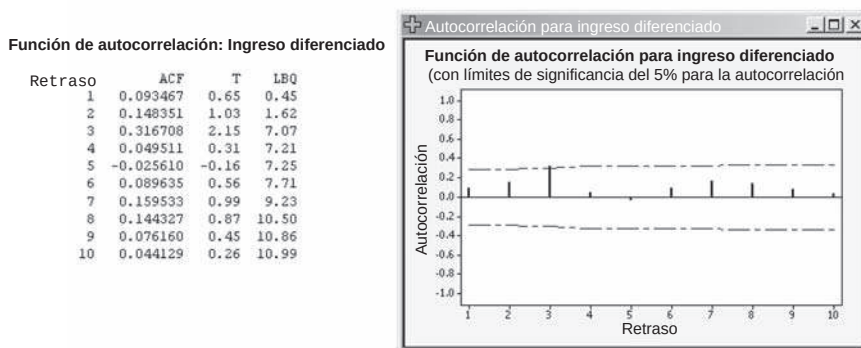


FIGURA 3-13 Función de autocorrelación para las primeras diferencias de los ingresos operativos de Sears para el ejemplo 3.4

TABLA 3-5 Ventas trimestrales de Coastal Marine, 1994-2006, para el ejemplo 3-5

<i>Año fiscal</i>	<i>Diciembre 31</i>	<i>Marzo 31</i>	<i>Junio 30</i>	<i>Septiembre 30</i>
1994	147.6	251.8	273.1	249.1
1995	139.3	221.2	260.2	259.5
1996	140.5	245.5	298.8	287.0
1997	168.8	322.6	393.5	404.3
1998	259.7	401.1	464.6	479.7
1999	264.4	402.6	411.3	385.9
2000	232.7	309.2	310.7	293.0
2001	205.1	234.4	285.4	258.7
2002	193.2	263.7	292.5	315.2
2003	178.3	274.5	295.4	286.4
2004	190.8	263.5	318.8	305.5
2005	242.6	318.8	329.6	338.2
2006	232.1	285.6	291.0	281.4

¿Los datos son estacionales?

Si una serie es estacional, un patrón relacionado con el calendario se repite a sí mismo durante un intervalo de tiempo específico (generalmente un año). Las observaciones de la misma posición, en diferentes periodos estacionales, tienden a estar relacionadas. Si se analizan datos trimestrales que tienen un patrón estacional, los primeros trimestres tienden a parecerse, los segundos trimestres tienden a parecerse, y así sucesivamente, y habrá un coeficiente de autocorrelación significativo en el retraso de tiempo 4. Si se analizan datos mensuales, aparecerá un coeficiente de autocorrelación significativo en el retraso de tiempo 12. Es decir, enero se correlacionará con otros enero, febrero se correlacionará con otros febreros y así sucesivamente. En el ejemplo 3.5 se analiza una serie que es estacional.

Ejemplo 3.5

Perkin Kendell es un analista de Coastal Marine Corporation. Él siempre ha creído que las ventas son estacionales. Perkin reúne los datos presentados en la tabla 3-5 de las ventas trimestrales de Coastal Marine Corporation, de 1994 a 2006, y los grafica como una serie de tiempo como la presentada en la figura 3-14. Después, calcula un intervalo de confianza de 95% para una muestra grande con el coeficiente de autocorrelación en el retraso de tiempo 1:

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

$$0 \pm .272$$

Luego, Perkin calcula los coeficientes de autocorrelación presentados en la figura 3-15. Él nota que los coeficientes de autocorrelación en los retrasos de tiempo 1 y 4 son significativamente diferentes de cero ($r_1 = .39 > .272$ y $r_4 = .74 > .333$). Concluye que las ventas de Coastal Marine son estacionales, sobre una base trimestral.

SELECCIÓN DE UNA TÉCNICA DE PRONÓSTICOS

En su mayor parte, este libro está dedicado a la explicación de varias técnicas de elaboración de pronósticos y a la demostración de su utilidad. Primero se examina el valioso trabajo de seleccionar, entre varias, una técnica de elaboración de pronósticos.

Veamos algunas de las preguntas que deben considerarse, antes de decidir sobre la técnica más adecuada para la elaboración del pronóstico de un problema específico:

- ¿Por qué es necesario un pronóstico?
- ¿Quién usará el pronóstico?

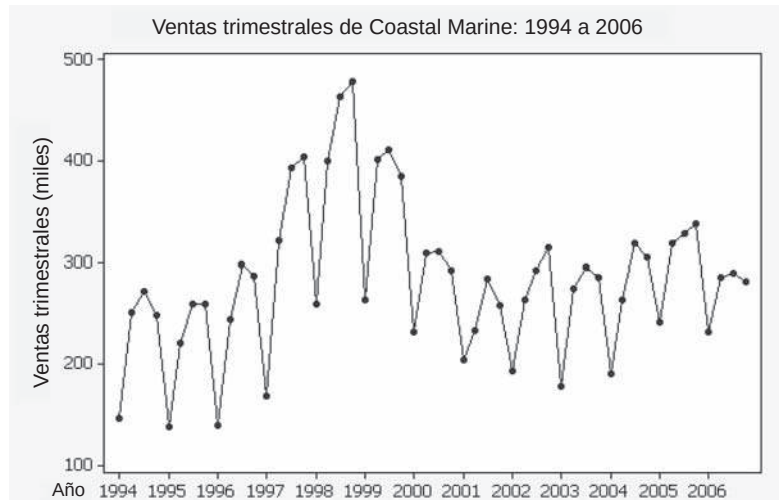


FIGURA 3-14 Gráfica de serie de tiempo de las ventas trimestrales de Coastal Marine para el ejemplo 3.5

Función de autocorrelación: Ventas

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

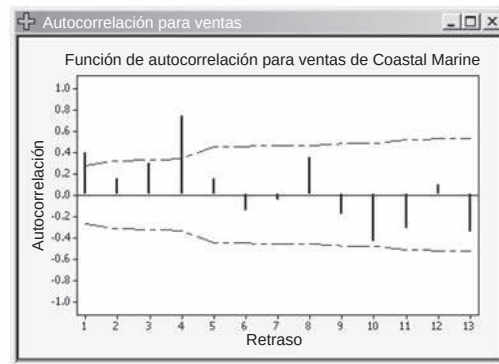


FIGURA 3-15 Función de autocorrelación de las ventas trimestrales de Coastal Marine para el ejemplo 3.5

- ¿Cuáles son las características de los datos disponibles?
- ¿Qué periodo se va pronosticar?
- ¿Cuáles son los requerimientos mínimos de datos?
- ¿Qué precisión se desea?
- ¿Cuánto costará el pronóstico?

Para elegir acertadamente la técnica de pronóstico, el pronosticador debe hacer lo siguiente:

- Definir la naturaleza del problema que se va a pronosticar.
- Explicar la naturaleza de los datos en investigación.
- Describir las capacidades y limitaciones de las técnicas de elaboración de pronósticos potencialmente útiles.
- Desarrollar algún criterio predeterminado, con el cual se pueda tomar la decisión.

Un factor importante que influye en la selección de la técnica de elaboración del pronóstico es la identificación y comprensión de patrones históricos en los datos. Si se pueden recono-

cer patrones de tendencia, cíclicos o estacionales, entonces se deben seleccionar las técnicas que sean capaces de extrapolar efectivamente tales patrones.

Técnicas de pronósticos para datos estacionarios

Se definió antes una serie *estacionaria* como aquella cuyo valor medio no cambia con el paso del tiempo. Tales situaciones surgen cuando los patrones de demanda que influyen en las series son relativamente estables. Es importante reconocer que los datos estacionarios no necesariamente varían de manera aleatoria alrededor de un nivel medio. Las series estacionarias pueden estar autocorrelacionadas; sin embargo, la naturaleza de la asociación es tal que los datos no se alejan de la media para cualquier periodo ampliado. En su forma más simple, el pronóstico de una serie estacionaria implica el uso de la historia disponible de la serie para estimar su valor medio, el cual se convierte así en el pronóstico de futuros periodos. Las técnicas más avanzadas implican actualizar la historia disponible de la serie para estimar su valor medio, lo que a su vez se convierte en el pronóstico de periodos futuros. Los pronósticos pueden actualizarse conforme se dispone de nueva información. La actualización es útil cuando las estimaciones iniciales no son confiables, o cuando se cuestiona la estabilidad del promedio. En este último caso, la actualización brinda algún grado de sensibilidad a los cambios potenciales en la estructura subyacente de la serie.

Las técnicas de pronósticos estacionarias se usan en las siguientes circunstancias:

- *Los factores que generan una serie se han estabilizado, y el ambiente en el cual existe la serie permanece relativamente sin cambios.* Algunos ejemplos son el número de interrupciones semanales de una línea de ensamble que tiene una tasa de producción uniforme, las unidades vendidas de un producto o servicio en la etapa de maduración de su ciclo de vida y el número de ventas resultantes de un nivel de esfuerzo constante.
- *Se necesita un modelo muy simple debido a la falta de datos para la explicación o implementación.* Un ejemplo se presenta cuando un negocio u organización es nuevo, y hay muy poca información histórica disponible.
- *La estabilidad puede obtenerse haciendo correcciones sencillas de factores tales como crecimiento demográfico o la inflación.* Algunos ejemplos son el cambio del ingreso por el ingreso per capita, y el cambio de ventas en dólares por cantidades constantes en dólares.
- *La serie puede convertirse en una serie estable.* Algunos ejemplos son la transformación de una serie mediante el uso de logaritmos, raíces cuadradas o diferencias.
- *La serie es un grupo de errores de pronóstico de una técnica de elaboración del pronóstico que se considera adecuada.* (Véase el ejemplo 3.7 de la p. 85).

Las técnicas que deben considerarse en la elaboración de pronósticos de serie estacionaria incluyen métodos informales, métodos de promedio simple, promedio móvil y modelos autorregresivos de promedio móvil (ARMA) así como los modelos de Box-Jenkins.

Técnicas de pronósticos para datos con una tendencia

En palabras sencillas, en una serie de tiempo una *tendencia* es un crecimiento o decrecimiento persistente de larga duración. Para una serie de tiempo con tendencia, el nivel de la serie no es constante. Es común que las series de tiempo de economía muestren una tendencia.

Las técnicas de pronóstico para datos con tendencia se usan en las siguientes circunstancias:

- *Un incremento en la productividad y nueva tecnología traen consigo cambios en el estilo de vida.* Algunos ejemplos son las demandas de componentes electrónicos, las cuales aumentaron con la llegada de las computadoras, y el uso del ferrocarril, el cual disminuyó con la aparición del avión.
- *Un incremento de la población causa aumentos en la demanda de bienes y servicios.* Ejemplos de esto son los ingresos por ventas de bienes de consumo, demanda de consumo de energía y uso de materias primas.
- *El poder de compra de la moneda afecta las variables económicas debido a la inflación.* Son ejemplos los salarios, los costos de producción y los precios.
- *Incremento de aceptación en el mercado.* Un ejemplo es el periodo de crecimiento en el ciclo de vida de un producto nuevo.

Las técnicas que deberían considerarse en el pronóstico de series con tendencia incluyen modelos de promedios móviles, de suavizamiento exponencial lineal de Holt, de regresión simple, de curvas de crecimiento, los modelos exponenciales y los modelos autorregresivos integrados de promedio móvil (ARIMA)(métodos de Box-Jenkins).

Técnicas de pronósticos para datos estacionales

Con anterioridad se definió una serie *estacional* como una serie de tiempo con un patrón de cambio que se repite a sí mismo año tras año. Una forma de desarrollar pronósticos estacionales es estimar índices estacionales de la historia de la serie. Por ejemplo, en datos mensuales, hay un índice para enero, un índice para febrero, y así sucesivamente. Estos índices se utilizan después para incluir la estacionalidad en los pronósticos o para eliminar tales efectos de los valores observados. El proceso que sigue se conoce como un ajuste de la estacionalidad de los datos y se analiza junto con los métodos de descomposición de la serie de tiempo en el capítulo 5.

Las técnicas para la elaboración de pronósticos con datos estacionales se usan en las siguientes circunstancias:

- *El clima influye en la variable de interés.* Ejemplos son el consumo eléctrico, las actividades de verano e invierno (por ejemplo, deportes como el esquí), el vestido y buenas temporadas en la agricultura.
- *El calendario anual influye en la variable de interés.* Son ejemplos las ventas al menudeo influidas por las vacaciones, fines de semana de tres días y calendarios académicos.

Las técnicas que deberían considerarse cuando se elaboren pronósticos con series estacionales incluyen modelos de descomposición clásica, Census *X*-12, suavizamiento exponencial de Winter, regresión múltiple y ARIMA (métodos de Box-Jenkins).

Técnicas de pronósticos para series cíclicas

El efecto cíclico se definió antes como la fluctuación con forma de onda alrededor de una tendencia. Los patrones cíclicos son difíciles de modelar porque sus patrones generalmente son inestables. Las fluctuaciones ondulatorias hacia arriba y hacia abajo de la tendencia rara vez se repiten en intervalos de tiempo fijos, en tanto que la magnitud de las variaciones también tiende a cambiar. Los métodos de descomposición (capítulo 5) pueden ampliarse para analizar los datos cíclicos. Sin embargo, debido al comportamiento irregular de los ciclos, el análisis del componente cíclico de una serie, si existe, a menudo requiere de la obtención de indicadores de coincidencia o indicadores económicos importantes.

Las técnicas para la elaboración de pronósticos con datos cíclicos se usan en las siguientes circunstancias:

- *El ciclo de negocios influye en la variable de interés.* Son ejemplos los factores económicos, de mercado y competitivos.
- *Ocurren cambios en el gusto popular.* Son ejemplos la moda, la música y la comida.
- *Sucedan cambios en la población.* Son ejemplos las guerras, las hambrunas, las epidemias y los desastres naturales.
- *Ocurren cambios en el ciclo de vida del producto.* Son ejemplos el lanzamiento, el crecimiento, la maduración, la saturación del mercado, y su declive.

Las técnicas que deberían considerarse cuando se elaboren pronósticos con series cíclicas incluyen la descomposición clásica, los indicadores de la economía, los modelos econométricos, la regresión múltiple y modelos ARIMA (métodos de Box-Jenkins).

Otros factores a considerar en la selección de una técnica de pronóstico

El horizonte de tiempo de un pronóstico tiene una relación directa con la selección de la técnica para pronosticar. En pronósticos de corto y mediano plazos, se puede aplicar una gama de técnicas cuantitativas. Sin embargo, conforme aumenta el horizonte del pronóstico, varias de estas técnicas se vuelven menos adecuadas. Por ejemplo, los modelos de promedios móviles, suavizamiento exponencial y ARIMA son pronosticadores deficientes de los cambios económicos drásticos;

mientras que los modelos econométricos resultan más útiles. Los modelos de regresión son apropiados para el corto, el mediano y el largo plazos. Las medias, los promedios móviles, la descomposición clásica y las proyecciones de la tendencia son técnicas cuantitativas adecuadas para horizontes de tiempo de corto y mediano plazos. Las técnicas más complejas de Box-Jenkins y econométricas también son recomendables para pronósticos de corto y mediano plazos. Los métodos cualitativos se usan a menudo para los horizontes de tiempo más largos (véase el capítulo 10).

La aplicabilidad de las técnicas de elaboración del pronóstico es algo que generalmente el pronosticador determina con base en la experiencia. Los gerentes frecuentemente necesitan pronósticos para un tiempo relativamente corto. Los métodos de suavizamiento exponencial, proyección de tendencia, modelos de regresión y descomposición clásica tienen una ventaja en esta situación (véase la tabla 3-6).

Los costos asociados con el uso de la computadora ya no son parte importante en la selección de una técnica. Las computadoras de escritorio (microprocesadores) y los paquetes de software para elaboración de pronósticos se han convertido en un lugar común en muchas organizaciones. Gracias a estos desarrollos, probablemente otros criterios minimizarán las consideraciones del costo informático.

A final de cuentas, el pronóstico se presentará a la gerencia para su aprobación y uso en el proceso de planeación. Por lo tanto, una fácil comprensión e interpretación de los resultados es una consideración importante. Los modelos de regresión, la proyección de tendencias, la descomposición clásica y las técnicas de suavizamiento exponencial, todas obtienen una calificación alta en dicho criterio.

Es importante señalar que la información presentada en la tabla 3-6 debería usarse como una guía para la selección de una técnica de elaboración de pronósticos. Es buena práctica

TABLA 3-6 Selección de una técnica de pronósticos

<i>Método</i>	<i>Patrón de datos</i>	<i>Horizonte de tiempo</i>	<i>Tipo del modelo</i>	<i>Datos mínimos requeridos</i>	
				<i>No estacionales</i>	<i>Estacionales</i>
Simple	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		
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 la curva S	T	I, L	TS	10	
Modelos de Gompertz	T	I, L	TS	10	
Curvas de crecimiento	T	I, L	TS	10	
Census 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 de series de tiempo	T, S	I, L	C		6 × s

Patrón de datos: ST, estacionario; T, de tendencia; S, estacional; C, cíclico

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

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

Estacional: S, longitud de la estacionalidad

Variable: V, número de variables

intentar más de un método para la elaboración del pronóstico de un problema específico, dejando fuera algunos datos recientes y luego calcular los pronósticos de estas observaciones excluidas, usando métodos diferentes. La efectividad de los métodos para estos casos de prueba excluidos puede determinarse usando una o más de las medidas de precisión definidas en las ecuaciones 3-7 a 3-11, que se estudian más adelante. Suponiendo un ajuste de los datos adecuado, el método más exacto (aquel con el menor error de pronóstico) es una selección razonable para “mejor” método. Aunque quizá no sea el mejor método en la siguiente situación.

Evaluación empírica de métodos para pronosticar

La investigación empírica ha encontrado que la exactitud en el pronóstico de los métodos sencillos a menudo es tan buena como la de las técnicas avanzadas o estadísticamente complejas (véase Fildes y otros, 1997; Makridakis y otros, 1993; y Makridakis y Hibon, 2000). Los resultados de M3-IJF Competition, donde diferentes expertos usaron su metodología favorita para pronosticar y obtuvieron cada uno pronósticos para 3,003 series de tiempo diferentes, tendieron a apoyar este hallazgo (Makridakis y Hibon, 2000). Parecería que cuanto más compleja es una técnica estadística, mejor debería predecir patrones de series de tiempo. Por desgracia, los patrones establecidos en las series de tiempo pueden cambiar en el futuro. De modo que tener el modelo que mejor represente los datos históricos (algo que los métodos complejos hacen bien) no necesariamente garantiza mayor exactitud en las predicciones del futuro. Por supuesto, la habilidad del pronosticador también desempeña un papel importante en el desarrollo de un buen pronóstico.

La M3-IJF Competition se realizó en 1997. Los pronósticos obtenidos con diferentes técnicas se compararon en toda la muestra de 3,003 series de tiempo, evaluando la exactitud con diversas medidas de error. La finalidad del estudio de 1997 fue verificar las cuatro conclusiones principales de la M-Competition original, sobre un conjunto de datos más grande (véase Makridakis y otros, 1982). Makridakis y Hibon (2000) resumieron la última competencia como sigue:

1. Al igual que en un estudio previo, los métodos avanzados o estadísticamente complejos no necesariamente dan como resultado pronósticos más precisos, que los métodos más simples.
2. Diferentes medidas de precisión producen resultados consistentes cuando se usan para evaluar diferentes métodos de elaboración de pronósticos.
3. La combinación de tres métodos de suavizamiento exponencial supera, en promedio, los métodos individuales que se combinan y funcionan bien en comparación con otros métodos.
4. La efectividad de los diferentes métodos de elaboración de pronósticos depende de la longitud del horizonte del pronóstico y la clase de datos (anuales, trimestrales, mensuales) que se analizan. Algunos métodos funcionan con más exactitud para horizontes cortos, en tanto que otros son más adecuados para los horizontes más grandes. Algunos métodos funcionan mejor con datos anuales y otros son más apropiados para datos trimestrales y mensuales.

Como parte de la selección final, cada técnica debe evaluarse en términos de su confiabilidad y aplicabilidad al problema que tenemos, de su eficiencia de costos y exactitud comparadas con técnicas similares, así como de su aceptación por parte de la gerencia. La tabla 3-6 resume las técnicas de elaboración de pronósticos adecuadas para patrones de datos específicos. Como señalamos, esta tabla representa un punto de inicio, es decir, métodos a tomar en cuenta —para datos con ciertas características. Finalmente, cualquier método seleccionado debería supervisarse continuamente para asegurarse de que hace adecuadamente el trabajo para el cual se aplica.

MEDICIÓN DEL ERROR DE PRONÓSTICO

Debido a que las técnicas cuantitativas de elaboración de pronósticos a menudo incluyen datos de series de tiempo, se desarrolló una notación matemática para referirse a cada periodo específico. La letra Y se usa para representar una variable de serie de tiempo, a menos que haya

más de una variable. El periodo asociado con una observación se identifica como un subíndice. De modo que Y_t se refiere al valor de la serie de tiempo en el periodo t . Los datos trimestrales para Coastal Marine Corporation presentados en el ejemplo 3.5 (véase la p. 76) se escribirían como $Y_1 = 147.6$, $Y_2 = 251.8$, $Y_3 = 273.1$, ..., $Y_{52} = 281.4$.

También debe desarrollarse notación matemática para distinguir entre un valor real de la serie de tiempo y el valor del pronóstico. Un símbolo $\hat{}$ (sombrero) se coloca arriba de un valor para indicar que se trata del pronóstico. El valor del pronóstico para Y_t es $\hat{Y}_t$. La exactitud de una técnica para la elaboración de un pronóstico a menudo se juzga por la comparación entre las series originales $Y_1, Y_2, \dots$, y los valores pronosticados de las series, $\hat{Y}_1, \hat{Y}_2, \dots$.

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

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

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

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

Hay varios métodos cuya finalidad es resumir los errores generados por una técnica específica de pronósticos. La mayoría de estas medidas son el promedio de alguna función de la diferencia entre su valor real y su valor pronosticado. Estas diferencias se conocen como residuos.

Un *residuo* es la diferencia entre un valor real observado y su valor de pronóstico.

La ecuación 3.6 se usa para calcular el error o residuo de cada periodo pronosticado.

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

donde

e_t = error de pronóstico en el periodo t

Y_t = valor real en el periodo t

$\hat{Y}_t$ = valor del pronóstico en el periodo t

Un método para evaluar una técnica de pronósticos usa la suma de los errores absolutos. La desviación media absoluta (*MAD*) mide la exactitud del pronóstico, promediando las magnitudes de los errores del pronóstico (los valores absolutos de los errores). *MAD* está en las mismas unidades que la serie original, y proporciona un tamaño promedio de los “errores” sin importar la dirección. 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*) es otro método para evaluar una técnica de elaboración de pronósticos. Cada error o residuo se eleva al cuadrado; luego éstos se suman y se dividen entre el número de observaciones. Este enfoque sanciona errores grandes en la elaboración de pronósticos, ya que los errores están elevados al cuadrado, lo cual es importante porque una técnica que produce errores moderados quizá sea preferible a una que usualmente tenga pequeños errores, pero ocasionalmente produce errores extremadamente grandes. El *MSE* está dado por la ecuación 3.8.

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

La raíz cuadrada del *MSE*, o la raíz cuadrada del error cuadrado medio (*RMSE*), también se usa para evaluar los métodos de elaboración de pronósticos. Tanto la *RMSE* como la *MSE* sancio-

nan los errores grandes pero tienen las mismas unidades de la serie que se está pronosticando, de modo que su magnitud se interpreta con mayor facilidad. La *RMSE* se presenta a continuación.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (3.9)$$

A veces es más útil calcular los errores del pronóstico en términos de porcentajes en vez de cantidades. El error porcentual absoluto medio (*MAPE*) se calcula obteniendo el error absoluto de cada periodo, dividiendo éste entre el valor real observado en ese periodo y promediando estos errores porcentuales absolutos. El resultado final se multiplica después por 100 y se expresa como porcentaje. Este enfoque es útil cuando el error relativo al tamaño respectivo del valor de la serie de tiempo es importante, para la evaluación de la exactitud del pronóstico. El *MAPE* es especialmente útil cuando los valores Y_t son grandes. El *MAPE* no tiene unidades de medición (es un porcentaje) y sirve para comparar la exactitud de la misma técnica o de otras técnicas en dos series completamente diferentes. La ecuación 3.10 muestra cómo se calcula el *MAPE*.

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

Note que el *MAPE* no puede calcularse si cualquiera de las Y_t es cero.

Algunas veces es necesario determinar si el método para pronosticar está sesgado (con pronósticos consistentemente altos o bajos). En estos casos, se usa el error porcentual medio (*MPE*), el cual se calcula obteniendo el error en cada periodo, dividiendo éste entre el valor real de ese periodo y luego promediando estos errores porcentuales. El resultado usualmente se multiplica por 100 y se expresa como un porcentaje. Si el enfoque del pronóstico no tiene sesgo, el *MPE* producirá un resultado que esté cercano a cero. Si el resultado es un porcentaje negativo grande, el método de elaboración del pronóstico está sobreestimando consistentemente. Si el resultado es un porcentaje positivo grande, el método de elaboración del pronóstico está subestimando consistentemente. El *MPE* está dado por:

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} \quad (3.11)$$

La decisión para usar una técnica de elaboración de pronósticos específica se basa, en parte, en la determinación de si la técnica producirá errores en el pronóstico que se consideren lo suficientemente pequeños. En efecto, es realista esperar que una buena técnica de elaboración de pronósticos produzca errores relativamente pequeños de manera consistente.

Las cinco medidas de precisión de un pronóstico recientemente descritas son usadas para:

- Comparar la exactitud de dos (o más) técnicas diferentes.
- Medir la utilidad o confiabilidad de una técnica en particular.
- Ayudar en la búsqueda de una técnica óptima.

El ejemplo 3.6 ilustrará cómo se calcula cada una de estas mediciones del error.

Ejemplo 3.6

La tabla 3.7 presenta los datos del número diario de clientes que solicitan reparaciones, Y_t , y un pronóstico de tales datos, $\hat{Y}_t$, para la estación de servicio Chevron de Gary. La técnica de elaboración del pronóstico utilizó el número de clientes atendidos en el periodo anterior como el pronóstico para el periodo actual. Esta sencilla técnica se estudiará en el capítulo 4. Se emplearon los siguientes cálculos para evaluar este modelo usando *MAD*, *MSE*, *RMSE*, *MAPE* y *MPE*.

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

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

$$RMSE = \sqrt{MSE} = \sqrt{23.5} = 4.8$$

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

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

TABLA 3-7 Datos usados en los cálculos para medir el error en el pronóstico del ejemplo 3.6

Tiempo t	Clientes Y_t	Pronóstico					
		$\hat{Y}_t$	Error e_t	$ e_t $	e_t^2	$ e_t /Y_t$	e_t/Y_t
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

La *MAD* indica que cada pronóstico está desviado un promedio de 4.3 clientes. El *MSE* de 23.5 (o la *RMSE* de 4.8) y el *MAPE* de 6.95% se compararían con el *MSE* (*RMSE*) y el *MAPE* de cualquier otro método usado para pronosticar esos datos. Finalmente, el pequeño *MPE* de 2.03% indica que la técnica no tiene sesgo: Puesto que el valor está cercano a cero, la técnica no sobreestima ni subestima consistentemente el número de clientes atendidos diariamente.

DETERMINACIÓN DE UNA TÉCNICA ADECUADA DE PRONÓSTICO

Antes de elaborar el pronóstico con una técnica particular, se debería evaluar lo adecuado de la elección. El pronosticador debe contestar las siguientes preguntas:

- ¿Los coeficientes de autocorrelación de los residuos son indicadores de una serie aleatoria? Esta pregunta puede contestarse examinando la función de autocorrelación para los residuos, tal como se presentó en el ejemplo 3.7.
- ¿La distribución de los residuos es aproximadamente normal? Esta pregunta puede contestarse analizando un histograma de los residuos o una gráfica de probabilidad normal.
- ¿Son significativas las estadísticas t para los valores estimados de los parámetros? La distribución t se revisó en el capítulo 2, y se presentan aplicaciones de las estadísticas t en el ejemplo 3.2 y en los capítulos 6 a 9.
- ¿Quiénes toman decisiones entienden la técnica y la emplean con facilidad?

El requisito básico de que el patrón residual sea aleatorio se verifica examinando los coeficientes de autocorrelación de los residuos. No debería haber coeficientes de autocorrelación significativos. El ejemplo 3.2 ilustra cómo los coeficientes de autocorrelación pueden usarse para determinar si una serie es aleatoria. El estadístico Q de Ljung-Box también se usa para

probar que las autocorrelaciones para todos los retrasos hasta el retraso K dado sean iguales a cero. El ejemplo 3.7 ilustra este procedimiento con los residuos de dos modelos ajustados.

Ejemplo 3.7

A Maggie Trimane, la analista de Sears, se le ha pedido el pronóstico de ventas para 2005. Los datos de 1995 a 2004 se presentan en la tabla 3-4. Primero, Maggie trata de pronosticar los datos usando un promedio móvil de cinco meses (véase el capítulo 4 para la descripción de esta técnica). Los residuos, las diferencias entre los valores reales y los valores pronosticados, se calculan y se guardan. Los coeficientes de autocorrelación para tales residuos se muestran en la figura 3-16. Un examen de estos coeficientes de autocorrelación indica que dos son significativamente diferentes de cero, $r_1 = .77$ y $r_2 = .58$. Los coeficientes de autocorrelación significativos indican alguna asociación o patrón en los residuos. Además, la función de autocorrelación misma tiene un patrón de coeficientes que declinan suavemente. Examinando las primeras 10 autocorrelaciones como un grupo, vemos que el estadístico Q para 10 retrasos es 73.90, mucho mayor que el valor 0.05 superior de una variable chi cuadrada, con 10 grados de libertad, 18.3. La hipótesis de que las primeras 10 autocorrelaciones son consistentes con aquéllas de una serie aleatoria se rechaza claramente para el nivel de 5%. Puesto que uno de los requisitos básicos de una buena técnica de pronósticos es que brinde un residuo o una serie de error que sea esencialmente aleatorio, Maggie juzga que la técnica del promedio móvil de cinco meses es inadecuada.

Ahora Maggie prueba con el suavizamiento exponencial lineal de Holt (véase el capítulo 4 para su descripción). La función de autocorrelación para la serie de residuos generada con esta técnica se presenta en la figura 3-17. Un examen de estos coeficientes de autocorrelación indica que ninguno es

Función de autocorrelación: Residuos MA

Retraso	ACF	T	LBQ
1	0.767355	5.15	28.30
2	0.577829	2.63	44.73
3	0.452556	1.80	55.04
4	0.302645	1.13	59.77
5	0.242298	0.88	62.87
6	0.260305	0.93	66.54
7	0.263140	0.92	70.40
8	0.213366	0.73	73.00
9	0.092468	0.31	73.50
10	-0.080688	-0.27	73.90

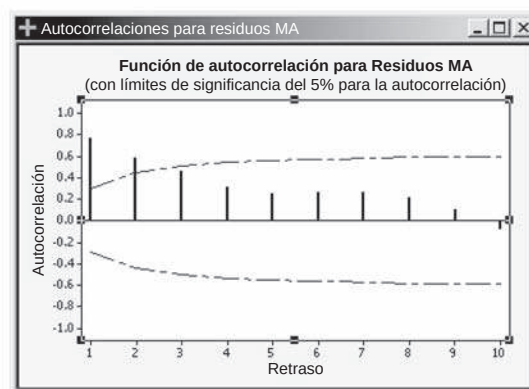


FIGURA 3-16 Función de autocorrelación para residuos con un patrón del ejemplo 3.7

Función de autocorrelación: Residuos Holt

Retraso	ACF	T	LBQ
1	-0.010799	-0.08	0.01
2	-0.045018	-0.32	0.12
3	0.180281	1.27	1.91
4	-0.104552	-0.71	2.53
5	-0.214739	-1.45	5.20
6	-0.031463	-0.20	5.25
7	0.113653	0.74	6.04
8	0.091499	0.59	6.55
9	0.003889	0.02	6.55
10	0.114144	0.73	7.40

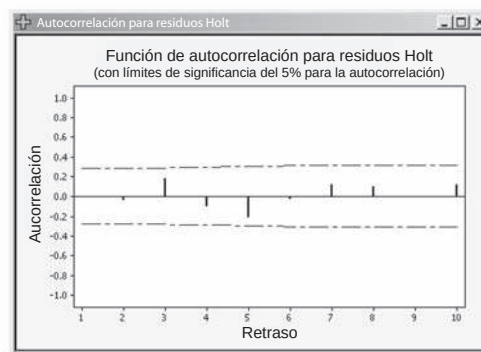


FIGURA 3-17 Función de autocorrelación para residuos que son esencialmente aleatorios del ejemplo 3.7

significativamente diferente de cero (al nivel de 5%). También se examina el estadístico Q para 10 retrasos de tiempo. El valor LBQ de 7.40 en el resultado Minitab es menor que el valor .05 superior de una variable chi cuadrada, con ocho grados de libertad, 15.5. (En este caso, los grados de libertad son iguales al número de retrasos que se prueban, menos el número de parámetros en el modelo de suavización exponencial lineal que se ha ajustado a los datos). Como grupo, las primeras 10 autocorrelaciones residuales no son diferentes de aquellas de una serie completamente aleatoria. Maggie decide considerar la técnica de suavizamiento exponencial lineal de Holt, como un posible modelo para elaborar el pronóstico de los ingresos operativos para Sears en 2005.

APLICACIÓN EN LA ADMINISTRACIÓN

Los conceptos en este capítulo brindan una base para la selección de una técnica de elaboración de pronósticos adecuada en una situación dada. Muchas de las técnicas más importantes para pronosticar se analizan y aplican en las situaciones de pronósticos en los siguientes capítulos. Es importante observar que en muchas situaciones prácticas, más de un método o modelo para pronosticar puede dar como pronósticos aceptables y casi sin diferencias. De hecho, se recomienda intentar varias técnicas razonables para pronosticar. A menudo se debe usar un criterio con base en la facilidad de uso, el costo, las condiciones ambientales externas, etcétera, para elegir un conjunto de pronósticos específico de entre, digamos, dos conjuntos de valores casi sin diferencias.

Los siguientes son unos cuantos ejemplos de situaciones que se presentan constantemente en el mundo de los negocios, donde una buena técnica para elaboración de pronósticos ayudaría en el proceso de toma de decisiones:

- Una compañía de bebidas gaseosas quiere proyectar la demanda mensual de su principal producto durante los próximos dos años.
- Una compañía de telecomunicaciones líder quiere pronosticar los pagos de dividendos trimestrales de su principal competidor por los próximos tres años.
- Una universidad necesita pronosticar las horas-crédito trimestrales de los estudiantes por los siguientes cuatro años, para desarrollar proyecciones del presupuesto para la legislatura estatal.
- Una empresa de contabilidad pública necesita pronósticos mensuales de ingresos por facturación, de modo que pueda planear las vacantes contables e iniciar el reclutamiento.
- El gerente de control de calidad de una fábrica que produce perfiles de aluminio necesita un pronóstico semanal de los defectos de producción para la alta gerencia de la compañía.
- Un banquero quiere ver las proyecciones mensuales de ingresos de un pequeño fabricante de bicicletas, quien está solicitando un crédito grande para triplicar su capacidad de producción.
- Una institución del gobierno federal necesita proyecciones anuales del promedio de millas por galón de los automóviles hechos en Estados Unidos durante los próximos 10 años, para efectuar recomendaciones regulatorias.
- Un gerente de recursos humanos necesita un pronóstico mensual de días de ausentismo en la fuerza laboral de la compañía para planear los gastos por horas extras.
- Un negocio de préstamos y ahorro requiere un pronóstico de préstamos incobrables por los siguientes dos años, en un intento por evitar la bancarrota.
- Una compañía que fabrica chips para computadora necesita un pronóstico de la industria sobre el número de computadoras personales vendidas para los próximos 5 años, con la finalidad de planear su presupuesto de investigación y desarrollo.
- Una compañía de Internet necesita pronosticar las solicitudes de servicio por los siguientes seis meses, para elaborar planes de reclutamiento de personal para sus *call centers*.

Glosario

Autocorrelación. Es la correlación entre una variable atrasada uno o más periodos consigo misma.

Correlograma o función de autocorrelación. Es una gráfica de las autocorrelaciones para varios retrasos en una serie de tiempo.

Corte transversal. Son observaciones recopiladas en un único periodo de tiempo.

Componente cíclico. Es la fluctuación ondulatoria alrededor de la tendencia.

Residuo. Es la diferencia entre un valor real y su valor pronosticado.

Componente estacional. Es un patrón de cambio que se repite año tras año.

Serie estacionaria. Es aquella cuyas propiedades estadísticas básicas, como la media y la varianza, permanecen constantes a través del tiempo.

Serie de tiempo. Una serie de tiempo consiste en datos que se recolectan, se registran o se observan durante incrementos de tiempo sucesivos.

Tendencia. Es el componente de largo plazo que representa el crecimiento o el decrecimiento en la serie de tiempo durante un periodo largo.

Fórmulas clave

Coefficiente de autocorrelación de 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

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

Estadístico Q de Ljung-Box

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

Modelo aleatorio

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

Estadístico t para probar la significancia de la autocorrelación del retraso 1

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

Error de pronóstico o residuo

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

Desviación media absoluta

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

Error cuadrático medio

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

Raíz cuadrada del error cuadrático medio

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (3.9)$$

Error porcentual absoluto medio

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

Error porcentual medio

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t} \quad (3.11)$$

Problemas

1. Explique las diferencias entre las técnicas de pronósticos cualitativas y cuantitativas.
2. ¿Qué es una serie de tiempo?
3. Describa cada uno de los componentes de una serie de tiempo.
4. ¿Qué es la autocorrelación?
5. ¿Qué mide un coeficiente de autocorrelación?
6. Explique cómo se usan los correlogramas para analizar las autocorrelaciones de varios retrasos de una serie de tiempo.
7. Indique si cada uno de los siguientes enunciados describe una serie estacionaria o una serie no estacionaria.
 - a) Una serie que muestra una tendencia
 - b) Una serie cuya media y varianza permanecen constantes en el tiempo
 - c) Una serie cuyo valor de la media cambia con el tiempo
 - d) Una serie que no tiene crecimiento ni decrecimiento
8. En seguida se presentan descripciones de varios tipos de series: aleatoria, estacionaria, de tendencia y estacional. Identifique el tipo de serie que describe cada inciso.
 - a) La serie tiene propiedades estadísticas básicas, como media y varianza, que permanecen constantes en el tiempo.
 - b) Los valores sucesivos de una serie de tiempo no están relacionados entre sí.
 - c) Existe una relación significativa entre cada valor sucesivo de una serie.

- d) Un coeficiente de autocorrelación significativo aparece en el retraso de tiempo 4 para datos trimestrales.
 - e) La serie no tiene crecimiento ni decrecimiento
 - f) En general, los coeficientes de autocorrelación son significativamente diferentes de cero, para varios de los primeros retrasos de tiempo y luego disminuyen gradualmente hacia cero, conforme se incrementa el número de retrasos.
9. Liste algunas de las técnicas de pronósticos que deban considerarse cuando se está pronosticando una serie estacionaria. Dé ejemplos de situaciones en que dichas técnicas serían aplicables.
 10. Liste algunas de las técnicas de pronósticos que deban considerarse cuando se está pronosticando una serie de tendencia. Mencione ejemplos de situaciones en que dichas técnicas serían aplicables.
 11. Liste algunas de las técnicas de pronósticos que deban considerarse cuando se está pronosticando una serie estacional. Mencione ejemplos de situaciones en que dichas técnicas serían aplicables.
 12. Liste algunas de las técnicas de pronósticos que deban considerarse cuando se está pronosticando una serie cíclica. Mencione ejemplos de situaciones en que dichas técnicas serían aplicables.
 13. En la tabla P-13, se presenta el número de matrimonios en Estados Unidos. Calcule las primeras diferencias para tales datos. Grafique los datos originales y los datos de diferencia como una serie de tiempo. ¿Hay tendencia en cualquiera de estas series? Explique su respuesta.
 14. Calcule un intervalo de confianza de 95% para el coeficiente de autocorrelación para el retraso de tiempo 1 para una serie que contiene 80 elementos.
 15. ¿Qué medida de precisión del pronóstico debería usarse en cada una de las siguientes situaciones?
 - a) El analista necesita determinar si el método de elaboración del pronóstico tiene un sesgo.
 - b) El analista cree que el tamaño o la magnitud de la variable del pronóstico es importante en la evaluación de la exactitud de éste.
 - c) El analista necesita sancionar errores grandes del pronóstico.

TABLA P-13

<i>Año</i>	<i>Matrimonios (miles)</i>	<i>Año</i>	<i>Matrimonios (miles)</i>
1985	2,413	1995	2,336
1986	2,407	1996	2,344
1987	2,403	1997	2,384
1988	2,396	1998	2,244
1989	2,403	1999	2,358
1990	2,443	2000	2,329
1991	2,371	2001	2,345
1992	2,362	2002	2,254
1993	2,334	2003	2,245
1994	2,362	2004	2,279

Fuente: Basada en *Statistical Abstract of the United States*, varios años.

16. ¿Cuál de los siguientes enunciados es verdadero en relación con las medidas de precisión usadas para evaluar pronósticos?
 - a) El *MAPE* toma en consideración la magnitud de los valores que se pronostican.
 - b) El *MSE* y la *RMSE* sancionan errores grandes.
 - c) El *MPE* se usa para determinar si, sistemáticamente, un modelo está prediciendo muy alto o muy bajo.
 - d) La ventaja del método de la *MAD* es que relaciona el tamaño del error con la observación real.
17. A Allie White, la jefa de la oficina de préstamos del Dominion Bank, le gustaría analizar el portafolios de préstamos del banco para los años 2001 a 2006. Los datos se presentan en la tabla P-17.
 - a) Calcule las autocorrelaciones para los retrasos de tiempo 1 y 2. Haga una prueba para determinar si estos coeficientes de autocorrelación son significativamente diferentes de cero, para el nivel de significancia de .05.
 - b) Use un programa de computadora para graficar los datos y calcule las autocorrelaciones para los primeros seis retrasos de tiempo. ¿Esta serie de tiempo es estacionaria?

TABLA P-17 Préstamos por trimestre (\$ millones) del Dominion Bank, 2001 a 2006				
<i>Calendario</i>	<i>Mar. 31</i>	<i>Jun. 30</i>	<i>Sep. 30</i>	<i>Dic. 31</i>
2001	2,313	2,495	2,609	2,792
2002	2,860	3,099	3,202	3,161
2003	3,399	3,471	3,545	3,851
2004	4,458	4,850	5,093	5,318
2005	5,756	6,013	6,158	6,289
2006	6,369	6,568	6,646	6,861

Fuente: Basado en los registros del Dominion Bank.

18. Para el problema anterior. Calcule las primeras diferencias de los datos de préstamos trimestrales del Dominion Bank.
 - a) Calcule el coeficiente de autocorrelación para el retraso de tiempo 1, usando los datos diferenciados.
 - b) Use un programa de cómputo para graficar los datos diferenciados y para calcular sus autocorrelaciones de los primeros seis retrasos de tiempo. ¿Esta serie es estacionaria?
19. Analice los coeficientes de autocorrelación para las series presentadas en las figuras 3-18 a 3-21. Describa brevemente cada serie.
20. A un analista le gustaría determinar si existe un patrón de ganancias por acción para la Price Company, la cual opera un negocio de autoservicio de venta al mayoreo/menudeo en varios estados con el nombre de Price Club. Los datos se presentan en la tabla P-20. Describa cualquiera de los patrones que existan en estos datos.
 - a) Encuentre el pronóstico de las ganancias trimestrales por acción de Price Club para cada trimestre, usando el enfoque sencillo (p. ej., el pronóstico del primer trimestre de 1987 es el valor del cuarto trimestre de 1986, es decir, .32).
 - b) Evalúe el pronóstico sencillo usando la *MAD*.

- c) Evalúe el pronóstico sencillo usando el MSE y la $RMSE$.
- d) Evalúe el pronóstico sencillo usando el $MAPE$.
- e) Evalúe el pronóstico sencillo usando el MPE .
- f) Escriba un reporte que resuma sus hallazgos.

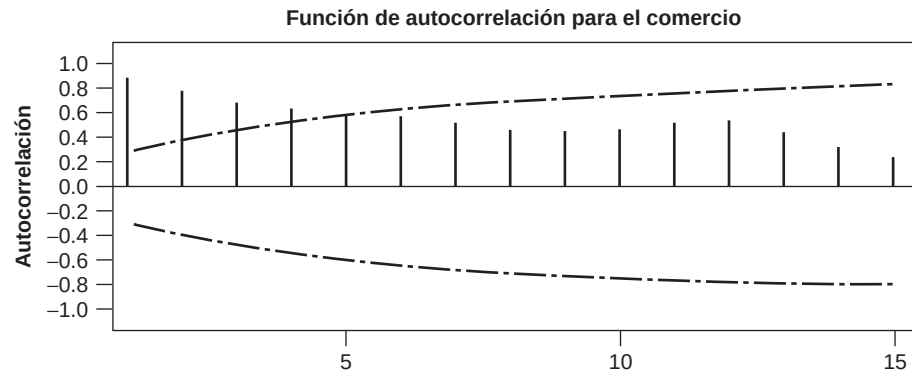


FIGURA 3-18 Función de autocorrelación para el problema 19

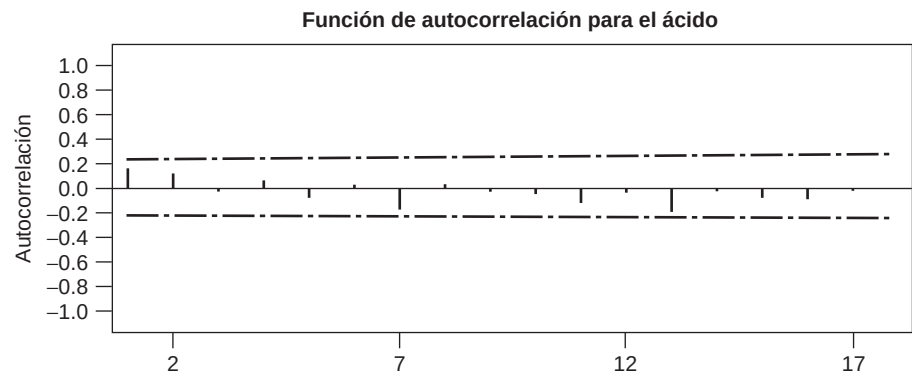


FIGURA 3-19 Función de autocorrelación para el problema 19

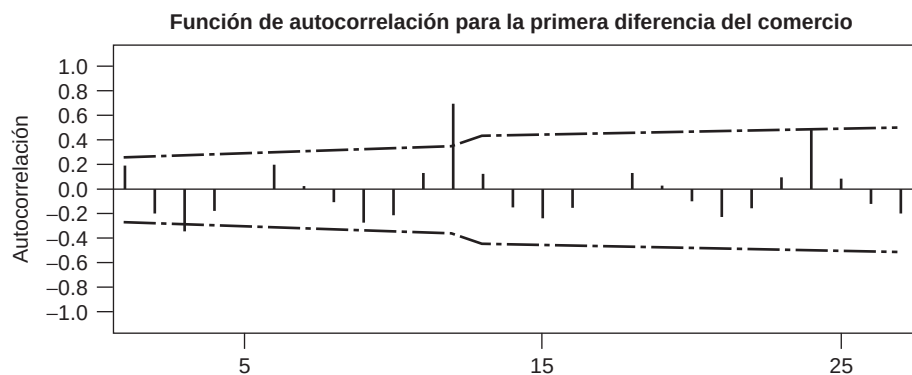


FIGURA 3-20 Función de autocorrelación para el problema 19

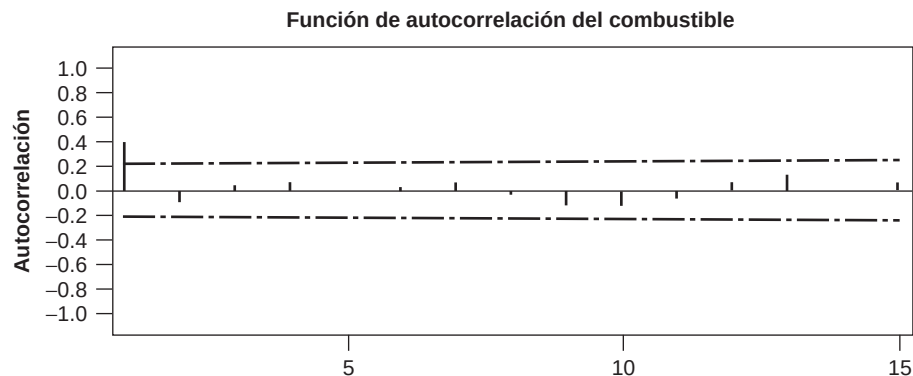


FIGURA 3-21 Función de autocorrelación para el problema 19

TABLA P-20 Ganancias trimestrales por acción de Price Club: trimestres de 1986 a 1993				
Año	1o	2o	3o	4o
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 alimenticio							
(Lea 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 semanales de un artículo alimenticio para 52 semanas consecutivas.
- a) Grafique los datos de ventas como una serie de tiempo.
 - b) ¿Cree que esta serie sea estacionaria o no estacionaria?
 - c) Usando Minitab o un programa similar, calcule las autocorrelaciones de la serie de ventas para los 10 primeros retrasos de tiempo. ¿El comportamiento de las autocorrelaciones es consistente con su respuesta al inciso b)? Explique por qué.

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, estimando c con la media muestral $\bar{Y}$ de modo que $\hat{Y}_t = \bar{Y}$. Calcule los residuos usando $e_t = Y_t - \hat{Y}_t = Y_t - \bar{Y}$.
 - Usando Minitab o un programa similar, calcule las autocorrelaciones de los residuos de la parte c para los primeros 10 retrasos de tiempo. ¿El modelo aleatorio es adecuado para los datos de ventas? Explique su respuesta.
23. La tabla P-23 presenta los ingresos trimestrales de Southwest Airlines, antes de partidas extraordinarias (millones de \$) para los años de 1988 a 1999.
- Grafique los datos de ingresos como una serie de tiempo y describa cualquier patrón existente.
 - ¿La serie es estacionaria o no estacionaria? Explique su respuesta.
 - Usando Minitab o un programa similar, calcule las autocorrelaciones de la serie de ingresos para los primeros 10 retrasos de tiempo. ¿El comportamiento de las autocorrelaciones es congruente con su elección realizada en el inciso b)? Explique su respuesta.
24. Esta pregunta se refiere al problema 23.
- Use Minitab o Excel para calcular las diferencias *cuartas* de los datos de ingresos en la tabla P-23. Las diferencias cuartas se calculan diferenciando las observaciones separadas por cuatro periodos. Con datos trimestrales, este procedimiento algunas veces es útil en la creación de una serie estacionaria, a partir de una serie no estacionaria (véase el capítulo 9). Por lo tanto, los datos diferenciados cuartos serán $Y_5 - Y_1 = 19.64 - .17 = 19.47$, $Y_6 - Y_2 = 19.24 - 15.13 = 4.11$, ..., y así sucesivamente.
 - Grafique las series de tiempo de las diferencias cuartas. ¿Estas series de tiempo parecen ser estacionarias o no estacionarias? Explique su respuesta.
25. Este problema se remite al problema 23.
- Considere un método de pronósticos sencillo, donde los ingresos del primer trimestre se usen para pronosticar los ingresos del primer trimestre del siguiente año, los ingresos del segundo trimestre se usen para pronosticar los ingresos del segundo trimestre del siguiente año, etcétera. Por ejemplo, un pronóstico de los ingresos del primer trimestre de 1998 lo proporcionan los ingresos del primer trimestre de 1997: 50.87 (véase la tabla P-23). Use este

TABLA P-23 Ingresos trimestrales de Southwest Airlines (millones de \$)				
Año	1o	2o	3o	4o
1988	0.17	15.13	26.59	16.07
1989	19.64	19.24	24.57	8.11
1990	5.09	23.53	23.04	-4.58
1991	-8.21	10.57	15.72	8.84
1992	13.48	23.48	26.89	27.17
1993	24.93	42.15	48.83	38.37
1994	41.85	58.52	58.62	20.34
1995	11.83	59.72	67.72	43.36
1996	33.00	85.32	60.86	28.16
1997	50.87	93.83	92.51	80.55
1998	70.01	133.39	129.64	100.38
1999	95.85	157.76	126.98	93.80

Fuente: Basada en *Compustat Industrial Quarterly Data Base*.

método sencillo para calcular los pronósticos de los ingresos trimestrales para los años 1998 a 1999.

b) Usando los pronósticos del inciso a), calcule *MAD*, *RMSE* y *MAPE*.

c) Dados los resultados en el inciso b) y la naturaleza de los patrones en la serie de ingresos, ¿cree usted que sea viable el método del pronóstico sencillo?
¿Pensaría en otro método sencillo que fuera mejor?

CASOS

CASO 3-1A MURPHY BROTHERS FURNITURE

En 1958 los hermanos Murphy abrieron una tienda de muebles en el centro de Dallas. Tuvieron mucho éxito durante varios años y ampliaron su cobertura de menudeo al oeste y al medio oeste. Para 1996 la cadena de sus mueblerías tenía presencia en 36 estados.

Julie Murphy, la hija de uno de los fundadores, se incorporó recientemente a la empresa. Su papá y su tío tenían buenos conocimientos en muchas áreas, aunque no en el área de las habilidades cuantitativas. En particular, ambos creían que no serían capaces de pronosticar con exactitud las futuras ventas de Murphy Brothers usando las técnicas computacionales modernas. Por tal razón, pidieron a Julie que los ayudara como parte de su nuevo empleo.

Julie pensó primero en usar las ventas en dólares de Murphy como su variable; pero se encontró con que faltaban varios años de información. Consultó a su padre, Glen, acerca de esto y él le dijo que por ese entonces él “creía que no era importante”. Julie explicó a Glen la importancia de los datos históricos y le pidió que guardara los datos en el futuro.

Julie pensó que las ventas de Murphy probablemente estaban muy relacionadas con las cifras de ventas nacionales y decidió buscar una variable adecuada en una de las diversas publicaciones federales. Después

de buscar en un número reciente de *Survey and Current Business*, encontró la historia de ventas mensuales de todas las tiendas minoristas en Estados Unidos, y decidió usar esta variable como sustituto de la variable que le interesaba: las ventas en dólares de Murphy Brothers. Ella creía que si era capaz de obtener pronósticos exactos de las ventas nacionales, podría relacionar tales pronósticos con las ventas propias de Murphy y conseguir los pronósticos que requería.

La tabla 3-8 presenta los datos que recopiló Julie, y la figura 3-22 ilustra una gráfica con los datos obtenidos con el programa de cómputo de Julie. Ella inició su análisis usando la computadora para desarrollar una gráfica de los coeficientes de autocorrelación.

Después de revisar la función de autocorrelación de la figura 3-23, era evidente para Julie que sus datos mostraban una tendencia. Los primeros coeficientes de autocorrelación eran muy grandes, y tendían hacia cero muy lentamente en el tiempo. Para hacer estacionaria la serie, de modo que pudiera considerar varios métodos de elaboración de pronósticos, Julie decidió diferenciar primero sus datos, para saber si la tendencia se podría eliminar. La función de autocorrelación para estos datos diferenciados se presenta en la figura 3-24.

PREGUNTAS

1. ¿Qué debería concluir Julie acerca de la serie de ventas al menudeo?
2. ¿Ha avanzado mucho Julie para encontrar una técnica de elaboración de pronósticos?
3. ¿Ha avanzado mucho Julie para encontrar una técnica de elaboración de pronósticos?
4. ¿Cómo sabrá Julie cuál técnica funcionará mejor?

TABLA 3-8 Ventas mensuales (miles de millones de \$) para todas las tiendas al menudeo, 1983 a 1995

	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
Mayo	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
Sept.	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: Basada en Survey of Current Business, varios años.

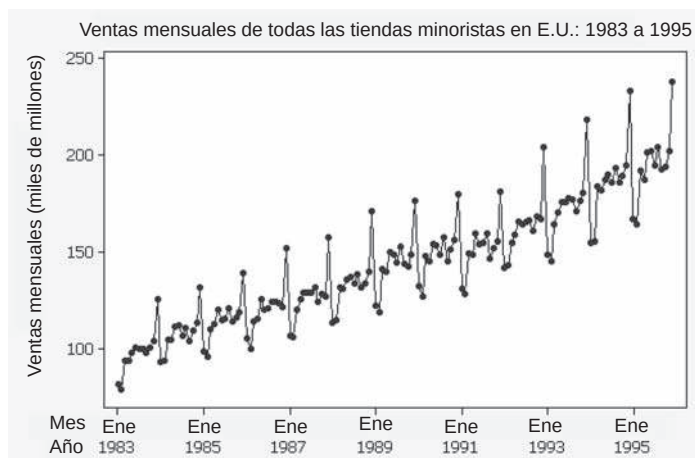


FIGURA 3-22 Gráfica de la serie de tiempo de ventas mensuales de todas las tiendas minoristas en Estados Unidos

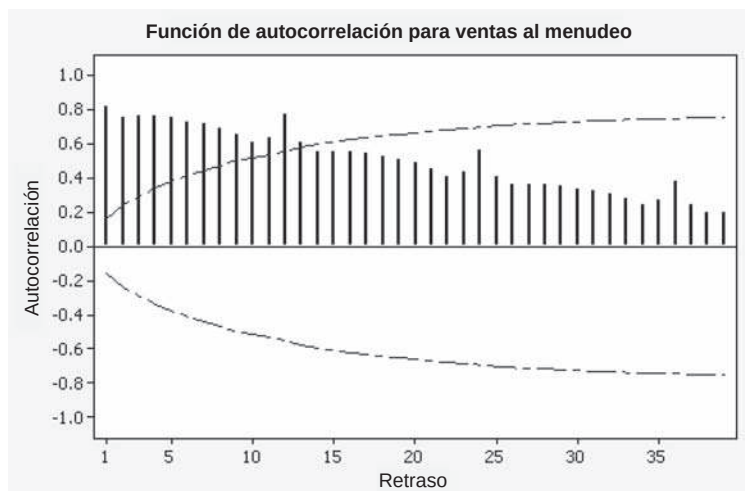


FIGURA 3-23 Función de autocorrelación para ventas mensuales de todas las tiendas minoristas en Estados Unidos

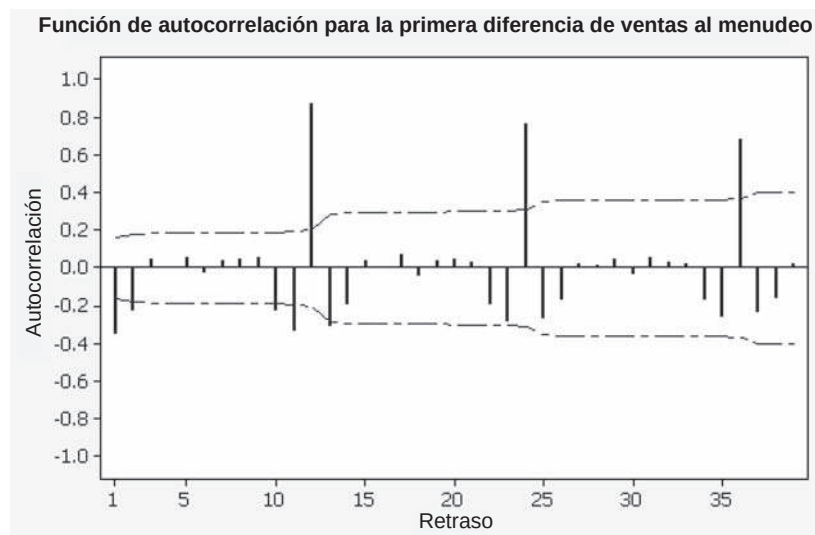


FIGURA 3-24 Función de autocorrelación de la primera diferencia para ventas mensuales en todas las tiendas minoristas de Estados Unidos

TABLA 3-9 Ventas mensuales de la mueblería Murphy Brothers, 1992 a 1995

	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>Mayo</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
1992	4,906	5,068	4,710	4,792	4,638	4,670	4,574	4,477	4,571	4,370	4,293	3,911
1993	5,389	5,507	4,727	5,030	4,926	4,847	4,814	4,744	4,844	4,769	4,483	4,120
1994	5,270	5,835	5,147	5,354	5,290	5,271	5,328	5,164	5,372	5,313	4,924	4,552
1995	6,283	6,051	5,298	5,659	5,343	5,461	5,568	5,287	5,555	5,501	5,201	4,826

Fuente: Registros de ventas de las tiendas Murphy Brothers.

CASO 3-1B MURPHY BROTHERS FURNITURE

Glen Murphy no quedó satisfecho con el regaño de su hija. Él decidió hacer una búsqueda intensiva en los registros de Murphy Brothers. Durante la investigación, él se emocionó al descubrir los datos de ventas de los pasados cuatro años, 1992 a 1995, como se presenta en

la tabla 3-9. Estaba sorprendido de averiguar que Julie no compartía su entusiasmo. Ella sabía que la obtención de datos reales de los últimos 4 años era un suceso positivo. El problema de Julie era que no estaba muy segura de qué hacer con los datos recién adquiridos.

PREGUNTAS

1. ¿Qué conclusiones debe obtener Julie acerca de los datos de ventas de Murphy Brothers?
2. ¿Cómo se compara el patrón de las ventas reales con el patrón de los datos de ventas al menudeo presentado en el caso 3-1A?
3. ¿Qué datos debería usar Julie para desarrollar un modelo de pronóstico?

CASO 3-2 MR. TUX

John Mosby, propietario de varias tiendas de alquiler Mr. Tux, está iniciando el pronóstico de la variable más importante de su negocio: las ventas mensuales en dólares (véase los casos de Mr. Tux al final de los capítulos 1 y 2). Una de sus empleadas, Virginia Perot, ha reunido los datos de ventas que se presentan en el caso 2-2. John decidió utilizar los 96 meses de datos que había recopilado. Él corrió los datos en Minitab y obtuvo la función de correlación mostrada en la figura 3-25. Puesto que todos los coeficientes de correlación son positivos y disminuyen muy lentamente, John concluyó que sus datos muestran una tendencia.

Luego, John le pidió al programa que calculara las primeras diferencias de los datos. La figura 3-26 presenta la función de autocorrelación para los datos diferenciados. Los coeficientes de autocorrelación para

los retrasos de tiempo 12 y 24, $r_{12} = .68$ y $r_{24} = .42$, respectivamente, son significativamente diferentes de cero.

Finalmente, John usa otro programa de cómputo para calcular el porcentaje de la varianza con los datos originales, explicado por los componentes de tendencia, estacionales y aleatorios.

El programa calcula el porcentaje de la varianza en los datos originales, explicado por los factores en el análisis:

FACTOR	EXPLICADO
Data	100
De tendencia	6
Estacional	45
Aleatorio	49

PREGUNTAS

1. Resuma los resultados del análisis de John en un párrafo que un gerente, no un pronosticador, pueda entender.
2. Describa los efectos de tendencia y estacionales que parecen estar presentes en los datos de ventas de Mr. Tux.
3. ¿Cómo explicaría usted la línea “Aleatoriedad, 49%”?
4. Considere las autocorrelaciones significativas, r_{12} y r_{24} , de los datos diferenciados. ¿Concluiría usted que las primeras ventas diferenciadas tienen un componente estacional? Si es así, ¿cuáles son las implicaciones para el pronóstico, digamos, de los cambios mensuales en ventas?

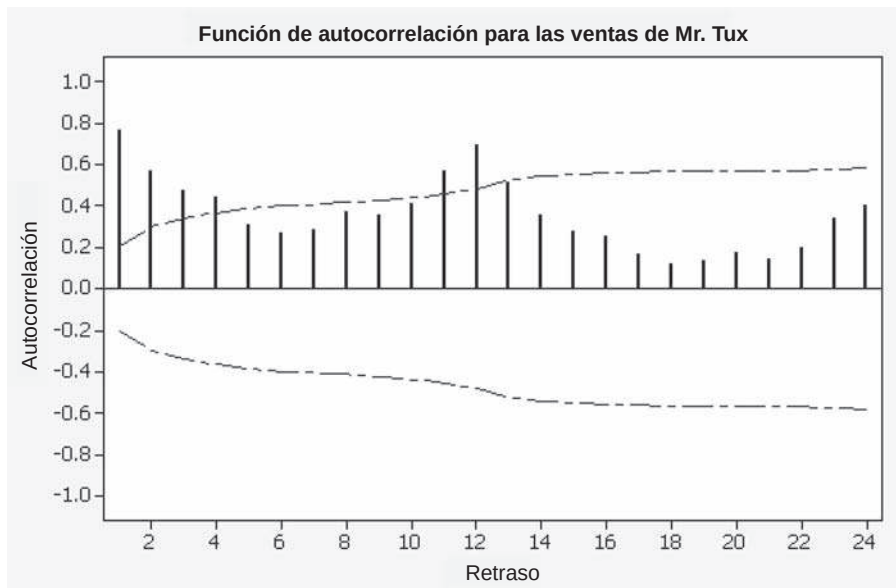


FIGURA 3-25 Función de autocorrelación para los datos de Mr. Tux

PREGUNTAS

1. Explique cómo usó Dorothy el análisis de autocorrelación para investigar el patrón de datos del número de clientes nuevos atendido por CCC.
2. ¿Qué concluyó ella después de terminar este análisis?
3. ¿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 de las operaciones de su compañía. Ella encontró varios meses de datos de ventas junto con diversas variables explicativas potenciales (repase esta situación en el ejemplo 1.1). Mientras su equipo de análisis estaba trabajando con los datos en un intento por pronosticar las ventas mensuales, se

impacientó y se preguntó cuáles de las variables explicativas eran las mejores para este propósito.

En el caso 2-3, Julie investigó las relaciones entre las ventas y las posibles variables explicativas. Se da cuenta ahora de que este paso fue prematuro porque ella incluso no conocía el patrón de las ventas (véase la tabla 3-11).

PREGUNTA

1. ¿Qué concluyó Julie acerca del patrón de datos de las ventas de Alomega?

TABLA 3-11 Ventas mensuales de las 27 tiendas Alomega, 2003 a 2006

<i>Mes</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
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
Mayo	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
Sept.	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

CASO 3-5 SURTIDO COOKIES

La compañía GAMESA es el productor más grande de galletas, botanas y pastelillos en México y América Latina.

Jaime Luna, un analista de SINTEC, una importante firma de consultoría en cadena de suministros establecida en México, está trabajando con GAMESA en un sistema de optimización de las rutas de reparto y del estacionamiento de vehículos, para todos los centros de distribución en México. Los pronósticos de la demanda de productos de GAMESA le ayudarán a elaborar un plan para la flotilla de camiones y requerimientos de almacenamiento, asociados con el sistema de optimización de las rutas de reparto y del estacionamiento de vehículos.

Como punto de partida, Jaime decidió enfocarse en uno de los productos principales de la División de Galletas de GAMESA. Él recopiló datos de la demanda agregada mensual (en kilogramos) de las galletas Surtido, de enero de 2000 a mayo del 2003. Los resultados se presentan en la tabla 3-12.

Surtido es una mezcla de diferentes galletas en una sola presentación. Jaime sabe que este producto se consume comúnmente en juntas de trabajo, fiestas y reuniones. También es popular durante las festividades navideñas. De modo que Jaime está bastante seguro de que habrá un componente estacional en la serie de tiempo de las ventas mensuales, pero no está seguro de si existirá una tendencia en las ventas. Él decidió graficar la serie de ventas y usar el análisis de autocorrelación para que le ayude a determinar si estos datos muestran una tendencia y un componente estacional.

Karen, uno de los miembros del equipo de Jaime, sugiere que los pronósticos de las ventas mensuales futuras podrían generarse usando simplemente el promedio de ventas de cada mes. Jaime decidió probar esta sugerencia “considerando” las ventas mensuales de 2003 para casos de prueba. Entonces, el promedio de ventas de enero de 2000 a 2002 se utilizará para pronosticar las ventas de enero de 2003, y así sucesivamente.

PREGUNTAS

1. ¿Qué debería concluir Jaime acerca del patrón de datos de las ventas de galletas Surtido?
2. ¿Qué aprendió Jaime acerca de la exactitud del pronóstico sugerencia de Karen?

TABLA 3-12 Ventas mensuales (en kg) de las galletas Surtido, de enero de 2000 a mayo de 2003				
Mes	2000	2001	2002	2003
Ene.	666,922	802,365	574,064	1,072,617
Feb.	559,962	567,637	521,469	510,005
Mar.	441,071	527,987	522,118	579,541
Abr.	556,265	684,457	716,624	771,350
Mayo	529,673	747,335	632,066	590,556
Jun.	564,722	658,036	633,984	
Jul.	638,531	874,679	543,636	
Ago.	478,899	793,355	596,131	
Sept.	917,045	1,819,748	1,606,798	
Oct.	1,515,695	1,600,287	1,774,832	
Nov.	1,834,695	2,177,214	2,102,863	
Dic.	1,874,515	1,759,703	1,819,749	

Aplicaciones de Minitab

El problema. En el ejemplo 3.4, Maggie Trymane, una analista de Sears, quiere pronosticar las ventas para 2005. Ella necesita determinar el patrón de los datos de ventas para los años 1955 a 2004.

Solución con Minitab

1. Introduzca los datos de Sears presentados en la tabla 3-4, en la columna C1. Puesto que los datos ya están almacenados en un archivo llamado Tab3-4.MTW, haga clic en

File>Open Worksheet

y doble clic en el archivo Ch3 de Minitab. Haga clic en Tab3-4.MTW y en Open para abrir el archivo. Los datos de Sears aparecerán en la columna C1.

2. Para construir una función de autocorrelación, haga clic en los siguientes menús, como se indica en la figura 3-27:

Stat>Time Series>Autocorrelation

3. Aparece la ventana de diálogo de la función de autocorrelación ilustrada en la figura 3-28.
 - a) Haga doble clic en la variable Revenue y ésta aparecerá a la derecha de Series.
 - b) Introduzca un encabezado en Title en el espacio adecuado y haga clic en OK. La función de autocorrelación resultante se presenta en la figura 3-11.
4. Para diferenciar los datos, haga clic en los siguientes menús:

Stat>Time Series>Differences

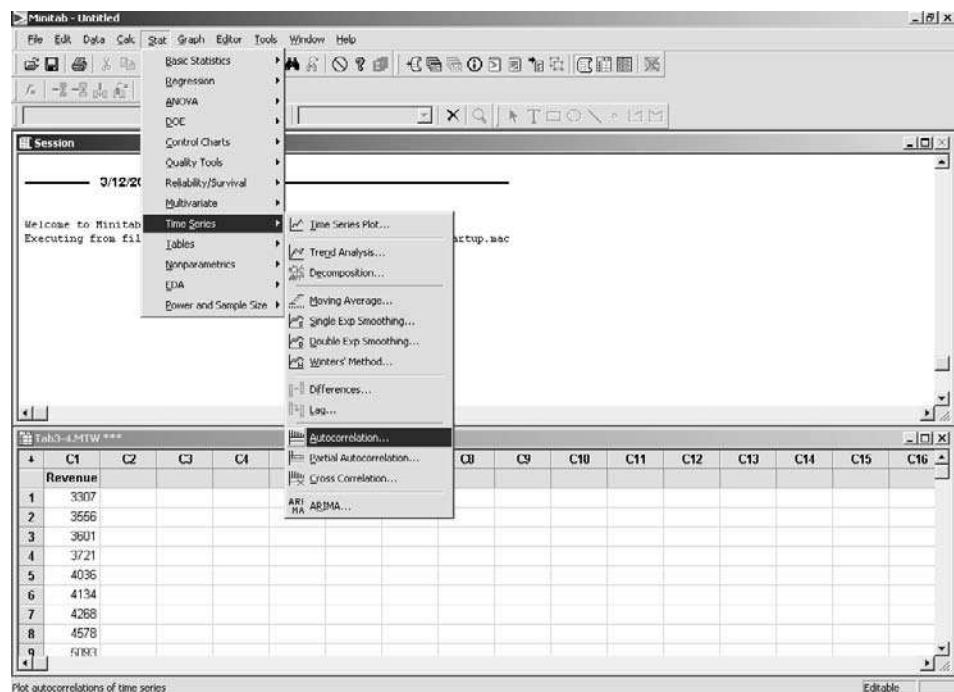


FIGURA 3-27 Menú de autocorrelación en Minitab

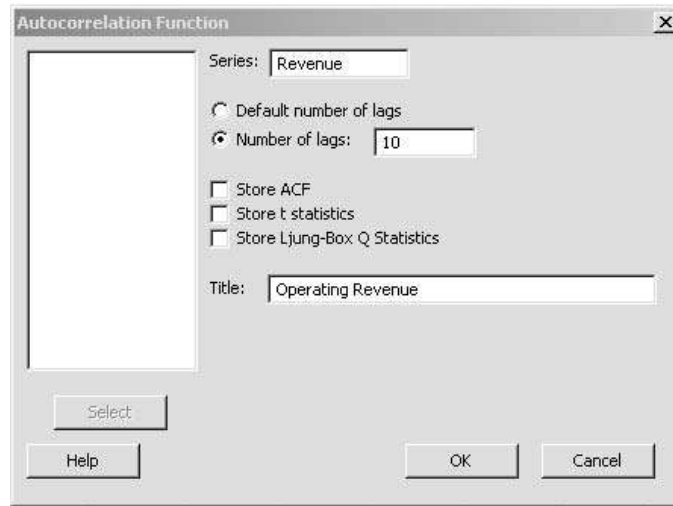


FIGURA 3-28 Ventana de diálogo de la función de autocorrelación en Minitab

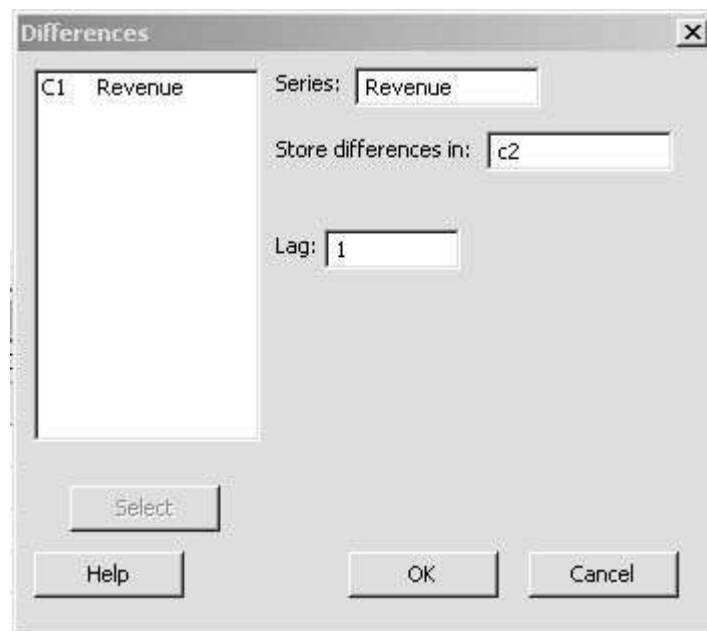


FIGURA 3-29 Ventana de diálogo de diferencias en Minitab

- La opción Differences está arriba de la opción Autocorrelation mostrada en la figura 3-27.
5. Aparece la ventana de diálogo Differences, ilustrada en la figura 3-29.
 - a) Haga doble clic en la variable Revenue y ésta aparecerá a la derecha de Series.
 - b) Utilice el tabulador para introducir las diferencias en el campo Store differences in: C2. Los datos diferenciados aparecerán ahora en la columna C2 de la hoja de cálculo.

Aplicaciones de Excel

El problema. Harry Vernon quiere usar Excel para calcular los coeficientes de autocorrelación y un correlograma para los datos presentados en la tabla 3-1. (Véase la p. 65).

Solución de Excel

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

File>New

2. Posicione el cursor en A1. Observe que, siempre que el cursor esté en una celda, ésta queda resaltada. Escriba el encabezado VERNON'S MUSIC STORE. Posicione el cursor en A2 y teclee NUMBER OF VCRS SOLD.
3. Posicione el cursor en A4 y teclee Month. Presione <enter> y la celda resaltará A5. Ahora introduzca cada mes, empezando con January (enero) en A5 y terminando con december (diciembre) en A16.
4. Posicione el cursor en B4 y teclee Y. Introduzca los datos de la tabla 3-1, empezando con la celda B5. Posicione el cursor en C4 y teclee Z.
5. Resalte las celdas B4:C16 y haga clic en los siguientes menús:

Insert>Name>Create

En la ventana de diálogo Create Names, haga clic en la ventana de diálogo Top row y luego 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 resaltándola y haciendo luego clic en el “llenado a mano”, en la esquina inferior derecha y arrástrelo hacia abajo a la celda C16. Con las celdas C5:C16 aún resaltadas, haga clic en el botón Decrease Decimal (indicado en la figura 3-30 en la mitad superior) hasta que se desplieguen tres posiciones decimales. El botón Decrease Decimal está en la barra de tareas Formatting. Esta barra de tareas puede desplegarse haciendo clic con el botón derecho del mouse en File y, luego, en Formatting con el botón izquierdo del mouse.

7. Introduzca las etiquetas LAG y ACF en las celdas E4 y F4. Para examinar los primeros seis retrasos de tiempo, introduzca los dígitos 1 a 6 en las celdas E5:E10.
8. Resalte F5 e introduzca la fórmula

$$=SUMPRODUCT(OFFSET(Z,E5,0,12-E5),OFFSET(Z,0,0,12-E5))/11$$

Resalte F5, haga clic en llenado a mano, en la esquina inferior derecha y arrástrelo a la celda F10. Con las celdas F5:F10 resaltadas, haga clic en el botón Decrease Decimal hasta que se desplieguen tres posiciones decimales. Se presentan los resultados en la figura 3-30.

9. Para desarrollar la función de autocorrelación, resalte las celdas F5:F10. Haga clic en la herramienta ChartWizard (presentada en la figura 3-31 en la mitad superior).
10. Aparece la ventana de diálogo ChartWizard pasos 1 a 4. En el paso 1, seleccione un tipo de gráfica haciendo clic en Column y luego en Next. En la ventana de diálogo del paso 2, haga clic en la ventana de diálogo Series. En el espacio siguiente a Name, teclee Corr. Dé clic en Next y aparecerá la ventana de diálogo 3. Con el título de Chart, borre Corr. Abajo del eje (X) Category, teclee Time Lags. Ahora haga clic en la ventana de diálogo Data Table y en la siguiente ventana a Show data table. Haga clic en Next para obtener la ventana de diálogo del paso 4 y luego en Finish para producir la función de autocorrelación presentada en la figura

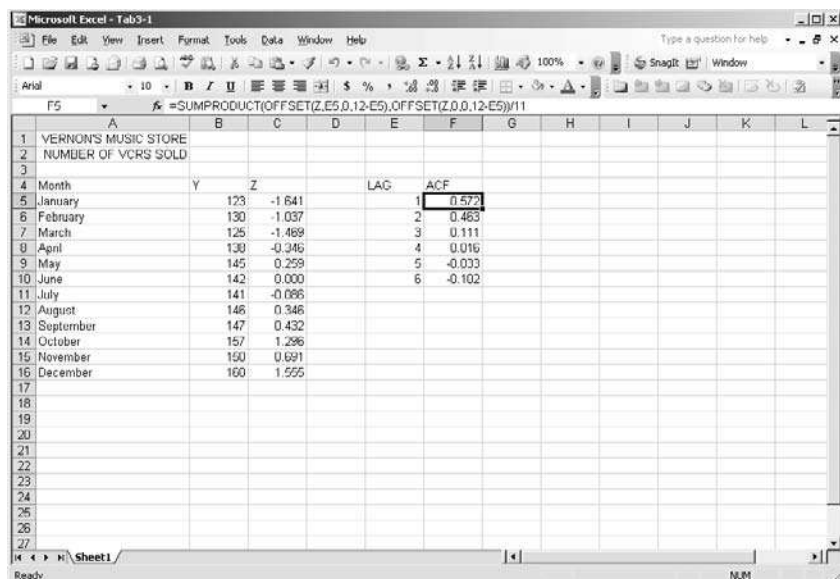


FIGURA 3-30 Hoja de cálculo Excel para los datos de las VCR

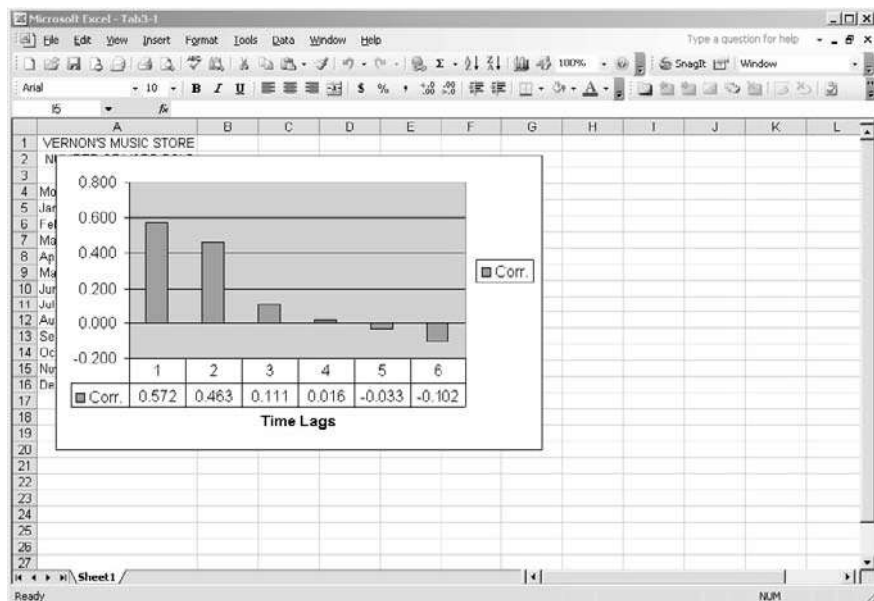


FIGURA 3-31 Resultado de Excel para la función de autocorrelación

3-31. Haga clic en una de las esquinas de la gráfica y muévela hacia el exterior para agrandar la función de autocorrelación.

11. Para salvar los datos y usarlos en el capítulo 9, haga clic en

File>Save As

En la ventana de diálogo Save As, teclee Tab3-1 en el espacio a la derecha de File name. Haga clic en Save y el archivo se guardará como Tab3-1.xls.

Referencias

- Armstrong, J. S., ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, Mass.: Kluwer, 2001.
- Diebold, F. X. *Elements of Forecasting*, 3rd ed. Cincinnati, Ohio: South-Western, 2004.
- Ermer, C. M. "Cost of Error Affects the Forecasting Model Selection." *Journal of Business Forecasting* 9 (Primavera de 1991): 10–12.
- Fildes, R., M. Hibon, S. Makridakis, and N. Meade. "The Accuracy of Extrapolative Forecasting Methods: Additional Empirical Evidence." *International Journal of Forecasting* 13 (1997): 13.
- Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and 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, J. K. Ord, and L. F. Simmons. "The M2-Competition: A Real Time Judgmentally Based Forecasting Study." *International Journal of Forecasting* 9 (1993): 5–30.
- Makridakis, S., and M. Hibon. "The M3-Competition: Results, Conclusions and Implications." *International Journal of Forecasting* 16 (2000): 451–476.
- Newbold, P., and T. Bos. *Introductory Business and Economic Forecasting*, 2nd ed. Cincinnati, Ohio: 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* 7 (Verano de 1989): 7–8.

MÉTODOS DE PROMEDIOS
MÓVILES Y DE SUAVIZACIÓN

Este capítulo describe tres enfoques sencillos para pronosticar una serie de tiempo: el método informal, el método de promedios y el método de suavización. Los métodos *informales* se usan para desarrollar modelos simples que suponen que los datos más recientes ofrecen las mejores predicciones del futuro. Los métodos *de promedio* generan pronósticos con base en un promedio de observaciones pasadas. Los métodos *de suavización* generan pronósticos con base en el promedio de valores pasados de una serie con una serie decreciente (exponencial) de ponderación o pesos.

La figura 4-1 presenta los métodos de pronósticos que se exponen en este capítulo. Visualícese usted mismo en una escala de tiempo. Suponga que se encuentra en el punto t de la figura 4-1 y puede mirar hacia atrás o hacia delante y observar la información pasada o futura de la variable de interés (Y_t). Después de que usted selecciona una técnica para la elaboración del pronóstico, la ajusta a los datos conocidos y obtiene los valores del pronóstico ($\hat{Y}_t$). Una vez que estos valores pronosticados están disponibles, los compara con las observaciones conocidas y calcula el error de pronóstico (e_t).

Una buena estrategia para evaluar los métodos de pronósticos implica los siguientes pasos:

- 1. La selección de un método de pronósticos se realiza con base en el análisis y la intuición del pronosticador acerca de la naturaleza de los datos.
- 2. El conjunto de datos se divide en dos secciones: una sección de inicio o de ajuste y una sección de prueba o de pronóstico.
- 3. La técnica seleccionada de pronóstico se usa para desarrollar valores de ajuste con la primera parte de los datos.
- 4. La técnica se usa para pronosticar la segunda sección, los pronósticos obtenidos se comparan con los datos y se evalúa el error de pronóstico (consulte el capítulo 3 para revisar las medidas de la precisión del pronóstico).
- 5. Se toma una decisión. La decisión puede ser usar la técnica en su forma actual, modificarla o desarrollar un pronóstico usando otra técnica y comparar los resultados.

FIGURA 4-1 Esquema de elaboración del pronóstico

	Datos pasados	Usted está aquí t	Periodos por pronosticar
	$\dots Y_{t-3}, Y_{t-2}, Y_{t-1},$	$Y_t,$	$\hat{Y}_{t+1}, \hat{Y}_{t+2}, \hat{Y}_{t+3}, \dots$
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

A menudo los negocios nuevos enfrentan el problema de tratar de elaborar un pronóstico con grupos muy reducidos de datos. Esta situación genera un problema real, puesto que muchas técnicas para la elaboración de un pronóstico requieren grandes cantidades de datos. Los pronósticos informales son una posible solución, puesto que se basan solamente en la información más reciente que está disponible.

Los pronósticos informales suponen que los periodos recientes son los mejores para predecir el futuro. El modelo más sencillo es:

$$\hat{Y}_{t+1} = Y_t \quad (4.1)$$

donde $\hat{Y}_{t+1}$ es el pronóstico realizado en el momento t (el origen del pronóstico) para el tiempo $t + 1$.

El pronóstico informal de cada periodo es la observación inmediata anterior. Se da el 100% del peso al valor actual de la serie. El pronóstico informal se conoce algunas veces como pronóstico “sin cambio”. En el pronóstico meteorológico de corto plazo a menudo se aplica el pronóstico “sin cambio”. Se supone que el clima de mañana se parecerá mucho al clima de hoy.

Puesto que el pronóstico informal (ecuación 4.1) descarta cualquier otra observación, este esquema detecta cambios muy rápidamente. El problema con este enfoque es que las fluctuaciones aleatorias se producen con la misma fidelidad que los cambios en los datos fundamentales.

Ejemplo 4.1

La figura 4-2 muestra las ventas trimestrales de serruchos del 2000 al 2006 de Acme Tool Company. Se usa la técnica informal para pronosticar que las ventas del siguiente trimestre serán iguales que las del trimestre anterior. La tabla 4-1 presenta los datos del 2000 al 2006. Si se usan los datos del 2000 al 2005 como parte inicial (de ajustes) y los datos del 2006 como la parte de prueba, el pronóstico para el primer trimestre del 2006 es:

$$\begin{aligned}\hat{Y}_{24+1} &= Y_{24} \\ \hat{Y}_{25} &= 650\end{aligned}$$

FIGURA 4-2 Gráfica de las series de tiempo de las ventas de serruchos de Acme Tool Company, 2000 a 2006, para el ejemplo 4.1

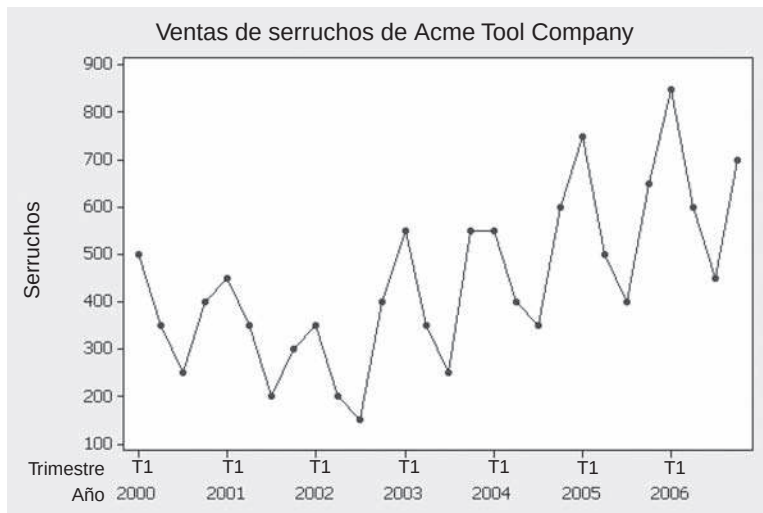


TABLA 4-1 Ventas de serruchos de Acme Tool Company, 2000 a 2006, para el ejemplo 4.1

<i>Año</i>	<i>Trimestre</i>	<i>t</i>	<i>Ventas</i>
2000	1	1	500
	2	2	350
	3	3	250
	4	4	400
2001	1	5	450
	2	6	350
	3	7	200
	4	8	300
2002	1	9	350
	2	10	200
	3	11	150
	4	12	400
2003	1	13	550
	2	14	350
	3	15	250
	4	16	550
2004	1	17	550
	2	18	400
	3	19	350
	4	20	600
2005	1	21	750
	2	22	500
	3	23	400
	4	24	650
2006	1	25	850
	2	26	600
	3	27	450
	4	28	700

El error de pronóstico se determina usando la ecuación 3.6. El error del periodo 25 es:

$$e_{25} = Y_{25} - \hat{Y}_{25} = 850 - 650 = 200$$

De modo similar, el pronóstico del periodo 26 es 850, con un error de -250 . La figura 4-2 indica que estos datos tienen una tendencia ascendente y que parece haber un patrón estacional (el primer trimestre y el cuarto son relativamente altos), de manera que se toma una decisión para modificar el modelo informal.

Un examen de los datos del ejemplo 4.1 nos lleva a concluir que los valores se incrementan con el tiempo. Cuando los valores de los datos aumentan con el tiempo, se dice que son de nivel *no estacionario* o que tienen una *tendencia*. Si se usa la ecuación 4.1, las proyecciones serán bajas de forma sistemática. Sin embargo, la técnica puede ajustarse para tomar en cuenta la tendencia sumando la diferencia entre este periodo y el último. La ecuación del pronóstico es:

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \quad (4.2)$$

La ecuación 4.2 toma en cuenta la magnitud del cambio que ocurre entre trimestres.

Ejemplo 4.1 (cont.)

Usando la ecuación 4.2, el pronóstico del primer trimestre del 2006 es:

$$\begin{aligned}\hat{Y}_{24+1} &= Y_{24} + (Y_{24} - Y_{24-1}) \\ \hat{Y}_{25} &= Y_{24} + (Y_{24} - Y_{23}) \\ \hat{Y}_{25} &= 650 + (650 - 400) \\ \hat{Y}_{25} &= 650 + 250 = 900\end{aligned}$$

El error de pronóstico con este modelo es:

$$e_{25} = Y_{25} - \hat{Y} = 850 - 900 = -50$$

En algunos casos, la tasa de cambio podría ser más apropiada que el valor absoluto del cambio. Si éste es el caso, es razonable elaborar el pronóstico de acuerdo con:

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} \quad (4.3)$$

Una revisión visual de los datos en la tabla 4-1 indica que puede existir una variación estacional. Las ventas en el primer trimestre y en el cuarto generalmente son mayores que las del segundo y tercer trimestres. Si el patrón estacional es fuerte, entonces una ecuación adecuada para el pronóstico de datos trimestrales sería

$$\hat{Y}_{t+1} = Y_{t-3} \quad (4.4)$$

La ecuación 4.4 indica que en el siguiente trimestre la variable tendrá el mismo valor que tenía en el trimestre correspondiente del año anterior.

La mayor debilidad de este enfoque es que ignora todo lo que ha ocurrido desde el año anterior y también cualquier tendencia. Existen varias maneras de introducir esta información más reciente. Por ejemplo, el analista puede combinar las estimaciones estacionales y de tendencia, y pronosticar el siguiente trimestre mediante:

$$\hat{Y}_{t+1} = Y_{t-3} + \frac{(Y_t - Y_{t-1}) + \dots + (Y_{t-3} - Y_{t-4})}{4} = Y_{t-3} + \frac{Y_t - Y_{t-4}}{4} \quad (4.5)$$

donde el término Y_{t-3} pronostica el patrón estacional, y el término restante promedia la magnitud del cambio de los cuatro trimestres anteriores y ofrece un estimado de la tendencia.

Los modelos de pronóstico informales de las ecuaciones 4.4 y 4.5 se aplican para datos trimestrales. Es posible hacer ajustes para los datos recolectados durante diferentes periodos de tiempo. Para datos mensuales, por ejemplo, el periodo estacional es 12, no 4, y el pronóstico para el siguiente periodo (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 modelos informales posibles están limitados sólo por el ingenio del analista, pero el uso de estas técnicas debe guiarse por el uso de un buen juicio.

Los métodos informales también se usan como base para hacer comparaciones con las cuales se evalúa el desempeño de métodos más complejos.

Ejemplo 4.1 (cont.)

Los pronósticos del primer trimestre de 2006, usando las ecuaciones 4.3, 4.4 y 4.5, son:

$$\begin{aligned}\hat{Y}_{24+1} &= Y_{24} \frac{Y_{24}}{Y_{24-1}} = Y_{24} \frac{Y_{24}}{Y_{23}} \\ \hat{Y}_{25} &= 650 \frac{650}{400} = 1,056\end{aligned} \quad (\text{Ecuación 4.3})$$

$$\begin{aligned}\hat{Y}_{24+1} &= Y_{24-3} = Y_{21} \\ \hat{Y}_{25} &= Y_{21} = 750\end{aligned} \quad (\text{Ecuación 4.4})$$

$$\begin{aligned}\hat{Y}_{24+1} &= Y_{24-3} + \frac{(Y_{24} - Y_{24-1}) + \dots + (Y_{24-3} - Y_{24-4})}{4} = Y_{24-3} + \frac{Y_{24} - Y_{2-4}}{4} \\ \hat{Y}_{25} &= Y_{21} + \frac{(Y_{24} - Y_{20})}{4} = 750 + \frac{650 - 600}{4} \\ \hat{Y}_{25} &= 750 + 12.5 = 762.5\end{aligned}\quad (\text{Ecuación 4.5})$$

MÉTODOS DE PRONÓSTICOS BASADOS EN PROMEDIOS

Con frecuencia, los administradores enfrentan el problema de actualizar de forma diaria, semanal o mensual los pronósticos de inventarios que contienen cientos o miles de artículos. Muchas veces no es posible desarrollar técnicas complejas para elaborar pronósticos de cada artículo. En vez de ello, se necesitan herramientas para pronosticar de manera sencilla, rápida, poco costosa y para el corto plazo.

Un administrador que enfrenta una situación como ésta probablemente use una técnica de promedios o una técnica de suavización. Estos tipos de técnicas usan alguna clase de promedio ponderado de observaciones pasadas para suavizar fluctuaciones de corto plazo. La suposición implícita de estas técnicas es que las fluctuaciones de los valores pasados representan puntos de partida aleatorios de alguna estructura subyacente. Una vez que se identifica esta estructura, es posible proyectarla hacia el futuro para generar un pronóstico.

Promedios simples

Los datos históricos pueden suavizarse de muchas maneras. El objetivo es usar los datos pasados para desarrollar un modelo de pronóstico para periodos futuros. En esta sección se presenta el método de promedios simples. Al igual que sucede con los métodos informales, se toma una decisión para usar los primeros puntos de datos t como la parte de inicialización y la parte restante de puntos como la parte de prueba. Luego, se usa la ecuación 4.6 para promediar (calcular la media de) la parte de inicialización de los datos y pronosticar el siguiente periodo.

$$\hat{Y}_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i \quad (4.6)$$

Cuando una nueva observación está disponible, el pronóstico del siguiente periodo, $\hat{Y}_{t+2}$, es el promedio o la media calculada empleando la ecuación 4.6 incluyendo esta nueva observación.

Cuando se pronostica un gran número de series simultáneamente (por ejemplo, en la administración de inventarios), el almacenamiento de datos suele ser un problema. La ecuación 4.7 resuelve este problema potencial. Sólo se requiere contar con el pronóstico más reciente y la observación más reciente conforme el tiempo transcurre.

$$\hat{Y}_{t+2} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t + 1} \quad (4.7)$$

El método de *promedios simples* es una técnica adecuada cuando los factores que producen la serie que se va a pronosticar se han estabilizado y el ambiente en el cual se encuentra la serie generalmente permanece sin cambios. Ejemplos de este tipo de series son: las ventas efectuadas como resultado de un nivel de esfuerzo constante de los vendedores; las ventas de un producto en la etapa de madurez de su ciclo de vida; y el número de citas semanales de un dentista, doctor o abogado cuyo número de clientes o pacientes es estable.

Un *promedio simple* usa la media de todas las observaciones históricas relevantes como el pronóstico para el siguiente periodo.

Ejemplo 4.2

La Spokane Transit Authority opera una flota de camionetas para transportar tanto personas mayores como discapacitadas. El registro de la gasolina comprada para esta flota de camionetas se presenta en la tabla 4-2. La cantidad real de gasolina consumida por una camioneta en un día cualquiera se caracteriza por la naturaleza aleatoria de las visitas y los destinos. La revisión de las compras de gasolina graficadas en la figura 4-3 indica que los datos son muy estables. Puesto que los datos parecen estacionarios, se usa el método de los promedios simples de las semanas 1 a 28 para pronosticar las compras de gasolina de las semanas 29 y 30. El pronóstico para la semana 29 es:

TABLA 4-2 Compras de gasolina de la Spokane Transit Authority para el ejemplo 4.2

Semana t	Galones Y_t	Semana t	Galones Y_t	Semana t	Galones Y_t
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

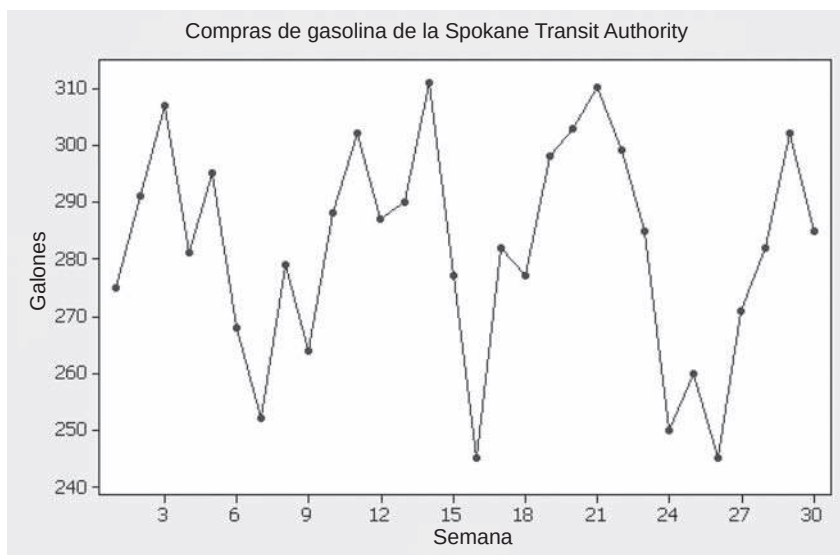


FIGURA 4-3 Gráfica de las series de tiempo de las compras semanales de gasolina de la Spokane Transit Authority para el ejemplo 4.2

$$\hat{Y}_{28+1} = \frac{1}{28} \sum_{i=1}^{28} Y_i$$

$$\hat{Y}_{29} = \frac{7,874}{28} = 281.2$$

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. El pronóstico calculado con la ecuación 4.7 es:

$$\hat{Y}_{28+2} = \frac{28\hat{Y}_{28+1} + Y_{28+1}}{28 + 1} = \frac{28\hat{Y}_{29} + Y_{29}}{29}$$

$$\hat{Y}_{20} = \frac{28(281.2) + 302}{29} = 281.9$$

El error de pronóstico es:

$$e_{30} = Y_{30} - \hat{Y}_{30} = 285 - 281.9 = 3.1$$

Usando el método de promedios simples, el pronóstico de galones de gasolina comprados para 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 sucede si en el analista está más interesado en las observaciones recientes? Se puede especificar un número constante de puntos de datos al inicio y se puede calcular una media con las observaciones más recientes. El término *promedios móviles* se utiliza para describir este enfoque. Conforme está disponible cada nueva observación, se calcula una nueva media sumando el valor más reciente y eliminando el valor más antiguo. Entonces se usa este promedio móvil para pronosticar el siguiente periodo. La ecuación 4.8 nos da el pronóstico de un 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

$\hat{Y}_{t+1}$ = valor pronosticado para el siguiente periodo

Y_t = valor real en el periodo t

k = número de términos en el promedio móvil

El promedio móvil para el periodo de tiempo t es la media aritmética de las k observaciones más recientes. En un promedio móvil, se asignan pesos iguales a cada observación. Conforme está disponible, cada nuevo punto de datos se incluye en el promedio y el punto de datos más antiguo se descarta. El porcentaje 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 promedio móvil maneja sólo los últimos k periodos de los datos conocidos; el número de puntos de datos en cada promedio no cambia conforme el tiempo avanza. El modelo de promedio móvil no maneja muy bien la tendencia o la estacionalidad, si bien es cierto que lo hace mejor que el método de promedio simple.

En un *promedio móvil*, el analista debe seleccionar el número de periodos, k . Un promedio móvil de orden 1, MA(1), tomaría la observación actual, Y_t , y la usaría para pronosticar Y para el siguiente periodo. Esto simplemente es el enfoque del pronóstico informal de la ecuación 4.1.

Un *promedio móvil* de orden k es el valor de la media de k observaciones consecutivas. El valor del promedio móvil más reciente indicará el pronóstico del siguiente periodo.

Ejemplo 4.3

La tabla 4-3 presenta la técnica del pronóstico del promedio móvil para los datos de la Spokane Transit Authority, usando un promedio móvil de cinco semanas. El pronóstico de promedio móvil para la semana 29 es:

$$\hat{Y}_{28+1} = \frac{Y_{28} + Y_{28-1} + \cdots + Y_{28-5+1}}{5}$$

$$\hat{Y}_{29} = \frac{Y_{28} + Y_{27} + Y_{26} + Y_{25} + Y_{24}}{5}$$

TABLA 4-3 Compras de gasolina de la Spokane Transit Authority para el ejemplo 4.3

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

$$\hat{Y}_{29} = \frac{282 + 271 + 245 + 260 + 250}{5} = \frac{1,308}{5} = 261.6$$

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 para la semana 31 es:

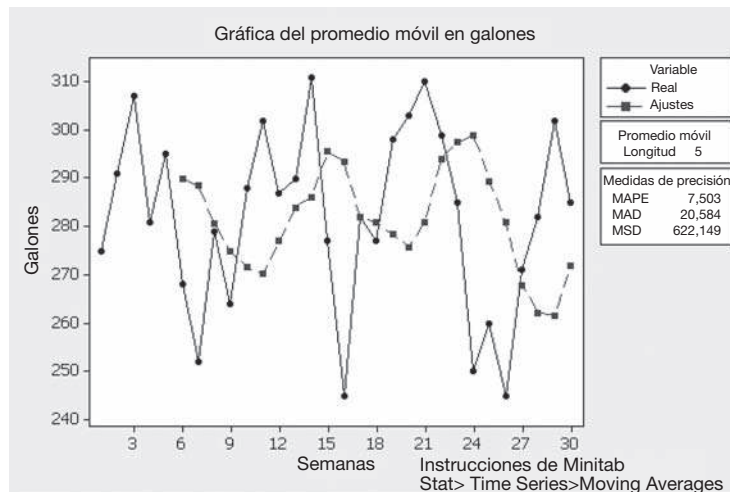
$$\begin{aligned}\hat{Y}_{30+1} &= \frac{Y_{30} + Y_{30-1} + \cdots + Y_{30-5+1}}{5} \\ \hat{Y}_{31} &= \frac{Y_{30} + Y_{29} + Y_{28} + Y_{27} + Y_{26}}{5} \\ \hat{Y}_{31} &= \frac{285 + 302 + 282 + 271 + 245}{5} = \frac{1,385}{5} = 277\end{aligned}$$

Se puede usar Minitab para calcular un promedio móvil de cinco semanas (para instrucciones específicas, véase la sección de aplicaciones de Minitab al final del capítulo). La figura 4-4 presenta el promedio móvil de cinco semanas graficado contra los datos reales, *MAPE*, *MAD* y *MSD* y las instrucciones básicas de Minitab. Observe que Minitab llama *MSD* (desviación cuadrática media) al error cuadrático medio.

La figura 4-5 presenta la función de autocorrelación de los residuos del método de promedio móvil para las cinco semanas. Los límites de error para las autocorrelaciones individuales centradas en cero y el estadístico *Q* de Ljung-Box (con seis grados de libertad, puesto que no se estiman los parámetros del modelo) indican que existen autocorrelaciones residuales significativas. Es decir, los residuos no son aleatorios. La asociación contenida en los residuos en ciertos retrasos de tiempo puede usarse para mejorar el modelo de pronóstico.

El analista debe usar su criterio en la determinación de los días, semanas, meses o trimestres en los cuales basar el promedio móvil. Cuanto menor es el número, mayor es el peso que se asigna a los periodos más recientes. Y a la inversa, cuanto mayor es el número, menor es el peso que se asigna a los periodos más recientes. Es más deseable un número pequeño cuando existen cambios repentinos en el nivel de la serie. Un número pequeño pone mayor peso a la

FIGURA 4-4 Promedio móvil de cinco semanas aplicado a las compras semanales de gasolina de la Spokane Transit Authority para el ejemplo 4.3



Función de autocorrelación: Residuos MA(5)

Retraso	ACF	T	LBQ
1	0.506290	2.53	7.21
2	0.078551	0.32	7.39
3	-0.298611	-1.21	10.13
4	-0.602830	-2.31	21.81
5	-0.642632	-2.06	35.74
6	-0.219452	-0.61	37.46

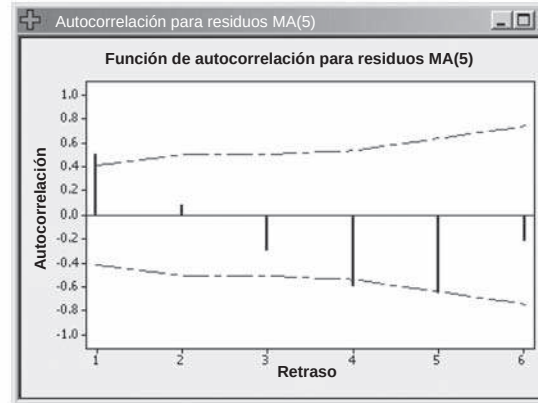


FIGURA 4-5 Función de autocorrelación para los residuos cuando se usa el método de promedio móvil de cinco semanas con datos de la Spokane Transit Authority para el ejemplo 4.3

historia reciente, lo cual facilita que los pronósticos alcancen más rápidamente el nivel real. Es deseable un número grande cuando hay fluctuaciones amplias y poco frecuentes en las series.

Los promedios móviles se usan a menudo con datos mensuales o trimestrales como ayuda para suavizar los componentes de una serie de tiempo, como se verá en el capítulo 5. Para datos trimestrales, un promedio móvil de cuatro trimestres, MA(4), genera un promedio de los cuatro trimestres, y para datos mensuales, un promedio móvil de 12 meses, MA(12), elimina o promedia los efectos estacionales. Cuanto más grande es el orden del promedio móvil, mayor es el efecto de suavización.

En el ejemplo 4.3 se usó la técnica del promedio móvil con datos estacionarios. En el ejemplo 4.4 vimos qué pasa cuando se usa el método del promedio móvil en datos con tendencia. La técnica del promedio móvil doble, la cual está diseñada para manejar datos con tendencia, se presenta a continuación.

Promedios móviles dobles

Una manera de pronosticar los datos de las series de tiempo que tienen una tendencia lineal es usar promedios móviles dobles. Este método hace lo que indica su nombre: se calcula un conjunto de promedios móviles y luego se calcula un segundo conjunto como un promedio móvil del primer conjunto.

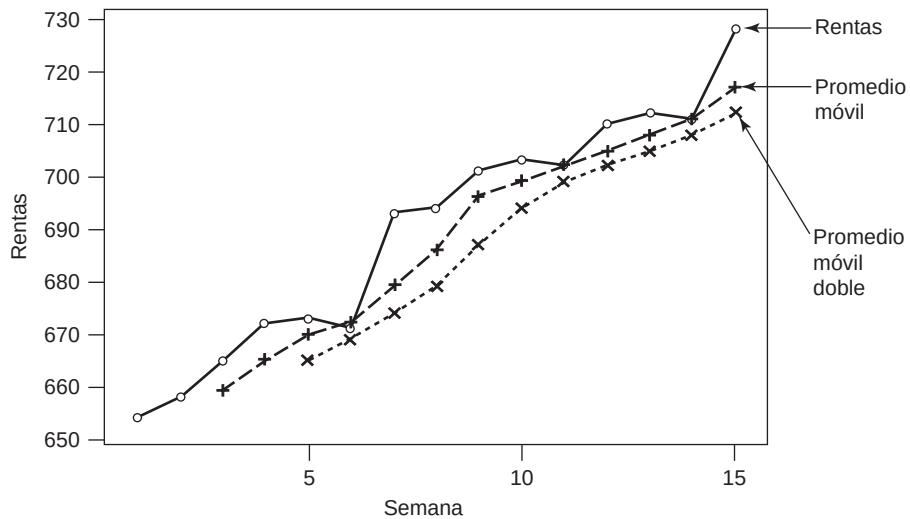
La tabla 4-4 presenta los datos de rentas semanales de Movie Video Store, junto con los resultados de la aplicación de un promedio móvil de tres semanas para pronosticar las ventas futuras. El examen de la columna de error en la tabla 4-4 indica que todas las entradas son positivas, lo cual significa que los pronósticos no alcanzaron la tendencia. El promedio móvil de tres semanas y el promedio móvil doble para estos datos se presentan en la figura 4-6. Observe cómo los promedios móviles de tres semanas se quedan retrasados con respecto a los valores reales de periodos similares. Esto ilustra lo que pasa cuando se emplea la técnica de promedios móviles con datos que muestran una tendencia. Advierta también que los pronósticos realizados mediante promedios móviles dobles se retrasan con respecto al primer conjunto de pronósticos casi tanto como éste se atrasa con respecto a los valores reales. La diferencia entre los dos conjuntos de promedios móviles se suma al promedio móvil de tres semanas para pronosticar los valores reales.

La construcción de un promedio móvil doble se resume en las ecuaciones 4.9 a 4.12. Primero, se usa la ecuación 4.8 para calcular el promedio móvil de orden k .

$$M_t = \hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \cdots + Y_{t-k+1}}{k}$$

TABLA 4-4 Rentas semanales de la Movie Video Store para el ejemplo 4.4

t	Unidades semanales rentadas Y_t	Variación total en tres semanas	Promedio móvil del pronóstico $\hat{Y}_{t+1}$	e
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	—

 $MSE = 133$ **FIGURA 4-6** Promedios móviles simple y doble de tres semanas de los datos de la Movie Video Store para el ejemplo 4.4

Entonces se usa la ecuación 4.9 para calcular el segundo promedio móvil:

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \cdots + M_{t-k+1}}{k} \quad (4.9)$$

Se emplea la ecuación 4.10 para desarrollar un pronóstico sumando al promedio móvil simple la diferencia entre el promedio móvil simple y el segundo promedio móvil.

$$a_t = M_t + (M_t - M'_t) = 2M_t - M'_t \quad (4.10)$$

La ecuación 4.11 es un factor de ajuste adicional, que es similar a la medida de cambio a lo largo de la serie:

$$b_t = \frac{2}{k-1}(M_t - M_t') \quad (4.11)$$

Finalmente, se usa la ecuación 4.12 para realizar el pronóstico de p periodos en el futuro.

$$\hat{Y}_{t+p} = a_t + b_t p \quad (4.12)$$

donde

k = número de periodos en el promedio móvil

p = número de periodos futuros por pronosticar

Ejemplo 4.4

La Movie Video Store opera varias tiendas de renta de películas en Denver, Colorado. La compañía está creciendo y necesita expandir su inventario para adaptar sus servicios a la creciente demanda. El presidente de la compañía asignó a Jill Ottenbreit la tarea de realizar un pronóstico de las rentas para el siguiente mes. Están disponibles los datos de las rentas de las últimas 15 semanas y se presentan en la tabla 4-5. En primer lugar, Jill intenta desarrollar un pronóstico usando un promedio móvil de tres semanas. El MSE para este modelo es 133 (véase la tabla 4-4). Puesto que los datos tienen una tendencia obvia, Jill descubre que sus pronósticos están subestimando de manera sistemática las rentas reales. Por esta razón decide intentar con un promedio móvil doble. Los resultados se presentan en la tabla 4-5. Para entender el pronóstico de la semana 16, se presentan los cálculos a continuación. Se emplea la ecuación 4.8 para calcular el promedio móvil de tres semanas (columna 3).

$$M_{15} = \hat{Y}_{15+1} = \frac{Y_{15} + Y_{15-1} + Y_{15-3+1}}{3}$$

$$M_{15} = \hat{Y}_{16} = \frac{728 + 711 + 712}{3} = 717$$

TABLA 4-5 Pronóstico de promedio móvil doble para Movie Video Store del ejemplo 4.4

(1) Tiempo t	(2) Ventas semanales Y_t	(3) Promedio móvil de tres semanas M_t	(4) Promedio móvil doble M_t'	(5) Valor de a	(6) Valor de b	(7) Pronóstico $a + bp$ ($p = 1$)	(8) e_t
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	—

$MSE = 63.7$

Luego se usa la ecuación 4.9 para calcular el promedio móvil doble (columna 4).

$$M'_{15} = \frac{M_{15} + M_{15-1} + M_{15-3+1}}{3}$$

$$M'_{15} = \frac{717 + 711 + 708}{3} = 712$$

Se emplea la ecuación 4.10 para calcular la diferencia entre los dos promedios móviles (columna 5).

$$a_{15} = 2M_{15} - M'_{15} = 2(717) - 712 = 722$$

Se utiliza la ecuación 4.11 para ajustar la pendiente (columna 6).

$$b_{15} = \frac{2}{3-1}(M_{15} - M'_{15}) = \frac{2}{2}(717 - 712) = 5$$

Se usa la expresión 4.12 para hacer el pronóstico de un periodo futuro (columna 7).

$$\hat{Y}_{15+1} = a_{15} + b_{15}p = 722 + 5(1) = 727$$

El pronóstico de 4 semanas en el futuro es:

$$\hat{Y}_{15+4} = a_{15} + b_{15}p = 722 + 5(4) = 742$$

Note que el *MSE* se ha reducido de 133 a 63.7.

Parece razonable que las observaciones más recientes son las que probablemente contienen la información más importante. En la siguiente sección se presenta un procedimiento que hace más énfasis en las observaciones más recientes.

MÉTODOS DE SUAVIZACIÓN EXPONENCIAL

Mientras que el método de promedios móviles toma en cuenta sólo las observaciones más recientes, la suavización exponencial simple ofrece un promedio móvil con peso exponencial para todos los valores previos observados. A menudo el modelo es adecuado para datos que no tienen una tendencia predecible ascendente o descendente. El objetivo es estimar el nivel real. Esta estimación de nivel se emplea luego como el pronóstico de valores futuros.

La suavización exponencial revisa continuamente un estimado a la luz de las experiencias más recientes. Este método se basa en promediar (suavizar) valores pasados de una serie de manera exponencialmente decreciente. La observación más reciente recibe el peso más grande, α (donde $0 < \alpha < 1$); la siguiente observación más reciente recibe menos peso, $\alpha(1 - \alpha)$; la observación de dos periodos en el pasado recibe incluso menos peso, $\alpha(1 - \alpha)^2$; y así sucesivamente.

En una representación de suavización exponencial, el nuevo pronóstico (para el tiempo $t + 1$) puede considerarse como la suma ponderada de la nueva observación (en el tiempo t) y el antiguo pronóstico (para el tiempo t). Se asigna el peso α ($0 < \alpha < 1$) al nuevo valor observado, y el peso $(1 - \alpha)$ al último pronóstico. Así,

$$\text{Nuevo pronóstico} = [\alpha \times (\text{nueva observación})] + [(1 - \alpha) \times (\text{último pronóstico})]$$

Más formalmente, la ecuación de suavización 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

α = constante de suavización ($0 < \alpha < 1$)

Y_t = nueva observación o el valor real de la serie en el periodo t

$\hat{Y}_t$ = último valor suavizado o el pronóstico del periodo t

La ecuación 4.13 se puede escribir 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 forma, el nuevo pronóstico ($\hat{Y}_{t+1}$) es el viejo ($\hat{Y}_t$) ajustado en α veces el error $Y_t - \hat{Y}_t$ en el pronóstico antiguo.

En la ecuación 4.13 la constante de suavización, α , sirve como el factor de ponderación. El valor de α determina el grado con el cual la observación actual influye en el pronóstico de la siguiente observación. Cuando α es cercano a 1, el nuevo pronóstico será, en esencia, la observación actual. (Asimismo, el nuevo pronóstico será el pronóstico anterior más un ajuste sustancial por cualquier error que haya ocurrido en el pronóstico precedente). A la inversa, cuando α es cercano a cero, el nuevo pronóstico será muy similar al pronóstico anterior, y la observación actual tendrá muy poco efecto.

La *suavización exponencial* es un procedimiento para revisar de forma continua un pronóstico a la luz de la experiencia más reciente.

Finalmente, la ecuación 4.13 supone que, para el tiempo t , $\hat{Y}_t = \alpha \hat{Y}_{t-1} + (1 - \alpha)Y_{t-1}$, y al sustituir en la ecuación 4.13 se obtiene:

$$\begin{aligned}\hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha)\hat{Y}_t = \alpha Y_t + (1 - \alpha)[\alpha Y_{t-1} + (1 - \alpha)\hat{Y}_{t-1}] \\ \hat{Y}_{t+1} &= \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + (1 - \alpha)^2\hat{Y}_{t-1}\end{aligned}$$

La sustitución continua (de $\hat{Y}_{t-1}$ y así sucesivamente) indica que $\hat{Y}_{t+1}$ puede escribirse como la suma de la Y actual y las anteriores con pesos que descienden exponencialmente

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2Y_{t-2} + \alpha(1 - \alpha)^3Y_{t-3} + \dots \quad (4.14)$$

Es decir, $\hat{Y}_{t+1}$ es un valor suavizado exponencialmente. La rapidez con la cual las observaciones pasadas pierden su efecto depende del valor de α , como se demuestra en la tabla 4-6.

Las ecuaciones 4.13 y 4.14 son equivalentes, pero la ecuación 4.13 se emplea generalmente para calcular el pronóstico de $\hat{Y}_{t+1}$ porque requiere menos almacenamiento de datos y se aplica fácilmente.

El valor asignado a α es la clave del análisis. Si se desea que las predicciones sean estables y las variaciones aleatorias se suavicen, se requiere un valor pequeño de α . Si se desea una respuesta rápida a un cambio real en el patrón de observaciones, un valor más grande de α es el apropiado. Un método para estimar α es un procedimiento iterativo que minimiza el error cuadrático medio (*MSE*) dado por la ecuación 3.8. Los pronósticos se calculan para, digamos, $\alpha = .1, .2, \dots, .9$, y se calcula la suma de los errores cuadráticos del pronóstico de cada uno de ellos.

TABLA 4-6 Comparación de constantes de suavización

Periodo	$\alpha = .1$		$\alpha = .6$	
	Cálculos	Peso	Cálculos	Peso
t		.100		.600
$t-1$	$.9 \times .1$	.090	$.4 \times .6$	.240
$t-2$	$.9 \times .9 \times .1$	.081	$.4 \times .4 \times .6$	.096
$t-3$	$.9 \times .9 \times .9 \times .1$	.073	$.4 \times .4 \times .4 \times .6$	.038
$t-4$	$.9 \times .9 \times .9 \times .9 \times .1$	.066	$.4 \times .4 \times .4 \times .4 \times .6$	.015
Todos los demás		.590		.011
	Totales	1.000		1.000

Se selecciona el valor de α que produzca el error más pequeño para usarlo en la generación de futuros pronósticos.

Para aplicar el algoritmo de la ecuación 4.13, se debe fijar un pronóstico. Una manera de fijar el primer pronóstico es tomarlo como la primera observación. El ejemplo 4.5 ilustrará este enfoque. Otro método es usar el promedio de las primeras cinco o seis observaciones del valor inicial suavizado.

Ejemplo 4.5

La técnica de suavización exponencial se presenta en la tabla 4-7 y en la figura 4-7 con datos de Acme Tool Company para los años 2000 a 2006, usando constantes de suavización de .1 y .6. Los datos del primer trimestre de 2006 se usarán como datos de prueba para ayudar a determinar el valor más adecuado de α (entre los dos considerados). La serie suavizada exponencialmente se calcula igualando la $\hat{Y}_1$ inicial a 500. Si existen datos anteriores, será posible usarlos para desarrollar una serie suavizada hasta el 2000 y luego usar esta experiencia como el valor inicial de la serie suavizada. Los cálculos que conducen al pronóstico de los periodos 3 y 4 se presentan a continuación:

1. Usando la ecuación 4-13, para el periodo de tiempo 2, el pronóstico para el periodo 3, con $\alpha = .1$ 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:

$$\begin{aligned}\hat{Y}_{3+1} &= \alpha Y_3 + (1 - \alpha) \hat{Y}_3 \\ \hat{Y}_4 &= .1(250) + .9(485) = 461.5\end{aligned}$$

En la tabla 4-7, cuando la constante de suavización es .1, el pronóstico para el primer trimestre de 2006 es 469, con un error cuadrático de 145,161. Cuando la constante de suavización es .6, el pronóstico para el primer trimestre de 2006 es 576, con un error cuadrático de 75,076. Con base en esta limitada evidencia, la suavización exponencial con $\alpha = .6$ funciona mejor que la suavización exponencial con $\alpha = .1$.

En la figura 4-7, advierta cómo son estables los valores suavizados para la constante de suavización .1. Con base en la minimización del error cuadrático medio *MSE* (el *MSE* se llama *MSD* en el resultado de Minitab), durante los primeros 24 trimestres, es mejor la constante de

TABLA 4-7 Valores de Acme Tool Company suavizados exponencialmente para el ejemplo 4.5

<i>Tiempo</i>		<i>Valor real</i>	<i>Valor suavizado</i>	<i>Error de pronóstico</i>	<i>Valor suavizado</i>	<i>Error de pronóstico</i>
<i>Año</i>	<i>Trimestres</i>	Y_t	$\hat{Y}_t(\alpha = .1)$	e_t	$\hat{Y}_t(\alpha = .6)$	e_t
2000	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
2001	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
2002	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
2003	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
2004	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
2005	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
2006	25	850	469.0	381.0	576.0	274.0

Nota: Los números entre paréntesis se refieren a las explicaciones que se dan en el texto del ejemplo 4.5.

suavización .6. Si se comparan los errores porcentuales absolutos medios (*MAPE*), la constante de suavización de .6 también es mejor. Resumiendo:

$$\alpha = .1 \quad MSE = 24,262 \quad MAPE = 38.9\%$$

$$\alpha = .6 \quad MSE = 22,248 \quad MAPE = 36.5\%$$

Sin embargo, tanto el *MSE* como el *MAPE* son grandes y, con base en estos estadísticos de resumen, es evidente que la suavización exponencial no representa bien estos datos. Como veremos, un método de suavización que permite la estacionalidad hace un mejor trabajo para predecir las ventas de serruchos de Acme Tool Company.

Un factor, diferente de la selección de α , que afecta los valores de pronósticos subsiguientes es la selección del valor inicial, $\hat{Y}_1$ de la serie suavizada. En la tabla 4-7 (véase el ejemplo 4.5), se utilizó $\hat{Y}_1 = Y_1$ como el valor inicial suavizado. Esta selección tiende a otorgar demasiado peso a Y_1 en los pronósticos posteriores. Por fortuna, la influencia del pronóstico inicial disminuye considerablemente conforme se incrementa t .

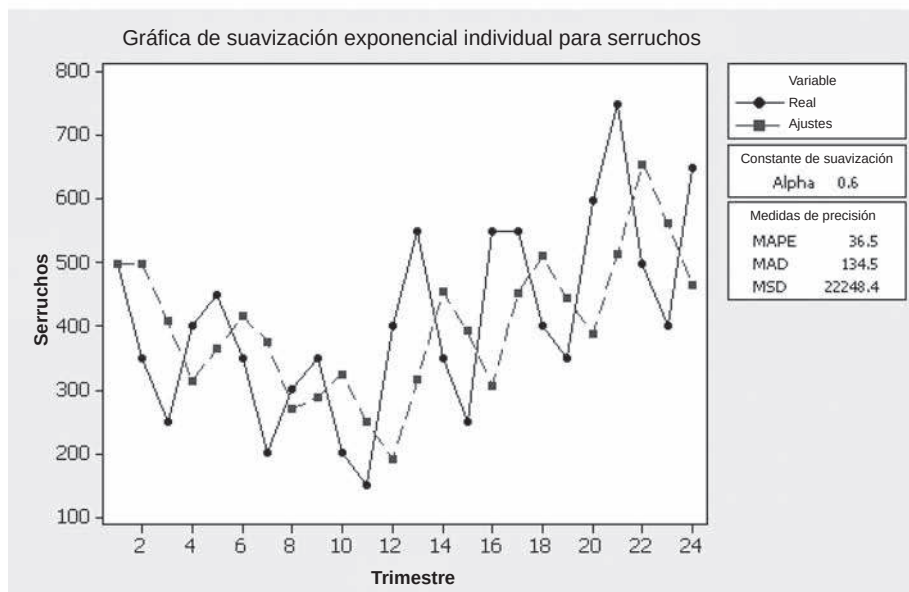
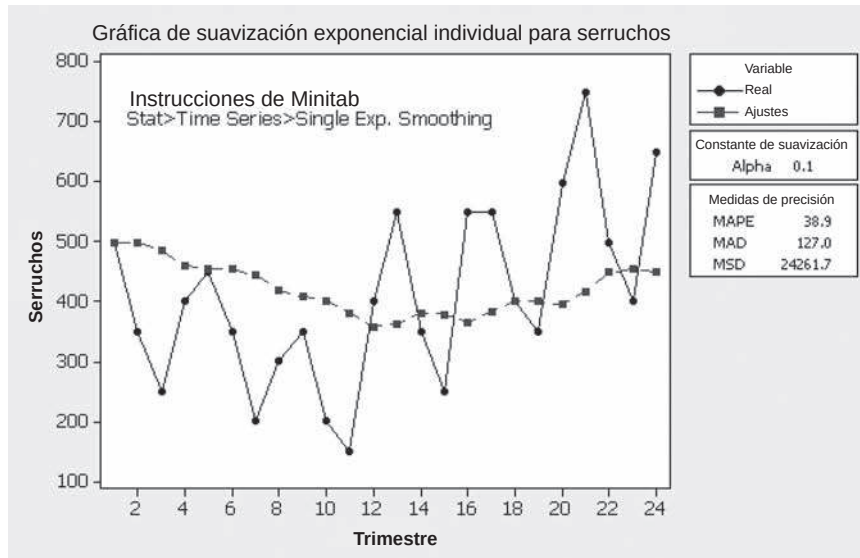


FIGURA 4-7 Suavización exponencial de los datos de Acme Tool Company del ejemplo 4.5: (arriba) $\alpha = .1$ y (abajo) $\alpha = .6$

Otro enfoque de inicialización del procedimiento de suavización es promediar las primeras k observaciones. Entonces, la suavización inicia con:

$$\hat{Y}_1 = \frac{1}{k} \sum_{t=1}^k Y_t$$

A menudo se selecciona un número relativamente pequeño para k . Por ejemplo, Minitab predetermina que $k = 6$.

Ejemplo 4.6

El cálculo del valor inicial como un promedio para los datos de Acme Tool Company presentados en el ejemplo 4.5 se muestra a continuación. Si se iguala k a 6, entonces el valor inicial es:

$$\hat{Y}_1 = \frac{1}{6} \sum_{t=1}^6 Y_t = \frac{1}{6}(500 + 350 + 250 + 400 + 450 + 350) = 383.3$$

A continuación se presentan el MSE y el $MAPE$ para cada α cuando se usa un valor inicial suavizado de 383.3:

$$\alpha = .1 \quad MSE = 21,091 \quad MAPE = 32.1\%$$

$$\alpha = .6 \quad MSE = 22,152 \quad MAPE = 36.7\%$$

El valor inicial $\hat{Y}_1 = 383.3$, trae consigo una disminución del MSE y el $MAPE$ para $\alpha = .1$, pero no tiene mucho efecto cuando $\alpha = .6$. Ahora el mejor modelo, basado en las medidas de resumen del MSE y el $MAPE$, parece ser uno que use $\alpha = .1$ en vez de $.6$.

La figura 4-8 presenta los resultados del ejemplo 4.5 cuando los datos se corren en Minitab (para instrucciones específicas, véase la sección de aplicaciones de Minitab al final del capítulo). La constante de suavización $\alpha = .266$ fue seleccionada automáticamente al minimizar el MSE . El MSE se redujo a 19,447, el $MAPE$ es igual al 32.2% y, aunque no se indica, el MPE es igual a -6.4% . El pronóstico para el primer trimestre de 2006 es 534.

La figura 4-9 presenta la función de autocorrelación para los residuos del método de suavización exponencial usando una alfa de .266. Cuando la prueba de Ljung-Box se realiza para seis retrasos de tiempo, el gran valor de LBQ (33.86) indica que el conjunto de las primeras seis autocorrelaciones residuales son más grandes de lo que se esperaría si los residuos fueran aleatorios. En particular, las autocorrelaciones de los residuos tan significativamente grandes en los retrasos 2 y 4 indican que la variación estacional en los datos que no se explica mediante un suavizamiento exponencial simple.

La suavización exponencial a menudo es un buen procedimiento para pronosticar cuando una serie de tiempo que no es aleatoria exhibe una tendencia en su comportamiento. Es útil desarrollar una medida que se pueda utilizar para determinar cuándo ha cambiado el patrón básico de una serie de tiempo. Una *señal de control* es un modo de supervisar un cambio. Una señal de control implica el cálculo de una medida de los errores de pronóstico en el tiempo y el establecimiento de límites, de manera que, cuando los errores caen fuera de esos límites, el encargado de efectuar el pronóstico recibe una señal de alerta.

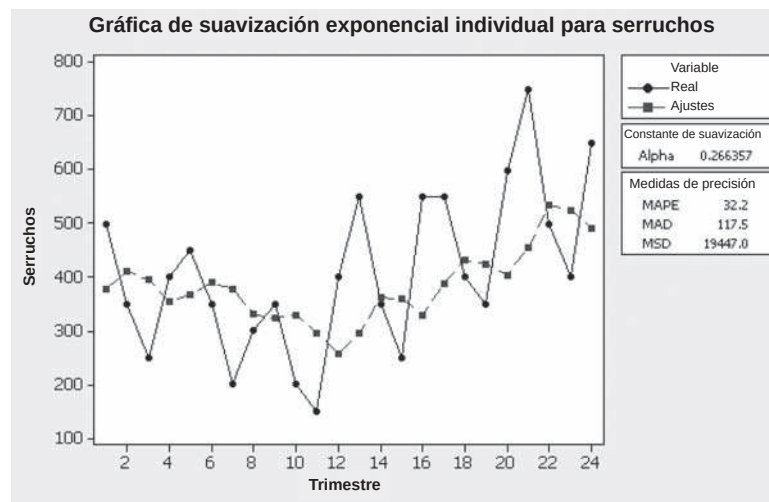


FIGURA 4-8 Suavización exponencial con $\alpha = .266$ para los datos de Acme Tool Company para el ejemplo 4.6

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	0.121421	0.59	0.40
2	-0.588941	-2.84	10.24
3	0.109122	0.41	10.59
4	0.646967	2.40	23.65
5	-0.006908	-0.02	23.65
6	-0.542625	-1.65	33.86

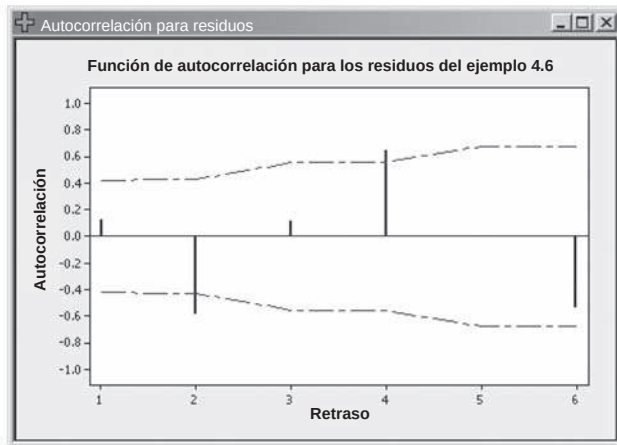


FIGURA 4-9 Función de autocorrelación para los residuos cuando se usa suavización exponencial con $\alpha = .266$ con datos de Acme Tool Company para el Ejemplo 4.6

Una *señal de control* implica el cálculo de una medida de errores de pronóstico en el tiempo y el establecimiento de límites, de manera que, cuando los errores acumulados caen fuera de esos límites, el encargado de efectuar el pronóstico recibe una señal de alerta.

Por ejemplo, podría utilizarse una señal de control para determinar cuándo debe modificarse el valor de la constante de suavización α . Como generalmente se pronostica un número grande de artículos, la práctica común es continuar con el mismo valor de α para varios periodos antes de intentar determinar si es necesaria una revisión. Desafortunadamente, la simplicidad de usar un modelo de suavización exponencial establecido es un fuerte motivador para no hacer un cambio. Pero en algún punto, puede ser necesario actualizar α o abandonar totalmente la suavización 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. Un sistema tal contiene un rango de desviaciones permitidas del pronóstico con respecto a los valores reales. Siempre que los pronósticos generados por la suavización exponencial caigan dentro de este rango, no se necesita ningún cambio de α . Sin embargo, si un pronóstico cae fuera del rango, el sistema señala la necesidad de actualizar α .

Por ejemplo, si las cosas van bien, la técnica de elaboración del pronóstico debería sobreestimar y subestimar con la misma frecuencia. Se puede desarrollar una señal de control con base en esta lógica.

Sea U igual al número de subestimaciones de los últimos n pronósticos fuera de los límites. En otras palabras, U es el número de errores entre los últimos k positivos. Si el proceso está bajo control, el valor esperado de U es $k/2$; aunque está implicada la variabilidad de la muestra, así que los valores cercanos a $k/2$ no son raros. Por otro lado, los valores que no están cercanos a $k/2$ indicarían que la técnica está generando pronósticos con sesgos.

Ejemplo 4.7

Suponga que Acme Tool Company decidió usar la técnica de suavización exponencial con α igual a .1, como se indica en el ejemplo 4.5 (véase la p. 121). Si el proceso está controlado, y el analista decide hacer un seguimiento de los últimos 10 valores de error, U tiene un valor esperado de 5. En realidad,

un valor de U igual a 2, 3, 4, 6, 7 u 8, no sería demasiado alarmante. Sin embargo, un valor de 0, 1, 9 o 10 sería motivo de preocupación, puesto que la probabilidad de obtener un valor tal, sólo por casualidad, sería de .022 (con base en la distribución binomial). Con esta información, se puede desarrollar un sistema de control con base en las siguientes reglas:

Si $2 \leq U \leq 8$, entonces el proceso está bajo control.

Si $U < 2$ o $U > 8$, entonces el proceso está fuera de control.

Suponga que, de los siguientes 10 pronósticos que usan esta técnica, sólo uno tiene un error positivo. Puesto que la probabilidad de obtener un solo error positivo de 10 posibles es muy baja (.011), se considera que el proceso está fuera de control (sobrestimando), y el valor de α debe cambiarse.

Otra manera de dar seguimiento a una técnica de pronósticos es determinar el rango en el que deben estar los errores de pronóstico. Esto puede realizarse usando el MSE que se estableció al determinar el tamaño óptimo de α . Si la técnica de suavización exponencial es razonablemente exacta, el error de pronóstico debe tener una distribución aproximadamente normal alrededor de una media igual a cero. Con esta condición, hay un 95% de probabilidad de que la observación real caiga dentro de aproximadamente dos desviaciones estándar del pronóstico. Usando la $RMSE$ como un estimado de la desviación estándar del error de pronóstico, se pueden determinar los límites de error aproximados a 95%. Los errores de pronóstico que caigan dentro de estos límites indican que no hay motivo de alarma. Los errores (particularmente una secuencia de errores) fuera de los límites sugieren un cambio. El ejemplo 4.8 ilustra este enfoque.

Ejemplo 4.8

En el ejemplo 4.6, sobre Acme Tool Company, se determinó que la α óptima era .266, con $MSE = 19,447$. Un estimado de la desviación estándar de los errores de pronóstico es $RMSE = \sqrt{19,447} = 139.5$. Si los errores de pronóstico tienen distribución aproximadamente normal alrededor de una media de cero, existe un 95% de probabilidad de que la observación real caiga dentro de dos desviaciones estándar del pronóstico, o dentro de:

$$\pm 2RMSE = \pm 2\sqrt{19,447} = \pm 2(139.5) = \pm 279$$

Para este ejemplo, el error permisible absoluto es 279. Si para cualquier pronóstico futuro la magnitud del error es mayor que 279, existen razones para creer que la constante α óptima de suavización debe actualizarse o que debe considerarse un método diferente para la elaboración del pronóstico.

La explicación anterior sobre las señales de control también se aplica a los métodos de suavización que serán discutidos en el resto del capítulo.

La suavización exponencial simple funciona bien cuando los datos tienen un nivel de cambio infrecuente. Siempre que existe una tendencia sostenida, la suavización exponencial permanecerá retrasada en el tiempo con respecto a los valores reales. La técnica de suavización exponencial lineal de Holt, la cual se diseñó para manejar datos con una tendencia bien definida, enfrenta este problema y se presenta a continuación.

Suavización exponencial ajustada a la tendencia: Método de Holt

En la suavización exponencial simple, se supone que el nivel de las series de tiempo varía ocasionalmente, y se requiere un estimado del nivel actual. En algunas situaciones, los datos observados tienen una tendencia clara y contienen información que permite anticipar movimientos futuros ascendentes. Cuando éste es el caso, se necesita una función de tendencia lineal del pronóstico. Puesto que las series de negocios y económicas rara vez exhiben una tendencia lineal fija, consideramos la posibilidad de modelar tendencias lineales locales en evolución con el tiempo. Holt (2004) desarrolló un método de suavización exponencial, conocida como la *suavización exponencial lineal de Holt*,¹ la cual toma en cuenta la evolución local lineal de las tendencias en una serie de tiempo y puede usarse para generar pronósticos.

¹La suavización exponencial lineal de Holt algunas veces se conoce como *suavización exponencial doble*.

Cuando se anticipa una tendencia en una serie de tiempo, se requiere un estimado de la pendiente actual, así como del nivel actual. La técnica de Holt suaviza directamente el nivel y la pendiente usando diferentes constantes de suavización para cada uno. Estas constantes de suavización proporcionan estimados del nivel y la pendiente que se adaptan en el tiempo conforme se dispone de nuevas observaciones. Una de las ventajas de la técnica de Holt es que ofrece un alto grado de flexibilidad en la selección de coeficientes con los cuales se controla el nivel y la tendencia.

Las tres ecuaciones usadas en el método de Holt son

1. La serie suavizada exponencialmente o nivel actual estimado

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.15)$$

2. El estimado de la tendencia

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.16)$$

3. El pronóstico para los p periodos del futuro

$$\hat{Y}_{t+p} = L_t + pT_t \quad (4.17)$$

donde

L_t = nuevo valor suavizado (estimado del nivel actual)

$\alpha = 1$, constante de suavización para el nivel ($0 < \alpha < 1$)

Y_t = nueva observación o valor real de la serie en el periodo t

β = constante de suavización para el estimado de la tendencia ($0 < \beta < 1$)

T_t = estimado de tendencia

p = periodos a pronosticar 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 de la suavización exponencial simple, la ecuación 4.13, excepto que se ha incorporado el término (T_{t-1}) para actualizar adecuadamente el nivel cuando existe una tendencia. Es decir, el nivel actual (L_t) se calcula tomando un promedio ponderado de dos estimaciones de nivel: una estimación está dada por la observación actual (Y_t), y la otra estimación está dada por la suma de la tendencia previa (T_{t-1}) y el nivel previamente suavizado (L_{t-1}). Si no existe una tendencia en los datos, no hay necesidad del término T_{t-1} en la ecuación 4.15, reduciéndola efectivamente a la ecuación 4.13. Tampoco es necesaria la ecuación 4.16.

Para efectuar la estimación de la tendencia, se usa una nueva constante de suavización, β . La ecuación 4.16 indica que la tendencia actual (T_t) es un promedio ponderado (con pesos β y $1 - \beta$) de dos tendencias estimadas: una estimación está dada por el cambio en el nivel de tiempo $t - 1$ a t ($L_t - L_{t-1}$), y la otra estimación es la tendencia previamente suavizada (T_{t-1}). La ecuación 4.16 es similar a la ecuación 4.15, excepto que la suavización se hace para la tendencia y no para los datos reales.

La ecuación 4.17 muestra el pronóstico para p periodos en el futuro. Para un pronóstico hecho en el tiempo t , la tendencia actual estimada (T_t) se multiplica por el número de periodos (p) que serán pronosticados, y luego el producto se suma al nivel actual (L_t). Note que los pronósticos para periodos futuros permanecen a lo largo de una línea recta con pendiente T_t e intersección en L_t .

Al igual que en la suavización exponencial simple, las constantes de suavización α y β pueden seleccionarse subjetivamente o generarse minimizando una medida de error de pronóstico tal como el *MSE*. Pesos grandes tienen como resultado cambios más rápidos en el componente; pesos pequeños tienen como resultado cambios menos rápidos. Por lo tanto, cuanto mayores son los pesos, los valores de suavizamiento seguirán más a los datos; cuanto menores son los pesos, los valores de suavizamiento seguirán más a los valores de suavizamiento previos.

Podríamos desarrollar una tabla de valores de α y β (por ejemplo, cada combinación de $\alpha = 0.1, 0.2, \dots, 0.9$ y $\beta = 0.1, 0.2, \dots, 0.9$) y luego seleccionar la combinación que dé el menor *MSE*. La mayoría de los paquetes de software para elaboración de pronósticos usan un algoritmo de optimización para minimizar el *MSE*. Podríamos insistir en que $\alpha = \beta$, proporcionando así cantidades iguales de suavización para el nivel y la tendencia. En el caso especial en que $\alpha = \beta$, el método de Holt es igual que la suavización exponencial doble de Brown.

Para iniciar, se deben determinar los valores iniciales de L y T en las ecuaciones 4.15 y 4.16. Un enfoque consiste en fijar la primera estimación del nivel suavizado igual a la primera observación. Luego se considera la tendencia igual a cero. Un segundo enfoque es usar el promedio de las primeras cinco o seis observaciones como el valor suavizado inicial de L . Luego, se estima la tendencia usando la pendiente de una línea ajustada a estas cinco o seis observaciones. Minitab desarrolla una ecuación de regresión usando la variable de interés como Y y el tiempo como la variable independiente X . La constante de esta ecuación es la estimación inicial del componente de nivel, y la pendiente o el coeficiente de regresión es la estimación inicial del componente de tendencia.

Ejemplo 4.9

En el ejemplo 4.6 la suavización exponencial simple no produjo pronósticos exitosos de las ventas de serruchos de Acme Tool Company. Como la figura 4-8 sugiere que podría haber una tendencia en estos datos, se usa la suavización exponencial lineal de Holt para desarrollar los pronósticos. Para iniciar los cálculos mostrados en la tabla 4-8, se necesita estimar dos valores iniciales, a saber, el valor del nivel inicial y el valor de la tendencia inicial. El estimado del nivel se iguala a la primera observación. La tendencia se iguala a cero. La técnica se presenta en la tabla 4-8 para $\alpha = .3$ y $\beta = .1$.

El valor de α que se usa aquí es cercano al valor óptimo de α ($\alpha = .266$) para la suavización exponencial simple del ejemplo 4.6. Se usa α en la suavización de los datos para eliminar la aleatoriedad y estimar el nivel. La constante de suavización β es como α , excepto que se usa para suavizar la tendencia en los datos. Ambas constantes de suavización se utilizan para promediar valores pasados y así eliminar la aleatoriedad. A continuación se presentan los cálculos para obtener el pronóstico del periodo 3.

1. Actualización de la serie o nivel suavizado exponencialmente:

$$\begin{aligned} L_t &= \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ L_2 &= .3Y_2 + (1 - .3)(L_{2-1} + T_{2-1}) \\ L_2 &= .3(350) + .7(500 + 0) = 455 \end{aligned}$$

2. Actualización de la estimación de la tendencia:

$$\begin{aligned} T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\ T_2 &= .1(L_2 - L_{2-1}) + (1 - .1)T_{2-1} \\ T_2 &= .1(455 - 500) + .9(0) = -4.5 \end{aligned}$$

3. Pronóstico de un periodo futuro:

$$\begin{aligned} \hat{Y}_{t+p} &= L_t + pT_t \\ \hat{Y}_{2+1} &= L_2 + 1T_2 = 455 + 1(-4.5) = 450.5 \end{aligned}$$

4. Determinación del error de pronóstico:

$$e_3 = Y_3 - \hat{Y}_3 = 250 - 450.5 = -200.5$$

El pronóstico para el periodo 25 se calcula como sigue:

1. Actualización de la serie o nivel suavizado exponencialmente:

$$\begin{aligned} L_{24} &= .3Y_{24} + (1 - .3)(L_{24-1} + T_{24-1}) \\ L_{24} &= .3(650) + .7(517.6 + 9.8) = 564.2 \end{aligned}$$

TABLA 4-8 Valores de las ventas suavizadas exponencialmente de Acme Tool Company, por el método de Holt, para el ejemplo 4.9

Año	t	Y_t	L_t	T_t	$\hat{Y}_{t+p}$	e_t
2000	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
2001	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
2002	9	350	307.4	-12.3	289.1	60.9
	10	200	266.6	-15.2	295.1	-95.1
	11	150	221.0	-18.2	251.4	-101.4
	12	400	262.0	-12.3	202.8	197.2
2003	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	-87.7
	16	550	379.1	1.8	305.9	244.1
2004	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
2005	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
2006	25	850	—	—	577.7	272.3

$MSE = 20,515.5$

2. Actualización del cálculo de la 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. Pronóstico de un periodo futuro:

$$\hat{Y}_{24+1} = L_{24} + 1T_{24}$$

$$\hat{Y}_{25} = 564.2 + 1(13.5) = 577.7$$

Con base en la minimización del MSE durante el periodo 2000 a 2006, la suavización lineal de Holt (con $\alpha = .3$ y $\beta = .1$) no reproduce los datos mejor que la suavización exponencial simple que usó una constante de suavización de .266. Una comparación de los $MAPE$ indica que son iguales. Cuando se comparan los pronósticos de las ventas actuales del primer trimestre de

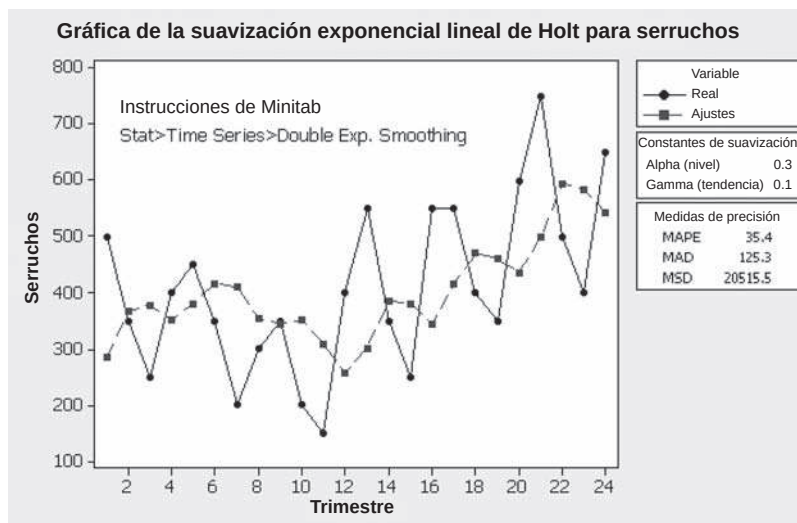


FIGURA 4-10 Suavización exponencial lineal de Holt de los datos de Acme Tool Company para el ejemplo 4.9

2006, una vez más la suavización de Holt y la suavización exponencial simple son comparables. Resumiendo:

$$\begin{array}{lll} \alpha = .266 & MSE = 19,447 & MAPE = 32.2\% \\ \alpha = .3, \beta = .1 & MSE = 20,516 & MAPE = 35.4\% \end{array}$$

La figura 4-10 presenta los resultados cuando el método de Holt con $\alpha = .3$ y $\beta = .1$ se corre en Minitab.² La función de autocorrelación para los residuos de la suavización exponencial lineal de Holt se presenta en la figura 4-11. Los coeficientes de autocorrelación en los retrasos 2 y 4 parecen significantes. También, cuando se calcula el estadístico Q de Ljung-Box para seis retrasos de tiempo, el valor grande del estadístico LBQ (36.33) indica que los residuos contienen autocorrelaciones grandes; no son aleatorios. Las grandes autocorrelaciones residuales en los retrasos 2 y 4 sugieren que puede estar presente un componente estacional en los datos de Acme Tool Company.

Los resultados en los ejemplos 4.6 y 4.9 (véanse las figuras 4-8 y 4-10) no son muy diferentes porque la constante α de suavización es aproximadamente la misma en ambos casos y la constante de suavización β en el ejemplo 4.9 es pequeña. (Para $\beta = 0$, la suavización lineal de Holt se convierte en una suavización exponencial simple).

Suavización exponencial ajustada a la tendencia y a la variación estacional: Método de Winters

La revisión de los datos de Acme Tool Company en la tabla 4-8 indica que las ventas son más altas durante el primero y el cuarto trimestres y más bajas durante el tercer trimestre de manera sistemática. Parece que existe un patrón estacional. El método de suavización exponencial lineal y estacional de tres parámetros de Winters, una extensión del método de Holt, podría representar mejor los datos y reducir el error de pronóstico. En el método de Winters, se emplea una ecuación adicional para estimar la estacionalidad. En la versión multiplicativa del método de Winters, la estimación de la estacionalidad está dada por un índice estacional y se calcula mediante la ecuación 4.20. Esta última indica que para calcular el componente esta-

²En el programa de Minitab, el parámetro de tendencia gama (γ) es idéntico a nuestra beta (β).

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	0.092547	0.45	0.23
2	-0.629408	-3.06	11.47
3	0.074355	0.27	11.63
4	0.601371	2.18	22.92
5	-0.070132	-0.22	23.08
6	-0.618267	-1.90	36.33

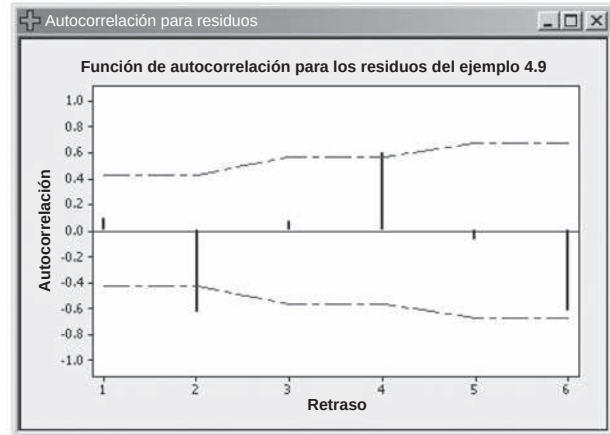


FIGURA 4-11 Función de autocorrelación de los residuos de la suavización exponencial lineal de Holt de los catos de Acme Tool Company para el Ejemplo 4.9

cional actual, S_t , el producto de γ y un estimado del índice estacional dado por Y_t/L_t se suma $(1 - \gamma)$ veces al componente estacional previo S_{t-s} . Este procedimiento es equivalente a suavizar los valores previos y actuales de Y_t/L_t . Y_t se divide entre el nivel actual estimado L_t , para crear un índice (razón) que pueda usarse de forma multiplicativa para ajustar un pronóstico que tome en cuenta los picos y valles estacionales.

Las cuatro ecuaciones usadas en la suavización (multiplicativa) de Winters son

1. Series suavizadas exponencialmente o nivel estimado:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (4.18)$$

2. Estimación de la tendencia:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (4.19)$$

3. Estimado de estacionalidad:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (4.20)$$

4. Pronóstico de p periodos futuros:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (4.21)$$

donde

L_t = nuevo valor suavizado (estimado de nivel actual)

α = constante de suavización del nivel

Y_t = nueva observación o valor real en el periodo t

β = constante de suavización para el estimado de tendencia

T_t = estimado de tendencia

γ = constante de suavización para el estimado de estacionalidad

S_t = estimado de estacionalidad

p = periodos futuros a pronosticarse

s = longitud de la estacionalidad

$\hat{Y}_{t+p}$ = el pronóstico para el periodo p en el futuro

La ecuación 4.18 actualiza las series suavizadas. Una ligera diferencia en esta ecuación la distingue de la correspondiente en el procedimiento de Holt, la ecuación 4.15. En la ecuación 4.18, Y_t se divide entre S_{t-s} , lo cual ajusta Y_t para la estacionalidad, eliminando así los efectos estacionales que pudieran existir en los datos originales Y_t .

Después de que la estimación de la tendencia y la estimación de la estacionalidad se han suavizado en las ecuaciones 4.19 y 4.20, se obtiene un pronóstico con la ecuación 4.21. Casi es igual que la fórmula correspondiente, la ecuación 4.17, usada para obtener un pronóstico con la suavización de Holt. La diferencia es que esta estimación para futuros periodos, $t + p$, se multiplica por S_{t-s+p} . Este índice estacional es el último disponible y, por consiguiente, se utiliza para ajustar el pronóstico con la estacionalidad.

Al igual que en la suavización exponencial lineal de Holt, los pesos α , β y γ pueden seleccionarse subjetivamente o generarse al minimizar una medida de error de pronóstico, como el *MSE*. El enfoque más común para determinar estos valores es usar un algoritmo de optimización para obtener las constantes óptimas de suavización.

Para iniciar el algoritmo de la ecuación 4.18, se deben establecer los valores iniciales para las series suavizadas, L_t , la tendencia T_t y los índices estacionales S_t . Un método consiste en igualar la primera estimación de la serie suavizada (nivel) a la primera observación. Luego, la tendencia se iguala a cero y cada índice estacional se fija en 1.0. Existen otros métodos para la inicialización de la estimación del nivel, tendencia y estacionalidad. Minitab, por ejemplo, desarrolla una ecuación de regresión usando la variable de interés como Y y el tiempo como la variable independiente X . La constante de esta ecuación es la estimación inicial de la serie suavizada o componente de nivel, y la pendiente o el coeficiente de regresión es la estimación inicial del componente de tendencia. Los valores iniciales de los componentes estacionales se obtienen a partir de una regresión de variable ficticia que usa datos sin tendencia (véase el capítulo 8).

Ejemplo 4.10

En la tabla 4-9 se presenta la técnica de Winters, para $\alpha = .4$, $\beta = .1$ y $\gamma = .3$ para los datos de Acme Tool Company. El valor de α es similar al utilizado en la suavización exponencial simple del ejemplo 4.6 y sirve para suavizar los datos y crear un nivel estimado. La constante de suavización β , se utiliza para crear una estimación suavizada de la tendencia. La constante de suavización γ se emplea para crear una estimación suavizada del componente estacional en los datos.

Es factible usar Minitab para resolver este ejemplo (para instrucciones específicas, véase la sección de aplicaciones de Minitab al final del capítulo).³ Los resultados se muestran en la tabla 4-9 y en la figura 4-12. El pronóstico para el primer trimestre de 2006 es 778.2. Los cálculos para obtener el valor del pronóstico para el primer trimestre de 2006, o periodo 25, se presentan a continuación.

1. La serie suavizada exponencialmente o nivel estimado:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$L_{24} = .4 \frac{Y_{24}}{S_{24-4}} + (1 - .4)(L_{24-1} + T_{24-1})$$

$$L_{24} = .4 \frac{650}{1.39628} + (1 - .4)(501.286 + 9.148)$$

$$L_{24} = .4(465.52) + .6(510.434) = 492.469$$

³En el programa de Minitab, el parámetro de tendencia gama (γ) es idéntico a nuestra beta (β) y el parámetro estacional delta (δ) es idéntico a nuestra gama (γ) de las ecuaciones 4.19 y 4.20, respectivamente.

TABLA 4-9 Valores suavizados exponencialmente de las ventas de Acme Tool Company, por el método de Winters, para el ejemplo 4.10

Año	t	Y_t	L_t	T_t	S_t	$\hat{Y}_{t+p}$	e_t
2000	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
2001	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
2002	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
2003	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
2004	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
2005	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
2006	25	850	—	—	—	778.179	—
	26	600	—	—	—	521.917	—
	27	450	—	—	—	393.430	—
	28	700	—	—	—	716.726	—

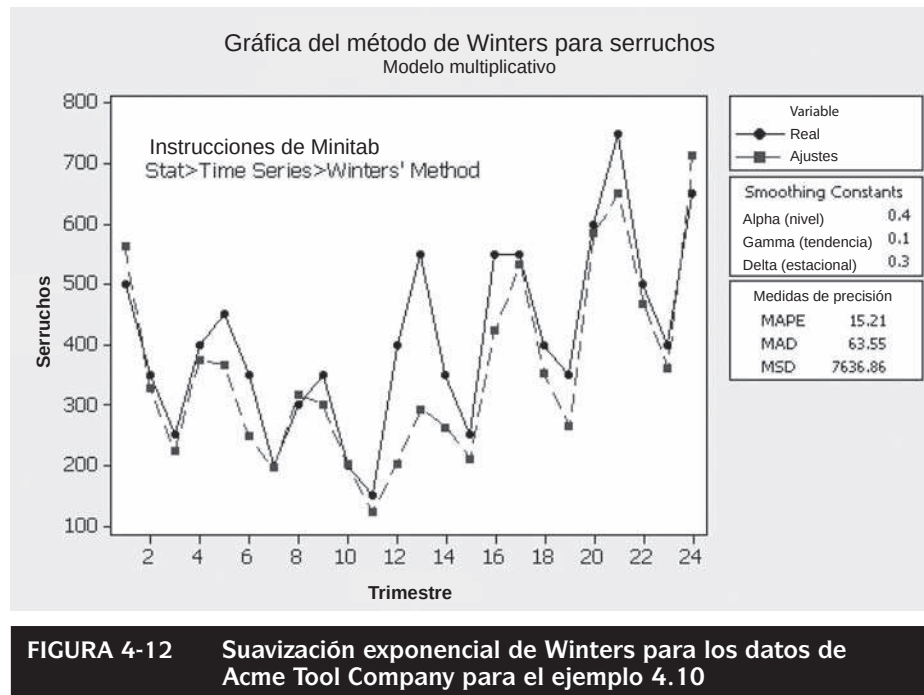
$MSE = 7,636.86$

2. El estimado de la tendencia:

$$\begin{aligned}
 T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\
 T_{24} &= .1(L_{24} - L_{24-1}) + (1 - .1)T_{24-1} \\
 T_{24} &= .1(492.469 - 501.286) + .9(9.148) \\
 T_{24} &= .1(-8.817) + .9(9.148) = 7.352
 \end{aligned}$$

3. El estimado de la estacionalidad:

$$\begin{aligned}
 S_t &= \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \\
 S_{24} &= .3 \frac{Y_{24}}{L_{24}} + (1 - .3)S_{24-4} \\
 S_{24} &= .3 \frac{650}{492.469} + .7(1.39628) \\
 S_{24} &= .3(1.3199) + .9774 = 1.3734
 \end{aligned}$$



4. El pronóstico para el primer periodo, $p = 1$ en el futuro:

$$\hat{Y}_{24+1} = (L_{24} + 1T_{24})S_{24-4+1}$$

$$\hat{Y}_{25} = (492.469 + 1(7.352))1.5569 = 778.17$$

Para los valores de los parámetros considerados, la técnica de Winters es mejor que los dos procedimientos de suavización anteriores en términos de la minimización del MSE . Cuando se comparan los pronósticos de las ventas reales para el primer trimestre de 2006, la técnica de Winters también parece hacer un mejor trabajo. La figura 4-13 muestra la función de autocorrelación para los residuos de suavización exponencial, con la técnica de Winters. Ninguno de los coeficientes de autocorrelación residual parece ser significativamente mayor que cero. Cuando se calcula el estadístico Q de Ljung-Box para seis retrasos de tiempo, el pequeño valor de LBQ (5.01) indica que la serie residual es aleatoria. El método de suavización exponencial de Winters parece arrojar pronósticos adecuados para los datos de Acme Tool Company.

El método de Winters ofrece una manera fácil de tomar en cuenta la estacionalidad cuando los datos muestran un patrón estacional. Un método alternativo consiste en eliminar la estacionalidad (desestacionalizar) primero o ajustar la estacionalidad de los datos. La desestacionalización es un proceso que elimina los efectos de la estacionalidad en los datos y se estudiará en el capítulo 5. El modelo de elaboración del pronóstico se aplica a los datos desestacionalizados y, si se requiere, se reinserta el componente de estacionalidad para obtener pronósticos exactos.

La suavización exponencial es una técnica habitual para pronósticos de corto plazo. Sus mayores ventajas son su bajo costo y su sencillez. Cuando se necesitan pronósticos para sistemas de inventario que contienen miles de artículos, los métodos de suavización son a menudo el único enfoque aceptable.

Los promedios móviles simples y la suavización exponencial basan los pronósticos en promedios ponderados de mediciones pasadas. El fundamento es que los valores pasados contienen información acerca de lo que ocurrirá en el futuro. Puesto que los valores pasados incluyen fluctuaciones aleatorias, así como información relacionada con el patrón subyacente

Función de autocorrelación para residuos de Winters

Retraso	ACF	T	LBQ
1	0.297749	1.46	2.41
2	-0.031543	-0.14	2.43
3	0.017610	0.08	2.44
4	-0.107084	-0.48	2.80
5	-0.257066	-1.15	4.97
6	0.033280	0.14	5.01

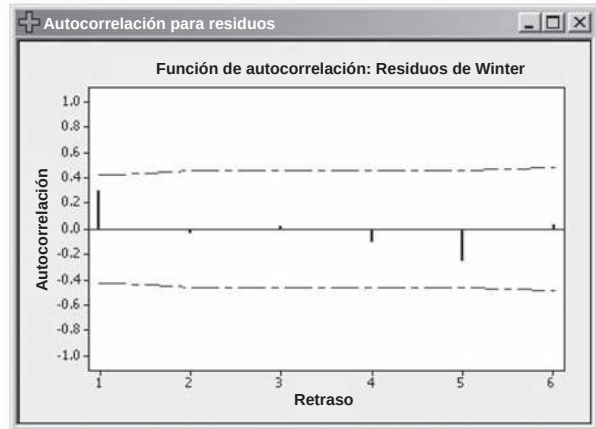


FIGURA 4-13 Función de autocorrelación del método de suavización exponencial multiplicativa de Winters de los residuos para los datos de Acme Tool Company para el ejemplo 4.10

de la variable, se hace un intento para suavizar los valores. La suavización supone que las fluctuaciones extremas representan la aleatoriedad en una serie de observaciones históricas.

Los promedios móviles son las medias de un cierto número, k , de valores de una variable. El promedio más reciente es entonces el pronóstico para el siguiente periodo. Este enfoque asigna un peso igual a cada valor pasado implicado en el promedio. No obstante, es una buena idea usar todos los datos pero haciendo énfasis en los valores más recientes. Los métodos de suavización exponencial son atractivos porque generan pronósticos usando todas las observaciones y asignando pesos que disminuyen exponencialmente conforme las observaciones se vuelven más antiguas.

APLICACIÓN A LA ADMINISTRACIÓN

Los pronósticos son uno de los elementos más importantes con los que cuentan los gerentes para apoyar el proceso de toma de decisiones. Prácticamente todas las decisiones operativas dependen hasta cierto punto de un pronóstico. El departamento de producción tiene que programar, para uno o dos meses, las necesidades de personal y los pedidos de materia prima; el departamento de finanzas debe determinar las mejores oportunidades de inversión; el mercadeo debe pronosticar la demanda de un nuevo producto. La lista de las aplicaciones de los pronósticos es muy extensa.

Los ejecutivos están profundamente conscientes de la importancia del pronóstico. De hecho, se dedica una gran cantidad de tiempo al estudio de las tendencias en asuntos económicos y políticos y en la manera como los acontecimientos pueden afectar la demanda de los productos o servicios. Un tema de interés es la importancia que dan los ejecutivos a los métodos de pronósticos cuantitativos en comparación con sus propias opiniones. Este tema es especialmente sensible cuando se trata de acontecimientos que tienen un efecto significativo en la demanda. Un problema con los métodos de pronóstico cuantitativos es que dependen de datos históricos. Por esta razón, probablemente son menos efectivos en la identificación de un cambio drástico que a menudo tiene como resultado demandas notoriamente más altas o más bajas.

Los métodos de suavización y promedio para la elaboración de pronósticos discutidos en este capítulo son útiles por su relativa sencillez. Estos métodos simples tienden a ser menos costosos, más fáciles de aplicar y más fáciles de comprender que los métodos complejos. A menudo, el costo y las dificultades potenciales asociadas con la construcción de modelos más complejos exceden por mucho cualquier aumento en la exactitud. Por esta razón, los negocios

pequeños encuentran prácticos los métodos simples. Los negocios que no tienen personal capaz de manejar modelos estadísticos también se inclinan por los métodos simples. Los directores de negocios a menudo enfrentan la necesidad de elaborar pronósticos de corto plazo para un buen número de diferentes artículos. Un ejemplo común es el gerente que debe programar la producción con base en algún pronóstico de la demanda de varios cientos de productos diferentes en una línea de producción. También los nuevos negocios sin grandes bases de datos históricos encuentran útiles estos enfoques simples.

Con una selección sensata del orden k , el método de promedio móvil puede hacer un buen trabajo de ajuste de los cambios en los niveles. Su actualización es de bajo costo y no requiere de un gran almacén de datos. El método de promedio móvil se usa más a menudo cuando son necesarios pronósticos repetitivos.

La suavización exponencial es una técnica difundida cuya fortaleza reside en la exactitud a corto plazo combinada con una actualización rápida y de bajo costo. La técnica es ampliamente usada cuando se necesitan pronósticos mensuales o semanales de un gran número, quizá miles, de artículos. El control de inventario es un ejemplo donde los métodos de suavización exponencial se usan de forma rutinaria.

Glosario

Suavización exponencial. Es un procedimiento de revisión continua de un pronóstico a la luz de la experiencia más reciente.

Promedio móvil. El valor medio de k observaciones consecutivas es un promedio móvil de orden k . Los valores más recientes del promedio móvil brindan un pronóstico para el siguiente periodo.

Promedio simple. Usa la media de todas las observaciones históricas relevantes como pronóstico para el siguiente periodo.

Señal de control. Implica el cálculo de una medida de errores de pronóstico a lo largo del tiempo y el establecimiento de límites de manera que, cuando el error acumulativo salga fuera de esos límites, se alerte al pronosticador.

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

$$\hat{Y}_{t+2} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t+1} \quad (4.7)$$

Promedio móvil para k periodos de tiempo

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-k+1}}{k} \quad (4.8)$$

Promedio móvil doble

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \cdots + M_{t-k+1}}{k} \quad (4.9)$$

$$a_t = 2M_t - M'_t \quad (4.10)$$

$$b_t = \frac{2}{k-1}(M_t - M'_t) \quad (4.11)$$

$$\hat{Y}_{t+p} = a_t + b_t p \quad (4.12)$$

Suavización exponencial simple

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t \quad (4.13)$$

Expresión alternativa equivalente:

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \alpha(1 - \alpha)^3 Y_{t-3} + \cdots \quad (4.14)$$

Suavización 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 futuro:

$$\hat{Y}_{t+p} = L_t + pT_t \quad (4.17)$$

Suavización multiplicativa de Winters

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:

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-s} \quad (4.20)$$

El pronóstico para p periodos futuros:

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (4.21)$$

Problemas

1. ¿Qué técnica de pronósticos revisa continuamente el estimado a la luz de las experiencias más recientes?
2. ¿Cuál técnica de pronósticos usa el valor del periodo actual como base del pronóstico para el siguiente periodo?
3. ¿Cuál técnica de pronósticos concede el mismo peso a cada observación?
4. ¿Cuál(es) técnica(s) de pronósticos debe(n) utilizarse si los datos muestran una tendencia?
5. ¿Cuál(es) técnica(s) de pronósticos debe(n) utilizarse si los datos son estacionales?
6. La Apex Mutual Fund invierte principalmente en acciones de tecnología. El precio del fondo al final de cada mes para los 12 meses de 2006 se presenta en la tabla P-6.
 - a) Obtenga el pronóstico del fondo común por cada mes usando un modelo informal (véase la ecuación 4.1). El valor para diciembre de 2005 fue de 19.00.
 - b) Evalúe este método de pronóstico usando la *MAD*.
 - c) Evalúe este método de pronóstico usando el *MSE*.
 - d) Evalúe este método de pronóstico usando el *MAPE*.
 - e) Evalúe este método de pronóstico usando el *MPE*.
 - f) Utilizando un modelo informal, pronostique el precio del fondo común para enero del 2007.
 - g) Redacte un escrito resumiendo sus hallazgos.
7. Refiérase al problema 6. Use un promedio móvil de tres meses para pronosticar el precio del fondo común para enero de 2007. ¿Es mejor este pronóstico que el pronóstico realizado usando el modelo informal? Explique.
8. A partir de la serie Y_t en la tabla P-8, responda lo siguiente:
 - a) ¿Cuál es el pronóstico para el periodo 9 usando un promedio móvil de cinco meses?
 - b) Si se usa la suavización exponencial simple con una constante de suavización de .4, ¿cuál es el pronóstico para el periodo 4?
 - c) En el inciso b), ¿cuál es el error de pronóstico para el periodo 3?
9. El rendimiento de un bono de obligación general de la ciudad de Davenport fluctúa con el mercado. Las cotizaciones mensuales para 2006 se indican en la tabla P-9.

TABLA P-6

<i>Mes</i>	<i>Precio del fondo común</i>
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

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

- Usando un promedio móvil de tres meses, obtenga el pronóstico del rendimiento mensual de los bonos de obligación a partir de abril.
 - Usando un promedio móvil de cinco meses, obtenga el pronóstico del rendimiento mensual de los bonos de obligación a partir de junio.
 - Evalúe estos métodos de pronósticos usando la *MAD*.
 - Evalúe estos métodos de pronósticos usando el *MSE*.
 - Evalúe estos métodos de pronósticos usando el *MAPE*.
 - Evalúe estos métodos de pronósticos usando el *MPE*.
 - Empleando la mejor técnica, pronostique el rendimiento para enero del 2007.
 - Redacte un escrito resumiendo sus hallazgos.
- Esta pregunta se refiere al problema 9. Use suavización exponencial con una constante de suavización de .2 y un valor inicial de 9.29 para pronosticar el rendimiento de enero de 2007. ¿Es mejor este pronóstico que el pronóstico realizado usando el mejor modelo de promedio móvil? Explique.
 - La Hughes Supply Company utiliza un método de administración del inventario para determinar las demandas mensuales de varios productos. Se tienen registrados los valores de la demanda de los últimos 12 meses de cada producto y están disponibles para futuros pronósticos. Los valores de la demanda de los 12 meses de 2006 de una pieza eléctrica se presentan en la tabla P-11.

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

Fuente: Con base en los registros de Hughes Supply Company.

Use una suavización exponencial con una constante de suavización de .5 y un valor inicial de 205 para pronosticar la demanda para enero de 2007.

12. La General American Investors Company, una compañía de administración de inversiones, invierte principalmente en acciones de mediana y alta calidad. Jim Campbell está estudiando el valor de sus activos por acción para esta compañía y le gustaría pronosticar esta variable para los trimestres restantes de 1996. Los datos se presentan en la tabla P-12.

Evalúe la capacidad para pronosticar el valor de los activos por acción de los siguientes métodos de pronósticos: informal, promedio móvil y suavización exponencial. Cuando compare las técnicas, tome en consideración que el valor real de los activos por acción para el segundo trimestre de 1996 fue de 26.47. Escriba un reporte para Jim indicando cuál método debe usar y por qué.

TABLA P-12 Activos por acción de la General American Investors Company, 1985 a 1996

<i>Año</i>	<i>Trimestre</i>			
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
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).

TABLA P-13 Ingresos de Southdown, 1986 a 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	74.7	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).

13. Southdown, Inc., el tercer productor de cemento más grande de Estados Unidos, está impulsando un programa de quema de residuos de combustible. El costo para Southdown totalizará cerca de los \$37 millones. Por esta razón, es extremadamente importante para la compañía tener un pronóstico exacto de los ingresos para el primer trimestre de 2000. Los datos se presentan en la tabla P-13.
 - a) Use suavización exponencial con una constante de suavización de .4 y un valor inicial de 77.4 para pronosticar los ingresos trimestrales del primer trimestre de 2000.
 - b) Use ahora una constante de suavización de .6 y un valor inicial de 77.4 para pronosticar los ingresos trimestrales del primer trimestre de 2000.
 - c) ¿Qué constante de suavización ofrece el mejor pronóstico?
 - d) Revise el inciso c). Examine las autocorrelaciones residuales. ¿Está usted satisfecho con la suavización exponencial simple para este ejemplo? Explique.
14. La Triton Energy Corporation explora para extraer petróleo y gas. El presidente de la compañía, Gail Freeman, quiere que el analista de la compañía pronostique las ventas por acción de la compañía para el año 2000. Éste será un pronóstico importante, puesto que los planes de reestructuración de Triton han encontrado inconvenientes. Los datos se presentan en la tabla P-14.

Determine cuál es el mejor método para la elaboración del pronóstico y pronostique las ventas por acción para el año 2000.
15. La Consolidated Edison Company vende energía eléctrica (82% de sus ingresos), gas (13%) y vapor (5%) en la ciudad de Nueva York y el condado de Westchester. A Bart Thomas, el encargado de elaborar pronósticos en la empresa, se le asigna la tarea de pronosticar los ingresos trimestrales de la compañía por el resto de 2002 y todo 2003. Él recopila los datos que se presentan en la tabla P-15.

Determine cuál es la mejor técnica para la elaboración del pronóstico y pronostique los ingresos trimestrales para el resto de 2002 y todo 2003.
16. Un fabricante que se especializa en refacciones no tiene sistema de pronósticos y fabrica sus productos con base en las últimas ventas mensuales. Están disponibles los datos de 24 meses de ventas y se presentan en la tabla P-16.

TABLA P-14 Ventas de Triton por acción, 1974 a 1999

<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).

TABLA P-15 Ingresos trimestrales para Consolidated Edison (millones de \$), 1985 a junio de 2002

<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).

a) Grafique los datos de ventas como una serie de tiempo. ¿Los datos son estacionales?

Pista: Para los datos mensuales el periodo estacional es $s = 12$. ¿Existe un patrón que tienda a repetirse cada 12 meses? (por ejemplo, ventas de verano relativamente bajas o ventas de otoño relativamente altas).

TABLA P-16

<i>Mes</i>	<i>Ventas</i>	<i>Mes</i>	<i>Ventas</i>
Enero de 2005	430	Enero 2006	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

- b) Use un modelo informal para generar pronósticos de ventas mensuales (por ejemplo, el pronóstico de febrero de 2005 está dado por el valor de enero de 2005, y así sucesivamente). Calcule el *MAPE*.
- c) Use suavización exponencial simple con una constante de suavización de .5 y un valor inicial suavizado de 430 para producir pronósticos de ventas mensuales. Calcule el *MAPE*.
- d) ¿Piensa que alguno de los modelos de los incisos b) y c) probablemente generará pronósticos de ventas exactos para las ventas mensuales futuras? Explique.
- e) Use Minitab y el método de suavización multiplicativa de Winters con constantes de suavización $\alpha = \beta = \gamma = .5$ para generar el pronóstico de enero de 2007. Guarde los residuos.
- f) Remítase al inciso e). Compare el *MAPE* del método de Winters que salió de la computadora con el *MAPE* de los incisos b) y c). ¿Cuál de los tres métodos para pronosticar prefiere y por qué?
- g) Revise el inciso e). Calcule la autocorrelación de los residuos (para seis retrasos de tiempo) con el procedimiento multiplicativo de Winters. ¿Las autocorrelaciones residuales sugieren que el procedimiento de Winters funciona bien para estos datos? Explique.
17. Considere las compras de gasolina de la Spokane Transit Authority de la tabla 4.2. En el ejemplo 4.3 se utilizó un promedio móvil de cinco semanas para suavizar los datos y generar los pronósticos.
- a) Utilice Minitab para suavizar los datos de la Spokane Transit Authority, usando un promedio móvil de cuatro semanas. ¿Cuál longitud de promedio móvil (cuatro o cinco semanas) parece representar mejor los datos? Explique.
- b) Utilice Minitab para suavizar los datos de Spokane Transit Authority usando la suavización exponencial simple. Compare sus resultados con los del inciso a). ¿Qué procedimiento, entre el promedio móvil de cuatro semanas o la suavización exponencial simple, prefiere para estos datos? Explique.
18. La tabla P-18 indica el número de terremotos severos anuales (aquellos con una magnitud en la escala de Richter de 7 grados o más) de 1900 a 1999.
- a) Utilice Minitab para suavizar los datos de terremotos con promedios móviles de órdenes de $k = 5, 10$ y 15 . Describa la naturaleza de la suavización conforme el orden del promedio móvil se incrementa. ¿Cree usted que podría haber un ciclo en estos datos? Si es así, dé un estimado de la duración (en años) del ciclo.

TABLA P-18 Número de terremotos severos, 1900 a 1999

Año	Número	Año	Número	Año	Número	Año	Número
1900	13	1927	20	1954	17	1981	14
1901	14	1928	22	1955	19	1982	10
1902	8	1929	19	1956	15	1983	15
1903	10	1930	13	1957	34	1984	8
1904	16	1931	26	1958	10	1985	15
1905	26	1932	13	1959	15	1986	6
1906	32	1933	14	1960	22	1987	11
1907	27	1934	22	1961	18	1988	8
1908	18	1935	24	1962	15	1989	7
1909	32	1936	21	1963	20	1990	12
1910	36	1937	22	1964	15	1991	11
1911	24	1938	26	1965	22	1992	23
1912	22	1939	21	1966	19	1993	16
1913	23	1940	23	1967	16	1994	15
1914	22	1941	24	1968	30	1995	25
1915	18	1942	27	1969	27	1996	22
1916	25	1943	41	1970	29	1997	20
1917	21	1944	31	1971	23	1998	16
1918	21	1945	27	1972	20	1999	23
1919	14	1946	35	1973	16		
1920	8	1947	26	1974	21		
1921	11	1948	28	1975	21		
1922	14	1949	36	1976	25		
1923	23	1950	39	1977	16		
1924	18	1951	21	1978	18		
1925	17	1952	17	1979	15		
1926	19	1953	22	1980	18		

Fuente: U.S. Geological Survey Earthquake Hazard Program.

- b) Use Minitab para suavizar los datos de los terremotos usando la suavización exponencial simple. Almacene los residuos y genere un pronóstico para el número de terremotos severos en el año 2000. Calcule las autocorrelaciones residuales. ¿La suavización exponencial simple ofrece un ajuste razonable de estos datos? Explique.
- c) ¿Existe un componente estacional en los datos del terremoto? ¿Por qué?
19. La tabla P-23 del capítulo 3 contiene los ingresos trimestrales antes de ingresos extraordinarios de la Southwest Airlines de 1988 a 1999.
- a) Con Minitab, suavice los datos de ingresos de la Southwest Airlines utilizando la suavización lineal de Holt y guarde los residuos. Calcule las autocorrelaciones de los residuos. ¿Parece que el procedimiento de suavización de Holt representa bien estos datos? Si no es así, ¿qué componente de la serie de tiempo (tendencia, ciclo, componente estacional) no toma en cuenta el método de Holt?
- b) Emplee Minitab para suavizar los datos de ingresos de Southwest Airlines con la suavización exponencial multiplicativa de Winters. Guarde los residuos y genere pronósticos de ingresos para los cuatro trimestres de 2000. Calcule las autocorrelaciones residuales. ¿La técnica de suavización de Winters se ajusta bien a los datos de ingresos? ¿El pronóstico parece razonable? Discútalos.

TABLA P-20 Ventas trimestrales para The Gap en los años fiscales 1980 a 2004

<i>Año</i>	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>
1980	65.3	72.1	107.1	125.3
1981	76.0	76.9	125.4	139.1
1982	84.4	86.0	125.4	139.1
1983	87.3	90.9	133.3	169.0
1984	98.0	105.3	141.4	189.4
1985	105.7	120.1	181.7	239.8
1986	160.0	164.8	224.8	298.4
1987	211.1	217.7	273.6	359.6
1988	241.3	264.3	322.8	423.7
1989	309.9	325.9	405.6	545.2
1990	402.4	405.0	501.7	624.7
1991	490.3	523.1	702.0	803.5
1992	588.9	614.1	827.2	930.2
1993	643.6	693.2	898.7	1,060.2
1994	751.7	773.1	988.3	1,209.8
1995	848.7	868.5	1,155.9	1,522.1
1996	1,113.2	1,120.3	1,383.0	1,667.9
1997	1,231.2	1,345.2	1,765.9	2,165.5
1998	1,719.7	1,905.0	2,399.9	3,029.8
1999	2,277.7	2,453.3	3,045.4	3,858.9
2000	2,732.0	2,947.7	3,414.7	4,579.1
2001	3,179.7	3,245.2	3,333.4	4,089.6
2002	2,890.8	3,268.3	3,645.0	4,650.6
2003	3,353.0	3,685.0	3,929.0	4,887.0
2004	3,668.0	3,721.0	3,980.0	4,898.0

Fuente: De acuerdo con *The Value Line Investment Survey* (Nueva York: Value Line, varios años) y archivos 10K de la Security and Exchange Comisión.

20. La tabla P-20 contiene las ventas trimestrales (en millones de \$) para The Gap de los años fiscales 1980 a 2004.

Grafique las ventas de The Gap como una serie de tiempo y examine sus características. El objetivo es generar pronósticos de ventas para los cuatro trimestres de 2005. Seleccione un método de suavización adecuado para la elaboración del pronóstico y justifique su selección.

CASOS

CASO 4-1 THE SOLAR ALTERNATIVE COMPANY⁴

The Solar Alternative Company está por entrar a su tercer año de operación. Bob y Mary Johnson, quienes dan clases de ciencias en la preparatoria local, fun-

daron la compañía. Los Johnson abrieron la compañía para complementar sus ingresos. Con base en sus investigaciones sobre sistemas de energía solar, lograron

⁴Este caso es contribución de William P. Darrow de Towson State University, Towson, Maryland.

integrar un sistema de energía solar que permitiera calentar el agua para uso doméstico. El sistema consiste en un tanque de almacenamiento de fibra de vidrio de 100 galones, dos paneles solares de 36 pies, controles electrónicos, tubería PVC y accesorios adicionales.

El periodo de recuperación del sistema es de 10 años. Si bien esta situación no es una oportunidad de inversión atractiva desde el punto de vista financiero, existe suficiente interés en lo novedoso del concepto como para generar un nivel de ventas moderado. Los Johnson ganan cerca de \$75 sobre un precio de \$2000 de un sistema instalado, después de gastos y costos. Los costos de material y equipo representan el 75% del costo total del sistema instalado. Una ventaja que ayuda a compensar el escaso margen de ganancias es que, por lo mismo, no genera competencia significativa entre los contratistas del ramo. Los Johnson operan el negocio fuera de su casa. Tienen una oficina en el sótano y su cochera la usan exclusivamente para almacenar los componentes y materiales del sistema. Como resultado, los gastos indirectos son mínimos. Los Johnson disfrutan de un modesto ingreso adicional por la operación de su compañía. El negocio también proporciona varias ventajas fiscales.

Bob y Mary están contentos con el crecimiento del negocio. Si bien las ventas son variables de un mes a otro, los resultados totales del segundo año fueron mucho mejores que los del primero. Muchos de los

clientes del segundo año eran vecinos de las personas que compraron el sistema durante el primer año. Aparentemente, después de ver que el sistema operó con éxito durante un año, varias personas estuvieron dispuestas a probar el concepto de energía solar. Las ventas se presentaron a lo largo del año. La demanda del sistema es mayor al final del verano y al principio del otoño, cuando la gente por lo general prepara sus casas para el invierno que se acerca.

Ante el probable crecimiento del negocio, los Johnson sienten que necesitan un pronóstico de las ventas para administrar efectivamente el año entrante. Por lo general, reciben los tanques de almacenamiento de 60 a 90 días después de hacer el pedido. Los paneles solares están disponibles en los anaqueles la mayor parte el año. Sin embargo, al final del verano y durante el otoño, el tiempo de entrega se puede prolongar de 90 a 100 días. Aunque la competencia es limitada, la pérdida de las ventas es una posibilidad real si al cliente potencial se le pide esperar algunos meses para hacer la instalación. Quizás lo más importante es hacer proyecciones exactas de las ventas para tener la ventaja de un descuento por una compra voluminosa. Cuando se combinan estos factores, con los altos costos de los componentes del sistema y el espacio limitado de almacenamiento en la cochera, se hace necesario desarrollar un pronóstico confiable. Las ventas históricas de la compañía para los primeros dos años se presentan en la tabla 4-10.

TABLA 4-10

<i>Mes</i>	<i>2005</i>	<i>2006</i>	<i>Mes</i>	<i>2005</i>	<i>2006</i>
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

TAREA

1. Identifique el modelo que Bob y Mary deberían utilizar como base para la planeación de su negocio en 2007 y explique por qué seleccionó este modelo.
2. Pronostique las ventas de 2007.

CASO 4-2 MR. TUX

John Mosby, propietario de varias tiendas de alquiler Mr. Tux, está empezando a pronosticar la variable más importante de su negocio: las ventas mensuales en dólares (véase los casos de Mr. Tux en capítulos anteriores). Uno de sus empleados, Virginia Perot, reunió los datos de ventas que se presentaron en el caso 2-2. Ahora, John quiere elaborar un pronóstico con base en estos datos de ventas empleando técnicas de promedio móvil y suavización exponencial.

John usó Minitab en el caso de estudio 3-2 para determinar que estos datos tienen una tendencia y un efecto estacional. Le han dicho que las técnicas de los promedios móviles simples y de la suavización exponencial no funcionarán con estos datos, pero él decidió averiguarlo por sí mismo.

Comenzó con un promedio móvil de tres meses. El programa calculó varias mediciones de resumen de errores de pronóstico. Estos valores resumen los errores encontrados en la predicción de valores de datos históricos reales usando un promedio móvil de tres meses. John decidió registrar dos de estas mediciones de error:

$$MAD = 54,373$$

$$MAPE = 47.0\%$$

La *MAD* (la desviación media absoluta) es el error absoluto promedio obtenido en la elaboración de pronósticos de valores pasados. Todo pronóstico que use el método de promedios móviles de tres meses es infructuoso con un promedio de error de 54,373. El *MAPE* (el error porcentual absoluto medio) indica el error como un porcentaje del valor real que se va a pronosticar. El error promedio al usar la técnica de promedio móvil de tres meses es del 47%, o casi tan grande como la mitad del valor que se va a pronosticar.

A continuación, John intenta la suavización exponencial simple. El programa le solicita que introduzca la constante de suavización α que habrá de utilizarse o que pida calcular el valor óptimo de α . John opta por

esto último y el programa encuentra que el valor óptimo de α es .867. De nuevo, él registra las mediciones apropiadas de error:

$$MAD = 46,562$$

$$MAPE = 44.0\%$$

John le pide al programa usar la suavización exponencial lineal de Holt para sus datos. Este programa usa el método de suavización exponencial, pero también puede tomar en cuenta la tendencia en los datos. John hace que el programa use una constante de suavización de .4 tanto para α como para β . Los dos resúmenes de mediciones de error para el método de Holt son:

$$MAD = 63,579$$

$$MAPE = 59.0\%$$

John está sorprendido de encontrar mediciones de error más grandes con esta técnica. Concluye que el problema es el aspecto estacional de los datos. La suavización exponencial multiplicativa de Winters es el siguiente método que John emplea. Este método puede tomar en cuenta tanto factores estacionales como tendencias. John usa constantes de suavización de $\alpha = .2$, $\beta = .2$ y $\gamma = .2$. Las mediciones de error son:

$$MAD = 25,825$$

$$MAPE = 22.0\%$$

Cuando John se sienta y empieza a estudiar los resultados de su análisis, se siente desilusionado. El método de Winters es una gran mejora; sin embargo, el *MAPE* es del 22%. Tenía la esperanza de que uno de los métodos que usó daría pronósticos exactos de los periodos pasados; luego, podría usar este método para pronosticar los niveles de ventas de los meses venideros del próximo año. Pero el promedio de errores absolutos (*MAD*) y los errores porcentuales (*MAPE*) de estos métodos lo inducen a buscar otra manera de realizar el pronóstico.

PREGUNTAS

1. Resuma el nivel de error de pronóstico del mejor método de pronósticos que John encontró usando Minitab.
2. John usó los valores que Minitab predeterminó para α , β , y γ . John cree que existen otras opciones para estos parámetros que implicarían mediciones del error más pequeñas. ¿Está usted de acuerdo?
3. Aun cuando se desilusionó con los resultados iniciales, esto es lo mejor que puede hacer con métodos de suavización. ¿Qué puede hacer John, por ejemplo, para determinar qué tan adecuada es la técnica de pronósticos de Winters?
4. Si bien no se calculó directamente en Minitab, el *MPE* (error porcentual medio) mide el sesgo del pronóstico. ¿Cuál es el valor ideal del *MPE*? ¿Cuál es la implicación de un signo negativo en el *MPE*?

CASO 4-3 CONSUMER CREDIT COUNSELING

La operación del Consumer Credit Counseling (CCC) se describió en el caso 1-2. El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que necesitaba pronosticar CCC era el número de clientes nuevos que serían atendidos por el resto de 1993. Marv proporcionó a Dorothy Mercer datos mensuales

del número de clientes nuevos atendidos por CCC durante el periodo de enero de 1985 a marzo de 1993 (véase el caso 3-3). Entonces Dorothy utilizó el análisis de autocorrelación para explorar el patrón de datos. Use los resultados de esta investigación para realizar las siguientes tareas.

TAREA

1. Desarrolle un modelo informal para pronosticar el número de clientes nuevos que atenderá CCC por el resto de 1993.
2. Desarrolle un modelo de promedio móvil para pronosticar el número de clientes nuevos que atenderá CCC por el resto de 1993.
3. Desarrolle un procedimiento de suavización exponencial para pronosticar el número de clientes nuevos que atenderá CCC por el resto de 1993.
4. Evalúe estos métodos de pronósticos usando las medidas de resumen de errores de pronósticos presentadas en el capítulo 3.
5. Seleccione el mejor modelo y pronostique los nuevos clientes para el resto de 1993.
6. Determine la idoneidad del modelo de pronóstico que usted seleccionó.

CASO 4-4 MURPHY BROTHERS FURNITURE

Julie Murphy sabe que las decisiones operativas más importantes dependen, en algún grado, de los pronósticos. En Murphy Brothers Furniture, los pronósticos de ventas tienen efecto en la incorporación de nuevas líneas de muebles o la eliminación de algunas existentes, en la planeación de compras, en el establecimiento de cuotas de ventas, y también en la toma de decisiones relacionadas con el personal, la publicidad y las finanzas. Específicamente, Julie está consciente de varias necesidades actuales de pronósticos. Sabe que el departamento de producción tiene que programar su plantilla de personal y definir los pedidos de materia prima para el siguiente mes o los siguientes dos. Ella también sabe que su papá, Glen Murphy, necesita iden-

tificar las mejores oportunidades de inversión y debe pronosticar la demanda de una nueva línea de muebles.

En el caso 3-1A, Julie Murphy usó las ventas mensuales de todas las tiendas minoristas de 1983 a 1995 (véase la tabla 3-8) para desarrollar un patrón de ventas en Murphy Brothers Furniture. En el caso de estudio 3-1B, Glen Murphy descubrió datos de ventas reales de los pasados cuatro años, de 1992 a 1995 (véase la tabla 3-9). Julie no estaba emocionada por el descubrimiento de su padre porque no estaba segura cuál conjunto de datos usar para desarrollar un pronóstico para 1996. Ella determinó que las ventas de todas las tiendas minoristas tenían algo del patrón de los datos de ventas reales de Murphy Brothers.

PREGUNTAS

1. ¿Alguno de los modelos de pronósticos estudiados en este capítulo funcionará con los datos de ventas nacionales?
2. ¿Alguno de los modelos de pronósticos estudiados en este capítulo funcionará con los datos de ventas reales de Murphy Brothers?
3. ¿Qué conjunto de datos y qué modelo de pronóstico debe usar Julie para pronosticar las ventas de 1996?

CASO 4-5 PROYECCIÓN DE CINCO AÑOS DE INGRESOS PARA DOWNTOWN RADIOLOGY

Hace algunos años, Downtown Radiology desarrolló un centro de imagenología médica que era el más completo y técnicamente más avanzado que cualquiera de los localizados en el área del oriente de Washington y el norte de Idaho, llamado el Inland Empire. El equipo planeado para el centro igualaba o superaba las capacidades de imagenología de todos los centros médicos en la región. El centro inicialmente contaba con un escáner de tomografía computarizada (TC) de la serie 9800 y equipo para obtener imágenes de resonancia magnética nuclear (MRI). El centro también tenía equipos de ultrasonido, medicina nuclear, angiografías por sustracción digital (DSA), mamografía, así como de radiología y fluoroscopia convencionales. Se hizo posible la participación accionaria mediante una oferta pública, y Downtown Radiology hizo una evaluación independiente del mercado. La Professional Marketing Associates, Inc., evaluó el mercado y elaboró una proyección de ingresos para cinco años.

DEFINICIÓN DEL PROBLEMA

El propósito de este estudio es elaborar un pronóstico de los ingresos para los próximos cinco años del centro de imagenología médica propuesto, suponiendo que usted es empleado de Professional Marketing Associates, Inc., en el año de 1984.

OBJETIVOS

Los objetivos de este estudio son:

- Identificar áreas de mercado para cada tipo de procedimiento médico que se ofrecerá en las nuevas instalaciones.
- Reunir y analizar los datos de ingresos existentes en el área del mercado para cada tipo de procedimiento médico ofrecido en las nuevas instalaciones.
- Identificar tendencias en la industria de cuidados de la salud que afectarán positiva o negativamente los ingresos de los procedimientos médicos que se ofrecerán en las nuevas instalaciones.
- Identificar los factores de negocio, marketing o planeación de la nueva empresa que afectarán positiva o negativamente las proyecciones de ingresos.
- Analizar los procedimientos médicos anteriores de Downtown Radiology cuando se compile una base de datos para el modelo del pronóstico que será desarrollado.
- Utilizar el modelo apropiado de pronóstico cuantitativo para obtener las proyecciones de ingresos del centro propuesto para cinco años.

METODOLOGÍA

Procedimientos médicos

Se realizaron los siguientes pasos con la finalidad de obtener la proyección de ingresos para los próximos cinco años. Se realizó un análisis de los procedimientos médicos anteriores. Se desarrolló el modelo de pronóstico adecuado y se usó para determinar un punto de inicio de la proyección de cada procedimiento médico.

1. Se determinó el área de mercado para cada tipo de procedimiento médico y el pronóstico de la población para 1986 y 1990.
2. Se estudiaron los patrones de recomendaciones médicas para determinar el porcentaje de doctores que sugerían los servicios de Downtown Radiology y el número promedio de recomendaciones por médico.
3. Las tarifas nacionales se obtuvieron del Centro Nacional de Estadísticas para la Salud. Estas tarifas se compararon con los números reales obtenidos por la Comisión Hospitalaria.
4. La participación en el mercado de Downtown Radiology se calculó con base en las tomografías computarizadas reales en el área de mercado. (La participación en el mercado de otros procedimientos médicos se determinó con base en la comparación de la participación de Downtown Radiology con las tarifas proporcionadas por el Centro Nacional de Estadísticas para la Salud).

Suposiciones

Se hicieron las siguientes suposiciones, las cuales fueron necesarias para desarrollar el pronóstico cuantitativo:

- El nuevo centro de imagenología estará en condiciones de operar con todo el equipo, a excepción del MRI, el 1 de enero de 1985.
- El equipo de resonancia magnética nuclear comenzará a funcionar en abril de 1985.
- La oferta de participación limitada será comercializada con éxito entre por lo menos 50 médicos en el área de servicio.
- Los médicos que tengan un interés financiero en el nuevo centro de imagenología aumentarán sus remisiones para el centro.
- No habrá otros equipos de MRI en el área de mercado antes de 1987.
- El nuevo centro de imagenología ofrecerá servicios a precios más bajos que la competencia.

- Tendrá lugar un esfuerzo efectivo de marketing, concentrado especialmente en grandes empleadores, compañías de seguros y sindicatos.
- El MRI reemplazará aproximadamente al 60% de las tomografías craneales que actualmente se practican con el escáner de TC durante los primeros seis meses de operación y al 70% durante los siguientes 12 meses.
- El público en general seguirá presionando a la industria del cuidado de la salud para mantener costos bajos.
- Los costos de los desembolsos en la industria del cuidado de la salud se elevaron 13.2% anualmente de 1971 a 1981. La Health Care Financing Administration estima que la tasa promedio anual de incremento se reducirá aproximadamente al 11% o 12% entre 1981 y 1990 (*Industry Surveys*, abril de 1984).
- Las compañías de seguros reembolsarán a los pacientes del 0% (en el peor de los casos) hasta el 100% (en el mejor de los casos) del costo de la imagenología por resonancia magnética (*Imaging News*, febrero de 1984).

Modelos

Se desarrolló un pronóstico para cada procedimiento médico, con base en la experiencia pasada, las cifras de la industria y supuestos razonables. Los modelos se desarrollaron con base en los supuestos anteriores; sin embargo, si los supuestos no son válidos, los modelos serán inexactos.

ANÁLISIS DE DATOS ANTERIORES

Oficina de rayos X

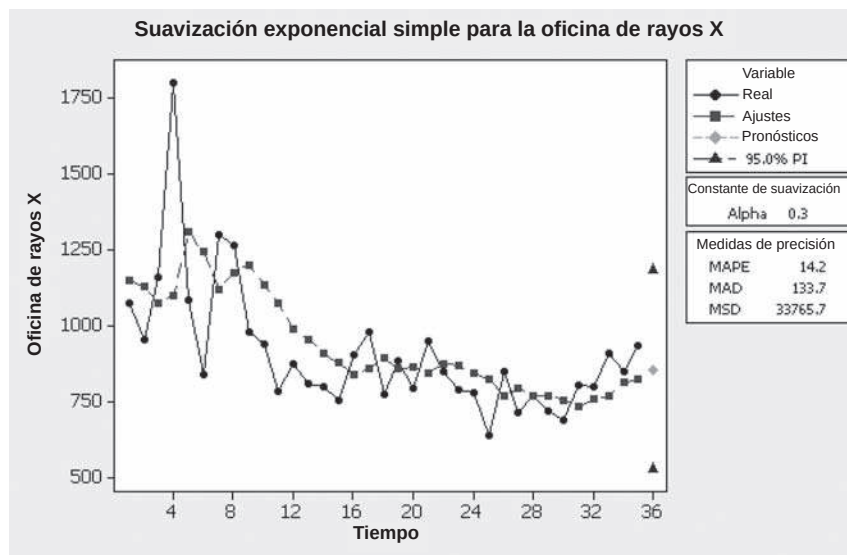
Se analizó el número de procedimientos médicos de rayos X practicados de julio de 1981 a mayo de 1984. Los datos incluyeron diagnóstico de rayos X, rayos X gastrointestinales, imagenología de tórax, inyecciones y procedimientos médicos especiales. La revisión de estos datos indicó que no existía un patrón de tendencia, ni tampoco estacional ni cíclico. Por esta razón, se optó por la suavización exponencial simple como el método adecuado de pronóstico. Se examinaron varias constantes de suavización y se encontró que una constante de .3 proporcionaba el mejor modelo. Los resultados se presentan en la figura 4-14. El pronóstico para el periodo 36, de junio de 1984, es de 855 procedimientos médicos de rayos X.

Oficina de ultrasonido

Se analizó el número de procedimientos médicos de ultrasonido realizados de julio de 1981 a mayo de 1984. La figura 4-15 muestra el patrón de datos. Nuevamente, no están presentes patrones de tendencia, ni tampoco estacionales ni cíclicos. Se determinó que la suavización exponencial con una constante de suavización $\alpha = .5$ ofrecía el mejor modelo. El pronóstico para junio de 1984 graficado en la figura 4-15 es de 127 procedimientos médicos de ultrasonido.

Se revisaron varios procedimientos médicos de ultrasonido ejecutados por las dos unidades móviles pertenecientes al Downtown Radiology de julio de 1981 a mayo de 1984. La figura 4-16 muestra el patrón de los datos.

FIGURA 4-14 Suavización exponencial simple: Downtown Radiology, rayos X



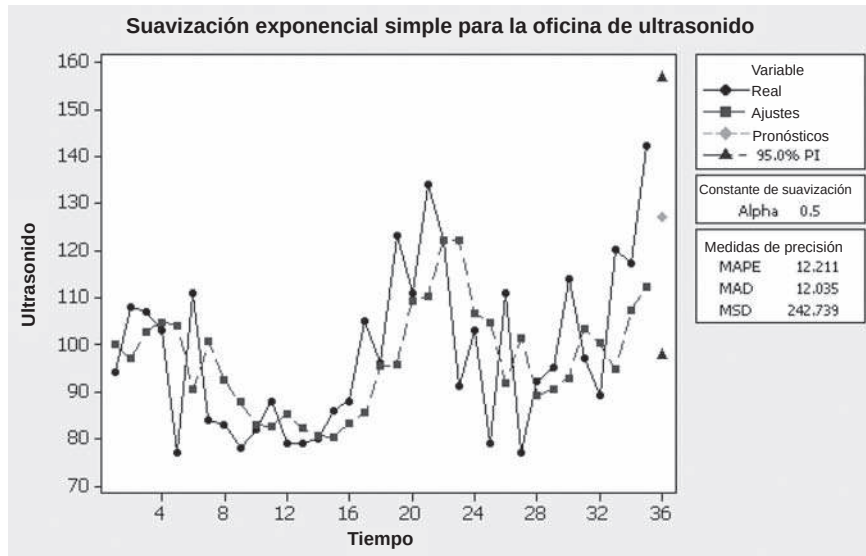


FIGURA 4-15 Suavización exponencial simple: Ultrasonido de Downtown Radiology

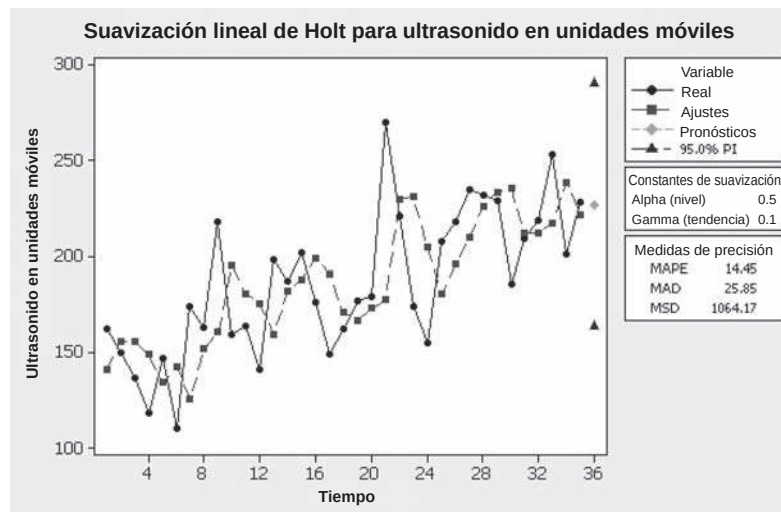


FIGURA 4-16 Suavización exponencial lineal de Holt: Ultrasonidos en unidades móviles de Downtown Radiology

Como es evidente, existe una tendencia ascendente y puede modelarse usando la suavización exponencial lineal de Holt con dos parámetros. Se usaron las constantes de suavización $\alpha = .5$ y $\beta = .1$ y el pronóstico del periodo 36, correspondiente a junio de 1984, es de 227.

Procedimientos de medicina nuclear

Se analizaron los procedimientos de medicina nuclear practicados en dos unidades móviles pertenecientes al Downtown Radiology de agosto de 1982 a mayo de

1984. La figura 4-17 muestra el patrón de datos. Los datos no fueron estacionales y no mostraron una tendencia, ni tampoco un patrón cíclico. Por esta razón, se eligió la suavización exponencial simple como el método apropiado de elaboración del pronóstico. Se encontró que un factor de suavización $\alpha = .5$ ofrecía el mejor modelo. El pronóstico del periodo 23, correspondiente a junio de 1984, es de 48 procedimientos médicos de medicina nuclear.

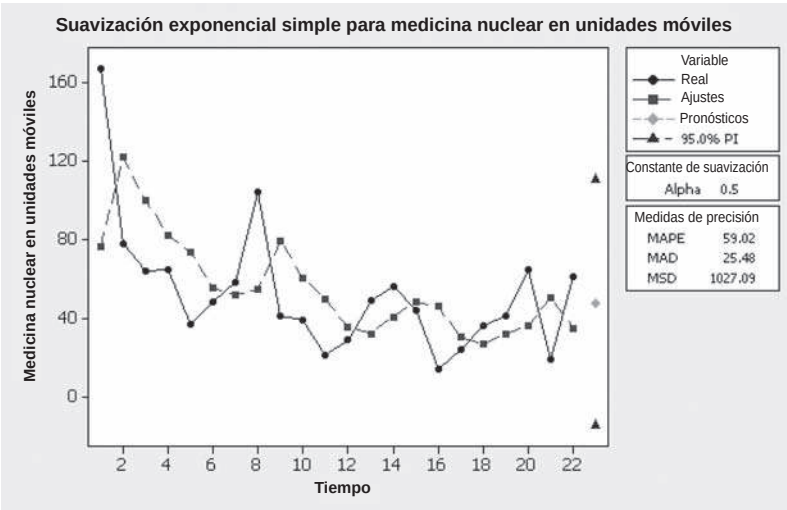


FIGURA 4-17 Suavización exponencial simple: Medicina nuclear en unidades móviles de Downtown Radiology

Oficina de tomografías computarizadas (TC)

También se analizó el número de tomografías computarizadas (TC) practicadas de julio de 1981 a mayo de 1984. No se encontró estacionalidad y el número de tomografías realizadas no parecía seguir una tendencia. Sin embargo, parecía estar presente un patrón cíclico. El hecho de saber cuántas tomografías se practicaron el mes pasado sería importante para pronosticar qué va a pasar este mes. Se examinó un modelo auto-

regresivo (véase los capítulos 8 y 9) y se comparó con un modelo de suavización exponencial con una constante de suavización $\alpha = 0.461$. La constante de suavización más grande asignó más peso en el pronóstico a la observación más reciente. Se determinó que el modelo de suavización exponencial era mejor que el modelo autorregresivo, y la figura 4-18 indica que la proyección del número de tomografías computarizadas para el periodo 36, junio de 1984, es de 221.

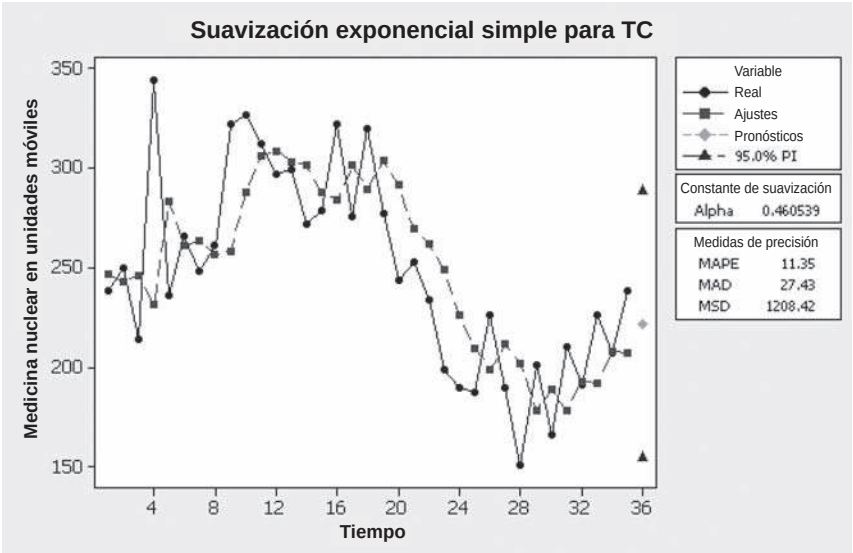


FIGURA 4-18 Suavización exponencial simple: Tomografías computarizadas de Downtown Radiology

ANÁLISIS DEL ÁREA DE MERCADO

Se determinaron las áreas de mercado de procedimientos médicos que actualmente realiza Downtown Radiology revisando los registros de los pacientes y los patrones de recomendaciones de los médicos. Se determinaron las áreas de mercado para procedimientos médicos *no* realizados actualmente en Downtown Radiology investigando la competencia y analizando las áreas geográficas que atienden.

Áreas de mercado de tomografías computarizadas (TC)

El área de mercado de las tomografías computarizadas para el centro de imagenología médico propuesto incluye los condados de Spokane, Whitman, Adams, Lincoln, Stevens y Pend Oreille, en Washington, y los condados de Bonner, Boundary, Kootenai, Benewah y Shoshone en Idaho. Con base en las proyecciones porcentuales pertinentes, el área del mercado de las tomografías computarizadas tendrá una población de 630,655 en 1985 y 696,018 en 1990.

Estimaciones cuantitativas

Para proyectar los ingresos, es necesario determinar ciertas estimaciones cuantitativas. La estimación más importante implica el número de médicos que participarán en la asociación limitada. La estimación utilizada en cálculos futuros es que participará, al menos, el 8% de la población de médicos del condado de Spokane.

La siguiente incertidumbre que debe ser cuantificada implica la determinación de cómo se verá afectado el patrón de recomendaciones con la participación de 50 médicos en la asociación limitada. Se supone que 30 de los doctores que actualmente recomiendan al Downtown Radiology se unirán a la asociación limitada. De esos 30, se supone que 10 no incrementarán sus recomendaciones y 20 duplicarán sus recomendaciones. También se supone que 20 doctores que nunca habían recomendado al Downtown Radiology se unirán a la asociación limitada y empezarán a recomendar Downtown Radiology a por lo menos la mitad de sus pacientes.

La cuantificación de las recomendaciones adicionales de los médicos debe clarificarse con algunas observaciones cualitativas. La estimación de que 50 doctores se unirán a la propuesta de asociación limitada es conservadora. Existe una gran posibilidad de que médicos de las áreas fuera del condado de Spokane se puedan unir. Tradicionalmente, el patrón de recomendaciones de los doctores cambia muy lentamente. Sin embargo, la súbita naturaleza competitiva del mercado local probablemente tendrá repercusión en las recomendaciones de los doctores. Si la asociación limitada se ofrece entre médicos especialistas con alto potencial para hacer recomendaciones radiológicas, el número de recomendaciones debe incrementarse más de lo proyectado. La variabilidad en el número de recomenda-

ciones por procedimiento médico es extremadamente grande. Unos cuantos médicos recomiendan un porcentaje extremadamente alto de procedimientos médicos hechos por Downtown Radiology. Si se reclutan unos cuantos médicos nuevos que hagan muchas recomendaciones, ellos pueden tener un mayor efecto sobre el número total de procedimientos médicos realizados por Downtown Radiology.

Finalmente, se debe estimar el efecto que tendrá un nuevo centro de imagenología en la participación de mercado de Downtown Radiology. El nuevo centro de imagenología tendrá el mejor equipamiento y estará preparado para cubrir el espectro total de procedimientos médicos al costo más bajo. El número de nuevos doctores que recomienden el centro debe crecer con base en la recomendación de boca a boca de los nuevos médicos inversionistas. Si las compañías de seguros, los grandes empleadores y/o los sindicatos entran en los acuerdos con el nuevo centro, Downtown Radiology debe estar en condiciones de incrementar su participación en el mercado por lo menos en un 4% en 1985, 2% en 1986 y 1% en 1987 y conservar su participación en el mercado durante 1988 y 1989. Este incremento en la participación de mercado será mencionado como el *efecto total de imagenología* en el resto de este reporte.

Proyecciones de ingresos

Se realizaron proyecciones de ingresos para cada procedimiento médico. En este caso de estudio se presentan sólo las proyecciones de las tomografías computarizadas.

Proyecciones de tomografías computarizadas (TC)

Con base en el modelo de suavización exponencial y lo que ha tenido lugar en los primeros cinco meses de 1984, el pronóstico de tomografías computarizadas para 1984 (de enero de 1984 a enero de 1985) es de 2,600.

El Centro Nacional de Estadísticas para la Salud reporta una tasa de 261 TC mensuales por cada 100,000 habitantes. Utilizando la cifra proyectada de la población de 630,655 para el área de mercado de TC, el mercado debe realizar 19,752 procedimientos durante todo 1985. El número real de TC realizadas en el área de mercado durante 1983 se estimó en 21,600. Esta estimación se basó en los procedimientos reales conocidos de Downtown Radiology (2,260), del Sacred Heart (4,970), del Deaconess (3,850), del Valley (2,300) y del Kootenai (1,820), así como en las estimaciones de Radiation Therapy (2,400) y Northwest Imaging (4,000). Si las estimaciones son exactas, Downtown Radiology tenía una participación en el mercado de aproximadamente el 10.5% en 1983. También se analizaron los valores reales de 1982, y se proyectó que Downtown Radiology tendría aproximadamente el 15.5% del mercado de TC durante ese año. Por lo tanto, se pronostica que Downtown Radiology tendrá un promedio del 13% del mercado.

Con base en las recomendaciones crecientes de los médicos que pertenecen a la sociedad limitada y en un análisis del promedio de recomendaciones de TC, se proyectó un incremento de 320 TC para 1985. Si se usan los valores reales de 1983, la tasa para el área de mercado de TC de Inland Empire es de 3,568 (21,600/6.054) por cada 100,000 habitantes. Si este patrón continúa, el número de TC en el área de mercado aumentará a 22,514 ($3,568 \times 6.31$) en 1985. Por consiguiente, la proyección de la participación en el mercado de Downtown Radiology será del 13% ($2,920/22,514$). Cuando se suma el 4% de incremento en la participación en el mercado basado en el total de imagenología, la participación en el mercado de Downtown Radiology se incrementa al 17% y el número proyectado de TC es de 3,827 ($22,514 \times .17$).

Sin embargo, la investigación parece indicar que las imágenes por resonancia magnética (MRI), con el tiempo, reemplazarán a un gran número de tomografías craneales (*Applied Radiology*, mayo/junio de 1983, y *Diagnostic Imaging*, febrero de 1984). El Centro Nacional de Estadísticas para la Salud indicaba que el 60% de todas las TC eran craneales. Los registros de Downtown Radiology mostraban que el 59% de sus TC en 1982 fueron de la cabeza y el 54% en 1983. Si el 60% de las TC de Downtown Radiology son craneales y considerando que el método de MRI sustituirá aproximadamente al 60% de éstas, se necesitan nuevas proyecciones de TC para 1985. Puesto que el MRI operará sólo la mitad del año, se pronostica una disminución de 689 ($3,827/2 \times .60 \times .60$) tomografías computarizadas.

El número proyectado de TC para 1985 es de 3,138. El costo promedio de una TC es de \$360, y el

TABLA 4-11 Ingresos proyectados de cinco años para TC

<i>Año</i>	<i>Procedimientos</i>	<i>Ingresos (\$)</i>
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

ingreso proyectado de TC es de \$1,129,680. La tabla 4-11 indica el ingreso proyectado de TC para los siguientes cinco años. Se estima que el costo de estos procedimientos médicos se incrementa en aproximadamente un 11% por año.

Sin el efecto de MRI, la proyección para las TC para 1986 se estima en 4,363 ($6.31 \times 1.02 \times 3,568 \times .19$). Sin embargo, si el 60% corresponde a TC craneales y la tecnología de MRI sustituye al 70% de las TC craneales, el número proyectado de TC disminuirá a 2,531 [$4,363 - (4,363 \times .60 \times .70)$].

La proyección de TC sin el efecto MRI para 1987 es de 4,683 ($6.31 \times 1.04 \times 3,568 \times .20$). El pronóstico con el efecto MRI es de 2,716 [$4,683 - (4,683 \times .60 \times .70)$].

La proyección de TC sin el efecto MRI para 1988 es de 4,773 ($6.31 \times 1.06 \times 3,568 \times .20$). El pronóstico con el efecto MRI es de 2,482 [$4,733 - (4,733 \times .60 \times .80)$].

La proyección de TC sin el efecto MRI para 1989 es de 4,863 ($6.31 \times 1.08 \times 3,568 \times .20$). El pronóstico con el efecto MRI es de 2,529 [$4,863 - (4,863 \times .60 \times .80)$].

PREGUNTA

1. El contador de Downtown Radiology proyectó unos ingresos considerablemente mayores en comparación con los que reportó Professional Marketing Associates. Puesto que la participación accionaria será posible mediante una oferta pública, la

administración de Downtown Radiology debe tomar una decisión en lo referente a la exactitud de las proyecciones de Professional Marketing Associates. A usted se le pide analizar el reporte. ¿Qué recomendaciones haría?

CASO 4-6 MINORISTA EN LA WEB

En el ejemplo 1.2 se presentó a Pat Niebuhr y su equipo, quienes son responsables del desarrollo de un plan global de dotación de personal para los centros de contacto de un minorista grande en la Web. Pat nece-

sita obtener un pronóstico mensual del total de los pedidos y los contactos totales por pedido (CPP) de los cuales le informará el departamento de finanzas, para así pronosticar el número de contactos semanales

con el cliente (por teléfono, correo electrónico, etcétera) que se establecen en los centros de contacto. Los centros de contacto están abiertos 24 horas al día, siete días a la semana y deben contar con el personal suficiente para mantener un alto nivel de servicio. El minorista reconoce que un excelente servicio al cliente probablemente ayudará a que éste regrese y compre de nuevo.

La ecuación clave para Pat y su equipo es:

$$\text{Contactos} = \text{Pedidos} \times \text{CPP}$$

Los datos históricos indican el porcentaje de contactos por cada día de la semana. Por ejemplo, históricamente el 9.10% de los contactos semanales ocurren los domingos, el 17.25% de los contactos semanales ocurren los lunes, y así sucesivamente. Teniendo en mente el número de domingos, lunes, etcétera, en un mes determinado, los pronósticos mensuales de contactos pueden convertirse a pronósticos semanales de contactos. Los pronósticos semanales son los que se utilizan para efectos de planeación de la plantilla.

Pat piensa que podría ser una buena idea usar los datos históricos para generar directamente pronósticos de pedidos y contactos por pedido. Él está interesado en determinar si éstos pronósticos son más exactos que los pronósticos que el departamento de finanzas derivó de las proyecciones de ingresos. En principio, Pat y su equipo están interesados en los patrones de datos de los pedidos históricos mensuales y de los contactos por pedido, así que decidieron graficar estas series de tiempo y analizar las autocorrelaciones. Los datos se presentan en la tabla 4-12 y están graficados en las figuras 4-19 y 4-20. Las autocorrelaciones se presentan en las figuras 4-21 y 4-22.

Pat está intrigado con las gráficas de las series de tiempo y las funciones de autocorrelación, y siente que un procedimiento de suavización podría ser la herramienta correcta para ajustar las series de tiempo a los pedidos y contactos por pedido, así como para la generación de pronósticos.

TABLA 4-12 Pedidos y contactos por pedido (CPP) del minorista en la Web, junio de 2001 a junio de 2003

<i>Mes</i>	<i>Pedidos</i>	<i>CPP</i>
01 jun.	3,155,413	0.178
01 jul.	3,074,723	0.184
01 ago.	3,283,838	0.146
01 sept.	2,772,971	0.144
01 oct.	3,354,889	0.144
01 nov.	4,475,792	0.152
01 dic.	5,944,348	0.152
02 ene.	3,742,334	0.174
02 feb.	3,681,370	0.123
02 mar.	3,546,647	0.121
02 abr.	3,324,321	0.117
02 mayo	3,318,181	0.116
02 jun.	3,181,115	0.129
02 jul.	3,022,091	0.131
02 ago.	3,408,870	0.137
02 sept.	3,501,779	0.140
02 oct.	3,712,424	0.144
02 nov.	4,852,090	0.129
02 dic.	7,584,065	0.124
03 ene.	4,622,233	0.136
03 feb.	3,965,540	0.116
03 mar.	3,899,108	0.111
03 abr.	3,670,589	0.108
03 mayo	3,809,110	0.101
03 jun.	4,159,358	0.105

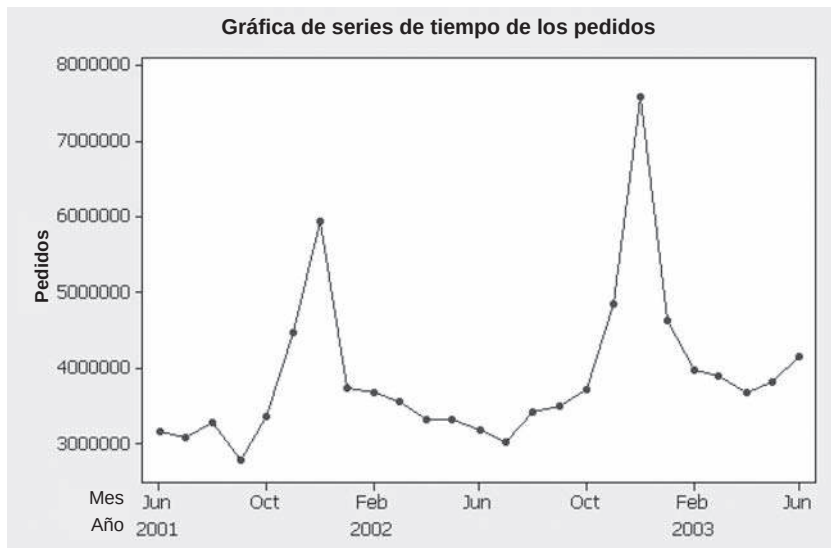


FIGURA 4-19 Gráfica de series de tiempo de pedidos, junio de 2001 a junio de 2003

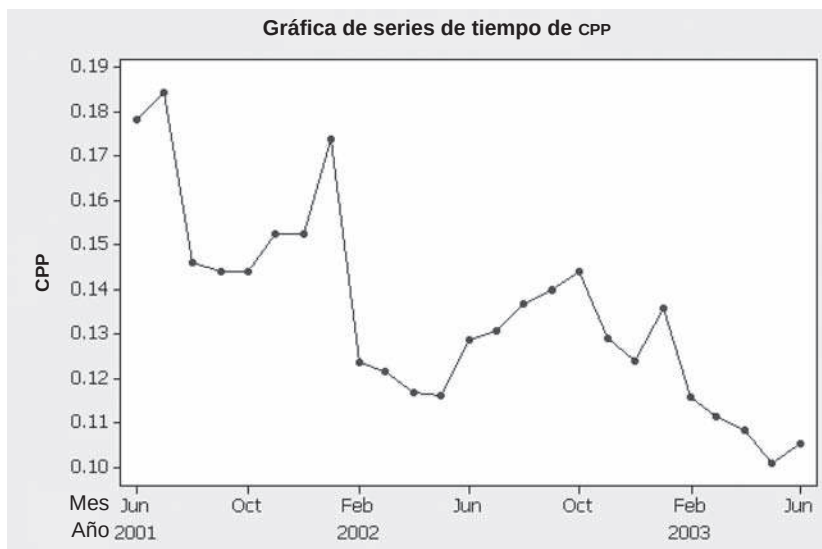


FIGURA 4-20 Gráfica de series de tiempo de contactos por pedido (CPP), junio de 2001 a junio de 2003

PREGUNTAS

1. ¿Qué aprendieron Pat y su equipo acerca de los patrones de datos de los pedidos y de los contactos por pedido en las gráficas de las series de tiempo y de las funciones de autocorrelación?
2. Ajuste un procedimiento de suavización apropiado para las series de tiempo de pedidos y genere pronósticos para los siguientes cuatro meses. Justifique su selección.

Función de autocorrelación para pedidos

Retraso	ACF	T	LBQ
1	0.449007	2.25	5.67
2	0.052358	0.22	5.75
3	-0.101155	-0.43	6.06
4	-0.196536	-0.82	7.31
5	-0.271190	-1.10	9.79
6	-0.187990	-0.73	11.04

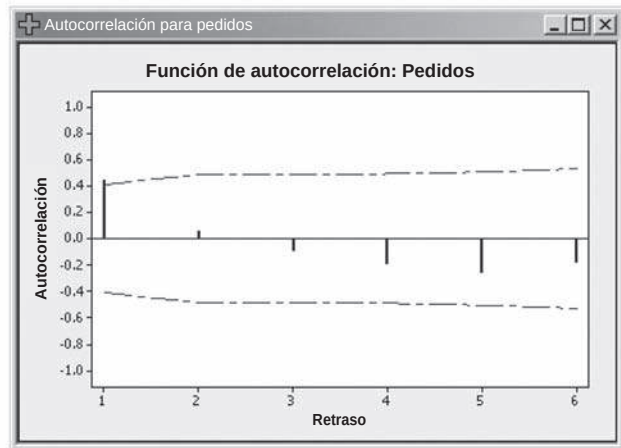


FIGURA 4-21 Función de autocorrelación para pedidos

Función de autocorrelación para CPP

Retraso	ACF	T	LBQ
1	0.637679	3.19	11.44
2	0.346901	1.29	14.97
3	0.195365	0.68	16.14
4	0.098064	0.34	16.45
5	0.084642	0.29	16.69
6	0.147248	0.50	17.46

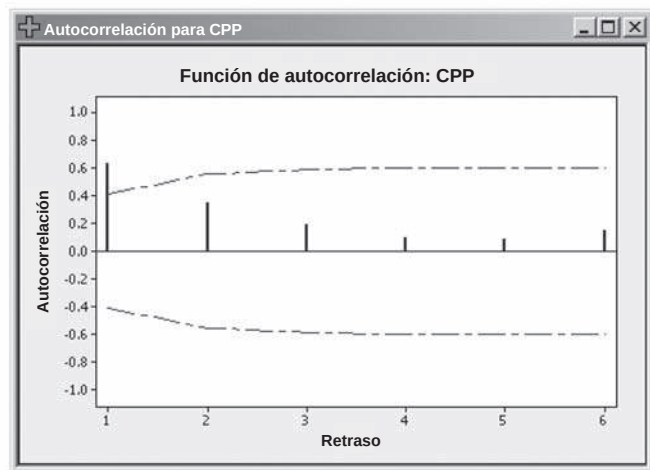


FIGURA 4-22 Función de autocorrelación para contactos por pedido (CPP)

- Ajuste un procedimiento de suavización apropiado para las series de tiempo de contactos por pedido y genere los pronósticos para los siguientes cuatro meses. Justifique su selección.
- Utilice los resultados de las preguntas 2 y 3 para generar pronósticos de contactos para los siguientes cuatro meses.
- Pat tiene acceso a una hoja de cálculo con la historia de contactos reales. Él está considerando elaborar el pronóstico de contactos directamente en vez de multiplicar los pronósticos de pedidos y contactos por pedido. ¿Esto parece razonable? ¿Por qué?
- Muchos pedidos contienen más de un artículo (unidad). ¿Sería mejor enfocarse en el número de unidades y contactos por unidad para obtener un pronóstico de contactos? Discútalos.

CASO 4-7 SOUTHWEST MEDICAL CENTER

Mary Beasley es responsable de dar seguimiento al número de visitas facturadas al grupo Medical Oncology de Southwest Medical Center. Su información personal sugiere que el número de visitas ha ido en aumento y que algunas épocas del año parecen estar más saturadas que otras. Algunos médicos están empezando a quejarse por la carga de trabajo y afirman que no siempre tienen tiempo suficiente para interactuar de forma individual con los pacientes. ¿Será necesario un aumento en la plantilla de médicos para manejar el aparente incremento en la demanda? Si es

así, ¿cuántos nuevos doctores habrá que contratar y/o reasignar a esta área?

Para comprender mejor la naturaleza de la demanda del servicio, Mary abre su hoja de cálculo de Excel y revisa el número total mensual de visitas facturadas de varios años fiscales anteriores. Los datos están listados en la tabla 4-13.

Una gráfica de series de tiempo con los datos de Mary se presenta en la figura 4-23. Como se esperaba, la serie de tiempo muestra una tendencia ascendente, pero Mary no está segura de si existe un componente

TABLA 4-13 Total de visitas facturadas por el Medical Oncology, año fiscal 1995 al año fiscal 2004

Año	Sept.	Oct.	Nov.	Dic.	Ene.	Feb.	Mar.	Abr.	Mayo	Jun.	Jul.	Ago.
FY1994-95	725	789	893	823	917	811	1,048	970	1,082	1,028	1,098	1,062
FY1995-96	899	1,022	895	828	1,011	868	991	970	934	784	1,028	956
FY1996-97	916	988	921	865	998	963	992	1,118	1,041	1,057	1,200	1,062
FY1997-98	1,061	1,049	829	1,029	1,120	1,084	1,307	1,458	1,295	1,412	1,553	1,480
FY1998-99	1,554	1,472	1,326	1,231	1,251	1,092	1,429	1,399	1,341	1,409	1,367	1,483
FY1999-00	1,492	1,650	1,454	1,373	1,466	1,477	1,466	1,182	1,208	1,132	1,094	1,061
FY2000-01	1,018	1,233	1,112	1,107	1,305	1,181	1,391	1,324	1,259	1,236	1,227	1,294
FY2001-02	1,083	1,404	1,329	1,107	1,313	1,156	1,184	1,404	1,310	1,200	1,396	1,373
FY2002-03	1,259	1,295	1,100	1,097	1,357	1,256	1,350	1,318	1,271	1,439	1,441	1,352
FY2003-04	1,339	1,351	1,197	1,333	1,339	1,307	—	—	—	—	—	—

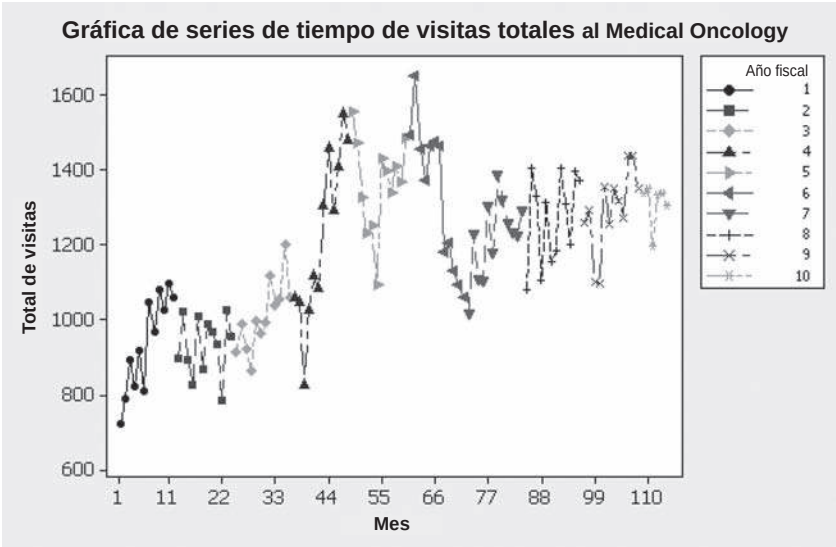


FIGURA 4-23 Gráfica de la serie de tiempo del total de visitas al Medical Oncology, año fiscal 1995 al año fiscal 2004

estacional en la serie de visitas totales. Ella decide investigar este asunto construyendo la función de autocorrelación. Si existe un componente estacional en los datos mensuales, Mary espera ver justamente grandes autocorrelaciones en los retrasos estacionales: 12, 24,

etcétera. Por un curso que tomó en su maestría en administración, Mary sabe que el procedimiento de suavización de Winters podría ser un buen camino para generar pronósticos de visitas futuras si existen componentes de tendencia y estacionales.

PREGUNTAS

1. ¿Qué revela el análisis de autocorrelación de Mary?
2. Ajuste un procedimiento de suavización apropiado para los datos de Mary, examine las autocorrelaciones residuales y genere pronósticos para el resto del año fiscal 2003-2004. ¿Estos pronósticos parecen razonables?
3. Considerando los resultados de la pregunta 2, ¿cree que es probable que otro método de elaboración del pronóstico genere “mejores” pronósticos? Discútalos.
4. ¿Usted cree que sería necesario el aumento de la plantilla médica para manejar la demanda futura? Escriba un breve reporte resumiendo el análisis de los datos de Mary y las implicaciones de la plantilla adicional.

CASO 4-8 SURTIDO COOKIES

En el caso 3-5, Jaime Luna investigó el patrón de datos de las ventas mensuales de Galletas Surtido (véase la tabla 3-12). En ese caso, Karin, uno de los miembros del equipo de Jaime, sugirió que los pronósticos de las ventas futuras para un mes determinado se podría generar usando sencillamente el promedio de ventas históricas de ese mes. Sin embargo, después de aprender algo acerca de los métodos de suavización, Jaime piensa que un procedimiento de suavización podría ser un mejor modo para elaborar pronósticos de las ventas

futuras. Jaime reconoce que los primeros pasos importantes para seleccionar un método de pronóstico son graficar las series de tiempo de las ventas y la realización de un análisis de autocorrelación. Él sabe que a menudo uno puede aprender mucho con tan sólo revisar una gráfica de series de tiempo. Más aún, las autocorrelaciones tienden a reforzar el patrón observado en la gráfica. Jaime está listo para iniciar con el objetivo de la generación de pronósticos de ventas mensuales de galletas para los meses restantes de 2003.

PREGUNTAS

1. ¿Qué patrón(es) observó Jaime en la gráfica de series de tiempo de las ventas de Surtido Cookies?
2. ¿Las autocorrelaciones son congruentes con el patrón (o los patrones) que Jaime observó en la gráfica de series de tiempo?
3. Seleccione y justifique un procedimiento de suavización apropiado para pronosticar las ventas futuras de galletas y elabore pronósticos para los meses restantes de 2003.
4. Use la sugerencia del promedio mensual histórico de Karin para elaborar pronósticos para los meses restantes de 2003. ¿Cuáles pronósticos prefiere, los de usted o los de Karin? ¿Por qué?

Aplicaciones de Minitab

El problema. En el ejemplo 4-3 se necesita pronosticar los datos de la Spokane Transit Authority usando un promedio móvil de cinco semanas.

Solución Minitab

1. Introduzca los datos de la Spokane Transit Authority mostrados en la tabla 4-2 (véase la p. 112), en la columna C1. Haga clic en los siguientes menús:

Stat>Time Series>Moving Average

2. Aparece la ventana de diálogo de Moving Average.
 - a) Haga doble clic en la variable Gallons y aparecerá a la derecha de Variable.
 - b) Puesto que queremos un promedio móvil de cinco meses, ponemos 5 para la longitud MA.
 - c) No haga clic en la ventana Center moving average. Usaremos un promedio móvil centrado para suavizar los datos en el capítulo 5.
 - d) Haga clic en Generate y ponga 1 a la derecha en número de pronósticos (Number of forecasts).
 - e) Haga clic en OK y aparecerá la figura 4-4.

El problema. En el ejemplo 4-6 se necesita pronosticar los datos de Acme Tool Company usando suavización exponencial simple.

Solución Minitab

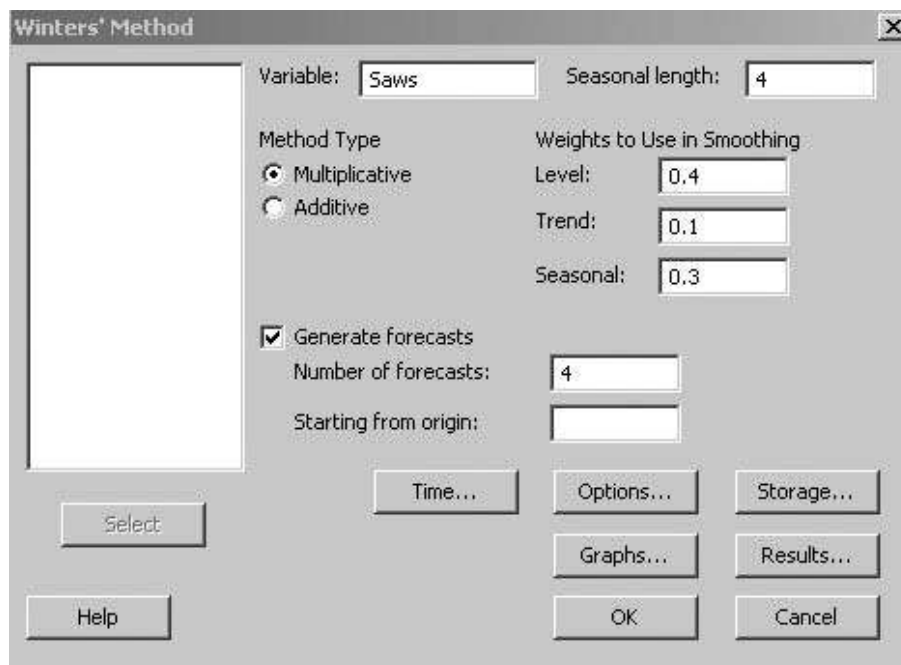
1. Introduzca los datos de Acme Tool Company presentados en la tabla 4-1 (véase la p. 109) de los años 2000 a 2006 en la columna C1. Haga clic en los siguientes menús:
Stat>Time Series>Single Exponential Smoothing
2. Aparece la ventana de diálogo de la suavización exponencial simple
 - a) Haga doble clic en la variable Saws y aparecerá a la derecha de Variable.
 - b) Debajo de Weight to Use in Smoothing, seleccione Optimal ARIMA y haga clic en OK. El resultado se presenta en la figura 4-8 (véase la p. 129).

El problema. En el ejemplo 4.10 se requiere pronosticar los datos de Acme Tool Company usando suavización exponencial ajustada para la tendencia y la estacionalidad.

Solución Minitab

1. Introduzca los datos de Acme Tool Company presentados en la tabla 4-1 (véase la p. 109) de los años 2000 a 2006 en la columna C1. Haga clic en los siguientes menús:
Stat>Time Series>Winters' Method

FIGURA 4-24 Ventana de diálogo de Minitab para el método de Winters



2. Aparece la ventana de diálogo del Winters Method como se presenta en la figura 4-24.
 - a) La variable de interés es Saws (serruchos).
 - b) Puesto que los datos son trimestrales, seleccione 4 para Seasonal (la estacionalidad).
 - c) Los Weights in Use in Smoothing son Level: 0.4; Trend: 0.1; y Seasonal: 0.3.
 - d) Haga clic en Generate y en Number of forecasts, seleccione 4.
 - e) Haga clic en Storage.
3. Aparece la ventana de diálogo Winter's Method Storage.
 - a) Haga clic en estimaciones Level, Trend y Seasonal, Fits (un periodo delante de los pronósticos) y Residuals.
 - b) Haga clic en OK tanto en la ventana de diálogo del Winter's Method Storage como en la ventana de diálogo Winters' Method. Los resultados se presentan en la tabla 4-9 y en la figura 4-12. El pronóstico para el primer trimestre de 2007 es 778.2.
4. Si desea guardar los datos para uso posterior, haga clic en los siguientes menús:
File>Save Worksheet As
5. Aparece la ventana de diálogo Save Worksheet As.
 - a) Teclee un nombre tal como Saws en el espacio de File Name.
 - b) El espacio Save as Type le permite seleccionar cómo quiere guardar su archivo. La mayoría de las veces usted seleccionará Minitab. Sin embargo, usted puede guardar su archivo de manera que puedan usarlo varios programas de software. Por ejemplo, usted podría guardarlo como un archivo de Excel. El archivo se guarda como Saws.xls y se utilizará en la sección de las aplicaciones de Excel.

Aplicaciones de Excel

El problema. En el ejemplo 4.5 los datos de Acme Tool Company se pronosticaron usando suavización exponencial simple con una constante de suavización igual a .6.

Solución de Excel

1. Abra el archivo que contiene los datos presentados en la tabla 4-1 (véase la p. 109) haciendo clic en los siguientes menús:

File>Open

Busque el archivo llamado Saws.xls.

2. Haga clic en los siguientes menús:

Tools>Data Analysis

Aparece la ventana de diálogo Data Analysis. Bajo Analysis Tools, seleccione Exponential Smoothing y haga clic en OK. Aparecerá la ventana de diálogo Exponential Smoothing presentada en la figura 4-25.

- a) Introduzca A2:A25 en la ventana de edición Input Range.
 - b) Revise la ventana Labels.
 - c) Introduzca .4 en el cuadro de edición del factor Damping, $(1 - \alpha)$ se definió como el complemento de la constante de suavización.
 - d) Introduzca B3 en la ventana de edición Output Range. (Esto pondrá al pronóstico $\hat{Y}_t$ opuesto al valor correspondiente en la columna A).
 - e) Revise la ventana Chart Output.
 - f) Ahora haga clic en OK.
4. Los resultados (columna B) y la gráfica se presentan en la figura 4-26. Advierta que la herramienta de análisis Exponential Smoothing pone fórmulas en la hoja de cálculo. La celda B5 aparece resaltada y la fórmula $= 0.6 \times A4 + 0.4 \times B4$ se presenta en la barra de herramientas de fórmula.

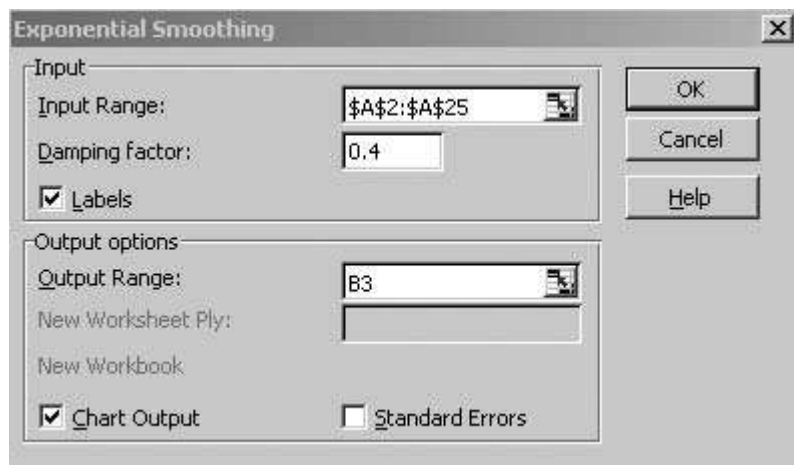


FIGURA 4-25 Ventana de diálogo de la Suavización exponencial de Excel

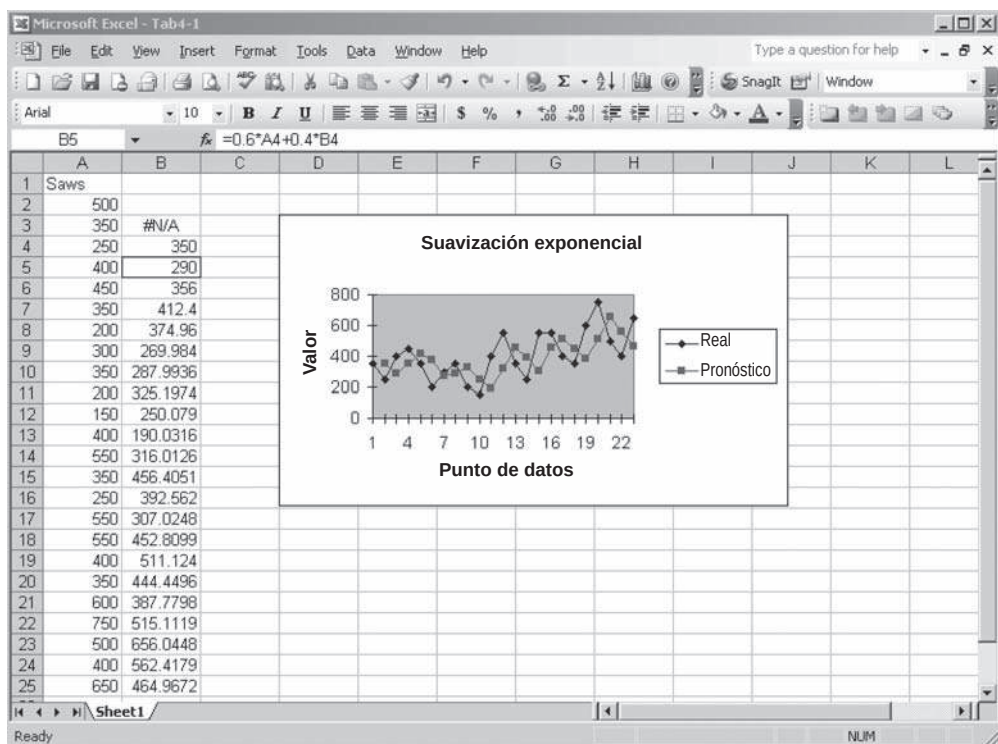


FIGURA 4-26 Suavización exponencial de Excel: Resultados del ejemplo 4.5

- Note que si bien Excel inicia la suavización exponencial de manera diferente al modo de inicio de la suavización en la tabla 4-7 (observe los primeros valores en la columna correspondiente a $\alpha = .6$), después de seis o siete iteraciones los números de la columna B

en la figura 4-26 y los números en la columna correspondientes a $\alpha = .6$ en la tabla 4-7 son exactamente los mismos. Las fluctuaciones debidas a las diferentes opciones de valores iniciales desaparecen rápidamente.

Referencias

- Aaker, D. A. y R. Jacobson. "The Sophistication of 'Naive' Modeling". *International Journal of Forecasting* 3(314)(1987): 449-452.
- Bowerman, B. L., R. T. O'Connell y A. B. Koehler. *Forecasting, Time Series and Regression*, 4a. ed. Belmont, CA: Thomson Brooks/Cole, 2005.
- Dalrymple, D. J. y B. E. King. "Selecting Parameters for Short-Term Forecasting Techniques". *Decision Sciences* 12(1981): 661-669.
- Gardner, E. S. Jr. "Exponential Smoothing: The State of the Art", *Journal of Forecasting* 4((1985): 1-28.
- 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 Seasonals and Trends by Exponentially Weighted Moving Averages". *International Journal of Forecasting* 20(2004): 5-10.
- Holt, C. C., F. Modigliani, J. F. Muth y H. A. Simon. *Planning Production Inventories and Work Force*. Englewood Cliffs, N.J.: Prentice Hall, 1960.
- Koehler, A. B., R. D. Snyder y D. K. Ord. "Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method". *International Journal of Forecasting* 17(2001): 269-286.
- Ledolter, J. y B. Abraham. "Some Comments on the Initialization of Exponential Smoothing", *Journal of Forecasting* 3(1)(1984): 79-84.
- Makridakis, S., S. C. Wheelwright y R. Hyndman. *Forecasting Method and Applications*. Nueva York: Wiley, 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*, 2a. ed. Cincinnati, Ohio: 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 hemos apuntado en capítulos anteriores, las observaciones de una variable Y que se recaban en el transcurso del tiempo se conocen como *datos de una serie de tiempo* o simplemente una *serie de tiempo*. Estas observaciones se registran a menudo en intervalos fijos de tiempo. Por ejemplo, Y podría representar ventas, y la serie de tiempo asociada podría ser una secuencia de cifras de ventas anuales. Otros ejemplos de series de tiempo incluyen ingresos trimestrales, niveles de inventario mensuales y tasas de tipo de cambio semanales. En general, las series de tiempo no se comportan como muestras aleatorias y requieren de métodos especiales para su análisis. Las observaciones de una serie de tiempo generalmente están relacionadas unas con otras (autocorrelacionadas). Esta dependencia genera patrones de variabilidad que pueden utilizarse para pronosticar valores futuros y ayudar en la administración de las operaciones de los negocios. Considere estas situaciones.

American Airlines (AA) compara las reservaciones actuales con pronósticos basados en proyecciones de patrones históricos. Si las reservaciones actuales están retrasadas o exceden las proyecciones, AA ajusta la proporción correspondiente a los asientos con descuento. Los ajustes se hacen para cada segmento de vuelo en el sistema de AA.

Un importador canadiense de flores compra sus productos a floricultores de Estados Unidos, México, Centroamérica y Sudamérica. Sin embargo, como estos proveedores adquieren parte de sus insumos y agroquímicos en Estados Unidos, todos los precios de venta se cotizan en dólares estadounidenses en el momento de la venta. Una factura no se paga inmediatamente, y puesto que la tasa de cambio entre el dólar canadiense y el estadounidense fluctúa, el costo para el importador en dólares canadienses se desconoce en el momento de la compra. Si la tasa de cambio no varía antes del pago de la factura, no existe riesgo monetario para el importador. Si la tasa se eleva, el importador pierde dinero por cada dólar de la compra. Si la tasa de cambio disminuye, el importador gana. El importador usa pronósticos semanales de la tasa de cambio entre el dólar canadiense y el dólar estadounidense para administrar el inventario de flores.

Si bien las series de tiempo a menudo se elaboran internamente y son únicas para cada organización, muchas series de interés en los negocios se pueden obtener de fuentes externas. Publicaciones como *Statistical Abstract of the United States*, *Survey of Current Business*, *Monthly Labor Review* y el *Federal Reserve Bulletin* contienen series de tiempo de todos tipos. Éstas y otras publicaciones dan a conocer datos de series de tiempo mensuales, trimestrales o anuales sobre precios, producción, ventas, empleo, desempleo, horas laboradas, combustible consumido, energía producida, ingresos, etcétera. En la actualidad están disponibles una gran cantidad de series de tiempo en la World Wide Web, en sitios correspondientes a oficinas gubernamentales, organizaciones de estadística, universidades e individuos.

Es importante que los administradores comprendan el pasado y usen datos históricos y su buen juicio para hacer planes inteligentes que satisfagan las demandas del futuro. Los pronósticos de series de tiempo elaborados adecuadamente ayudan a eliminar parte de la incertidumbre asociada con el futuro y puede ayudar a la dirección de una empresa en la definición de estrategias alternativas.

La alternativa, por supuesto, no es planear anticipadamente. Sin embargo, en un ambiente de negocios dinámico, la falta de planeación podría ser desastrosa. Un fabricante de grandes computadoras que hace algunos años hubiera ignorado la tendencia hacia las computadoras personales y estaciones de trabajo habría perdido una gran parte de su participación en el mercado bastante rápido.

Si bien centraremos nuestra atención en un enfoque basado en un modelo de análisis de series de tiempo que depende fundamentalmente de los datos, es muy importante una revisión subjetiva de las actividades de la elaboración de pronósticos. El pasado es relevante en la búsqueda de claves acerca del futuro sólo en la medida en que las condiciones causales previas continúen su efecto en el periodo siguiente. En las actividades económicas y de negocios, las condiciones causales pocas veces permanecen constantes. La multitud de factores causales en el trabajo tienden a cambiar constantemente, de manera que las relaciones entre el pasado, el presente y el futuro deben reevaluarse continuamente.

Las técnicas de las series de tiempo ofrecen un enfoque conceptual para la elaboración de pronósticos que ha probado ser muy útil. Los pronósticos se elaboran con la ayuda de un conjunto de procedimientos formales específicos y los juicios que se derivan de ello se exponen de manera explícita.

DESCOMPOSICIÓN

Un enfoque para el análisis de los datos de las series de tiempo incluye un intento por identificar los componentes que influyen en cada uno de los valores de una serie. Este procedimiento de identificación se llama *descomposición*. Cada componente se identifica por separado. Las proyecciones de cada uno de los componentes se pueden usar luego de manera combinada para elaborar pronósticos de valores futuros de la serie de tiempo. Los métodos de descomposición se usan para pronósticos tanto de corto como de largo plazos. También se emplean para exhibir de manera simple el crecimiento o la declinación subyacente de una serie, así como para ajustar la serie al eliminar uno o más de los componentes.

El análisis de una serie de tiempo a través de la descomposición de sus componentes tiene una larga historia. Sin embargo, recientemente, los métodos de descomposición para pronosticar han perdido algo de su prestigio. Con frecuencia, las proyecciones de los componentes individuales en el futuro y la recombinación de esas proyecciones para elaborar un pronóstico de la serie subyacente no funcionan muy bien en la práctica. La dificultad radica en la obtención de pronósticos exactos de los componentes. El desarrollo de procedimientos para la elaboración de pronósticos basados en un modelo más flexible (algunos de los cuales examinaremos en capítulos posteriores) ha hecho de la descomposición una herramienta fundamental para la comprensión de las series de tiempo y no un método de pronóstico en sí mismo.

Para comprender la descomposición, iniciaremos con los cuatro componentes de las series de tiempo que se presentaron en el capítulo 3. Éstos son el componente de la tendencia, el componente cíclico, el componente estacional y el componente irregular o aleatorio.

1. *Tendencia*. La tendencia es el componente que representa el crecimiento (o la declinación) subyacente en una serie de tiempo. La tendencia se produce, por ejemplo, como resultado de la inflación, de cambios demográficos persistentes o de cambios tecnológicos e incrementos en la productividad. El componente de la tendencia se denota como T .
2. *Componente cíclico*. Es una serie de fluctuaciones en forma de onda o ciclos de más de un año de duración. Las condiciones cambiantes de la economía generalmente producen ciclos. La letra C representa el componente cíclico.

En la práctica, los ciclos son difíciles de identificar, de manera que con frecuencia se les considera como parte de la tendencia. En este caso, el crecimiento (o la declinación) general subyacente del componente se llama *tendencia-ciclo* y se denota como T . Usamos la notación de la tendencia porque con frecuencia es difícil separar el componente cíclico de la tendencia.

3. *Componente estacional*. Los componentes estacionales se encuentran comúnmente 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 un año tras otro. Los

patrones estacionales ocurren por la influencia del tiempo climatológico o por acontecimientos relacionados con el calendario, como las vacaciones escolares y los días feriados nacionales. S representa al componente estacional.

4. *Componente irregular.* El componente irregular consiste en fluctuaciones impredecibles o aleatorias. Estas fluctuaciones son el resultado de incontables hechos que de forma individual tal vez no sean particularmente importantes, pero cuyos efectos combinados suelen ser considerables. I representa al componente irregular.

Para estudiar los componentes de una serie de tiempo, el analista debe considerar cómo se relacionan los componentes con la serie original. Esta tarea se lleva a cabo especificando 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 los valores de la serie de tiempo como una suma de los componentes se llama *modelo aditivo de los componentes*. Un modelo que trata los valores de las series de tiempo como el producto de los componentes se llama *modelo multiplicativo de los componentes*. Ambos modelos se denominan algunas veces *modelos de componentes no observados*, puesto que, en la práctica, si bien observamos los valores de la serie de tiempo, no sucede lo mismo con los valores de los componentes. El método de análisis de las series de tiempo descrito en este capítulo implica un intento por estimar los valores de los componentes, dada la serie observada. Estas estimaciones pueden utilizarse para pronosticar o desplegar la serie sin el inconveniente de las fluctuaciones estacionales. Este último proceso se llama *ajuste estacional*.

Es difícil tratar con el componente cíclico de una serie de tiempo. En la medida en que los ciclos puedan determinarse a partir de los datos históricos, tanto su duración (medida en años) como su magnitud (diferencias entre altas y bajas) están lejos de ser constantes. Esta falta de un patrón consistente de la onda hace que se dificulte la distinción entre los ciclos y las tendencias que evolucionan con suavidad. En consecuencia, para simplificar, supondremos que cualquier ciclo en los datos es parte de la tendencia. De manera que inicialmente consideraremos sólo tres componentes: T , S e I . Una explicación breve de un modo de manejar las fluctuaciones cíclicas en el método de descomposición para el análisis de las series de tiempo se incluye en la sección de variaciones cíclicas e irregulares de este capítulo (véase la p. 180).

Los dos modelos más sencillos que relacionan los valores observados (Y_t) de una serie de tiempo con los componentes de la tendencia (T_t), estacionales (S_t) e irregulares (I_t) son el modelo de componentes aditivo:

$$Y_t = T_t + S_t + I_t \quad (5.1)$$

y el modelo de componentes multiplicativo:

$$Y_t = T_t \times S_t \times I_t \quad (5.2)$$

El modelo de componentes aditivo funciona mejor cuando la serie de tiempo sometida a análisis tiene aproximadamente la misma variabilidad a lo largo de toda la serie. Es decir, en esencia, todos los valores de la serie caen dentro de una banda de anchura constante centrada en la tendencia.

El modelo de componentes multiplicativo funciona mejor cuando la variabilidad de la serie de tiempo aumenta con el nivel.¹ Es decir, los valores de la serie se dispersan conforme la tendencia aumenta, y el conjunto de observaciones tiene la apariencia de un megáfono o embudo. En la figura 5-1 se representan una serie de tiempo con variabilidad constante y una serie de

¹Es posible convertir una descomposición multiplicativa en una descomposición aditiva trabajando con los logaritmos de los datos. A partir de la ecuación 5.2 y de las propiedades de los logaritmos, tenemos que $\log Y = \log (T \times S \times I) = \log T + \log S + \log I$. La descomposición de datos logarítmicos se analiza en el problema 5.15.

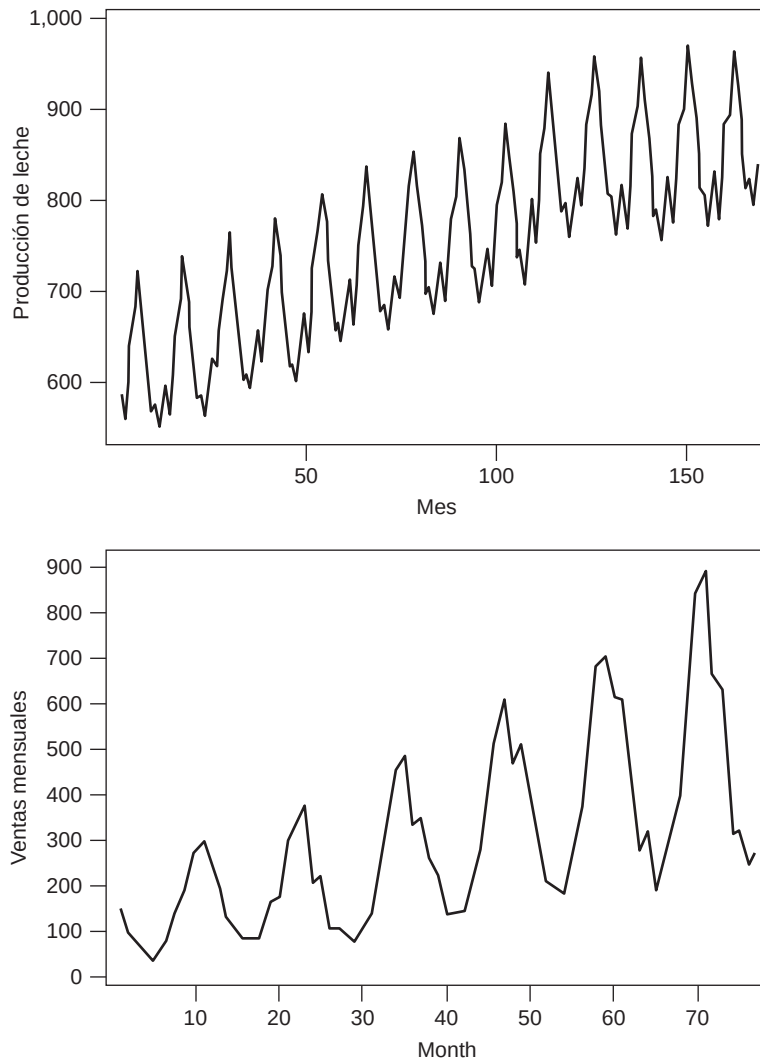


FIGURA 5-1 Serie de tiempo con variabilidad constante (arriba) y una serie de tiempo con variabilidad creciente con el nivel (abajo)

tiempo cuya variabilidad aumenta con el nivel. Estas dos series mensuales tienen una tendencia creciente y un patrón estacional claramente definido.²

Tendencia

Las tendencias son movimientos a largo plazo en una serie de tiempo que en ocasiones pueden describirse mediante una línea recta o una curva suave. Ejemplos de los factores básicos que producen o influyen en la tendencia de una serie de tiempo son un cambio en la población, un cambio de precio, un cambio tecnológico, un incremento en la productividad y los ciclos de vida de un producto.

²Existen variantes de los modelos de descomposición (véanse las ecuaciones 5.1 y 5.2) que contienen términos tanto multiplicativos como aditivos. Por ejemplo, algunos paquetes de software realizan la descomposición “multiplicativa” usando el modelo $Y = T \times S + I$.

Un incremento en la población podría causar que las ventas al menudeo de una comunidad aumenten anualmente durante varios años. Más aún, es posible que las ventas en dólares actuales se eleven durante el mismo periodo como resultado de incrementos generales en los precios de bienes al menudeo, aun cuando el volumen físico de los bienes vendidos no cambie.

En ocasiones los cambios tecnológicos provocan que una serie de tiempo se desplace hacia arriba o hacia abajo. El desarrollo de chips de alta velocidad para computadoras, el perfeccionamiento de los dispositivos de memoria y la mejora en las pantallas, acompañados de avances en la tecnología de telecomunicaciones, han dado como resultado incrementos drásticos en el uso de computadoras personales y teléfonos celulares. Por supuesto, los mismos desarrollos tecnológicos han traído consigo una tendencia descendente en la producción de calculadoras mecánicas y teléfonos convencionales.

Los incrementos en la productividad que podrían deberse a cambios tecnológicos le dan una pendiente ascendente a muchas series de tiempo. Cualquier medida de la producción total, como las ventas de los fabricantes, se ve afectada por cambios en la productividad.

Para las series de tiempo de negocios y económicas, lo mejor es ver la tendencia (o tendencia-ciclo) como un cambio uniforme en el tiempo. En raras ocasiones podemos suponer que la tendencia realmente queda representada mediante alguna función tan simple como una línea recta durante el todo el periodo en el cual observamos la serie de tiempo. Sin embargo, a menudo es conveniente ajustar una curva de tendencia a una serie de tiempo por dos razones: **1.** proporciona cierto indicio 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, si aumenta o disminuye como una línea recta—, entonces se representa por la ecuación:

$$\hat{T}_t = b_0 + b_1t \quad (5.3)$$

en donde $\hat{T}_t$ es el valor pronosticado de la tendencia para el tiempo t . El símbolo t representa el tiempo, la variable independiente, y ordinariamente toma los valores 1, 2, 3, ... correspondientes a periodos consecutivos. El coeficiente de la pendiente, b_1 , es el incremento o decremento promedio de T para cada incremento de un periodo en el tiempo.

Las ecuaciones de tendencia en el tiempo, incluyendo la tendencia en línea recta, se pueden ajustar a los datos usando el *método de mínimos cuadrados*. Recuerde del capítulo 2 que el método de mínimos cuadrados selecciona los valores de los coeficientes en la ecuación de tendencia (b_0 y b_1 en el caso de la línea recta), de manera que los valores estimados de la tendencia ($\hat{T}_t$) estén cerca de los valores reales (Y_t) conforme se miden con el criterio de la suma del cuadrado de los residuos.

$$SSE = \sum (Y_t - \hat{T}_t)^2 \quad (5.4)$$

Ejemplo 5.1

Los datos anuales de los registros de vehículos nuevos de pasajeros en Estados Unidos de 1960 a 1992 se presentan en la tabla 5-1 y están graficados en la figura 5-2. Los valores de 1960 a 1992 se emplean para desarrollar la ecuación de la tendencia. Los registros de vehículos son la variable dependiente, y la variable independiente es el tiempo t , codificado como 1960 = 1, 1961 = 2 y así sucesivamente.

La línea de tendencia ajustada tiene la ecuación:

$$\hat{T}_t = 7.988 + .0687t$$

La pendiente de la ecuación de tendencia indica que se estima que los registros se incrementan en un promedio de 68,700 cada año. La figura 5-3 presenta la línea recta de la tendencia ajustada de los datos reales. La figura 5-3 también muestra los pronósticos de registros de automóviles nuevos para los años 1993 y 1994 ($t = 34$ y $t = 35$), los cuales se obtuvieron extrapolarando la línea de la tendencia. Más adelante, diremos algo más acerca del pronóstico de la tendencia.

Los valores estimados de la tendencia para los registros de vehículos de pasajeros de 1960 a 1992 se presentan en la tabla 5-1, en la columna $\hat{T}$. Por ejemplo, la ecuación de la tendencia estima que los registros en 1992 ($t = 33$) fueron:

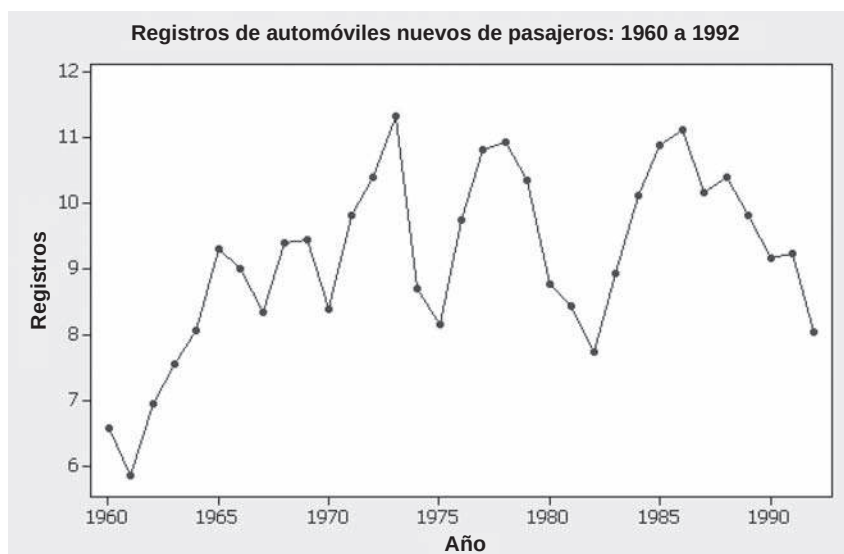
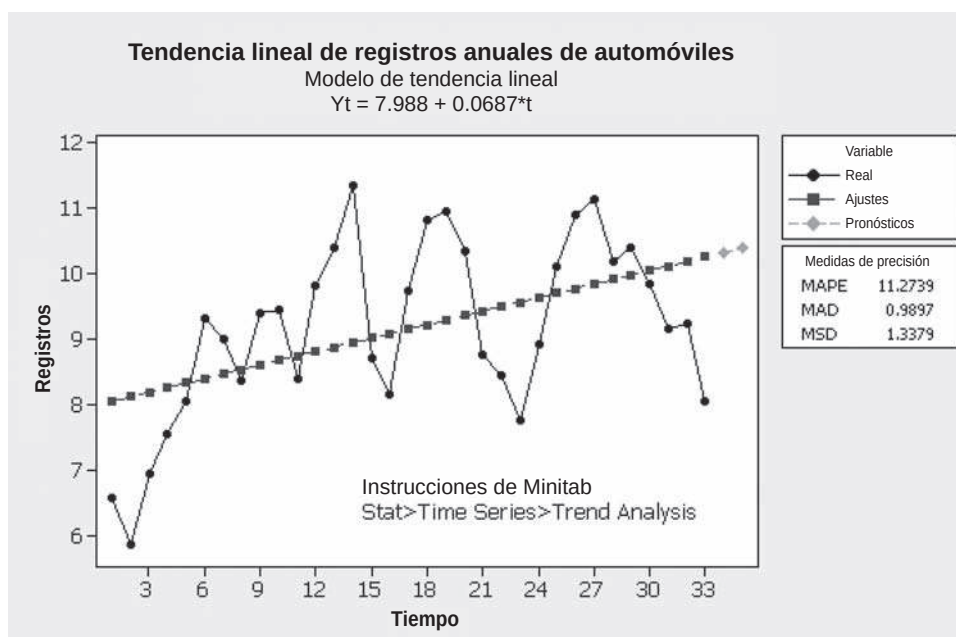
$$\hat{T}_{33} = 7.988 + .0687(33) = 10.255$$

TABLA 5-1 Registro de vehículos nuevos de pasajeros en Estados Unidos, 1960 a 1992, para el ejemplo 5.1

<i>Año</i>	<i>Registros (millones) Y</i>	<i>Tiempo t</i>	<i>Tendencia estimada (millones) $\hat{T}$</i>	<i>Error (millones) $Y - \hat{T}$</i>
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 Departamento de Comercio de Estados Unidos, *Survey of Current Business* (varios años).

es decir, 10,255,000 registros. Sin embargo, los registros de automóviles nuevos de pasajeros fueron en realidad 8,054,000 en 1992. De manera que, para 1992, la ecuación de la 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 usaron para calcular las medidas de ajuste, *MAD*, *MSD* y *MAPE*, presentadas en la figura 5-3. Éstas son las mismas medidas de precisión del pronóstico dadas por las ecuaciones 3.7, 3.8 y 3.10, respectivamente, en el capítulo 3. (Los comandos de Minitab que se usaron para generar los resultados en el ejemplo 5.1 se especifican en la sección de Aplicaciones de Minitab al final de este capítulo).

**FIGURA 5-2** Serie de tiempo de registros de automóviles para el ejemplo 5.1**FIGURA 5-3** Tendencia lineal de las series de tiempo de registros de automóviles para el ejemplo 5.1

Curvas de tendencia no lineal

El ciclo de vida de un producto nuevo tiene tres etapas: Introducción, crecimiento, y madurez y saturación. En la figura 5-4 se observa una curva que representa las ventas (en dólares o unidades) durante el ciclo de vida de un producto nuevo. El tiempo, representado sobre el eje horizontal, puede variar de días a años, dependiendo de la naturaleza del mercado. Con estos datos no funcionaría una línea recta para modelar la tendencia. Los modelos lineales suponen que una variable aumenta (o disminuye) por una cantidad constante en cada periodo de

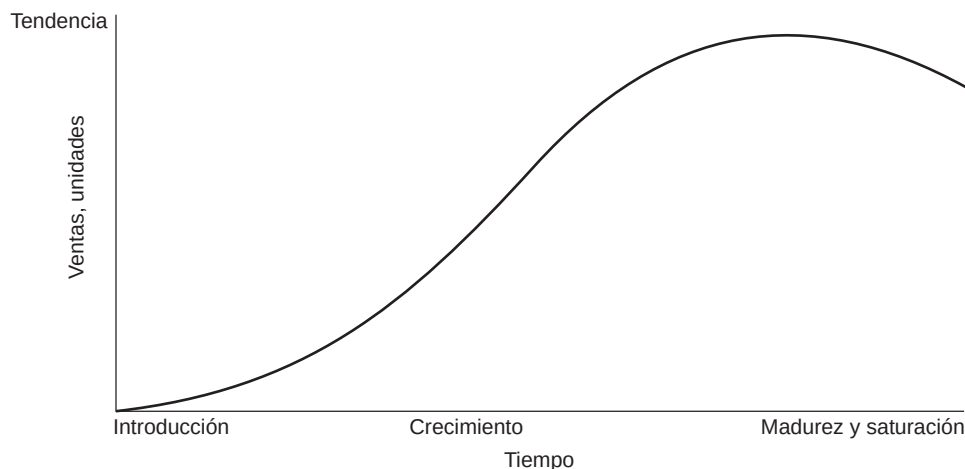


FIGURA 5-4 Ciclo de vida típico de un producto nuevo

tiempo. Los incrementos por periodo en la curva del ciclo de vida del producto son muy diferentes dependiendo de la etapa del ciclo. Se necesita una curva diferente de una línea recta para modelar la tendencia del ciclo de vida de un producto nuevo.

Una función simple que toma en cuenta la curvatura es la ecuación de la tendencia cuadrática:

$$\hat{T}_t = b_0 + b_1t + b_2t^2 \quad (5.5)$$

Como ilustración de lo anterior, la figura 5-5 presenta una curva de tendencia cuadrática ajustada a los datos de los registros de automóviles de pasajeros del ejemplo 5.1 usando el criterio *SSE*. Es posible proyectar la tendencia cuadrática más allá de los datos para, digamos, dos años adicionales, 1993 y 1994. Consideraremos las implicaciones de esta proyección en la siguiente sección, dedicada al pronóstico de la tendencia.

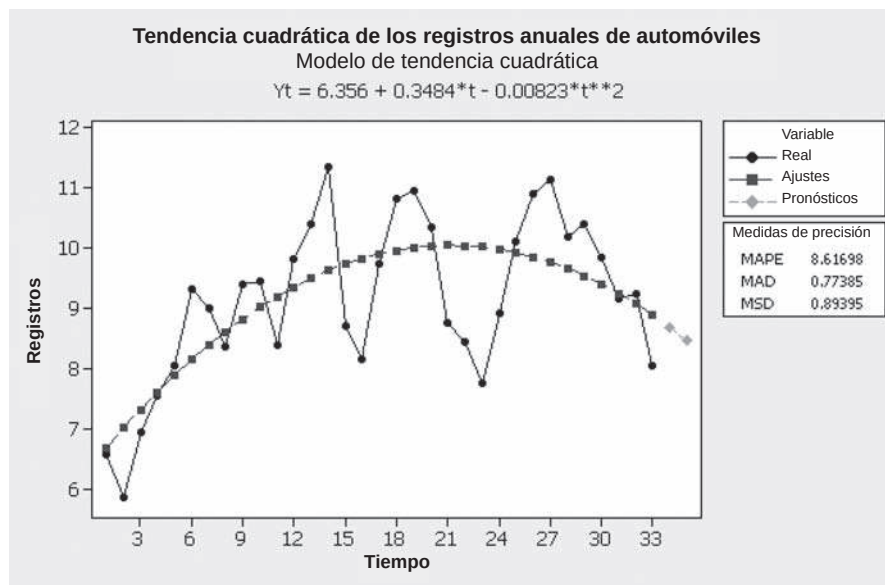


FIGURA 5-5 Curva de tendencia cuadrática para la serie de tiempo de los registros de automóviles para el ejemplo 5.1

Con base en las medidas de precisión *MAPE*, *MAD* y *MSD*, una tendencia cuadrática parece ser una mejor representación de la dirección general de las series de registros de vehículos que la tendencia lineal de la figura 5-3. ¿Cuál es el modelo de tendencia adecuado? Antes de considerar este asunto, presentaremos unas cuantas curvas adicionales de tendencia que han probado su utilidad.

Cuando una serie de tiempo inicia lentamente y luego parece que se incrementa a un ritmo creciente de manera (véase la figura 5-4) que la diferencia porcentual entre una observación y otra es constante, es posible ajustar una tendencia exponencial. La tendencia exponencial está dada por:

$$\hat{T}_t = b_0 b_1^t \quad (5.6)$$

El coeficiente b_1 está relacionado con la tasa de crecimiento. Si la tendencia exponencial se ajusta a los datos anuales, se estima que la tasa de crecimiento anual es $100(b_1 - 1)\%$.

La figura 5-6 indica el número de vendedores empleados por el fondo común de inversión de una compañía particular durante varios años consecutivos. El incremento en el número de vendedores no es constante. Parece como si los mayores incrementos en el número de vendedores se presentaron en los últimos años.

Una curva de tendencia exponencial que se ajusta a los datos de los vendedores tiene la ecuación

$$\hat{T}_t = 10.016(1.313)^t$$

lo que implica una tasa de crecimiento anual de alrededor del 31%. En consecuencia, si el modelo estima 51 vendedores en 2005, el aumento en 2006 sería $16(51 \times .31)$, para un estimado total de 67. Esto se puede comparar con el valor real de 68 vendedores.

Una tendencia lineal ajustada a los datos de los vendedores indicaría un incremento promedio constante de alrededor de nueve vendedores por año. Esta tendencia sobreestima el incremento real en los primeros años y subestima el crecimiento en el último año. No modela la tendencia aparente de los datos tan bien como lo hace la curva exponencial.

Es claro que la extrapolación de una tendencia exponencial con una tasa de crecimiento del 31% rápidamente producirá algunas cifras muy grandes. Éste es un problema potencial del modelo de tendencia exponencial. ¿Qué pasa cuando la economía entra en calma y los precios de las acciones comienzan a retraerse? La demanda de vendedores del fondo de inversión disminuirá, y el número de vendedores podría incluso declinar. El pronóstico de la tendencia mediante la curva exponencial será demasiado alto.

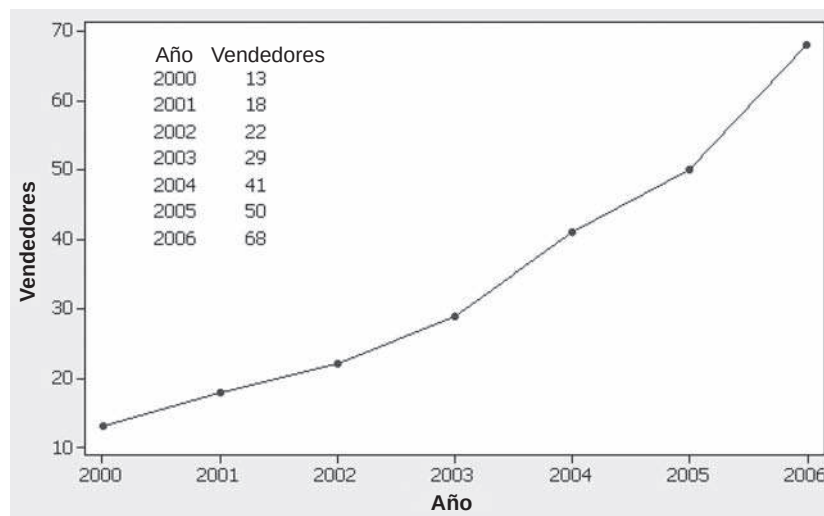


FIGURA 5-6 Gráfica de vendedores del fondo común de inversión

Las curvas de crecimiento del tipo logístico o de Gompertz representan la tendencia de muchas industrias y líneas de productos cuyo crecimiento registra una tasa de declinación conforme maduran. Si los datos graficados reflejan una situación en la cual las ventas inician bajas, luego se incrementan conforme el producto gana popularidad y finalmente disminuyen conforme se alcanza la saturación, la curva de Gompertz o el modelo logístico de Pearl-Reed podrían ser adecuados. La figura 5-7 muestra una comparación de las formas generales de a) la curva de Gompertz y b) el modelo logístico de Pearl Reed. Observe que la curva de logística es muy similar a la curva de Gompertz, aunque tiene una pendiente ligeramente más suave. La figura 5-7 indica cómo las intersecciones en Y y los valores máximos para estas curvas están relacionados con algunos de los coeficientes en sus formas funcionales. Las fórmulas para estas líneas de tendencia son complejas y no están dentro del alcance de este libro. Muchos paquetes estadísticos de software, incluyendo Minitab, permiten ajustar varios de los modelos de tendencias presentados en esta sección.

Si bien existen algunos criterios objetivos para seleccionar una tendencia apropiada, en general, la elección correcta es un asunto de buen juicio y, por lo tanto, requiere de experiencia y sentido común por parte del analista. Como veremos en la siguiente sección, la línea o curva que mejor se ajusta a un conjunto de puntos de datos podría carecer de sentido cuando se proyecta como una tendencia hacia el futuro.

Pronóstico de la tendencia

Ahora suponga que estamos en el momento $t = n$ (fin de la serie) y queremos usar un modelo de tendencia para pronosticar el valor de Y , p pasos adelante. El periodo en el cual hacemos el pronóstico, n en este caso, se llama el *origen del pronóstico*. El valor p se conoce como *tiempo principal*. Para el modelo de tendencia lineal, podemos generar un pronóstico evaluando $\hat{T}_{n+p} = b_0 + b_1(n + p)$.

Al emplear la línea de tendencia ajustada para los datos del registro de vehículos del ejemplo 5.1, un pronóstico de la tendencia para 1993 ($t = 34$) realizado 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 análoga, el pronóstico para dos pasos adelante $p = 2$ (1994) 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.

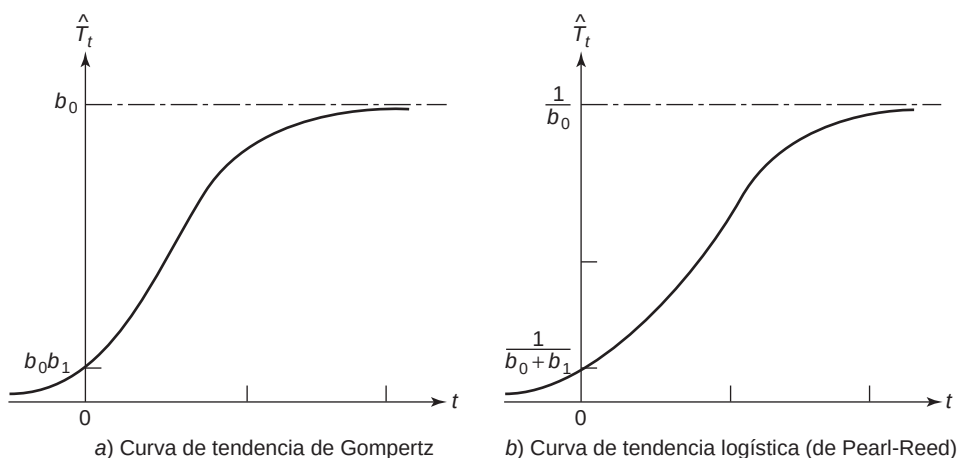


FIGURA 5-7 Curvas de crecimiento en forma de S

La figura 5-5 ilustra la curva de tendencia cuadrática ajustada a los datos del registro de automóviles. Al emplear la ecuación presentada en la figura, podemos calcular pronósticos de la tendencia para 1993 y 1994 estableciendo que $t = 33 + 1 = 34$ y $t = 33 + 2 = 35$. El lector podrá verificar que $\hat{T}_{33+1} = 8.688$ y $\hat{T}_{33+2} = 8.468$. Estos números se graficaron en la figura 5-5 como extrapolaciones de la curva de tendencia cuadrática.

Como los registros de vehículos se miden en millones, los dos pronósticos de tendencia obtenidos a partir de la curva cuadrática son muy diferentes de los pronósticos obtenidos a partir de la ecuación de tendencia lineal. Más aún, se mueven en dirección opuesta. Si tuviéramos que extrapolar las tendencias lineal y cuadrática para periodos de tiempo adicionales, sus diferencias serían más evidentes.

El ejemplo del registro de vehículos ilustra por qué debemos tener mucho cuidado al usar curvas de tendencia ajustadas para pronosticar tendencias futuras. Dos ecuaciones que representan razonablemente bien la serie de tiempo observada, tal vez den resultados muy diferentes cuando se proyectan hacia periodos futuros. Estas diferencias podrían ser sustanciales para tiempos principales largos (pronósticos a largo plazo).

Los modelos de la curva de tendencia se basan en las siguientes suposiciones:

1. Se ha seleccionado la curva de tendencia correcta.
2. La curva que se ajusta al pasado es indicativa del futuro.

Estas suposiciones sugieren que el buen juicio y la experiencia desempeñan un papel esencial en la selección y el uso de una curva de tendencia. Al usar una curva de tendencia para pronosticar, debemos estar en condiciones de argumentar que se ha seleccionado la tendencia correcta y que, con toda probabilidad, el futuro será como el pasado.

Existen criterios objetivos para seleccionar una curva de tendencia. En el capítulo 9 analizaremos dos de estos criterios: el criterio de información Akaike (AIC, por las siglas de *Akaike information criterion*) y el criterio de información Bayesiano (BIC, por las siglas de *Bayesian information criterion*). Sin embargo, aun cuando éstos y otros criterios ayudan a identificar un modelo adecuado, no reemplazan el buen juicio.

Estacionalidad

Un patrón estacional es aquel que se repite un año tras otro. Para datos anuales, la estacionalidad no tiene sentido porque no hay oportunidad de modelar un patrón con datos que se recaban una vez por año. Sin embargo, las series de tiempo que consisten en observaciones semanales, mensuales o trimestrales a menudo presentan 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 de los niveles medios y bajos. Los planes de marketing, por ejemplo, deben tomar en cuenta los patrones estacionales esperados en las compras de los consumidores.

Se han desarrollado varios métodos para medir la variación estacional. La idea básica en todos estos métodos es, primero, estimar y eliminar la tendencia de la serie original y, después, suavizar el componente irregular. Teniendo en mente nuestros modelos de descomposición, esto deja datos que contienen sólo variación estacional. Luego, se recolectan y se resumen los valores estacionales para obtener un número (generalmente un *número índice*) para cada intervalo del año observado (semana, mes, trimestre, etcétera).

Así, la identificación del componente estacional en una serie de tiempo difiere del análisis de tendencia por lo menos en dos formas:

1. La tendencia se determina directamente a partir de los datos originales, pero el componente estacional se determina indirectamente después de eliminar los otros componentes de los datos, de manera que sólo quede la estacionalidad.
2. La tendencia se representa mediante la curva, o ecuación, mejor ajustada, mientras que debe calcularse un valor estacional por separado para cada intervalo observado del año (semana, mes, trimestre) y a menudo tiene la forma de un número índice.

Si se emplea una descomposición aditiva, las estimaciones de la tendencia y los componentes estacionales e irregulares se suman para generar la serie original. Si se usa una descomposición multiplicativa se deben multiplicar los componentes individuales para reconstruir la serie original, y en esta formulación, el componente estacional se representa por una colección de números índices. Estos números indican qué periodos del año son relativamente bajos y qué periodos son relativamente altos. Los índices estacionales definen el patrón estacional.

Los *números índices* son porcentajes que indican los cambios en el tiempo.

Con los datos mensuales, por ejemplo, un índice estacional de 1.0 para un mes particular significa que el valor esperado para ese mes es de 1/12 del total del año. Un índice de 1.25 para un mes diferente implica que se espera que la observación para ese mes sea un 25% mayor que 1/12 del total anual. Un índice mensual de 0.80 indica que el nivel esperado de actividad para ese mes es 20% menor que 1/12 del total del año, y así sucesivamente. Los números índices indican las altas y bajas esperadas en los niveles de actividad durante el curso del año después de haber eliminado los efectos que se deben a la tendencia (o tendencia-ciclo) y los componentes irregulares.

Para resaltar la estacionalidad, primero debemos estimar y eliminar la tendencia. Esta última puede estimarse con una de las curvas de tendencia que explicamos antes, o bien, usando un promedio móvil, como vimos en el capítulo 4.

Al suponer un modelo de descomposición multiplicativo, la *proporción del promedio móvil* es un método que se emplea comúnmente para medir la variación estacional. En este método, la tendencia se estima usando un promedio móvil centrado. Ilustramos el método de la proporción del promedio móvil con las ventas mensuales de la Cavanaugh Company, que se muestran en la figura 5.1 (abajo) en el siguiente ejemplo.

Ejemplo 5.2

Para ilustrar el método de proporción del promedio móvil, utilizamos los datos de dos años de ventas mensuales de la Cavanaugh Company.³ La tabla 5-2 indica las ventas mensuales de enero de 2004 a diciembre de 2005 para ilustrar los cálculos desde el principio. Cuando tenemos datos mensuales, el primer paso es calcular un promedio móvil de 12 meses (para datos trimestrales se calcularía un promedio móvil de cuatro meses). Como todos los meses del año están incluidos en los cálculos de este promedio móvil, se eliminan los efectos que se deben al componente estacional, de manera que el promedio móvil sólo contenga la tendencia y los componentes irregulares.

Los pasos (identificados en la tabla 5-2) para calcular los índices estacionales por el método de la proporción del promedio móvil son los siguientes:

- Paso 1.** Partiendo del principio de la serie, calcule el total móvil de 12 meses y coloque este total para enero de 2004 hasta diciembre de 2004 entre junio y julio de 2004.
- Paso 2.** Calcule el total móvil de dos años, de manera que los promedios subsecuentes estén centrados en julio, en lugar de que esto ocurra entre un mes y otro.
- Paso 3.** Puesto que el total de dos años contiene los datos de 24 meses (enero de 2004 una vez, febrero de 2004 a diciembre de 2004 dos veces, y enero de 2005 una vez), este total se centra en julio de 2004.
- 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 dividiendo el valor real de julio entre el promedio móvil centrado de 12 meses.⁴

³Se omitieron las unidades, y las fechas y nombres se alteraron para proteger la identidad de la compañía.

⁴Ésta es la operación de proporción del promedio móvil que da su nombre al procedimiento.

TABLA 5-2 Ventas de la Cavanaugh Company, 2004 a 2005, para el ejemplo 5.2

<i>Periodo</i>	<i>Ventas</i>	<i>Total móvil de 12 meses</i>	<i>Total móvil de dos años</i>	<i>Promedio móvil centrado de 12 meses</i>	<i>Índice estacional</i>
2004					
Enero	518	1 4,869 } ²	9,833 } ³	409.7 } ⁴	0.60 } ⁵
Febrero	404				
Marzo	300				
Abril	210				
Mayo	196				
Junio	186				
Julio	247				
Agosto	343				
Septiembre	464				
Octubre	680				
Noviembre	711				
Diciembre	610				
2005					
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				

Repita los pasos 1 a 5 iniciando con el segundo mes de la serie, agosto de 2004, y así sucesivamente. El proceso termina cuando ya no sea posible calcular un total móvil completo de 12 meses.

Puesto que existen varias estimaciones (correspondientes a diferentes años) del índice estacional para cada mes, deben sumarse para obtener un solo número. Se usa la mediana, y no la media, como la medida de resumen. Al emplear la mediana se elimina la influencia de los datos que son inusualmente grandes o pequeños para un mes de un año en particular. Un resumen de las proporciones estacionales, junto con el valor de la mediana para cada mes, se presenta en la tabla 5-3.

Los índices estacionales mensuales para cada año deben sumar 12, de manera que la mediana de cada mes debe ajustarse para obtener el conjunto final de índices estacionales.⁵ Puesto que este multiplicador debe ser mayor que 1 si el total de razones de medianas antes del ajuste es menor que 12, y menor que 1 si el total es mayor que 12, el multiplicador se define como

$$\text{Multiplicador} = \frac{12}{\text{Total real}}$$

⁵Los índices mensuales deben sumar 12 de manera que el total anual esperado iguale al total real del año.

TABLA 5-3 Resumen de índices estacionales mensuales de Cavanaugh Company para el ejemplo 5-2

<i>Mes</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>Mediana</i>	<i>Índice estacional ajustado (Mediana × 1.0044)</i>
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

Usando la información de la tabla 5-3,

$$\text{Multiplicador} = \frac{12}{11.948} = 1.0044$$

La columna final de la tabla 5-3 contiene el índice estacional final de cada mes, calculado haciendo el ajuste (al multiplicar por 1.0044) para cada una de las proporciones de las medianas.⁶ Los índices estacionales finales presentados en la figura 5-8 representan el componente estacional en una descomposición multiplicativa de las ventas de la serie de tiempo de la Cavanaugh Company. La esta-

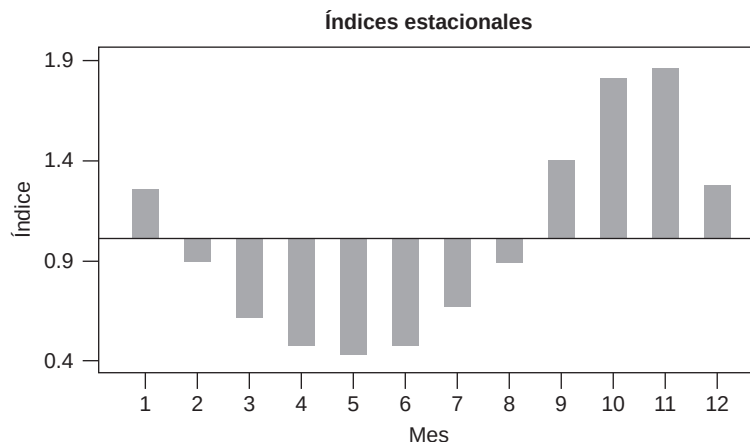


FIGURA 5-8 Índices estacionales de la Cavanaugh Company para el ejemplo 5.2

⁶En ocasiones los índices estacionales se multiplican por 100 y se expresan como porcentajes.

cionalidad en las ventas es evidente en la figura 5-8. Las ventas de esta compañía son periódicas, con ventas relativamente bajas al final de la primavera y relativamente altas al final del otoño.

Nuestro análisis de la serie de ventas del ejemplo 5.2 supuso que el patrón estacional permanecía constante de un año a otro. Si el patrón estacional parece modificarse con el tiempo, entonces la estimación del componente estacional con el conjunto completo de datos puede conducir a resultados engañosos. En este caso, es mejor: **1.** usar sólo datos recientes (de los últimos años) para estimar el componente estacional, o bien, **2.** usar un modelo de series de tiempo que permita analizar la evolución de la estacionalidad. En un capítulo posterior estudiaremos los modelos que permiten conocer la evolución de la estacionalidad.

El análisis estacional ilustrado en el ejemplo 5.2 es adecuado para un modelo de descomposición multiplicativo. Sin embargo, el enfoque general esquematizado en los pasos 1 a 5 funciona para una descomposición aditiva si, en el paso 5, la estacionalidad se calcula restando la tendencia de la serie original en vez de dividirla entre la tendencia (promedio móvil) para obtener un índice. En la descomposición aditiva, el componente estacional se expresa en las mismas unidades que la serie original.

Además, es evidente en nuestro ejemplo de ventas que el hecho de utilizar un promedio móvil centrado para determinar la tendencia da como resultado algunos valores faltantes en los extremos de la serie. Esto es particularmente problemático si el objetivo es la elaboración de un pronóstico. Para pronosticar valores futuros usando un método de descomposición, se deben usar métodos alternativos para la estimación de la tendencia.

Los resultados de un análisis estacional sirven para: **1.** eliminar la estacionalidad en los datos; **2.** pronosticar valores futuros; **3.** evaluar las posiciones actuales en ventas, inventarios y embarques, por ejemplo; y **4.** programar la producción.

Datos ajustados a la estacionalidad

Una vez que se ha aislado el componente estacional, se puede usar para calcular *datos ajustados a la estacionalidad*. Para una descomposición aditiva, los datos ajustados a la estacionalidad se calculan restando el componente estacional:

$$Y_t - S_t = T_t + I_t \quad (5.7a)$$

Para una descomposición multiplicativa, los datos ajustados a la estacionalidad se calculan dividiendo las observaciones originales entre el componente estacional:

$$\frac{Y_t}{S_t} = T_t \times I_t \quad (5.7b)$$

La mayoría de las series económicas que publican las oficinas gubernamentales están ajustadas estacionalmente porque la variación estacional no es de interés primordial. Más bien, lo que resulta de interés es el patrón de la actividad económica, independientemente de las fluctuaciones estacionales normales. Por ejemplo, los registros de vehículos nuevos podrían aumentar un 10% de mayo a junio, pero, ¿este aumento es un indicador de que las ventas de vehículos nuevos están completando un trimestre representativo? La respuesta es “no” si el 10% de aumento es común en ese momento del año fundamentalmente a causa de factores estacionales.

En una investigación relacionada con la obtención de datos ajustados estacionalmente, Bell y Hillmer (1984) encontraron que una amplia variedad de usuarios valoran los ajustes estacionales. Ellos identificaron tres motivos para hacer ajustes estacionales:

1. Los ajustes estacionales permiten la comparación confiable de valores en diferentes momentos.
2. Es más fácil comprender las relaciones entre las variables económicas o de negocios una vez que se ha eliminado de los datos el complicado factor de la estacionalidad.
3. El ajuste estacional es un elemento útil para elaborar pronósticos de corto plazo de los valores futuros de una serie de tiempo.

Bell y Hillmer concluyeron que “el ajuste estacional permite simplificar los datos de manera que los usuarios que no son expertos en estadística puedan interpretarlos más fácilmente, sin que esto implique una pérdida significativa de información” (p. 301).

Variaciones cíclicas e irregulares

Los ciclos son fluctuaciones de largo plazo en forma de onda, que se manifiestan sobre todo en los indicadores macroeconómicos de un país. Como hemos visto, hasta el grado en que se pueden medir, los ciclos no muestran un patrón consistente. Sin embargo, es posible lograr cierta comprensión del comportamiento cíclico de una serie de tiempo si se eliminan los componentes de tendencia y estacionalidad mediante la descomposición multiplicativa.⁷

$$\frac{Y_t}{T_t \times S_t} = \frac{T_t \times C_t \times S_t \times I_t}{T_t \times S_t} = C_t \times I_t \quad (5.8)$$

Se puede usar un promedio móvil para suavizar las irregularidades, I_t , dejando el componente cíclico C_t . Para eliminar el problema del centrado que se presenta cuando se usa un promedio móvil con un número par de periodos, las irregularidades se suavizan usando un promedio móvil con un número impar de periodos. Para datos mensuales, funcionará un promedio móvil de 5, 7, 9 o incluso 11 periodos. Para datos trimestrales, se puede calcular una estimación de C utilizando un promedio móvil de tres periodos de los valores.⁸

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 la serie de tiempo después de que se eliminaron los demás componentes. Algunas veces se le conoce como *residuo* o *error*. En una descomposición multiplicativa, tanto los componentes cíclicos como los irregulares se expresan en forma de índices.

Ejemplo de resumen

Una razón para la descomposición de una serie de tiempo es aislar y examinar los componentes de la serie. Después de que el analista está listo para revisar uno por uno los componentes de tendencia, estacionales, cíclicos e irregulares, podrá comprender los patrones en los valores de los datos originales. Además, una vez que se han aislado los componentes, es posible recombinarlos o sintetizarlos para elaborar pronósticos de valores futuros de las series de tiempo.

Ejemplo 5.3

En el ejemplo 3.5, Perkin Kendell, el analista de Coastal Marine Corporation, empleó el análisis de autocorrelación para determinar que las ventas eran estacionales sobre una base trimestral. Ahora utilizará la descomposición para entender la variable de las ventas trimestrales. Perkin emplea Minitab (véase la sección de aplicaciones de Minitab al final del capítulo) para elaborar la tabla 5-4 y la figura 5-9. Para conservar el patrón estacional actual, sólo se analizaron los datos de venta (Y) de los últimos siete años (2000 a 2006).

Los datos originales se presentan en la gráfica superior izquierda de la figura 5-10. La tendencia se calculó usando el modelo lineal: $\hat{T}_t = 261.24 + .759t$. Puesto que 1 representaba el primer trimestre de 2000, la tabla 5-4 muestra un valor de la tendencia igual a 262.000 para ese periodo, en tanto que las ventas estimadas (columna T) registraron un incremento de .759 cada trimestre.

⁷Advierta que hemos agregado el componente cíclico C a la descomposición multiplicativa de la ecuación 5.2.

⁸Para datos anuales, no existe componente estacional, y el componente cíclico $\times$ el componente irregular se obtiene eliminando simplemente la tendencia de la serie original.

TABLA 5-4 La descomposición multiplicativa de las ventas de Coastal Marine para el ejemplo 5.3

<i>t</i>	Año	Trimestre	Ventas	<i>T</i>	<i>SCI</i>	<i>S</i>	<i>TCI</i>	<i>CI</i>	<i>C</i>	<i>I</i>
1	2000	1	232.7	262.000	0.888	0.780	298.458	1.139	—	—
2		2	309.2	262.759	1.177	1.016	304.434	1.159	1.118	1.036
3		3	310.7	263.518	1.179	1.117	278.239	1.056	1.078	0.980
4		4	293.0	264.276	1.109	1.088	269.300	1.019	1.022	0.997
5	2001	1	205.1	265.035	0.774	0.780	263.059	0.993	0.960	1.034
6		2	234.4	265.794	0.882	1.016	230.787	0.868	0.940	0.924
7		3	285.4	266.553	1.071	1.117	255.582	0.959	0.906	1.059
8		4	258.7	267.312	0.968	1.088	237.775	0.890	0.924	0.962
9	2002	1	193.2	268.071	0.721	0.780	247.796	0.924	0.927	0.998
10		2	263.7	268.830	0.981	1.016	259.635	0.966	0.954	1.012
11		3	292.5	269.589	1.085	1.117	261.940	0.972	1.003	0.969
12		4	315.2	270.348	1.166	1.088	289.705	1.072	0.962	1.114
13	2003	1	178.3	271.107	0.658	0.780	228.686	0.844	0.970	0.870
14		2	274.5	271.866	1.010	1.016	270.269	0.994	0.936	1.062
15		3	295.4	272.625	1.084	1.117	264.537	0.970	0.976	0.994
16		4	286.4	273.383	1.048	1.088	263.234	0.963	0.942	1.022
17	2004	1	190.8	274.142	0.696	0.780	244.718	0.893	0.933	0.957
18		2	263.5	274.901	0.959	1.016	259.438	0.944	0.957	0.986
19		3	318.8	275.660	1.157	1.117	285.492	1.036	0.998	1.038
20		4	305.3	276.419	1.104	1.088	280.605	1.015	1.058	0.960
21	2005	1	242.6	277.178	0.875	0.780	311.156	1.123	1.089	1.031
22		2	318.8	277.937	1.147	1.016	313.886	1.129	1.104	1.023
23		3	329.6	278.696	1.183	1.117	295.164	1.059	1.100	0.963
24		4	338.2	279.455	1.210	1.088	310.844	1.112	1.078	1.032
25	2006	1	232.1	280.214	0.828	0.780	297.689	1.062	1.059	1.004
26		2	285.6	280.973	1.016	1.016	281.198	1.001	0.996	1.005
27		3	291.0	281.732	1.033	1.117	260.597	0.925	0.947	0.977
28		4	281.4	282.491	0.996	1.088	258.639	0.916	—	—

La gráfica superior derecha en la figura 5-10 presenta los datos sin la tendencia. Estos datos también se muestran en la columna *SCI* de la tabla 5-4. El valor sin la tendencia para el primer trimestre de 2000 fue⁹

$$SCI = \frac{Y}{T} = \frac{232.7}{262.000} = .888$$

Los datos ajustados a la estacionalidad se presentan en la columna *TCI* de la tabla 5-4 y en la figura 5-10. El valor ajustado a la estacionalidad del primer trimestre de 2000 fue

$$TCI = \frac{232.7}{.77967} = 298.458$$

Las ventas en el primer trimestre de 2005 fueron 242.6. Sin embargo, la revisión de la columna ajustada a la estacionalidad indica que las ventas para ese trimestre en realidad fueron altas cuando los datos se ajustaron al hecho de que el primer trimestre, comúnmente, es muy débil.

⁹Para simplificar la notación en este ejemplo, omitimos el subíndice *t* en los datos originales *Y*, así como en cada uno de sus componentes *T*, *S*, *C* e *I*. También omitimos el signo de multiplicación $\times$ entre los componentes, puesto que resulta claro que estamos considerando una descomposición multiplicativa.

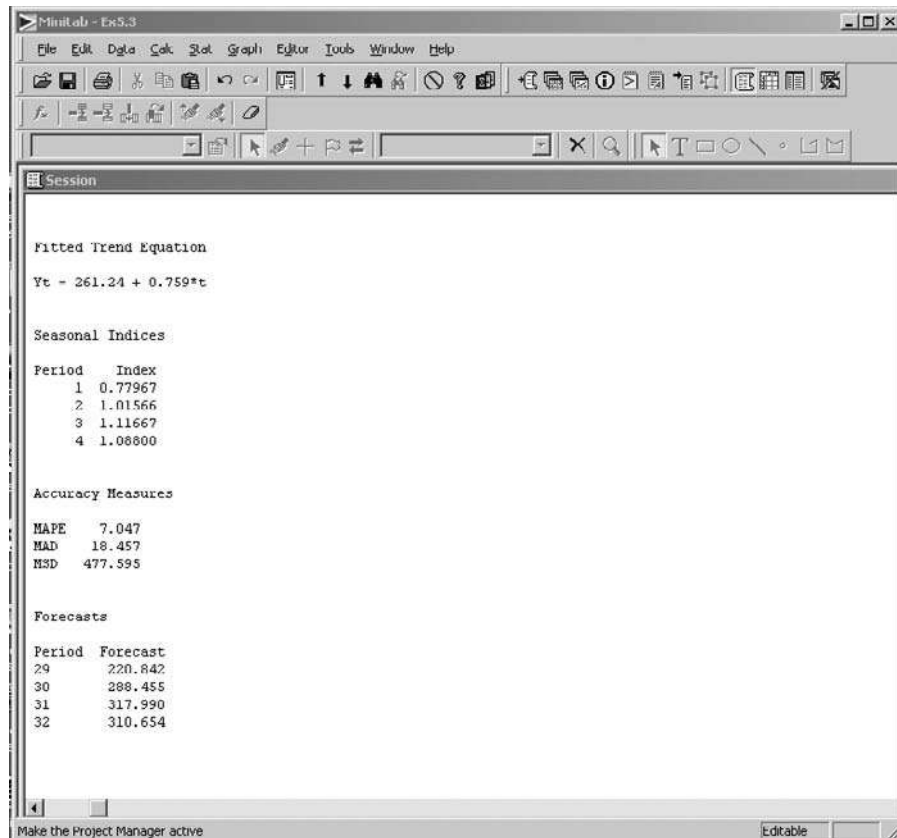


FIGURA 5-9 Resultado de Minitab para la descomposición de las ventas trimestrales de Coastal Marine para el ejemplo 5.3

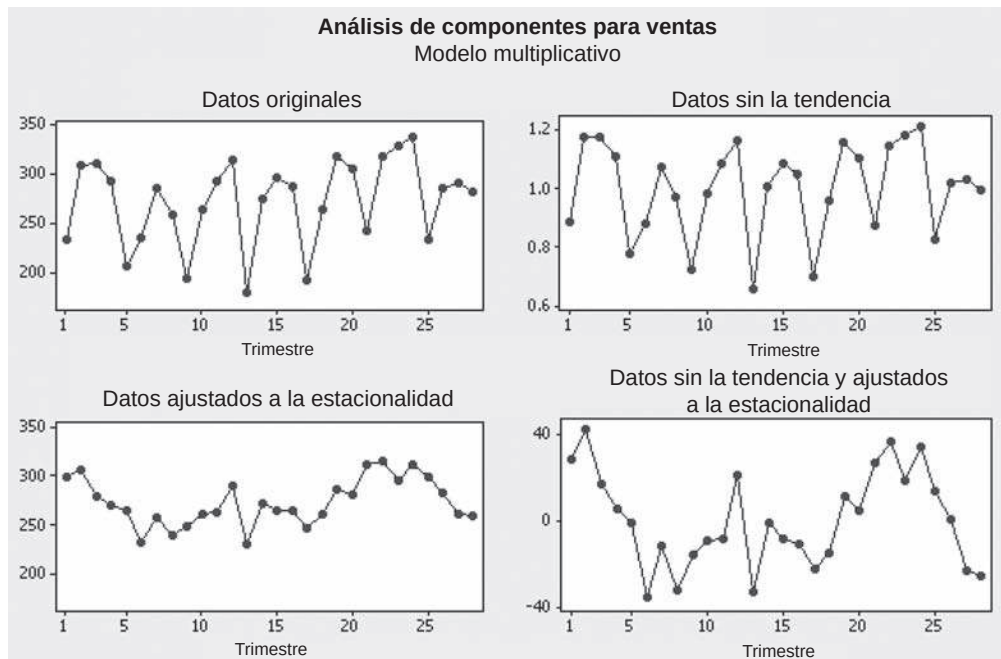


FIGURA 5-10 Análisis de componentes de las ventas de Coastal Marine para el ejemplo 5.3

Los índices estacionales en la figura 5-9 fueron

Primer trimestre = .77967 → 78.0 %
 Segundo trimestre = 1.01566 → 101.6%
 Tercer trimestre = 1.11667 → 111.7%
 Cuarto trimestre = 1.08800 → 108.8%

La gráfica de la izquierda en la figura 5-11 muestra los componentes estacionales en relación con 1.0. Vemos que las ventas del primer trimestre están un 22% por debajo del promedio, las ventas del segundo trimestre están cerca de lo esperado, las ventas del tercer trimestre están casi un 12% por arriba del promedio, mientras que las ventas del cuarto trimestre están casi un 9% por arriba de lo normal.

El valor cíclico-irregular del primer trimestre de 2000 fue¹⁰

$$CI = \frac{Y}{TS} = \frac{232.7}{(262.000)(.77967)} = 1.139$$

Para calcular la columna cíclica, se calculó un promedio móvil de tres periodos. El valor para el segundo trimestre de 2000 fue

$$\begin{array}{r} 1.139 \\ 1.159 \\ 1.056 \\ \hline 3.354 \end{array} \quad 3.354/3 = 1.118$$

Observe que la suavización de la columna *C* es comparada con la columna *CI*. La razón es que al usar el promedio móvil se han suavizado las irregularidades.

Finalmente, se calculó la columna *I*. Por ejemplo, para el segundo trimestre de 2000,

$$I = \frac{CI}{C} = \frac{1.159}{1.118} = 1.036$$

Al examinar la columna *I* se observa que existieron algunos cambios importantes en el componente irregular. El índice irregular cayó del 111.4% en el cuarto trimestre de 2002 al 87% en el primer

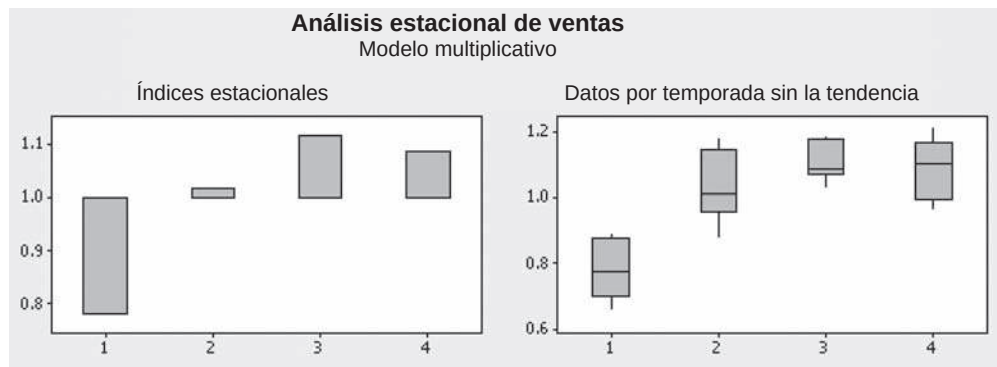


FIGURA 5-11 Análisis estacional de las ventas de Coastal Marine para el ejemplo 5.3

¹⁰Minitab calcula el componente cíclico × el componente irregular (o simplemente el componente irregular si no se considera el componente cíclico) restando de los datos originales el componente tendencia × el componente estacional. 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. Más aún, Minitab ajusta la línea de tendencia a los datos ajustados estacionalmente. Es decir, se realiza el ajuste estacional antes de determinar la tendencia.

trimestre de 2003, y luego aumentó al 106.2% en el segundo trimestre de 2003. Este comportamiento es resultado de las ventas inusualmente bajas del primer trimestre de 2003.

Indicadores de negocios

Los índices cíclicos nos ayudan a responder las siguientes preguntas:

1. ¿La serie es cíclica?
2. Si es así, ¿qué tan extremo es el ciclo?
3. ¿La serie sigue el estado general de la economía (ciclo de negocios)?

Una manera de investigar los patrones cíclicos es a través del estudio de los indicadores de negocios. Un *indicador de negocios* es una serie de tiempo relacionada con los negocios que se emplea como ayuda para evaluar el estado general de la economía, particularmente en lo que se refiere al ciclo de negocios. Muchos empresarios y economistas siguen sistemáticamente los movimientos de estas series estadísticas para obtener información económica y de negocios en forma de una imagen desplegada que se considera actualizada, detallada, relativamente objetiva, y la cual puede leerse y comprenderse en poco tiempo.

Los *indicadores de negocios* son series de tiempo relacionadas con los negocios, las cuales ayudan a evaluar el estado general de la economía.

La lista más importante de indicadores estadísticos se originó durante la fuerte recesión que sufrieron los negocios entre 1937 y 1938. En esa época, el Secretario del Tesoro, Henry Morgenthau, pidió a la Oficina Nacional de Investigación Económica (National Bureau of Economic Research, NBER) diseñar un sistema que indicara cuándo terminaría la recesión. Bajo el liderazgo de Wesley Michael y Arthur F. Burns, los economistas del NBER seleccionaron 21 series que, según su comportamiento en el pasado, prometían ser indicadores confiables de la reactivación económica. Desde entonces, la lista de indicadores de negocios se ha revisado varias veces. La lista actual contiene 21 indicadores: 10 clasificados como adelantados, 4 como coincidentes y 7 como retrasados.

1. *Indicadores adelantados.* En la práctica, los componentes de la serie adelantada se estudian para anticipar los momentos de cambio en la economía. El *Survey of Current Business* publica esta lista cada mes, junto con los valores reales de cada serie en varios meses pasados y el año más reciente. Además, se calcula un índice compuesto de indicadores adelantados para cada mes y año; el valor mensual más reciente se reporta a menudo en la prensa no especializada para indicar la dirección general del futuro de la economía. Ejemplos de indicadores adelantados son los permisos de construcción y los índices bursátiles.
2. *Indicadores coincidentes.* Los cuatro indicadores coincidentes proporcionan una medida de cómo está funcionando actualmente la economía de Estados Unidos. Un índice de estas cuatro series se calcula cada mes. Ejemplos de indicadores coincidentes son la producción industrial y las ventas de la industria y el comercio.
3. *Indicadores retrasados.* El nombre de estos indicadores se debe a que tienden a retrasarse con respecto a la situación general de la economía, tanto en tiempos de bonanza como de crisis. También se calcula un índice compuesto para esta lista. Ejemplos de indicadores retrasados son las tasas de interés preferencial y el índice de desempleo.

Los ciclos implican momentos de inflexión. Es decir, estos momentos indican el inicio de una caída o un crecimiento en el ciclo de negocios. Los indicadores adelantados cambian de dirección antes de que se presente un viraje en la actividad general de negocios, los indicadores coincidentes cambian al mismo tiempo que la economía general, y los indicadores retrasados experimentan un cambio después de que se manifiestan modificaciones en la

economía general. Sin embargo, es difícil identificar los puntos de inflexión cíclicos en el momento en que ocurren, puesto que no todas las áreas de la economía crecen al mismo tiempo durante periodos de expansión, ni tampoco se contraen de manera simultánea durante periodos de contracción. Por eso, tal vez pasen varios meses antes de que se identifique con certeza un verdadero ascenso o descenso cíclico.

Los indicadores adelantados son la herramienta más útil de predicción, puesto que intentan señalar cambios económicos de forma anticipada. En cambio, los indicadores coincidentes y los de comportamiento retrasado son de mínimo interés desde el punto de vista de los pronósticos, pero sirven para evaluar la efectividad de la política económica actual y pasada y así contribuir a formular futuras políticas.

En su artículo titulado “Early Warning Signals for the Economy”, Geoffrey H. Moore y Julius Shiskin (1976) dicen lo siguiente acerca de la utilidad de los indicadores del ciclo de negocios:

A partir de los registros, parece claro que los indicadores del ciclo empresarial son útiles en la interpretación de la tónica actual de los negocios y de las perspectivas a corto plazo. Pero a causa de sus limitaciones, los indicadores deben usarse junto con otros datos y con plena conciencia de los antecedentes de los negocios, de las expectativas y la confianza del consumidor, de las políticas gubernamentales, así como de los sucesos internacionales. También debemos advertir que los indicadores son a menudo difíciles de interpretar, que los analistas difieren en sus interpretaciones y que las señales que envían los indicadores tal vez no se interpreten correctamente. Los indicadores ofrecen una imagen sensible y reveladora del flujo y reflujo de las mareas económicas, la cual servirá a un analista hábil del escenario económico, político e internacional para mejorar sus probabilidades de hacer un pronóstico válido de las tendencias económicas a corto plazo. Si el analista está consciente de las limitaciones de los indicadores y está al tanto del mundo que lo rodea, encontrará en estos datos referencias útiles para hacer un balance de la economía y sus necesidades. (p. 81)

Por lo general, los componentes cíclicos de las series de tiempo individuales no se ajustan con mucho rigor —y en ocasiones no se ajustan en absoluto— al ciclo de negocios tal como se identifica con los indicadores NBER. Sin embargo, si se calcula un componente cíclico para una serie de tiempo determinada, siempre debe graficarse contra el tiempo para obtener alguna información de la magnitud y duración de cualquier ciclo que pudiera existir. Además, la gráfica debe examinarse para saber si existe alguna relación con las altas y bajas de la actividad económica en general.

La discusión hasta ahora indica cómo se pueden separar los factores que crean la variación en una serie de tiempo para estudiarlos individualmente. *Análisis* es el proceso de separar las partes de una serie de tiempo; *síntesis* es el proceso de reunir de nuevo las partes. A continuación pondremos juntos otra vez los componentes de la serie de tiempo para hacer pronósticos.

PRONÓSTICO DE UNA SERIE DE TIEMPO ESTACIONAL

En la elaboración del pronóstico de una serie estacional, se invierte el proceso de descomposición. En vez de separar la serie en sus componentes individuales para su examen, recombina los componentes para elaborar los pronósticos de periodos futuros. Usaremos el modelo multiplicativo y los resultados del ejemplo 5.3 para desarrollar los pronósticos de las ventas de Coastal Marine Corporation.

Ejemplo 5.4

Es posible desarrollar pronósticos para los cuatro trimestres de 2007 de las ventas de Coastal Marine Corporation usando la tabla 5-4.

1. *Tendencia*. La ecuación de la tendencia trimestral es $\hat{T}_t = 261.24 + .759t$. El pronóstico de origen es el cuarto trimestre de 2006, o periodo $t = n = 28$.

Las ventas del primer trimestre de 2007 ocurren en el periodo $t = 28 + 1 = 29$. Esta notación indica que estamos pronosticando el periodo $p = 1$ a partir del final de la serie de tiempo. Al establecer $t = 29$, la proyección de la tendencia es

$$\hat{T}_{29} = 261.24 + .759(29) = 283.251$$

2. *Estacionalidad*. El índice estacional para el primer trimestre, .77967, se presenta en la figura 5-9.
3. *Componente cíclico*. La proyección cíclica debe determinarse a partir del patrón cíclico estimado (si existe) y de cualquier otra información generada por los indicadores de la economía general para 2007. La proyección del patrón cíclico para periodos futuros está llena de incertidumbre y, como indicamos antes, para efectos del pronóstico se supone generalmente que está incluida en la tendencia. En el desarrollo de este ejemplo, daremos al índice cíclico un valor de 1.0.
4. *Componente irregular*. Las fluctuaciones irregulares representan una variación aleatoria que no es explicable a través de otros componentes. Para efectos del pronóstico, el componente irregular se fija en el valor promedio de 1.0.¹¹

El pronóstico del primer trimestre de 2007 es

$$\hat{Y}_{29} = T_{29} \times S_{29} \times C_{29} \times I_{29} = (283.251)(.77967)(1.0)(1.0) = 220.842$$

Los pronósticos para el resto de 2007 son

Segundo trimestre = 288.455
 Tercer trimestre = 317.990
 Cuarto trimestre = 310.654

Los pronósticos ajustados por descomposición multiplicativa para las ventas de Coastal Marine Corporation —junto con los pronósticos de 2007— se presenta en la figura 5-12. En la figura podemos ver que el ajuste, construido a partir de los componentes de tendencia y estacional, representa los datos reales razonablemente bien. Sin embargo, el ajuste no es bueno para los dos últimos trimestres de 2006, los periodos 27 y 28. Los pronósticos de 2007 imitan el patrón del ajuste.

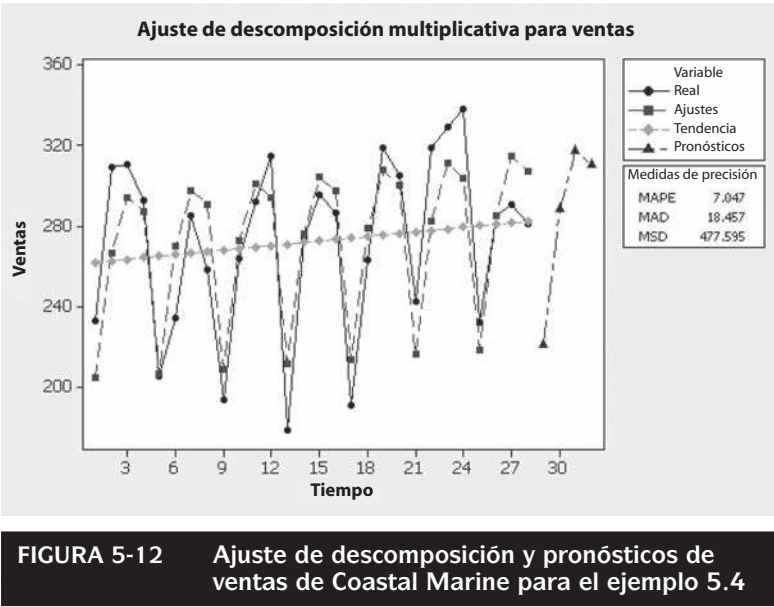


FIGURA 5-12 Ajuste de descomposición y pronósticos de ventas de Coastal Marine para el ejemplo 5.4

¹¹Para los pronósticos generados a partir de un modelo aditivo, el índice de irregularidad se fija en un valor promedio de 0.

Los pronósticos que se elaboran a partir un modelo aditivo o un modelo de descomposición multiplicativo reflejan la importancia de los componentes individuales. Si una variable es altamente estacional, entonces los pronósticos tendrán un fuerte patrón estacional. Si, además, existe una tendencia, los pronósticos seguirán un patrón estacional superpuesto a la tendencia extrapolada. Si un componente domina el análisis, por sí solo daría un pronóstico exacto y práctico de corto plazo.

MÉTODO DE DESCOMPOSICIÓN DE CENSO II

Los métodos de descomposición de las series de tiempo tienen una larga historia. En las décadas de 1920 y 1930, el Federal Reserve Board y la National Bureau of Economic Research estuvieron fuertemente implicados en el ajuste estacional y la suavización de las series de tiempo económicas. Sin embargo, antes del desarrollo de las computadoras, los cálculos de descomposición eran laboriosos y la aplicación práctica de los métodos era limitada. A principios de la década de 1950, Julius Shiskin, jefe de estadística económica de la Oficina de Censos de Estados Unidos (U.S. Bureau of the Census), desarrolló un programa de cómputo de gran escala para descomponer series de tiempo. El primer programa de cómputo era similar, en esencia, a los métodos manuales que se usaban hasta entonces; un año después, se sustituyó por un programa mejorado conocido como el Método II. Con el paso de los años, continuaron las mejoras al Método II. La versión actual del programa de descomposición de series de tiempo del Census Bureau se conoce como X-12-ARIMA. Este programa está disponible sin cargo alguno en el Census Bureau y se utiliza ampliamente en oficinas gubernamentales y compañías privadas.¹²

La descomposición del Censo II por lo general es multiplicativa, puesto que la mayoría de las series de tiempo de naturaleza económica tienen una variación estacional que se incrementa con el nivel de las series. La descomposición también supone tres componentes: tendencia-ciclo, estacional e irregular.

El método del Censo II repite una serie de pasos hasta que los componentes logran aislarse con éxito. Muchos de los pasos implican la aplicación de promedios móviles ponderados a los datos. Esto da como resultado una pérdida inevitable de datos al principio y al final de la serie a causa del promedio. La parte ARIMA del X-12-ARIMA tiene la capacidad de ampliar la serie original en ambas direcciones con pronósticos, de manera que la mayoría de las observaciones se ajusten usando promedios móviles totalmente ponderados. Estos pronósticos se generan mediante un modelo ARIMA de series de tiempo (véase el capítulo 9).

Los pasos de cada iteración del método Censo II, tal como se implementaron en el X-12-ARIMA, se describen a continuación. Quizá parezca que el método es complicado, ya que son muchos los pasos que deben realizarse. Sin embargo, la idea básica es muy simple: aislar uno por uno los componentes de tendencia-ciclo, estacional e irregular. Las numerosas iteraciones están diseñadas para mejorar la estimación de cada componente. Para un estudio más detallado, consulte las siguientes referencias: *Forecasting: Methods and Applications* (Makridakis, Wheelwright y Hyndman, 1998), *Forecasting: Practice and Process for Demand Management* (Levenbach y Cleary, 2006), y “New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program” (Findley *et al.*, 1998).

- Paso 1.** Se aplica un promedio móvil de s periodos a los datos originales para obtener un estimado aproximado del componente tendencia-ciclo. (Para datos mensuales, $s = 12$; para datos trimestrales, $s = 4$; etcétera).
- Paso 2.** Se calculan las razones entre los datos originales y estos valores de promedios móviles como una descomposición multiplicativa clásica, como se ilustra en el ejemplo 5.2.

¹²La versión para PC del programa X-12-ARIMA puede descargarse del sitio Web del U.S. Census Bureau. En el momento en que se escribió este libro, la dirección del sitio de la página para efectuar la descarga era www.census.gov/srd/www/x12a/x12down_pc.html.

- Paso 3.** Las razones del paso 2 contienen tanto el componente estacional como el irregular. También incluyen los valores extremos resultantes de eventos inusuales como huelgas o guerras. Las razones se dividen entre una estimación aproximada del componente estacional para obtener una estimación del componente irregular. Un valor demasiado grande del término irregular indica un valor extremo en los datos originales. Se identifican estos valores extremos y se ajustan las razones del paso 2 como corresponda. Esto elimina efectivamente valores que no se ajustan al patrón de los datos restantes. Los valores faltantes al principio y al final de las series también se reemplazan con las estimaciones de esta etapa.
- Paso 4.** Las razones obtenidas a partir de los datos modificados (con los valores extremos reemplazados y las estimaciones de los valores faltantes) se suavizan usando un promedio móvil para eliminar la variación irregular. Esto permite obtener una estimación preliminar del componente estacional.
- Paso 5.** Los datos originales se dividen entre el componente estacional preliminar del paso 4 para obtener la serie preliminar ajustada estacionalmente, la cual contiene el componente tendencia-ciclo y el componente irregular. En símbolos,

$$\frac{Y_t}{S_t} = \frac{T_t \times S_t \times I_t}{S_t} = T_t \times I_t$$

- Paso 6.** El componente tendencia-ciclo se estima aplicando un promedio móvil ponderado a la serie preliminar ajustada estacionalmente. Este promedio móvil elimina la variación irregular y genera una curva suave que indica el componente preliminar de tendencia-ciclo en los datos.
- Paso 7.** Se repite el paso 2 con esta nueva estimación de la tendencia-ciclo. Es decir, las nuevas razones que contienen sólo los componentes estacionales e irregulares se obtienen dividiendo las observaciones originales entre la tendencia-ciclo del paso 6. Éstas son las razones finales estacionales-irregulares. 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 usando las nuevas razones calculadas en el paso 7.
- Paso 9.** Se repite el paso 4 para obtener una nueva estimación 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 del paso 10 entre la tendencia-ciclo obtenida en el paso 6 para obtener el componente irregular estimado.
- Paso 12.** Los valores extremos del componente irregular se sustituyen como en el paso 3. Se obtiene una serie de datos modificados al multiplicar la tendencia-ciclo, el componente estacional y el componente irregular ajustado. Estos datos reproducen los datos originales, a excepción de los valores extremos.

Se repiten los 12 pasos anteriores, iniciando con los datos modificados del paso 12 en vez de los datos originales. La longitud de los promedios móviles usados en los diferentes pasos cambiará dependiendo de la variabilidad en los datos.

La serie final ajustada estacionalmente se determina dividiendo el componente estacional final entre los datos originales. El resultado contiene sólo el producto del componente tendencia-ciclo y el componente irregular.

Los valores de cada uno de los componentes finales se imprimen y se grafican. Existe un conjunto de pruebas de diagnóstico que permiten determinar si la descomposición fue correcta.

El programa X-12-ARIMA posee características adicionales que no hemos descrito. Por ejemplo, es posible hacer ajustes para considerar diferentes números de días de operación o los efectos de días no laborables; también es posible estimar y reemplazar los valores faltantes dentro de la serie, eliminar los efectos de los valores atípicos antes de la descomposición, o bien, 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 utilizada en el pronóstico de acontecimientos futuros entrelazados de alguna manera con la economía. Las compañías manufactureras están sumamente interesadas en los ciclos de prosperidad y declinación tanto de la economía nacional como de economías extranjeras, de manera que puedan predecir mejor la demanda de sus productos, la cual, a la vez, tendrá efectos sobre sus niveles de inventario, necesidades de personal, flujos de efectivo y casi todas las demás actividades dentro de la empresa.

La complejidad de estos problemas es enorme. Tomemos, por ejemplo, el problema de la predicción de la demanda de petróleo y sus derivados. A finales de la década de 1960, el precio del petróleo por barril era muy bajo y parecía haber una demanda mundial insaciable de gas y petróleo. Luego vino la sacudida de los precios del petróleo al principio y mediados de la década de 1970. ¿Cuál sería la demanda futura del petróleo? ¿Qué sucedería con los precios? Desde luego, empresas como Exxon y General Motors estaban muy interesadas en las respuestas a estas preguntas. Si los precios continuaban la escalada, ¿disminuiría la demanda de automóviles grandes? ¿Cuál sería la demanda de energía eléctrica? En general, los analistas predijeron que la demanda de energía y, por consiguiente, de petróleo sería muy inelástica, de manera que los precios continuarían superando a la inflación. Sin embargo, estas predicciones no tomaron en cuenta un profundo descenso en el ciclo de negocios a principios de la década de 1980 y una mayor elasticidad que la anunciada en la demanda de energía por parte de los consumidores. Alrededor de 1980, el mundo empezó a ver un exceso de petróleo en el mercado y los precios cayeron de forma drástica. Por ese tiempo, parecía difícil creer que los consumidores estuvieran beneficiándose realmente, una vez más, de la guerra de precios de la gasolina. Para cuando esta edición se escribía, disturbios importantes en el Medio Oriente generaron de nuevo una escasez de petróleo. El precio de un barril de petróleo y el costo de un galón de gasolina en Estados Unidos alcanzaron cifras muy altas y sin precedentes.

La demanda de petróleo no sólo se ve influida por acontecimientos cíclicos de largo plazo, sino además por sucesos estacionales y aleatorios; esto se refleja también en la mayoría de los pronósticos de demanda de cualquier tipo de producto o servicio. Por ejemplo, considere los sectores de servicios y de ventas al menudeo. Hemos sido testigos de un desplazamiento continuo del empleo del sector manufacturero hacia los sectores de servicios y de ventas al menudeo. Sin embargo, las ventas al menudeo (ya sea en tiendas, por catálogo o por Internet) son negocios extremadamente estacionales y cíclicos, por lo que las proyecciones de demanda y de inventario son esenciales para los minoristas, quienes cada vez utilizarán más el análisis de las series de tiempo.

Las compañías manufactureras tienen una necesidad continua de proyecciones estadísticas de acontecimientos futuros. Fuimos testigos del crecimiento explosivo en los campos de la tecnología y las telecomunicaciones durante la década de 1990 y de la contracción sustancial de estas industrias al principio del siglo XXI. Tal crecimiento y contracción fueron el resultado, en gran medida, de proyecciones de demanda que nunca se materializaron por completo. Las preguntas que todas las empresas manufactureras deben plantearse son: ¿Cuál será la tasa de inflación en el futuro? ¿Cómo se verán afectados los ajustes del costo de la vida que se incluyen en los contratos laborales de la compañía? ¿Qué repercusión tendrán estos ajustes en los precios y en la demanda? ¿Cuáles serán las habilidades gerenciales requeridas en 2025? ¿Cuál será el efecto del gasto gubernamental y las estrategias fiscales?

¿Cómo será la población de gente joven en el futuro? ¿Cuál será la mezcla étnica? Estos asuntos afectan a casi todos los segmentos de la economía. Los demógrafos están siguiendo muy de cerca las tasas de fertilidad actual y emplean casi todas las técnicas de pronósticos

disponibles de series de tiempo para tratar de hacer proyecciones de las variables demográficas. Errores de cálculo muy pequeños tendrán fuertes repercusiones en todos los ámbitos, desde la producción de juguetes para niños hasta la solidez financiera del sistema de seguridad social. Es interesante ver cómo los demógrafos están mirando los ciclos de negocios de largo plazo (ciclos de 20 años o más) al tratar de predecir qué hará esa generación de mujeres en edad fértil en lo que se refiere a la maternidad. ¿Tendrán uno o dos hijos como las familias de las décadas de 1960 y 1970, o regresarán a la concepción de dos o tres hijos, como lo hicieron las generaciones anteriores? Estas decisiones determinarán la composición de edad de la población para los próximos 50 o 75 años.

Los especialistas en ciencias políticas están interesados en usar el análisis de las series de tiempo para estudiar los patrones variables de los gastos gubernamentales en los programas de defensa y de asistencia social. Como es evidente, estas tendencias tienen gran repercusión en el futuro de todas las industrias.

Finalmente, un microcosmos interesante para la aplicación del análisis de las series de tiempo es el ámbito legal. Los abogados están usando cada vez más testigos expertos para testificar acerca del valor presente de los ingresos futuros de una persona o una empresa, el costo en que se incurre por la pérdida de un trabajo a causa de la discriminación, y el efecto sobre el mercado de una huelga ilegal. Estas preguntas a menudo se contestan mejor a través del uso acertado del análisis de las series de tiempo.

La tecnología satelital y la World Wide Web han hecho posible la acumulación y transmisión de información de manera casi instantánea. La proliferación de computadoras personales, la disponibilidad de programas estadísticos de cómputo fáciles de usar y el acceso creciente a las bases de datos han traído al escritorio el procesamiento de la información. La supervivencia de los negocios durante periodos de mayor cambio competitivo requiere la toma rápida de decisiones con base en el manejo de datos. El análisis de las series de tiempo y la elaboración de pronósticos desempeñan un papel importante en estos procesos de toma de decisiones.

APÉNDICE: ÍNDICE DE PRECIOS

Varias de las series de tiempo referentes a producción, ventas y otras situaciones económicas contienen datos disponibles sólo en dólares. Estos datos se ven afectados tanto por las cantidades físicas de bienes vendidos como por sus precios. La inflación y una gran variación de precios en el tiempo pueden causar problemas de análisis. Por ejemplo, un aumento en la cantidad en dólares puede esconder una disminución en las ventas de unidades cuando los precios están inflados. Así, a menudo es necesario saber cuánto del cambio del valor en dólares representa un cambio real en la cantidad física y cuánto se debe al cambio en el precio causado por la inflación. En estos casos es deseable expresar los valores en dólares constantes.

El concepto de *poder de compra* es importante. El poder actual de compra de \$1 se define como sigue:

$$\text{Actual poder de compra de \$1} = \frac{100}{\text{Índice de precios al consumidor}} \quad (5.10)$$

De esta forma, si en noviembre de 2006 el índice de precios al consumidor (considerando 2002 como 100) alcanza 150, el poder de compra actual del dólar del consumidor en noviembre de 2006 es

$$\text{Actual poder de compra de \$1} = \frac{100}{150} = .67$$

El dólar de 2006 compraba sólo dos tercios de los bienes y servicios que pudieron comprarse con un dólar del periodo base (2002).

Para expresar los valores de dólares en términos de *dólares constantes*, se usa la ecuación 5.11.

$$\text{Valor deflactado del dólar} = (\text{Valor del dólar}) = (\text{Poder de compra de \$1}) \quad (5.11)$$

Suponga que las ventas de automóviles se elevaron de \$300,000 en 2005 a \$350,000 en 2006, mientras que el índice de precios de automóviles nuevos (considerando 2002 como base) se elevó de 135 a 155. Las ventas deflactadas de 2005 y 2006 serían

$$\text{Ventas deflactadas de 2005} = (\$300,000) \left(\frac{100}{135} \right) = \$222,222$$

$$\text{Ventas deflactadas de 2006} = (\$350,000) \left(\frac{100}{155} \right) = \$225,806$$

Observe que las ventas en dólares reales tuvieron un aumento considerable de \$350,000 – \$300,000 = \$50,000. Sin embargo, las ventas deflactadas se incrementaron sólo en \$225,806 – \$222,222 = \$3,584.

El propósito de la deflación en los valores en dólares es eliminar el efecto de los cambios en el precio. Este ajuste se llama *deflación del precio* y se utiliza para expresar una serie en dólares constantes.

La *deflación del precio* es el proceso que permite expresar los valores de una serie en dólares 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 mercancías cuyos valores van a ser deflactados. Por ejemplo, las ventas de una zapatería deberían deflactarse mediante un índice de precios del calzado, no mediante un índice general de precios. Para valores monetarios deflactados que representan más de un tipo de mercancía, el analista debe desarrollar un índice de precios combinando los índices de precios pertinentes en una mezcla correcta.

Ejemplo 5.5

El señor Burnham desea estudiar el crecimiento a largo plazo de la Burnham Furniture Store. La tendencia a largo plazo se debe evaluar usando el volumen físico de las ventas. Si no es posible hacer esta evaluación, los cambios de precio reflejados en las ventas en dólares seguirán patrones inconsistentes y solamente ocultarán el patrón de crecimiento real. Si se usan las ventas en dólares, es necesario dividir las ventas en dólares reales entre un índice de precios adecuado para obtener las ventas medidas en dólares constantes.

El índice de precios al consumidor (IPC) no es conveniente para Burnham porque contiene elementos como renta, alimentos y servicios personales que la tienda no vende, pero algunos componentes de este índice sí son apropiados. Burnham está consciente de que el 70% de las ventas son de muebles y el 30% de electrodomésticos. Por lo tanto, puede multiplicar el componente de muebles al menudeo del IPC por .70, multiplicar el componente de electrodomésticos por .30 y luego sumar los resultados para obtener un índice de precios combinados. La tabla 5-5 ilustra este enfoque, en el cual los cálculos para 1999 son

$$90.1(.70) + 94.6(.30) = 91.45$$

Las ventas de 1999 se deflactaron en términos del poder de compra de 2002, de manera que

$$\text{Ventas deflactadas en 1999} = (42.1) \left(\frac{100}{91.45} \right) = 46.0$$

La tabla 5-5 indica que, si bien las ventas reales crecieron regularmente de 1999 a 2006, los volúmenes físicos permanecieron estables de 2004 a 2006. Como es evidente, los incrementos en las ventas se debieron a los aumentos de precios generados, a la vez, por la tendencia inflacionaria de la economía.

TABLA 5-5 Datos de las ventas de la Burnham Furniture Store, 1999 a 2006, para el ejemplo 5-5

<i>Año</i>	<i>Ventas de Burnham (en miles de \$)</i>	<i>Índice de precios de muebles al menudeo (2002=100)</i>	<i>Índice de precios de electrodomésticos al menudeo (2002=100)</i>	<i>Índice de precios^a (2002=100)</i>	<i>Ventas deflactadas^b (en miles de 2002)</i>
1999	42.1	90.1	94.6	91.45	46.0
2000	47.2	95.4	97.2	95.94	49.2
2001	48.4	97.2	98.4	97.56	49.6
2002	50.6	100.0	100.0	100.00	50.6
2003	55.2	104.5	101.1	103.48	53.3
2004	57.9	108.6	103.2	106.98	54.1
2005	59.8	112.4	104.3	109.97	54.4
2006	60.7	114.0	105.6	111.48	54.4

^aIntegrado al combinar precios de muebles (70%) y precios de electrodomésticos (30%).

^bVentas divididas entre el índice de precios por 100.

Glosario

Indicadores de negocios. Series de tiempo relacionadas con los negocios, las cuales ayudan a evaluar el estado general de la economía.

Números índice. Porcentajes que muestran cambios a través del tiempo.

Deflación de precios. Proceso que permite expresar los valores de una serie en dólares constantes.

Fórmulas clave

Descomposición aditiva de series de tiempo

$$Y_t = T_t + S_t + I_t \quad (5.1)$$

Descomposición multiplicativa de series de tiempo

$$Y_t = T_t \times S_t \times I_t \quad (5.2)$$

Tendencia lineal

$$\hat{T}_t = b_0 + b_1t \quad (5.3)$$

Criterio del cuadrado de los errores (con tendencia $\hat{T}$)

$$SSE = \sum (Y_t - \hat{T}_t)^2 \quad (5.4)$$

Tendencia cuadrática

$$\hat{T}_t = b_0 + b_1t + b_2t^2 \quad (5.5)$$

Tendencia exponencial

$$\hat{T}_t = b_0b_1^t \quad (5.6)$$

Datos ajustados estacionalmente (descomposición aditiva)

$$Y_t - S_t = T_t + I_t \quad (5.7a)$$

Datos ajustados estacionalmente (descomposición multiplicativa)

$$\frac{Y_t}{S_t} = T_t \times I_t \quad (5.7b)$$

Componente cíclico-irregular (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 de compra actual de \$1

$$\frac{100}{\text{Índice de precios al consumidor}} \quad (5.10)$$

Valor del dólar deflactado

$$(\text{valor del dólar}) \times (\text{poder de compra de \$1}) \quad (5.11)$$

Problemas

1. Explique el concepto de descomposición de una serie de tiempo.
2. Explique cuándo es más adecuada una descomposición multiplicativa que una descomposición aditiva.
3. ¿Cuáles son algunos factores básicos que influyen en el ciclo de tendencia de la mayoría de las variables?
4. ¿Qué clase de modelo de tendencia debería emplearse en cada uno de los siguientes casos?
 - a) La variable aumenta a una tasa constante.
 - b) La variable aumenta a una tasa constante hasta que alcanza la saturación y se nivela.
 - c) La variable aumenta en una cantidad constante.
5. ¿Cuáles son algunos factores básicos que influyen en el componente estacional de la mayoría de las variables?
6. Las estimaciones de crecimiento de ventas e ingresos de Value Line para compañías individuales se derivan de las correlaciones de ventas, ingresos y dividendos de los componentes apropiados de National Income Accounts, como inversiones en bienes de capital. Jason Black, un analista de Value Line, está revisando la tendencia de la variable de las inversiones en bienes de capital de 1977 a 1993. Los datos se presentan en la tabla P-6.
 - a) Grafique los datos y determine el modelo de tendencia adecuado para los años de 1977 a 1993.
 - b) Si el modelo adecuado es lineal, calcule el modelo de tendencia lineal para los años de 1977 a 1993.

TABLA P-6 Inversiones en bienes de capital (miles de millones de \$), 1977 a 1993

<i>Año</i>	<i>Miles de millones de \$</i>	<i>Año</i>	<i>Miles de millones de \$</i>	<i>Año</i>	<i>Miles de millones de \$</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 ^a

^aEstimado de Value Line

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1988, 1990, 1994).

- c) ¿Cuál es el incremento promedio anual en inversiones en bienes de capital desde 1977?
- d) Estime el valor de la tendencia para inversiones en bienes de capital en 1994.
- e) Compare su estimación de la tendencia con la de Value Line.
- f) ¿Qué factor(es) influye(n) en la tendencia de la inversión en bienes de capital?
7. Una compañía estadounidense grande está considerando hacer recortes en su publicidad en TV y, en vez de ello, entregar a sus clientes videos del negocio. Esta acción se está revisando después de que el presidente de la compañía leyó recientemente un artículo en el periódico que se refería a los videos para atraer clientes como “el arma para lograr ventas” en la actualidad. Algo que al presidente le gustaría investigar antes de emprender esta acción es la historia de la publicidad en TV en Estados Unidos, especialmente en relación con el ciclo de tendencia.
- La tabla P-7 indica el gasto total en dólares en publicidad televisiva en Estados Unidos (en millones de dólares).
- a) Grafique la serie de tiempo de los gastos en publicidad televisiva en Estados Unidos.
- b) Ajuste una tendencia lineal a los datos de publicidad y grafique la línea ajustada sobre la gráfica de la serie de tiempo.
- c) Pronostique los gastos en publicidad televisiva en dólares para 1998.
- d) De acuerdo con los resultados del inciso b), ¿cree que podría haber un componente cíclico en los gastos en publicidad televisiva en dólares? Explique.
8. Suponga que los siguientes índices estacionales específicos para marzo están dados como porcentajes y se obtuvieron 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 usando la mediana?

TABLA P-7

<i>Año</i>	<i>Y</i>	<i>Año</i>	<i>Y</i>
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*, varios años.

9. El valor esperado de la tendencia para octubre es de \$850. Suponiendo un índice estacional para octubre de 1.12 (112%) y el modelo multiplicativo dado por la ecuación 5.2, ¿cuál sería el pronóstico para octubre?
10. Los siguientes porcentajes específicos para los índices estacionales corresponden al mes de 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 cerca de Portland, Maine, ha llevado registro de sus ventas mensuales durante varios años, pero nunca ha analizado estos datos. El centro vacacional calcula los índices estacionales para sus ventas mensuales. ¿Cuáles de los siguientes enunciados acerca del índice son correctos?
- La suma de los 12 números índices mensuales, expresados como porcentajes, debería ser 1,200.
 - Un índice de 85 para mayo indica que las ventas son 15% más bajas que las ventas mensuales promedio.
 - Un índice de 130 para enero indica que las ventas están un 30% por arriba de las ventas mensuales promedio.
 - El índice para cualquier mes debe estar entre 0 y 200.
 - El índice porcentual promedio para cada uno de los 12 meses debe ser 100.
12. En la preparación de un reporte para June Bancock, gerente de Kula Department Store, usted incluye las cifras de ventas del último año (en miles de dólares) presentadas en la tabla P-12. Al verlas, el señor Bancock dice: “Este reporte confirma lo que le he estado diciendo: el negocio cada vez va mejor”. ¿Es correcta esta afirmación? ¿Por qué?
13. Los niveles de ventas trimestrales (medidos en millones de dólares) de Goodyear Tire se presentan en la tabla P-13. ¿Parece haber un efecto estacional significativo en estos niveles de ventas? Analice esta serie de tiempo para obtener los cuatro índices estacionales y determine la magnitud del componente estacional en las ventas de Goodyear.

TABLA P-12

<i>Mes</i>	<i>Ventas (en miles de \$)</i>	<i>Índice estacional ajustado (%)</i>
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: Con base en los registros de Kula Department Store.

TABLA P-13

Año	Trimestre			
	1	2	3	4
1985	2,292	2,450	2,363	2,477
1986	2,063	2,358	2,316	2,366
1987	2,268	2,533	2,479	2,625
1988	2,616	2,793	2,656	2,746
1989	2,643	2,811	2,679	2,736
1990	2,692	2,871	2,900	2,811
1991	2,497	2,792	2,838	2,780
1992	2,778	3,066	3,213	2,928
1993	2,874	3,000	2,913	2,916
1994	2,910	3,052	3,116	3,210
1995	3,243	3,351	3,305	3,267
1996	3,246	3,330	3,340 ^a	3,300 ^a

^aEstimados de Value LineFuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1988, 1989, 1993, 1994, 1996).

- a) Para pronosticar, ¿usaría el componente de la tendencia, el componente de la estacionalidad o ambos?
- b) Elabore un pronóstico para el tercero y el cuarto trimestres de 1996.
- c) Compare sus pronósticos con los de Value Line.
14. Las ventas mensuales de Cavanaugh Company, representadas en la figura 5.1 (abajo), se especifican en la tabla P-14.
- a) Realice una descomposición multiplicativa de la serie de tiempo de las ventas de Cavanaugh Company, suponiendo los componentes de tendencia, estacional e irregular.
- b) Para pronosticar, ¿usaría el componente de tendencia, el componente estacional o ambos?
- c) Obtenga pronósticos para el resto de 2006.
15. Construya una tabla similar a la tabla P-14 con los logaritmos naturales de las ventas mensuales. Por ejemplo, el valor para enero de 2000 es $\ln(154) = 5.037$.
- a) Haga una descomposición aditiva de $\ln(\text{ventas})$, considerando el modelo

$$Y = T + S + I.$$

TABLA P-14

Mes	2000	2001	2002	2003	2004	2005	2006
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	

- b) Para pronosticar, ¿usaría el componente de la tendencia, el componente estacional o ambos?
- c) Elabore pronósticos para $\ln(\text{ventas})$ para los meses restantes de 2006.
- d) Tome los antilogaritmos de los pronósticos calculados en el inciso c) para obtener pronósticos de las ventas reales para el resto de 2006.
- e) Compare los pronósticos del inciso d) con los del inciso c) del problema 14. ¿Cuál conjunto de pronósticos prefiere usted? ¿Por qué?
16. La tabla P-16 indica las ventas trimestrales (en millones de dólares) de Disney Company del primer trimestre de 1980 al tercer trimestre de 1995.
- a) Haga una descomposición multiplicativa de la serie de tiempo que integran las ventas trimestrales de Disney.
- b) ¿Parece haber una tendencia significativa? Discuta la naturaleza del componente estacional.
- c) ¿Usaría usted ambos componentes, el de tendencia y el estacional, para pronosticar?
- d) Pronostique las ventas para el cuarto trimestre de 1995 y los cuatro trimestres de 1996.
17. La demanda mensual de gasolina (en miles de barriles/día) de la Yukong Oil Company de Corea del Sur para el periodo que va de enero de 1986 a septiembre de 1996 se indica en la tabla P-17.
- a) Grafique la serie de tiempo de la demanda de gasolina. ¿Piensa usted que sería adecuada una descomposición aditiva o una multiplicativa para esta serie de tiempo? Explique.
- b) Realice un análisis de descomposición de 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 datos que representan las ventas mensuales (en miles de millones de dólares) de todas las tiendas minoristas en Estados Unidos. Con base en los datos de 1994, ejecute un análisis de descomposición de esta serie. Haga comentarios acerca de los tres componentes de la serie. Pronostique las ventas al menudeo para 1995 y compare sus resultados con los valores reales que aparecen en la tabla.

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	1,013.4	1,043.6
1989	1,037.9	1,167.6	1,345.1	1,288.2
1990	1,303.8	1,539.5	1,712.2	1,492.4
1991	1,439.0	1,511.6	1,739.4	1,936.6
1992	1,655.1	1,853.5	2,079.1	2,391.4
1993	2,026.5	1,936.8	2,174.5	2,727.3
1994	2,275.8	2,353.6	2,698.4	3,301.7
1995	2,922.8	2,764.0	3,123.6	

TABLA P-17

<i>Mes</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>	<i>1994</i>	<i>1995</i>	<i>1996</i>
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	

TABLA P-18

<i>Mes</i>	<i>1988</i>	<i>1989</i>	<i>1990</i>	<i>1991</i>	<i>1992</i>	<i>1993</i>	<i>1994</i>	<i>1995</i>
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
Diciembrer	171.0	176.5	179.7	181.0	204.1	218.3	233.3	238.0

Fuente: Con base en *Survey of Current Business*, 1989, 1993, 1996.

TABLA P-19

<i>Mes</i>	<i>Índice estacional ajustado</i>	<i>Mes</i>	<i>Índice estacional ajustado</i>
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: Con base en los registros del Mount Spokane Resort Hotel.

19. Los índices estacionales ajustados que aparecen en la tabla P-19 reflejan el volumen cambiante de los negocios del Mount Spokane Resort Hotel, el cual ofrece servicios a familias en el verano y a esquiadores entusiastas durante los meses de invierno. No se esperan variaciones cíclicas bruscas durante 2007.

TABLA P-24

		<i>Volumen de ventas</i>	<i>Índice de precios de productos primarios (2001 = 100)</i>
2005	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

- a) Si 600 turistas se alojaron en el hotel en enero de 2007, ¿cuál sería una estimación razonable para febrero?
- b) La ecuación de la tendencia mensual es $\hat{T} = 140 + 5t$ donde $t = 0$ representa el 15 de enero de 2001. ¿Cuál es el pronóstico para cada mes de 2007?
- c) ¿Cuál es el número promedio de nuevos turistas por mes?
20. Analice el desempeño del índice compuesto de los indicadores con comportamiento adelantado como un barómetro de la actividad empresarial en años recientes.
21. ¿Cuál es la posición actual del ciclo de negocios? ¿Se está expandiendo o contrayendo? ¿Cuándo se presentará el siguiente punto de inflexión?
22. ¿Cuál es el propósito de la deflación de una serie de tiempo que se mide en dólares?
23. En el periodo base de junio, el precio de una cantidad seleccionada de bienes era de \$1,289.73. En los meses más recientes, el índice de precios para estos bienes era de 284.7. ¿Cuánto sería el costo de los bienes seleccionados si se compraron en el mes más reciente?
24. Desarrolle una deflación de los volúmenes de ventas en dólares de la tabla P-24 usando el índice de precios de los productos primarios. Estos índices se aplican a todos los productos primarios, considerando que 2001 = 100.
25. La tabla P-25 contiene el número (en miles) de varones de 16 años de edad en adelante que fueron empleados en Estados Unidos para los meses de enero de 1993 a octubre de 2003.

TABLA P-25

<i>Año</i>	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>Mayo</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
1993	63,344	63,621	64,023	64,482	65,350	66,412	67,001	66,861	65,808	65,961	65,779	65,545
1994	64,434	64,564	64,936	65,492	66,340	67,230	67,649	67,717	66,997	67,424	67,313	67,292
1995	65,966	66,333	66,758	67,018	67,227	68,384	68,750	68,326	67,646	67,850	67,219	67,049
1996	66,006	66,481	66,961	67,415	68,258	69,298	69,819	69,533	68,614	69,099	68,565	68,434
1997	67,640	67,981	68,573	69,105	69,968	70,619	71,157	70,890	69,890	70,215	70,328	69,849
1998	68,932	69,197	69,506	70,348	70,856	71,618	72,049	71,537	70,866	71,219	71,256	70,930
1999	69,992	70,084	70,544	70,877	71,470	72,312	72,803	72,348	71,603	71,825	71,797	71,699
2000	71,862	72,177	72,501	73,006	73,236	74,267	74,420	74,352	73,391	73,616	73,497	73,338
2001	72,408	72,505	72,725	73,155	73,313	74,007	74,579	73,714	73,483	73,228	72,690	72,547
2002	71,285	71,792	71,956	72,483	73,230	73,747	74,210	73,870	73,596	73,513	72,718	72,437
2003	71,716	72,237	72,304	72,905	73,131	73,894	74,269	74,032	73,715	73,979		

Fuente: Con base en estadísticas de la fuerza laboral del Current Population Survey.

Use Minitab para hacer una descomposición multiplicativa de estos datos y genere los pronósticos para los siguientes 12 meses. ¿Parece adecuada una descomposición multiplicativa para este caso? Explique. ¿Existe un fuerte componente estacional en estos datos? ¿Le sorprende? ¿Los pronósticos parecen razonables?

26. Remítase al problema 25. La descomposición multiplicativa en Minitab supone de forma predeterminada una tendencia lineal. Grafique los datos de los varones empleados de la tabla P-25 y examine los años de 1993 a 2000 y de 2001 a 2003. ¿Se trata de una tendencia lineal adecuada? Si no es así, ¿puede usted sugerir una curva de tendencia que podría ser adecuada? Ajuste su curva de tendencia sugerida y guarde los residuos. Calcule la función de autocorrelación de los residuos. ¿Las autocorrelaciones residuales sugieren un componente estacional? Explique.
27. La tabla P-27 indica las ventas trimestrales (en millones de dólares) de las tiendas Wal-Mart de 1990 a 2004. Use Minitab para hacer una descomposición multiplicativa de la serie de tiempo de las ventas de Wal-Mart de los años 1990 a 2003 y genere pronósticos de los cuatro trimestres de 2004. ¿Es adecuada una descomposición multiplicativa para los datos de Wal-Mart? Explique. ¿Existe un fuerte componente estacional? ¿Le sorprende? Compare los pronósticos trimestrales de 2004 con las ventas reales. ¿Los resultados refuerzan la selección de una descomposición multiplicativa?
28. Remítase al problema 27. La descomposición multiplicativa en Minitab supone de manera predeterminada una tendencia lineal. Ajuste y grafique una línea de tendencia lineal para las ventas de Wal-Mart. ¿Es adecuada la tendencia lineal para estos datos? Si no es así, ¿podría sugerir una curva de tendencia que resulte adecuada? Ajuste su curva de tendencia sugerida y guarde los residuos. Calcule las autocorrelaciones de los residuos. ¿Las autocorrelaciones residuales sugieren un componente estacional? Explique.

TABLA P-27

Año	Trimestre			
	1	2	3	4
1990	6,768	7,544	7,931	10,359
1991	9,281	10,340	10,628	13,639
1992	11,649	13,028	13,684	17,122
1993	13,920	16,237	16,827	20,361
1994	17,686	19,942	20,418	24,448
1995	20,440	22,723	22,914	27,550
1996	22,772	25,587	25,644	30,856
1997	25,409	28,386	28,777	35,386
1998	29,819	33,521	33,509	40,785
1999	35,129	28,913	40,899	51,868
2000	43,447	46,588	46,181	57,079
2001	48,565	53,187	53,185	64,735
2002	52,126	56,781	55,765	66,905
2003	57,224	63,231	63,036	75,190
2004	65,443	70,466	69,261	82,819

Fuente: Base de datos de S&P Compustat North American Industrial Quarterly.

CASOS

CASO 5-1 THE SMALL ENGINE DOCTOR¹³

The Small Engine Doctor es el nombre de un negocio fundado por Thomas Brown, quien es un cartero del servicio postal de Estados Unidos. Él ha sido un reparador de todo desde su niñez, ya que, desde entonces, tiene la costumbre de examinar los aparatos domésticos inservibles para entender por qué se descompusieron. Conforme Tom creció y se convirtió en un habitante típico de los suburbios, adquirió numerosos artículos de equipo de jardinería. Cuando Tom averiguó acerca de un curso de reparación de máquinas pequeñas que se impartiría en una universidad local, no dejó pasar la oportunidad. Tom inició la reparación de pequeñas máquinas desarmando su propio equipo de jardinería para revisarlo y luego volverlo a ensamblar. Tan pronto como terminó su curso de reparación de máquinas, empezó a reparar podadoras de pasto, arados giratorios, sopladores de nieve y otros equipos de jardinería de amigos y vecinos. En este proceso adquirió varios manuales de equipo y herramientas especiales.

No pasó mucho tiempo para que Tom decidiera convertir su pasatiempo en un negocio de tiempo parcial. Publicó un anuncio en un boletín comercial suburbano bajo el nombre de Small Engine Doctor (Doctor de máquinas pequeñas). Durante los últimos dos años, el negocio ha crecido lo suficiente para complementar su sueldo regular. Si bien el crecimiento fue bienvenido, conforme el negocio está por entrar a su tercer año de operación, existen varias preocupaciones. Tom opera su negocio fuera de su casa. El sótano está dividido en una sala familiar, un taller y una oficina. Originalmente, el área de oficina era utilizada para manejar la publicidad, procesar pedidos y llevar la contabilidad. Todas las reparaciones de las máquinas se hacían en el taller. La política de Tom ha sido tener en inventario un limitado número de refacciones y solicitar el resto de éstas conforme se necesita. Éste parecía ser el único modo práctico de manejar la gran variedad de refacciones necesarias en la reparación de las máquinas fabricadas por una docena o más de empresas fabricantes de equipo para jardinería.

Las piezas de repuesto han sido el problema más grave durante el funcionamiento del negocio. Tom inició su negocio comprando piezas de repuesto a vendedores de equipo. Esta práctica tuvo varias desventajas. Primero, él tenía que comprar las piezas a precios de menudeo. Segundo, la mayoría de las veces el vendedor

tenía que reservar el pedido de una o más partes para cualquier trabajo de reparación. Las partes ordenadas al fabricante tenían tiempos de entrega de 30 a 120 días. Como consecuencia, Tom cambió su política y comenzó a ordenar las piezas de repuesto directamente a la fábrica. Se dio cuenta de que los gastos de embarque y manejo se comían sus utilidades, aun cuando el precio de la pieza era sólo del 60% de su precio al menudeo. Sin embargo, los dos problemas más importantes que implicó la sustitución de piezas fueron las pérdidas de ventas y el espacio de almacenamiento. Tom atrajo clientes por la calidad de su servicio y los precios razonables de reparación, los cuales eran posibles gracias a que sus gastos indirectos eran bajos. Por desgracia, muchos clientes potenciales acudían con los vendedores de equipo en vez de esperar varios meses por la reparación. El problema más apremiante era el espacio de almacenamiento. Mientras que un equipo estaba en espera de las piezas de repuesto, tenía que almacenarse en el local. No pasó mucho tiempo para que tanto su taller como su estacionamiento para un solo coche estuvieran saturados de equipos mientras él esperaba las piezas de repuesto. En el segundo año de operación, Tom realmente tuvo que suspender la publicidad como una táctica para limitar el número de clientes ante la falta de espacio de almacenamiento.

Tom ha considerado tener existencias en inventario para su tercer año de operación. Esta práctica reducirá costos de compras haciendo posible obtener descuentos y plazos de envío más favorables. También espera ofrecer mucho mejor tiempo de respuesta a los clientes, mejorando tanto el flujo de efectivo como las ventas. Los riesgos de esta estrategia son los costos que acarrea el manejo de inventarios sin control y la obsolescencia de las piezas.

Antes de aventurarse a surtir un inventario de piezas de repuesto, Tom quiere tener un pronóstico confiable de la actividad del negocio para el próximo año. Él confía lo suficiente en su conocimiento de la mezcla del producto para usar un pronóstico agregado de órdenes de reparación de los clientes como base para el encargo selectivo de las piezas de repuesto. El pronóstico se complica por los patrones estacionales de la demanda y una tendencia hacia arriba de las ventas.

Tom planea desarrollar un pronóstico de ventas para el tercer año de operación. Los datos de las ventas de los primeros dos años se presentan en la tabla 5-6.

¹³Este caso es contribución de William P. Darrow de la Universidad Estatal de Towson, en Towson, Maryland.

TABLA 5-6 Historia de ventas de Small Engine Doctor

<i>Mes</i>	<i>2005 (unidades)</i>	<i>2006 (unidades)</i>	<i>Mes</i>	<i>2005 (unidades)</i>	<i>2006 (unidades)</i>
Enero	5	21	Julio	28	46
Febrero	8	20	Agosto	20	32
Marzo	10	29	Septiembre	14	27
Abril	18	32	Octubre	8	13
Mayo	26	44	Noviembre	6	11
Junio	35	58	Diciembre	26	52

Tarea

1. Grafique los datos en un horizonte de tiempo de dos años de 2005 a 2006. Una los puntos de datos para trazar una gráfica de serie de tiempo.
2. Desarrolle una ecuación de tendencia lineal usando la regresión lineal y grafique los resultados.
3. Estime los factores de ajuste estacional para cada mes dividiendo la demanda promedio para los meses correspondientes entre el promedio de los pronósticos correspondientes de la línea de tendencia. Grafique los valores ajustados y los pronósticos de 2007 que se obtienen a partir del producto de la tendencia por el componente estacional.
4. Suavice las series de tiempo usando la suavización exponencial lineal de Holt con tres grupos de constantes de suavización: $(\alpha = .1, \beta = .1)$, $(\alpha = .25, \beta = .25)$ y $(\alpha = .5, \beta = .5)$. Grafique los tres conjuntos de valores suavizados en la gráfica de la serie de tiempo. Genere pronósticos para el final del tercer año de cada una de las posibilidades consideradas de suavización exponencial ajustada a la tendencia.
5. Calcule los valores de la *MAD* para los dos modelos que visualmente parecen tener los mejores ajustes (los pronósticos más exactos del siguiente periodo).
6. Si tuviera que restringir su selección a uno de los modelos de las preguntas 2 y 4, identifique el modelo que usaría para la planeación de su negocio en 2007 y discuta por qué seleccionó ese modelo por encima de los demás.

CASO 5-2 MR. TUX

John Mosby ha estado mirando hacia delante de la descomposición de su serie de tiempo, las ventas mensuales en dólares. Él sabe que la serie tiene un fuerte efecto estacional y le gustaría medirlo por dos razones. Primero, el ejecutivo bancario está renuente a autorizarle pagos mensuales variables de su préstamo. John le explicó que, a causa de la estacionalidad de las ventas y el flujo de efectivo mensual, le gustaría hacer pagos extras en algunos meses y pagos reducidos —o inclusive no pagar— en otros. Su banquero quiere ver una prueba de lo que John asegura acerca de que sus ventas tienen un fuerte efecto estacional.

Segundo, John quiere pronosticar sus ventas mensuales. Necesita estos pronósticos para efectos de planeación, especialmente desde que su negocio está creciendo. Tanto banqueros como empresas quieren

algún pronóstico sólido sobre el cual basar sus decisiones de inversión. John sabe que su negocio está mejorando y que las perspectivas son alentadoras, pero los inversionistas quieren documentación.

Los volúmenes de ventas mensuales de Mr. Tux de 1999 a 2005 se introducen en Minitab. Puesto que 1998 fue el primer año del negocio, los volúmenes de ventas fueron extremadamente bajos comparados con el resto de los años. Por esta razón, John decide eliminar estos valores del análisis. Los índices estacionales se presentan en la tabla 5-7. El resto del reporte de resultados de la computadora se presenta en la tabla 5-8.

John no está sorprendido de ver los índices estacionales que aparecen en la tabla 5-7 y está contento de tener cifras contundentes para mostrar a su banquero. Después de revisar estas cifras, el banquero está de

TABLA 5-7 Resumen de índices estacionales mensuales para Mr. Tux

<i>Descomposición de la serie de tiempo</i>	
<i>Ecuación de la línea de tendencia</i>	
$\hat{T}_t = 12,133 + 3,033t$	
<i>Índice estacional</i>	
<i>Periodo</i>	<i>Índice</i>
1	0.3144
2	0.4724
3	0.8877
4	1.7787
5	1.9180
6	1.1858
7	1.0292
8	1.2870
9	0.9377
10	0.8147
11	0.6038
12	0.7706
<i>Exactitud del modelo</i>	
<i>MAPE:</i>	20
<i>MAD:</i>	21,548
<i>MSD:</i>	9.12E + 08

TABLA 5-8 Cálculo de los componentes de corto plazo para Mr. Tux

<i>t</i>	<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	1999	Enero	16,850	15,166	1.1111	53,589	3.5336	—	—
2		Febrero	12,753	18,198	0.7708	26,997	1.4835	—	—
3		Marzo	26,901	21,231	1.2671	30,306	1.4274	2.1387	0.6674
4		Abril	61,494	24,264	2.5344	34,572	1.4249	1.7545	0.8121
5		Mayo	147,862	27,297	5.4169	77,092	2.8242	1.7567	1.6077
6		Junio	57,990	30,329	1.9120	48,902	1.6124	1.7001	0.9484
7		Julio	51,318	33,362	1.5382	49,862	1.4946	1.5397	0.9707
8		Agosto	53,599	36,395	1.4727	41,647	1.1443	1.2142	0.9424
9		Septiembre	23,038	39,428	0.5843	24,568	0.6231	1.0325	0.6035
10		Octubre	41,396	42,460	0.9749	50,815	1.1968	0.8550	1.3997
11		Noviembre	19,330	45,493	0.4249	32,014	0.7037	0.8161	0.8623
12		Diciembre	22,707	48,526	0.4679	29,466	0.6072	0.9305	0.6526
13	2000	Enero	15,395	51,559	0.2986	48,961	0.9496	0.7912	1.2002
14		Febrero	30,826	54,592	0.5647	65,257	1.1954	0.8418	1.4201
15		Marzo	25,589	57,624	0.4441	28,828	0.5003	1.0439	0.4793
16		Abril	103,184	60,657	1.7011	58,011	0.9564	1.0273	0.9309
17		Mayo	197,608	63,690	3.1027	103,029	1.6177	0.8994	1.7985
18		Junio	68,600	66,723	1.0281	57,850	0.8670	0.9945	0.8719
19		Julio	39,909	69,755	0.5721	38,777	0.5559	0.9685	0.5740
20		Agosto	91,368	72,788	1.2553	70,994	0.9754	0.8308	1.1740

TABLA 5-8 (Continuación)

<i>t</i>	<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>
21	2001	Septiembre	58,781	75,821	0.7753	62,684	0.8267	0.7927	1.0430
22		Octubre	59,679	78,854	0.7568	73,257	0.9290	0.8457	1.0986
23		Noviembre	33,443	81,887	0.4084	55,387	0.6764	0.8515	0.7944
24		Diciembre	53,719	84,919	0.6326	69,709	0.8209	0.8567	0.9582
25		Enero	27,773	87,952	0.3158	88,327	1.0043	0.7935	1.2657
26		Febrero	36,653	90,985	0.4029	77,592	0.8528	0.9102	0.9370
27		Marzo	51,157	94,018	0.5441	57,632	0.6130	0.9609	0.6380
28		Abril	217,509	97,050	2.2412	122,285	1.2600	0.9401	1.3403
29		Mayo	206,229	100,083	2.0606	107,523	1.0743	0.9579	1.1216
30		Junio	110,081	103,116	1.0676	92,830	0.9003	1.0187	0.8837
31		Julio	102,893	106,149	0.9693	99,973	0.9418	0.9658	0.9751
32		Agosto	128,857	109,182	1.1802	100,124	0.9170	0.9875	0.9287
33	2002	Septiembre	104,776	112,214	0.9337	111,732	0.9957	0.9858	1.0100
34		Octubre	111,036	115,247	0.9635	136,299	1.1827	0.9743	1.2139
35		Noviembre	63,701	118,280	0.5386	105,499	0.8920	0.9516	0.9373
36		Diciembre	82,657	121,313	0.6814	107,260	0.8842	0.9131	0.9683
37		Enero	31,416	124,345	0.2527	99,913	0.8035	0.8246	0.9745
38		Febrero	48,341	127,378	0.3795	102,335	0.8034	0.8507	0.9444
39		Marzo	85,651	130,411	0.6568	96,492	0.7399	0.8951	0.8266
40		Abril	242,673	133,444	1.8185	136,432	1.0224	0.9331	1.0957
41		Mayo	289,554	136,477	2.1216	150,967	1.1062	0.9914	1.1158
42		Junio	164,373	139,509	1.1782	138,614	0.9936	1.0314	0.9634
43		Julio	160,608	142,542	1.1267	156,051	1.0948	1.0312	1.0616
44		Agosto	176,096	145,575	1.2097	136,829	0.9399	0.9960	0.9437
45	2003	Septiembre	142,363	148,608	0.9580	151,815	1.0216	1.0405	0.9819
46		Octubre	114,907	151,640	0.7578	141,051	0.9302	1.0306	0.9026
47		Noviembre	113,552	154,673	0.7341	188,061	1.2159	1.0468	1.1615
48		Diciembre	127,042	157,706	0.8056	164,856	1.0453	1.0502	0.9953
49		Enero	51,604	160,739	0.3210	164,118	1.0210	1.1464	0.8906
50		Febrero	80,366	163,771	0.4907	170,129	1.0388	1.0779	0.9637
51		Marzo	208,938	166,804	1.2526	235,383	1.4111	1.0210	1.3821
52		Abril	263,830	169,837	1.5534	148,327	0.8734	1.0273	0.8501
53		Mayo	252,216	172,870	1.4590	131,500	0.7607	0.9815	0.7751
54		Junio	219,566	175,903	1.2482	185,157	1.0526	0.8819	1.1936
55		Julio	149,082	178,935	0.8332	144,852	0.8095	0.9135	0.8862
56		Agosto	213,888	181,968	1.1754	166,194	0.9133	0.9359	0.9759
57	2004	Septiembre	178,947	185,001	0.9673	190,828	1.0315	0.9281	1.1114
58		Octubre	133,650	188,034	0.7108	164,059	0.8725	0.9857	0.8852
59		Noviembre	116,946	191,066	0.6121	193,682	1.0137	0.9929	1.0210
60		Diciembre	164,154	194,099	0.8457	213,015	1.0975	0.9609	1.1422
61		Enero	58,843	197,132	0.2985	187,140	0.9493	1.0356	0.9167
62		Febrero	82,386	200,165	0.4116	174,405	0.8713	1.0261	0.8492
63		Marzo	224,803	203,198	1.1063	253,256	1.2464	0.9701	1.2847
64		Abril	354,301	206,230	1.7180	199,190	0.9659	1.0295	0.9382
65		Mayo	328,263	209,263	1.5687	171,149	0.8179	1.0488	0.7798
66		Junio	313,647	212,296	1.4774	264,495	1.2459	1.0395	1.1985

TABLA 5-8 (Continuación)

<i>t</i>	Año	Mes	Ventas	<i>T</i>	<i>SCI</i>	<i>TCI</i>	<i>CI</i>	<i>C</i>	<i>I</i>
67		Julio	214,561	215,329	0.9964	208,473	0.9682	1.0231	0.9463
68		Agosto	337,192	218,361	1.5442	262,003	1.1999	1.0177	1.1790
69		Septiembre	183,482	221,394	0.8288	195,664	0.8838	0.9721	0.9092
70		Octubre	144,618	224,427	0.6644	177,522	0.7910	0.9862	0.8020
71		Noviembre	139,750	227,460	0.6144	231,449	1.0175	0.9398	1.0828
72		Diciembre	184,546	230,493	0.8007	239,476	1.0390	1.0367	1.0022
73	2005	Enero	71,043	233,525	0.3042	225,940	0.9675	1.1142	0.8684
74		Febrero	152,930	236,558	0.6465	323,742	1.3686	1.1005	1.2436
75		Marzo	250,559	239,591	1.0458	282,272	1.1781	1.0602	1.1112
76		Abril	409,567	242,624	1.6881	230,261	0.9490	1.0518	0.9023
77		Mayo	394,747	245,656	1.6069	205,813	0.8378	0.9559	0.8765
78		Junio	272,874	248,689	1.0973	230,111	0.9253	0.9492	0.9748
79		Julio	230,303	251,722	0.9149	223,768	0.8890	0.9211	0.9651
80		Agosto	375,402	254,755	1.4736	291,693	1.1450	0.9169	1.2488
81		Septiembre	195,409	257,787	0.7580	208,383	0.8084	0.9599	0.8421
82		Octubre	173,518	260,820	0.6653	212,998	0.8166	1.0337	0.7900
83		Noviembre	181,702	263,853	0.6887	300,928	1.1405	—	—
84		Diciembre	258,713	266,886	0.9694	335,719	1.2579	—	—

acuerdo en que John haga pagos dobles sobre su préstamo en abril, mayo, junio y agosto, y que no haga pagos en enero, febrero, noviembre y diciembre. El ejecutivo bancario le pide una copia de los índices estacionales para mostrárselos a su jefe e incluye la copia en el expediente del préstamo de John.

Para el pronóstico de los primeros seis meses de 2006, John empieza por proyectar valores de la tendencia usando la ecuación de la tendencia $\hat{T}_t = 12,133 + 3,033t$. La tendencia estimada para enero de 2006 es

$$\hat{T}_{85} = 12,133 + 3,033(85) = 269,938$$

A continuación, John obtiene el índice estacional de la tabla 5-7. El índice para enero es 31.44%. John ha estado leyendo *The Wall Street Journal* y viendo los programas de análisis de negocios con regularidad, de manera que ya tiene una idea de la naturaleza general de la economía y su curso en el futuro. También pertenece a un club de servicio de negocios que organiza con regularidad pláticas con expertos en la economía local. Cuando estudia la columna *C* del resultado de su computadora, la cual contiene la historia cíclica de su serie, piensa en cómo pronosticar este valor para los primeros seis meses de 2006. Como los pronósticos de los expertos nacionales y locales anuncian una mejora en los negocios para 2006 y puesto que el último valor de *C* para octubre del 2005 se ha elevado (103.4%), decide usar los siguientes valores de *C* para sus pronósticos:

2005	Noviembre	104
	Diciembre	105
2006	Enero	105
	Febrero	106
	Marzo	107
	Abril	109
	Mayo	110
	Junio	111

Para el valor irregular (*I*) de estos meses, John no prevé ningún acontecimiento inusual, excepto en marzo de 2006. En ese mes, él planea mantener una exhibición abierta al público y reducir las tarifas en una de sus tiendas que está terminando de remodelar. Como resultado de esta promoción, que estará acompañada de publicidad en radio y TV, espera que las ventas en esa tienda sean 50% más altas de lo normal. Por lo que se refiere a sus ventas mensuales totales, cree que este efecto tendrá como resultado un incremento del 15% en sus ventas totales.

Usando todos los cálculos que realizó, junto con los resultados de su computadora, John hace los pronósticos de las ventas de Mr. Tux para los primeros seis meses de 2006, los cuales se muestran en la tabla 5-9.

Después de estudiar los pronósticos de 2006, John se inquieta al ver el rango tan amplio de las proyecciones de las ventas mensuales, que van de \$89,112 a \$595,111. Si bien sabía que su volumen mensual tenía una considerable variabilidad, le preocupa la amplitud

TABLA 5-9 Pronósticos de Mr. Tux

<i>Ventas</i>	<i>Pronóstico</i>	=	<i>T</i>	×	<i>S</i>	×	<i>C</i>	×	<i>I</i>
Enero	89,112	=	269,938		.3144		1.05		1.00
Febrero	136,689	=	272,971		.4724		1.06		1.00
Marzo	301,483	=	276,004		.8877		1.07		1.15
Abril	540,992	=	279,037		1.7787		1.09		1.00
Mayo	595,111	=	282,070		1.9180		1.10		1.00
Junio	375,263	=	285,103		1.1858		1.11		1.00

de las fluctuaciones. John ha estado pensando acerca de la expansión de su ubicación actual en Spokane hacia el área de Seattle. Él descubrió recientemente que existen varios grandes acontecimientos en Seattle que lo hacen diferente de su mercado actual de Spokane. En particular, la celebración de bailes formales de regreso a casa será fastuosa en Seattle, no en Spokane. Puesto que esto tendrá lugar en el otoño, cuando su negocio de Spokane tiene menor actividad (véase los índices estacionales de octubre y noviembre), John ve la ventaja de nivelar su negocio entrando al mercado de Seattle.

Su preocupación real, sin embargo, se enfoca en los peores meses, enero y febrero. Él ha considerado recientemente comprar una máquina de fabricación de camisas para esmoquin que vio en una feria de productos, pensando en que podría dedicarse a esa actividad durante los meses de invierno. Si la reacción de los compradores potenciales de las camisas para ese periodo es positiva, él estaría dispuesto a intentarlo. Por así decirlo, los índices estacionales calculados por su computadora han enfocado su atención en los cambios extremos de los niveles de ventas mensuales.

Preguntas

- Suponga que el ejecutivo bancario de John le pide dos enunciados, para mostrarlos a su jefe, que justifiquen la petición de John de realizar pagos extras del préstamo en algunos meses y dejar de pagar en otros. Escriba estos dos enunciados.
- Suponga que John hará el doble de negocios en Seattle que en Spokane el próximo año. Determine los índices estacionales de Seattle que serían ideales para equilibrar los ingresos mensuales de Mr. Tux.
- Ignorando Seattle, ¿qué volumen de camisas tendría que confeccionar John en su máquina para que tanto enero como febrero resulten con un nivel “promedio”?

CASO 5-3 CONSUMER CREDIT COUNSELING

El Consumer Credit Counseling (CCC) se describió en el 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 serían atendidos durante el resto de 1993. Marv entregó a Dorothy Mercer datos mensuales del número de clientes nuevos atendidos por CCC durante el periodo de enero de 1985 a marzo de 1993 (véase el caso 3-3).

Dorothy le dio a usted estos datos y le pidió que realizara un análisis de descomposición de la serie de tiempo. Ella enfatizó que quería comprender plenamente los componentes de tendencia y estacionalidad. Dorothy quiere conocer la importancia de cada componente. También quiere saber si en los datos aparecen algunas irregularidades inusuales. Por último, le solicitó elaborar el pronóstico para el resto de 1993.

Tarea

Escriba un reporte para Dorothy con la información que le solicitó.

CASO 5-4 MURPHY BROTHERS FURNITURE

En el caso 4-4, Julie Murphy desarrolló un modelo informal que combinaba estimaciones estacionales y de tendencia (similar a la ecuación 4.5). Una de las razones principales para seleccionar este modelo informal fue su sencillez. Julie sabía que su padre, Glen, necesitaba comprender el modelo de pronóstico que utilizaba la compañía.

Suponga que ahora estamos en octubre de 2002, y hay muchos cambios. Glen Murphy se retiró. Julie ha tomado algunos cursos de negocios en la universidad local, incluyendo uno de pronósticos de negocios. Murphy Brothers Furniture construyó una fábrica en Dallas y empezó a fabricar su propia línea de muebles en octubre de 1995.

En la tabla 5-10 se presentan los datos mensuales de ventas de Murphy Brothers Furniture de 1996 a octubre de 2002. Como indica el patrón de estos datos en la figura 5-13, las ventas han aumentado notoriamente desde 1996. Por desgracia, la figura 5-13 también indica que uno de los problemas de la demanda es que es un tanto estacional. La política general de la compañía es contratar dos turnos durante el verano y los primeros meses del otoño, y luego trabajar un solo turno durante el resto del año. De este modo, se conformarán inventarios sustanciales al final del verano y los meses del otoño hasta que la demanda comience a

mejorar en noviembre y diciembre. Ante estos requerimientos de producción, Julie está muy ansiosa por elaborar pronósticos de corto plazo para la compañía que se apoye en la mejor información disponible referente a la demanda.

Para efectos del pronóstico, Julie decidió usar sólo los datos reunidos desde 1996, el primer año completo en que Murphy Brothers fabricó su propia línea de muebles (tabla 5-10). Julie puede ver (tabla 5-13) que sus datos tienen tanto tendencia como estacionalidad. Por esta razón, decide usar un método de descomposición de series de tiempo para analizar su variable de ventas.

Como la figura 5-13 indica que la serie de tiempo que Julie está analizando tiene aproximadamente la misma variabilidad a lo largo de la serie, ella decide usar un modelo de componentes aditivos para elaborar el pronóstico. Así, corre el modelo $Y_t = T_t + S_t + I_t$. En la tabla 5-11 se presenta un resumen de los resultados. Julie observa el patrón de autocorrelación de los residuos (véase la figura 5-14) para verificar la aleatoriedad. Los residuos no son aleatorios y el modelo no parece ser el adecuado.

Julie está atorada. Ella intentó un modelo informal que combinaba estimaciones estacionales y de tendencia, la suavización exponencial de Winters y la descomposición clásica. Finalmente, Julie decide hacer un ajuste para dejar la estacionalidad fuera de los datos,

TABLA 5-10 Ventas mensuales de Murphy Brothers Furniture, 1996 a 2002

	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>Mayo</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
1996	4,964	4,968	5,601	5,454	5,721	5,690	5,804	6,040	5,843	6,087	6,469	7,002
1997	5,416	5,393	5,907	5,768	6,107	6,016	6,131	6,499	6,249	6,472	6,946	7,615
1998	5,876	5,818	6,342	6,143	6,442	6,407	6,545	6,758	6,485	6,805	7,361	8,079
1999	6,061	6,187	6,792	6,587	6,918	6,920	7,030	7,491	7,305	7,571	8,013	8,727
2000	6,776	6,847	7,531	7,333	7,685	7,518	7,672	7,992	7,645	7,923	8,297	8,537
2001	7,005	6,855	7,420	7,183	7,554	7,475	7,687	7,922	7,426	7,736	8,483	9,329
2002	7,120	7,124	7,817	7,538	7,921	7,757	7,816	8,208	7,828			

Fuente: Registros de ventas de Murphy Brothers Furniture.

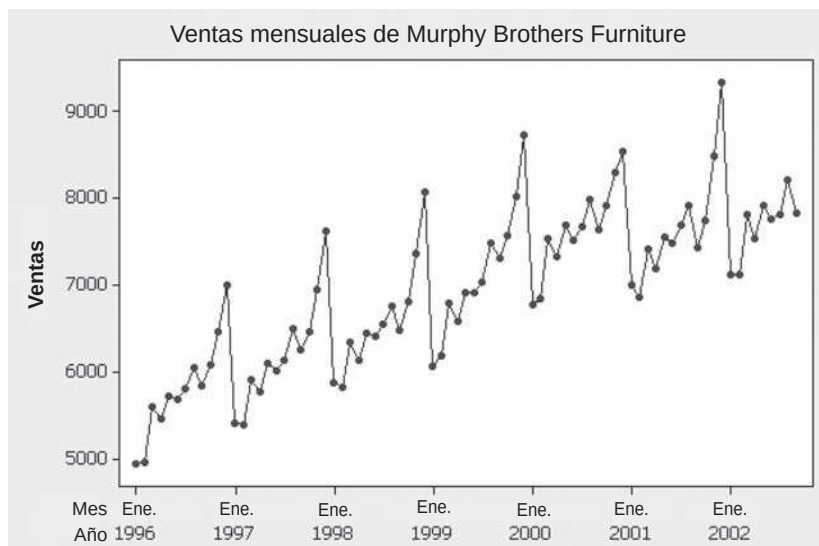


FIGURA 5-13 Ventas mensuales de Murphy Brothers Furniture, 1996 a 2002

TABLA 5-11 Resumen del modelo de descomposición para Murphy Brothers Furniture

Descomposición de las series de tiempo

Ecuación de la línea de tendencia

$$\hat{T}_t = 5,672 + 31.4t$$

Índice estacional

<i>Periodo</i>	<i>Índice</i>
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
<i>Exactitud</i>	<i>Medidas</i>
MAPE:	1.9
MAD:	135.1
MSD:	30,965.3

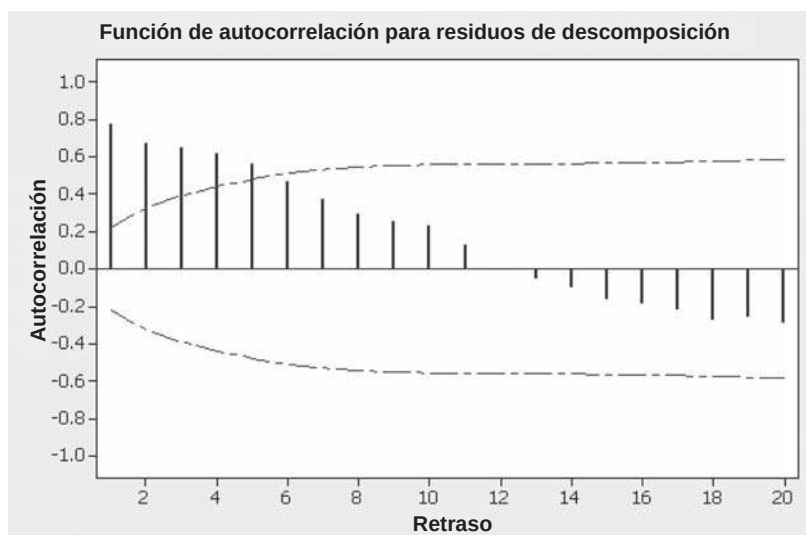


FIGURA 5-14 Función de autocorrelación para los residuos usando un modelo de descomposición de series de tiempo para Murphy Brothers Furniture

TABLA 5-12 Ventas mensuales ajustadas estacionalmente de Murphy Brothers Furniture, 1996 a 2002

	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>Mayo</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
1996	5,621	5,671	5,745	5,821	5,775	5,863	5,847	5,818	5,901	5,941	5,857	5,767
1997	6,091	6,096	6,051	6,135	6,161	6,189	6,174	6,277	6,307	6,326	6,334	6,380
1998	6,551	6,521	6,486	6,510	6,496	6,580	6,588	6,536	6,543	6,659	6,749	6,844
1999	6,736	6,890	6,936	6,954	6,972	7,093	7,073	7,269	7,363	7,425	7,401	7,492
2000	7,451	7,550	7,675	7,700	7,739	7,691	7,715	7,770	7,703	7,777	7,685	7,302
2001	7,680	7,558	7,564	7,550	7,608	7,648	7,730	7,700	7,484	7,590	7,871	8,094
2002	7,795	7,827	7,961	7,905	7,975	7,930	7,859	7,986	7,886			

de manera que pueda aplicar las técnicas de pronóstico que no manejan datos estacionales. Julie desestacionaliza los datos sumando o restando el índice estacional para el mes apropiado. Por ejemplo, suma 674.60 a los

datos de cada enero y resta 1,234.63 de los datos de cada diciembre. La tabla 5-12 presenta los datos ajustados estacionalmente.

Tarea

1. Con base en los datos de 2001 en la tabla 5-12, desarrolle un modelo para pronosticar los datos de ventas ajustados a la estacionalidad y generar pronósticos para los primeros nueve meses de 2002.
2. De acuerdo con los pronósticos del inciso 1, pronostique las ventas para los primeros nueve meses de 2002 sumando o restando el índice estacional apropiado de la tabla 5-11. ¿Son exactos estos pronósticos cuando se comparan con los valores reales?
3. Pronostique las ventas de octubre de 2002 usando el mismo procedimiento del inciso 2.
4. Compare el patrón de los datos de ventas al menudeo presentados en el caso 3-1A con el patrón de los datos de ventas reales de 1992 a 1995 presentados en el caso 4-4 y con el patrón de los datos de ventas reales de 1996 a 2001 presentados en este caso.

CASO 5-5 AAA WASHINGTON¹⁴

En 1993 AAA Washington era uno de los dos clubes regionales de conductores de automóviles afiliado a la Asociación Americana de Automovilistas (AAA o triple A) que operaban en el estado de Washington. En esa época, el 69% de las personas que pertenecían a algún club de automovilistas eran socios de la Asociación Americana del Automóvil, lo que hizo de ésta el club de conductores de automóviles más grande de Norteamérica. AAA era una asociación nacional que atendía a sus socios individuales a través de una federación de aproximadamente 150 clubes regionales que optaban por afiliarse a la asociación nacional. La asociación estableció un número mínimo de normas que los clubes afiliados tenían que cumplir para mantener su afiliación. Cada club regional era dirigido localmente por su propio consejo y equipo de administración. Los miembros del consejo de administración local y los gerentes eran responsables del reclutamiento y la retención de socios dentro de sus territorios y de garantizar la salud financiera del club regional. Más allá del cumplimiento de las normas mínimas establecidas por AAA, cada club regional era libre para definir qué productos y servicios adicionales ofrecería y cómo los cobraría.

AAA Washington fue fundada en 1904. Su territorio estaba integrado por los 26 condados de Washington al oeste del río Columbia. El club ofrecía a sus socios una variedad de servicios relacionados con la conducción de automóviles y con los viajes en este medio de transporte. Los beneficios a los socios brindados en cooperación con la asociación nacional incluían servicios de emergencia en el camino; servicio clasificado de alojamiento, restaurantes y talleres de reparación de automóviles; viajes guiados a los alojamientos aprobados por AAA, restaurantes, campamentos y puntos de interés; y asistencia legal a los miembros. Además de estos servicios, AAA Washington ofrecía a sus socios planes de protección ampliada para servicios de emergencia en el camino; servicios financieros, incluyendo tarjetas de crédito afines, líneas de crédito personal, cuentas de cheques y ahorro, depósitos a plazo, y cheques de viajero de American Express sin cargo; acceso a una flota de camionetas para diagnóstico de automóviles para determinar el estado del vehículo de un socio; una agencia de viajes; y

una agencia de seguros. El club brindaba estos servicios a través de una red de oficinas localizadas en Bellevue, Bellingham, Bremerton, Everett, Lynnwood, Olympia, Renton, Seattle, Tacoma, las tres ciudades (Pasco, Richland y Kennewick), Vancouver, Wenatchee y Yakima, en Washington.

La investigación del club ha demostrado de forma constante que el servicio de emergencia en el camino era la razón principal por la que la gente se unía a AAA. La importancia del servicio de emergencia en el camino para la seguridad de los socios se reflejaba en los tres tipos de membresías ofrecidas por AAA Washington: Básica, AAA Plus y AAA Plus RV. La membresía básica proporcionaba a los socios cinco millas de arrastre con grúa desde el punto en el que su vehículo se había descompuesto. AAA Plus ofrecía a los socios 100 millas de arrastre con grúa desde el punto en el que su vehículo se había descompuesto. AAA Plus RV brindaba a los socios 100 millas de servicio de arrastre a los socios que tenían vehículos recreativos (campers), automóviles de pasajeros y camiones ligeros. El servicio de emergencia en el camino era el gasto operativo individual más grande del club. Se había proyectado que para el siguiente año fiscal, el otorgamiento del servicio de emergencia en el camino costaría \$9.5 millones, el 37% del presupuesto operativo anual del club.

Michael DeCoria, un graduado de la carrera de contaduría (CPA) con una maestría en administración (MBA) de la Universidad del Este de Washington, se unió recientemente al equipo de administración del club como vicepresidente de operaciones. Una de las responsabilidades que Michael asumió era la administración del servicio de emergencia en el camino. Al principio de la evaluación de la operación de este servicio, DeCoria descubrió que los costos se habían incrementado a un ritmo más rápido de lo que podía justificarse con la tasa de inflación o el incremento en las afiliaciones al club. Michael empezó por analizar la manera en que el club prestaba el servicio de emergencia en el camino para determinar si los costos podrían controlarse más de cerca en esta área.

El servicio de emergencia en el camino se otorgaba en una de cuatro modalidades: a través de la flota de servicio de AAA Washington, compañías contratistas, reembolso recíproco y reembolso directo. Los

¹⁴Este caso fue proporcionado por Steve Branton, ex alumno y graduado de MBA, de la Universidad del Este de Washington.

vehículos de la flota de servicio de AAA Washington respondían a las llamadas de solicitud de servicio de los socios que tenían problemas en el área del centro de Seattle. Dentro del área de servicio de AAA Washington, pero fuera del centro de Seattle, el servicio era proporcionado por las compañías comerciales de grúas que había contratado AAA Washington. Los socios tramitaban ambos tipos de servicio llamando al centro de atención del club. Si un socio tenía problemas fuera del área de AAA Washington, podía llamar al club AAA local afiliado para recibir el servicio. El club afiliado pagaba por este servicio y luego lo facturaba a AAA Washington para un reembolso recíproco a través de un servicio de compensación proporcionado por la aso-

ciación nacional. Finalmente, los socios podían hacer contacto directamente con una compañía de grúas de su elección, pagando por el servicio de la grúa y luego enviando un requerimiento de reembolso por parte del club. AAA Washington reembolsaba directamente al socio el costo real del arrastre o \$50, lo que fuera menor. Después de un examen cuidadoso de los cuatro métodos de prestación del servicio del club, Michael concluyó que el club estaba controlando el costo de la entrega del servicio tan estrechamente como era posible.

Otra fuente posible de incremento de costos era un aumento en el uso del servicio. El número de afiliados había aumentado constantemente por varios años, pero el aumento en el costo era más de lo que podría

TABLA 5-13 Volumen de llamadas mensuales de emergencias en el camino de AAA Washington

<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
1989	Enero	23,327	1992	Septiembre	19,647
	Febrero	24,050		Octubre	22,013
	Marzo	24,010		November	22,375
	Abril	19,735		Diciembre	22,727
	Mayo	20,153		Enero	22,367
	Junio	19,512		Febrero	21,155
	Julio	19,892		Marzo	21,209
	Agosto	20,326		Abril	19,286
1990	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
	Abril	20,155		Diciembre	26,039
	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			

atribuirse tan sólo al aumento de afiliados. Entonces Michael revisó para ver si existía un aumento en el uso del servicio sobre una base per cápita. Descubrió que entre el año fiscal 1990 y el año fiscal 1991, el número promedio de llamadas de emergencia por socio para solicitar el servicio creció en 3.28%, de un promedio de

0.61 llamadas por socio a 0.63 llamadas. A DeCoria le preocupaba el hecho de que si esta tendencia continuaba, podría haber un efecto negativo en las finanzas del club; así que reunió los datos sobre el volumen de las llamadas del servicio de emergencia en el camino, los cuales se presentan en la tabla 5-13.

Tarea

1. Realice la descomposición de una serie de tiempo con los datos de llamadas de emergencia para solicitar el servicio de asistencia en el camino de AAA.
2. Escriba un memorando a DeCoria en el que resuma los hechos importantes de los cambios en

las llamadas de los servicios de emergencia en el camino que usted descubrió con el análisis de la descomposición de la serie de tiempo.

CASO 5-6 ALOMEGA FOOD STORES

En el ejemplo 1.1, Julie Ruth, presidenta de Alomega Food Stores, recolectó datos sobre las ventas mensuales de su compañía junto con otras variables que podrían estar relacionadas con las ventas (revise el ejemplo 1.1). En el caso 2-3 se explicó cómo Julie usó su programa de Minitab para calcular la ecuación de regresión múltiple usando la mejor predicción de las ventas mensuales.

Después de revisar los resultados de este análisis de regresión, incluyendo el bajo valor de r^2 (36%), Julie decidió intentar la descomposición de series de tiempo de una variable individual: las ventas mensuales. La figura 5-15 presenta la gráfica de los datos de ventas que obtuvo (véase el caso 3-4). Parece que las ventas estuvieran ampliamente distribuidas alrededor de la línea de tendencia para los pronósticos exactos. Confirmó

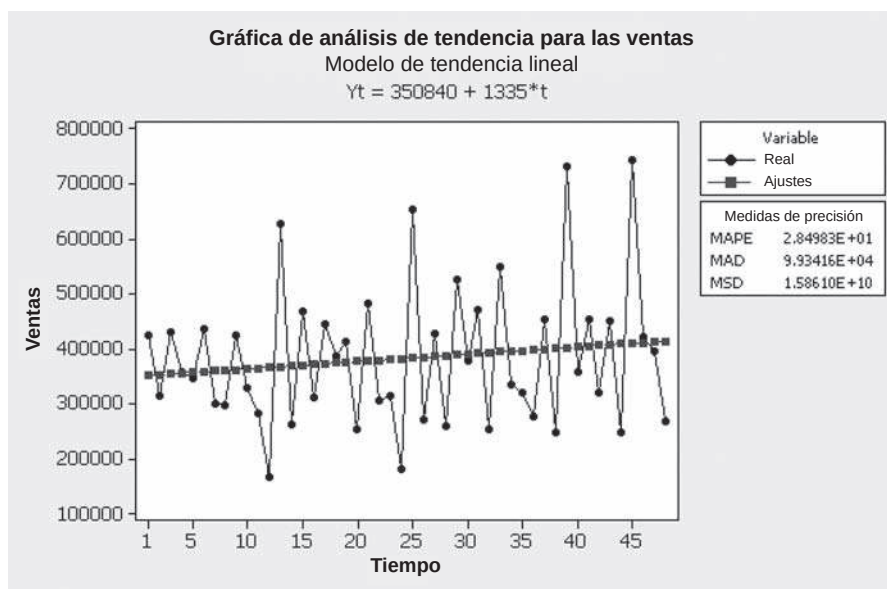


FIGURA 5-15 Análisis de tendencia de las ventas de Alomega Food Stores

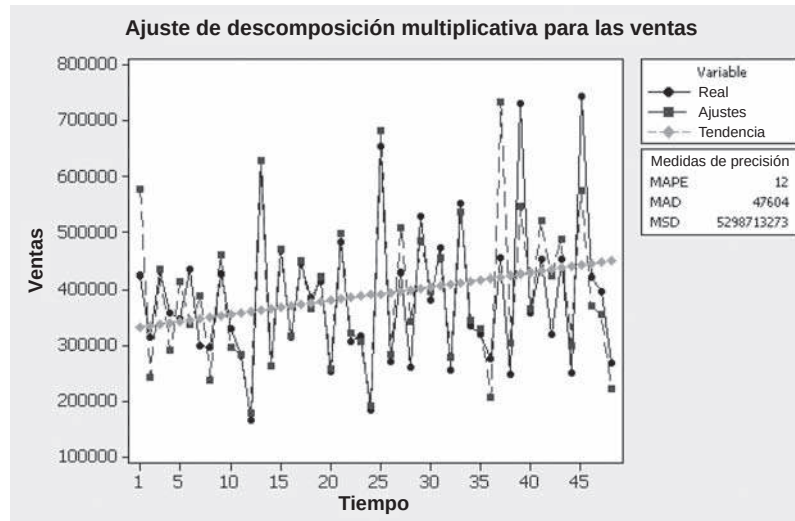


FIGURA 5-16 Ajuste de descomposición de las ventas de Alomega Food Stores

esta impresión cuando observó el valor de 28 del *MAPE*. Ella interpretó que esto significaba que el porcentaje de error promedio entre los valores reales y la línea de tendencia era 28%, un valor que consideró muy alto. Luego intentó una descomposición multiplicativa de los datos. Los resultados se presentan en la figura 5-16.

Además de la ecuación de la tendencia presentada en el reporte, Julie estaba interesada en los índices (mensuales) estacionales que calculó el programa. Ella notó que el mes con las ventas más bajas era diciembre (mes 12, índice = 0.49) y el más alto era enero (mes 1, índice 1.74). Ella estaba consciente de la gran diferencia entre diciembre y enero, pero no se dio cuenta qué tan extrema era.

También notó que el *MAPE* había caído al 12%, una mejora definitiva sobre el valor obtenido cuando sólo se usó la ecuación de la tendencia.

Finalmente, Julie hizo que el programa elaborara pronósticos de los siguientes 12 meses, usando las proyecciones de la ecuación de la tendencia modificada por los índices estacionales. Pensó que podría usarlos como pronósticos para efectos de planeación, pero se preguntaba si otro método podría arrojar mejores pronósticos. También estaba preocupada por lo que su gerente de producción, Jackson Tilson, podría decir acerca de sus pronósticos, especialmente porque había expresado su preocupación por el uso de la computadora para hacer predicciones (véase este comentario al final del ejemplo 1.1).

Pregunta

1. ¿Qué podría decir Jackson Tilson acerca de los pronósticos de Julie?

CASO 5-7 SURTIDO COOKIES

Las ventas mensuales (en kilogramos) de las galletas Surtido se reportaron en el caso 3-5. En el caso 4-8, Jaime Luna examinó las autocorrelaciones de las ventas de galletas para determinar si podría existir un

componente estacional. También consideró un método de suavización para generar pronósticos de las ventas futuras. Un miembro del equipo de Jaime tiene alguna experiencia con el análisis de descomposición y sugiere

que se intente una descomposición multiplicativa de los datos de ventas de las galletas. No sólo se obtendría una indicación de la tendencia en las ventas, sino que también se podrían ver los índices estacionales. Estos últimos son piezas importantes de información para determinar los requerimientos de la flotilla de camiones y de almacenamiento (inventario) durante el año. Jaime

no tiene mucha fe en un procedimiento que intente estimar componentes que no pueden observarse directamente. Sin embargo, puesto que Minitab está disponible, está de acuerdo en hacer un intento de descomposición y utilizarla con la finalidad de generar pronósticos de ventas para los meses restantes de 2003.

Preguntas

1. Realice una descomposición multiplicativa de los datos de ventas de Surtido Cookies, guarde los residuos y genere pronósticos de ventas para los meses restantes de 2003.
2. ¿De qué se enteró Jaime en relación con la tendencia en ventas? ¿Qué le dicen los índices estacionales?
3. Calcule las autocorrelaciones residuales. Revisando las autocorrelaciones residuales y los pronósticos de ventas por el resto de 2003, ¿debe Jaime modificar sus creencias acerca del valor del análisis de descomposición? Explique.

CASO 5-8 SOUTHWEST MEDICAL CENTER

Mary Beasley y su objetivo de pronosticar el número de visitas futuras facturables al Medical Oncology de Southwest Medical Center fueron presentados en el caso 4-7. En ese caso, Mary se enteró de que había un componente estacional en sus datos y consideró el uso del método de suavización de Winters para generar los pronósticos de las visitas futuras. Mary no estaba totalmente satisfecha con este análisis, puesto que parecía haber algunas autocorrelaciones residuales significativas. Más aún, Mary estaba interesada en aislar los índices estacionales porque quería tener total eviden-

cia de que algunos periodos del año estaban más saturados que otros. (La programación de los médicos nunca es una tarea fácil). Un colega de Mary le sugirió que considerara el análisis de descomposición, una técnica con la cual Mary no estaba familiarizada. Sin embargo, como tenía disponible un paquete de software estadístico que incluía la descomposición, Mary estaba dispuesta a intentarlo. También se dio cuenta de que tenía que entender la descomposición lo suficientemente bien para vender los resultados a la administración central si era necesario.

Preguntas

1. Redacte un breve escrito para explicar a Mary la descomposición de una serie de tiempo..
2. Ejecute una descomposición multiplicativa del total de visitas facturables, guarde los residuos y genere los pronósticos de las visitas para los siguientes 12 meses, usando febrero del año fiscal 2003-2004 como el pronóstico de origen.
3. Interprete el componente de la tendencia para Mary. ¿Qué aprendió ella de los índices estacionales?
4. Calcule las autocorrelaciones residuales. Dadas las autocorrelaciones residuales y los pronósticos, ¿debe Mary estar contenta con el análisis de descomposición del total de visitas facturables? Explique.

Aplicaciones de Minitab

El problema. En el ejemplo 5.1 se desarrolló una ecuación de la tendencia para el registro anual de automóviles nuevos de pasajeros en Estados Unidos de 1960 a 1992.

Solución de Minitab

1. Después de introducir los datos de registro de automóviles nuevos de pasajeros en la columna C1 de la hoja de trabajo, haga clic en los siguientes menús para correr el análisis de la tendencia:

Stat>Time Series>Trend Analysis

2. Aparece la ventana de diálogo Trend Analysis.
 - a) La variable es Cars.
 - b) Haga clic en Linear para Model Type.
 - c) Haga clic en Generate forecasts, y coloque un 1 en Number of forecasts para el pronóstico de 1993.
 - d) Haga clic en Options. En el espacio provisto para TITLE, escriba Linear Trend for Annual Car Registrations.
 - e) Haga clic en OK en la ventana de diálogo de Options. Haga clic en OK otra vez y aparecerá la gráfica que se presenta en la figura 5-3.

El problema. La tabla 5-1 se construyó con la finalidad de mostrar la tendencia estimada y los errores calculados para los datos del registro de automóviles nuevos de pasajeros (véase la p. 170).

Solución de Minitab

1. La columna C1 está titulada como Year, la C2 como Y, la C3 como t, la C4 como Estimates, y la C5 está titulada Error. Haga clic en los siguientes menús para crear los años:

Calc>Make Patterned Data>Simple Set of Numbers

2. Aparece la ventana de diálogo Simple Set of Numbers.
 - a) Se deben dar las siguientes respuestas:
Almacenar datos con patrón en: C1
Desde el primer valor: 1960
Hasta el último valor: 1992
En pasos de: 1
 - b) Haga clic en OK y aparecerán los años en C1.
 - c) Introduzca los datos de automóviles nuevos de pasajeros registrados en C2.
3. Los datos codificados de tiempo t se introducen en C3, usando la ventana de diálogo Simple Set of Numbers.
4. En C4 se introducen las estimaciones de la tendencia haciendo clic en los mismos menús de análisis de la tendencia usados para resolver el ejemplo 5.1, con un paso adicional.
5. Haga clic en el menú Storage y obtendrá la ventana de diálogo Trend Analysis Storage.
 - a) Abajo de Storage, haga clic en Fits (línea de tendencia) y Residuals (datos sin tendencia).
 - b) Haga clic en OK en esta ventana de diálogo y luego en la ventana Trend Analysis. Las estimaciones de la tendencia aparecerán en C4 y los errores (residuos) aparecerán en C5.

El problema. En los ejemplos 5.3 y 5.4, Perkin Kendell, el analista de Coastal Marine Corporation, quería pronosticar las ventas trimestrales de 2007.

Solución de Minitab

1. Introduzca los años correspondientes en la columna C1, los trimestres en C2 y los datos en C3. Para correr un modelo de descomposición, haga clic en los siguientes menús:

Stat>Time Series>Decomposition

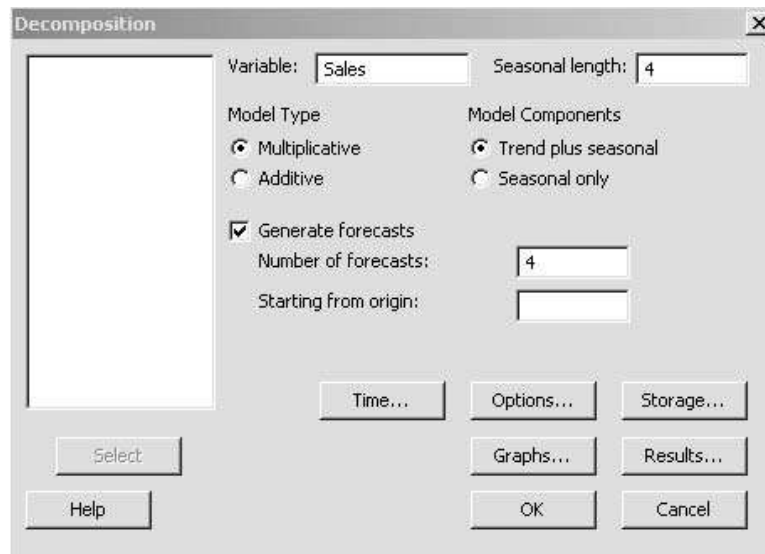


FIGURA 5-17 Ventana de diálogo de Descomposición en Minitab

2. Aparece la ventana de diálogo Decomposition mostrada en la figura 5-17.
 - a) La variable es C3 o Sales.
 - b) Como los datos son trimestrales, la duración estacional (Seasonal length) es 4.
 - c) El tipo de modelo es multiplicativo y los componentes del modelo son el de tendencia y el estacional (Trend plus seasonal).
 - d) Haga clic en Options. El periodo estacional inicial es 1. Haga clic en OK.
 - e) Haga clic en Generate forecasts y escriba el número 4 en Number of forecasts.
3. Haga clic en el menú Storage y obtendrá la ventana de diálogo Decomposition-Storage mostrada en la figura 5-18.
 - a) Debajo del menú Storage, seleccione Trend line, Detrended data, Seasonals y Seasonally adjusted data.
 - b) Haga clic en OK tanto para esta ventana de diálogo como en la ventana de diálogo Decomposition. La tabla 5-4 presenta las estimaciones de la tendencia en la columna C4 (titulada como *T*), los datos sin tendencia en la columna C5 (titulada como *SCI*), las estacionalidades en la columna C6 (titulada como *S*), y los datos ajustados estacionalmente en la columna C7 (titulada como *TCI*).
4. Las figuras 5-10, 5-11 y 5-12 aparecen en la pantalla y pueden imprimirse una por una usando los siguientes menús:

File>Print Graph
5. Después de imprimir las gráficas, haga clic en

File>Print Session Window

y se imprimirán los pronósticos presentados en la figura 5-9.

Las columnas *CI*, *C* e *I* se pueden calcular usando Minitab, como se demuestra en los siguientes pasos.

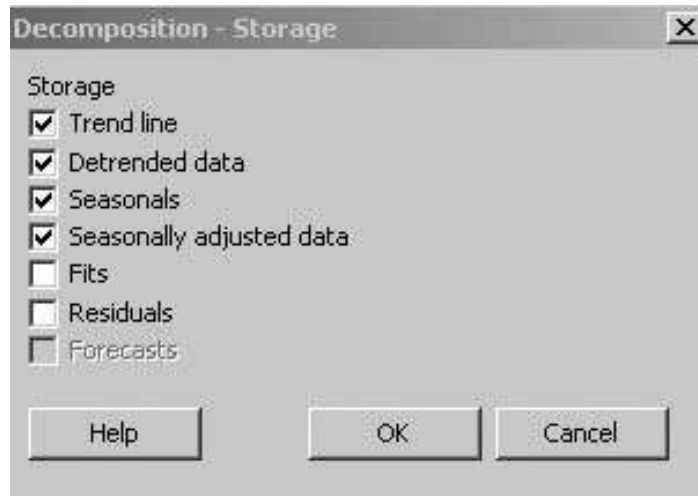


FIGURA 5-18 Ventana de diálogo Descomposición-almacenamiento en Minitab

6. Etiquete la columna C8 CI. Usando el menú Calc>Calculator de Minitab, genere la columna C8 dividiendo la columna C7 entre la columna C4, así que $CI = TCI/T$.
7. Etiquete la columna C9 C. Genere la columna C9 seleccionando un promedio móvil centrado de orden 3 de los valores en la columna C8. Use el menú

Stat>Time Series>Moving Average

Asegúrese de verificar la ventana Center the moving averages.

8. Etiquete la columna C10I. Usando el menú Calc>Calculador de Minitab, genere la columna C10 dividiendo la columna C8 entre la columna C9, así que $I = CI/C$. Se completan ahora las columnas restantes de la tabla 5-4.

Aplicaciones de Excel

El problema. La figura 5-6 presenta los datos y la gráfica correspondientes a los vendedores del fondo común (véase la p. 173). Se necesita un modelo de tendencia exponencial para ajustar estos datos.

Solución de Excel

1. Introduzca Y en A1 y los datos de los vendedores en A2:A8.
2. Introduzca X en B1 y la variable codificada X (1, 2, 3, 4, 5, 6, 7) en B2:B8.
3. Introduzca Log Y en C1, y produzca los logaritmos (de base 10) de la variable Y introduciendo la fórmula = LOG10(A2) en la celda C2. Copie esta fórmula para el resto de la columna.
4. Haga clic en los siguientes menús para calcular la tendencia exponencial:

Tools>Data Analysis

5. Aparece la ventana de diálogo Data Analysis. Debajo de Analysis Tools, seleccione Regression y haga clic en OK. Aparecerá la ventana de diálogo de Regresión mostrada en la figura 5-19.

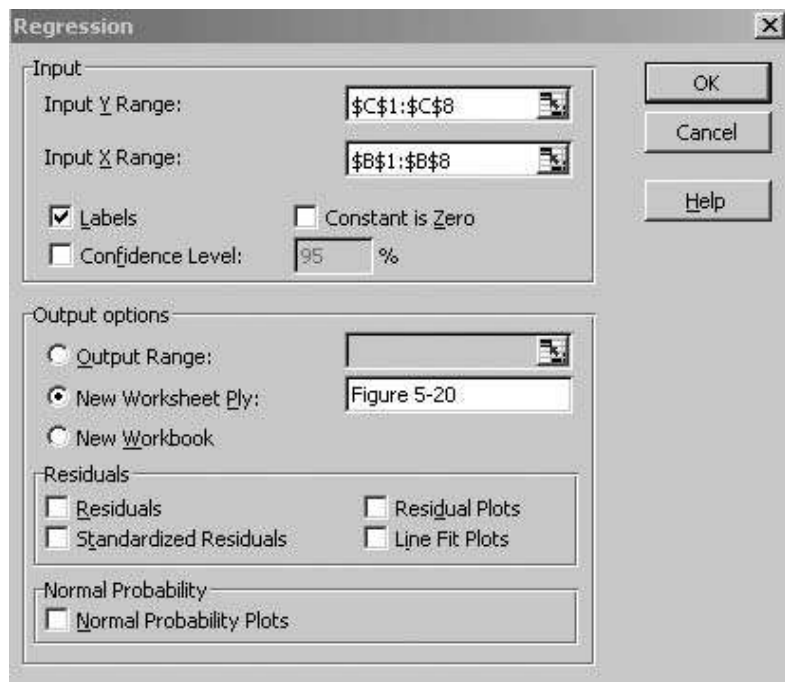


FIGURA 5-19 Ventana de diálogo para la regresión de Excel

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Type a question for help

Snagit Window

Arial 10 B I U

A22 fx

1	SUMMARY OUTPUT							
2								
3	Regression Statistics							
4	Multiple R	0.9984						
5	R Square	0.9969						
6	Adjusted R Square	0.9963						
7	Standard Error	0.0157						
8	Observations	7						
9								
10	ANOVA							
11		df	SS	MS	F	Significance F		
12	Regression	1	0.392106065	0.392106	1600.788	1.8389E-07		
13	Residual	5	0.001224728	0.000245				
14	Total	6	0.393330793					
15								
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%
17	Intercept	1.00069	0.01323	75.65363	7.64E-09	0.96669	1.03469	0.96669
18	X	0.11834	0.00296	40.00985	1.84E-07	0.11073	0.12594	0.11073
19								
20								
21								
22								
23								
24								
25								

Figure 5-20 Sheet1 Sheet2 Sheet3

Ready NUM

FIGURA 5-20 Resultado de Excel de la regresión para el ejemplo de los vendedores del fondo común

- a) Introduzca C1:C8 como el rango de Y (Input Y Range).
- b) Introduzca B1:B8 como el rango de X (Input X Range).
- c) Seleccione la casilla de verificación Labels.
- d) Introduzca la figura 5-20 como el nombre de New Worksheet Ply. Haga clic en OK.

La figura 5-20 representa el resultado de Excel de un modelo exponencial de los vendedores del fondo común. La ecuación es

$$\log \hat{T} = 1.00069 + 0.11834t$$

Los antilogaritmos de los coeficientes de regresión en esta ecuación son

$$b_0 = \text{antilog } 1.00069 = 10.016$$

$$b_1 = \text{antilog } 0.11834 = 1.313$$

Por lo tanto, la ecuación de la tendencia exponencial ajustada es

$$\hat{Y} = (10.016)(1.313)^t$$

Referencias

- Bell, W. R. y S. C. Hillmer. "Issues Involved with the Seasonal Adjustment of Economic Time Series", *Journal of Business and Economic Statistics* 2 (1984):291-320.
- Bowerman, B. L., R. T. O'Connell y A. B. Koehler. *Forecasting, Time Series and Regresión*, 4a. ed. Belmont, Calif.: Thomson Brooks/Cole, 2005.
- Diebold, F. X. *Elements of Forecasting*, 3a. ed. Cincinnati, Ohio: 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* 16 (1998): 127-152.
- Levenbach, H. y J. P. Cleary. *Forecasting: Practice and Process for Demand Management*. Belmont, Calif.: Thomson Brooks/Cole, 2006.
- Makridakis, S., S. C. Wheelwright y R. J. Hyndman. *Forecasting Methods and Applications*, 3a. ed. Nueva York: Wiley, 1998.
- Moore, G. H. y J. Shiskin. "Early Warning Signals for the Economy". En *Statistics: A Guide to Business and Economics*, J. M. Tanur *et al.*, eds. San Francisco: Holden-Day, 1976.

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 vimos cómo ajustar una línea recta a pares de observaciones de las dos variables usando el método de los mínimos cuadrados. En este capítulo se estudian con detalle los modelos de regresión lineal simple (línea recta). Una vez que se establece una relación lineal, el conocimiento de la variable independiente servirá para pronosticar la variable dependiente.

Con la finalidad de revisar el análisis de la relación entre dos variables, presentado en el capítulo 2, considere el siguiente ejemplo.

Ejemplo 6.1

Suponga que el señor Bump observa el precio y volumen de los galones de leche vendidos durante 10 semanas seleccionadas al azar. Los datos que recolectó se presentan en la tabla 6-1.

Primero, elabora un diagrama de dispersión de los datos, que se muestra en la figura 6-1. En este diagrama parece que existe una relación lineal negativa entre Y , el número de galones de leche vendidos, y X , el precio por galón. Al parecer, conforme el precio sube, el volumen baja.

Ahora, Bump desea medir el grado de esta relación aparente calculando el coeficiente de correlación de la muestra, r . Con base en la ecuación 2.10 y los cálculos de la tabla 6-2, él obtiene

$$\begin{aligned} r &= \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2}\sqrt{n\sum Y^2 - (\sum Y)^2}} \\ &= \frac{10(149.3) - (14.4)(112)}{\sqrt{10(21.56) - (14.4)^2}\sqrt{10(1,488) - (112)^2}} \\ &= \frac{-119.8}{138.7} = -.86 \end{aligned}$$

TABLA 6-1 Datos de los galones de leche para el ejemplo 6.1

<i>Semana</i>	<i>Nivel de ventas semanal, Y (miles de galones)</i>	<i>Precio de venta 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

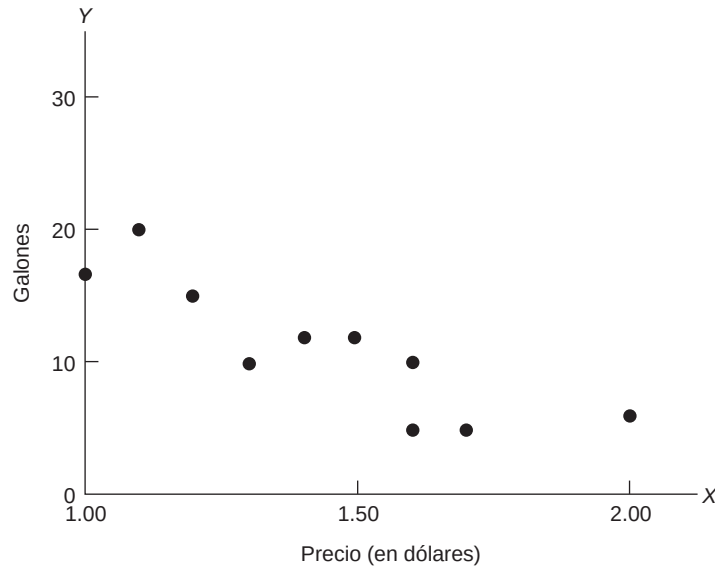


FIGURA 6-1 Diagrama de dispersión para el ejemplo 6.1

TABLA 6-2 Cálculos necesarios para el ejemplo 6.1

	<i>Y</i>	<i>X</i>	<i>XY</i>	<i>X</i> ²	<i>Y</i> ²
<i>n</i> = 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 $-.86$ indica una relación negativa bastante fuerte entre *Y* y *X*: conforme el precio del galón de leche aumenta, el número de galones vendidos disminuye.

La pregunta que se puede hacer ahora es: ¿En qué medida descende el volumen conforme el precio se eleva? Esta pregunta sugiere la graficación de una línea recta por los puntos de datos representados en el diagrama de dispersión. Después de que se ha dibujado la línea, la pendiente de ésta indicará la disminución promedio del volumen *Y* por cada dólar de incremento en el precio *X*.

LÍNEA DE REGRESIÓN

El señor Bump podría dibujar una línea recta por los puntos de datos, intentando ajustar la línea a los puntos tanto como sea posible. Sin embargo, alguien más podría dibujar una línea diferente. Un procedimiento más adecuado es obtener la mejor línea recta usando un criterio que, para un conjunto determinado de datos, produzca siempre la misma línea recta sin impor-

tar quién esté haciendo el ajuste. Como apuntamos en el capítulo 2, un procedimiento muy conocido para determinar la mejor línea recta ajustada es el criterio de los mínimos cuadrados.

La línea que mejor se ajusta a una colección de puntos de datos X - Y es la línea que minimiza la suma de los cuadrados de las distancias entre los puntos y la línea, medidas verticalmente, es decir, en la dirección Y . Esta línea se conoce como la *línea de los mínimos cuadrados* o *línea de regresión ajustada* y su ecuación se llama *ecuación de regresión ajustada*.

La línea recta ajustada es de la forma $\hat{Y} = b_0 + b_1X$. El primer término del segundo miembro, b_0 , es la *intersección con Y* , y el segundo término, b_1 , es la *pendiente*. Recuerde que la pendiente representa la cantidad de cambio de Y cuando se incrementa X en una unidad. El objetivo inmediato es determinar los valores de b_0 y b_1 .

Recuerde del capítulo 2 (véase la ecuación 2.11) que el *método de los mínimos cuadrados* selecciona los valores de b_0 y b_1 que minimizan los errores de la suma de los cuadrados (distancias).

$$SSE = \sum (Y - \hat{Y})^2 = \sum (Y - b_0 - b_1X)^2 \quad (6.1)$$

Usando un poco de cálculo, es posible obtener expresiones algebraicas específicas para los valores de los mínimos cuadrados. En particular,

$$b_1 = \frac{n\sum XY - \sum X\sum Y}{n\sum X^2 - (\sum X)^2} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \quad (6.2)$$

$$b_0 = \frac{\sum Y}{n} - \frac{b_1\sum X}{n} = \bar{Y} - b_1\bar{X} \quad (6.3)$$

Como usted podrá adivinar, el coeficiente de la pendiente de los 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 entre sí y tienen el mismo signo.

Las diferencias entre los valores Y realmente observados y los correspondientes valores Y ajustados, las $\hat{Y}$'s, se llaman *residuos*. Los residuos son las distancias verticales (positivas o negativas) entre los puntos de datos y la línea de mínimos cuadrados.

Tenemos la identidad

$$\text{Observación} = \text{Ajuste} + \text{Residuo}$$

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 residuos representan las desviaciones de ese patrón. La separación en ajustes más residuos se aplica también a otros patrones diferentes a los de una línea recta y los utilizaremos repetidamente en capítulos posteriores.

Ejemplo 6.2

Los coeficientes de mínimos cuadrados para ajustar una línea recta a los datos del señor Bump (véase la figura 6-1) se calculan fácilmente mediante las ecuaciones 6.2 y 6.3 y la información de la tabla 6-2. Así, tenemos

$$b_1 = \frac{10(149.3) - (14.4)(112)}{10(21.56) - (14.4)^2} = \frac{-119.8}{8.24} = -14.54$$

$$b_0 = \frac{112}{10} - (-14.54)\frac{14.4}{10} = 11.2 + 14.54(1.44) = 32.14$$

La línea de mínimos cuadrados ajustada tiene la ecuación

$$\hat{Y} = b_0 + b_1X$$

$$\hat{Y} = 32.14 - 14.54X \quad (6.6)$$

Esta ecuación se llama la *ecuación de regresión ajustada*.

Ahora, el señor Bump desea interpretar los valores en esta ecuación. La intersección con Y , b_0 , es el valor de Y cuando X es igual a cero. Una interpretación estricta sugeriría que el número promedio de galones vendidos cuando $X = 0$ (es decir, cuando el precio del galón de leche es cero) es de 32,140 galones. Esta interpretación no concuerda con el sentido común, puesto que uno esperaría que se “vendiera” más leche si ésta fuera gratis. El problema ilustrado aquí implica la predicción de un valor de Y con base en un valor de X acerca del cual no se ha recolectado evidencia muestral. Es decir, ninguno de los puntos de la muestra tiene un valor X en cero o cerca de cero. En este caso, como en muchos casos de análisis de regresión, no es posible tener una interpretación útil de la intersección con Y .

En términos más generales, con frecuencia no conviene predecir valores de Y para cualquier X más allá del rango de las X recopiladas en los datos muestrales. Una función de regresión debería interpretarse como una aproximación útil al comportamiento en el mundo real *de la región en la cual existen datos*. La extrapolación más allá de los datos requiere una desmedida suposición de que la naturaleza de la relación entre X y Y no cambia.¹

El coeficiente de la pendiente, b_1 , se puede interpretar como el cambio promedio de Y que ocurre cuando X se incrementa en una unidad. En este ejemplo, Y disminuye un promedio de 14.54 (esto es, se venden 14,540 galones menos) cuando X se incrementa en 1 (es decir, cuando el costo de un galón de leche se incrementa en \$1). Cada dólar de incremento en el precio de un galón de leche reduce la cantidad comprada en un promedio de 14,540 galones. Para expresar este enunciado en unidades más significativas, la evidencia muestral indica que cada incremento de 1 centavo en el precio de un galón de leche reduce la cantidad comprada en un promedio de 145.4 galones.

La relación X - Y se puede ilustrar dibujando la línea recta que mejor se ajusta a los puntos de datos en un diagrama de dispersión. El resultado se presenta en la figura 6-2.

Observe que las distancias verticales de los puntos a la línea se representan como líneas punteadas. Si estas distancias se elevan al cuadrado y se suman, el resultado sería menor que el calculado para cualquier otra línea que pudiera dibujarse a través de los puntos.² Así, de acuerdo con el procedimiento de los mínimos cuadrados, esta línea representa el mejor ajuste posible para los puntos de los 10 datos muestrales.

Cuando los puntos en el diagrama de dispersión no se ubican exactamente sobre la línea, como en la figura 6-2, es necesario recurrir a ideas estadísticas para el estudio de la relación entre dos variables. Pensamos en los datos de un diagrama de dispersión como una muestra de observaciones sobre una relación subyacente que se mantiene en la población de valores X - Y . El modelo estadístico de la regresión lineal simple supone que para todos los valores de X , los valores observados de la variable dependiente, Y , están normalmente distribuidos

¹En el capítulo 5 encontramos una situación similar cuando analizamos la utilidad de la extrapolación de curvas de tendencia para pronosticar valores futuros de una serie de tiempo.

²Para los datos de Bump, la suma más pequeña de las distancias elevadas al cuadrado es $SSE = 59.14$.

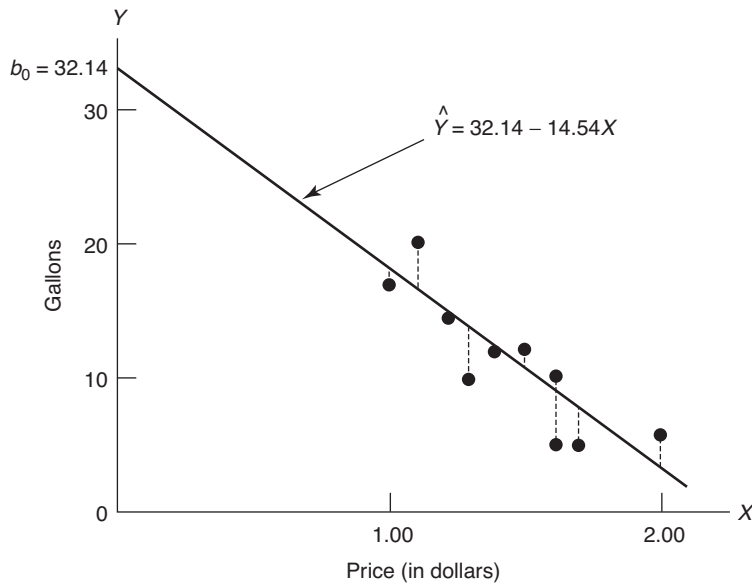


FIGURA 6-2 Línea de regresión ajustada de los datos del señor Bump para el ejemplo 6.2

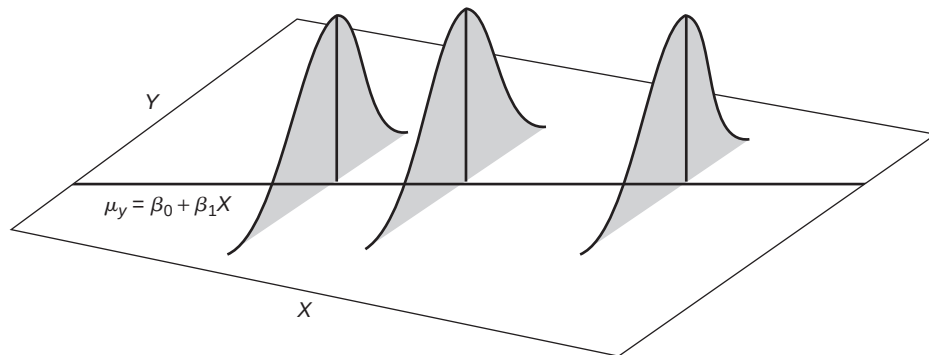


FIGURA 6-3 Modelo estadístico de la regresión lineal simple

alrededor de la media, $\mu_y = \beta_0 + \beta_1 X$, la cual depende linealmente de X . Es decir, conforme X cambia, las medias de la distribución de los valores posibles de Y permanecen a lo largo de la línea recta. Esta línea se conoce como la *línea de regresión de población*. Las Y observadas variarán alrededor de estas medias por la influencia de factores no medidos. El modelo supone que esta variación, medida por la desviación estándar, σ , es la misma para todos los valores de X . Finalmente, la desviación (distancia) entre un valor de Y y su media se conoce como *error* y se representa mediante la letra griega épsilon.

En el modelo de regresión lineal simple, la variable de respuesta o dependiente Y es la suma de esta media y la desviación aleatoria (ϵ) en relación con la media. Las desviaciones (ϵ) representan la variación de Y que se debe a otros factores no observados, los cuales impiden a los valores X - Y ubicarse exactamente sobre una línea recta en el diagrama de dispersión.

El modelo estadístico de regresión lineal simple se ilustra en la figura 6-3.

Modelo estadístico de regresión lineal

La variable de respuesta o dependiente, Y , está relacionada con la variable controlada o independiente, X , mediante

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Aquí, $\beta_0 + \beta_1 X$ es la respuesta promedio para una X dada. Se supone que las desviaciones (ε) son independientes y están normalmente distribuidas, con una media igual a 0 y una desviación estándar σ . Las constantes desconocidas son β_0 , β_1 y σ .

La línea de regresión ajustada (de la muestra), $\hat{Y} = b_0 + b_1 X$, puede considerarse como una estimación de la línea de regresión de la población, $m_y = \beta_0 + \beta_1 X$, en tanto que los residuos $e = Y - \hat{Y}$, pueden considerarse como estimaciones de los componentes del error, ε . Esto implica la siguiente correspondencia:

$$\text{Población: } Y = \beta_0 + \beta_1 X + \varepsilon$$

$$\text{Muestra: } Y = b_0 + b_1 X + e$$

La mayoría de los paquetes de software de computadora ejecutan los análisis de correlación y regresión. Las instrucciones para usar Excel en los ejemplos 6.1 y 6.2 se presentan en la sección de aplicaciones de Excel al final de este capítulo.

ERROR ESTÁNDAR DE LA ESTIMACIÓN

Una vez que se ha calculado la línea recta ajustada, el señor Bump podría estar interesado en medir el grado en el que los puntos de datos muestrales se dispersan alrededor de la función de regresión ajustada. Es posible desarrollar una medida de dispersión análoga a la desviación estándar de la muestra. Esta medida, llamada *error estándar de la estimación*, mide la dispersión de los puntos de datos en la dirección Y alrededor de la línea ajustada. El error estándar de la estimación se denota por $s_{y \cdot x}$ y está dado 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 reales de Y difieren de los valores estimados o $\hat{Y}$. Para muestras relativamente grandes, esperaríamos que alrededor del 67% de las diferencias $Y - \hat{Y}$ estuvieran a una distancia inferior de $s_{y \cdot x}$ del 0, y que alrededor del 95% de estas diferencias estuvieran dentro de $2s_{y \cdot x}$ del 0.

El error estándar de la estimación es similar a la desviación estándar de la muestra presentada en el capítulo 2. Puede utilizarse para calcular la desviación estándar de la población. De hecho, $s_{y \cdot x}$ estima la desviación estándar, σ , del término de error, ε , en el modelo estadístico de regresión lineal simple. De manera equivalente, $s_{y \cdot x}$ estima la desviación estándar común, σ , de la distribución normal de los valores Y alrededor de la línea de regresión de la población, $\mu_y = \beta_0 + \beta_1 X$, para cada X (véase la figura 6-3).

Un análisis de regresión con un error estándar pequeño de la estimación significa que todos los puntos de datos se ubican muy cerca de la línea de regresión ajustada.³ Si el error estándar de la estimación es grande, los puntos de datos están considerablemente dispersos alrededor de la línea ajustada.

Para efectos de cálculo, la ecuación 6.7 se convierte en

$$s_{y \cdot x} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum XY}{n - 2}} \quad (6.8)$$

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 de $\hat{Y}$ en el rango 3-18 (véase la figura 6-2), $s_{y \cdot x}$ es moderadamente grande e indica que una cantidad apreciable de la variación en Y (galones vendidos) no se explica por X (precio). Investigaremos más adelante esta afirmación, en la sección sobre el coeficiente de determinación.

PRONÓSTICO DE Y

Luego, la línea de regresión ajustada puede usarse para estimar el valor de Y para un valor dado de X . Si queremos obtener un *pronóstico puntual* o un pronóstico para un valor dado de X , simplemente evaluamos la función de regresión estimada para X .

Ejemplo 6.3

Suponga que el señor Bump desea pronosticar la cantidad de leche vendida si el precio se fijara en \$1.63. De acuerdo con la ecuación 6.6, el pronóstico es

$$\begin{aligned} \hat{Y} &= 32.14 - 14.54X \\ \hat{Y} &= 32.14 - 14.54(1.63) = 8.440 \end{aligned}$$

o 8,440 galones. Observe 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 se da cuenta de que es improbable que los valores reales de Y correspondientes a las posiciones de las X se ubiquen exactamente sobre la línea de regresión. De hecho, estarán distribuidos a uno y otro lado de la línea conforme a la medición $s_{y \cdot x}$. Más aún, la línea de regresión (ajustada) de la muestra es una estimación de la línea de regresión de la población basada en una muestra de 10 puntos de datos. Otras muestras aleatorias de 10 producirían diferentes líneas de regresión ajustada, en forma similar al caso en el que muchas muestras extraídas de la misma población tienen diferentes medias muestrales.

Entonces, existen dos fuentes de incertidumbre asociadas con un pronóstico puntual generado por una ecuación de regresión ajustada:

1. La incertidumbre que se debe a la dispersión de los puntos de datos alrededor de la línea de regresión de la muestra.
2. La incertidumbre que se debe a la dispersión de la línea de la regresión de la muestra alrededor de la línea de regresión de la población.

³Si todos los puntos de datos se localizan exactamente sobre la línea ajustada, $Y = \hat{Y}$ para toda X y $s_{y \cdot x} = 0$.

Es posible construir 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 pronosticada en relación con el valor real de Y para un valor dado de X .

El error estándar del pronóstico es

$$s_f = \sqrt{s_{y \cdot x}^2 + s_{y \cdot x}^2 \left(\frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2} \right)}$$

$$s_f = s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2}} \quad (6.9)$$

El primer término debajo del primer radical en la ecuación 6.9, $s_{y \cdot x}^2$, mide la dispersión de los puntos de datos alrededor de la línea de regresión de la muestra (primera fuente de incertidumbre). El segundo término debajo del radical mide la dispersión de la línea de regresión de la muestra alrededor de la línea de regresión de la población (segunda fuente de incertidumbre). Advierta que el error estándar del pronóstico depende de X , el valor de X para el cual se desea un pronóstico de Y . También, observe que s_f es el más pequeño cuando $X = \bar{X}$, puesto que el numerador del tercer término debajo del radical en la ecuación 6.9 (abajo) será $(\bar{X} - \bar{X})^2 = 0$.⁴ Si todo permanece igual, cuanto más lejos está X de $\bar{X}$, más grande es el error estándar del pronóstico.

Si el modelo estadístico de regresión lineal simple es el adecuado, un intervalo de predicción de Y está dado por

$$\hat{Y} \pm ts_f \quad (6.10)$$

donde t es un punto porcentual de la distribución t de Student con $df = n - 2$. Si el tamaño de la muestra es grande ($n \geq 30$), el punto porcentual t puede sustituirse por el correspondiente punto porcentual Z de la distribución normal estándar. Por ejemplo, en esencia, un intervalo de predicción del 95% para una muestra grande para Y es

$$\hat{Y} \pm 2s_f \quad (6.11)$$

Ejemplo 6.4

Gráficamente, en el caso que analiza el señor Bump, el intervalo de predicción del 95% para Y con varios valores de X se vería como el de 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 para $X = 1.63$ es

$$s_f = 2.72 \sqrt{1 + \frac{1}{10} + \frac{(1.63 - 1.44)^2}{.824}} = 2.72(1.069) = 2.91$$

A partir del ejemplo 6.3, $\hat{Y} = 8.440$ cuando $X = 1.63$, y usando la ecuación 6.10, un intervalo de predicción del 95% para Y es

$$\hat{Y} \pm ts_f = 8.44 \pm 2.306(2.91) = 8.44 \pm 6.71$$

⁴Para la opción $X = \bar{X}$, el pronóstico de Y es $\hat{Y} = \bar{Y}$.

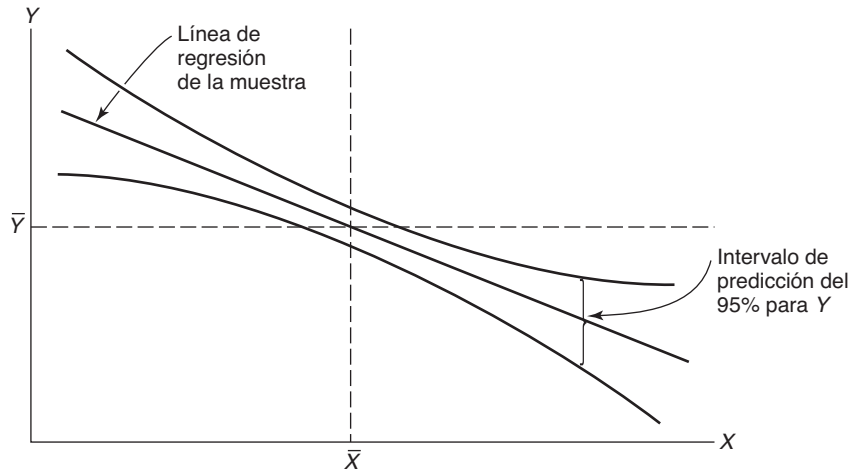


FIGURA 6-4 Intervalo de predicción de los datos del señor Bump para el ejemplo 6.4

TABLA 6-3 Cálculo de $\Sigma(X - \bar{X})^2$ 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
$\Sigma(X - \bar{X})^2 = .8240$	

o (1.73, 15.15); es decir, de 1,730 a 15,150 galones. Aquí, $2.306 = t_{0.025}$ es el punto que deja un área de 2.5% a la derecha de la distribución t con 8 grados de libertad $df = 8$.

El intervalo de predicción es muy extenso —tanto, que prácticamente carece de valor para pronosticar Y — a causa del reducido tamaño de la muestra y del valor relativamente grande de s_f . La cantidad de incertidumbre reflejada por el intervalo de predicción tan grande no es evidente en el pronóstico puntual calculado a partir de la función de regresión ajustada. La ventaja principal de la estimación del intervalo es que da cuenta explícitamente de la incertidumbre asociada con el pronóstico.

En general, es peligroso utilizar la función de regresión ajustada para predecir los valores de Y más allá del rango de los datos disponibles. El señor Bump tiene justificación al tratar de pronosticar Y cuando $X = 1.63$ porque algunos de los valores originales de X están cercanos a 1.63. Por otro lado, tal vez no sería conveniente pronosticar Y cuando $X = 3.00$. No se han recopilado valores de X tan grandes y, por esa razón, cualquier pronóstico que implique valores de X como éstos sería muy sospechoso.⁵ Para calcular la cantidad de leche vendida cuando el precio por galón es de \$3.00, el señor Bump tiene que suponer que el modelo lineal sigue siendo válido. Tal vez tenga una buena razón para hacer esta suposición, pero no tiene evidencia directa para apoyarla.

⁵Más aún, el error estándar del pronóstico sería grande, puesto que la cifra $(X - \bar{X})^2$ sería relativamente grande.

Resulta útil terminar esta sección revisando los supuestos subyacentes del modelo estadístico de regresión lineal.

1. *Para un valor dado de X , la población de los valores de Y está normalmente distribuida alrededor de la línea de regresión de la población.* Esta condición se evidencia en la figura 6-3. En la práctica, se obtienen resultados razonablemente exactos siempre que los valores de Y estén distribuidos aproximadamente de manera normal.
2. *La dispersión de los puntos de datos de la población 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 más grande o más pequeña conforme los valores de los puntos de datos de X se incrementan. Una violación de este supuesto se llama *heteroscedasticidad*; un ejemplo de esta condición y su solución se encuentra en el capítulo 8.
3. *Los términos de error (ϵ) son independientes uno del otro.* Esta suposición implica una muestra aleatoria de datos X - Y . Cuando los puntos de datos X - Y se registran en el tiempo, este supuesto se viola a menudo. En vez de ser independientes, las observaciones consecutivas están *serialmente correlacionadas*. Los métodos para manejar el problema de la correlación serial se estudiarán en el capítulo 8.
4. *Existe una relación lineal entre X y Y en la población.* Existen extensiones de regresiones lineales simples para manejar las relaciones X - Y que no son lineales; algunas de ellas se estudiarán más adelante en este capítulo.

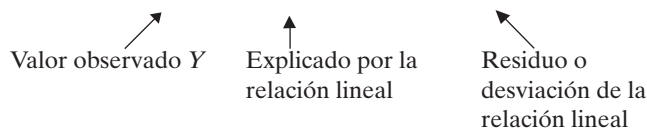
DESCOMPOSICIÓN DE LA VARIANZA

De acuerdo con la ecuación 6.5

$$Y = \hat{Y} + (Y - \hat{Y})$$

o

$$Y = b_0 + b_1X + (Y - b_0 - b_1X)$$



En una situación ideal, en la cual todos los puntos se ubican exactamente sobre la línea ajustada, todos los residuos son cero y los valores de Y están totalmente explicados por la relación lineal con X .

Al restar $\bar{Y}$ de ambos lados de la expresión anterior,

$$(Y - \bar{Y}) = (\hat{Y} - \bar{Y}) + (Y - \hat{Y})$$

Con un poco de álgebra es posible demostrar que la suma de los cuadrados da por 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 cuadrados* (del inglés, *sum of squares*) y T , R y E significan *total*, *regresión* y *error*, respectivamente. Estas sumas de cuadrados tienen 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 están relacionados por

$$n - 1 = 1 + (n - 2) \quad (6.13)$$

Si no existe una relación lineal, Y no depende de X , y la variación de Y queda descrita por la varianza de la muestra:

$$s_y^2 = \frac{1}{n - 1} \sum (Y - \bar{Y})^2$$

Si, por otro lado, Y está relacionada con X , algunas de las diferencias en los valores de Y se deben a esta relación.

La regresión de la suma de los cuadrados, SSR , mide la parte de la variación de Y que se explica por la relación lineal. La suma de los errores al cuadrado, SSE , es la variación restante de Y , o la variación de Y que no se explica por la relación lineal.

Descomposición de la variabilidad

$$\begin{array}{ccccc} SST & = & SSR & + & SSE \\ \text{Variabilidad} & & \text{Variabilidad} & & \text{Residuo o} \\ \text{total de } Y & & \text{explicada por} & & \text{variabilidad} \\ & & \text{la relación lineal} & & \text{no explicada} \end{array}$$

La sumas de los cuadrados asociadas con la descomposición de la variabilidad de Y y sus correspondientes grados de libertad se pueden arreglar como se muestra en la tabla 6-4, que constituye un ejemplo de una *tabla ANOVA* (*analysis of variance*, análisis de varianza).

La columna final de 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. De forma similar, el error cuadrático 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 la regresión de línea recta

<i>Fuente</i>	<i>Suma de 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 Residuos de los datos del señor Bump con el factor de predicción $\bar{Y}$ para el ejemplo 6.5

<i>Y real</i>	<i>Y pronosticada ($\bar{Y}$)</i>	<i>Residuo ($Y - \bar{Y}$)</i>	$(Y - \bar{Y})^2$
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	233.60
			0.0

Observe que en la ecuación 6.7 el error cuadrático medio es

$$MSE = \frac{SSE}{n - 2} = \frac{\sum (Y - \hat{Y})^2}{n - 2} = s_{y \cdot x}^2$$

el cuadrado del error estándar de la estimación. Las razones de los cuadrados medios se utilizarán más adelante con otra finalidad en este capítulo.

Ejemplo 6.5

El análisis del señor Bump se inició con el conocimiento tan sólo de 10 cantidades del volumen de ventas semanales (la variable Y). Si no hubiera más información disponible, el señor Bump podría utilizar el promedio muestral, $\bar{Y} = 11.2$, como un factor de predicción de los galones de leche vendidos cada semana. Los errores, o residuos, asociados con este pronóstico son $Y - \bar{Y}$, y la suma de los errores cuadráticos es $\sum (Y - \bar{Y})^2$. Observe que esta última cantidad, $\sum (Y - \bar{Y})^2$, es exactamente SST , la suma total de cuadrados que se introdujo en la ecuación 6.12. Por lo tanto, SST mide la variabilidad de Y alrededor de un factor de predicción que usa sólo los valores Y en sus cálculos.⁶ Los pronósticos ($\bar{Y}$), los residuos ($Y - \bar{Y}$), y la suma total de cuadrados ($SST = \sum (Y - \bar{Y})^2$) se muestran en la tabla 6-5.⁷

El señor Bump también tiene información acerca de una variable X , el precio por galón de leche, que está relacionada con Y , el volumen semanal de leche vendida. (Recuerde el ejemplo 6.1, donde $r = -.86$). Con esta variable adicional, él espera explicar parte de la variación (diferencias) de los valores de Y más allá de lo que explica el factor de predicción, $\bar{Y}$.

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 construir una tabla similar a la tabla 6-5, si se usa $\hat{Y}$ para predecir Y . El resultado es la tabla 6-6.⁸

Una comparación de las tablas 6-5 y 6-6 indica que el uso de $\hat{Y}$ como factor de predicción de Y tiene como resultado residuos generalmente más pequeños (en valor absoluto) y una suma de cuadrados residual (error) considerablemente menor que cuando se usa $\bar{Y}$ como factor de predicción. El uso de la variable relacionada X reduce los errores de predicción o pronóstico. Es decir, el conocimiento de X ayuda a explicar aún más las diferencias en las Y . ¿Qué tanto ayuda X ? La descomposición de la variabilidad ofrece una respuesta a esta pregunta.

⁶Si el análisis se detuviera en este punto, la variabilidad de Y se mediría por la varianza muestral, $s_y^2 = \sum (Y - \bar{Y})^2 / (n - 1)$, en vez de $SST = \sum (Y - \bar{Y})^2$. La varianza muestral es la medida habitual de variabilidad para mediciones de una variable individual.

⁷Los residuos, $Y - \bar{Y}$, siempre suman cero porque el promedio $\bar{Y}$ es el centro matemático de los valores de Y .

⁸Si se incluye un término de intersección en la ecuación de regresión, la suma de los residuos $\sum (Y - \hat{Y})$ siempre es cero.

TABLA 6-6 Residuos de los datos del señor Bump con el factor de predicción $\hat{Y}$ para el ejemplo 6.5

X	Y	$Y(\hat{Y})$ pronosticada usando $\hat{Y} = 32.14 - 14.54X$	Residual ($Y - \hat{Y}$)	($Y - \hat{Y}$) ²
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

De acuerdo con las tablas 6-5 y 6-6 y la ecuación 6-12,

$$SST = \Sigma(Y - \bar{Y})^2 = 233.60$$

$$SSE = \Sigma(Y - \hat{Y})^2 = 59.41$$

y, en consecuencia,

$$SSR = \Sigma(\hat{Y} - \bar{Y})^2 = 233.60 - 59.41 = 174.19$$

La descomposición de la variabilidad es

SST	=	SSR	+	SSE
233.60	=	174.19	+	59.41
Variación total		Variación explicada		Variación no explicada

De la variabilidad restante después del pronóstico de Y con $\bar{Y}$, el señor Bump ve que una proporción

$$\frac{SSR}{SST} = \frac{174.19}{233.60} = .75$$

de ella ha sido explicada por la relación de Y con X . Una proporción, $1 - .75 = .25$, de la variación de Y con respecto a $\bar{Y}$ permanece sin explicación. Desde esta perspectiva, el conocimiento de la variable relacionada X da como resultado mejores predicciones de Y de las que pueden obtenerse a partir de $\bar{Y}$, una cantidad que no depende de X .

La descomposición de la variabilidad de los datos del señor Bump puede obtenerse a partir del análisis de una tabla de varianza, como se demuestra en el siguiente ejemplo.

Ejemplo 6.6

El señor Bump construye una tabla ANOVA para sus datos usando el formato de la tabla 6-4 y los cálculos de las sumas de los cuadrados del ejemplo 6.5, como se indica en la tabla 6-7.

La descomposición de la variabilidad se presenta claramente en la columna de la suma de cuadrados. Observe que, dentro del redondeo del error, $MSE = 7.43 = (2.72)^2 = s_{y \cdot x}^2$.

TABLA 6-7 Tabla ANOVA de los datos del señor Bump para el ejemplo 6.6

<i>Fuente</i>	<i>Suma de 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 dada por la ecuación 6.12 y se muestra gráficamente para la línea 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 $Y - \bar{Y}$ sencillamente reflejara la variación aleatoria. Sin embargo, Y depende de X de una 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 correlación negativa bastante fuerte ($r = -.86$). De la distancia vertical total $Y - \bar{Y}$, la cantidad $\hat{Y} - \bar{Y}$ es “explicada”, por lo tanto, por la variación de X , mientras que la distancia vertical restante, $Y - \hat{Y}$, “no se explica” por el movimiento de X .

Como se indicó en la sección anterior, SST mide la variación total alrededor de $\bar{Y}$, y la parte del total que se explica por la variación de X es SSR . La variación restante o no explicada es el SSE . La razón entre la variación explicada y la variación total se llama *coeficiente de determinación* de la muestra y se representa por r^2 .

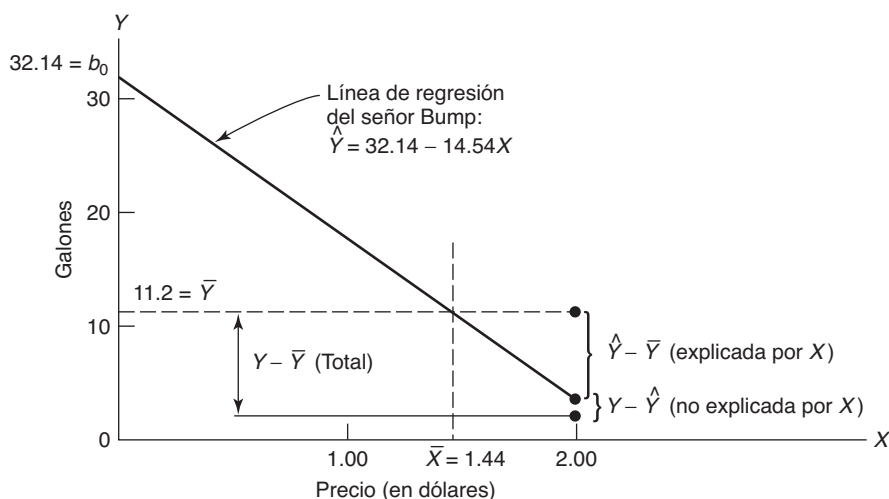


FIGURA 6-5 Variación explicada y variación no explicada en los datos del señor Bump

$$\begin{aligned}
 r^2 &= \frac{\text{Variación explicada}}{\text{Variación total}} = \frac{SSR}{SST} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} \\
 &= 1 - \frac{\text{Variación no explicada}}{\text{Variación total}} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} \quad (6.14)
 \end{aligned}$$

El *coeficiente de determinación* mide el porcentaje de variabilidad de Y que se puede explicar a través del conocimiento de la variabilidad (diferencias) de la variable independiente X .

Ejemplo 6.7

El coeficiente de determinación, r^2 , para los datos del señor Bump se calculó en el ejemplo 6.5, aunque no se le llamó así. El coeficiente de determinación también se encuentra fácilmente disponible en la tabla ANOVA, es decir, la tabla 6-7, del ejemplo 6.6. Recuerde que

$$SST = \sum(Y - \bar{Y})^2 = 233.60$$

$$SSR = \sum(\hat{Y} - \bar{Y})^2 = 174.19$$

$$SSE = \sum(Y - \hat{Y})^2 = 59.41$$

y

$$r^2 = \frac{174.19}{233.60} = .746$$

De forma alternativa, r^2 puede calcularse como

$$r^2 = 1 - \frac{59.41}{233.60} = 1 - .254 = .746$$

Alrededor del 75% de la variabilidad en galones de leche vendida (Y) puede explicarse por las diferencias en el precio por galón (X). Alrededor del 25% de la variabilidad en la cantidad de leche vendida *no puede* explicarse por el precio. Esta porción de la variabilidad debe explicarse por factores que no se consideraron en este análisis de regresión (por ejemplo, la cantidad de publicidad, la disponibilidad de productos sustitutos o la calidad de la leche).

Para la regresión lineal, r^2 es el cuadrado de r , el coeficiente de correlación.

$$\text{Coeficiente de determinación} = (\text{Coeficiente de correlación})^2$$

$$r^2 = (r)^2$$

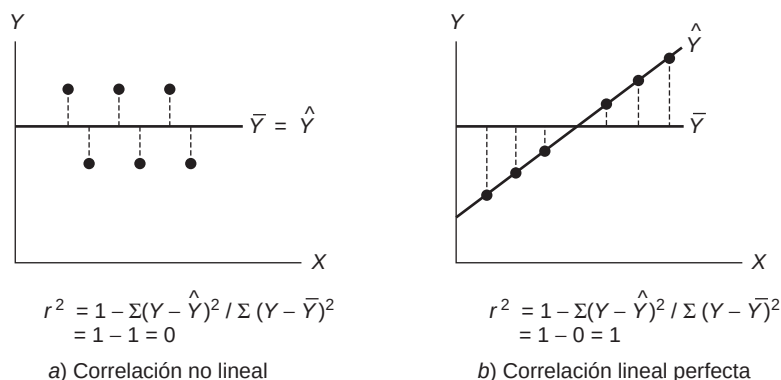
De manera que, para los datos del señor Bump e ignorando el redondeo del error,

$$.746 = (-.863)^2$$

¿Por qué es necesario identificar específicamente tanto r como r^2 en un análisis de regresión? La respuesta es que expresan información diferente.

El coeficiente de correlación revela la fortaleza y dirección de la relación lineal. En el caso de los datos que recopiló el señor Bump, existe una relación negativa ($r = -.86$). En otros casos, el valor de r podría estar indicando una relación positiva. Como veremos en el siguiente capítulo, es útil identificar la naturaleza de las relaciones que existen entre ciertos pares de variables cuando se está manejando un gran conjunto de variables. Observe que, cuando el coeficiente de correlación está elevado al cuadrado, el valor siempre es positivo, de manera que la naturaleza de la relación se pierde.

El coeficiente de determinación, r^2 , mide la fuerza de la relación entre Y y X de un modo diferente a como lo hace r . El valor de r^2 mide el *porcentaje* de variabilidad de Y que se explica

FIGURA 6-6 Valores extremos de r^2

por las diferencias de X . Esta útil interpretación se puede generalizar a las relaciones entre Y y más de una X .

La figura 6-6 ilustra los dos casos extremos para r^2 , $r^2 = 0$ y $r^2 = 1$. En el primer caso, *nada* de variabilidad de Y se explica por X : el diagrama de dispersión sugiere que no existe una relación lineal entre X y Y . Cuando $r^2 = 1$, *toda* la variabilidad de Y se explica cuando se conoce X : todos los puntos de datos muestrales se ubican sobre la línea de regresión ajustada.

PRUEBA DE HIPÓTESIS

La línea de regresión ajustada se genera por los valores X - Y de la *muestra*. El modelo estadístico de regresión lineal simple sugiere que la relación lineal entre Y y X se mantiene para *todas* las opciones de pares X - Y . Es decir, existe una relación verdadera entre X y Y de la forma $\mu_y = \beta_0 + \beta_1 X$. Dada la evidencia muestral, ¿podemos concluir que la relación verdadera se mantiene para todas las X y Y ?

Considere la hipótesis

$$H_0 : \beta_1 = 0$$

donde β_1 es la pendiente de la línea de regresión de la población. Advierta que, si esta hipótesis es verdadera, no existe relación entre Y y X en la población. Por otro lado, si no se rechaza H_0 , significa que a pesar del hecho de que la muestra ha producido una línea ajustada con un valor diferente de cero para b_1 , debemos concluir que no hay suficiente evidencia para indicar que Y esté relacionada con X . Es decir, no podemos excluir la posibilidad de que la línea de regresión de la población sea plana (horizontal).⁹

¿Cómo es posible que β_1 sea cero mientras b_1 es diferente de cero? Considere la figura 6-7 donde se presenta una población de puntos de datos de los cuales se selecciona una muestra de cinco datos (los puntos de datos muestrales se indican mediante $\times$). Como lo sugiere el diagrama de dispersión, si se seleccionan suficientes puntos de datos muestrales, se hará obvio que la población de puntos de datos tiene una línea de regresión con pendiente cero. Sin embargo, los cinco puntos de datos seleccionados de manera aleatoria se ubican bastante cerca de una línea de regresión con tendencia ascendente. A partir de esta evidencia se podría concluir erróneamente que X y Y están relacionadas linealmente de manera positiva. Sin embargo, si se prueba la hipótesis $\beta_1 = 0$ con los datos muestrales, quizás no pueda rechazarla el encargado de efectuar el pronóstico.

⁹Una línea de regresión plana de la población (esto es, $\beta_1 = 0$) también es equivalente al enunciado $H_0: \rho = 0$, donde ρ es el coeficiente de correlación de la población.

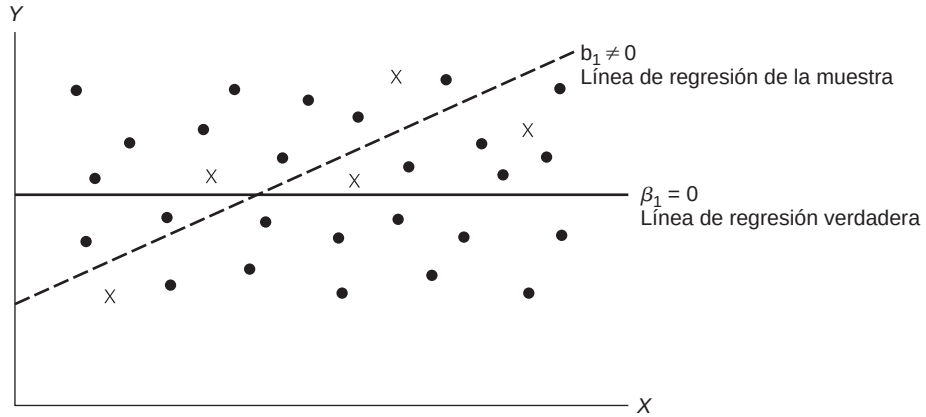


FIGURA 6-7 Población y puntos de datos muestrales

Si $H_0: \beta_1 = 0$ es verdadera, el estadístico de prueba t con valor de $t = \frac{b_1}{s_{b_1}}$ tiene una distribución t con $df = n - 2$, donde S_{b_1} es la desviación estándar estimada (o error estándar) de b_1 , dado por $s_{b_1} = s_{y \cdot x} / \sqrt{\sum (X - \bar{X})^2}$.

Este resultado ofrece un modo de probar $H_0: \beta_1 = 0$, como se ilustra en el siguiente ejemplo.

Ejemplo 6.8

Al señor Bump le gustaría probar

$$H_0: \beta_1 = 0 \text{ (no hay relación lineal)}$$

contra

$$H_1: \beta_1 \neq 0 \text{ (una relación lineal con pendiente diferente de cero)}$$

Él calcula¹⁰

$$s_{b_1} = s_{y \cdot x} / \sqrt{\sum (X - \bar{X})^2} = 2.72 / \sqrt{.824} = 3.00$$

y formula la razón (estadístico de prueba)

$$t = \frac{b_1}{s_{b_1}} = \frac{-14.54}{3.00} = -4.8$$

¿El valor $t = -4.8$ es un resultado inusual si H_0 es verdadera? El señor Bump verifica t en la tabla para $n - 2 = 8$ grados de libertad y obtiene

$$t_{.005} = 3.355$$

$$t_{.025} = 2.306$$

Puesto que $|t| = 4.8 > 3.355$, el señor Bump rechaza H_0 para el nivel de significancia del 1%. Concluye que su regresión es significativa, puesto que $t = -4.8$ es sumamente improbable si H_0 es verdadera. Un valor t de esta magnitud se presentará menos de una vez en 100 si no existiera una relación lineal entre Y y X .

¹⁰Los valores usados en los siguientes cálculos se obtuvieron con anterioridad en este capítulo.

Para tamaños muestrales muy grandes, es posible rechazar H_0 y concluir que existe una relación lineal entre X y Y aun cuando r^2 sea pequeña, digamos, del 10%. De forma similar, para muestras pequeñas y una r^2 muy grande —digamos, del 95%— es posible concluir que la regresión es significativa. Una r^2 pequeña significa que es improbable que la ecuación de regresión ajustada tenga mucho poder de predicción. Por otro lado, una r^2 grande con un tamaño muestral muy pequeño podría dejar incómodo al analista y requerir mayor evidencia muestral antes de que la función ajustada se use para pronosticar. Puede haber una diferencia entre la significancia estadística y la significancia práctica. A veces, es necesario un buen juicio aparejado con conocimiento del tema para determinar si una función de regresión ajustada es una herramienta útil para pronosticar.

Una prueba alternativa de $H_0: \beta_1 = 0$ está disponible en la tabla ANOVA.

Si los supuestos del modelo estadístico de la regresión lineal son adecuados y si la hipótesis nula $H_0: \beta_1 = 0$ es verdadera, la razón

$$F = \frac{\text{Cuadrado medio de la regresión}}{\text{Error cuadrático medio}} = \frac{MSR}{MSE} \quad (6.15)$$

tiene una distribución F con $df = 1, n - 2$. Cuando H_0 es verdadera, tanto el MSR como el MSE son estimadores de σ^2 , la varianza del término de error (ε) en el modelo estadístico de regresión de línea recta. Por otro lado, si $H_0: \beta_1 \neq 0$ es verdadera, el numerador F en la razón tiende a ser mayor que el denominador. De esta forma, razones grandes de F son congruentes con la hipótesis alternativa.

En el modelo de regresión de línea recta, la prueba de la hipótesis $H_0: \beta_1 = 0$ contra $H_0: \beta_1 \neq 0$ puede apoyarse en la razón $F = \frac{MSR}{MSE}$ con $df = 1, n - 2$. En el nivel α , la región de rechazo es $F > F_\alpha$.

Como se discutirá en el próximo capítulo, la prueba F se puede extender para verificar la significancia de los 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. En esta tabla,

$$F = \frac{MSR}{MSE} = \frac{174.19}{7.43} = 23.4$$

y con $\delta_1 = 1$ y $\delta_2 = 8$ grados de libertad

$$F_{.05} = 5.32$$

$$F_{.01} = 11.26$$

Puesto que $F = 23.4 > 11.26$, $H_0: \beta_1 = 0$ se rechaza con un nivel del 1%. La regresión es significativa.

No es accidental que los resultados de la prueba F sean congruentes con los resultados de la prueba t del ejemplo 6.8. De hecho, dentro del error de redondeo,

$$F = 23.4 = (-4.8)^2 = t^2$$

Más aún,

$$F_{.01} = 11.26 = (3.355)^2 = (t_{.005})^2$$

Así, para un nivel de significancia dado, la prueba t rechaza $H_0: \beta_1 = 0$ siempre que la prueba F la rechace y viceversa. Esta relación entre las pruebas t y F se cumple sólo para el modelo de regresión de línea recta.

El estadístico F también puede expresarse en términos del coeficiente de determinación, r^2 .

$$F = \frac{r^2(n-2)}{1-r^2} \quad (6.16)$$

Si todo lo demás permanece igual, el valor de F se incrementa conforme r^2 aumenta. Cuanta mayor sea la cantidad de variación de Y que tenga explicación mediante la función de regresión ajustada, más probabilidad hay de que la prueba F considere a la regresión como significativa.

ANÁLISIS DE RESIDUOS

El ajuste de un modelo por los mínimos cuadrados, la construcción de intervalos de predicción y la prueba de hipótesis no completan un estudio de regresión. Estos pasos son sólo la mitad de la historia: las inferencias que pueden hacerse *cuando el modelo supuesto es adecuado*. En la mayoría de los estudios no resulta obvio cuál es el modelo correcto.

Las inferencias suelen ser muy engañosas si los supuestos realizados en la formulación del modelo son extremadamente incompatibles con los datos. Es esencial verificar con cuidado los datos para obtener indicios de cualquier violación de los supuestos. Recuerde que los supuestos del modelo de regresión de línea recta son los siguientes:

1. La relación entre las variables es lineal.
2. Los errores son independientes.
3. Los errores tienen una varianza constante.
4. Los errores están distribuidos normalmente.

La información sobre la variación que no puede explicarse mediante la función de regresión ajustada está contenida en los residuos, $e = Y - \hat{Y}$. Para calificar los méritos de un modelo tentativo, podemos examinar los residuos graficándolos de varias maneras.

1. Grafique un histograma de los residuos.
2. Grafique los residuos contra los valores ajustados.
3. Grafique los residuos contra la variable explicativa.
4. Grafique los residuos en el tiempo si los datos son cronológicos.

Un histograma de los residuos permite verificar el supuesto de normalidad. Comúnmente, las desviaciones moderadas de una curva en forma de campana no afectan las conclusiones de las pruebas o los intervalos de predicción basados en la distribución t , particularmente si el conjunto de datos es grande. Por lo regular, una violación del supuesto de normalidad no es tan seria como una violación de cualquier otro de los supuestos.

Si una gráfica de los residuos contra los valores ajustados indica que la naturaleza general de la relación entre Y y X forma una curva en vez de una línea recta, una transformación conveniente de los datos puede reducir una relación no lineal a una que sea aproximadamente lineal. Examinaremos las transformaciones de las variables en una sección posterior de este capítulo.

Una transformación también ayuda a estabilizar la varianza. La figura 6-8 presenta una gráfica residual que indica que la dispersión de los residuos se incrementa conforme la magnitud de los valores ajustados aumenta. Es decir, la variabilidad de los puntos de datos alrededor de la línea de los mínimos cuadrados es más grande para valores ajustados de Y grandes que para valores pequeños. Esto implica que tal vez no se cumpla el supuesto de la varianza constante. En esta situación, relacionar el logaritmo de Y con X puede generar una relación residual que sea más congruente con una varianza constante.

El supuesto de independencia es el más importante. La falta de independencia podría distorsionar en forma drástica las conclusiones obtenidas de las pruebas t . El supuesto de independencia es particularmente riesgoso en los datos de las series de tiempo, que son frecuentes en problemas de pronósticos económicos y de negocios.

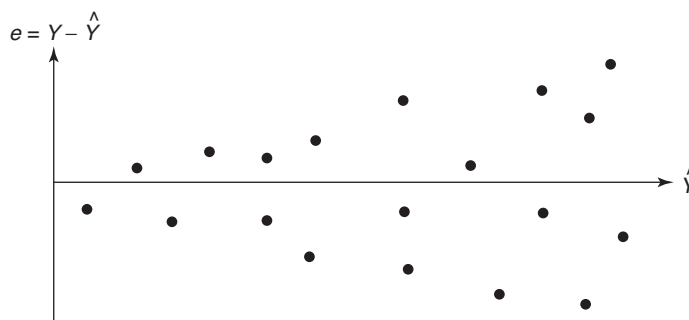


FIGURA 6-8 Una gráfica residual con dispersión creciente

Para los residuos de las series de tiempo —es decir, para los residuos que se producen al usar métodos de regresión con datos ordenados en el tiempo— la independencia puede verificarse con una gráfica de los residuos en el tiempo. No debe haber patrones sistemáticos, tales como una sucesión de valores altos seguida de una sucesión de valores bajos. Además, las autocorrelaciones muestrales de los residuos

$$r_k(e) = \frac{\sum_{t=k+1}^n e_t e_{t-k}}{\sum_{t=1}^n e_t^2} \quad k = 1, 2, \dots, K \quad (6.17)$$

donde n es el número de residuos y K comúnmente es $n/4$, deberían ser todas pequeñas. En particular, la independencia existe si los coeficientes de autocorrelación residual se encuentran todos en el intervalo $0 \pm 2/\sqrt{n}$ para todos los retrasos de tiempo k .

Ejemplo 6.10

Usando los valores ajustados y los residuos presentados en la tabla 6-6, el señor Bump construyó un histograma de los residuos y una gráfica de los residuos contra los valores ajustados. Éstos y otros resultados se presentan en la figura 6-9.

El histograma está centrado en cero y, si bien es simétrico, no parece tener forma de campana. Sin embargo, con sólo 10 observaciones, un histograma como el ilustrado en la figura 6-9 no es raro para datos con distribución normal. El supuesto de normalidad parece razonable.

Los puntos en la gráfica de probabilidad normal de la figura 6-9 se ubican muy cerca de la línea recta. Como señalamos en el capítulo 2, este comportamiento como línea recta sugiere un buen ajuste entre los datos (en este caso, los residuos) y una distribución normal. La gráfica de probabilidad normal sugiere que no hay razón para dudar del supuesto de normalidad.

La segunda gráfica de la primera fila en la figura 6-9 también se ve bien. Cuando se grafican los residuos contra los valores ajustados, la dispersión alrededor del cero en la dirección vertical debe ser aproximadamente igual para todos los valores a lo largo del eje horizontal. Es decir, las magnitudes de los residuos para valores ajustados pequeños deben ser aproximadamente iguales que las magnitudes de los residuos para valores ajustados intermedios y aproximadamente iguales que las magnitudes de los residuos para valores ajustados grandes. Este comportamiento ideal sugiere dos cosas: **1.** La relación subyacente entre Y y X es lineal, y **2.** la variabilidad del error es constante (las Y para diferentes valores de X tienen la misma dispersión alrededor de la línea de regresión).

El señor Bump está contento de que la gráfica de los residuos contra los valores ajustados no esté “arqueada” —por ejemplo, con una sucesión de residuos positivos seguida de una sucesión de residuos negativos, seguida, a la vez, de una sucesión de residuos positivos. Este comportamiento

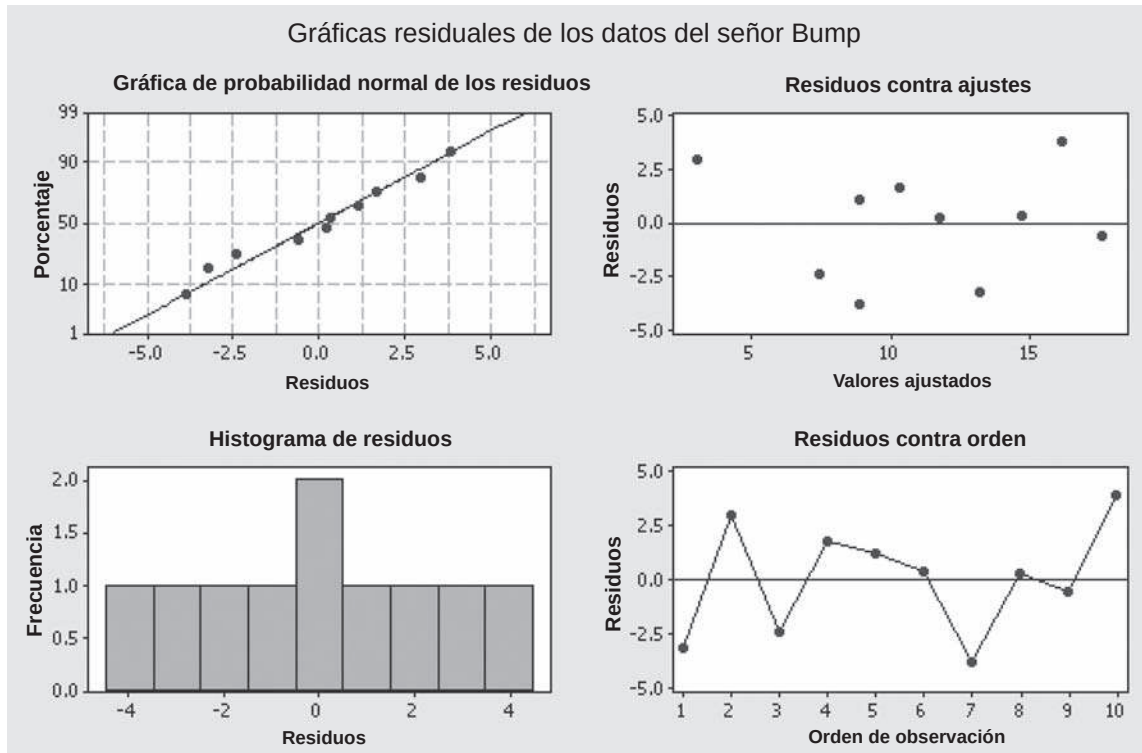


FIGURA 6-9 Gráficas residuales de los datos del señor Bump para el ejemplo 6.10

sugeriría una relación no lineal entre Y y X . Él también está complacido porque su gráfica no tiene la forma cónica como la de la gráfica de la figura 6-8, lo cual indica variabilidad inconstante (creciente).

Si bien las Y representan ventas semanales de leche, las semanas fueron seleccionadas al azar y no estaban ordenadas en el tiempo. Por consiguiente, la graficación de los residuos en el tiempo o el cálculo de las autocorrelaciones residuales no eran apropiados.

El señor Bump está satisfecho con su análisis residual. Considera que su modelo de regresión de línea recta describe adecuadamente la relación entre el volumen de las ventas semanales y el precio.

RESULTADO DE COMPUTADORA

El problema del análisis de regresión del señor Bump (con los datos de la tabla 6-1) se corre en Minitab (para instrucciones específicas, véase la sección de aplicaciones de Minitab al final del capítulo), y la salida se presenta en la tabla 6-8.

Para explicar la terminología usada en el resultado de la computadora, se presentan las definiciones y los cálculos en la siguiente lista. Estas definiciones y cálculos se teclearon para obtener la tabla 6-8.

1. Correlación = $-.86$. El coeficiente de correlación (r) de la muestra indica la relación entre X y Y , es decir, precio y ventas, respectivamente.
2. El coeficiente de regresión (Coef) = -14.54 . Este valor (b_1) es el cambio de Y (las ventas) cuando X (el precio) se incrementa en una unidad. Cuando el precio aumenta \$1, las ventas estimadas disminuyen en 14,539 unidades.

TABLA 6-8 Resultado de Minitab para los datos del señor Bump

Correlations: Sales, Price

Pearson correlation of Sales (Y) and Price (X) = -0.863 (1)

Regression Analysis: Sales versus Price

The regression equation is

Sales (Y) = 32.1 - 14.5 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 = 2.72545 (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 (12)	0.001
Residual Error	8	59.42 (10)	7.43 (6)		
Total	9	233.60 (11)			

3. El error estándar del coeficiente de regresión (SE Coef) = 3.0. Este valor es la estimación de la desviación estándar del valor del coeficiente de regresión (b_1).

$$s_{b_1} = s_{y \cdot x} / \sqrt{\sum (X - \bar{X})^2} = 2.725 / \sqrt{.824} = 2.725 / .908 = 3.002$$

4. Se calcula el valor de $t = -4.84$. El valor de t calculado se utiliza para probar si el coeficiente de regresión de la población (β_1) es significativamente diferente de cero.

$$t = \frac{b_1}{s_{b_1}} = \frac{-14.54}{3.002} = -4.84$$

5. La constante = 32.14. Este valor es la intersección con Y (b_0). Por lo tanto, la ecuación completa de regresión es

$$\hat{Y} = 32.14 - 14.54X$$

6. El error estándar de la estimación = 2.725. El error estándar de la estimación indica que los valores de Y caen generalmente a una distancia de unas 2.725 unidades de la línea de regresión.

$$s_{y \cdot x} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{SSE}{n - 2}} = \sqrt{MSE} = \sqrt{7.43} = 2.725$$

7. El valor p , .001, es la probabilidad de obtener un valor t tan grande como $|t| = |-4.84| = 4.84$ por azar si $H_0: \beta_1 = 0$ es verdadera. Puesto que el valor p es extremadamente pequeño, se concluye que el coeficiente de la pendiente de la regresión es significativamente diferente de 0.

8. $r^2 = (R - Sq) = .746$. La línea de regresión ajustada explica el 74.6% de la varianza en el volumen de las ventas.

$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{59.42}{233.60} = 1 - .254 = .746$$

9. La r^2 ajustada ($R-Sq(adj)$) = .714. Se ajusta la r^2 para los grados de libertad adecuados.

$$\begin{aligned}\bar{r}^2 &= 1 - \frac{\Sigma(Y - \hat{Y})^2/(n - 2)}{\Sigma(Y - \bar{Y})^2/(n - 1)} = 1 - \frac{SSE/(n - 2)}{SST/(n - 1)} = 1 - \frac{59.42/8}{233.60/9} \\ &= 1 - \frac{7.428}{25.956} = 1 - .286 = .714\end{aligned}$$

10. La suma residual de cuadrados (Error residual) = 59.42. La suma de residuos al cuadrado es la suma de las diferencias al cuadrado entre las Y reales y las Y pronosticadas ($\hat{Y}$'s).

$$SSE = \Sigma(Y - \hat{Y})^2 = 59.42$$

11. Suma total de cuadrados = 233.60. Este valor es la suma de las desviaciones al cuadrado de las Y con respecto a su media.

$$SST = \Sigma(Y - \bar{Y})^2 = 233.60$$

12. El análisis de la varianza y la razón F : La razón F ($23.45 = 174.18/7.43$) en esta tabla ANOVA prueba la hipótesis nula de que la regresión no es significativa; es decir, $H_0: \beta_1 = 0$. Un valor F grande permitirá el rechazo de esta hipótesis, sugiriendo una regresión *significativa*. El valor F (23.45) aumenta conforme una porción más grande de la suma total de las desviaciones al cuadrado (SST) se explica por la regresión. En este caso, el valor tabulado de F ($df = 1, 8; \alpha = .01$) es 11.26. La hipótesis de regresión no significativa con un nivel de significancia del 1% se rechaza, puesto que $F = 23.45 > 11.26$. (Véase el ejemplo 6.9).

TRANSFORMACIONES DE VARIABLES

Si bien el modelo de regresión lineal simple supone una relación lineal entre X y Y , en general, un modelo de regresión lineal se refiere a un modelo que es lineal en la incógnita β . Siempre que la función de regresión sea lineal en las β (por ejemplo, no está presente), las variables de predicción (las X) pueden adoptar varias formas y entonces la metodología de regresión estándar sigue siendo adecuada. Es posible utilizar los modelos de regresión para modelar relaciones complejas entre Y y X (o varias X), o bien, para modelar una relación lineal entre Y y alguna función (transformación) de X .

Cuando un diagrama de dispersión indica que existe una relación no lineal entre Y y X , existen dos métodos básicos para tratar este caso. El primero consiste en ajustar los datos a una función de regresión que se grafique como una curva y utilizar la función ajustada para elaborar el pronóstico. El segundo método implica la transformación de la variable X a otra forma, de manera que la relación resultante con Y sea lineal.

Cuatro de las transformaciones más comunes (funciones) que se emplean para generar nuevas variables de predicción son el recíproco, el logaritmo, la raíz cuadrada y el cuadrado.

$$\frac{1}{X}, \log X, \sqrt{X}, X^2$$

Cuando cada una de estas variables se grafica 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 esta nueva variable pueden tratarse con el modelo lineal explicado en este capítulo, incluyendo el cálculo del coeficiente de correlación y la ecuación de regresión ajustada.

En el siguiente ejemplo se usa Minitab (véase la sección de aplicaciones de Minitab al final del capítulo) para graficar una relación simple entre X y Y que parece ser no lineal. Entonces se dan instrucciones al programa para que calcule las cuatro transformaciones descritas anteriormente. Luego, estas variables se grafican contra Y para generar los puntos de datos que se indican.

Ejemplo 6.11

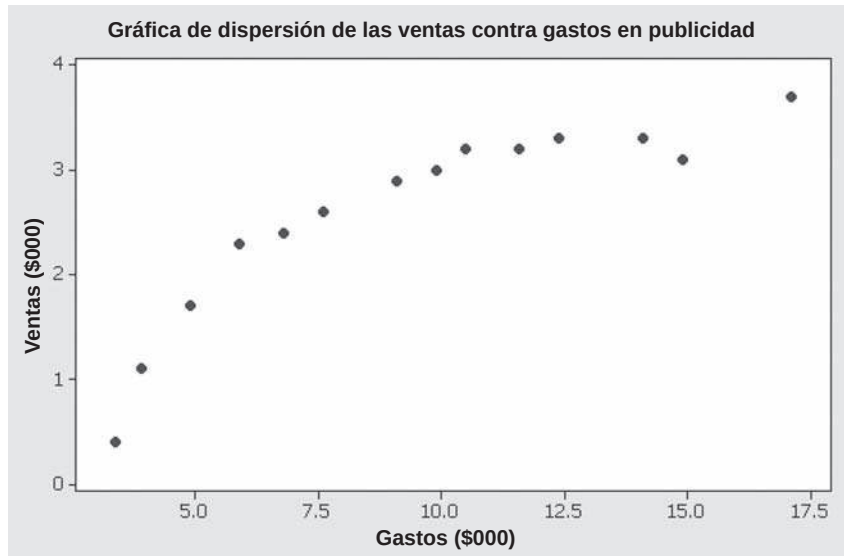
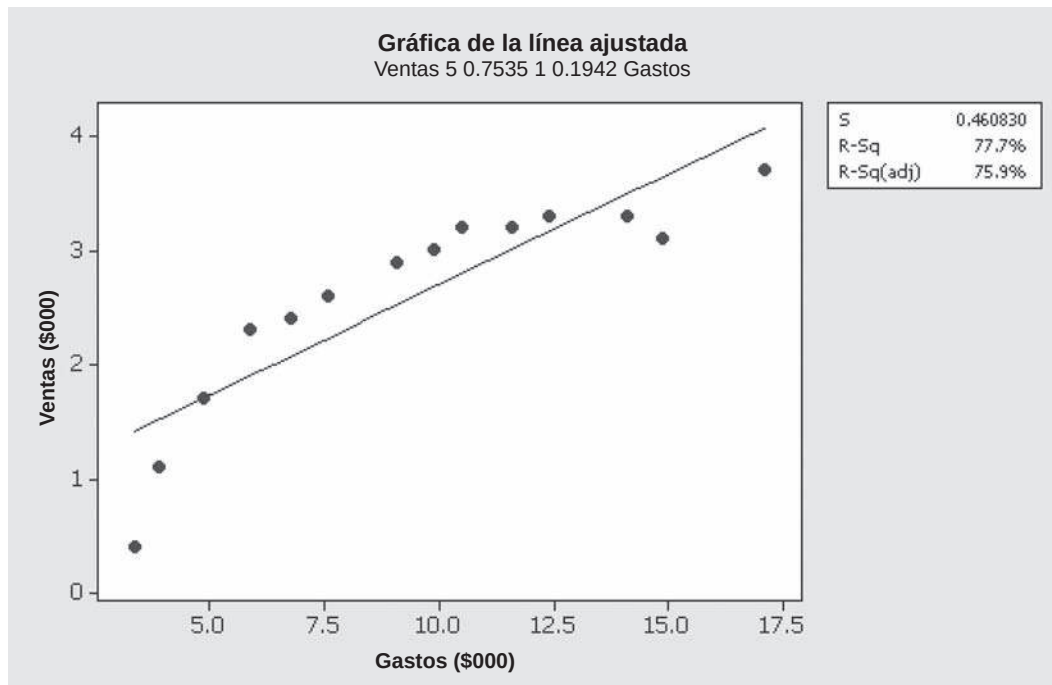
Gilbert García posee una cadena de ferreterías en Chicago, Illinois. Él está interesado en predecir sus ventas mensuales usando el conocimiento que tiene acerca de los gastos de publicidad mensual correspondientes. Gil sospecha que las ventas se incrementarán conforme la cantidad gastada en publicidad se incrementa. Sin embargo, también cree que después de cierto punto, las ventas empezarán a incrementarse a un ritmo más lento. Gil siente que, después de gastar cierta cantidad en publicidad, alcanzará un punto donde se obtendrán pocas ventas a partir de gastos ulteriores en publicidad.

A partir de los registros de la compañía, Gil seleccionó una muestra aleatoria de datos de 14 meses. Los datos aparecen en la figura 6-10, mientras que el diagrama de dispersión de los datos se presenta en la figura 6-11. Gil advierte que las ventas parecen nivelarse después de que se gasta cierta cantidad en publicidad. Ajusta la ecuación de regresión lineal presentada en la figura 6-12 y se percata de que la ecuación explica el 77.7% de la variabilidad en las ventas.

En la figura 6-13 se observa una gráfica de los residuos contra los valores ajustados de la línea recta ajustada. Esta gráfica indica que una línea recta no describe adecuadamente la relación entre las ventas y los gastos de publicidad. Todos los residuos son negativos para pequeños valores pronosticados, todos son positivos para los valores pronosticados de nivel medio, y todos son negativos otra

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	Sales	Expenditures	LOGX	SQRTX	X*X	1/X				
1	1.1	3.9	1.36098	1.97484	15.21	0.256410				
2	1.7	4.9	1.58924	2.21359	24.01	0.204082				
3	2.6	7.6	2.02815	2.75681	57.76	0.131579				
4	2.4	6.8	1.91692	2.60768	46.24	0.147059				
5	2.3	5.9	1.77495	2.42899	34.81	0.169492				
6	2.9	9.1	2.20827	3.01662	82.81	0.109890				
7	0.4	3.4	1.22378	1.84391	11.56	0.294118				
8	3.2	11.6	2.45101	3.40588	134.56	0.086207				
9	3.3	14.1	2.64617	3.75500	198.81	0.070922				
10	3.1	14.9	2.70136	3.86005	222.01	0.067114				
11	3.2	10.5	2.35138	3.24037	110.25	0.095238				
12	3.0	9.9	2.29253	3.14643	98.01	0.101010				
13	3.7	17.1	2.83908	4.13521	292.41	0.058480				
14	3.3	12.4	2.51770	3.52136	153.76	0.080645				
15										

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

vez para valores pronosticados grandes. Los residuos no están distribuidos de manera uniforme alrededor de la línea de regresión estimada. Se ve claramente que una línea recta no refleja la curvatura de los datos.

Luego, Gil considera varias transformaciones de X (véase la figura 6-10). Construye un diagrama de dispersión múltiple con gráficas de las ventas (Y) contra los logaritmos de los gastos en publicidad ($\log X$), ventas (Y) contra las raíces cuadradas de los gastos en publicidad ($\sqrt{X}$), ventas

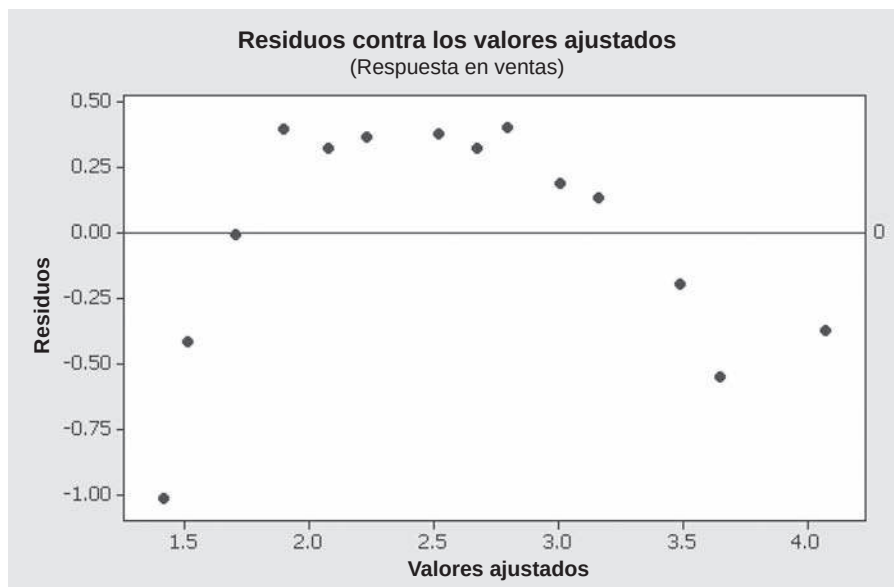


FIGURA 6-13 Gráfica de residuos contra valores ajustados para la línea recta ajustada del ejemplo 6.11

(Y) contra los gastos en publicidad elevados al cuadrado (X^2) y, finalmente, ventas (Y) contra los recíprocos de los gastos en publicidad ($1/X$). El resultado se presenta en la figura 6-14.

Después de examinar el diagrama de dispersión múltiple, Gil concluye que las relaciones entre las ventas y los logaritmos de gastos en publicidad, las ventas y las raíces cuadradas de gastos en publicidad, y las ventas y los gastos en publicidad elevados al cuadrado muestran todos algún grado de curvatura. Estas relaciones no son lineales. Sin embargo, la relación entre las ventas y los recíprocos de los gastos en publicidad parece ser lineal con una pendiente negativa. Usando Minitab, Gil ajusta un modelo de regresión lineal simple a los datos de las ventas y los recíprocos de los gastos en publicidad. Así, determina la ecuación ajustada $\hat{Y} = 4.29 - 12.7(1/X)$, como se presenta en la tabla 6-9, y ve que r^2 es igual al 98.1%. Un análisis residual indica que este modelo lineal es apropiado.

CURVAS DE CRECIMIENTO

Las curvas de crecimiento se presentaron en el capítulo 5 (véase la figura 5-7) en nuestro estudio para modelar la tendencia en la descomposición de una serie de tiempo. A menudo, la tendencia (cambio a largo plazo) es de interés por sí misma. Un analista, por ejemplo, tal vez esté interesado en la proyección de los costos del uso de la computadora hasta bien entrado el futuro sin considerar la tecnología que podría traer cambios en el costo. Como otro ejemplo, un analista de seguros de vida quizás esté interesado en proyecciones a largo plazo de la expectativa de vida de la población de Estados Unidos sin considerar las condiciones económicas y ambientales que podrían ser responsables de cambios en la expectativa de vida.

Las curvas de crecimiento son relaciones curvilíneas entre la variable de interés y el tiempo. Las curvas de crecimiento se ajustan comúnmente a los datos anuales, puesto que en estos casos se requieren pronósticos a largo plazo.

Aun cuando los pronósticos son probablemente imprecisos cuando los ajustes de las curvas de crecimiento a los datos históricos son extrapolados para predecir el futuro, este método de pronóstico puede ser de gran ayuda para los administradores, puesto que concentra su atención en aspectos a largo plazo de los negocios. Más aún, las curvas de crecimiento indican la tasa anual de crecimiento que debe mantenerse con la finalidad de alcanzar los niveles proyectados

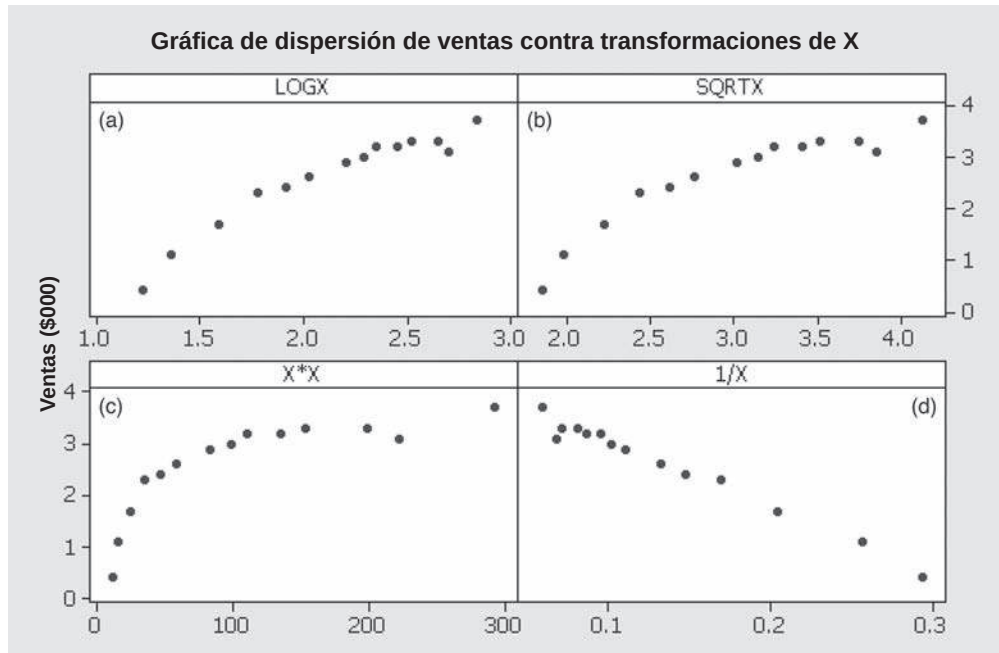


FIGURA 6-14 Diagrama de dispersión de ventas (Y) contra $a)$ $\log X$, $b)$ raíz cuadrada de X , $c)$ X al cuadrado y $d)$ recíproco de X

TABLA 6-9 Resultado de Minitab para el ejemplo 6-11

Regression Analysis: Sales versus 1/X

The regression equation is

$$\text{Sales} = 4.29 - 12.7 \, 1/X$$

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 = 0.134159 R-Sq = 98.1% R-Sq(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			

para el futuro. Esta tasa de crecimiento anual puede o no ser razonable y tal vez sea motivo de debate en una reunión de ejecutivos o en un “comité de expertos”.

Si una variable medida en el tiempo se incrementa en el mismo *porcentaje* cada periodo, se dice que presenta un *crecimiento exponencial*. Si una variable se incrementa en la misma *cantidad* cada periodo, se dice que registra un *crecimiento lineal*. Algunas veces una transformación simple convierte una variable con crecimiento exponencial en una variable con crecimiento lineal. Si éste es el caso, los métodos de regresión presentados en este capítulo servirán para modelar el crecimiento exponencial.

Por ejemplo, suponga que una variable Y , medida anualmente, sigue la curva de tendencia exponencial (véase la ecuación 5.6).

$$Y = b_0 b_1^t$$

Aquí, $100(b_1 - 1)\%$ es el incremento porcentual anual de Y . Tomando los logaritmos de ambos lados de la ecuación de la tendencia exponencial, tenemos

$$\log Y = \log(b_0 b_1^t) = \log b_0 + t \log b_1 = \tilde{b}_0 + \tilde{b}_1 t$$

Así, $\log Y$ contra el tiempo (t) grafica una línea recta con pendiente $\tilde{b}_1 = \log b_1$ e intersección $\tilde{b}_0 = \log b_0$. $\log Y$ se incrementa una cantidad constante, $\tilde{b}_1$, de un año a otro. Como se ilustra en el siguiente ejemplo, es posible obtener los pronósticos de Y y una estimación de la tasa de crecimiento, b_1 , a partir de la ecuación de la línea recta que relaciona $\log Y$ con t .

Ejemplo 6.12

Jill Johnson, la asistente administrativa de una universidad importante, está contenta con la aprobación de un sistema de inscripciones por Internet para los cursos. Ella recolectó los datos de la tabla 6-10 y ve que el uso del sistema ha crecido de forma increíble. Si la tasa actual de crecimiento continúa, todo el campus estará registrándose por Internet para los cursos dentro de un año y el sistema anterior de inscripciones por teléfono se eliminará por completo.

Los datos de la tabla 6-10 indican los números de estudiantes que usan el sistema de inscripciones por Internet para el periodo de 2002 a 2006. La última columna de la tabla contiene los logaritmos de base 10 de los números de usuarios.

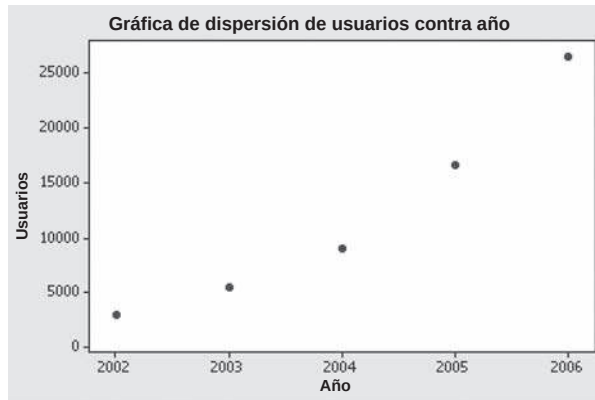
Jill ha tenido alguna experiencia con la regresión lineal simple y se pregunta si este método servirá para pronosticar el número de usuarios en 2007 y para hacer una estimación de la tasa de crecimiento anual. Una gráfica del número de usuarios contra el año (véase la figura 6-15a) rápidamente le sugiere a Jill que un modelo de regresión lineal simple no es adecuado para sus datos. El crecimiento del uso parece ser exponencial, con un incremento porcentual anual de alrededor del 70%. Sin embargo, el aumento en el número de usuarios cada año es completamente diferente. Los incrementos varían en magnitud desde una modesta cifra de 2,500 durante el periodo de 2002 a 2003 a una cifra considerable de 10,000 durante el periodo de 2005 a 2006.

Jill siempre ha sido buena manejando logaritmos y recuerda que un cambio porcentual constante en una variable durante cierto periodo es equivalente a un incremento constante en la magnitud del logaritmo de la variable durante el mismo periodo de tiempo. Con esto en mente, Jill transforma los números de las inscripciones por Internet de los usuarios a los cursos tomando sus logaritmos; luego, grafica los logaritmos (usuarios) contra el año (véase la figura 6-15b).

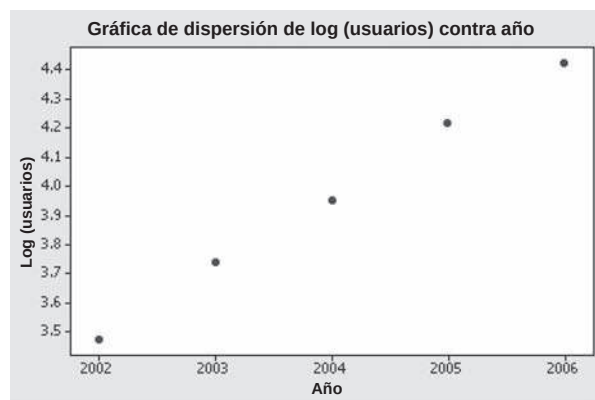
Como ella sospechaba, cuando los números de los usuarios se transforman usando sus logaritmos, el crecimiento exponencial se convierte en un crecimiento lineal. La curva que representa el crecimiento exponencial en la figura 6-15a se convierte en una línea en la figura 6-15b) gracias a la transformación logarítmica. En la figura 6-15b), el incremento en la magnitud de los logaritmos (usuarios) cada año es aproximadamente el mismo, alrededor de .23 por año. Jill ajusta una línea recta a los datos de la figura 6-15b) con los logaritmos (usuarios) como la variable dependiente Y y el tiempo (t) como la variable independiente. Los resultados se presentan en la figura 6-16.

TABLA 6-10 Datos para el ejemplo 6.12

<i>Año</i>	<i>Tiempo</i>	<i>Usuarios</i>	<i>Log (Usuarios)</i>
2002	1	3,000	3.477
2003	2	5,500	3.740
2004	3	9,000	3.954
2005	4	16,500	4.217
2006	5	26,500	4.423



a)



b)

FIGURA 6-15 Diagramas de dispersión de los datos de Jill Johnson para el ejemplo 6.12

Jill calcula el pronóstico para el año de 2007 usando la ecuación ajustada $\log \hat{Y} = 3.252 + .2369t$ con $t = 6$:

$$\log \hat{Y}_{2007} = 3.252 + .2369(6) = 4.673$$

$$\hat{Y}_{2007} = 10^{4.673} = 47,100$$

Jill pronostica que 47,100 estudiantes usarán el sistema de inscripción por Internet en 2007. Puesto que la universidad espera una matrícula total de 45,000 estudiantes, Jill concluye que probablemente todos los estudiantes van a usar el nuevo sistema de inscripciones a los cursos y que el sistema telefónico anterior puede descontinuarse.

Finalmente, Jill calcula que la tasa de crecimiento anual de los usuarios del sistema de inscripciones por Internet es

$$100(\text{antilog}(.2369) - 1)\% = 100(10^{.2369} - 1)\% = 100(1.725 - 1)\% = 72.5\%$$

Jill admite que con la capacidad de inscripción de la universidad, no tiene sentido extrapolar el aumento del número de usuarios del sistema de inscripción por Internet más allá de 2007.

En algunas situaciones, una curva de crecimiento lineal puede ajustarse directamente a los datos originales.

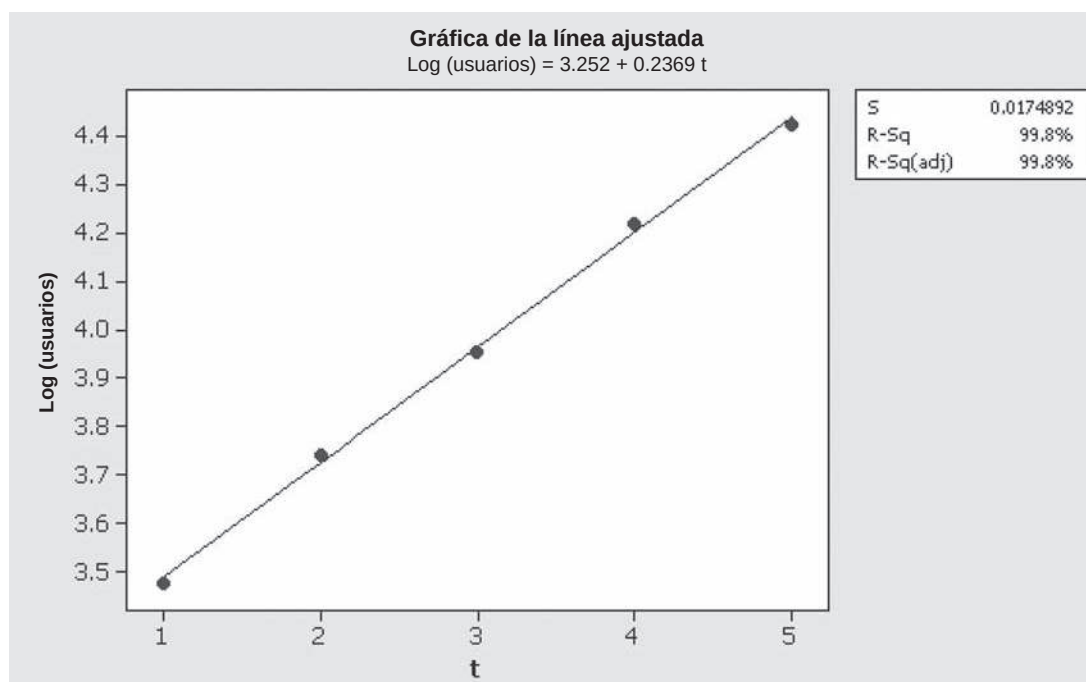


FIGURA 6-16 Línea ajustada de regresión para el ejemplo 6.12

Ejemplo 6.13

Suponga que en 1999 una compañía relacionada con la prestación de servicios de salud para personas mayores está interesada en una proyección a largo plazo de la expectativa de vida en el momento del nacimiento en Estados Unidos, de manera que pueda formular una estrategia corporativa de largo plazo. La compañía obtiene proyecciones de esta variable desde 1970 a 1997 del *Statistical Abstract of the United States*. Estos datos se presentan en la tabla 6-11. Después de codificar los años como 1 = 1970, 2 = 1971, ..., 28 = 1997 y teclear los datos en Minitab, la rutina Regresión-Gráfica de línea ajustada produce la gráfica de la figura 6-17.

La compañía no está interesada en examinar las razones de un incremento obvio en la longevidad durante varios años pasados, sino en una proyección lejana en el futuro de la línea de regresión ajustada. Tampoco está interesada en especular acerca de si la tendencia ascendente continuará, o por qué medios se podría alcanzar un incremento como éste. Sólo quiere completar el siguiente enunciado: “Si las tendencias actuales continúan, la expectativa de vida en el momento del nacimiento será de Y en un futuro de X años”. Se selecciona el año de 2050 ($X = 53$ años más allá de 1997) como el año objetivo en el futuro. Al usar $28 + 53 = 81$ como el valor para el tiempo en la ecuación de regresión que se presenta en la figura 6-17 y al despejar Y, la compañía obtiene una expectativa de vida de 86.6 años. Esta edad obviamente representa un aumento sustancial en la longevidad para las próximas décadas y la compañía empieza a trabajar sobre planes estratégicos para tomar ventaja de este incremento esperado en el tiempo de vida.

APLICACIÓN A LA ADMINISTRACIÓN

El análisis de regresión es la herramienta estadística que se utiliza más ampliamente en la dirección de una empresa cuando existe la necesidad de evaluar el efecto de una sola variable independiente sobre una variable dependiente. El análisis de regresión —junto con el análisis de correlación— ayuda al responsable de hacer el pronóstico a conocer las relaciones entre variables. El analista puede determinar tanto la importancia como la dirección de la relación entre variables.

La mayoría de los problemas que requieren el análisis de regresión implican la versión más compleja, llamada *análisis de regresión múltiple* (que estudiaremos en el siguiente capítulo), porque la mayoría de las relaciones ameritan el estudio de la relación entre una variable

TABLA 6-11 Expectativa de vida al nacer de la población total de Estados Unidos, ejemplo 6.13			
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

Fuente: *Statiscal Abstract of the United States*, 119a. ed., 1999.

*Interpoladas.

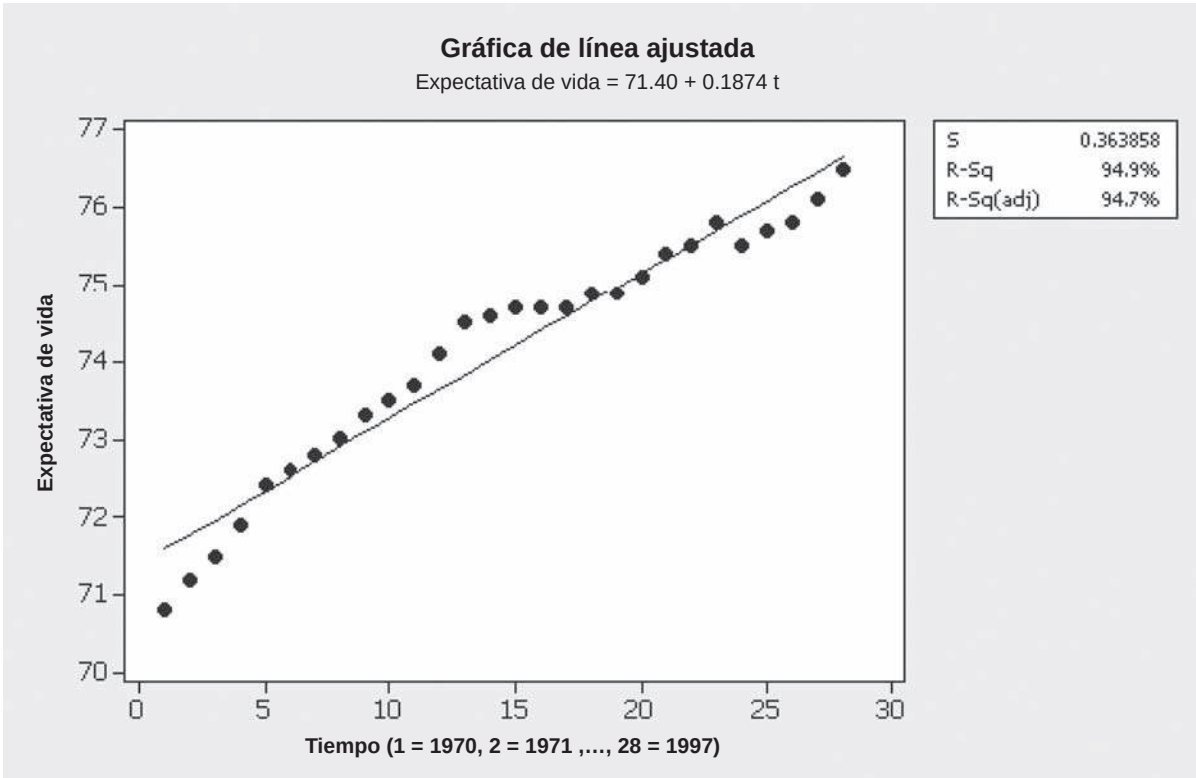


FIGURA 6-17 Línea de regresión ajustada para el ejemplo 6.13

dependiente y más de una variable independiente. Sin embargo, la regresión simple y el análisis de correlación simple se utilizan con frecuencia. Los siguientes son unos cuantos ejemplos de situaciones que implican la regresión múltiple:

- *Consumo de productos.* Un fabricante desea pronosticar cuánta cerveza bebe una persona a la semana considerando variables como ingreso, edad, educación y clase demográfica.
- *Ventas.* Un minorista quiere pronosticar las ventas de un producto de una tienda contra las ventas de otra sobre la base de diferencias de precio, el ingreso relativo de la comunidad circundante, la simpatía relativa del personal de la tienda, y el número y fortaleza de los competidores en cada mercado.
- *Precios de las acciones.* Un analista bursátil de una firma regional de corretaje desea pronosticar el precio de una nueva emisión de acciones de una empresa local sobre la base de la economía regional, el ingreso, la población y el prestigio de la compañía.
- *Deudas incobrables.* Un contador necesita pronosticar las deudas incobrables que una empresa podría enfrentar el siguiente año fiscal sobre la base del número de personas desempleadas, créditos pendientes, tasas de interés y ventas esperadas.
- *Necesidades de empleo.* El director de personal de una gran empresa manufacturera desea pronosticar los requerimientos de personal para el año venidero tomando como base la edad promedio de sus empleados, su escala de salarios comparada con los de la localidad, la expectativa de nuevos contratos de ventas y la disponibilidad de puestos de trabajo competitivos.
- *Demanda en centros comerciales.* El gerente de un nuevo centro comercial quiere anticipar la demanda analizando el ingreso de la localidad, el tamaño de la población, así como la proximidad y el tamaño de centros comerciales competidores.

Una vez que se determina la relación entre las variables independiente y dependiente, la administración puede, en algunos casos, tratar de controlar la variable dependiente con base en este conocimiento. Por ejemplo, suponga que un gerente de marketing determina que existe una relación positiva significativa entre los gastos de publicidad y las ventas. La ecuación de regresión podría ser

$$\text{Ventas} = \$43,000 + .3(\text{Gastos de publicidad})$$

A partir de esta ecuación, el gerente de marketing puede tratar de controlar las ventas aumentando o disminuyendo los gastos de publicidad en una cantidad que maximice las utilidades. Siempre que el gerente tenga control sobre la variable independiente y exista la relación causa-efecto implicada por la ecuación de regresión, es posible controlar parcialmente la variable dependiente. La ecuación de regresión y el coeficiente de determinación ayudan a la administración de una compañía a determinar si tal control vale la pena.

Glosario

Coefficiente de determinación. Mide el porcentaje de variabilidad de Y que puede explicarse a través del conocimiento de la variabilidad (diferencias) de la variable independiente X .

Línea de regresión ajustada. Es la línea que mejor se ajusta a la colección de puntos de datos $X-Y$. Minimiza la suma de las distancias elevadas al cuadrado desde los puntos hasta la línea medidas verticalmente, es decir, en la dirección Y .

Error estándar de la estimación. Mide la cantidad por la que los valores reales de Y difieren de los valores estimados. Es una estimación de la desviación estándar del término de error, ε , en el modelo de regresión lineal simple.

Error estándar del pronóstico (s_f). Mide la variabilidad de la Y pronosticada en relación con la Y real para un valor dado de X .

Fórmulas clave

Suma de errores al cuadrado

$$SSE = \sum (Y - \hat{Y})^2 = \sum (Y - b_0 - b_1X)^2 \quad (6.1)$$

Método de los mínimos cuadrados: fórmula de la pendiente

$$b_1 = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2} \quad (6.2)$$

Método de los mínimos cuadrados: fórmula de la intersección con Y

$$b_0 = \frac{\sum Y}{n} - \frac{b_1 \sum X}{n} = \bar{Y} - b_1 \bar{X} \quad (6.3)$$

Relación entre el coeficiente de la pendiente y el coeficiente de correlación

$$b_1 = \frac{\sqrt{\sum (Y - \bar{Y})^2}}{\sqrt{\sum (X - \bar{X})^2}} r \quad (6.4)$$

Observación = Ajuste + residuo

$$Y = \hat{Y} + (Y - \hat{Y}) \quad (6.5)$$

Ecuación de regresión ajustada

$$\hat{Y} = b_0 + b_1X \quad (6.6)$$

Error estándar de la estimación: fórmula de definición

$$s_{y \cdot x} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} \quad (6.7)$$

Error estándar de la estimación: fórmula de cálculo

$$s_{y \cdot x} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b_1 \sum XY}{n - 2}} \quad (6.8)$$

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 de predicción

$$\hat{Y} \pm ts_f \quad (6.10)$$

Intervalo de predicción del 95% para una muestra grande

$$\hat{Y} \pm 2s_f \quad (6.11)$$

Descomposición de la suma de los cuadrados y grados de libertad

$$\begin{aligned} SST &= SSR + SSE \\ \sum (Y - \bar{Y})^2 &= \sum (\hat{Y} - \bar{Y})^2 + \sum (Y - \hat{Y})^2 \end{aligned} \quad (6.12)$$

$$df: n - 1 = 1 + (n - 2) \quad (6.13)$$

Coefficiente de determinación

$$r^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} \quad (6.14)$$

Estadístico t para probar $H_0: \beta_1 = 0$

$$t = \frac{b_1}{s_{b_1}} \quad (\text{véase el recuadro sombreado en la p. 237})$$

Error estándar del coeficiente de regresión

$$s_{b_1} = s_{y \cdot x} / \sqrt{\sum(X - \bar{X})^2} \quad (\text{véase el recuadro sombreado en la p. 237})$$

Estadístico F

$$F = \frac{\text{Cuadrado medio de regresión}}{\text{Error cuadrático medio}} = \frac{MSR}{MSE} \quad (6.15)$$

Relación del estadístico F con el coeficiente de determinación

$$F = \frac{r^2(n-2)}{1-r^2} \quad (6.16)$$

Coefficiente de autocorrelación residual

$$r_k(e) = \frac{\sum_{t=k+1}^n e_t e_{t-k}}{\sum_{t=1}^n e_t^2} \quad k = 1, 2, \dots, K \quad (6.17)$$

Problemas

Nota: La mayoría de los problemas siguientes contienen datos que van a ser manipulados usando procedimientos de análisis de regresión. Si bien es posible, incluso útil, trabajar uno o dos de estos análisis a mano, es importante que usted aprenda cómo usar el software de computadora para resolver problemas como éstos. En el siguiente capítulo usted aprenderá cómo se efectúa un análisis de regresión múltiple, en el cual no es posible resolver los problemas a mano. Por eso, usted debe familiarizarse con el software de análisis de regresión cuando resuelva los siguientes problemas. Si usted tiene acceso a Minitab o Excel, vea las instrucciones de uso en la sección final de este capítulo.

- ¿Cuál de las siguientes situaciones es incongruente?
 - $\hat{Y} = 499 + .21X$ y $r = .75$
 - $\hat{Y} = 100 + .9X$ y $r = -.70$
 - $\hat{Y} = -20 + 1X$ y $r = .40$
 - $\hat{Y} = -7 - 4X$ y $r = -.90$
- Las ganancias en miles de millones de dólares de American Telegraph and Telephone (AT&T) se estiman usando el producto interno bruto (PIB). La ecuación de regresión es $\hat{Y} = .078 + .06X$, donde el PIB se mide en miles de millones de dólares.
 - Interprete la pendiente.
 - Interprete la intersección con Y .

TABLA P-3

<i>Y</i> (\$)	<i>X</i> (\$)	<i>Y</i> (\$)	<i>X</i> (\$)
1,250	41	1,300	46
1,380	54	1,400	62
1,425	63	1,510	61
1,425	54	1,575	64
1,450	48	1,650	71

3. Considere los datos de la tabla P-3 donde X = gastos semanales de publicidad y Y = ventas semanales.
 - a) ¿Existe una relación significativa entre los gastos de publicidad y las ventas?
 - b) Defina la ecuación de predicción.
 - c) Pronostique las ventas para un gasto de publicidad de \$50.
 - d) ¿Qué porcentaje de variación de las ventas puede explicarse con la ecuación de predicción?
 - e) Defina la cantidad de variación no explicada.
 - f) Obtenga la cantidad de variación total.
4. En la tabla P-4 se presenta el tiempo requerido para hacer el cobro a los clientes en un supermercado y los valores correspondientes a las compras. Conteste los incisos a), b), e) y f) del problema 3 usando estos datos. Proporcione un estimado puntual y un intervalo estimado del 99% para Y si $X = 3.0$.
5. A Lori Franz, supervisora de mantenimiento de la Baltimore Transit Authority, le gustaría determinar si existe una relación positiva entre el costo anual de mantenimiento de un autobús y su antigüedad. Si existe una relación, Lori siente que puede hacer un mejor trabajo al pronosticar el presupuesto de mantenimiento anual del autobús. Ella recopiló los datos que se presentan en 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 línea de mínimos cuadrados.
 - e) Pruebe la significancia del coeficiente de la pendiente para el nivel de significancia de .05. ¿Es significativa la correlación? Explique.
 - f) Pronostique el costo anual de mantenimiento para un autobús de cinco años de antigüedad.
6. Andrew Vazsonyi es el gerente de la cadena de supermercados Spendwise. A él le gustaría pronosticar las ventas de libros en edición rústica (libros por semana) con base en la canti-

TABLA P-4

<i>Tiempo requerido para el pago (minutos)</i>	<i>Valor de la compra (\$)</i>	<i>Tiempo requerido para el pago (minutos)</i>	<i>Valor de la compra (\$)</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

TABLA P-5

<i>Autobús</i>	<i>Costo de mantenimiento (\$)</i> <i>Y</i>	<i>Antigüedad (años)</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-6

<i>Semana</i>	<i>Número de libros vendidos</i> <i>Y</i>	<i>Pies de espacio de anaqueles</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

dad de espacio de exhibición disponible (pies) en los anaqueles. Andrew recopila los datos de una muestra de 11 semanas, los cuales se presentan en la tabla P-6.

- Grafique 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 del coeficiente de la pendiente en el nivel de significancia de .10. ¿Es significativa la correlación? Explique.
 - Grafique los residuos contra los valores ajustados. De acuerdo con esta gráfica, ¿el modelo de regresión lineal simple es adecuado para estos datos?
 - Pronostique las ventas de libros en edición rústica para una semana durante la cual existen cuatro pies de espacio en los anaqueles.
- En la tabla P-7 se presenta la información correspondiente a un negocio de ventas por correo que opera en 12 ciudades.
 - Determine la línea de regresión ajustada.
 - Calcule el error estándar de la estimación.
 - Determine la tabla ANOVA.
 - ¿Qué porcentaje de la variación de ventas por correo se explica por el número de catálogos distribuidos?

TABLA P-7

<i>Ciudad</i>	<i>Número de órdenes de ventas por correo (en miles)</i> <i>Y</i>	<i>Número de catálogos distribuidos (en miles)</i> <i>X</i>	<i>Ciudad</i>	<i>Número de órdenes de ventas por correo (en miles)</i> <i>Y</i>	<i>Número de catálogos distribuidos (en miles)</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

- e) Haga una prueba para determinar si el coeficiente de regresión de la pendiente es significativamente diferente de cero. (Use un nivel de significancia de .01).
- f) Pruebe la significancia de la regresión usando el estadístico F de la tabla ANOVA. (Use un nivel de significancia de .01). ¿Es congruente el resultado con el del inciso e)? ¿Debería serlo?
- g) Pronostique el número de ventas por correo recibidas cuando se distribuyen 10,000 catálogos; utilice un intervalo de predicción del 90%.
8. En un estudio de inversiones y tasas de interés se recabaron los datos de la tabla P-8 durante un periodo de 10 años.
- a) ¿Es significativa la relación entre estas variables?
- b) ¿Puede desarrollarse una ecuación efectiva de predicción?
- c) Si la tasa de interés promedio es del 4% dentro de cinco años, ¿puede pronosticarse la inversión anual?
- d) Calcule e interprete r^2 .
- e) Discuta la correlación y la causalidad en este ejemplo.
9. La ABC Investment Company está en el negocio de hacer licitaciones sobre inversiones ofrecidas por varias empresas que quieren financiamiento adicional. La ABC ha tabulado sus ofertas en las últimas 25 licitaciones en términos de su porcentaje de valor nominal. También se tabularon las ofertas del principal competidor de ABC, como un porcentaje del valor nominal. ABC ahora se pregunta si está usando el mismo razonamiento que su competidor en la preparación de las ofertas. En otras palabras, ¿es posible que ABC pronostique las ofertas de su competidor a partir de las propias? Si no es así, entonces el competidor debe estar evaluando las licitaciones de manera diferente. Los datos se presentan en la tabla P-9.

TABLA P-8

<i>Inversión anual (en miles de \$)</i>	<i>Tasa de interés promedio (%)</i>	<i>Inversión anual (en miles de \$)</i>	<i>Tasa de interés promedio (%)</i>
1,060	4.8	2,050	3.8
940	5.1	2,070	3.7
920	5.9	2,030	4.5
1,110	5.1	1,780	4.9
1,590	4.8	1,420	6.2

TABLA P-9

<i>Licitación</i>	<i>Oferta de ABC</i>	<i>Oferta del competidor</i>	<i>Licitación</i>	<i>Oferta de ABC</i>	<i>Oferta 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			

- a) ¿Hasta qué grado están usando las dos firmas el mismo razonamiento en la preparación de sus ofertas?
- b) Pronostique la oferta del competidor si las ofertas de ABC tienen el 101% de valor nominal. Construya para ambos un estimado puntual y un intervalo de la predicción.
- c) En el inciso b), ¿cuál es la probabilidad de que ABC gane esta licitación particular? (La oferta más baja gana).
10. Evalúe los siguientes enunciados.
- a) Una r^2 alta se refiere a una regresión significativa.
- b) Un tamaño muestral muy grande en un problema de regresión siempre arrojará resultados útiles.
11. Suponga que usted fue contratado por Ed Bodganski, dueño de la American Precast Company, como analista de tiempo parcial. Ed estaba muy contento cuando usted determinó que existe una relación positiva entre los permisos de construcción expedidos y la cantidad de trabajo disponible para su compañía. Ahora él se pregunta si es posible usar el conocimiento de las tasas de interés sobre las hipotecas de primer grado para predecir el número de permisos de construcción que se expedirán cada mes. Usted recopila una muestra aleatoria de nueve meses de datos, como se presenta en la tabla P-11.

TABLA P-11

<i>Mes</i>	<i>Permisos de construcción Y</i>	<i>Tasa de interés (%) 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 la pendiente con un nivel de significancia de .05.
 - d) Cuando la tasa de interés se incrementa en 1%, ¿cuál es el decremento promedio en el número de permisos de construcción expedidos?
 - e) Calcule el coeficiente de determinación.
 - f) Redacte un enunciado que Ed pueda comprender y en el cual se interprete el número calculado en el inciso e).
 - g) Redacte un escrito para Ed explicándole los resultados de su análisis.
12. Considere la población de 140 observaciones que se presenta en la tabla P-12. La Marshall Printing Company desea estimar la relación entre el número de copias producidas por una técnica de impresión de offset (X) y el costo de la mano de obra directa asociada (Y). Seleccione una muestra aleatoria de 20 observaciones.
- a) Construya un diagrama de dispersión.
 - b) Calcule el coeficiente de correlación de la muestra.
 - c) Determine la línea de regresión ajustada.
 - d) Grafique la línea ajustada sobre 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, β_1 , de la línea de regresión de la población es cero.
 - h) Calcule un pronóstico puntual y un intervalo de predicción del 90% para el costo de la mano de obra directa si el proyecto implica 250 copias.
 - i) Examine los residuos. ¿Parece que un modelo de regresión lineal simple es adecuado para estos datos? Explique.
13. Harry Daniels es un ingeniero de control de calidad de Specific Electric Corporation, una empresa dedicada a fabricar motores eléctricos. Uno de los pasos en el proceso de manufactura implica el uso de una fresadora automática para hacer las ranuras en el eje de los motores. Cada lote de ejes de motor se prueba y todos los ejes que no tengan las dimensiones requeridas se desechan. La fresadora debe reajustarse al comenzar a trabajar con cada nuevo lote porque su cabeza cortadora se desgasta ligeramente durante la producción. A Harry se le asigna el trabajo de pronosticar cómo afecta el tamaño de un lote al número de ejes defectuosos en el lote, de manera que pueda seleccionar el mejor tamaño de lote. Él recopila los datos del tamaño promedio del lote de los 13 lotes considerados en la tabla P-13 y le pide a usted analizarla.
- a) Grafique los datos en un diagrama de dispersión.
 - b) Ajuste un modelo de regresión lineal simple.
 - c) Haga una prueba de significancia del coeficiente de la pendiente.
 - d) Examine los residuos.
 - e) Desarrolle un modelo curvilíneo ajustando un modelo de regresión lineal simple para alguna transformación de la variable independiente.
 - f) Haga una prueba de la significancia del coeficiente de la pendiente de la variable transformada.
 - g) Examine los residuos.
 - h) Pronostique el número de defectos para un tamaño de lote de 300 ejes.
 - i) ¿Cuál modelo prefiere, el del inciso b) o el del inciso e)?
 - j) Redacte un escrito para Harry resumiendo sus resultados.
14. Los datos de la tabla P-14 fueron recopilados como parte de un estudio de evaluación de bienes inmuebles. Los números son observaciones de X = valor tasado (en miles de dólares) en los libros del valuador de la ciudad y Y = valor de mercado (precio de venta en miles de dólares) para $n = 30$ terrenos que se vendieron en un año particular en cierta área geográfica.

TABLA P-12

<i>Obs.</i>	<i>Y</i>	<i>X</i>	<i>Obs.</i>	<i>Y</i>	<i>X</i>	<i>Obs.</i>	<i>Y</i>	<i>X</i>
(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 defectos</i> <i>Y</i>	<i>Tamaño del lote</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

TABLA P-14

<i>Terreno</i>	<i>Valor tasado</i>	<i>Mercado</i>	<i>Terreno</i>	<i>Valor tasado</i>	<i>Mercado</i>
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

- Grafique el valor de mercado contra el valor tasado en los libros en un diagrama de dispersión.
 - Suponiendo un modelo de regresión lineal simple, determine la línea de mínimos cuadrados que relaciona el valor de mercado con el valor tasado en los libros.
 - Determine r^2 e interprete su valor.
 - ¿Es significativa la regresión? Explique.
 - Pronostique el valor de mercado de una propiedad a partir de un valor tasado en los libros de 90.5. ¿Hay algún peligro al hacer esta predicción?
 - Examine los residuos. ¿Puede usted identificar observaciones que tienen gran influencia en la localización de la línea de mínimos cuadrados?
15. En la tabla P-15 se presentan los costos de jugadores (X) y gastos operativos (Y) para $n = 26$ de los principales equipos de la liga mayor de béisbol durante la temporada 1990-1991.

TABLA P-15

<i>Equipo</i>	<i>Costos de jugadores (millones de \$)</i>	<i>Gastos operativos Y (millones de \$)</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

- Suponiendo un modelo de regresión lineal simple, determine la ecuación para la línea recta ajustada.
 - Determine r^2 y haga comentarios acerca de la intensidad de la relación lineal.
 - Pruebe la significancia de la regresión para un nivel de significancia de .10.
 - ¿Se puede concluir que, como regla general, los gastos operativos son aproximadamente el doble de los costos de los jugadores? Analícelo.
 - Pronostique los gastos operativos, con un intervalo de predicción del 95% de una muestra grande, si los costos de los jugadores son de 30.5 millones de dólares.
 - Usando los residuos como guía, identifique cualquier observación atípica. Es decir, ¿algunos equipos tienen costos de jugadores inusualmente bajos o altos en comparación con los gastos operativos?
16. La tabla P-16 contiene datos de consumo de periódicos de 23 ciudades (Y) y el número de familias en la ciudad (X) durante un año particular.
- Grafique el consumo de periódicos contra el número de familias en un diagrama de dispersión.
 - ¿Es adecuado el modelo de regresión lineal simple para los datos de la tabla P-16? Asegúrese de que su respuesta incluya un análisis de los residuos.
 - Considere una transformación logarítmica del consumo de periódicos y un modelo de regresión lineal simple que relacione $Y = \log$ (consumo de periódicos) con $X =$ número de familias. Ajuste este modelo.

TABLA P-16

<i>Ciudad</i>	<i>Consumo de periódico Y (toneladas métricas)</i>	<i>Número de familias X</i>	<i>Ciudad</i>	<i>Consumo de periódico Y (toneladas métricas)</i>	<i>Número de familias 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			

- d) Examine los residuos de la regresión del inciso c). ¿Cuál modelo es mejor, el del inciso b) o el del inciso c)? Justifique su respuesta.
- e) Usando la función ajustada del inciso c), pronostique la cantidad de periódico consumida en un año si una ciudad tiene 100,000 familias.
- f) ¿Podría indicar otras variables que probablemente tengan influencia sobre la cantidad de periódico consumida durante un año?
17. La Outback Steakhouse creció explosivamente durante sus primeros años de operación. Los números de sucursales de Outback Steakhouse para el periodo de 1988 a 1993 se presentan abajo.

Año	1988	1989	1990	1991	1992	1993
Núm. de sucursales	2	9	23	49	87	137

Fuente: Outback Steakhouse.

- a) ¿Parece haber un crecimiento lineal o exponencial en el número de sucursales?
- b) Calcule la tasa de crecimiento anual de Outback Steakhouse durante el periodo de 1988 a 1993.
- c) Pronostique el número de sucursales de Outback para 2007. ¿Parece razonable este número? Explique.
18. On The Double, una cadena de centros de copiado instalados dentro de planteles universitarios, inició operaciones con una sola tienda en 1993. En la tabla P-18 está registrado el número de centros de copiado en operación, Y , durante 14 años consecutivos.
- a) Grafique el número de centros de copiado contra el año. ¿El crecimiento experimentado por On The Double es lineal o exponencial?
- b) Determine la tasa de crecimiento anual de On The Double.
- c) Pronostique el número de centros de copiado en operación para 2012. ¿Parece razonable este número? ¿Por qué?
19. En la tabla P-19 se presenta el número de empleados (X) y las utilidades por empleado (Y) para $n = 16$ empresas editoriales. Los empleados se registran en miles, y las utilidades por

TABLA P-18

<i>Año</i>	<i>Periodo</i>	<i>Centros de copiado</i>
1993	1	1
1994	2	2
1995	3	2
1996	4	6
1997	5	10
1998	6	16
1999	7	25
2000	8	41
2001	9	60
2002	10	97
2003	11	150
2004	12	211
2005	13	382
2006	14	537

TABLA P-19

<i>Empresa</i>	<i>Utilidades (en miles de \$)</i>	<i>Empleados (en miles)</i>
1	33.5	9.4
2	31.4	6.3
3	25.0	10.7
4	23.1	7.4
5	14.2	17.1
6	11.7	21.2
7	10.8	36.8
8	10.5	28.5
9	9.8	10.7
10	9.1	9.9
11	8.5	26.1
12*	8.3	70.5
13	4.8	14.8
14	3.2	21.3
15	2.7	14.6
16	-9.5	26.8

*Dun and Bradstreet.

empleado en miles de dólares. A continuación se presenta una parte del resultado de Minitab de un análisis de regresión lineal.

The regression equation is

Profits = 18.0 - 0.271 Employees

Predictor	Coef	SE Coef	T	P
Constant	17.954	4.457	4.03	0.001
Employees	-0.2715	0.1726	-1.57	0.138

s = 10.6112 R - Sq = 15.0% R - Sq(adj) = 9.0%

- a) Identifique la estimación de los mínimos cuadrados de los coeficientes de la pendiente y de la intersección.
- b) Pruebe la hipótesis $H_0: \beta_1 = 0$ con $\alpha = .10$. ¿Parece haber una relación entre las utilidades por empleado y el número de empleados? Explique.
- c) Identifique e interprete r^2 .
- d) ¿Esta función de regresión ajustada es una buena herramienta para pronosticar las utilidades por empleado para un número dado de empleados de una empresa editorial? Discútalos.
20. Con respecto al problema 19, la observación 12 corresponde a Dun y Bradstreet. Haga de nuevo el análisis de regresión lineal omitiendo esta observación. ¿Cambian sus conclusiones de los incisos b) y d) del problema 19? ¿Qué implica esto, si acaso, con respecto a la influencia de una observación individual sobre el análisis de regresión cuando el número de observaciones es bastante pequeño? ¿Cree usted razonable despedir a Dun y Bradstreet? Analícelo.
21. En el contexto del ejemplo 2-13 se presenta la tabla 2-28 con los datos sobre el costo real (Y) y el costo estimado (X), en millones de dólares, para $n = 26$ proyectos de construcción. La figura 2-17 muestra la gráfica de la línea ajustada para estos datos.
- a) Reanalice los datos de costos usando un modelo de regresión lineal simple. Elabore una gráfica de los residuos contra los valores ajustados.
- b) ¿Es significativa la regresión del costo real sobre el costo estimado? Justifique su respuesta.
- c) Identifique e interprete r^2 .
- d) ¿Cuáles son los valores apropiados de los coeficientes de la pendiente y de la intersección de la función de regresión de la población si el costo estimado es un factor de predicción perfecto del costo real esperado? ¿Son congruentes los coeficientes estimados en la ecuación de la línea recta ajustada con estos valores? Discútalos.
- e) Considere la gráfica de los residuos contra los valores ajustados. ¿Parece como si los costos estimados fueran, en general, factores de predicción más exactos de los costos reales de los proyectos relativamente baratos que de los proyectos de alto costo? ¿Por qué?
22. Con respecto al problema 21, los datos de costo pueden transformarse tomando los logaritmos naturales tanto de los costos estimados como de los costos reales. Considere un modelo de regresión lineal simple para los costos transformados. A continuación se presenta una porción del resultado de Minitab para este análisis.

The regression equation is

$\text{LnActual} = 0.003 + 0.968 \text{ LnEstimate}$

Predictor	Coef	SE Coef	T	P
Constant	0.0026	0.1625	0.02	0.987
LnEstimate	0.96811	0.06583	14.71	0.000

$S = 0.412124$ $R - \text{Sq} = 90.0\%$ $R - \text{Sq}(\text{adj}) = 89.6\%$

- a) ¿Es significativa la regresión? Explique.
- b) Identifique e interprete r^2 .
- c) ¿Los datos muestrales son congruentes con los valores de la población, $\beta_0 = 0$ y $\beta_1 = 1$? Explique. ¿Cuál es la significancia de estas opciones particulares para β_0 y β_1 ?
- d) Elabore un pronóstico del costo real cuando el costo estimado es de \$24 millones.

CASOS

CASO 6-1 TIGER TRANSPORT

Tiger Transport Company es una empresa de transporte que traslada enseres domésticos tanto localmente como a nivel nacional. Actualmente está interesada en el precio cobrado por transportar cargas pequeñas a largas distancias. La empresa está satisfecha con sus tarifas para cargas completas; estos precios están basados en los costos variables de los conductores, el combustible y el servicio de mantenimiento, más los costos indirectos y las utilidades. Sin embargo, hay dudas sobre la tarifa apropiada que debe cobrar por las mercancías adicionales necesarias para llenar el camión. Para realizar el pronóstico de las necesidades futuras de combustible y elaborar presupuestos de largo plazo, la compañía desea determinar el costo de agregar carga a un camión parcialmente lleno.

La compañía considera que el único costo adicional en que se incurre si se agrega carga al camión es el costo del combustible extra, pues se reducirían las millas que recorre el camión por galón. Como uno de los factores que se usarían para determinar las tarifas de cargas pequeñas, a la compañía le gustaría conocer el monto de los costos asociados al consumo adicional de combustible.

Usted está recién graduado de la escuela de negocios y trabaja en el departamento de contabilidad de costos; se le asigna la tarea de investigar este problema y comunicar a la alta dirección las consideraciones

necesarias para tomar una decisión sensata sobre las tarifas. Usted inicia suponiendo que todos los camiones son iguales; de hecho, son casi idénticos en términos de tamaño, capacidad de peso bruto y tamaño del motor. Usted también supone que todos los conductores recorrerán el mismo millaje en un trayecto largo. El jefe del departamento de contabilidad de Tiger considera que estas suposiciones son razonables.

Entonces usted tiene sólo una variable que podría afectar el millaje por galón para camiones en grandes distancias: el peso de la carga. Usted encuentra que el departamento de contabilidad tiene registros de cada viaje realizado por un camión de Tiger durante varios años. Estos registros incluyen el peso total de la carga, la distancia recorrida y el número de galones de diesel consumidos. Estas dos últimas cifras conforman una razón: las millas por galón para el viaje.

Usted se lecciona como su población los viajes realizados durante los últimos cuatro años; hay un total de 5,428 viajes. Entonces usted selecciona 40 números de una tabla de números aleatorios y puesto que los viajes están registrados uno tras otro, usted asigna el 1 al primer viaje, 2 al segundo y así sucesivamente. De esta forma, sus 40 números aleatorios producen una selección aleatoria de 40 viajes que se van a examinar. El peso de la carga y las millas por galón aparecen registrados en la tabla 6-12.

TABLA 6-12 Datos del peso de la carga y millas por galón de Tiger Transport

<i>Peso</i> (miles de libras)	<i>Millas</i> por galón	<i>Peso</i> (miles de libras)	<i>Millas</i> por galón	<i>Peso</i> (miles de libras)	<i>Millas</i> por galón	<i>Peso</i> (miles de libras)	<i>Millas</i> 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 Resultados de computadora del análisis de regresión de 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			

Puesto que su computadora de escritorio tiene el software de un paquete de análisis de regresión, usted ajusta un modelo de regresión lineal simple a los datos de la tabla 6-12. El reporte resultante se presenta en la tabla 6-13.

Después de estudiar el reporte en la tabla 6-13, usted concluye que los datos muestrales han generado una ecuación de regresión útil. Esta conclusión está basada en una r^2 relativamente alta (76%), un valor grande negativo de t (-10.9) y un valor alto de F (119). Del reporte de resultados usted obtiene la ecuación de la línea ajustada

$$\hat{Y} = 8.8484 - .0604X$$

donde Y se mide en millas por galón y X en miles de libras. La pendiente de la ecuación de regresión ($-.0604$) se interpreta como sigue: *Cada 1,000 libras adicionales de carga reducen el millaje de un camión un promedio de .0604 millas por galón.*

La compañía paga aproximadamente \$2.55 por galón de combustible diesel. Por lo tanto, usted puede calcular el costo de transportar una carga de 1,000 libras adicionales una distancia de 100 millas de la siguiente manera:

Promedio de millas por galón = 4.7 (de la tabla 6-12)

$$\text{Costo de 100 millas} = \frac{100(2.55)}{4.7} = \$54.26$$

Costo del mismo viaje con carga adicional de

$$1,000 \text{ libras es } \frac{100(2.55)}{(4.7 - .0604)} = \$54.96$$

Por lo tanto,

$$\begin{aligned} \text{Aumento de costo de 1,000 libras transportadas} \\ 100 \text{ millas} &= \$.70 \end{aligned}$$

Ahora usted cree que ha completado parte de su tarea, a saber, la determinación de los costos en efectivo de combustible asociados con el peso de la carga adicional de un camión que está parcialmente lleno. Desde luego, usted se da cuenta de que hay otros factores relacionados con una decisión de tarifa para cargas pequeñas.

TAREA

1. Redacte un escrito para la alta dirección de la compañía que resuma el análisis. Incluya comentarios sobre la forma en la cual su trabajo mejorará

los pronósticos de necesidades de combustible e ingresos por camión.

CASO 6-2 BUTCHER PRODUCTS, INC.

Gene Butcher es dueño y presidente de Butcher Products, Inc., una compañía pequeña que fabrica tubos de fibra de vidrio para cables de instalaciones eléctricas. Gene ha estudiado el número de unidades manufacturadas diariamente durante los dos años y medio anteriores y está preocupado por la amplia variabilidad de estas cifras. Para pronosticar adecuadamente el rendimiento de producción, los costos e ingresos, Gene necesita establecer una relación entre la producción y alguna otra variable.

Con base en su experiencia en la compañía, Gene no puede identificar una razón para la variabilidad en la producción sino hasta que piensa en las condiciones del estado del tiempo. Su razonamiento es que la temperatura exterior puede tener algo que ver con la productividad de su fuerza de trabajo y la producción diaria obtenida.

Él selecciona de manera aleatoria varios días de sus archivos y registra los números de unidades de conductos producidos para cada uno de esos días. Luego, va a la estación meteorológica local y, para cada uno de los días seleccionados, registra la temperatura más alta del día. Ahora está listo para efectuar un estudio de correlación entre estas dos variables al darse cuenta de que la producción probablemente está relacionada con la desviación con respecto a una temperatura ideal y no con la temperatura misma. Es decir, él piensa que un día que es muy caluroso o muy frío tendría un efecto negativo sobre la producción cuando se le compara con un día que tiene una temperatura ideal. Él decide convertir sus lecturas de la temperatura en desviaciones con respecto a los 65 grados Fahrenheit, una temperatura que considera ideal para la obtención de un alto rendimiento. Sus datos aparecen a continuación; Y representa el número de unidades producidas,

mientras que X representa la diferencia absoluta (se eliminaron los signos negativos) entre la temperatura más alta del día y los 65 grados ideales:

Y	X	Y	X
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 ejecuta un análisis de regresión lineal simple usando la computadora de su compañía y el programa de Minitab.

Gene está complacido de ver los resultados de su análisis de regresión, tal como se presentan en la tabla 6-14. Los valores t correspondientes a los coeficientes de la intersección estimada y la pendiente son grandes (en valor absoluto) y sus valores p son muy pequeños. Ambos coeficientes (552 y -8.9) en la ecuación de regresión de la muestra son claramente significativos.

Al considerar r^2 , Gene se siente algo decepcionado por encontrar que este valor, si bien es satisfactorio, no es tan alto como esperaba (64.2%). Sin embargo, determina que es lo suficientemente alto para empezar a pensar en formas de incrementar los niveles de la producción diaria.

TABLA 6-14 Pantalla de resultado del análisis de regresión de Butcher Products, Inc.

Regression Analysis: Y versus X

The regression equation is

$$Y = 552 - 8.91 X$$

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	132,758	132,758	37.62	0.000
Residual Error	21	74,109	3,529		
Total	22	206,866			

PREGUNTAS

1. ¿Cuántas unidades pronosticaría usted para un día en el cual la temperatura más alta es de 89 grados Fahrenheit?
2. ¿Cuántas unidades pronosticaría usted para un día en el cual la temperatura más alta es de 41 grados Fahrenheit?
3. Con base en los resultados del análisis de regresión que se presentó antes, ¿qué acción aconsejaría usted a Gene para incrementar la producción diaria?
4. ¿Cree usted que Gene ha desarrollado una herramienta de pronóstico efectiva?

CASO 6-3 ACE MANUFACTURING

Ace Manufacturing Company da empleo a varios miles de personas en la fabricación de teclados, estuches para equipo y cables para la industria de las computadoras pequeñas. Recientemente, el presidente de Ace ha estado preocupado por la tasa de ausentismo entre los empleados de la compañía y pidió al departamento de personal que investigue este asunto. El departamento de personal se da cuenta de que un método efectivo para pronosticar el ausentismo fortalecería mucho su capacidad para planear adecuadamente.

Bill McGone, el director de personal, decide echar un vistazo a unos cuantos expedientes del personal en un intento por evaluar el problema. Decide seleccionar 15 expedientes de manera aleatoria y registra el número de faltas durante el año fiscal pasado, junto con la edad del empleado. Después de leer un artículo en una revista especializada sobre personal, McGone cree que la edad podría tener un efecto significativo en el ausentismo. Si descubre que la edad y las faltas mani-

fiestan una buena correlación en su pequeña muestra, tomará una muestra de 200 o 300 expedientes y formulará una buena ecuación de predicción.

La siguiente tabla contiene los valores de los datos recopilados en la muestra inicial. El número de faltas durante el pasado año fiscal está representado por Y , 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 están correlacionados las faltas al trabajo y la edad? ¿Podría generalizarse esta correlación para la fuerza laboral completa?
2. ¿Cuál es la ecuación para pronosticar los días de ausentismo usando la edad como variable de predicción?
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 las faltas al trabajo y la edad? Al responder esta pregunta use los procedimientos estadísticos pertinentes para fundamentar su respuesta.
5. Suponga que una persona recién contratada tiene 24 años. ¿Cuántas faltas pronosticaría usted en el caso de esta persona durante el año fiscal?
6. ¿Debe proceder Bill McGone a tomar una muestra más grande de los empleados de la compañía con base en los resultados preliminares de su muestra?
7. ¿Se ha desarrollado un método efectivo de pronóstico?

CASO 6-4 MR. TUX

John Mosby ha escuchado que el análisis de regresión a menudo se usa para pronosticar las variables de las series de tiempo y, como tiene una computadora personal con un paquete de software de regresión, decide intentarlo. Los volúmenes de ventas mensuales de Mr. Tux de 1998 a 2005 son la variable dependiente. Como primer intento, decide usar el número del periodo como el factor de predicción, o variable X . Su primer valor de ventas Y , \$6,028, tendrá un valor X de 1, el segundo tendrá un valor de X de 2, etcétera. Su razonamiento es que la tendencia ascendente que él sabe que existe en sus datos será tomada en cuenta usando un valor siempre creciente de X para explicar sus datos de ventas.

Después de realizar el análisis de regresión en su computadora, John registra los siguientes valores:

$$t = 11.01 \quad r^2 = .563$$

$$F = 121.14 \quad \hat{Y} = -6,495 + 2,729.2X$$

El valor alto de t le indica a John que la pendiente de su línea de regresión ajustada (2,729.2) es significativa; es decir, rechaza la hipótesis nula porque la pendiente de la línea de regresión de la población es cero. El valor alto de F es congruente con este resultado

(John recuerda que $F = t^2$ para la regresión de línea recta), y la hipótesis nula de que la regresión no es significativa debe rechazarse.

John está decepcionado con el valor relativamente bajo de r^2 (56.3%). Él esperaba un valor más alto de manera que su ecuación de regresión simple pudiera utilizarse para pronosticar con exactitud sus ventas. Se da cuenta de que este valor bajo puede ser resultado de la estacionalidad de sus ventas mensuales, un hecho que conocía antes de iniciar sus actividades de pronóstico. Una variación estacional considerable tendría como resultado que los puntos de datos mensuales no se aglomerasen alrededor de la línea de regresión lineal, arrojando un valor insatisfactorio de r^2 .

Hay algo más que preocupa a John acerca de los resultados de su regresión. En el reporte de resultados de la computadora se encuentra este enunciado: Durbin-Watson = .99. John no entiende este resultado y llama a su profesor de estadística de la universidad. Después de que le describe los valores de su reporte de resultados de la regresión, el profesor le dice: “Tengo una clase en este momento, pero el bajo valor del estadístico de Durbin-Watson significa que uno de los supuestos de tu análisis de regresión no se cumple”.

PREGUNTAS

1. Comente acerca de la creencia de John de que sus ventas mensuales son altamente estacionales y, por lo tanto, traen consigo un valor “bajo” de r^2 .
2. ¿Cuál es su opinión con respecto a la idoneidad del método de pronóstico de John?
3. ¿De qué manera violan los datos de John uno de los supuestos del análisis de regresión?

CASO 6-5 CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se describió en el caso 1-2.

El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de nuevos clientes que serían atendidos durante el resto de 1993. Marv proporcionó a Dorothy Mercer los datos mensuales del

número de clientes nuevos atendidos por CCC en el periodo de enero de 1985 a marzo de 1993 (véase el caso 3-3). En el caso 3-3, Dorothy utilizó el análisis de autocorrelación para explorar el patrón de datos. En el caso 4-3 utilizó los métodos de promedios móviles y de suavización exponencial para pronosticar los meses restantes de 1993.

Año	Ene.	Feb.	Mar.	Abr.	Mayo	Jun.	Jul.	Ago.	Sept.	Oct.	Nov.	Dic.
1985	110	108	105	104	104	102	103	101	102	102	99	99
1986	102	105	106	107	105	106	105	105	103	105	103	101
1987	108	105	105	108	104	108	105	105	103	105	104	104
1988	104	109	109	103	103	104	99	102	101	101	102	102
1989	107	106	109	105	104	103	103	106	105	106	107	99
1990	103	106	110	108	110	105	105	106	107	107	111	112
1991	114	116	118	119	118	120	120	121	119	121	119	120
1992	122	118	123	118	118	120	122	120	122	123	124	122
1993	125	125	130									

Dorothy se pregunta si podría utilizar el análisis de regresión para desarrollar un buen modelo de pronóstico. Le pidió a Marv que pensara en algunas variables posibles de predicción. Marv sentía que el número de personas con cupones canjeables por alimentos podría estar relacionado con el número de clientes nuevos atendidos.

Dorothy sólo pudo obtener los datos del número de personas con cupones canjeables por alimentos a partir de enero de 1989 y hasta diciembre de 1992. A continuación se presentan esos datos.

Marv también estaba familiarizado con un índice de actividad de negocios calculado para el condado por el Consejo de Desarrollo Económico local. El índice de

1989	24,450 25,163	24,761 25,184	25,397 25,417	25,617 25,411	25,283 25,565	25,242 26,543
1990	26,784 28,161	27,044 27,936	27,567 28,423	28,080 28,366	28,142 29,029	28,412 29,035
1991	29,254 30,863	29,962 31,288	30,499 31,492	30,879 31,577	30,995 31,912	31,356 32,050
1992	32,383 33,556	32,625 33,751	33,499 33,777	34,076 33,769	34,191 34,077	33,788 34,232

actividad económica era un indicador de los cambios relativos en las condiciones generales de los negocios para la región. Los datos de este índice se encuentran en la parte superior de la página.

PREGUNTAS

1. Determine si existe una relación significativa entre el número de clientes nuevos atendidos y el número de personas con cupones canjeables por alimentos y/o el índice de actividad de negocios. No olvide la posibilidad de transformar los datos.

2. Desarrolle una ecuación de regresión y utilícela para pronosticar el número de clientes nuevos 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 predicción del número de clientes nuevos?

5. Los datos consisten en una serie de tiempo. ¿Significa esto que se ha violado el supuesto de la independencia?

6. Suponga que usted desarrolló una buena ecuación de regresión. ¿Usaría usted esta ecuación para hacer un pronóstico del resto de 1993? Explique su respuesta.

CASO 6-6 AAA WASHINGTON

En el caso 5-5 se dio una visión general de AAA Washington cuando se pidió a los estudiantes que prepararan la descomposición de una serie de tiempo de las llamadas de servicio de emergencia en el camino recibidas por el club durante cinco años. La descomposición de la serie de tiempo realizada en el caso 5-5

demostró que el patrón que Michael DeCoria había observado en el volumen de llamadas de emergencia en el camino probablemente era de naturaleza cíclica. A Michael le gustaría pronosticar el volumen de llamadas de emergencia en el camino para años futuros.

Otra investigación hecha por el club identificó varios factores que tenían efecto en el volumen de llamadas al servicio de emergencia en el camino. Entre estos factores se encuentran el promedio de temperatura diaria y la cantidad de lluvia registrada en un día. Esta investigación indica que las llamadas al servicio de emergencia en el camino se incrementan conforme aumentan las precipitaciones y conforme declina la temperatura del día. El club también cree que el número total de llamadas de emergencia que recibe es

dependiente del número de miembros en el club. Finalmente, Michael cree que el número de llamadas recibidas está relacionado con el ciclo general de la economía. La tasa de desempleo para el estado de Washington se utiliza como una buena medición sustituta de la economía general del estado de Washington. En la tabla 6-15 se incluyen los datos correspondientes a la tasa de desempleo, la temperatura mensual promedio, la precipitación mensual y el número de miembros del club.

TABLA 6-15 Datos para AAA Washington

<i>Año</i>	<i>Mes</i>	<i>Llamadas</i>	<i>Tasa de desempleo (%)</i>	<i>Temp. (° F)</i>	<i>Lluvia (pulgadas)</i>	<i>Miembros</i>
1988	Mayo	20,002	5.7867	55.1	3.75	
	Junio	21,591	5.7592	59	1.95	
	Julio	22,696	5.5718	63.8	0.89	
	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	5.863	45.4	8.42	393,115
	Diciembre	23,529	6.1349	41	4.44	392,631
1989	Enero	23,327	7.5474	40.3	4.3	396,975
	Febrero	24,050	7.8157	34.3	3.18	395,186
	Marzo	24,010	7.139	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	1.3	401,949
	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	41.8	4.72	415,342
1990	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	3.4	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
	Julio	20,804	4.2559	66.7	0.59	437,969
	Aosto	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

TABLA 6-15 (continuación)

<i>Año</i>	<i>Mes</i>	<i>Llamadas</i>	<i>Tasa de desempleo (%)</i>	<i>Temp. (° F)</i>	<i>Lluvia (pulgadas)</i>	<i>Miembros</i>
1991	Diciembre	24,861	5.5102	33.9	7.93	446,455
	Enero	23,441	6.8901	37.9	4.4	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	6.128	49.1	5.69	446,430
	Mayo	19,077	5.8146	54.3	2.12	450,001
	Junio	19,141	5.948	58.2	1.61	452,303
	Julio	20,883	5.9026	65.4	0.51	456,551
	Agosto	20,709	5.7227	66	2.8	455,747
	Septiembre	19,647	5.6877	60.9	0.2	456,764
	Octubre	22,013	6.2922	51	1.7	462,340
	Noviembre	22,375	7.0615	46.2	6.5	460,492
1992	Diciembre	22,727	7.437	42.4	3.45	465,361
	Enero	22,367	8.4513	43	7.26	465,492
	Febrero	21,155	8.7699	46	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	0.6	459,019
	Junio	20,276	7.0478	63.6	1.84	463,665
	Julio	20,795	7.108	64.9	1.41	463,775
	Agosto	21,126	6.7824	65	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	
1993	Diciembre	26,039	8.246	37.8	4.38	
	Enero	26,127	9.5301	34.9	4.08	
	Febrero	20,067	9.279			
	Marzo	19,673	8.6802			
	Abril	19,142	7.7815			

Una conversación con el director del centro de llamadas del servicio de emergencia en el camino condujo a dos observaciones importantes: **1.** Los fabricantes de automóviles parecen diseñar vehículos que funcionan mejor a 65 grados Fahrenheit y **2.** el volumen de llamadas parece incrementarse notoriamente cuando la temperatura promedio disminuye unos

cuantos grados a partir de una temperatura promedio en el rango de los 30 grados en la escala que cuando se registra una disminución similar a partir de una temperatura promedio en el rango de los 60 grados en la escala. Esta información sugiere que el efecto de la temperatura sobre el servicio de emergencia en el camino no es lineal.

PREGUNTAS

1. Corra cuatro modelos de regresión lineal simple usando el número total de llamadas al servicio de emergencia en el camino como la variable dependiente, y la tasa de desempleo, la temperatura, la lluvia y el número de miembros como las cuatro variables independientes. ¿Cualquiera de estas cuatro variables independientes sería útil para la predicción del número total de llamadas al servicio?
2. Establezca una nueva variable de temperatura y relaciónela 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 funcionar mejor a los 65 grados Fahrenheit, entonces cada grado por arriba o por debajo de los 65 grados haría que los vehículos operaran menos confiablemente. Para llevar a cabo una transformación de los datos de temperatura que simule este efecto, comience por restar los 65 grados de los valores promedio mensuales de la temperatura. Esto reposiciona el “cero” en los 65 grados Fahrenheit. ¿Se deben usar los valores absolutos de esta nueva variable de temperatura?
3. Desarrolle un diagrama de dispersión. ¿Existe una relación lineal entre las llamadas y la nueva variable de la temperatura?
4. Si existe una relación no lineal entre las llamadas y la nueva variable de la temperatura, desarrolle el mejor modelo.

Aplicaciones de Minitab

El problema. En el ejemplo 6.2, el señor Bump quería correr un análisis de regresión con los datos presentados en la tabla 6-1.

Solución de Minitab

1. Introduzca los datos de la tabla 6-1 en la hoja de cálculo: los galones de las ventas de leche van en la columna C1 y los precios de venta en la columna C2.
2. Para correr un modelo de regresión, haga clic en los siguientes menús:
Stat>Regression>Regression
3. Aparecerá la ventana de diálogo de Regression que se presenta en la figura 6-18.
 - a) Se selecciona Sales como respuesta (Response) o variable dependiente.
 - b) Se selecciona Price como factor de predicción (Predictor) o variable independiente.
4. Para almacenar los residuos y las estimaciones $\hat{Y}$, haga clic en Storage. Aparece la ventana de diálogo Regression-Storage presentada en la figura 6-19..
 - a) Haga clic en Residuals debajo de Diagnostic Measures para almacenar los residuos en C3.
 - b) Haga clic en Fits debajo de Characteristics of Estimated Equation para almacenar los valores pronosticados de $\hat{Y}$ en C4.
 - c) Haga clic en OK para cerrar la ventana de diálogo Regression-Storage.
5. Para correr gráficas de residuos, haga clic en Graphs en la ventana de diálogo de Regression. Aparecerá la ventana de diálogo Regression Graphs presentada en la figura 6-20.
 - a) Haga clic en Four in one para incluir las cuatro gráficas.
 - b) Haga clic en OK para cerrar la ventana de diálogo de Regression-Graphs.
 - c) Haga clic en OK en la ventana de diálogo Regression y se presentará el análisis de regresión desplegado en la tabla 6-8 en la ventana de sesión y sobre la pantalla aparecerá la gráfica de la figura 6-9.

El problema. En el ejemplo 6.11 Gilbert García quería pronosticar las ventas usando los gastos de publicidad.

Solución de Minitab

1. Introduzca los datos de la figura 6-10 en la hoja de cálculo: las ventas van en la columna C1 y los gastos de publicidad en la columna C2.

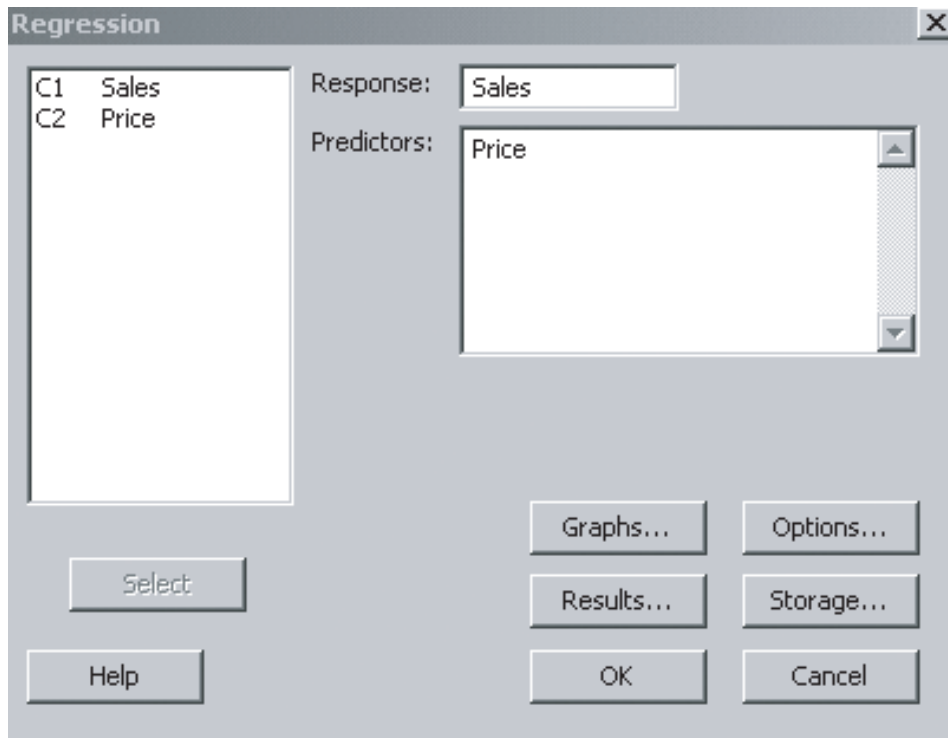


FIGURA 6-18 Ventana de diálogo Regression de Minitab

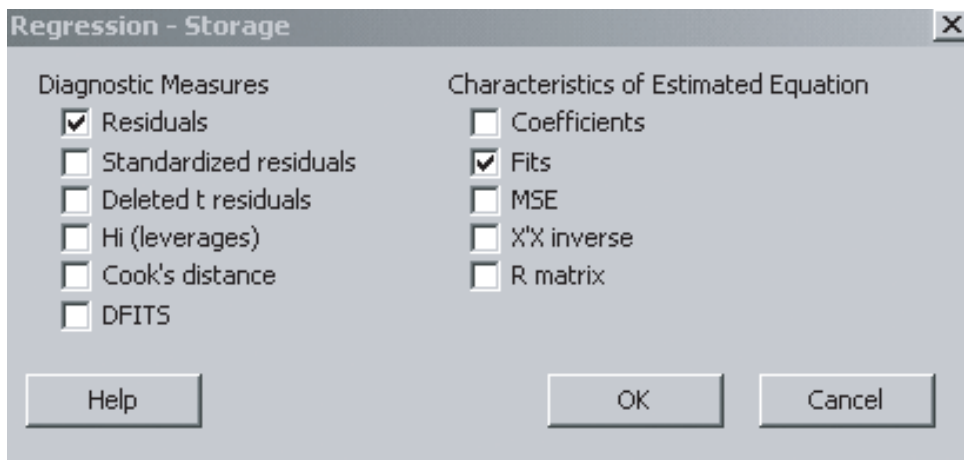


FIGURA 6-19 Ventana de diálogo Regression-Storage de Minitab

2. Para desarrollar el diagrama de dispersión que se observa en la figura 6-11, haga clic en Graph>Scatterplot
3. Aparecerán unas opciones de Scatterplots. Seleccione Simple y haga clic en OK.
4. Aparecerá la ventana de diálogo Scatterplot-Simple.

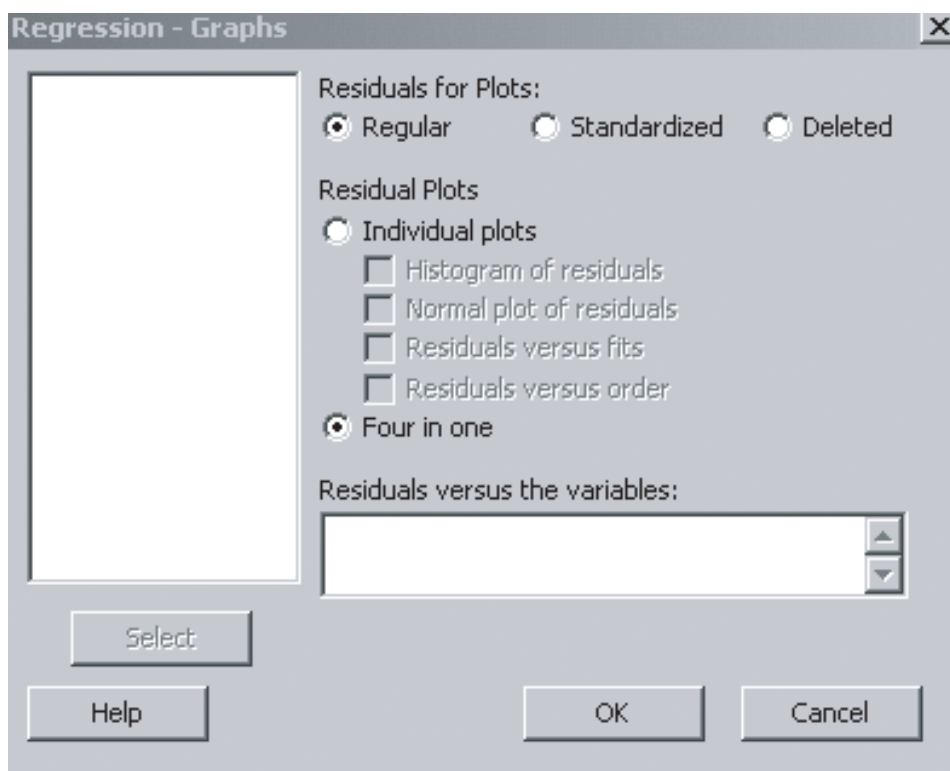


FIGURA 6-20 Ventana de diálogo Regression-Graphs de Minitab

- a) Seleccione C1, Sales, como la variable Y y C2, gastos de publicidad, como la variable X .
- b) Haga clic en OK y aparecerá el diagrama de dispersión que se aprecia en la figura 6-11.
5. Para correr un modelo ajustado como el de la figura 6-12, haga clic en
`Stat>Regression>Fitted Line Plot`
6. Aparecerá la ventana de diálogo Fitted Line Plot.
 - a) La variable dependiente Response (Y) es Sales.
 - b) La variable independiente o Predictor (X) es Advertising Expenditures.
 - c) El tipo de modelo de regresión es lineal (Linear).
 - d) Haga clic en OK. Los resultados se muestran en la figura 6-12.
7. Luego, convierta la variable X al logaritmo natural de X haciendo clic en los siguientes menús:
`Calc>Calculator`
8. Aparecerá la ventana de diálogo Calculator que se observa en la figura 6-21.
 - a) Introduzca C3 en el espacio junto a Store result in variable.
 - b) Para ejecutar la transformación, resalte Natural log (log base e) en la pantalla de Functions
 - c) Haga clic en Select y aparecerá LN (number) debajo de Select en el espacio de Expression.
 - d) Puesto que los gastos de publicidad constituyen la variable que se va a transformar, C2 sustituye a "number" en esta expresión.
 - e) Haga clic en OK y aparecerá el logaritmo natural de X en la celda C3 de la hoja de cálculo de los datos.

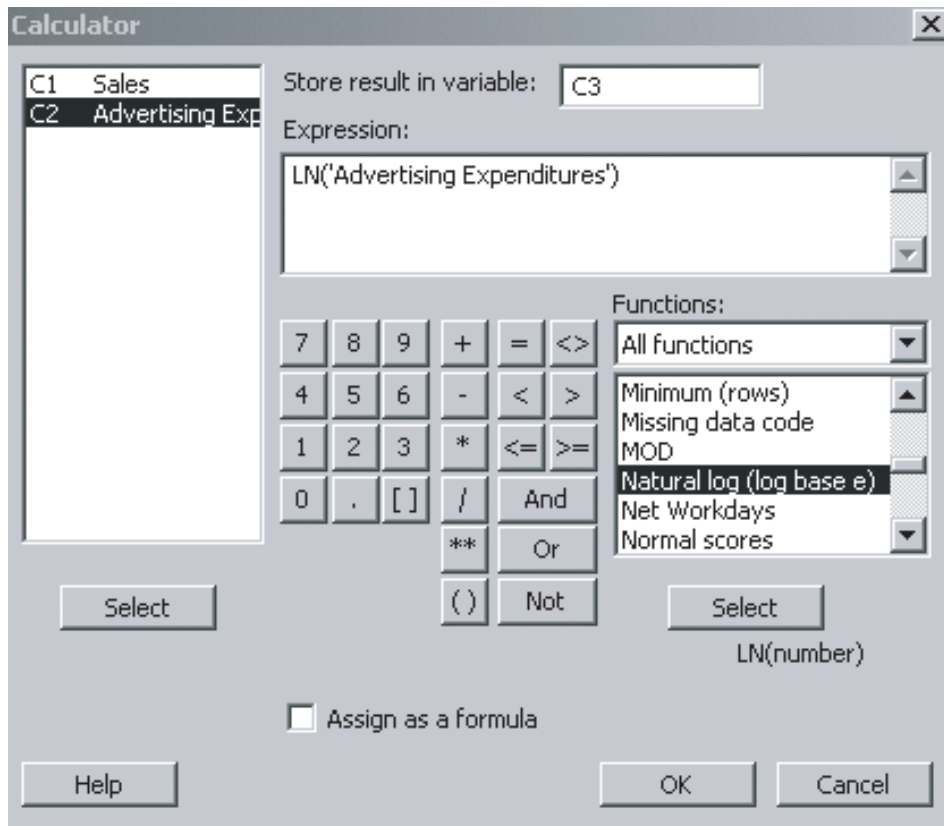


FIGURA 6-21 Ventana de diálogo Calculator de Minitab

9. Las transformaciones para la raíz cuadrada de X , el cuadrado de X y el recíproco de X también se llevan a cabo usando la ventana de diálogo Calculator.
10. La hoja de cálculo completa de Minitab se presenta en la figura 6-10.

Aplicaciones de Excel

El problema. En la situación del señor Bump, del ejemplo 6.1, se usa el análisis de regresión para determinar si el precio de venta sirve para pronosticar las ventas semanales de galones de leche.

Solución de Excel

1. Introduzca las ventas semanales (véase la tabla 6-1) de A1 a A10 y el precio de venta de B1 a B10 en la hoja de cálculo.
2. Haga clic en los siguientes menús para ejecutar el análisis de regresión:
Tools>Data Analysis
3. Aparecerá la ventana de diálogo Data Analysis. Debajo de Analysis Tools, seleccione Regression y haga clic en OK. Aparecerá la ventana de diálogo Regression de la figura 6-22.
 - a) Introduzca A1:A10 en Input Y Range.
 - b) Introduzca B1:B10 en Input X Range.
 - c) Haga clic en Output Range e introduzca C1 en el siguiente espacio.
 - d) Haga clic en OK, y aparecerá la salida presentada en la figura 6-23.

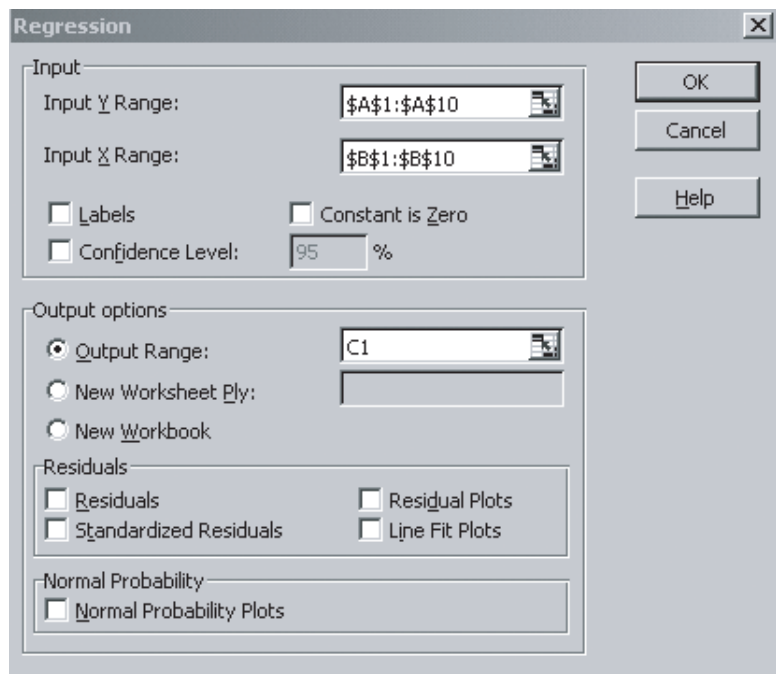


FIGURA 6-22 Ventana de diálogo Regression en Excel

Microsoft Excel - ExcelAppCh6

	A	B	C	D	E	F	G	H	I
1	10	1.3	SUMMARY OUTPUT						
2	6	2							
3	6	1.7	Regression Statistics						
4	12	1.5	Multiple R	0.863					
5	10	1.6	R Square	0.746					
6	16	1.2	Adjusted R Square	0.714					
7	5	1.6	Standard Error	2.725					
8	12	1.4	Observations	10					
9	17	1							
10	20	1.1	ANOVA						
11				df	SS	MS	F	Significance F	
12			Regression	1	174.175	174.175	23.448	0.0013	
13			Residual	8	59.425	7.428			
14			Total	9	233.6				
15									
16				Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
17			Intercept	32.136	4.409	7.289	0.0001	21.970	42.902
18			X Variable 1	14.539	3.002	-4.842	0.0013	21.462	7.615
19									
20									
21									
22									
23									
24									
25									

Ready NUM

FIGURA 6-23 Salida de Regresion de Excel para el ejemplo 6.1

Referencias

- Abraham, B. y J. Ledolter. *Introduction to regression Modeling*. Belmont, Calif.: Thomson Brooks/Cole, 2006.
- Draper, N. y H. Smith. *Applied Regression Analysis*, 3a. ed. Nueva York: Wiley, 1998.
- Flaherty, W. P. "Using Regression Analysis to Pick the Best Targets," *M&A* (marzo-abril de 1991): 47-49.
- Kutner, M. H., C. J. Nachtsheim y J. Neter. *Applied Linear Regression Models*, 4a. ed. Nueva York: McGraw-Hill, 2004.
- Moore, D. S., G. P. McCabe, W. M. Duckworth y S. L. Sclove. *The Practice of Business Statistics*. Nueva York: Freeman, 2003.

ANÁLISIS DE REGRESIÓN
MÚLTIPLE

En la regresión lineal simple se investiga la relación entre una variable independiente y una variable dependiente. A menudo la relación entre dos variables permite pronosticar con exactitud la variable dependiente a partir del conocimiento de la variable independiente. Por desgracia, muchos escenarios de elaboración de pronósticos en la vida real no son tan simples. Con frecuencia se necesita más de una variable independiente para pronosticar con exactitud una 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 utilizan en la regresión múltiple. Sin embargo, surgen algunos conceptos nuevos, ya 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.

DIVERSAS VARIABLES EXPLICATIVAS

Para ejemplificar, regresemos al problema en el cual se pronostica el volumen de las ventas de galones de leche a partir del conocimiento del precio por galón. El señor Bump enfrenta el problema de hacer un pronóstico que no es totalmente exacto. Él puede explicar casi el 75% de las diferencias en los galones vendidos de leche usando una variable independiente. Sin embargo, el 25% ($1 - r^2$) de la variación total no encuentra explicación. En otras palabras, por la evidencia muestral, el señor Bump conoce el 75% de lo que debe saber para pronosticar perfectamente el volumen de ventas. Para hacer un trabajo con mayor exactitud, necesita obtener otra variable explicativa que le permita explicar más de la variación total. Si el señor Bump puede reducir la variación no explicada, su pronóstico tendrá menos incertidumbre y más exactitud.

Es necesario investigar para conocer otra variable independiente que esté relacionada con el volumen de galones vendidos de leche. Sin embargo, esta nueva variable independiente, o predictiva, no debe estar muy estrechamente relacionada con la variable independiente que ya se está usando (el precio por galón). Si las dos variables independientes están profundamente relacionadas una con otra, explicarán la misma variación, y el hecho de agregar una segunda variable no mejorará el pronóstico.¹ En campos como la econometría y la estadística aplicada, existe gran preocupación por este problema de intercorrelación entre variables independientes, conocido a

¹Las variables explicativas interrelacionadas contienen, en esencia, la misma información y, por lo tanto, no aportan “nueva” información acerca del comportamiento de la variable dependiente. De manera ideal, los efectos de cada una de las variables explicativas sobre la variable dependiente no deben estar relacionados unos con otros.

menudo como *multicolinealidad*. La solución más sencilla al problema de dos variables independientes profundamente relacionadas consiste en *no* usarlas juntas. Estudiaremos el problema de la multicolinealidad más adelante en este capítulo

MATRIZ DE CORRELACIÓN

El señor Bump determina que los gastos en publicidad podrían ayudar a mejorar su pronóstico del volumen semanal de ventas. Él investiga las relaciones entre el gasto en publicidad, el volumen de ventas y el precio por galón examinando una matriz de correlación. La *matriz de correlación* se elabora calculando los coeficientes de correlación simple de cada combinación de pares de variables.

En la tabla 7-1 se presenta un ejemplo de una matriz de correlación. El coeficiente de correlación que indica la relación entre las variables 1 y 2 se representa como r_{12} . Observe que el primer subíndice, 1, hace referencia a la fila y el segundo subíndice, 2, hace referencia a la columna de la tabla. Esta convención permite determinar, con una mirada, la relación entre cualquier par de variables. Por supuesto, la correlación entre, digamos, la variable 1 y la variable 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 siempre es 1, de manera que, por ejemplo, $r_{11} = r_{22} = r_{33} = 1$.

El señor Bump ingresa sus datos en la computadora, y los resultados de la matriz de correlación se muestran en la tabla 7-2. Un análisis de las relaciones entre el gasto en publicidad, el volumen de ventas y el precio por galón indica que la nueva variable independiente debe contribuir a mejorar el pronóstico. La matriz de correlación indica que el gasto en publicidad tiene una relación positiva alta ($r_{13} = .89$) con la variable dependiente, el volumen de ventas, y una relación negativa moderada ($r_{23} = -.65$) con la variable independiente, el precio por galón. Esta combinación de relaciones debería permitir que los gastos en publicidad expliquen algo de la variación total del volumen de ventas que todavía no se ha explicado mediante el precio por galón. Como veremos, cuando se usan tanto el precio por galón como el gasto en publicidad para estimar el volumen de ventas R^2 aumenta al 93.2%.

El análisis de la matriz de correlación es un paso inicial importante en la resolución de cualquier problema que implique múltiples variables independientes

TABLA 7-1 Matriz de correlación

Variables	Variables		
	1	2	3
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 de los datos del señor Bump

Variables	Variables		
	Ventas 1	Precio 2	Publicidad 3
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 puede representarse mediante Y , y la variable independiente mediante X . En el análisis de regresión múltiple, las variables independientes se denotan mediante X con subíndices. Así, las variables dependientes se siguen representando por Y y las variables independientes por $X_1, X_2, \dots, X_k$. Una vez que se ha determinado el conjunto inicial de variables independientes, la relación entre Y y estas X se expresa como un *modelo de regresión múltiple*.

En el modelo de regresión múltiple, la respuesta promedio se toma como una función lineal de las variables explicativas:

$$\mu_Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (7.1)$$

Esta expresión es la *función de regresión múltiple de la población*. Al igual que en el caso de la regresión lineal simple, no podemos observar directamente la función de regresión de la población porque los valores observados de Y varían alrededor de sus medias. Cada combinación de valores para todas las X define la media de una subpoblación de respuestas Y . Suponemos que las Y de cada una de estas subpoblaciones están distribuidas de manera normal alrededor de sus medias con la misma desviación estándar, σ .

Los datos de la regresión lineal simple consisten en observaciones (X_i, Y_i) de las dos variables. En la regresión múltiple, los datos en cada caso consisten en una observación de la respuesta y una observación de cada una de las variables independientes. La i -ésima observación de la j -ésima variable explicativa se representa por X_{ij} . Con esta notación, los datos de la regresión múltiple tienen la forma que se ilustra en la tabla 7-3. Es conveniente referirse a los datos del i -ésimo caso simplemente como la i -ésima observación. Con esta convención, n es el número de observaciones y k es el número de variables explicativas.

Modelo estadístico para la regresión múltiple

La respuesta, Y , es una variable aleatoria que está relacionada con las variables independientes (predictivas), $X_1, X_2, \dots, X_k$, por

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

donde

1. Para la i -ésima observación, $Y = Y_i$ y $X_1, X_2, \dots, X_k$ se definen para los valores $X_{i1}, X_{i2}, \dots, X_{ik}$.

TABLA 7-3 Estructura de datos de la regresión múltiple

<i>Caso</i>	<i>Variables explicativas</i>				<i>Respuesta</i>
	X_1	X_2		X_k	Y
1	X_{11}	X_{12}		X_{1k}	Y_1
2	X_{21}	X_{22}		X_{2k}	Y_2
.	.	.		.	.
.	.	.		.	.
i	X_{i1}	X_{i2}		X_{ik}	Y_i
.	.	.		.	.
.	.	.		.	.
n	X_{n1}	X_{n2}		X_{nk}	Y_n

2. Las ε 's son componentes de error que representan las desviaciones de la respuesta de la relación verdadera. Son variables aleatorias no observables que explican los efectos de otros factores sobre la respuesta. Se supone que los errores son independientes y todos están distribuidos de manera normal con una media 0 y desviación estándar desconocida σ .
3. Los coeficientes de regresión, $\beta_0, \beta_1, \dots, \beta_k$, que en conjunto definen la función de regresión son desconocidos.

Teniendo los datos, es posible estimar los coeficientes de regresión usando el principio de mínimos cuadrados. Las estimaciones de los mínimos cuadrados se representan por $b_0, b_1, \dots, b_k$ y la función de regresión estimada por

$$\hat{Y} = b_0 + b_1X_1 + \dots + b_kX_k \quad (7.2)$$

Los residuos, $e = Y - \hat{Y}$, son estimaciones del componente de error y similares al escenario de la regresión lineal simple; la correspondencia entre la población y la muestra es

$$\text{Población: } Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k + \varepsilon$$

$$\text{Muestra: } Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

Por lo general, los cálculos en el análisis de regresión múltiple se realizan usando paquetes de software como Minitab y Excel (véanse las secciones de aplicaciones de Minitab y Excel al final del capítulo).

Ejemplo 7.1

Para los datos presentados en la tabla 7-4, el señor Bump considera un modelo de regresión múltiple que relaciona el volumen de ventas (Y) con el precio (X_1) y la publicidad (X_2):

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \varepsilon$$

El señor Bump determina la función de regresión ajustada:

$$\hat{Y} = 16.41 - 8.25X_1 + .59X_2$$

Los valores de los mínimos cuadrados — $b_0 = 16.41$, $b_1 = -8.25$, y $b_2 = .59$ — minimizan la suma de errores al cuadrado:

$$SSE = \sum_i (Y_i - b_0 - b_1X_{i1} - b_2X_{i2})^2 = \sum_i (Y_i - \hat{Y}_i)^2$$

para todas las opciones posibles de b_0, b_1 , y b_2 . Aquí, la función que mejor ajusta es un plano (véase la figura 7-1). Los puntos de datos se grafican en tres dimensiones a lo largo de los ejes Y, X_1 , y X_2 . Los puntos caen por encima y por debajo del plano de manera tal que $\sum (Y - \hat{Y})^2$ es un mínimo.

La función de regresión ajustada servirá para pronosticar las ventas de la siguiente semana. Si los planes requieren un precio por galón de \$1.50 y gastos de publicidad de \$1,000, el pronóstico es de 9.935 miles de galones; es decir,

$$\hat{Y} = 16.41 - 8.25X_1 + .59X_2$$

$$\hat{Y} = 16.41 - 8.25(1.5) + .59(10) = 9.935$$

TABLA 7-4 Datos del señor Bump para el ejemplo 7.1			
Semana	Ventas (en miles) Y	Precio por galón (\$) X ₁	Publicidad (en cientos de \$) X ₂
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
Medias	11.2	1.44	11.4

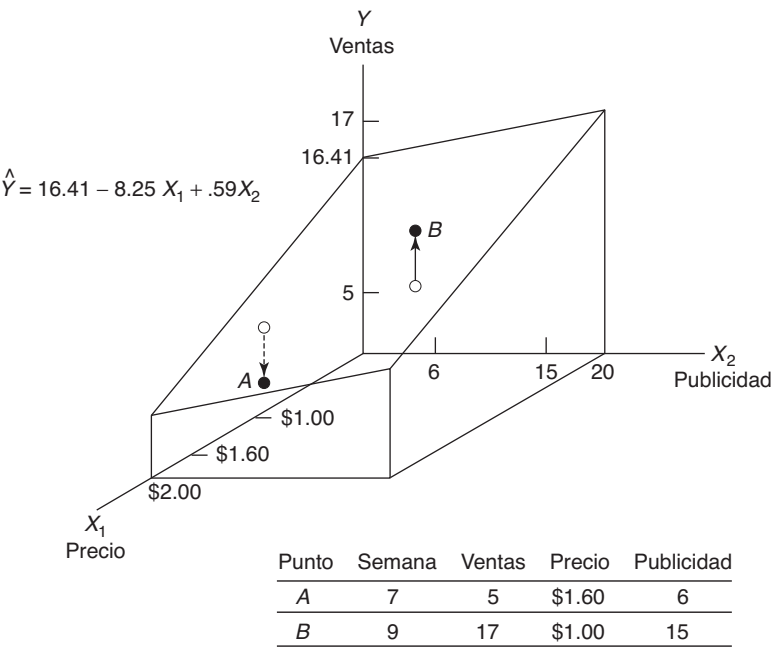


FIGURA 7-1 Plano de regresión ajustado de los datos del señor Bump; ejemplo 7.1

INTERPRETACIÓN DE LOS COEFICIENTES DE REGRESIÓN

Considere la interpretación de b_0 , b_1 , y b_2 en la función de regresión ajustada del señor Bump. Otra vez el valor de b_0 es la intersección con Y . Sin embargo, ahora se interpreta como el valor de $\hat{Y}$ cuando tanto X_1 como X_2 son iguales a cero. Los coeficientes b_1 y b_2 se conocen como los *coeficientes de la regresión parcial* o *neta*. Cada uno mide el cambio promedio de Y por unidad de cambio en las variables independientes relevantes. Sin embargo, como la función de regre-

sión está midiendo la influencia simultánea de todas las variables independientes sobre Y , el efecto parcial o neto de X_1 (o de cualquier X) debe medirse por separado de la influencia de otras variables. Por lo tanto, se dice que b_1 mide el cambio promedio de Y por unidad de cambio de X_1 , cuando se mantienen constantes las demás variables independientes.

El *coeficiente de regresión parcial o neto* mide el cambio promedio de la variable dependiente por unidad de cambio en la variable independiente relevante, cuando se mantienen constantes las demás variables independientes.

En este ejemplo, el valor b_1 de -8.25 indica que por cada incremento de 1 centavo en el precio de un galón de leche cuando los gastos de publicidad se mantienen *constantes*, la cantidad comprada se reduce un promedio de 82.5 galones. De manera análoga, el valor b_2 de $.59$ significa que, si los gastos de publicidad se incrementan \$100 cuando el precio por galón se mantiene constante, el volumen de ventas aumentará en un promedio de 590 galones.

Ejemplo 7.2

Para ilustrar los efectos netos de las X individuales sobre la respuesta, considere la situación en la cual el precio es de \$1.00 por galón y se gastan \$1,000 en publicidad. Así

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25X_1 + .59X_2 \\ &= 16.41 - 8.25(1.00) + .59(10) \\ &= 16.41 - 8.25 + 5.9 = 14.06\end{aligned}$$

El pronóstico de las ventas es de 14,060 galones de leche.

¿Cuál es el efecto sobre las ventas del incremento de un centavo en el precio si se gastan \$1,000 en publicidad?

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25(1.01) + .59(10) \\ &= 16.41 - 8.3325 + 5.9 = 13.9775\end{aligned}$$

Advierta que las ventas disminuyen 82.5 galones ($14.06 - 13.9775 = .0825$).

¿Cuál es el efecto sobre las ventas de un aumento de \$100 en publicidad si el precio permanece constante en \$1.00

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25(1.00) + .59(11) \\ &= 16.41 - 8.25 + 6.49 = 14.65\end{aligned}$$

Note que las ventas se incrementan 590 galones ($14.65 - 14.06 = .59$).

INFERENCIAS EN MODELOS DE REGRESIÓN MÚLTIPLE

La inferencia en los modelos de regresión múltiple es análoga a la de la regresión lineal simple. La estimación de mínimos cuadrados de los parámetros del modelo, la estimación de los errores estándar, las estadísticas t que se usan para examinar la significancia de los términos individuales del modelo de regresión, y el estadístico F que permite verificar la significancia de la regresión se obtienen empleando paquetes estándar de software estadístico. La determinación manual de estas cantidades en un análisis de regresión múltiple de cualquier tamaño es impráctica, y se debe usar una computadora para efectuar los cálculos.

Como en el capítulo 6, cualquier observación Y puede escribirse como

$$\text{Observación} = \text{Ajuste} + \text{Residuo}$$

o

$$Y = \hat{Y} + (Y - \hat{Y})$$

donde

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k$$

es la función de regresión ajustada. Recuerde que $\hat{Y}$ es una estimación de la función de regresión de la población. Representa la parte de Y que se explica por la relación de Y con las X . El residuo, $Y - \hat{Y}$, es una estimación del componente de error del modelo. Representa la parte de Y que no explican las variables explicativas.

La descomposición de la suma de cuadrados y los grados de libertad asociada 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 el SSE , la variación no explicada o el error. La información de la ecuación 7.3 puede obtenerse haciendo un análisis de la tabla de varianza (ANOVA), la cual se analizará en una sección posterior.

Error estándar de la estimación

El error estándar de la estimación es la desviación estándar de los residuos. Mide la dispersión típica de los valores de Y alrededor de la función de regresión ajustada.² 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 = el número de observaciones

k = el número de variables independientes en la función de regresión

$SSE = \Sigma(Y - \hat{Y})^2$ = suma de los cuadrados de los residuos

$MSE = SSE/(n - k - 1)$ = el cuadrado medio de los residuos

El *error estándar de la estimación* es la desviación estándar de los residuos. Mide la cantidad en que los valores reales de (Y) difieren de los valores estimados ($\hat{Y}$). Para muestras relativamente grandes, esperaríamos que alrededor del 67% de las diferencias $Y - \hat{Y}$ estén dentro de $s_{y \cdot x's}$ de cero, y alrededor del 95% de estas diferencias estén dentro de $2 s_{y \cdot x's}$ de cero.

Ejemplo 7.3

Las cantidades requeridas para calcular el error estándar de la estimación para los datos del señor Bump se indican en la tabla 7-5.

²El error estándar de la estimación es una estimación de σ , la desviación estándar del término de error, ε , en el modelo de regresión múltiple.

TABLA 7-5 Residuos del modelo de los datos del señor Bump para el ejemplo 7.3

<i>Y</i>	<i>X</i> ₁	<i>X</i> ₂	<i>Y</i> pronosticada (<i>Ŷ</i>) usando $\hat{Y} = 16.406 - 8.248X_1 + .585X_2$	Residuo (<i>Y</i> - <i>Ŷ</i>)	(<i>Y</i> - <i>Ŷ</i>) 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

$$s_{y \cdot x'} = \sqrt{\frac{15.90}{10 - 2 - 1}} = \sqrt{2.27} = 1.51$$

Con una sola variable explicativa X_1 = precio, el error estándar de la estimación era $s_{y \cdot x} = 2.72$ (véase el capítulo 6). Con la variable explicativa adicional, X_2 = publicidad, el señor Bump redujo el error estándar de la estimación en casi un 50%. Las diferencias entre los volúmenes reales de leche vendida y los pronósticos obtenidos a partir de la ecuación de regresión ajustada son considerablemente más pequeñas con dos variables explicativas que con una sola. Es decir, la ecuación con dos variables explicativas está mucho muy cerca de reproducir el valor real de las Y que la ecuación con una sola variable explicativa.

Significancia de la regresión

La tabla ANOVA basada en la descomposición de la variación total de Y (SST) en sus partes explicadas (SSR) y no explicadas (SSE) (véase la ecuación 7.3) se presenta en la tabla 7-6.

Considere la hipótesis $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$. Esta hipótesis significa que Y no está relacionada con ninguna de las X (el coeficiente de cada X es cero). La prueba de H_0 se conoce como prueba de *significancia de la regresión*. Si los supuestos del modelo de regresión son apropiados y H_0 es verdadera, la razón

$$F = \frac{MSR}{MSE}$$

tiene una distribución F con $df = k, n - k - 1$. De este modo, la razón F se puede usar como prueba de la significancia de la regresión.

TABLA 7-6 Tabla ANOVA de regresión múltiple

<i>Fuente</i>	<i>Suma de cuadrados</i>	<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	SST	$n - 1$		

En la regresión lineal simple, sólo existe una variable explicativa. En consecuencia, la prueba de significancia de la regresión usando la razón F de la tabla ANOVA es equivalente a la prueba t de dos colas para la hipótesis de que la pendiente de la línea de regresión es cero. 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 la significancia de todas las X de manera colectiva.

Prueba F de la significancia de la regresión

En el modelo de regresión múltiple, las hipótesis

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$H_1: \text{por lo menos una } b_j \neq 0$$

se prueban mediante la razón F :

$$F = \frac{MSR}{MSE}$$

con $df = k, n - k - 1$. Con un nivel de significancia α , la región de rechazo es

$$F > F_\alpha$$

donde F_α es el punto que deja un área de α a la derecha de 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 , está dado por

$$\begin{aligned} R^2 &= \frac{SSR}{SST} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2} \\ &= 1 - \frac{SSE}{SST} = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} \end{aligned} \quad (7.5)$$

y tiene la misma forma e interpretación que la r^2 de la regresión lineal simple. Representa la proporción de variación en la respuesta, Y , explicada por la relación de Y con las X .

Un valor de $R^2 = 1$ indica que todas las Y observadas caen exactamente sobre la función de regresión ajustada. Toda la variación en la respuesta se explica mediante la regresión. Un valor de $R^2 = 0$ dice que $\hat{Y} = \bar{Y}$ —es decir, $SSR = 0$ —y nada de la variación en Y se explica mediante la regresión. En la práctica, $0 \leq R^2 \leq 1$, y el valor de R^2 debe interpretarse dentro de los extremos 0 y 1.

La cantidad

$$R = \sqrt{R^2} \quad (7.6)$$

se llama *coeficiente de correlación múltiple* y es la correlación entre las respuestas, Y , y los valores ajustados, $\hat{Y}$. Puesto que los valores ajustados predicen las respuestas, R siempre es positiva, de manera que $0 \leq R \leq 1$.

Para la regresión múltiple,

$$F = \frac{R^2}{1 - R^2} \left(\frac{n - k - 1}{k} \right) \quad (7.7)$$

de manera que, si todo lo demás permanece igual, las regresiones significativas (razones F grandes) están asociadas con valores relativamente grandes de R^2 .

El coeficiente de determinación (R^2) siempre se incrementa al agregar una variable independiente, X , a la función de regresión, incluso si esta variable adicional no es importante.³ Por esta razón, algunos analistas prefieren interpretar R^2 ajustada al número de términos en la función de regresión. El coeficiente ajustado de determinación, $\bar{R}^2$, está dado por

$$\bar{R}^2 = 1 - (1 - R^2) \left(\frac{n - 1}{n - k - 1} \right) \quad (7.8)$$

Al igual que R^2 , $\bar{R}^2$ es una medida de la proporción de la variabilidad en la respuesta, Y , explicada por la regresión. Es posible demostrar que $0 \leq \bar{R}^2 \leq R^2$. Cuando el número de observaciones (n) es grande en relación 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 $\bar{R}^2$ y R^2 .

Ejemplo 7.4

Usando la suma total de cuadrados de la tabla 7-6 y la suma de cuadrados de los residuos del ejemplo 7.3, la descomposición de la suma de cuadrados para el problema del señor Bump es

$$\begin{aligned} SST &= SST + SSE \\ \sum (Y - \bar{Y})^2 &= \sum (\hat{Y} - \bar{Y})^2 + \sum (Y - \hat{Y})^2 \\ 233.6 &= 217.7 + 15.9 \end{aligned}$$

Por consiguiente, usando ambas formas de la ecuación 7.5 para ilustrar 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 del 93% de la variación en el volumen de ventas es explicado por la regresión, es decir, la relación entre las ventas y el precio y los gastos de publicidad. Además, la correlación entre las ventas y las ventas ajustadas es de alrededor de .965, lo que indica un acuerdo cercano entre las valores reales y los pronosticados. En la tabla 7-7 se presenta un resumen del análisis de los datos del señor Bump al respecto.

Variables explicativas individuales

El coeficiente de una X individual en la función de regresión mide el efecto parcial o neto de esa X sobre la respuesta, Y , manteniendo constantes las demás X de la ecuación. Si la regresión se considera como significativa, entonces es de interés examinar la significancia de las variables explicativas individuales. El asunto es éste: Dadas las otras X , ¿es importante el efecto de esta X particular, o es posible eliminar de la función de regresión este término X ? Esta pregunta puede contestarse examinando un valor apropiado de t .

³Aquí, “no es importante” significa “no significativo”. Es decir, el coeficiente de X no es significativamente diferente de cero (véase la sección que sigue sobre las variables individuales de explicativas).

TABLA 7-7 Resumen del análisis de los datos del señor Bump para el ejemplo 7.4

<i>Variables usadas para explicar la variabilidad de Y</i>	R^2	$\Sigma(Y - \hat{Y})^2$
Ninguna	.00	233.6
Precio	.75	59.4
Precio y gastos de publicidad	.93	15.9

Si $H_0: \beta_j = 0$ es verdadera, el estadístico de prueba, t , para $t = b_j/s_{b_j}$ tiene una distribución t con $df = n - k - 1$.⁴

Para evaluar la significancia del término j -ésimo, $j = 0, 1, \dots, k$, en la función de regresión, el estadístico de prueba t se compara con un punto porcentual de una distribución t con $n - k - 1$ grados de libertad. Para una prueba en el nivel de significancia α de

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

si $|t| > t_{\alpha/2}$ se rechaza H_0 . Aquí, $t_{\alpha/2}$ es el punto que deja un área $\alpha/2$ a la derecha de una distribución t con $df = n - k - 1$.

Se debe tener cuidado en la eliminación de la función de regresión de aquellas variables explicativas que se consideran insignificantes por la prueba t ($H_0: \beta_j = 0$ no puede rechazarse). Si las X están relacionadas (multicolinealidad), los coeficientes de los mínimos cuadrados y los valores correspondientes de t pueden cambiar, algunas veces de manera apreciable, si una X individual se elimina de la función de regresión. Por ejemplo, una X que previamente era insignificante puede convertirse en significativa. En consecuencia, si existen varios valores t pequeños (insignificantes), las variables explicativas deben eliminarse de una en una (comenzando con la variable que tiene el valor t más pequeño) en vez de hacerlo por grupos. El proceso se detiene cuando la regresión es significativa y todas las variables explicativas tienen estadísticos t grandes (significativos).

Pronóstico de una respuesta futura

Un pronóstico, $\hat{Y}^*$, de una respuesta futura, Y , para valores nuevos de las X —digamos, $X_1 = X_1^*$, $X_2 = X_2^*, \dots, X_k = X_k^*$ —está dado por la evaluación de la función de regresión ajustada para las X :

$$\hat{Y}^* = b_0 + b_1X_1^* + b_2X_2^* + \dots + b_kX_k^* \quad (7.9)$$

Con un nivel de confianza $1 - \alpha$, un intervalo de predicción para Y toma la forma

$$\hat{Y}^* \pm t_{\alpha/2} \times (\text{Error estándar del pronóstico})$$

El error estándar del 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, un intervalo de predicción *aproximado* de $100(1 - \alpha)\%$ para una nueva respuesta Y es

$$(\hat{Y}^* - t_{\alpha/2}s_{y \cdot x}, \hat{Y}^* + t_{\alpha/2}s_{y \cdot x}) \quad (7.10)$$

⁴Aquí, b_j es el coeficiente de mínimos cuadrados para la j -ésima variable explicativa, X_j , y S_{b_j} es la estimación de la desviación estándar (error estándar). Estos dos estadísticos se obtienen normalmente a través de un software como Minitab.

RESULTADO DE LA COMPUTADORA

En la tabla 7-8 se presenta el resultado de computadora para el problema del señor Bump. La revisión de resultado trae consigo las siguientes observaciones (las explicaciones aparecen en la tabla 7-8).

1. Los coeficientes de regresión son -8.25 para el precio y $.585$ para gastos de publicidad. La ecuación de regresión ajustada es $\hat{Y} = 16.4 - 8.25X_1 + .585X_2$.
2. La ecuación de regresión explica el 93.2% de la variación en el volumen de ventas.
3. El error estándar de la estimación es 1.5072 galones. Este valor es una medida de la cantidad en que los valores reales difieren de los valores ajustados.
4. En el capítulo 6, el coeficiente de la pendiente de regresión se sometió a prueba para determinar si era diferente de cero. En la presente situación, el estadístico t grande de -3.76 para la variable de precio, X_1 , y su pequeño valor p (.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 puede eliminarse de la función de regresión. De manera análoga, el estadístico t grande de 4.38 para la variable de publicidad, X_2 , y su pequeño valor p (.003) indican que el coeficiente de publicidad es significativamente diferente de cero (se rechaza $H_0: \beta_2 = 0$). Dada la variable del precio, X_1 , en la función de regresión, la variable de publicidad no puede eliminarse de la función de regresión. (Como un punto de referencia para la magnitud de los valores t , con siete grados de libertad, la tabla B-3 indica $t_{.01} = 2.998$.) En resumen, los coeficientes de ambas variables explicativas son significativamente diferentes de cero.
5. El valor p .007 es la probabilidad de obtener un valor de t por lo menos tan grande como -3.76 si la hipótesis $H_0: \beta_1 = 0$ es verdadera. Puesto que esta probabilidad es extremadamente pequeña, es improbable que H_0 sea verdadera, y se rechaza. El coeficiente del precio es significativamente diferente de cero. El valor p .003 es la probabilidad de obtener un valor t por lo menos tan grande como 4.38 si $H_0: \beta_2 = 0$ es verdadera. Puesto que un valor t de esta magnitud es extremadamente improbable, se rechaza, H_0 . El coeficiente de publicidad es significativamente diferente de cero.

TABLA 7-8 Resultado de Minitab para los datos del señor Bump

Correlations: Y, X1, X2

	Y	X1
X1	-0.863	
X2	0.891	-0.654 (6)

Regression Analysis: Y versus X1, X2

The regression equation is

$$Y = 16.4 - 8.25 X1 + 0.585 X2 \quad (1)$$

Presictor	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)

$$S = 1.50720 \quad (3) \quad R\text{-Sq} = 93.2\% \quad (2) \quad R\text{-Sq}(\text{adj}) = 91.2\% \quad (9)$$

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. La matriz de correlación se presentó en la tabla 7-2.
7. La descomposición de la suma de cuadrados, $SST = SSR + SSE$ (suma total de cuadrados = suma de cuadrados de la regresión + suma de cuadrados de los errores), se presentó en el ejemplo 7.4.
8. El valor F calculado (47.92) se utiliza para probar la significancia de la regresión. La razón grande F y su valor pequeño p (.000) indican que la regresión es significativa (se rechaza $H_0: \beta_1 = \beta_2 = 0$). La razón F se calcula mediante

$$F = \frac{MSR}{MSE} = \frac{108.85}{2.27} = 47.92$$

Como una referencia para la magnitud de la razón F , la tabla B-5 proporciona el punto superior de 1% de una distribución F con dos y siete grados de libertad como $F_{.01} = 9.55$. La función de regresión explica una cantidad significativa de la variabilidad en las ventas, Y .

9. El cálculo de la R^2 corregida o ajustada, $\bar{R}^2$, 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 qué tan bien predice el rendimiento en el trabajo una prueba de aptitud. Ocho mujeres y siete hombres realizaron la prueba, la cual mide la destreza para manipular objetos minúsculos. Cada sujeto pasó por un mes de entrenamiento intensivo como ensamblador de aparatos electrónicos, seguido de otro mes como ensamblador real, durante el cual su productividad fue evaluada por medio de un índice cuyos valores iban de 0 a 10 (cero significa improductivo).

Los datos se muestran en la tabla 7-9. Un diagrama de dispersión se presenta en la figura 7-2. Cada mujer trabajadora se representa por un cero y cada trabajador masculino se representa por un 1.

Al observar la figura 7-2 de inmediato se hace evidente que la relación de esta prueba de aptitud con el rendimiento en el trabajo sigue dos patrones distintos, uno aplicable a las mujeres y otro a los hombres.

Algunas veces es necesario determinar cómo se relaciona una variable dependiente con una variable independiente cuando un factor *cualitativo* influye en la situación. Esta relación se establece creando una *variable ficticia*. Existen muchas maneras de identificar cuantitativamente las clases de una variable cualitativa. En este texto se usan los valores 0 y 1.

Las *variables ficticias*, o indicadores ficticios, se utilizan para determinar las relaciones entre variables independientes cualitativas y una variable dependiente.

La técnica de la variable ficticia se ilustra en la figura 7-3. Los puntos de datos de las mujeres se presentan como 0; y los 1 representan a los hombres. Se construyen dos líneas paralelas en el diagrama de dispersión. La de arriba se ajusta los datos de las mujeres; la de abajo se ajusta los puntos de datos de los hombres.

Cada una de estas líneas se obtuvo a partir de una función de regresión ajustada de la forma

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

TABLA 7-9 Datos de la variable ficticia de ensambladores de aparatos electrónicos para el ejemplo 7.5

<i>Sujeto</i>	<i>Calificación del desempeño laboral Y</i>	<i>Calificación en la prueba de aptitud X_1</i>	<i>Género X_2</i>
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)
Totales	87	1,093	

$\bar{Y}_F$ = 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 para las mujeres = 64

$\bar{X}_M$ = Media del resultado de la prueba de aptitud para los hombres = 83

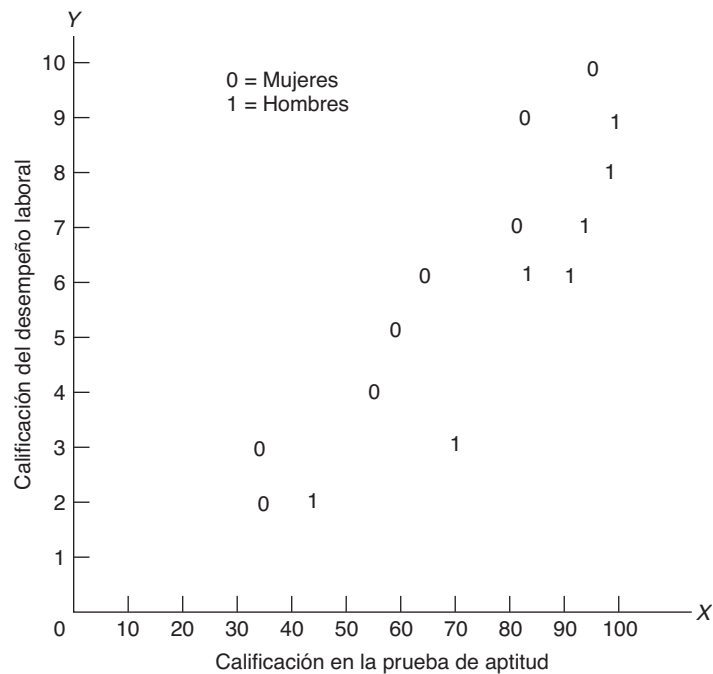


FIGURA 7-2 Diagrama de dispersión de los datos para el ejemplo 7.5

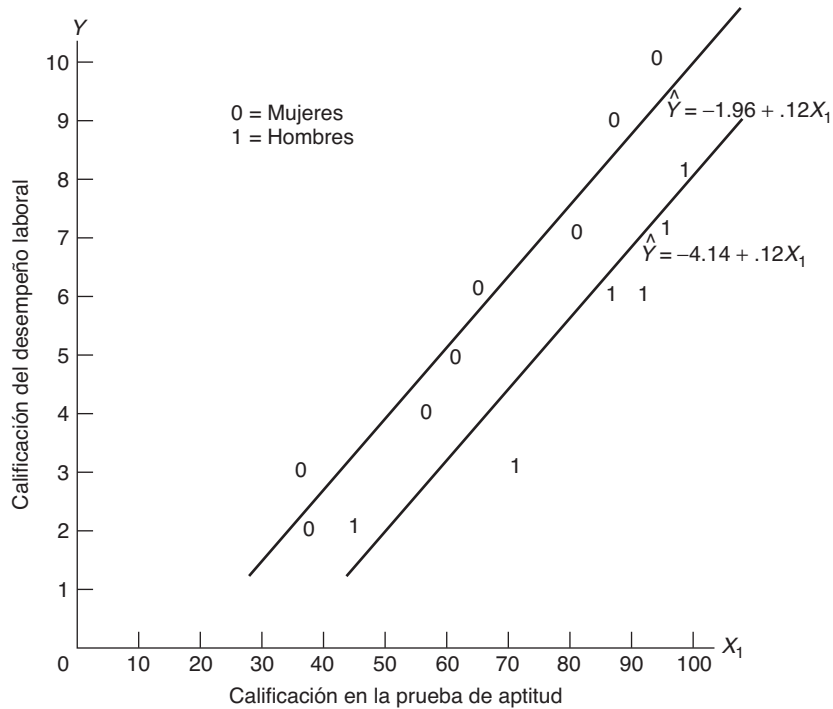


FIGURA 7-3 Líneas de regresión correspondientes a las variables ficticias para el ejemplo 7.5

donde

X_1 = calificación de la prueba

$X_2 = \begin{cases} 0 & \text{para mujeres} \\ 1 & \text{para hombres} \end{cases}$ variable ficticia

La ecuación individual es equivalente a las siguientes dos ecuaciones:

$$\hat{Y} = b_0 + b_1X_1 \quad \text{para mujeres}$$

$$\hat{Y} = b_0 + b_1X_1 + b_2 = (b_0 + b_2) + b_1X_1 \quad \text{para hombres}$$

Observe que b_2 representa el efecto de un varón sobre el rendimiento en el trabajo, y b_1 representa el efecto de las diferencias en las calificaciones de las pruebas de aptitud (se supone que el valor de b_1 es el mismo tanto para hombres como para mujeres). El punto importante es que una ecuación de regresión múltiple producirá las dos líneas de estimación mostradas en la figura 7-3. La línea de arriba es la relación estimada para mujeres, y la línea de abajo es la relación estimada para hombres. Uno podría imaginarse a X_2 como una variable “apagador” que está “encendido” cuando se hace una observación para un hombre y “apagado” cuando se hace para una mujer.

Ejemplo 7.6

La ecuación de regresión múltiple estimada para los datos del ejemplo 7.5 se presenta en los resultados de Minitab de la tabla 7-10. Ésta es

$$\hat{Y} = -1.96 + .12X_1 - 2.18X_2$$

TABLA 7-10 Resultados de Minitab del 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

$$\text{Rating} = -1.96 + 0.120 \text{ Test} - 2.18 \text{ Gender}$$

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

$$\hat{Y} = -1.96 + .12X_1 - 2.18(0) = -1.96 + .12X_1 \quad \text{para mujeres}$$

y

$$\hat{Y} = -1.96 + .12X_1 - 2.18(1) = -4.14 + .12X_1 \quad \text{para hombres}$$

Estas ecuaciones se pueden interpretar de la siguiente manera: el valor del coeficiente de regresión $b_1 = .12$, el cual es la pendiente de cada una de las líneas, es el incremento promedio estimado de la calificación del desempeño laboral por cada unidad de incremento en la calificación de la prueba de aptitud. Este coeficiente se aplica tanto a los hombres como a las mujeres.

El otro coeficiente de regresión, $b_2 = -2.18$, se aplica sólo a los hombres. Para un hombre que realiza la prueba, la calificación del desempeño laboral estimada se reduce, en relación con las mujeres que hacen la prueba, en 2.18 unidades cuando la calificación de aptitud se mantiene constante.

Un examen de los promedios de las variables Y y X_1 clasificadas por género, nos ayuda a comprender este resultado. La tabla 7-9 indica que la media de las calificaciones del desempeño laboral fueron aproximadamente iguales para los hombres, 5.86, y las mujeres, 5.75. Sin embargo, los hombres obtuvieron una calificación significativamente más alta (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 una calificación de 70, la calificación del desempeño laboral estimada de la mujer sería 2.18 más alta que la del hombre, puesto que

$$\text{Mujer: } \hat{Y} = -1.96 + .12X_1 = -1.96 + .12(70) = 6.44$$

$$\text{Hombre: } \hat{Y} = -4.14 + .12X_1 = -4.14 + .12(70) = 4.26$$

Una mirada a la matriz de correlación de la tabla 7-10 ofrece una visión interesante. Existe una relación lineal fuerte entre el rendimiento en el trabajo y la prueba de aptitud porque $r_{12} = .876$. Si solamente se usara la calificación de la prueba de aptitud para predecir el rendimiento, se explicaría alrededor del 77% ($.876^2 = .767$) de la variación en las calificaciones del rendimiento en el trabajo.

El coeficiente de correlación $r_{13} = .02$ indica que prácticamente no existe relación entre el género y el rendimiento en el trabajo. Esta conclusión también se hace evidente a partir del hecho de que los índices de rendimiento promedio para hombres y mujeres son aproximadamente iguales (5.86 contra 5.75). A primera vista, uno podría concluir que el conocimiento de que un candidato sea

hombre o mujer no es una información útil. Sin embargo, la relación moderada, $r_{23} = .428$, entre el género y la calificación en la prueba de aptitud indica que la prueba podría hacer una discriminación entre sexos. Los hombres parecen desempeñarse mejor en la prueba que las mujeres (83 contra 64). Quizás en la prueba se requiere algún elemento de fortaleza que no se requiere en el trabajo.

Cuando se usan tanto los resultados de la prueba como el género para pronosticar el rendimiento en el trabajo, se explica el 92% de la varianza. Este resultado sugiere que ambas variables hacen una contribución valiosa al rendimiento pronosticado. La calificación de la prueba de aptitud explica el 77% de la varianza, y el género considerado junto con la calificación de la prueba de aptitud agrega otro 15%. Los estadísticos t calculados, 11.86 (valor- $p = .000$) y -4.84 (valor- $p = .000$), para la calificación de la prueba de aptitud y el género, respectivamente, indican que ambas variables explicativas deben incluirse en la función final de regresión.

MULTICOLINEALIDAD

En muchos problemas de regresión, los datos se registran rutinariamente en vez de ser generados por posiciones previamente elegidas de las variables independientes. En estos casos, las variables independientes son a menudo linealmente dependientes o *multicolineales*. Por ejemplo, en el trabajo de avalúos, el precio de venta de una casa puede estar relacionado con variables explicativas como la edad, el espacio habitable en pies cuadrados, el número de baños, el número de habitaciones que no son baños, las dimensiones del terreno y un índice de la calidad de la construcción. El espacio habitable, el número de cuartos y el número de baños, desde luego, deben “moverse juntos”. Si una de estas variables aumenta, las otras generalmente se incrementarán.

Si esta dependencia lineal es menos perfecta, todavía se puede obtener la estimación de mínimos cuadrados de los coeficientes del modelo de regresión. Sin embargo, estas estimaciones tienden a ser inestables —sus valores pueden cambiar drásticamente con cambios ligeros en los datos— e inflarse (sus valores son más grandes de lo esperado). En particular, los coeficientes individuales pueden tener el signo equivocado, y los estadísticos t para juzgar la significancia de los términos individuales pueden ser insignificantes, aun cuando la prueba F indique que la relación es significativa. Finalmente, el cálculo de los estimados de mínimos cuadrados es sensible a errores de redondeo.

La *multicolinealidad* es la situación en la cual las variables independientes de una ecuación de regresión múltiple están sumamente intercorrelacionadas. Es decir, existe una relación lineal entre dos o más variables independientes.

La fortaleza de la multicolinealidad se mide por el factor de inflación de la varianza (VIF , por las siglas de variance inflation factor):⁵

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, 2, \dots, k \quad (7.11)$$

Aquí, R_j^2 es el coeficiente de determinación de la regresión de la j -ésima variable independiente con el resto $k - 1$ de las variables independientes. Para dos variables independientes $k = 2$, R_j^2 es el cuadrado de su correlación muestral, r .

Si la j -ésima variable explicativa, X_j , no está relacionada con el resto de las X , $R_j^2 = 0$ y el $VIF_j = 1$. Si existe una relación, entonces $VIF_j > 1$. Por ejemplo, cuando R_j^2 es igual a .90, $VIF_j = 1/(1 - .90) = 10$.

Un VIF próximo a 1 sugiere que la multicolinealidad no es un problema para esa variable independiente. Su coeficiente estimado y el valor t asociado no cambiarán mucho conforme las otras variables independientes se agregan o se eliminan de la ecuación de regresión. Un VIF

⁵El factor de inflación de la varianza (VIF) toma su nombre del hecho de que $s_{b_j} \propto VIF_j$. La desviación estándar estimada (error estándar) del coeficiente de mínimos cuadrados, b_j , se incrementa conforme VIF_j aumenta.

mucho más grande que 1 indica que el coeficiente estimado de esa variable independiente es inestable. Su valor y su estadístico t asociado pueden cambiar considerablemente conforme las otras variables independientes se agregan o se eliminan de la ecuación de regresión. Un VIF grande significa, en esencia, que existe información redundante entre las variables explicativas. La información expresada por una variable con un VIF grande se explica por las variables explicativas restantes. Por lo tanto, la multicolinealidad hace difícil la interpretación del efecto de una variable explicativa individual sobre la respuesta (variable dependiente).

Ejemplo 7.7

Un componente grande del costo de poseer una empresa periodística es el papel. Las editoriales de periódicos están interesadas en los factores que determinan el consumo anual de papel. En un estudio (véase Johnson y Wichern, 1997), los datos sobre el consumo anual de papel (Y), el número de periódicos en una ciudad (X_1), el logaritmo⁶ del número de familias en una ciudad (X_2), y el logaritmo de las ventas totales al menudeo en una ciudad (X_3) se recopilaron para 15 ciudades $n = 15$. El arreglo de correlación para las tres variables explicativas y el resultado de Minitab de un análisis de regresión que relaciona el consumo de papel con las variables explicativas se encuentran en la tabla 7-11.

El estadístico F (18.54) y su valor p (.000) indican claramente que la regresión es significativa. El estadístico t de cada una de las variables independientes es pequeño con un valor p relativamente grande. Debe concluirse, por ejemplo, que la variable LnFamily no es significativa, siempre que las otras variables explicativas permanezcan en la función de regresión. Esto sugiere que el término $\beta_2 X_2$ puede eliminarse de la función de regresión si los términos restantes, $\beta_1 X_1$ y $\beta_3 X_3$, se conservan. De manera análoga, parece que $\beta_3 X_3$ puede eliminarse si $\beta_1 X_1$ y $\beta_2 X_2$ permanecen en la función de regresión. El valor t (1.69) asociado con los periódicos es marginalmente significativo, pero el término $\beta_1 X_1$ también podría eliminarse si las otras variables explicativas permanecen en la ecuación. Aquí, la regresión es significativa, pero cada una de las variables explicativas no es significativa. ¿Por qué?

La columna VIF de la tabla 7-11 brinda la respuesta. Puesto que $VIF = 1.7$ para periódicos (Papers), esta variable explicativa está muy débilmente relacionada (VIF cerca de 1) con las variables explicativas restantes, LnFamily y LnRetSales. El $VIF = 7.4$ de LnFamily es relativamente grande, lo

TABLA 7-11 Resultado de Minitab para el ejemplo 7.7: tres variables explicativas

Correlations: Papers, LnFamily, LnRetSales

	Papers	LnFamily
LnFamily	0.600	
LnRetSales	0.643	0.930

Regression Analysis: Newsprint versus Papers, LnFamily, LnRetSales

The regression equation is

$$\text{Newsprint} = -56388 + 2385 \text{ Papers} + 1859 \text{ LnFamily} + 3455 \text{ LnRetSales}$$

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

$$S = 1849 \quad R\text{-Sq} = 83.8\% \quad R\text{-Sq}(\text{adj}) = 79.0\%$$

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			

⁶Los logaritmos del número de familias y de las ventas totales se emplean para hacer los números menos positivamente sesgados y más manejables.

TABLA 7-12 Resultado de Minitab para el ejemplo 7.7:
dos variables explicativas**Regression Analysis: Newsprint versus Papers, LnRetSales**

The regression equation is

$$\text{Newsprint} = -59766 + 2393 \text{ Papers} + 5279 \text{ LnRetSales}$$

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			

que indica que esta variable se encuentra linealmente relacionada con las variables explicativas restantes. Además, el $VIF = 8.1$ de LnRetSales indica que LnRetSales está relacionada con el resto de las variables explicativas. Puesto que Papers está débilmente relacionada con LnFamily y LnRetSales, la relación entre las variables explicativas es esencialmente la relación entre LnFamily y LnRetSales. De hecho, la correlación muestral entre LnFamily y LnRetSales es $r = .93$, lo que indica una asociación lineal fuerte.

Las variables LnFamily y LnRetSales son muy similares en su capacidad para explicar el consumo de papel. Sólo necesitamos una, no ambas, en la función de regresión. El resultado de Minitab de un análisis de regresión eliminando LnFamily (estadístico t más pequeño) de la función de regresión se presenta en la tabla 7-12.

Advierta que el coeficiente de Papers es aproximadamente igual para las dos regresiones. Sin embargo, los coeficientes de LnRetSales son considerablemente diferentes (3,455 para las variables explicativas $k = 3$ y 5,279 para variables explicativas $k = 2$). Además, para la segunda regresión, la variable LnRetSales es claramente significativa ($t = 4.51$ con valor $p = .001$). Con Papers en el modelo, LnRetSales es una variable explicativa adicional importante del consumo de papel. Las R^2 para las dos regresiones son casi iguales, aproximadamente .83, como lo son los errores estándar de las estimaciones, $s_{y \cdot x's} = 1,849$ y $s_{y \cdot x's} = 1,820$, respectivamente. Por último, el VIF común = 1.7 para los dos factores explicativos en el segundo modelo indica que la multicolinealidad ya no es un problema. Como lo confirma un análisis residual, para las variables consideradas, la regresión de Newsprint sobre Papers y LnRetSales es completamente adecuada.

Si la estimación de los efectos por separado de las variables explicativas es importante y la multicolinealidad parece ser un problema, ¿qué se debe hacer? Existen varias maneras para manejar la multicolinealidad severa, como veremos en seguida. Ninguna de ellas es completamente satisfactoria o factible

- Cree nuevas variables X (llamándolas $\tilde{X}$) luego escale todas las variables independientes de acuerdo con la fórmula

$$\tilde{X} = \frac{X_{ij} - \bar{X}_j}{\sqrt{\frac{1}{n} \sum_i (X_{ij} - \bar{X}_j)^2}} \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, n \quad (7.12)$$

Cada una de estas variables nuevas tiene una media muestral de 0 y la misma desviación estándar muestral. Los cálculos de la regresión con las nuevas X son menos sensibles al error de redondeo en presencia de la multicolinealidad severa.

- Identifique y elimine una o más de las variables independientes redundantes de la función de regresión. (Este método se usó en el ejemplo 7.7).

- Considere otros procedimientos de estimación que no sean los mínimos cuadrados.⁷
- Realice una regresión de la respuesta, Y , sobre las nuevas X que no están correlacionadas unas con otras. Es posible elaborar combinaciones lineales de las X originales que no están correlacionadas.⁸
- Seleccione cuidadosamente las variables independientes potenciales al principio del estudio. Trate de eliminar las variables que “dicen lo mismo”

SELECCIÓN DE LA “MEJOR” ECUACIÓN DE REGRESIÓN

¿Cómo se desarrolla la mejor ecuación de regresión múltiple para pronosticar una variable de interés? El *primer paso* implica la selección de un conjunto completo de variables explicativas potenciales. Se debe incluir cualquier variable que pudiera agregar exactitud al pronóstico. En la selección de la ecuación final, el analista generalmente se enfrenta al dilema de obtener el pronóstico más exacto al menor costo. En otras palabras, cuando elegimos las variables explicativas para incluirlas en la ecuación final, debemos evaluarlas usando los siguientes dos criterios antagónicos:

1. El analista desea que la ecuación incluya tantas variables explicativas como sea posible.⁹
2. Puesto que obtener y monitorear la información de un gran número de X es una labor costosa, la ecuación debe incluir menos variables explicativas como sea posible. Por lo general, la ecuación más simple es la mejor.

La selección de la mejor ecuación de regresión, por lo regular, implica un compromiso entre estos extremos, y el criterio personal será una parte obligatoria de cualquier solución.

Después de recopilar una lista aparentemente completa de variables explicativas potenciales, el *segundo paso* es eliminar las variables independientes que no parezcan adecuadas. Una variable independiente: **1.** quizá no sea fundamental para el problema (debe haber alguna relación posible entre la variable dependiente y la variable independiente), **2.** podría estar sujeta a errores de medición grandes, **3.** puede ser un duplicado de otra variable independiente (multicolinealidad), o **4.** tal vez sea difícil medirla con exactitud (los datos exactos no están disponibles o sería muy costoso obtenerlos).

El *tercer paso* es recortar la lista de variables explicativas de manera que se obtenga la “mejor” selección de variables independientes. En el material que sigue se analizan las técnicas actualmente en uso. No se puede decir que alguno de los procedimientos de búsqueda permita obtener el “mejor” conjunto de variables independientes. De hecho, no es frecuente que exista un solo conjunto “mejor”. Para agravar la confusión, las diferentes técnicas no necesariamente conducen a la misma ecuación final de predicción. El proceso completo de selección de la variable es muy subjetivo. La ventaja principal de los procedimientos automáticos de búsqueda es que los analistas pueden enfocar su atención en las áreas fundamentales del problema.

Para estudiar diferentes procedimientos de búsqueda, se presenta un ejemplo sencillo que incluye cinco variables independientes potenciales.

Ejemplo 7.8

Pam Weigand, la directora de personal de Zurenko Pharmaceutical Company, está interesada en pronosticar si un candidato particular se convertirá en un buen vendedor. Ella decide usar las primeras ventas mensuales como la variable dependiente (Y) y selecciona para el análisis las siguientes variables independientes

⁷Los procedimientos alternativos para la estimación de los parámetros de la regresión están más allá del alcance de este libro. Al lector interesado se le recomienda consultar el trabajo de Draper y Smith (1988).

⁸Otra vez, los procedimientos para la creación de combinaciones lineales de las X que no están correlacionadas se encuentran más allá del alcance de este libro. Draper y Smith (1988) analizan estas técnicas.

⁹Recuerde que siempre que se agrega una variable explicativa nueva a la ecuación de regresión múltiple, R^2 aumenta. Por ello, es importante que una nueva variable explicativa haga una contribución *significativa* a la ecuación de regresión.

X_1 = prueba de aptitud para ventas

X_2 = edad, en años

X_3 = prueba de ansiedad

X_4 = experiencia en años

X_5 = promedio de calificaciones (*grade point average*)

La directora de personal recopila los datos presentados en la tabla 7-13, y asigna a su analista la tarea de obtener el “mejor” conjunto de variables independientes para pronosticar la habilidad para trabajar en ventas.

El primer paso es obtener una matriz de correlación para todas las variables usando un programa de computadora. Esta matriz le dará el conocimiento esencial acerca de las relaciones básicas entre las variables.

El examen de la matriz de correlación en la tabla 7-14 revela que la calificación del examen de aptitud para ventas, la edad, la experiencia y el GPA están relacionados positivamente con la habilidad para trabajar en ventas y tienen potencial como buenas variables explicativas. La calificación en la prueba de ansiedad indica una correlación negativa baja con las ventas y probablemente no es una variable explicativa importante. El análisis posterior indica que la edad está correlacionada modera-

TABLA 7-13 Datos de la Zurenko Pharmaceutical para el ejemplo 7.8

<i>Ventas de un mes (unidades)</i>	<i>Calificación en la prueba de aptitud</i>	<i>Edad (años)</i>	<i>Calificación en la prueba de ansiedad</i>	<i>Experiencia (años)</i>	<i>GPA de la preparatoria</i>
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, experiencia, GPA

Correlations: Sales, Aptitude, Age, Anxiety, Experience, GPA					
	Sales	Aptitude	Age	Anxiety	Experience
Aptitude	0.676				
Age	0.798	0.228			
Anxiety	-0.296	-0.222	-0.287		
Experience	0.550	0.350	0.540	-0.279	
GPA	0.622	0.318	0.695	-0.244	0.312

damente tanto con el GPA como con la experiencia. La presencia de estas interrelaciones debe considerarse en el intento por encontrar el mejor conjunto posible de variables explicativas.

Se presentan dos procedimientos: todas las regresiones posibles y la regresión por pasos.

Todas las regresiones posibles

El procedimiento requiere la investigación de todas las ecuaciones de regresión posibles que implican las variables independientes potenciales. El analista inicia con una ecuación que no contiene variables independientes y luego procede al análisis de todas las combinaciones posibles para seleccionar el mejor conjunto de variables explicativas.

Con el *enfoque de todas las regresiones posibles* se pueden usar diferentes criterios para comparar varias ecuaciones de regresión. Aquí se analizará sólo la técnica de la R^2 la cual contempla cuatro pasos

Este procedimiento requiere primero el ajuste de cada modelo de regresión posible que incluye la variable dependiente y cualquier número de variables independientes. Cada variable independiente puede estar o no en la ecuación (dos posibles resultados), y este hecho es verdad para todas las variables independientes. Por lo tanto, existen en total 2^k ecuaciones (donde k es el número de variables independientes). De esta forma, si existen ocho variables independientes a considerar ($k = 8$), habrá que examinar $2^8 = 256$ ecuaciones.

El segundo paso del 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 la Zurenko Pharmaceutical Company se presentan en la tabla 7-15. Observe que la tabla 7-15 está dividida en seis conjuntos de resultados de las ecuaciones de regresión. Este agrupamiento coincide con el número de parámetros contenidos en cada ecuación.

El tercer paso consiste en la selección de la mejor variable (o variables) independiente(s) para cada agrupamiento de parámetros. La ecuación con la R^2 más alta se considera la mejor. Con base en los resultados del ejemplo 7.9, la mejor ecuación de cada grupo listado en la tabla 7-15 se presenta en la tabla 7-16

El cuarto paso supone tomar la decisión subjetiva: “¿Cuál ecuación es la mejor?”. Por un lado, el analista desea la R^2 más grande posible; por otro, quiere la ecuación más sencilla posible. El enfoque de todas las regresiones posibles supone que el número de puntos de datos, n , excede el número de parámetros, $k + 1$.

Ejemplo 7.10

El analista está intentando encontrar el punto en el cual agregar variables independientes para el problema de la Zurenko Pharmaceutical ya no vale la pena porque conduce a un incremento muy pequeño en R^2 . Los resultados en la tabla 7-16 indican claramente que no es necesario sumar varia-

TABLA 7-15 Valores de R^2 para todas las regresiones posibles de Zurenko Pharmaceutical; ejemplo 7.9

<i>Variables independientes empleadas</i>	<i>Número de parámetros</i>	<i>Grados de libertad de los errores</i>	R^2
Ninguno	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 de la Zurenko Pharmaceutical; ejemplo 7.9

<i>Número de parámetros</i>	<i>Variables independientes</i>	<i>Grados de libertad de los errores</i>	R^2
1	Ninguno	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

bles después de la prueba de aptitud para ventas (X_1) y la edad (X_2). Por lo tanto, la ecuación final de regresión ajustada tiene la forma

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2$$

y explica el 89.48% de la variación en Y .

Draper y Smith (1998) resumen así el procedimiento de todas las regresiones posibles:

En general, el análisis de todas las regresiones no está justificado. Aunque significa que el investigador ha “considerado todas las posibilidades”, también significa que ha examinado un gran número de ecuaciones de regresión que una persona inteligente a menudo rechazaría de inmediato. El tiempo de computadora utilizado es un despilfarro y tan sólo el trabajo físico de examinar todos los reportes de resultados de la computadora es enorme cuando se examinan más de unas cuantas variables. Es preferible alguna clase de selección de un procedimiento que acorte esta tarea. (p. 333)

Regresión por pasos

El procedimiento de regresión por pasos agrega al modelo una variable independiente a la vez. Cuando se utiliza este procedimiento se puede manejar un número grande de variables independientes en una sola corrida en la computadora.

La regresión por pasos puede describirse mejor listando los pasos básicos (el algoritmo) para realizar los cálculos.

1. Se consideran todas las regresiones *simples* posibles. La variable explicativa que explica la proporción significativa más grande de la variación de Y (es decir, tiene la correlación más grande con la respuesta) es la primera variable que se introduce en la ecuación de regresión.
2. La siguiente variable en introducirse a la ecuación es la que (fuera de aquellas no incluidas) hace la mayor contribución significativa a la suma de cuadrados de la regresión. La significancia de la contribución se determina mediante una prueba F . El valor del estadístico F que debe sobrepasarse antes de que la contribución de una variable se considere significativa a menudo se conoce como F de entrada (F to enter).
3. Una vez que una variable adicional se ha incluido en la ecuación, se hace una prueba de significancia de las contribuciones individuales a la suma de cuadrados de regresión de las otras variables *que ya están en la ecuación* usando pruebas F . Si el estadístico F es menor que un valor llamado F de salida (F to remove), la variable se elimina de la ecuación de regresión.
4. Los pasos 2 y 3 se repiten hasta que todas las adiciones posibles sean no significativas y todas las eliminaciones posibles sean significativas. En este punto, la selección se detiene.

La *regresión por pasos* permite a las variables explicativas entrar o salir de la función de regresión en diferentes etapas de su desarrollo. Una variable independiente se elimina del modelo si 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 una variable puede entrar y cuándo una variable se elimina. Puesto que los estadísticos F usados en la regresión por pasos son tales que $F = t^2$ donde t es el estadístico t para probar la significancia de una variable explicativa, $F = 4$ (correspondiente a $|t| = 2$) es una opción común tanto para F de entrada y la F de salida. Una F de entrada de 4 es esencialmente equivalente a probar la significancia de una variable explicativa con un nivel de 5%. El programa por pasos de Minitab permite al usuario seleccionar un nivel α para introducir y eliminar variables, o

bien, el valor F para introducir y eliminar variables. Usar un valor α de .05 es aproximadamente equivalente a emplear $F = 4$. Los valores actuales predeterminados de Minitab son $\alpha = .15$ y $F = 4$.

El resultado del procedimiento por pasos es un modelo que contiene sólo variables independientes con valores t que son significativos en el nivel especificado. Sin embargo, el desarrollo del procedimiento por pasos no garantiza que la regresión por pasos permita seleccionar, por ejemplo, las tres mejores variables explicativas. Además, un método de selección automática no es capaz de indicar cuándo son útiles las transformaciones de las variables, ni necesariamente evita el problema de la multicolinealidad. Finalmente, la regresión por pasos no puede crear variables importantes que no sean suministradas por el usuario. Es necesario pensar cuidadosamente acerca de la colección de variables independientes que se suministra a un programa de regresión por pasos.

El procedimiento por pasos se ilustra en el ejemplo 7.11.

Ejemplo 7.11

Vamos a “resolver” el problema de Zurenko Pharmaceutical usando la regresión por pasos.

Pam examina la matriz de correlación presentada en la tabla 7-14 y decide que, cuando ella corra el análisis por pasos, la variable de la edad entrará primero al modelo porque tiene la correlación más grande con las ventas ($r_{1,3} = .798$) y explicará el 63.7% ($.798^2$) de la variación en las ventas.

Ella observa que la calificación en la prueba de aptitud probablemente será la segunda en entrar al modelo porque está fuertemente relacionada con las ventas ($r_{1,2} = .676$) pero no está fuertemente relacionada con la variable de la edad ($r_{2,3} = .228$) ya incluida en el modelo.

Pam también advierte que las otras variables probablemente no calificarán como buenas variables explicativas. La calificación en la prueba de ansiedad no será una buena variable explicativa porque no está bien relacionada con las ventas ($r_{1,4} = -.296$). Las variables de experiencia y GPA podrían tener potencial como buenas variables explicativas ($r_{1,5} = .550$ y $r_{1,6} = .622$, respectivamente). Sin embargo, ambas variables explicativas tienen un problema potencial de multicolinealidad con la variable de la edad ($r_{3,5} = .540$ y $r_{3,6} = .695$, respectivamente).

Los comandos de Minitab para correr un análisis de regresión por pasos para este ejemplo se describen en la sección de aplicaciones de Minitab al final del capítulo. La salida para esta regresión por pasos se presenta en la tabla 7-17. El análisis por pasos se aplica de acuerdo con los siguientes pasos.

TABLA 7-17 Regresión por pasos para el ejemplo 7.11: Ventas contra aptitud, edad, ansiedad, experiencia y 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
Mallows Cp	57.4	0.2

Paso 1. El modelo después del paso 1 es

$$\text{Ventas} = -100.85 + 6.97 (\text{Edad})$$

Como Pam estableció, la variable de la edad se introduce primero al modelo y explica el 63.7% de la varianza en las ventas. Puesto que el valor p de .000 es menor que el valor α de .05, se agrega la edad al modelo. Recuerde que el valor p es la probabilidad de obtener un estadístico t tan grande como 7.01 por casualidad. La regla de decisión de Minitab que Pam seleccionó fue introducir la variable si el valor p era menor que $\alpha = .05$.

Observe que $t = 7.01 < 2.048$, 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 este modo, en el nivel de significancia .05, la hipótesis $H_0: \beta_1 = 0$ se rechaza en favor de $H_1: \beta_1 \neq 0$. Puesto que $t^2 = F$ o $2.048^2 = 4.19$, una F de entrada de 4 también es esencialmente equivalente a la prueba de significancia de una variable explicativa en el nivel del 5%. En este caso, puesto que el coeficiente de la variable de la edad es a todas luces significativamente diferente de cero, la edad entra a la ecuación de regresión y el procedimiento ahora continúa hacia 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 el 89.48% de la variación en las ventas.

La hipótesis nula y la hipótesis alternativa para determinar si el coeficiente de regresión de la calificación en 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 .000 es menor que el valor α de .05, y la calificación en la prueba de aptitud se agrega al modelo. El coeficiente de regresión de la calificación de la prueba de aptitud es significativamente diferente de cero y la probabilidad de que esto ocurra tan sólo por error de muestreo es aproximadamente de cero. Este resultado significa que la calificación en la prueba de aptitud es una variable importante cuando se utiliza en conjunto con la edad.

El estadístico crítico t basado en 27 ($n - k - 1 = 30 - 2 - 1$) grados de libertad es 2.052.¹⁰ La razón t calculada que se encuentra en el resultado de Minitab es 8.13, una cifra mayor que 2.052. Usando la prueba t , la hipótesis nula también se rechaza.

Advierta que el valor p del estadístico t de la variable de la edad, .000, sigue siendo muy pequeño. La edad aún es una variable explicativa significativa de las ventas. El procedimiento ahora avanza hacia el paso 3.

Paso 3. La computadora ahora considera agregar una tercera variable explicativa, en vista de que X_1 (la edad) y X_2 (la calificación en la prueba de aptitud) están en la ecuación de regresión. Ninguna de las variables independientes restantes es significativa (con un valor p menor que .05) cuando corre en combinación con X_1 y X_2 , de manera que el procedimiento por pasos concluye.

El modelo final que seleccionó Pam mediante el procedimiento de los pasos es el modelo de dos variables explicativas descrito en el paso 2.

Notas finales sobre la regresión por pasos

La técnica de regresión por pasos es extremadamente fácil de usar. Por desgracia, también es extremadamente fácil hacer un uso inadecuado de esta técnica. Los analistas que desarrollan un modelo de regresión a menudo generan un conjunto grande de variables independientes potenciales y luego dejan que el procedimiento por pasos determine cuáles son significativas. El problema es que, cuando se analiza un grupo grande de variables independientes, se realizan muchas pruebas t y es probable que se presente el error tipo I (agregar una variable no significativa). Es decir, el modelo final podría contener una variable que no está linealmente relacionada con la variable dependiente y que se introdujo al modelo de manera fortuita.

¹⁰Otra vez, puesto que $2.052^2 = 4.21$, el uso de una F de entrada de 4 es aproximadamente equivalente que hacer una prueba de significancia de una variable explicativa con un nivel de .05.

Como se mencionó anteriormente, existe otro problema implicado en la selección inicial de las variables independientes potenciales. Cuando se seleccionan estas variables, a menudo se omiten términos de alto orden (curvilíneos, no lineales e interacción) para dejar manejable el número de variables. Así, es posible que muchas variables importantes se omitan del modelo desde el principio. Es obvio que la selección intuitiva de las variables independientes iniciales por parte del analista es crucial para tener éxito en el desarrollo de un modelo de regresión.

DIAGNÓSTICOS DE REGRESIÓN Y ANÁLISIS RESIDUAL

Un análisis de regresión no está completo sino hasta que uno está convencido de que el modelo es una representación adecuada de los datos. Es imperativo examinar la idoneidad del modelo antes de que se convierta en parte del mecanismo de toma de decisiones.

Un examen de los residuos es un componente crucial de la determinación de la idoneidad del modelo. Varias gráficas residuales diseñadas para verificar los supuestos del modelo se presentaron en el capítulo 6. Por otra parte, si se usan los modelos de regresión con datos de una serie de tiempo (lo que se analizará en el capítulo 8), es necesario calcular las autocorrelaciones residuales para verificar el supuesto de independencia. Las inferencias (y decisiones) con base en modelos que no satisfacen los supuestos de la regresión pueden ser sumamente desorientadoras. Por ejemplo, podría concluirse que la manipulación de una variable explicativa producirá un cambio específico en la respuesta cuando, de hecho, no sucede así. Se puede concluir también que es muy probable que un pronóstico (95% de confianza) esté dentro del 2% de la respuesta futura cuando, de hecho, la confianza real es mucho menor, y así podríamos seguir dando ejemplos.

En esta sección se analizarán algunas herramientas adicionales que permiten evaluar un modelo de regresión. Estas herramientas están diseñadas para identificar observaciones que son atípicas o extremas (observaciones que están muy alejadas del resto de los datos). Las observaciones atípicas a menudo se esconden tras el proceso de ajuste y es difícil detectarlas en un examen de las gráficas de residuos. Sin embargo, pueden tener un papel importante al determinar la función de regresión ajustada. Es fundamental estudiar las observaciones atípicas para decidir si deben conservarse o eliminarse y, en el primer caso, habrá que decidir si su influencia debe reducirse en el proceso de ajuste o si debe revisarse la función de regresión.

El apalancamiento h_{ii} ofrece una medida de la influencia del i -ésimo punto de datos sobre la ubicación de la función de regresión ajustada. El apalancamiento sólo depende de las variables explicativas; no depende de la respuesta Y . Para la regresión lineal simple con una variable explicativa X ,

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2} \quad (7.13)$$

Con k variables explicativas, la expresión para el i -ésimo apalancamiento es más complicada; sin embargo, uno puede demostrar que $0 < h_{ii} < 1$ y que el apalancamiento medio es $\bar{h} = (k + 1)/n$.

Si el i -ésimo punto de datos tiene un apalancamiento alto (h_{ii} está cercano a 1), la respuesta ajustada $\hat{Y}_i$, en estas X se encuentra casi completamente determinada por Y_i , con muy poca influencia de los datos restantes. El punto de datos con apalancamiento alto también es un valor atípico entre las X (lejos de otras combinaciones de valores X).¹¹ Una regla empírica sugiere que h_{ii} es lo suficientemente grande para verificar si $h_{ii} \geq 3(k + 1)/n$.

La detección de valores atípicos o extremos de Y está basada en el tamaño de los residuos, $e = Y - \hat{Y}$. Residuos grandes indican un valor de Y que está “lejos” de su valor pronosticado o

¹¹Lo contrario no necesariamente es cierto. Es decir, un valor atípico entre las X puede no tener un punto de apalancamiento alto.

ajustado, $\hat{Y}$. Un residuo grande se manifestará en un histograma de los residuos con un valor lejos de cero (en cualquier dirección). Un residuo grande se manifestará en una gráfica de residuos contra los valores ajustados como un punto muy arriba o muy abajo del eje horizontal.

Los paquetes de software, como Minitab, marcan los puntos de datos con valores extremos de Y calculando residuos “estandarizados” e identificando puntos con residuos estandarizados grandes.

Una estandarización se basa en el hecho de que los residuos 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 punto de datos. El *residuo estandarizado*¹² es entonces

$$\frac{e_i}{s_{e_i}} = \frac{e_i}{s_{y \cdot x's} \sqrt{1 - h_{ii}}} \quad (7.14)$$

Todos los residuos estandarizados tienen una varianza de 1. Un residuo estandarizado se considera grande (la respuesta extrema) si

$$\left| \frac{e_i}{s_{e_i}} \right| > 2$$

Los valores Y correspondientes a los puntos de datos con residuos estandarizados grandes pueden influir fuertemente en la ubicación de la función de regresión ajustada.

Ejemplo 7.12

Los salarios de los directores ejecutivos en Estados Unidos son de interés en virtud de su relación con los salarios de los directores ejecutivos de las empresas internacionales y de los profesionales mejor pagados fuera de Estados Unidos. Además, para una empresa particular, la remuneración directa o indirecta de un director ejecutivo influye en los salarios de los gerentes de menor nivel. El salario del director ejecutivo varía considerablemente de una empresa a otra, pero los datos sugieren que el salario se puede explicar en términos de las ventas de la empresa, de la experiencia y el nivel educativo del director ejecutivo, así como de su participación como accionista en la compañía. En un estudio que incluyó a 50 empresas, los investigadores se propusieron desarrollar un modelo de regresión múltiple para vincular el sueldo de los directores ejecutivos con diversas variables explicativas, como ventas, utilidades, edad, experiencia, antecedentes profesionales y participación accionaria.

Después de la eliminación de variables explicativas sin importancia, la función de regresión ajustada fue

$$\hat{Y} = 5.52 - .467X_1 + .263X_2$$

donde

Y = logaritmo de la remuneración del director general

X_1 = variable indicadora del nivel educativo

X_2 = logaritmo de las ventas de la compañía

Minitab identificó tres observaciones a partir de este análisis de regresión que tenían residuos estandarizados grandes o apalancamiento grande.

¹²Algunos paquetes de software llaman *Studentized* al residuo estandarizado de la ecuación 7.14.

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 residuos estandarizados grandes. La función de regresión ajustada está prediciendo una remuneración (en logaritmos) de estos dos directores ejecutivos muy elevada. Un examen del conjunto completo de datos indica que cada uno de estos directores ejecutivos posee porcentajes relativamente grandes de las acciones de su compañía. El caso 14 posee más del 10% de acciones de la empresa, y el caso 33 posee más del 17% de acciones. Estos individuos reciben la mayor parte de su remuneración en forma de pagos a largo plazo, como incentivos accionarios, y no en forma de salario y bonos anuales. Puesto que la cantidad de acciones poseídas (o valor de las acciones) no está incluida como variable en la función de regresión, no puede utilizarse para ajustar el pronóstico de la remuneración determinada por la educación del director ejecutivo y las ventas de la compañía. Si bien la educación y las ventas (en logaritmo) no pronostican la remuneración de estos dos directores ejecutivos tan bien como las otras, no parece haber una razón para eliminarlas del estudio.

La observación 25 se distingue porque el apalancamiento de este punto de datos es mayor que $3(k+1)/n = 3(3)/50 = .18$. Este director ejecutivo no tiene grados académicos (Educación = 0) pero se encuentra en una compañía con ventas relativamente grandes ($\text{LnSales} = 9.394$). La combinación (0, 9.394) está lejos del punto $(\bar{X}_1, \bar{X}_2)$; por lo tanto, es un valor atípico en 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 residuo estandarizado para este punto de datos es pequeño, lo cual indica que la remuneración (log) pronosticada o ajustada está cerca del valor real]. Este director ejecutivo particular tiene 30 años de experiencia en el cargo, más experiencia que todos los directores ejecutivos, a excepción de uno, en el conjunto de datos. Esta observación es influyente, por lo que no hay razón para eliminarla.

El apalancamiento nos indica si una observación tiene variables explicativas inusuales, y un residuo estandarizado nos dice si una observación tiene una respuesta inusual. Estas cantidades pueden combinarse en una medida integral de influencia conocida como distancia de Cook. Las *distancias de Cook* se obtienen en la mayoría de paquetes de software estadístico, pero una explicación adicional está más allá del alcance de este libro.¹³

ADVERTENCIAS DEL PRONÓSTICO

Terminaremos este estudio de la regresión múltiple con algunos comentarios generales, los cuales están orientados a la aplicación práctica del análisis de regresión.

Sobreajuste de la prueba

El *sobreajuste* se refiere al hecho de agregar variables independientes a la función de regresión que, en gran medida, dan cuenta de todas las excentricidades de los datos muestrales bajo análisis.

Cuando se aplica un modelo sobreajustado a nuevos conjuntos de datos seleccionados de la misma población, no pronostica tan bien como el ajuste inicial podría sugerir.

Es más probable que ocurra el sobreajuste cuando el tamaño de la muestra es pequeño, especialmente si se incluye un gran número de variables independientes en el modelo. Algunos especialistas

¹³Draper y Smith (1998) aportan una buena explicación de la distancia de Cook.

sugieren que deben existir por lo menos 10 observaciones por cada variable independiente. (Si existen cuatro variables independientes, se sugiere un tamaño n de la muestra de por lo menos 40).

Una manera de protegerse contra el sobreajuste es desarrollar la función de regresión de una parte de los datos y luego aplicarla a una porción de la muestra. Se usa la función de regresión ajustada para pronosticar las respuestas de otra porción de datos y calcular los errores de pronóstico. Si los errores de pronóstico son sustancialmente mayores que los errores de ajuste medidos, digamos, por los errores cuadráticos medios comparables, entonces ha ocurrido un sobreajuste.

Regresiones útiles, valores grandes calculados de F

Una regresión que es estadísticamente significativa no necesariamente es útil. Con un tamaño muestral relativamente grande (por ejemplo, cuando n es grande en relación con k , el número de variables explicativas), no es extraño obtener una razón F significativa y una R^2 pequeña. Es decir, la regresión es significativa, pero sólo explica una pequeña porción de la variación en la respuesta. Una regla general sugiere que con un nivel de significancia de .05, la razón F debe ser por lo menos cuatro veces el valor crítico correspondiente antes de que la regresión pueda usarse para fines de predicción.¹⁴

El criterio de “cuatro veces” surge del argumento de que el rango de las predicciones (sobre todas las X) debe ser alrededor de cuatro veces el error de predicción (promedio) antes de que la regresión pueda generar una interpretación que valga la pena.¹⁵

Como ejemplo, con $k = 3$ variables explicativas, $n = 25$ observaciones, y un nivel de significancia de .05, la F calculada a partir de una tabla ANOVA tendría que exceder el valor crítico $F = 3.07$ (véase la tabla B-5 con $\delta_1 = k = 3$, $\delta_2 = n - k - 1 = 21$ grados de libertad) para que la regresión sea significativa. (Con base en la ecuación 7.7, la F crítica = 3.07 corresponde a una R^2 de alrededor del 30%, una cifra que no es particularmente grande). Sin embargo, la regla de las “cuatro veces” sugiere que la F calculada debería exceder $4(3.07) = 12.28$ para que la regresión valga la pena desde un punto de vista práctico.

APLICACIÓN A LA ADMINISTRACIÓN

El análisis de regresión múltiple se usa ampliamente para ayudar a pronosticar la actividad de varios segmentos de la economía. Muchos de los reportes y pronósticos acerca del futuro de la economía que aparecen en el *Wall Street Journal*, *Fortune*, *Business Week* y otras fuentes similares se basan en modelos econométricos (de regresión). El gobierno de Estados Unidos usa con frecuencia el análisis de regresión para pronosticar ingresos, gastos, niveles de ingreso, tasas de interés, índices de natalidad, desempleo y requerimientos para obtener beneficios de seguridad social, así como una multitud de sucesos. De hecho, prácticamente todas las grandes oficinas del gobierno de Estados Unidos utilizan las herramientas descritas en este capítulo.

De manera similar, las entidades de negocios han adoptado y, cuando es necesario, modificado el análisis de regresión como ayuda para el pronóstico de acontecimientos futuros. Pocas empresas pueden sobrevivir en el ambiente actual sin un pronóstico bastante exacto de las ventas, los gastos, los requerimientos de capital y los flujos de efectivo para el futuro. Si bien las empresas pequeñas o menos complejas pueden funcionar con pronósticos intuitivos, las empresas más grandes y/o más complejas recurren al análisis de regresión para estudiar las relaciones entre diversas variables y determinar cómo tales variables podrían afectar su futuro.

Por desgracia, la gran notoriedad que adquiere el análisis de regresión por su utilidad como una herramienta de pronóstico del futuro tiende a opacar un activo igualmente importante: su capacidad para ayudar a evaluar y controlar el presente. Como la ecuación de regresión ajustada brinda al investigador información sobre la *fortaleza y dirección*, la administración puede evaluar y modificar las estrategias en curso.

¹⁴Algunos autores argumentan que la regla de “cuatro veces” no es suficiente y que debería sustituirse por un criterio de “diez veces”.

¹⁵Esto supone que no se detectan otros defectos en el ajuste.

Suponga, por ejemplo, que un fabricante de mermelada quiere saber hacia dónde dirigir sus esfuerzos de marketing cuando lanza al mercado un nuevo sabor. Se puede usar el análisis de regresión para ayudar a determinar el perfil de los grandes consumidores de mermelada. Por ejemplo, tal vez una compañía quiera predecir el número de sabores de mermelada que una ama de casa podría tener en la alacena en un momento dado sobre la base de un número de variables independientes, como las siguientes:

- Número de niños que hay en casa
- Edad de los niños
- Sexo de los niños
- Propietarios de casas contra arrendadores
- Tiempo invertido en compras
- Ingreso

Incluso una reflexión superficial sobre el ejemplo de la mermelada rápidamente lleva al investigador a reconocer que el análisis de regresión tiene muchas posibilidades de uso en estudios de segmentación de mercado. De hecho, muchas compañías usan la regresión para estudiar segmentos de mercado y determinar cuáles variables parecen tener efecto sobre la participación en el mercado, la frecuencia de compra, la propiedad del producto y la lealtad a la marca y al producto, así como sobre otras muchas áreas.

Los especialistas en ciencias agrícolas utilizan el análisis de regresión para explorar la relación entre el rendimiento de la tierra (por ejemplo, número de fanegas de maíz por acre) y el tipo y la cantidad de fertilizante, la precipitación, la temperatura, los días de sol y las plagas de insectos. Las granjas modernas están equipadas con mini y micro computadoras completas con paquetes de software para ayudar en este proceso.

Los investigadores médicos usan el análisis de regresión para buscar vínculos entre la presión sanguínea y variables independientes como la edad, la clase social, el peso, el tabaquismo y la raza. También estudian el efecto de las comunicaciones, el número de visitas y la edad del paciente en relación con la satisfacción del servicio médico.

Los directores de personal exploran la relación de los niveles de salario de los empleados con la ubicación geográfica, las tasas de desempleo, el crecimiento de la industria, la afiliación a sindicatos, el tipo de industria y salarios competitivos. Los analistas financieros buscan las causas de los altos precios de las acciones analizando el rendimiento de dividendos, las ganancias por acción, la partición accionaria, las expectativas del consumidor en cuanto a tasas de interés, los niveles de ahorro y las tasas de inflación.

Los gerentes de publicidad a menudo tratan de estudiar el efecto de los presupuestos en publicidad, la selección de medios, la redacción del mensaje, la frecuencia del anuncio y la selección del anunciador en relación con el cambio de actitud del consumidor. De manera análoga, los especialistas en marketing intentan determinar las ventas a partir de los gastos en publicidad, los niveles de precio, los gastos competitivos en marketing y los ingresos disponibles del consumidor, así como una gran diversidad de variables.

Citemos un ejemplo final adicional para ilustrar la versatilidad del análisis de regresión. Los analistas del mercado inmobiliario han encontrado que el análisis de regresión es muy útil en la localización exacta de áreas geográficas con abundancia o escasez de tipos específicos de tiendas minoristas. Por ejemplo, una cadena de ferreterías podría buscar una ciudad potencial en la cual ubicar una nueva sucursal desarrollando un modelo de regresión diseñado para predecir las ventas de herramientas en cualquier ciudad específica. Los investigadores concentrarían sus esfuerzos en aquellas ciudades donde el modelo predijo ventas más altas que las reales (como se puede determinar consultando muchas fuentes). La hipótesis es que las ventas de artículos de ferretería no alcanzan todo su potencial en esas ciudades.

En resumen, el análisis de regresión brinda a los gerentes una herramienta poderosa y versátil para estudiar las relaciones entre una variable dependiente y múltiples variables independientes. El objetivo es comprender mejor y quizás controlar acontecimientos del presente, así como predecir mejor hechos futuros.

Glosario

Variables ficticias. Las variables o indicadores ficticios sirven para determinar las relaciones entre variables independientes cualitativas y una variable dependiente.

Multicolinealidad. La multicolinealidad es la situación en la cual las variables independientes en una ecuación de regresión múltiple están estrechamente intercorrelacionadas. Es decir, existe una relación lineal entre dos o más variables independientes.

Regresión múltiple. La regresión múltiple implica el uso de más de una variable independiente para predecir una variable dependiente.

Sobreajuste. Se refiere al hecho de agregar variables independientes a la función de regresión que, en gran medida, explican todas las excentricidades de los datos muestrales bajo análisis.

Coefficiente de regresión parcial o neto. El coeficiente de regresión parcial, o neto, mide el cambio promedio de la variable dependiente por unidad de cambio en la variable independiente relevante, cuando se mantienen constantes las demás variables independientes.

Error estándar de la estimación. El error estándar de la estimación es la desviación estándar de los residuos. Mide la cantidad de valores reales (Y) que difieren de los valores estimados ($\hat{Y}$).

Regresión por pasos. La regresión por pasos permite a las variables explicativas entrar o salir de la función de regresión en diferentes etapas de su desarrollo. Una variable independiente se elimina del modelo si no continúa haciendo una contribución significativa cuando se agrega una nueva variable.

Fórmulas clave

Función de regresión múltiple de 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

$$s_{y \cdot x's} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - k - 1}} = \sqrt{\frac{SSE}{n - k - 1}} = \sqrt{MSE} \quad (7.4)$$

Estadístico F para prueba de significancia de la regresión

$$F = \frac{MSR}{MSE} \quad (\text{véase el recuadro de la p. 289})$$

Coefficiente 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 el estadístico 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

$$\bar{R}^2 = 1 - (1 - R^2) \left(\frac{n - 1}{n - k - 1} \right) \quad (7.8)$$

Estadístico t para la prueba de $H_0: \beta_j = 0$

$$t = \frac{b_j}{s_{b_j}} \quad (\text{véase el recuadro de la p. 291})$$

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 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 inflación 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

$$\tilde{X} = \frac{X_{ij} - \bar{X}_j}{\sqrt{\sum_i (X_{ij} - \bar{X}_j)^2}} \quad j = 1, 2, \dots, k; \quad i = 1, 2, \dots, n \quad (7.12)$$

Apalancamiento (una variable explicativa)

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum (X_i - \bar{X})^2} \quad (7.13)$$

Residuo estandarizado

$$\frac{e_i}{s_{e_i}} = \frac{e_i}{s_{y \cdot x' s} \sqrt{1 - h_{ii}}} \quad (7.14)$$

Problemas

1. ¿Cuáles son las características de una buena variable explicativa?
2. ¿Cuáles son los supuestos asociados con el 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. Su ecuación de la regresión múltiple estimada 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 variable	Número de 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:
 - a) Matriz de correlación
 - b) R^2
 - c) Multicolinealidad
 - d) Residuo
 - e) Variable ficticia
 - f) Regresión por pasos
7. La mayoría de las soluciones por computadora para la regresión múltiple inician con una matriz de correlación. El examen de esta matriz a menudo es el primer paso cuando se analiza un problema de regresión que implica más de una variable independiente. Conteste las siguientes preguntas relacionadas con la matriz de correlación de la tabla P-7.
 - a) ¿Por qué todas las entradas que se encuentran sobre la diagonal principal son iguales a 1.00?
 - b) ¿Por qué la mitad de la matriz debajo de la diagonal principal se encuentra en blanco?
 - c) Si la variable 1 es la variable dependiente, ¿cuáles variables independientes tienen el grado más alto de asociación lineal con la variable 1?
 - d) ¿Qué clase de asociación existe entre las variables 1 y 4?
 - e) ¿Esta matriz de correlación indica alguna evidencia de multicolinealidad?
 - f) En su opinión, ¿cuál variable o variables deberían incluirse en el mejor modelo de pronóstico? Explique.
 - g) Si los datos presentados en esta matriz de correlación se corren en un programa de regresión por pasos, ¿cuál variable independiente (2, 3, 4, 5 o 6) será la primera en entrar en la función de regresión?
8. A Jennifer Dahl, supervisora de la cadena de descuento Circle O, le gustaría pronosticar el tiempo que se tardan los empleados en cobrarle a un cliente. Ella decide usar las siguientes variables independientes: número de artículos comprados e importe total de compra. Recopila datos para una muestra de 18 clientes, los cuales se presentan en la tabla P-8.
 - a) Determine la mejor ecuación de regresión.
 - b) Cuando se compra un artículo adicional, ¿cuál es el incremento promedio en el tiempo de cobro?
 - c) Calcule el residuo para el cliente 18.
 - d) Calcule el error estándar de la estimación.
 - e) Interprete el inciso d) en términos de las variables usadas en el problema.
 - f) Elabore un pronóstico del tiempo de cobro si un cliente compra 14 artículos que suman un total de \$70.
 - g) Calcule un intervalo estimado de 95% para su pronóstico del inciso f).
 - h) ¿Qué debería concluir Jennifer?
9. La tabla P-9 contiene datos de gastos de alimentación, ingreso anual y tamaño de la familia para una muestra de 10 familias.

TABLA P-8

<i>Cliente</i>	<i>Tiempo de cobro (minutos) Y</i>	<i>Importe (\$) X₁</i>	<i>Número de artículos X₂</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 de comida (en cientos de \$) Y</i>	<i>Ingreso anual (en miles de \$) X₁</i>	<i>Tamaño de la familia X₂</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

- Elabore la matriz de correlación para las tres variables de la tabla P-9. Interprete las correlaciones en la matriz.
 - Ajuste un modelo de regresión múltiple que relacione los gastos de comida con los ingresos y el tamaño de la familia. Interprete los coeficientes de regresión parcial de ingreso y tamaño de la familia. ¿Tienen sentido?
 - Calcule los factores de inflación de la varianza (*VIF*) para las variables independientes. ¿La multicolinealidad es un problema para estos datos? Si es así, ¿cómo podría usted modificar el modelo de regresión?
10. Las ventas de cerveza en la tienda Shapiro One-Stop se analizan considerando la temperatura y el número de personas en la calle (de 21 años de edad en adelante) como variables independientes. Se selecciona una muestra aleatoria de 20 días, y se miden las siguientes variables:

TABLA P-10 Resultado de Minitab					
Correlations					
		Y		X1	
X1	0.827				
X2	0.822	0.680			
Regression Analysis					
The regression equation is					
$Y = -26.7 + .782 X1 + .068 X2$					
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 = el número de empaques de 6 cervezas que se venden diariamente

X_1 = la temperatura diaria más alta

X_2 = el tráfico diario de personas

Se analizan los datos usando regresión múltiple. El resultado parcial de la computadora es la tabla P-10.

- Analice la matriz de correlación.
 - Pruebe la hipótesis $H_0: \beta_j = 0, j = 1, 2$, con un nivel de significancia de .01.
 - Pronostique el volumen de cerveza vendida si la temperatura más alta es de 60 grados Fahrenheit y el tráfico es de 500 personas.
 - Calcule R^2 , e interprete su significado en términos de este problema.
 - Calcule el error estándar de la estimación.
 - Explique cómo se ven afectadas las ventas de cerveza por el incremento de un grado en la temperatura más alta.
 - Redacte sus conclusiones de este análisis en relación con la exactitud de la ecuación del pronóstico y también sobre las contribuciones de las variables independientes.
11. Una compañía de taxis está interesada en la relación entre el millaje, medido en millas por galón, y la antigüedad de los automóviles de su flota. Los 12 vehículos son de la misma marca y del mismo tamaño y están en buenas condiciones de operación gracias a un mantenimiento sistemático. La compañía emplea tanto a hombres como mujeres como conductores y se cree que algo de la variabilidad en el millaje puede deberse a las diferencias en las técnicas de manejo entre los grupos de conductores de uno y otro sexo. De hecho, si todo lo demás se mantiene igual, las mujeres tienden a obtener mejor millaje que los hombres. Se registran los datos generados por la asignación al azar de los 12 automóviles a cinco mujeres y siete hombres, y se calculan las millas por galón después de recorrer 300 millas. Los datos aparecen en la tabla P-11.
- Elabore un diagrama de dispersión con Y como el eje vertical y X_1 como el eje horizontal. Identifique los puntos correspondientes a los conductores hombres y mujeres, respectivamente.

TABLA P-11

<i>Millas por galón Y</i>	<i>Antigüedad del auto (años) X_1</i>	<i>Sexo (0 = hombre, 1 = mujer) X_2</i>
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

b) Ajuste el modelo de regresión

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

interprete el coeficiente de mínimos cuadrados, b_2 .

- c) Calcule los valores ajustados de cada uno de los pares (X_1, X_2) y grafique los valores ajustados en el diagrama de dispersión. Dibuje líneas rectas a través de los valores ajustados para los conductores hombres y las conductoras mujeres, respectivamente. Especifique las ecuaciones para estas dos líneas rectas.
- d) Suponga que se ignora el género. Ajuste el modelo de regresión lineal simple, $Y = \beta_0 + \beta_1 X_1 + \varepsilon$, y grafique la línea recta ajustada sobre el diagrama de dispersión. ¿Es importante incluir los efectos del género en este caso? Explique.
12. El gerente de ventas de un distribuidor grande de partes automotrices, Hartman Auto Supplies, quiere desarrollar un modelo para pronosticar desde mayo las ventas anuales totales de una región. Si es posible pronosticar las ventas regionales, entonces es factible pronosticar las ventas totales de la compañía. El número de minoristas en la región en cuyos inventarios hay refacciones de la compañía y el número de automóviles registrados para cada región al primero de mayo son las dos variables independientes investigadas. Los datos aparecen en la tabla P-12.
- a) Analice la matriz de correlación.
- b) ¿Qué tanto de error está implicado en el pronóstico de la región 1?
- c) Pronostique las ventas anuales para la región 12, dados 2,500 minoristas y 20.2 millones de automóviles registrados.
- d) Discuta la exactitud del pronóstico elaborado en el inciso c).
- e) Indique cómo encontró el error estándar de la estimación.
- f) Dé una interpretación de los coeficientes de regresión parciales. ¿Son razonables estos coeficientes de regresión?
- g) ¿Cómo se puede mejorar esta ecuación de regresión?
13. El director de ventas de Hartman Auto Supplies decide investigar una nueva variable independiente, el ingreso personal por región (véase el problema 12). Los datos para esta nueva variable se presentan en la tabla P-13.
- a) ¿El ingreso personal por región hace una contribución al pronóstico de las ventas?

TABLA P-12

<i>Región X_2</i>	<i>Ventas anuales (millones de \$) Y</i>	<i>Número de minoristas X_1</i>	<i>Número de automóviles registrados (\$ millones)</i>
1	52.3	2,011	24.6
2	26.0	2,850	22.1
3	20.2	650	7.9
4	16.0	480	12.5
5	30.0	1,694	9.0
6	46.2	2,302	11.5
7	35.0	2,214	20.5
8	3.5	125	4.1
9	33.1	1,840	8.9
10	25.2	1,233	6.1
11	38.2	1,699	9.5

TABLA P-13

<i>Región</i>	<i>Ingreso personal (miles de millones de \$)</i>	<i>Región</i>	<i>Ingreso personal (miles de millones de \$)</i>
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		

- b) Pronostique las ventas anuales para la región 12 para un ingreso personal de \$40 mil millones y las cifras correspondientes a los minoristas y los automóviles registrados de acuerdo con el inciso c) del problema 12
- c) Discuta la exactitud del pronóstico elaborado en el inciso b).
- d) ¿Cuáles variables independientes incluiría usted en su modelo final de pronóstico? ¿Por qué?
14. Nelson Corporation decide desarrollar una ecuación de regresión múltiple para pronosticar el desempeño del área de ventas. Se entrevista a 14 vendedores de una muestra aleatoria y se les aplica una prueba de aptitud. También se calcula un índice de esfuerzo realizado por cada vendedor sobre la base de una razón entre el millaje recorrido en el automóvil de su compañía y el total de millaje proyectado para una cobertura adecuada de territorio. El análisis de regresión produce los siguientes resultados:

$$\hat{Y} = 16.57 + \underset{(.05)}{.65} X_1 + \underset{(1.69)}{20.6} X_2$$

Las cantidades entre paréntesis son los errores estándar de los coeficientes de regresión parciales. El error estándar de la estimación es 3.56. La desviación estándar de la variable de ventas es $s_y = 16.57$. Las variables son

Y = el desempeño de las ventas, en miles
 X_1 = la calificación en la prueba de aptitud
 X_2 = el índice de esfuerzo

- a) ¿Los coeficientes de regresión parciales son significativamente diferentes de cero en el nivel de significancia de .01?
- b) Interprete el coeficiente de regresión parcial para el índice de esfuerzo.
- c) Pronostique el desempeño en las ventas de un vendedor que obtiene una calificación de 75 en la prueba de aptitud y un índice de esfuerzo de .5.
- d) Calcule la suma de los cuadrados residuales, $\Sigma(Y - \hat{Y})^2$.
- e) Calcule la suma total de cuadrados $\Sigma(Y - \bar{Y})^2$.
- f) Calcule R^2 , e interprete este número en términos del problema.
- g) Calcule el coeficiente ajustado de determinación, $\bar{R}^2$.
15. Cabe esperar que las compras con tarjeta de crédito sean diferentes de las compras en efectivo en la misma tienda. La tabla P-15 indica las ventas diarias brutas y los artículos vendidos que se pagan en efectivo, así como las ventas diarias brutas en efectivo y los artículos vendidos que se pagan con tarjetas de crédito en la misma tienda de consignación por 25 días consecutivos.
- a) Elabore un diagrama de dispersión de las ventas brutas diarias, Y , contra artículos vendidos que se pagan en efectivo, X_1 . Usando otros símbolos, o bien, colores diferentes, agregue las ventas brutas diarias y los artículos vendidos que se pagan con tarjeta de crédito, X_2 . Compare visualmente la relación entre las ventas y el número de artículos vendidos que se pagan en efectivo con los pagados con tarjeta de crédito.

TABLA P-15

<i>Día</i>	<i>Ventas brutas en efectivo (\$)</i>	<i>Número de artículos</i>	<i>Ventas brutas con tarjeta de crédito (\$)</i>	<i>Número de artículos</i>
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

b) Defina la variable ficticia

$$X_2 = \begin{cases} 1 & \text{si la compra es en efectivo} \\ 0 & \text{si la compra es con tarjeta de crédito} \end{cases}$$

y ajuste el modelo de regresión

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- c) Analice el ajuste del inciso b). Asegúrese de incluir un análisis de los residuos. ¿Está usted satisfecho con su modelo?
- d) Con base en el modelo ajustado del inciso b), elabore un pronóstico de las ventas diarias para un individuo que compra 25 artículos y paga en efectivo. Para una muestra grande, construya un intervalo de predicción del 95% 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, una para las ventas en efectivo y otra para las ventas con tarjeta de crédito, para los datos de la tabla P-15? Discútalos.
16. Cindy Lawson acaba de comprar un equipo de la liga mayor de béisbol. Ella ha estado recibiendo muchos consejos acerca de lo que debe hacer para convertir a su equipo en ganador. Cindy le pide a usted que estudie este problema y escriba un reporte. Usted decide usar el análisis de regresión múltiple para determinar cuáles datos estadísticos son importantes para desarrollar un equipo ganador (considerado como tal por el número de juegos ganados durante la temporada 1991). Usted reúne los datos de seis estadísticos del *Sporting News*

TABLA P-16

<i>Equipo</i>	<i>Victorias</i>	<i>ERA</i>	<i>SO</i>	<i>BA</i>	<i>Carreras</i>	<i>HR</i>	<i>SB</i>
Gigantes	75	4.03	905	.246	649	141	95
Mets	77	3.56	1,028	.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	1,033	.244	605	79	125
Dodgers	93	3.06	1,028	.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	1,003	.255	702	126	97
Rangers	85	4.47	1,022	.270	829	177	102
Atléticos	84	4.57	892	.248	760	159	151
Reales	82	3.92	1,004	.264	727	117	119
Ángeles	81	3.69	990	.255	653	115	94
Gemelos	95	3.69	876	.280	776	140	107

NOTA: *ERA* significa promedio de carreras ganadas (*earned run average*); *SO* significa ponches (*strikeouts*); *BA* significa promedio de bateo (*batting average*); *HR* significa *homeruns*; y *SB* significa bases robadas (*stolen bases*).

1992 *Baseball Yearbook*, los cuales se muestran en la tabla P-16, y corre un programa de regresión por pasos, suponiendo un modelo de regresión múltiple con “Victorias” como la variable dependiente.

- a) Discuta la importancia de cada variable independiente.
 - b) ¿Cuál ecuación debe usar Cindy para pronosticar las victorias?
 - c) Escriba un reporte para Cindy.
 - d) Recopile datos de los números más recientes del *Sporting News Baseball Yearbook* u otra fuente de datos estadísticos de béisbol. Corra una regresión por pasos y compare sus resultados.
17. La señora Haight, una agente inmobiliaria, desea pronosticar la importancia de cuatro factores en la determinación de los precios de terrenos. Ella recopila datos sobre el precio, el área, la elevación y la inclinación, y, con base en ello, califica la vista de 50 lotes. Ingresa los datos en un programa de correlación y obtiene la matriz de correlación que se presenta en la tabla P-17. Luego, Haight ingresa los datos en un programa de regresión múltiple por pasos.
- a) Determine cuál variable entrará en primer lugar, segundo, tercero y último lugar al modelo.
 - b) ¿Qué variable o variables serán incluidas en la mejor ecuación de predicción?
18. En la tabla P-18 se listan: la calificación de dos exámenes parciales, X_1 y X_2 ; el promedio actual de calificaciones (GPA), X_3 ; y la calificación del examen final, Y , de 20 estudiantes en la clase de estadística para los negocios.
- a) Ajuste un modelo de regresión múltiple para predecir la puntuación del examen final a partir de las calificaciones de los exámenes parciales y el GPA. ¿Es significativa la regresión? Explique.
 - b) Pronostique la calificación del examen final de un estudiante con calificaciones parciales de 86 y 77 y un GPA de 3.4.
 - c) Calcule los *VIF* y examine los estadísticos t para probar la significancia de las variables explicativas individuales. ¿La multicolinealidad es un problema? Explique.

TABLA P-17

Variable	Variable				
	Precio	Área	Elevación	Pendiente	Vista
Precio	1.00	.59	.66	.68	.88
Área		1.00	.04	.64	.41
Elevación			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

- d) Calcule el apalancamiento medio. ¿Algunas de las observaciones son puntos de apalancamiento alto?
- e) Calcule los residuos estandarizados. Identifique cualquier observación con residuo estandarizado grande. ¿El modelo ajustado predice una respuesta muy alta o muy baja de estas observaciones?
19. Con respecto a los datos de la tabla P-18, obtenga el “mejor” modelo de regresión usando el procedimiento de regresión por pasos y el procedimiento de todas las regresiones posibles. Compare los resultados. ¿Se siente usted confiado al usar un modelo de regresión para predecir la calificación del examen final con menos variables independientes que las tres originales?
20. Recuerde el ejemplo 7.12. El conjunto completo de datos relacionado con la remuneración del director general aparece en el apéndice C. (Consulte las páginas 533-545). Use la regresión por pasos para seleccionar el “mejor” modelo con $k = 3$ variables explicativas. Ajuste el modelo de pasos e interprete los coeficientes calculados. Examine los residuos. Identifique y explique cualquier observación con influencia. Si usted tuviera que elegir entre este modelo y el modelo de pronóstico con $k = 2$ analizado en el ejemplo 7-12, ¿cuál de ellos seleccionaría? ¿Por qué?
21. La tabla P-21 contiene el número de cuentas (en miles) y los activos (en miles de millones de dólares) de 10 corredurías de bolsa *online*. Grafique los activos contra el número de cuentas. Investigue la posibilidad de que la relación sea curva elaborando una regresión múltiple para pronosticar los activos; considere el número de cuentas y el número de cuentas elevado al cuadrado como variables independientes.
- a) Proporcione la función de regresión ajustada. ¿Es significativa la regresión? Explique.
- b) Pruebe la significancia del coeficiente del término elevado al cuadrado. Resuma su conclusión.
- c) Corra otra vez el análisis sin el término cuadrático (elevado al cuadrado). Explique por qué el coeficiente del número de cuentas no es el mismo que el que usted obtuvo en el inciso a).
22. Existen catadores de queso que determinan la calidad de estos productos; las puntuaciones que otorgan se suman para obtener una variable dependiente denominada simplemente “sabor”. Las variables independientes (explicativas) son tres sustancias químicas presentes en el queso: ácido acético, sulfuro de hidrógeno (H_2S), y ácido láctico. Los 15 casos del conjunto de datos se presentan en la tabla P-22. Analice estos datos usando métodos de regresión múltiple. Asegúrese de incluir solamente variables independientes significativas en su modelo final e interprete R^2 . Incluya un análisis de los residuos.
23. Remítase al problema 22. Usando su función final de regresión ajustada, pronostique Sabor (calidad) para Acético = 5.750, H_2S = 7.300, y Láctico = 1.85. (Es posible que no se

TABLA P-21

<i>Activos (miles de millones de) X</i>	<i>Número de cuentas (en miles) Y</i>
219.0	2,500
21.1	909
38.8	615
5.5	205
160.0	2,300
19.5	428
11.2	590
5.9	134
1.3	130
6.8	125

TABLA P-22

<i>Caso</i>	<i>Sabor Y</i>	<i>Acético X₁</i>	<i>H₂S X₂</i>	<i>Láctico X₃</i>
1	40.9	6.365	9.588	1.74
2	15.9	4.787	3.912	1.16
3	6.4	5.412	4.700	1.49
4	18.0	5.247	6.174	1.63
5	38.9	5.438	9.064	1.99
6	14.0	4.564	4.949	1.15
7	15.2	5.298	5.220	1.33
8	32.0	5.455	9.242	1.44
9	56.7	5.855	10.199	2.01
10	16.8	5.366	3.664	1.31
11	11.6	6.043	3.219	1.46
12	26.5	6.458	6.962	1.72
13	0.7	5.328	3.912	1.25
14	13.4	5.802	6.685	1.08
15	5.5	6.176	4.787	1.25

TABLA P-24 Categorías contables para la liga mayor de béisbol (en millones de \$)

<i>GtReceipt</i>	<i>MediaRev</i>	<i>StadRev</i>	<i>TotRev</i>	<i>PlayerCt</i>	<i>OpExpens</i>	<i>OpIncome</i>	<i>FranValu</i>	<i>Franchise</i>
19.4	61.0	7.4	90.0	29.8	59.6	30.4	200	NyYank
26.6	32.5	18.0	79.3	36.0	72.0	7.3	180	LADodge
22.9	50.0	16.0	91.1	35.2	70.4	20.7	170	NYMets
44.5	30.0	12.0	88.7	29.7	62.4	26.3	160	TorBIJay
24.5	40.5	14.3	81.5	35.4	70.8	10.7	160	BosRdSox
19.0	24.4	5.0	50.6	15.8	39.5	11.1	140	BaltOrio
27.5	25.7	19.3	78.0	18.0	60.0	18.0	140	ChWhSox
19.9	25.0	12.0	59.1	23.2	46.4	12.7	132	StLCard
22.8	27.5	12.0	64.5	29.0	58.0	6.5	132	ChCubs
19.0	25.5	14.0	61.5	20.7	47.6	13.9	123	TxRanger
16.9	20.5	14.0	53.6	30.4	60.8	-7.2	117	KCRoyals
15.2	23.2	7.6	48.2	21.7	43.4	4.8	115	PhilPhil
25.7	27.0	10.0	64.9	39.2	66.6	-1.7	115	OakAthlt
19.0	27.9	5.0	54.1	34.3	61.7	-7.6	103	CalAngls
15.5	26.2	5.0	48.9	33.3	53.3	-4.4	99	SFGiants
17.1	24.4	5.3	49.0	27.1	48.8	0.2	98	CinnReds
15.6	23.1	7.5	48.4	24.4	48.8	-0.4	96	SDPadres
10.6	25.2	8.0	46.0	12.1	31.5	14.5	95	HousAstr
16.2	21.9	5.5	45.8	24.9	49.8	-4.0	87	PittPirt
15.6	28.8	5.0	51.6	31.1	54.4	-2.8	85	DetTiger
15.4	18.9	3.8	40.3	20.4	40.8	-0.5	83	AtBraves
18.2	20.5	3.2	44.1	24.1	48.2	-4.1	83	MinnTwin
15.5	22.0	5.0	44.7	17.4	41.8	2.9	79	SeatMar
14.2	19.4	3.0	38.8	26.4	50.2	-11.4	77	MilBrews
9.5	23.7	6.6	42.0	19.5	46.8	-4.8	77	ClevIndn
10.7	23.5	3.0	39.4	21.8	43.6	-4.2	75	MonExpos

Fuente: M. Ozanian and S. Taub, "Big Leagues, Bad Business," *Financial World*, July 7, 1992, pp. 34-51.

requieran las tres variables independientes). Si bien n es pequeña en este caso, elabore el intervalo de predicción del 95% para una muestra grande. ¿Siente usted que el análisis de regresión ha producido una herramienta útil para pronosticar la calidad del queso? Explique

24. Las cifras contables de 1991 de la liga mayor de béisbol aparecen en la tabla P-24. Todas las cifras están en millones de dólares. Las variables numéricas son GtReceit (entradas con boleto), MediaRev (ingreso promedio), StadRev (ingreso del estadio), TotRev (ingreso total), PlayerCt (costos de jugadores), OpExpens (gastos operativos), OpIncome (ingreso operativo = ingreso total – gastos operativos) y FranValu (valor de la franquicia).
 - a) Elabore una matriz de correlación para las variables GtReceit, MediaRev, ..., FranValu. A partir de la matriz de correlación, ¿puede usted determinar una variable que sea un buen factor para explicar FranValu? Discútalos.
 - b) Use la regresión por pasos para elaborar un modelo que permita pronosticar el valor de la franquicia usando las variables restantes. ¿Está sorprendido con el resultado? Explique.
 - c) ¿Podría usted concluir que, como regla general, el valor de la franquicia es aproximadamente el doble del ingreso total? Explique
 - d) Los costos de los jugadores son una gran parte de los costos operativos. Desarrolle una ecuación para pronosticar los gastos operativos a partir de los costos de jugadores. Comente acerca de la fortaleza de la relación. Usando los residuos como guía, identifique los equipos que tienen costos de jugadores inusualmente bajos o inusualmente altos como un componente de los gastos operativos.
 - e) Considere las variables diferentes a FranValu. De acuerdo con sus definiciones, ¿existen grupos de variables que sean multicolineales? Si es así, identifique esos grupos.

CASOS

CASO 7-1 EL MERCADO DE BONOS¹⁶

Judy Johnson, vicepresidente financiera de una gran compañía privada que presta servicios públicos en el noroeste del país, se enfrentó con un problema de financiamiento. La compañía necesitaba dinero para pagar deudas de corto plazo por vencer y para continuar la construcción de una planta activada por carbón.

La principal preocupación de Judy era la estimación del mercado de bonos de 10 o 30 años; la compañía necesitaba decidir si usar financiamiento de capital o deuda a largo plazo. Para tomar esta decisión, la institución necesitaba un pronóstico confiable de la tasa de interés que pagaría en el momento de la emisión de los bonos.

Judy convocó a una junta al personal de finanzas para discutir el problema de los bonos. Un miembro de su equipo, Ron Peterson, con una maestría en administración, dijo que, en su opinión, se podía desarrollar un modelo de regresión múltiple para pronosticar las tasas de los bonos. Puesto que la vicepresidente no estaba familiarizada con la regresión múltiple, condujo la discusión en otra dirección. Después de una hora de interacción improductiva, Judy pidió a Ron un reporte al respecto para el siguiente lunes.

Ron sabía que la clave para el desarrollo de un buen modelo de pronóstico es la identificación de las variables independientes que se relacionan con la tasa

¹⁶Los datos de este caso fueron proporcionados por Dorothy Mercer, una estudiante de la maestría en administración de la Universidad del Este de Washington. El análisis fue realizado por los estudiantes de maestría Tak Fu, Ron Hand, Dorothy Mercer, Mary Lou Redmond y Harold Wilson.

de interés pagada por las empresas en el momento de la emisión de los bonos. Después de discutir el problema con varias personas de la institución, Ron decidió investigar las siguientes variables: un índice (Moody's) para la calidad de bonos de una empresa de servicios públicos, la tasa de ganancias de renta fija de una empresa de servicios públicos, las tasas de interés de los bonos del Tesoro de Estados Unidos, el vencimiento de los bonos (a 10 o 30 años) y la prima de interés preferencial al momento de la emisión.

Ron recopiló los datos que podrían correlacionarse con las tasas de interés de bonos para la emisión de bonos de la empresa durante los dos años anteriores. Al principio, no estaba seguro de cómo manejar las tasas de los bonos. Finalmente, decidió considerar sólo empresas cuyas tasas fueran iguales o ligeramente superiores a las de su compañía. Esta decisión le proporcionó una muestra de 93 emisiones para analizar. Pero él estaba preocupado acerca de la validez de utilizar la calidad de los bonos como datos de tipo intervalo. Ron llamó a su ex profesor de estadística y entendió que las variables ficticias resolverían el problema. Así, codificó las clasificaciones de los bonos de la siguiente manera:

$X_1 = 1$ si la clasificación de los bonos es A; 0 de lo contrario

$X_2 = 1$ si la clasificación de los bonos es AA; 0 de lo contrario

Si la clasificación del bono es BAA, tanto X_1 como X_2 son 0. El siguiente paso para Ron fue seleccionar un programa de regresión múltiple de la computadora e introducir los datos. Se incluyeron las siguientes variables en la ecuación completa del modelo:

Variable 1: Y = la tasa de interés pagada por la empresa en el momento de la emisión de bonos

Variable 2: $X_1 = 1$ si la clasificación de los bonos de la empresa es A

Variable 3: $X_2 = 1$ si la clasificación de los bonos de la empresa es AA

Variable 4: X_3 = la tasa de ganancia de la empresa contra los cargos fijos

Variable 5: X_4 = las tasas de bonos del tesoro de los Estados Unidos (para 10 y 30 años) en el momento de la emisión de los bonos

Variable 6: X_5 = vencimiento del bono (10 o 30 años)

Variable 7: X_6 = prima de interés preferencial en el momento de la emisión

Los datos reales se presentan en el apéndice A.

Ron decidió analizar la matriz de correlación mostrada en la tabla 7-18. No se sorprendió de encontrar una relación positiva alta entre la tasa de interés pagada por la empresa en el momento de la emisión de los bonos y la tasa $r = .883$ de los bonos del tesoro de Estados Unidos. También esperaba una relación positiva bastante alta entre la variable dependiente y la tasa preferencial ($r = .596$). No estaba muy sorprendido de descubrir que estas dos variables explicativas estuvieran también relacionadas una con otra (multicolinealidad potencial $r = .713$). La relación negativa entre la variable dependiente y la duración del plazo de los bonos (10 o 30 años), $r = -.221$, también era un resultado que tenía sentido para Ron.

En seguida, Ron corrió un modelo completo que contenía todas las variables explicativas. El examen de los valores t y/o los valores p calculados, los cuales se presentan en la tabla 7-19, indicaron que quizá la variable de la tasa de ingresos a cargos fijos y seguramente la variable de tasa de interés preferencial no estaban haciendo una contribución al pronóstico de la tasa de interés pagada por la institución en el momento de la emisión de los bonos.

Para verificar esta interpretación de los resultados de la regresión completa, Ron decidió hacer una regresión por pasos. La salida se presenta en la tabla 7-20.

Si bien el valor p , .035, asociado con el valor t para la razón entre ingresos y cargos fijos era menor que .05, Ron decidió eliminar la variable razón (Ratio) de su ecuación de regresión. Un argumento adicional para apoyar esta decisión lo dio la pequeña correlación ($r = .097$) de la razón con la variable dependiente Rate. Los resultados de Minitab del modelo final de Ron se presentan en la tabla 7-21.

El reporte de Ron para Judy incluía los siguientes comentarios:

1. El mejor modelo, Interés = $-1.28 - .929 \text{ AA} - 1.18 \text{ AA} + 1.23 \text{ tasas de bonos} + 0.0615 \text{ Plazo}$, explica el 90.6% de la variación de la tasa de interés.
2. El error estándar de la estimación es .53 con base en $n = 93$ observaciones. Por lo tanto, para un pronóstico determinado, $\hat{Y}$, el intervalo de predicción del 95% de los valores reales de la variable es $\hat{Y} \pm 2(.53)$ o $\hat{Y} \pm 1.06$.
3. Todos los coeficientes de las variables independientes son significativos (valores muy pequeños de p) y parecen ser confiables. La multicolinealidad no es un problema.

Ron estaba muy satisfecho con su trabajo y pensaba que Judy también lo estaría.

TABLA 7-18 Matriz de correlación para el estudio del mercado de bonos

	<i>Correlaciones</i>					
	<i>Tasa de interés</i>	<i>A</i>	<i>AA</i>	<i>Razón</i>	<i>Tasas de bonos</i>	<i>Plazo</i>
A	−0.347					−0.314
AA	0.173	−0.399				
Razón	0.097	0.037	0.577			
Tasas de bonos	0.883	−0.256	0.291	0.253		
Plazo	−0.221	0.278	0.010	0.094	−0.477	
Tasa preferencial	0.596	−0.152	0.342	0.255	0.713	

TABLA 7-19 Corrida del modelo completo para el estudio del mercado de bonos**Regression Analysis**

The regression equation is

$$\text{Interest Rates} = -1.03 - 0.829 A - 0.889 AA - 0.242 \text{ Ratio} + \text{Bond Rates} \\ + 0.0628 \text{ Maturity} - 0.0031 \text{ Prime}$$

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

$$S = 0.5235 \quad R\text{-Sq} = 91.1\% \quad R\text{-sq(adj)} = 90.5\%$$

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.

PREGUNTA

1. ¿Qué preguntas cree usted que le hará Judy a Ron?

TABLA 7-20 Regresión por pasos para el estudio del mercado de bonos

Stepwise Regression: Int rate versus A, AA, . . .

Alpha-to-Enter: 0.05 Alpha-to-Remove: 0.05

Response is Int rate on 6 predictors, with N = 93

Step	1	2	3	4	5
Constant	1.9818	-0.7641	-0.7372		-1.0244
Bondrate	1.029	1.173	1.267	1.231	1.253
T-Value	17.95	20.44	23.43	24.78	27.21
P-Value	0.000	0.000	0.000	0.000	0.000
Maturity		0.0439	0.0537	0.058	0.0629
T-Value		5.25	7.02	8.37	9.60
P-Value		0.000	0.000	0.000	0.000
Ratio			-0.547		-0.241
T-Value			-5.08	-5.29	-2.14
P-Value			0.000	0.000	0.035
A				-0.56	-0.83
T-Value				-4.49	-6.28
P-Value				0.000	0.000
AA					-0.89
T-Value					-4.10
P-Value					0.000
S	0.800	0.704	0.623	0.565	0.521
R-Sq	77.97	83.14	86.93	89.37	91.09
R-Sq(adj)	77.73	82.76	86.49	88.88	90.58
Mallows Cp	123.6	75.7	41.2	19.6	5.0

TABLA 7-21 Modelo final para el 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	-4.42R
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 echó un vistazo a AAA Washington cuando se pidió a los estudiantes preparar una descomposición de la serie de tiempo de las llamadas al servicio de emergencia en el camino recibidas en el club durante cinco años. El análisis realizado en el caso 5-5 mostró que el patrón del volumen de llamadas del servicio de emergencia en el camino tenía algo de cíclico en su naturaleza. En el caso 6-6 se investigaron cuatro variables: tasa de desempleo, temperatura promedio diaria, precipitación y número de miembros en el club.

La temperatura promedio diaria, la precipitación y, quizás, la tasa de desempleo se identificaron como variables significativas. El número de miembros en el club no era una variable significativa.

Una conversación con el director del centro de llamadas del servicio de emergencia en el camino trajo consigo dos observaciones importantes: **1.** Las fábricas de automóviles parecen diseñar vehículos que operan mejor a 65 grados Fahrenheit y **2.** el volumen de llamadas

parece incrementarse más marcadamente cuando la temperatura promedio baja unos cuantos grados a partir de una temperatura promedio en el rango de los 30 grados de lo que se incrementa cuando una caída similar ocurre con una temperatura promedio en el rango de los 60 grados. Esta información sugería que el efecto de la temperatura sobre el servicio de emergencia en el camino no era lineal.

Michael DeCoria estableció en el caso 6-6 que creía que el número de llamadas recibidas en el servicio estaba relacionado con el ciclo general de la economía y que la tasa de desempleo en el estado de Washington es una buena medida sustituta del estado general de la economía en el estado de Washington. Ahora ha observado que la tendencia cíclica de la serie de tiempo parece quedarse rezagada con respecto al ciclo general de la economía. Los datos sobre la temperatura promedio mensual y la tasa de desempleo en el estado de Washington se presentan en la tabla 7-22.

TABLA 7-22 Datos de AAA Washington

<i>Año</i>	<i>Mes</i>	<i>Llamadas</i>	<i>Tasa de desempleo</i>	<i>Temp.</i>
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
	Junio	21,591	5.7592	59
	Julio	22,696	5.5718	63.8
	Agosto	21,509	5.2939	63.8
	Septiembre	22,123	5.4709	59.1
	Octubre	21,449	5.5049	54.6
	Noviembre	23,475	5.8630	45.4
	Diciembre	23,529	6.1349	41
1989	Enero	23,327	7.5474	40.3
	Febrero	24,050	7.8157	34.3
	Marzo	24,010	7.1390	43.2
	Abril	19,735	6.2637	52.5
	Mayo	20,153	5.8332	55.3
	Junio	19,512	5.8077	62.4
	Julio	19,892	5.6713	62.9

TABLA 7-22 (Continuación)				
<i>Año</i>	<i>Mes</i>	<i>Llamadas</i>	<i>Tasa de desempleo</i>	<i>Temp.</i>
1990	Agosto	20,326	5.4977	63.5
	Septiembre	19,378	5.2989	60.9
	Octubre	21,263	5.6028	51.9
	Noviembre	21,443	5.9143	46.2
	Diciembre	23,366	5	41.8
	Enero	23,836	6.1917	41.8
	Febrero	23,336	6.3775	38.9
	Marzo	22,003	5.7234	46.3
	Abril	20,155	4.7792	51.7
	Mayo	20,070	4.5715	54.9
	Junio	19,588	4.3899	59.8
	Julio	20,804	4.2559	66.7
1991	Agosto	19,644	3.9359	66.4
	Septiembre	17,424	3.9048	61.9
	Octubre	20,833	4.4294	50.4
	Noviembre	22,490	5.1523	45.8
	Diciembre	24,861	5.5102	33.9
	Enero	23,441	6.8901	37.9
	Febrero	19,205	7.0308	46.9
	Marzo	20,386	6.7186	43.4
	Abril	19,988	6.1280	49.1
	Mayo	19,077	5.8146	54.3
	Junio	19,141	5.9480	58.2
	Julio	20,883	5.9026	65.4
1992	Agosto	20,709	5.7227	66
	Septiembre	19,647	5.6877	60.9
	Octubre	22,013	6.2922	51
	Noviembre	22,375	7.0615	46.2
	Diciembre	22,727	7.4370	42.4
	Enero	22,367	8.4513	43
	Febrero	21,155	8.7699	46
	Marzo	21,209	8.0728	48.9
	Abril	19,286	7.2392	52.7
	Mayo	19,725	7.0461	58.3
	Junio	20,276	7.0478	63.6
	Julio	20,795	7.1080	64.9
1993	Agosto	21,126	6.7824	65
	Septiembre	20,251	6.7691	58.4
	Octubre	22,069	7.5896	53.2
	Noviembre	23,268	7.9908	44.8
	Diciembre	26,039	8.2460	37.8
	Enero	26,127	9.5301	34.9

PREGUNTAS

1. Desarrolle una ecuación de regresión múltiple usando la tasa de desempleo y la temperatura promedio mensual para pronosticar las llamadas al servicio de emergencias en el camino.
2. Defina una nueva variable de temperatura y relaciónela con el servicio de emergencia en el camino. 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 a 65 grados Fahrenheit, entonces cada grado por arriba o por debajo de 65 grados debe hacer que los vehículos operen menos confiablemente. Para llevar a cabo una transformación de los datos de temperatura que simule este efecto, empiece por sustraer 65 de los valores de temperatura promedio mensual. Esto reposiciona el “cero” a 65 grados Fahrenheit. ¿Deben usarse los valores absolutos de esta nueva variable de temperatura?

3. Defina una nueva variable de tasa de desempleo y relaciónela con el servicio de emergencia en el camino. Dé al desempleo un efecto retrasado en el servicio de emergencia en el camino usando la tasa de desempleo para el mes: **1.** tres meses antes del mes

actual y **2.** 11 meses antes del mes actual de acuerdo con los datos de la variable independiente de desempleo. ¿Qué modelo es mejor para pronosticar? ¿Los signos de los coeficientes de las variables independientes son los que usted esperaría que fueran? ¿Los coeficientes de las variables independientes son significativamente diferentes de cero?

4. Desarrolle una ecuación de regresión múltiple usando la variable de temperatura promedio transformada creada en el paso 2 y la variable de desempleo retrasada creada en el paso 3. ¿Es éste un buen modelo? ¿Se ha violado cualquiera de los supuestos subyacentes?

CASO 7-3 BÉISBOL DE FANTASÍA (A)

Era febrero y John Hanke, un profesor retirado de estadística, se estaba preparando para otra temporada de béisbol de fantasía. En años pasados, sus amigos beisbolistas siempre le decían que seguramente utilizaba sus conocimientos de estadística para obtener alguna ventaja. Por desgracia, eso nunca fue verdad. La enseñanza, la investigación, las publicaciones y los comités de trabajo lo habían mantenido ocupadísimo. Ahora que Hanke está retirado, finalmente tiene el tiempo para aplicar sus conocimientos de estadística en el proceso anual de reclutamiento de jugadores. Cada director técnico tiene \$260 con los cuales hacer ofertas y comprar 23 jugadores (14 bateadores y 9 lanzadores). Cada equipo es clasificado (con base en datos estadísticos reales del jugador de la temporada pasada) en ocho categorías estadísticas. El doctor Hanke estaba muy ocupado en la selección de jugadores que rendirían bien en tres de las cuatro categorías de lanzamiento. En años pasados, su equipo de lanzadores, especialmente sus lanzadores abridores, habían sido el hazmerreír de la liga. La temporada 2007 iba a ser diferente. Él

intentó desarrollar modelos para pronosticar con exactitud los rendimientos de lanzadores abridores.

Las tres categorías que Hanke deseaba investigar eran victorias (VICTORIAS), promedio de carreras ganadas (ERA), y bases por bolas y hits otorgados por entrada (WHIP). Destinó una considerable cantidad de tiempo descargando datos estadísticos de béisbol para lanzadores abridores de la temporada 2006.¹⁷ Intentó desarrollar un modelo de regresión múltiple para pronosticar cada una de las tres categorías de interés. A menudo había predicado a sus estudiantes que la selección de la variable inicial era el aspecto más importante en el desarrollo de un modelo de regresión. Él sabía que, si no tenía buenas variables explicativas, no obtendría ecuaciones útiles de predicción. Después de una cantidad considerable de trabajo, el doctor Hanke seleccionó las cinco variables explicativas potenciales que siguen. También decidió incluir solamente lanzadores abridores que habían lanzado por lo menos 100 entradas durante la temporada. Una parte de los datos de los 138 lanzadores abridores seleccionados se presentan en la tabla 7-23.

TABLA 7-23 Estadísticas de lanzamiento de 138 lanzadores abridores

<i>Lanzador</i>	<i>EQUIPO</i>	<i>ERA</i>	<i>WHIP</i>	<i>CMD</i>	<i>K/9</i>	<i>HR/9</i>	<i>OBA</i>	<i>LANZAMIENTOS</i>
1 Armas, Jr., T	WAS	5.03	1.50	5.7	1.5	1.1	339	1
2 Arroyo, B	CIN	3.29	1.19	6.9	2.9	1.2	289	1
3 Batista, M	ARI	4.58	1.53	4.8	1.3	0.8	343	1
4 Beckett, J	BOS	5.01	1.29	6.9	2.1	1.6	307	1
5 Bedard, E	BAL	3.76	1.35	7.8	2.5	0.7	316	0

¹⁷Los datos estadísticos de lanzamientos de béisbol se descargaron del sitio Web BaseballHQ de Ron Shandler, en www.BaseballHQ.com.

TABLA 7-23 (Continuación)

<i>Lanzador</i>		<i>EQUIPO</i>	<i>ERA</i>	<i>WHIP</i>	<i>CMD</i>	<i>K/9</i>	<i>HR/9</i>	<i>OBA</i>	<i>LANZAMIENTOS</i>
6	Benson, K	BAL	4.82	1.40	4.3	1.5	1.6	325	1
7	Blanton, J	OAK	4.82	1.54	5.0	1.8	0.8	345	1
8	Bonderman, J	DET	4.08	1.30	8.5	3.2	0.8	308	1
9	Bonser, B	MIN	4.22	1.28	7.5	3.5	1.6	304	1
10	Buchholz, T	HOU	5.89	1.25	6.1	2.3	1.7	299	1
11	Buehrle, M	CHW	4.99	1.45	4.3	2.0	1.6	331	0
12	Burnett, A	TOR	3.98	1.30	7.8	3.0	0.9	309	1
13	Bush, D	MIL	4.41	1.14	7.1	4.4	1.1	280	1
14	Byrd, P	CLE	4.88	1.51	4.4	2.3	1.3	341	1
15	Cabrera, D	BAL	4.74	1.58	9.5	1.5	0.7	351	1
16	Cain, M	SF	4.15	1.28	8.4	2.1	0.8	305	1
17	Capuano, C	MIL	4.03	1.25	7.1	3.7	1.2	299	0
18	Carpenter, C	STL	3.09	1.07	7.5	4.3	0.9	268	1
19	Carrasco, H	LAA	3.41	1.20	6.5	2.7	0.9	291	1
20	Chacon, S	PIT	6.36	1.72	5.1	1.0	1.9	370	1
	:	:	:	:	:	:	:	:	:
	:	:	:	:	:	:	:	:	:
119	Tomko, B	LAD	4.73	1.35	6.1	2.6	1.4	317	1
120	Trachsel, S	NYM	4.97	1.60	4.3	1.0	1.3	354	1
121	Vargas, C	ARI	4.83	1.41	6.6	2.4	1.4	326	1
122	Vazquez, J	CHW	4.84	1.29	8.2	3.3	1.0	307	1
123	Verlander, J	DET	3.63	1.33	6.0	2.1	1.0	313	1
124	Wakefield, T	BOS	4.63	1.33	5.8	1.8	1.2	313	1
125	Wang, C	NYN	3.63	1.31	3.1	1.5	0.5	309	1
126	Washburn, J	SEA	4.67	1.35	5.0	1.9	1.2	317	0
127	Weaver, J	STL	5.76	1.51	5.6	2.3	1.8	341	1
128	Weaver, J	LAA	2.56	1.03	7.7	3.2	1.1	261	1
129	Webb, B	ARI	3.10	1.13	6.8	3.6	0.6	279	1
130	Westbrook, J	CLE	4.17	1.43	4.6	2.0	0.6	329	1
131	Williams, W	SD	3.65	1.29	4.5	2.1	1.3	306	1
132	Willis, D	FLA	3.87	1.42	6.4	1.9	0.8	327	0
133	Woods, J	SEA	4.20	1.60	5.7	1.2	1.0	354	0
134	Wright, J	SF	5.19	1.48	4.6	1.2	0.9	336	1
135	Wright, J	NYN	4.49	1.52	5.4	1.5	0.6	343	1
136	Young, C	SD	3.46	1.13	8.2	2.4	1.4	279	1
137	Zambrano, C	CHC	3.41	1.29	8.8	1.8	0.8	307	1
138	Zito, B	OAK	3.83	1.40	6.1	1.5	1.1	324	0

*El conjunto completo de datos está disponible en el sitio Web: www.prenhall.com/hanke under Chapter 7 Case 7-3.

Las variables se definen de la siguiente forma:

ERA: Promedio de carreras ganadas o número de carreras permitidas por partido (en nueve entradas)

WHIP: Número de bases por bolas más hits cedidos por entrada lanzada

CMD: Dominio de lanzamientos, la razón de ponches/bases por bolas

K/9: Bateadores que un lanzador poncha por partido (en nueve entradas lanzadas)

HR/9: *Homeruns* de la oposición por partido (en nueve entradas lanzadas)

OBA: Promedio de bateo de la oposición

THROWS: Lanzador diestro (1) o lanzador zurdo (0)

El siguiente paso en el análisis fue la creación de la matriz de correlación presentada en la tabla 7-24. Hanke encontró que las correlaciones entre ERA y WHIP, y entre ERA y OBA eran iguales, .825. Más aún,

TABLA 7-24 Correlaciones: ERA, THROWS, WHIP, K/9, CMD, HR/9, OBA

	ERA	THROWS	WHIP	K/9	CMD	HR/9
THROWS	0.036					
WHIP	0.825	-0.039				
K/9	-0.445	-0.061	-0.503			
CMD	-0.462	0.121	-0.664	0.570		
HR/9	0.562	0.023	0.341	-0.238	-0.184	
OBA	0.825	-0.039	0.998	-0.510	-0.673	0.341

la correlación entre WHIP y OBA, .998, indicaba que estas variables tenían una relación lineal fuerte. En consecuencia, los lanzadores que estaban bien en una de estas variables debían estar bien en las otras dos. Hanke determinó que el ERA es el mejor indicador del rendimiento y decidió correr una regresión para saber qué tan bien pronosticaban el ERA las variables en este conjunto. Sin embargo, sabía que la alta correlación entre WHIP y OBA crearían un problema de multicolinealidad, de manera que sólo una de estas variables sería requerida en la función de regresión. Hanke seleccionó el OBA lanzando una moneda al aire y corrió una regresión con el ERA como variable dependiente y el resto de variables, con excepción de WHIP, como variables independientes. Los resultados se presentan en la tabla 7-25.

De la tabla 7-25 se desprende que las variables independientes THROWS y K/9 no son significativas, dadas las otras variables en la función de regresión. Más aún, los pequeños *VIF* sugieren que THROWS y K/9 se pueden eliminar juntas y los coeficientes de las variables restantes no cambiarán mucho. La tabla 7-26 presenta el resultado de cuando tanto THROWS como K/9 se dejan fuera del modelo. La R^2 es 78.1%, y la ecuación se ve bien. El estadístico *t* de cada una de las variables explicativas es grande con un valor *p* muy pequeño. Los *VIF* son pequeños para las tres variables explicativas, lo que indica que la multicolinealidad ya no es problema.

Hanke determinó que tiene un buen modelo y desarrolló las gráficas residuales que se presentan en la figura 7-4.

TABLA 7-25 Resultado de la regresión de Minitab usando todas las variables explicativas a excepción de WHIP**Regression Analysis: ERA versus THROWS, K/9, CMD, HR/9, OBA**

The regression equation is

$$\text{ERA} = -5.54 + 0.0779 \text{ THROWS} - 0.0303 \text{ K/9} + 0.125 \text{ CMD} + 0.840 \text{ HR/9} + 0.0280 \text{ OBA}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-5.5439	0.7395	-7.50	0.000	
THROWS	0.07792	0.08507	0.92	0.361	1.042
K/9	-0.03028	0.02977	-1.02	0.311	1.602
CMD	0.12517	0.04671	2.68	0.008	2.181
HR/9	0.8404	0.1198	7.02	0.000	1.151
OBA	0.027991	0.002025	13.82	0.000	2.068

S = 0.437305 R-Sq = 78.4% R-Sq(adj) = 77.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	5	91.783	18.357	95.99	0.000
Residual Error	132	25.243	0.191		
Total	137	117.026			

TABLA 7-26 Resultado de la regresión final de Minitab para pronosticar el ERA**Regression Analysis: ERA versus CMD, HR/9, OBA**

The regression equation is

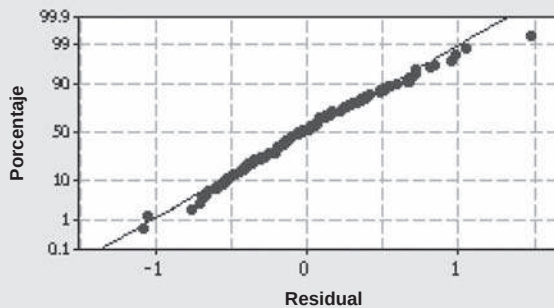
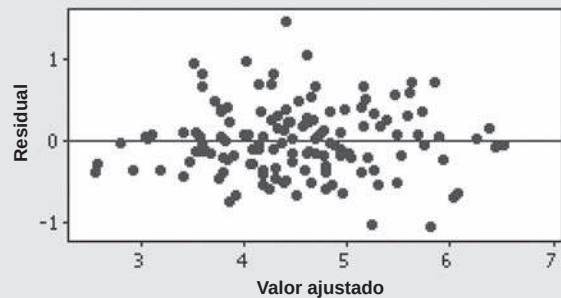
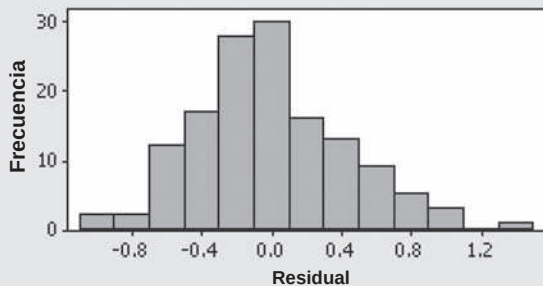
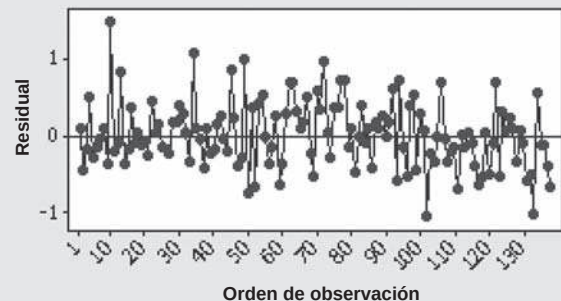
$$\text{ERA} = -5.81 + 0.113 \text{ CMD} + 0.857 \text{ HR/9} + 0.0284 \text{ OBA}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-5.8088	0.6862	-8.46	0.000	
CMD	0.11298	0.04290	2.63	0.009	1.837
HR/9	0.8573	0.1191	7.20	0.000	1.137
OBA	0.028425	0.001997	14.23	0.000	2.009

S = 0.437628 R-Sq = 78.1% R-Sq(adj) = 77.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	91.363	30.454	159.01	0.000
Residual Error	134	25.663	0.192		
Total	137	117.026			

Gráficas residuales para ERA (variables explicativas: CMD, HR/9, OBA)**Gráfica de probabilidad normal****Contra ajustes****Histograma****Contra orden****FIGURA 7-4 Gráficas residuales del pronóstico del ERA**

PREGUNTAS

1. Haga comentarios sobre el modelo que desarrolló Hanke para pronosticar el ERA. Examine las gráficas residuales de la figura 7-4 y determine si este modelo es válido.
2. ¿Existen algunas relaciones no lineales entre las variables explicativas y el promedio de carreras ganadas? Si es así, desarrolle un modelo nuevo incluyendo la transformación apropiada de la variable.
3. Desarrolle un modelo para pronosticar el ERA usando la variable explicativa WHIP en vez de OBA. ¿Qué modelo prefiere usted, el de OBA como variable explicativa o el de WHIP como variable explicativa? ¿Por qué

CASO 7-4 BÉISBOL DE FANTASÍA (B)¹⁸

Ahora que el doctor Hanke siente que ha desarrollado con éxito un modelo para pronosticar el ERA, está listo para ocuparse de las victorias (WINS).¹⁹ Sin embargo, el consenso de los expertos indicaba que el proyecto estaba destinado al fracaso. Los comentarios en el sitio Web BaseballHQ de Ron Shandler decían: “No hay manera de proyectar con exactitud a los lanzadores año tras año” (Bill James); “Tu activo más valioso es un lanzador abridor con el que puedas contar. El único problema es que no puedes contar con ninguno de ellos” (Peter Golenbock); “¿En qué otro sitio del reino de los deportes fantásticos puede usted tener peores augurios sobre el éxito que en el mundo maravilloso de los lanzadores?” (Rod Beaton); “Los lanzadores abridores son el grupo de personas menos confiables, más impredecibles y más desagradables del mundo; desde el punto de vista estadístico, así es” (John Benson)

Hanke investigó y encontró un modelo estadístico que podía utilizarse. De acuerdo con BaseballHQ de Ron Shandler, cuatro variables tienen un efecto importante sobre las victorias totales de lanzamiento:

1. Ofensiva del equipo
2. Efectividad del lanzamiento

3. Apoyo de lanzadores de reserva
4. Defensa del equipo

Hanke estaba listo para desarrollar una base de datos predictivos.

Para proyectar una ofensiva del equipo, seleccionó carreras anotadas por un equipo (RUNS)²⁰ como la variable más importante. Para indicar qué tan bueno era un equipo en todo, Hanke seleccionó las victorias del equipo (TMWINS). Seis de estas siete variables usadas en el caso 7-3 se utilizaron para indicar la efectividad de un lanzador: ERA, CMD, K/9, HR/9, OBA y THROWS. Para la defensa del equipo, utilizó el número total de errores del equipo (ERRORS). Para el apoyo de lanzadores de reserva, intentó con salvamentos (SAVES).

Un conjunto parcial de datos para las variables de efectividad en el lanzamiento están dadas en la tabla 7-23. Algunos de los datos de las variables restantes se presentan en la tabla 7-27. Los datos restantes tanto para la efectividad en el lanzamiento como datos estadísticos del equipo están disponibles en el sitio Web: www.prenhall.com/hanke en el capítulo 7, caso 7-4.

¹⁸El caso 7-4 está basado en un modelo desarrollado por BaseballHQ de Ron Shandler.

¹⁹Los datos estadísticos de lanzamiento de béisbol fueron descargados del sitio Web de BaseballHQ de Ron Shandler en www.BaseballHQ.com.

²⁰Los datos estadísticos para RUNS, TMWINS y SAVES son de SportsTicket® y aparecen en www.sports.espn.go.com/mlb/stats.

TABLA 7-27 Datos estadísticos del equipo para lanzadores abridores

<i>Lanzador</i>		<i>WINS</i>	<i>TM WINS</i>	<i>RUNS</i>	<i>ERRORS</i>	<i>SAVES</i>
1	Armas Jr., T	9	71	746	131	32
2	Arroyo, B	14	80	749	128	36
3	Batista, M	11	76	773	104	34
4	Beckett, J	16	86	820	66	46
5	Bedard, E	15	70	768	102	35
6	Benson, K	11	70	768	102	35
7	Blanton, J	16	93	771	84	54
8	Bonderman, J	14	95	822	106	46
9	Bonser, B	7	96	801	84	40
10	Buchholz, T	6	82	735	80	42
11	Buehrle, M	12	90	868	80	46
12	Burnett, A.	10	87	809	99	42
13	Bush, D	12	75	730	117	43
14	Byrd, P	10	78	870	118	24
15	Cabrera, D	9	70	768	102	35
16	Cain, M	13	76	746	91	37
17	Capuano, C	11	75	730	117	43
18	Carpenter, C	15	83	781	98	38
19	Carrasco, H	7	89	766	124	50
20	Chacon, S	7	67	691	104	39
:	:	:	:	:	:	:
:	:	:	:	:	:	:
119	Tomko, B	8	88	820	115	40
120	Trachsel, S	15	97	834	104	43
121	Vargas, C	12	76	773	104	34
122	Vazquez, J	11	90	868	80	46
123	Verlander, J	17	95	822	106	46
124	Wakefield, T	7	86	820	66	46
125	Wang, C	19	97	930	104	43
126	Washburn, J	8	78	756	88	47
127	Weaver, J	8	83	781	98	38
128	Weaver, J	11	89	766	124	50
129	Webb, B	16	76	773	104	34
130	Westbrook, J	15	78	870	118	24
131	Williams, W	12	88	731	92	50
132	Willis, D	12	78	758	126	41
133	Woods, J	7	78	756	88	47
134	Wright, J	6	76	746	91	37
135	Wright, J	11	97	930	104	43
136	Young, C	11	88	731	92	50
137	Zambrano, C	16	66	716	106	29
138	Zito, B	16	93	771	84	54

*El conjunto completo de datos está disponible en el sitio Web: www.prenhall.com/hanke en el capítulo 7, caso 7-4.

PREGUNTA

1. El consenso de los expertos indicaba que el proyecto estaba destinado al fracaso. ¿Estaban en lo correcto?

Aplicaciones de 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 cuál candidato se convertiría en un buen vendedor.

Solución de Minitab

1. Si 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 un archivo, introdúzcalos en una hoja de cálculo. Para correr la regresión por pasos, haga clic en los siguientes menús:

Stat>Regression>Stepwise

2. Aparece la ventana de diálogo Stepwise Regression presentada en la figura 7-5.
 - a) La variable dependiente o respuesta (Response) es C1 o Sales.
 - b) Las variables explicativas son C2 a C6 o Aptitude-GPA.
 - c) Haga clic en Methods para indicar el nivel α .

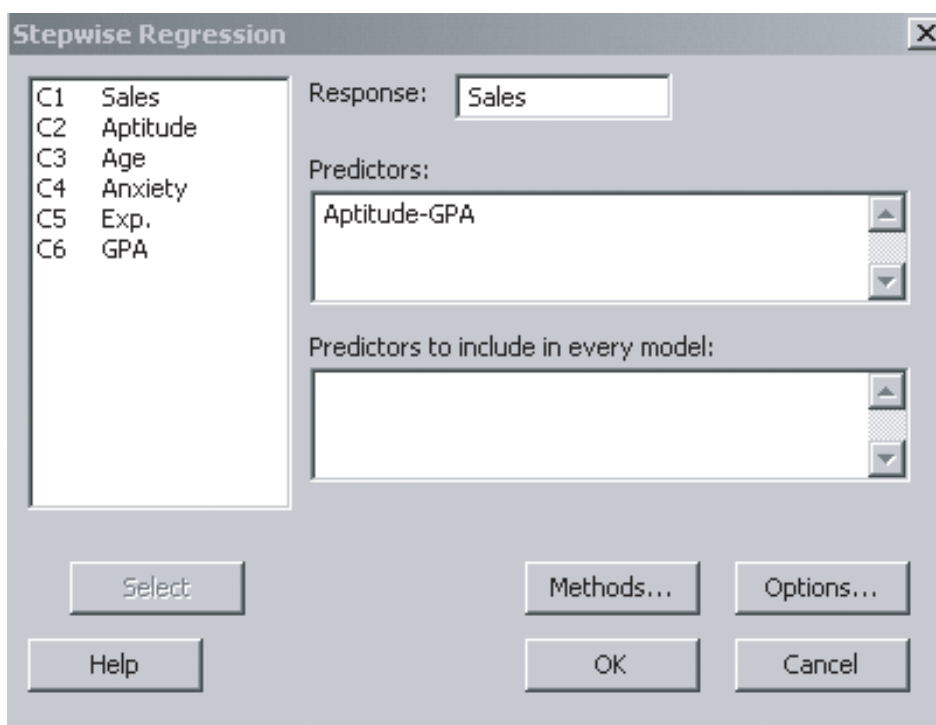


FIGURA 7-5 Ventana de diálogo de la regresión por pasos de Minitab

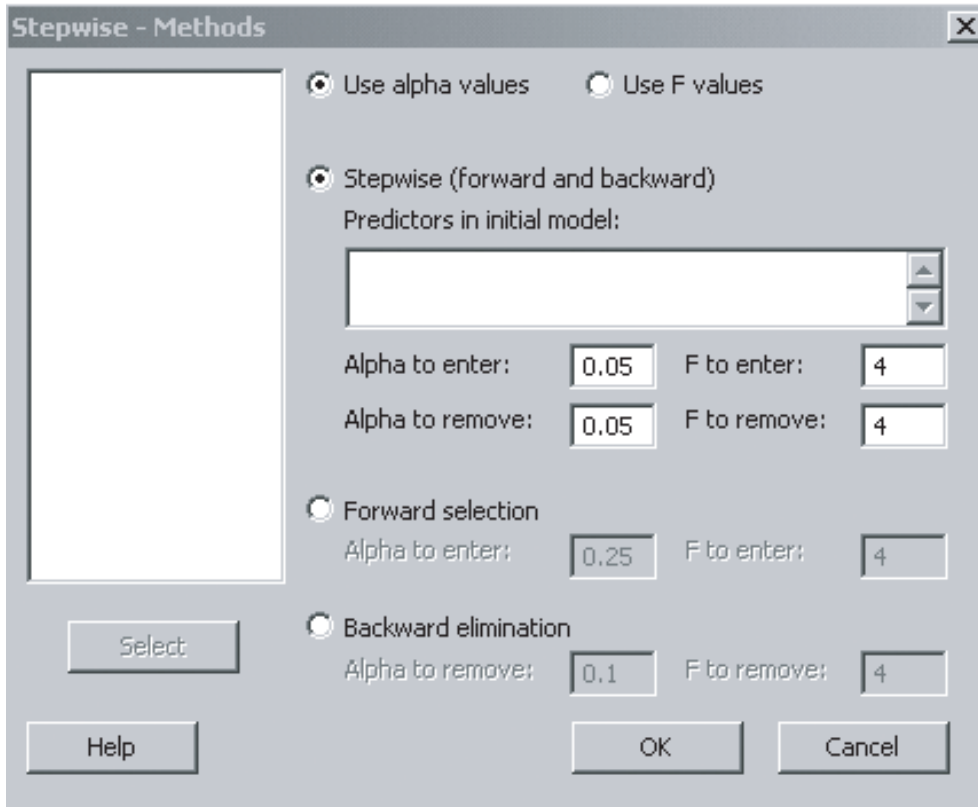


FIGURA 7-6 Ventana de diálogo de Métodos por pasos de Minitab

3. Aparece la ventana de diálogo Stepwise-Methods de la figura 7-6.
 - a) Haga clic en Use alpha values.
 - b) Haga clic en Stepwise (forward and backward).
 - c) Cambie el Alpha to enter y el Alpha to remove de .15 a .05.
 - d) Haga clic en OK y luego en OK de la ventana de diálogo Stepwise Regression. Los resultados se presentan en la tabla 7-17.

Aplicaciones de Excel

El problema. En el ejemplo 7.1 se usó el análisis de la regresión múltiple para determinar si el precio de venta y los gastos de publicidad podían usarse para pronosticar las ventas mensuales de galones de leche (véase la tabla 7-4).

Solución de Excel

1. Introduzca el título Sales en A1, Price en B1, y Adv. En C1. Introduzca las ventas semanales de A2 a A11, precios de venta de B2 a B11, y gastos de publicidad de C2 a C11 de la hoja de cálculo. Las dos variables explicativas deben estar en columnas adyacentes.
2. Haga clic en los siguientes menús:

Tools>Data Analysis

3. Aparecerá la ventana de diálogo Data Analysis. Debajo de Analysis Tools, seleccione Regression. Aparecerá la ventana de diálogo Regression de la figura 6-22.
 - a) Introduzca de A1 a A2 en Input Y Range (rango Y de entrada).
 - b) Introduzca de B1 a C11 en Input X Range para las dos variables explicativas: precio de venta y gastos de publicidad.
 - c) Haga clic en Labels.
 - d) Haga clic en Output Range e introduzca D1.
 - e) Haga clic en OK y aparecerá el resultado en la hoja de cálculo.

Referencias

Abraham, B. y J. Ledolter. <i>Introduction to Regression Modeling</i> . Belmont, Calif.: Thomson Brooks/Cole, 2006.	Johnson, R. A. y D. W. Wichern. <i>Business Statistics: Decision Making with Data</i> . New York: Wiley, 1997.
Belsley, D. A. <i>Conditioning, Diagnostics, Collinearity y Weak Data in Regression</i> . New York: Wiley, 1991.	Kutner, M. H., C. J. Nachtsheim, y J. Neter. <i>Applied Linear Regression Models</i> , 4a. ed. New York: McGraw-Hill, 2004.
Bowerman, B. L., R. T. O'Connell y A. B. Koehler. <i>Forecasting, TimeSeries and Regression</i> , 4a. ed. Belmont, Calif: Thomson Brooks/Cole, 2005.	
Draper, N. R. y H. Smith. <i>Applied Regression Analysis</i> , 3a. ed. New York: Wiley, 1998.	

REGRESIÓN CON DATOS
DE SERIES DE TIEMPO

Muchas aplicaciones de pronósticos en los negocios y en la economía implican series de tiempo. Los modelos de regresión se pueden ajustar a los datos mensuales, trimestrales o anuales, empleando las técnicas descritas en los capítulos anteriores. Sin embargo, como los datos recopilados a lo largo del tiempo son proclives a mostrar tendencias, patrones estacionales, etcétera, las observaciones en diferentes periodos de tiempo están relacionadas o autocorrelacionadas. Es decir, en las series de tiempo, la muestra de observaciones no puede considerarse una muestra aleatoria. Es posible que surjan problemas de interpretación cuando los métodos de regresión estándar se aplican a las observaciones que están relacionadas una con otra en el transcurso del tiempo. El ajuste de los modelos de regresión a las series de tiempo debe realizarse con sumo cuidado.

SERIES DE TIEMPO Y EL PROBLEMA DE AUTOCORRELACIÓN

Los modelos de regresión discutidos en los capítulos 6 y 7 suponen que los errores, ε , son variables aleatorias independientes (o no correlacionadas). Esto significa que los diferentes valores de la variable de respuesta, Y , pueden estar relacionados con los valores de las variables explicativas, las X , pero no entre sí. Las interpretaciones habituales de los resultados de un análisis de regresión dependen en gran parte del supuesto de independencia.

En las series de tiempo el supuesto de independencia rara vez se cumple. Considere el precio base anual del modelo nuevo de un automóvil. ¿Imagina el caos que existiría si los precios de los automóviles nuevos de un año al siguiente no estuvieran relacionados unos con otros (es decir, si fueran independientes)? En un mundo como ése, los precios se determinarían como los números extraídos de una tabla de números aleatorios. El conocimiento del precio en un año no le diría a usted nada acerca del precio para el año siguiente. En el mundo real, el precio vigente en el año en curso está relacionado (correlacionado) con el precio del año anterior, y quizás con el precio de dos años atrás, y así sucesivamente. Es decir, los precios de los diferentes años están autocorrelacionados; no son independientes.

Recordemos lo siguiente a partir del estudio de autocorrelación en el capítulo 3:

Existe *Autocorrelación* cuando las observaciones sucesivas en el tiempo están relacionadas unas con otras.

La autocorrelación ocurre porque el efecto de una variable explicativa sobre la respuesta está distribuido en el tiempo. Por ejemplo, un incremento salarial afectará su consumo (o ahorro) no sólo en el periodo actual, sino también durante varios periodos futuros. Un cambio en el precio afectará las ventas en el periodo actual y en periodos futuros. Un contrato laboral en el presente podría afectar el costo de producción que se registrará en algún tiempo. Las relaciones tienden a ser dinámicas (evolucionan) en el tiempo, no estáticas.

Desde una perspectiva de pronóstico, la autocorrelación no es tan mala. Si los valores de una respuesta, Y , en un periodo están relacionados con los valores Y de periodos anteriores, entonces las Y previas se pueden utilizar para predecir las futuras Y .¹ En un esquema de regresión, la autocorrelación se maneja “reparando” el modelo de regresión estándar. Para acomodar la autocorrelación, algunas veces es necesario cambiar la mezcla de variables explicativas y/o la forma de la función de regresión. Sin embargo, por lo general, la autocorrelación se maneja cambiando la naturaleza del término de error.

Una clase común de autocorrelación, llamada algunas veces *correlación serial de primer orden*, es una en la cual el término de error en el periodo actual está directamente relacionado con el término de error del periodo anterior. En este caso, el subíndice t representa el tiempo, y 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

ε_t = el error en el tiempo

ρ = el parámetro (retraso 1 del coeficiente de autocorrelación) que mide la correlación entre términos de error adyacentes

ν_t = el término de error independiente normalmente distribuido con media 0 y varianza σ_v^2

La ecuación 8.2 dice que el nivel de un término de error (ε_{t-1}) afecta directamente el nivel del término de error siguiente (ε_t). La magnitud del coeficiente de autocorrelación, ρ , donde $-1 \leq \rho < 1$, indica la fortaleza de la correlación serial. Si ρ es cero, entonces no existe correlación serial, y los términos de error son independientes ($\varepsilon_t = \nu_t$).

La figura 8-1 ilustra el efecto de la correlación serial positiva en un modelo de regresión lineal simple. Suponga que la relación verdadera entre Y y X , indicada por la línea continua en la figura, aumenta con el tiempo. Si el primer valor de Y está por encima de la línea de regresión verdadera, entonces los valores siguientes de Y probablemente estarán por arriba de la línea a causa de la autocorrelación positiva (si el primer error es positivo, el segundo error probablemente será positivo, y así sucesivamente). Con el tiempo, puede haber una secuencia de Y debajo de la línea de regresión verdadera (un error negativo probablemente irá seguido de otro error negativo). Los datos están “inclinados” con respecto a la relación verdadera X - Y . Sin embargo, la línea de mínimos cuadrados, por su propia naturaleza, pasará a través de las observaciones, como indica la línea punteada de la figura. Usar la línea punteada para hacer inferencias acerca de la línea continua o usar la línea punteada para generar pronósticos de Y para el futuro podría ser engañoso.

También debe quedar claro en la figura 8-1 que, en este caso, la dispersión alrededor de la línea de mínimos cuadrados es más estrecha de lo que es alrededor de la línea de regresión verdadera. Por consiguiente, el error estándar de la estimación, $s_{y,x}$, subestimaré la variabilidad de las Y alrededor de la línea de regresión verdadera o, de manera equivalente, subestimaré la desviación estándar, σ , del término de error, ε .

Una autocorrelación fuerte podría hacer que dos variables no relacionadas parezcan estar relacionadas. Los procedimientos de regresión estándar aplicados a las observaciones en estas variables pueden generar una regresión significativa. En este caso, la relación estimada es falsa,

¹Esta idea se considerará de nuevo en una sección posterior de este capítulo, cuando se estudien los modelos autorregresivos, y se desarrollará con mayor profundidad en el capítulo 9.

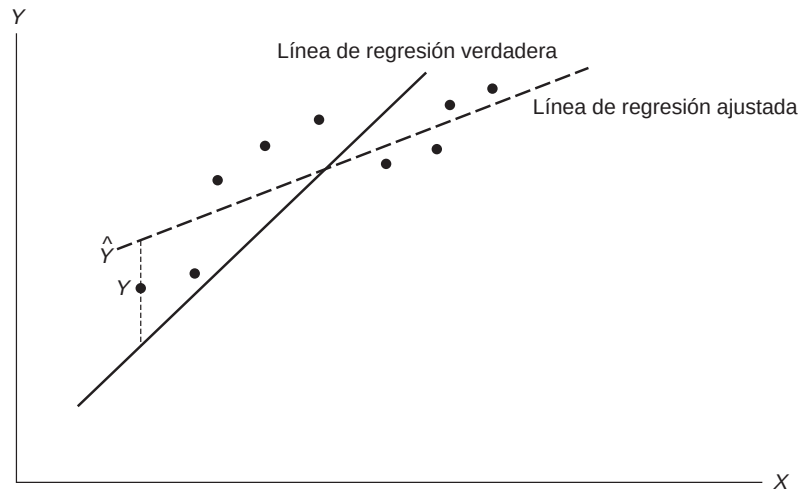


FIGURA 8-1 Correlación serial positiva y procedimiento de los mínimos cuadrados

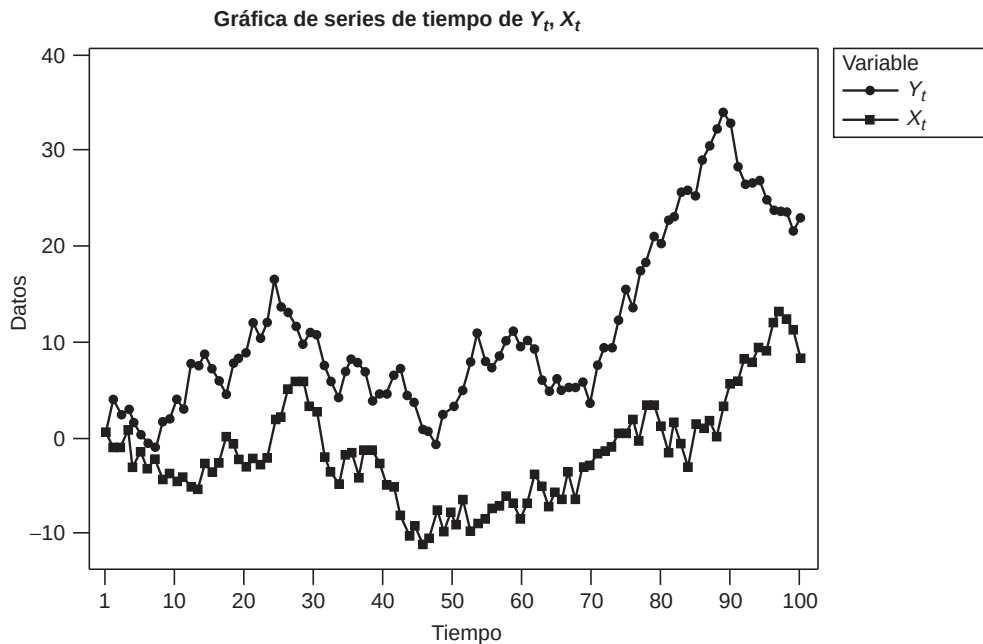


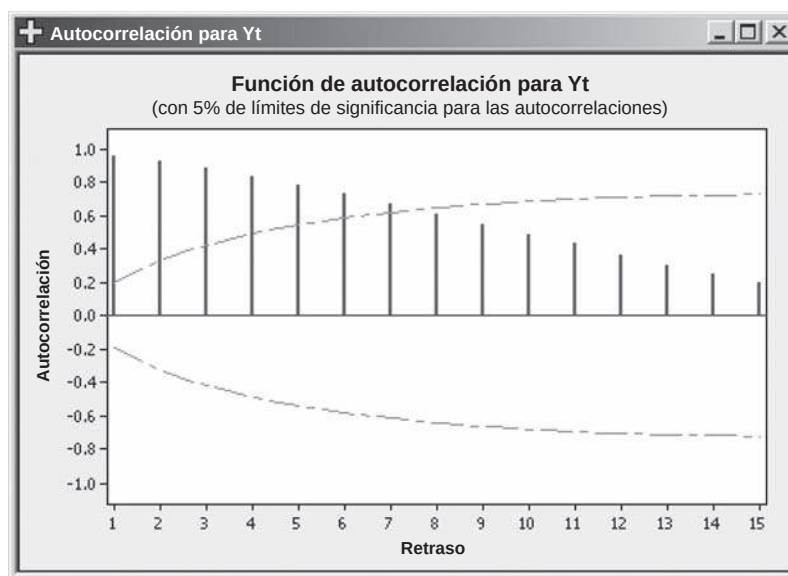
FIGURA 8-2 Gráficas de series de tiempo de dos series no relacionadas, Y_t (arriba) y X_t (abajo)

y un examen de los residuos por lo general sacará a la luz el problema.² Sin embargo, si se hace una aplicación poco crítica de los procedimientos estándar, la regresión falsa podría pasar inadvertida, dando como resultado una inadecuada interpretación de los resultados.

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 se formaron de tal manera que la primera (Y_t) no está relacionada con la

²Dos (o más) series de tiempo autocorrelacionadas pueden estar relacionadas, pero se debe tener especial cuidado para descubrir la relación. Más adelante en este capítulo, en la sección de series de tiempo cointegradas, se considerará brevemente un caso especial.

Función autocorrelación: Y_t

Retraso	ACF	T	LBQ	Retraso	ACF	T	LBQ
1	0.959355	9.59	94.83	9	0.543683	1.61	584.51
2	0.923743	5.48	183.64	10	0.486047	1.40	611.29
3	0.878880	4.12	264.86	11	0.427587	1.21	632.24
4	0.833189	3.38	338.62	12	0.360773	1.01	647.33
5	0.783946	2.87	404.61	13	0.299414	0.83	657.84
6	0.730737	2.48	462.55	14	0.245812	0.68	665.01
7	0.666258	2.13	511.24	15	0.199546	0.55	669.78
8	0.601754	1.84	551.38				

FIGURA 8-3 Autocorrelaciones para la serie Y_t

segunda (X_t).³ Al mismo tiempo, cada secuencia de observaciones está altamente autocorrelacionada. Las autocorrelaciones para la primera serie se presentan en la figura 8-3. Las autocorrelaciones para la segunda serie (no presentadas) son muy similares.

La figura 8-2 indica que las dos series de tiempo parecen moverse juntas. De hecho, sería posible relacionar la serie de arriba con la serie de abajo, usando un modelo de regresión lineal simple. En la figura 8-4 se presenta un diagrama de dispersión de los datos junto con la línea de mínimos cuadrados. El estadístico R^2 también se muestra en la figura. La regresión estimada es significativa (un análisis completo proporciona una $F = 83.08$ y un valor $p = .000$), y X_t explica alrededor del 46% de la variabilidad de Y_t . Sin embargo, la serie Y_t fue generada independientemente de la serie X_t . Es decir, las X no se emplearon para generar las Y . La regresión estimada en este caso es falsa.

Un examen de los residuos revelaría problemas con este análisis de regresión. Por ejemplo, las autocorrelaciones residuales son grandes para varios retrasos, lo que indica que el supuesto de errores independientes es incorrecto y que el modelo de regresión inicial debe modificarse. En este caso, si la modificación se hace correctamente, la relación falsa entre Y_t y X_t desaparecerá.

Si los modelos de regresión se usan con datos (series de tiempo) autocorrelacionados, es especialmente importante examinar los residuos. Si esto no se hace, es posible obtener conclusiones sin justificación. La falla no está en el procedimiento de mínimos cuadrados, sino en la

³La primera serie se elaboró seleccionando una muestra aleatoria de 100 valores de una distribución normal con media cero y una desviación estándar de 2 y luego haciendo sumas parciales. Por ejemplo, la primera observación en la serie fue el primer valor seleccionado en la muestra, la segunda observación en la serie fue la suma de los primeros dos valores, la tercera observación fue la suma de los primeros tres valores, y así sucesivamente. La segunda serie se elaboró de igual forma, iniciando con una muestra aleatoria diferente de 100 valores de la misma distribución normal.

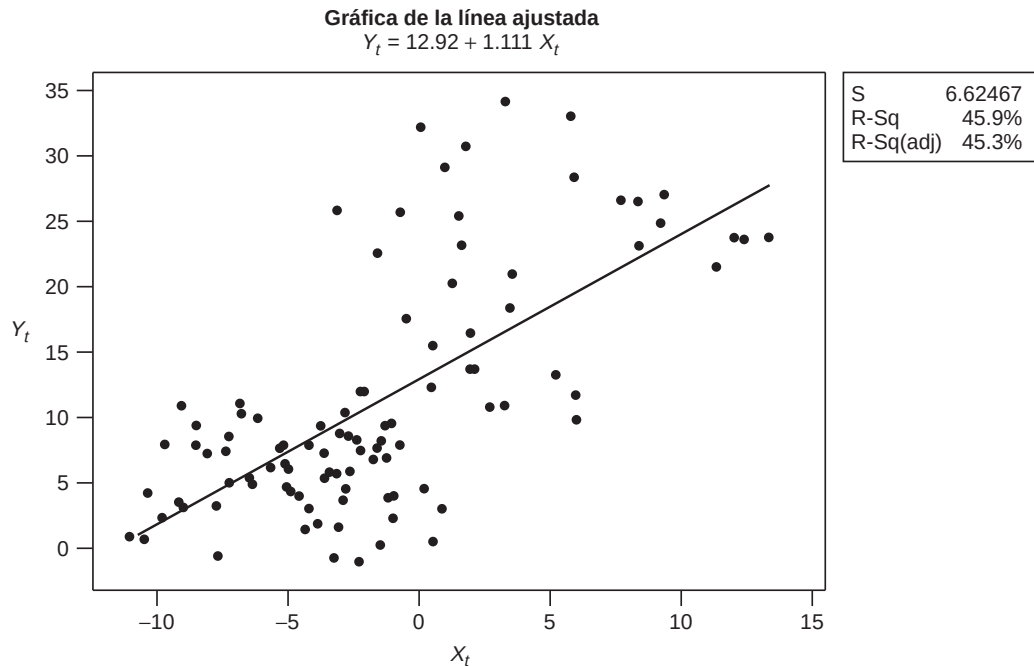


FIGURA 8-4 Resultados de la regresión lineal simple para dos series de tiempo Y_t y X_t altamente autocorrelacionadas

aplicación del modelo de regresión estándar a una situación que no corresponde a los supuestos de regresión habituales. Entre los problemas técnicos que surgen se incluyen los siguientes:

1. El error estándar de la estimación puede subestimar seriamente la variabilidad de los términos de error.
2. Las inferencias habituales basadas en los estadísticos t y F no son aplicables en sentido estricto.
3. Los errores estándar de los coeficientes de regresión subestiman la variabilidad de los coeficientes de regresión estimados. Esto puede dar como resultado una regresión falsa.

AUTOCORRELACIÓN Y LA PRUEBA DE DURBIN-WATSON

Como vimos en el capítulo 3, es posible examinar la autocorrelación creando los coeficientes de autocorrelación y comparándolos con sus errores estándar (véanse las ecuaciones 3.1 y 3.2). Además, el estadístico Q de Ljung-Box (ecuación 3.3) brinda una prueba completa para la autocorrelación significativa. En el contexto actual, aplicaríamos estos procedimientos a los residuos de un análisis de regresión de una serie de tiempo para verificar la aleatoriedad. La ecuación 6.17 da los coeficientes de autocorrelación residual.

El examen directo de las autocorrelaciones residuales es una buena práctica y debe utilizarse. Sin embargo, Minitab y otros programas de computadora tienen la opción de calcular directamente un estadístico que es útil para detectar la correlación serial de primer orden, o autocorrelación del retraso 1, conocida como el estadístico de *Durbin-Watson* (DW). Una

prueba para la correlación serial significativa de primer orden con base en este estadístico se conoce como la *prueba de Durbin-Watson*.⁴

Considere

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$$

Las hipótesis sometidas a prueba son

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

La hipótesis alternativa es $\rho > 0$, puesto que las series de tiempo de negocios y económicas generalmente tienden a presentar una autocorrelación positiva.

Si un modelo de regresión no explica adecuadamente la correlación, los residuos estarán autocorrelacionados. Como resultado, la prueba de Durbin-Watson se aplica usando los residuos del análisis de regresión.

El estadístico de 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

$$e_t = Y_t - \hat{Y}_t = \text{el residuo para el periodo } t$$

$$e_{t-1} = Y_{t-1} - \hat{Y}_{t-1} = \text{el residuo para el periodo } t - 1$$

Para la correlación serial positiva, los residuos sucesivos tienden a ser parecidos, y la suma de las diferencias al cuadrado en el numerador del estadístico de Durbin-Watson será relativamente pequeña. Los valores pequeños del estadístico de Durbin-Watson son coherentes con la correlación serial positiva.

El coeficiente de autocorrelación, ρ , se puede estimar por la autocorrelación residual de retraso 1, $r_1(e)$ (véase la ecuación 6.17), y con un poco de maniobra matemática, el estadístico de Durbin-Watson puede relacionarse con $r_1(e)$. Para muestras de moderadas a grandes,

$$DW = 2(1 - r_1(e)) \quad (8.4)$$

Puesto que $-1 < r_1(e) < 1$, la ecuación 8.4 indica que $0 < DW < 4$. Para $r_1(e)$ cerca de cero, el estadístico DW estará cerca de 2. La autocorrelación residual positiva de retraso 1 estará relacionada con valores de DW menores de 2, y la autocorrelación residual negativa del retraso 1 estará asociada con los valores de DW arriba de 2.

Una prueba útil, aunque no definitiva, de la correlación serial puede realizarse comparando el valor calculado del estadístico de Durbin-Watson con los límites inferior (d_L) y superior (d_U). Las reglas de decisión son las siguientes:

1. Cuando el estadístico de Durbin-Watson es mayor que el límite superior (d_U); el subíndice U proviene de la designación en inglés “upper bound”, que significa límite superior), el coeficiente de autocorrelación ρ , es igual a cero (no existe autocorrelación positiva).
2. Cuando el estadístico de Durbin-Watson es menor que el límite inferior (d_L); el subíndice L proviene de la designación en inglés “lower bound”, que significa límite inferior), el coeficiente de autocorrelación, ρ , es mayor que cero (existe autocorrelación positiva).
3. Cuando el estadístico de Durbin-Watson permanece dentro de los límites inferior y superior, la prueba no es concluyente (no sabemos si existe autocorrelación positiva).

⁴Véase Durbin y Watson (1951). Esta prueba no es aplicable directamente si la ecuación de regresión no contiene un término constante.

La prueba de Durbin-Watson permite determinar si la autocorrelación positiva en el retraso 1 está presente.

Si $DW > d_U$, se concluye que $H_0: \rho = 0$.

Si $DW < d_L$, se concluye que $H_1: \rho > 0$.

Si DW permanece dentro de 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 presentan en la tabla B-6. Para obtener 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 B-6, el tamaño de la muestra se encuentra en la columna de la izquierda, y el número de variables independientes se determina en la parte superior de cada columna. Si se usaran tres variables independientes, por ejemplo, buscaríamos en la columna $k = 3$.⁵

Como se indicó en la ecuación 8.4, es posible inferir el signo y la magnitud del coeficiente de autocorrelación residual de retraso 1 a partir del estadístico DW y viceversa. Por lo tanto, para el escenario en que la prueba de Durbin-Watson no es concluyente, la significancia de la correlación serial se puede investigar comparando $r_1(e)$ con $\pm 2/\sqrt{n}$.⁶ Si $r_1(e)$ cae en el intervalo $0 \pm 2/\sqrt{n}$, se concluye que la autocorrelación es pequeña y puede ignorarse.

Ejemplo 8.2

Suponga que un analista se comprometió a sacar adelante la planeación de Reynolds Metals Company, una compañía productora de aluminio, y desea establecer una base cuantitativa para la proyección de ventas futuras. Puesto que la compañía vende regionalmente, una medida del ingreso personal disponible en la región debe estar estrechamente relacionada con las ventas. La tabla 8-1 contiene las ventas y el ingreso para el periodo de 1986 a 2006. También contiene las columnas necesarias para calcular el estadístico DW (véase la sección de aplicaciones de Minitab al final del capítulo). Los residuos provienen de la línea de mínimos cuadrados ajustada a los datos, como se ve en la figura 8-5.

Antes de usar la línea de mínimos cuadrados para pronosticar, el analista ejecuta la prueba de Durbin-Watson para la correlación serial positiva. Los cálculos para las tres últimas columnas de 1987 son como sigue:

$$\begin{aligned} e_t - e_{t-1} &= -47.53 - (-76.36) = 28.83 \\ (e_t - e_{t-1})^2 &= 28.83^2 = 831.17 \\ e_t^2 &= (-47.53)^2 = 2,259.1 \end{aligned}$$

El estadístico de Durbin-Watson se calcula como

$$DW = \frac{\sum_{t=2}^{21} (e_t - e_{t-1})^2}{\sum_{t=1}^{21} e_t^2} = \frac{1,926,035.14}{2,210,641.78} = .87$$

Usando un nivel de significancia de .01 para una muestra de $n = 21$ y $k = 1$ variable independiente, se obtiene

$$\begin{aligned} d_L &= .97 \\ d_U &= 1.16 \end{aligned}$$

⁵También es posible probar la autocorrelación negativa. En este caso $H_1: \rho < 0$, y el estadístico de prueba DW , se compara con $4 - d_L$ y $4 - d_U$. 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 para DW entre $4 - d_U$ y $4 - d_L$.

⁶Si no existe autocorrelación, el error estándar de $r_1(e)$ es aproximadamente $1/\sqrt{n}$ (véase la explicación de autocorrelación en el capítulo 3 y la de autocorrelación residual en el capítulo 6).

TABLA 8-1 Datos de las ventas de Reynolds Metals y los cálculos de la prueba de Durbin-Watson para el ejemplo 8.2

Año	Ventas Y_t	Ingreso X_t	Residuos e_t	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	e_t^2
1986	295	273.4	-76.36	—	—	5,830.85
1987	400	291.3	-47.53	28.83	831.17	2,259.10
1988	390	306.9	-123.91	-76.38	5,833.90	15,353.69
1989	425	317.1	-132.32	-8.41	70.73	17,508.58
1990	547	336.1	-91.16	41.16	1,694.15	8,310.15
1991	555	349.4	-139.76	-48.60	2,361.96	19,532.86
1992	620	362.9	-132.20	7.56	57.15	17,476.84
1993	720	383.9	-121.56	10.64	113.21	14,776.83
1994	880	402.8	-41.98	79.58	6,332.98	1,762.32
1995	1,050	437.0	-17.51	24.47	598.78	306.60
1996	1,290	472.2	72.71	90.22	8,139.65	5,286.74
1997	1,528	510.4	148.16	75.45	5,692.70	21,951.39
1998	1,586	544.5	61.06	-87.10	7,586.41	3,728.32
1999	1,960	588.1	249.53	188.47	35,520.94	62,265.22
2000	2,118	630.4	227.54	-21.99	483.56	51,774.45
2001	2,116	685.9	-10.62	-238.16	56,720.19	112.78
2002	2,477	742.8	108.26	118.88	14,132.45	11,720.23
2003	3,199	801.3	581.33	473.07	223,795.22	337,944.57
2004	3,702	903.1	651.16	69.83	4,876.23	424,009.35
2005	3,316	983.6	-77.38	-728.54	530,770.53	5,987.66
2006	2,702	1,076.7	-1,087.54	-1,010.16	1,020,423.23	1,182,743.25
Totales					1,926,035.14	2,210,641.78

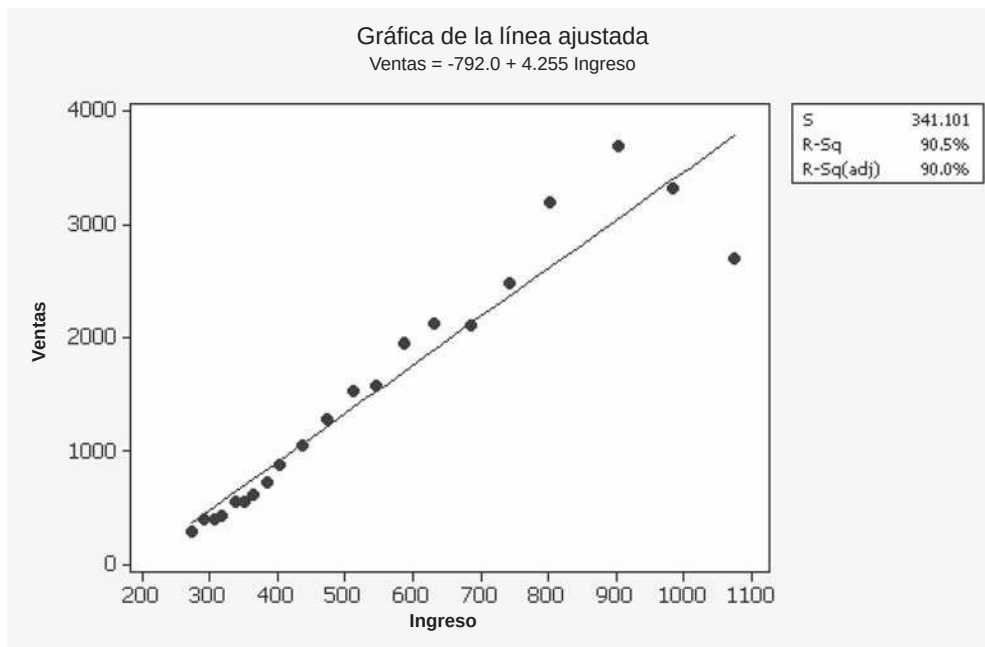


FIGURA 8-5 Gráfica de regresión de los datos de Reynolds Metals para el ejemplo 8.2

Como $DW = .87$ cae por debajo de $d_L = .97$ se rechaza la hipótesis nula $H_0: \rho = 0$ y se concluye que los errores están positivamente correlacionados ($\rho > 0$). El modelo de regresión debe modificarse antes de utilizarlo para pronosticar.

SOLUCIONES A LOS PROBLEMAS DE AUTOCORRELACIÓN

Después de que se ha detectado la autocorrelación en la regresión de una serie de tiempo, es necesario eliminarla, o modelarla, antes de evaluar la efectividad de la función de regresión. El método apropiado para tratar la autocorrelación depende en primer lugar de la razón que la causó. La autocorrelación puede surgir por un error tal como la omisión de una variable, o porque los términos independientes del error están correlacionados en un modelo especificado correctamente de otra forma.

La solución al problema de correlación inicia con una evaluación de la especificación del modelo. ¿Es correcta la forma funcional? ¿Se omitieron algunas variables importantes? ¿Existen efectos con algún patrón en el tiempo que pudieran haber introducido la autocorrelación en los errores?

Puesto que la causa principal de los errores autocorrelacionados en el modelo de regresión es la omisión de una o más variables clave, el mejor enfoque para resolver el problema es encontrarlas. En ocasiones este esfuerzo se conoce como una mejora de la especificación del modelo. La especificación del modelo no sólo implica la obtención de variables explicativas importantes, sino también la introducción correcta de estas variables en la función de regresión. Por desgracia, no siempre es posible mejorar la especificación del modelo porque una variable importante que falta no puede cuantificarse o, si es cuantificable, los datos no están disponibles. Por ejemplo, tal vez uno sospeche que la inversión en el negocio en periodos futuros está relacionada con la actitud de los inversionistas potenciales. Sin embargo, es difícil cuantificar la variable “actitud”. No obstante, siempre que es posible, el modelo debe especificarse de acuerdo con una sólida comprensión teórica.

Sólo después de que se ha revisado cuidadosamente la especificación de la ecuación, se debe considerar la posibilidad de un ajuste. Se estudiarán varias técnicas para eliminar la autocorrelación.

Un método para eliminar la autocorrelación consiste en agregar a la función de regresión una variable omitida que explique la asociación en la respuesta de un periodo al siguiente.

Otro método de eliminación de la autocorrelación implica la noción general de diferenciación. En este método, el modelo de regresión se especifica en términos de *cambios* y no de *niveles*. Por ejemplo, usando los datos de Reynolds Metals, el cambio en las ventas de un año al siguiente (este cambio es de 105 para el periodo de 1986 a 1987) está relacionado con el cambio correspondiente en el ingreso (este cambio es de 17.9 para el periodo de 1986 a 1987). En ocasiones, las variables originales se pueden expresar en términos de logaritmos, y los cambios en los logaritmos se usan en la regresión. Este procedimiento es equivalente a la regresión del *cambio porcentual* en la respuesta sobre los cambios porcentuales en la variable explicativa. Finalmente, en vez de usar las diferencias simples o las primeras diferencias en el modelo de regresión, es posible que se requieran *diferencias generalizadas* para eliminar la autocorrelación.

El método del modelo autorregresivo para eliminar la autocorrelación genera las variables explicativas usando la variable de respuesta, Y , retrasada uno o más periodos. En el caso del modelo autorregresivo de primer orden, la única variable explicativa es la variable Y retrasada un periodo. Usando otra vez los datos de Reynolds Metals, el valor usado para predecir las ventas de 1987 son las ventas de 1986 (295).

Los siguientes ejemplos ilustran estos métodos para eliminar la autocorrelación.

Error de especificación del modelo (omisión de una variable)

El ejemplo 8.3 ilustra cómo la incorporación de una variable faltante puede eliminar la correlación serial.

Ejemplo 8.3

Novak Corporation desea desarrollar un modelo de pronóstico para la proyección de las ventas futuras. Como la corporación tiene puntos de venta por toda la región, el ingreso personal disponible en una región amplia se elige como una posible variable explicativa. La tabla 8-2 indica las ventas de Novak de 1990 a 2006. La tabla también contiene datos sobre el ingreso personal disponible y el desempleo en la región.

De la tabla 8-3, el estadístico de Durbin-Watson es .72, y usando un nivel de significancia de .01 con $n = 17$ y $k = 1$, la tabla B-6 nos da

$$\begin{aligned}d_L &= .87 \\d_U &= 1.10\end{aligned}$$

Como $DW = .72 < d_L = .87$, existe una correlación serial positiva. Una variable clave que explique la asociación restante en las ventas de un año al siguiente podría estar faltando en el modelo. Este resultado puede ser cierto aun cuando el resultado de Minitab diga que el ingreso disponible para gastar explica el 99.5% de la variabilidad en las ventas.

La tasa de desempleo puede ser una variable explicativa importante de las ventas que está faltando. La tabla 8-4 presenta los resultados del análisis de regresión cuando se agrega la tasa de desempleo al modelo. Ahora el modelo ajustado explica el 99.9% de la variabilidad en las ventas. Si bien la intercepción no es significativa, el modelo reajustado sin la intercepción deja el estadístico de Durbin-Watson de 1.98 prácticamente sin cambio.

Con un nivel de significancia de $n = 17$, y $k = 2$, la tabla B-6 nos da

$$\begin{aligned}d_L &= .77 \\d_U &= 1.25\end{aligned}$$

Puesto que $DW = 1.98 > d_U = 1.25$, no hay evidencia de correlación serial de primer orden. Más aún, la autocorrelación residual del retraso 1 es $r_1(e) = .005$, lo cual está bien dentro de $\pm 2/\sqrt{n} = \pm 2/\sqrt{17} = \pm .49$ de cero. Observe que $DW = 1.98 \cong 2(1 - r_1(e)) =$, de manera que los resultados de la prueba de Durbin-Watson son congruentes con la ausencia de autocorrelación residual en un retraso 1, como se esperaba.

TABLA 8-2 Datos de las ventas de Novak para los ejemplos 8.3 y 8.6

<i>Fila</i>	<i>Año</i>	<i>Ventas (millones de \$) (Y)</i>	<i>Ingreso (millones de \$)</i>	<i>Tasa de desempleo</i>	<i>Y-retrasada</i>
1	1990	8.0	336.1	5.5	—
2	1991	8.2	349.4	5.5	8.0
3	1992	8.5	362.9	6.7	8.2
4	1993	9.2	383.9	5.5	8.5
5	1994	10.2	402.8	5.7	9.2
6	1995	11.4	437.0	5.2	10.2
7	1996	12.8	472.2	4.5	11.4
8	1997	13.6	510.4	3.8	12.8
9	1998	14.6	544.5	3.8	13.6
10	1999	16.4	588.1	3.6	14.6
11	2000	17.8	630.4	3.5	16.4
12	2001	18.6	685.9	4.9	17.8
13	2002	20.0	742.8	5.9	18.6
14	2003	21.9	801.3	5.6	20.0
15	2004	24.9	903.1	4.9	21.9
16	2005	27.3	983.6	5.6	24.9
17	2006	29.1	1076.7	8.5	27.3

TABLA 8-3 Resultados de Minitab para las ventas de Novak e ingreso personal disponible para el ejemplo 8.3**Regression Analysis: Sales (Y) versus Income**

The regression equation is

$$\text{Sales (Y)} = -1.50 + 0.0292 \text{ Income}$$

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 = 0.476669 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

TABLA 8-4 Resultados de Minitab para las ventas de Novak, ingreso disponible y tasa de desempleo para el ejemplo 8.3**Regression Analysis: Sales (Y) versus Income, Rate**

The regression equation is

$$\text{Sales (Y)} = -0.014 + 0.0297 \text{ Income} - 0.350 \text{ Rate}$$

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 = 0.219930 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 función $\hat{Y} = -.014 + .03X_1 - .35X_2$ sirve para predecir las ventas de Novak con el conocimiento de que los errores son independientes.⁷ Se usan las estimaciones de los expertos acerca del ingreso personal disponible (\$1,185 millones) y la tasa de desempleo (7.8%) para la región con la finalidad de generar un pronóstico de las ventas de Novak para 2007. El pronóstico es

$$\hat{Y} = -.014 + .03(1,185) - .35(7.8) = 32.7$$

o \$32.7 millones.

⁷Las gráficas residuales comunes indican que no hay razón para dudar de cualquiera de los supuestos del modelo de regresión.

Regresión con diferencias

Para datos altamente correlacionados, modelar los cambios en vez de los niveles a menudo permite eliminar la correlación serial. Es decir, en vez de formular la ecuación de regresión en términos de Y y $X_1, X_2, \dots, X_k$, la ecuación de la regresión se escribe en términos de las diferencias $Y'_t = Y_t - Y_{t-1}$ y $X'_{t1} = X_{t1} - X_{t-1,1}$, $X'_{t2} = X_{t2} - X_{t-1,2}$, etcétera. Se deben considerar las diferencias cuando el estadístico de Durbin-Watson asociado con la regresión que implica las variables originales está cerca de cero.⁸

Un fundamento para la diferenciación surge del siguiente argumento. Suponga que las ecuaciones 8.1 y 8.2 se cumplen, de manera que

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

con

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t$$

donde

ρ = la correlación entre errores consecutivos

v_t = el 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 ambos lados de esta ecuación por ρ y luego restando de la ecuación 8.1 se obtiene

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \quad (\text{Ecuación 8.1})$$

$$\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{t-1} + \rho \varepsilon_{t-1} \quad (\text{multiplicando } Y_{t-1} \text{ por } \rho)$$

$$Y_t - \rho Y_{t-1} = \beta_0 - \rho \beta_0 + (\beta_1 X_t - \rho \beta_1 X_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}) \quad (\text{restando})$$

o

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t \quad (8.5)$$

donde la “prima” indica las *diferencias generalizadas*

$$\begin{aligned} Y'_t &= Y_t - \rho Y_{t-1} \\ X'_t &= X_t - \rho X_{t-1} \end{aligned} \quad (8.6)$$

El modelo de la ecuación 8.5 tiene errores, v_t , que están distribuidos independientemente con la media igual a cero y una varianza constante. Por lo tanto, los métodos de regresión comunes se pueden aplicar a este modelo.

⁸Un patrón de autocorrelación para la variable Y o las variables X , tal como el presentado en la figura 8-3, también indica que una función de regresión con diferencias puede eliminar (o reducir enormemente) los problemas causados por la autocorrelación serial.

Si la correlación entre errores consecutivos es fuerte (ρ está cerca de 1), las diferencias generalizadas son, en esencia, diferencias simples o primeras diferencias:

$$\begin{aligned} Y'_t &= Y_t - Y_{t-1} \\ X'_t &= X_t - X_{t-1} \end{aligned} \quad (8.7)$$

y el término de la intercepción en el modelo (ecuación 8.5) es cercano a cero (desaparece).

Usando modelos de regresión elaborados con diferencias generalizadas a menudo permite eliminar la correlación serial. Si la correlación serial es especialmente fuerte, se pueden usar las diferencias simples. El ejemplo 8.4 indica lo que pasaría si se ignora la autocorrelación fuerte.

Ejemplo 8.4

Hace algunos años, Fred Gardner tenía el compromiso de pronosticar las ventas en miles de dólares de Sears Roebuck para la región oeste. Él había seleccionado el ingreso personal disponible de la región como su variable independiente. Fred quería relacionar las ventas con el ingreso disponible usando un modelo logarítmico de regresión lineal, puesto que le permitiría estimar también la elasticidad del ingreso de las ventas. La elasticidad mide el cambio porcentual en las ventas para un 1% de cambio en el ingreso.

El modelo logarítmico de regresión lineal supone que el ingreso está relacionado con las ventas mediante la ecuación

$$\text{Ventas} = \gamma(\text{Ingreso})^{\beta_1}$$

Al tomar logaritmos naturales (Ln) de ambos lados de la ecuación se obtiene

$$\text{Ln}(\text{Ventas}) = \text{Ln } \gamma + \beta_1 \text{Ln}(\text{Ingreso})$$

Agregando un término de error para explicar la influencia de otras variables diferentes al ingreso por las ventas, la expresión anterior se convierte en un *modelo logarítmico de regresión lineal* de la forma

$$\text{Ln } Y_t = \beta_0 + \beta_1 \text{Ln } X_t + \varepsilon_t \quad (8.8)$$

donde

$\text{Ln } Y_t = \text{Ln}(\text{Sales}) =$ logaritmo natural de las ventas

$\text{Ln } X_t = \text{Ln}(\text{Income}) =$ logaritmo natural del ingreso

$\varepsilon_t =$ término de error

$\beta_0 = \text{Ln } \gamma =$ coeficiente de intercepción

$\beta_1 =$ coeficiente de la pendiente = la elasticidad del ingreso de las ventas

La tabla 8-5 indica las ventas de Sears, el ingreso disponible, sus logaritmos y las diferencias en los logaritmos de las ventas e ingreso disponible para el periodo de 1976 a 1996.

Una parte del resultado de Minitab para la regresión de Ln(Ventas) sobre Ln(Ingreso) se presenta en la tabla 8-6. Fred notó que el 99.2% de la variabilidad en el logaritmo de las ventas de Sears para la región oeste puede explicarse por su relación con el logaritmo del ingreso disponible para la misma región. La regresión es altamente significativa. También se estima que la elasticidad del ingreso es $b_1 = 1.117$, con un error estándar de $s_{b_1} = .023$. Sin embargo, el estadístico de Durbin-Watson de .50 es pequeño y menor que $d_L = .97$, el más bajo del nivel del valor crítico .01 para $n = 21$ y $k = 1$. Fred concluyó que la correlación entre errores sucesivos es positiva y grande (cercana a 1).

En virtud de la correlación serial grande, Fred decidió modelar los cambios o las diferencias en los logaritmos de las ventas y los ingresos, respectivamente. Él sabía que el coeficiente de la pendiente en el modelo de las diferencias era el mismo del modelo original que implica logaritmos. Por lo tanto, podía estimar la elasticidad del ingreso directamente. El coeficiente de intercepción en el modelo de regresión de las diferencias probablemente sería pequeño y se omitió. Los resultados de Minitab para los cambios se presentan en la tabla 8-7.

La tabla 8-7 indica que la regresión es significativa. La elasticidad del ingreso se estima como $b_1 = 1.010$, con un error estándar de $s_{b_1} = .093$. La elasticidad estimada, b_1 , no cambió demasiado en relación con la primera regresión (un 1% de incremento en el ingreso disponible trae consigo un 1%

TABLA 8-5 Ventas de Sears e ingreso disponible en Estados Unidos, 1976 a 1996, junto con los datos transformados para el ejemplo 8.4

Año	Ventas (en miles de \$)	Ingreso (en miles de \$)	Ln Y_t	Ln X_t	Diferencias	
	Y_t	X_t			Y'_t	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

TABLA 8-6 Resultado de Minitab para la regresión de los logaritmos de las ventas de Sears sobre los logaritmos del ingreso disponible para el ejemplo 8.4

Regression Analysis: Ln(Sales) versus Ln(Income)

The regression equation is

$$\text{Ln(Sales)} = 1.82 + 1.12 \text{ Ln(Income)}$$

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

$$S = 0.0436806 \quad R\text{-Sq} = 99.2\% \quad R\text{-Sq(adj)} = 99.2\%$$

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 Resultado de Minitab para la regresión de los cambios en los logaritmos de las ventas de Sears sobre los cambios en 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 = 0.0297487

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

Función de autocorrelación: Residuos

Retraso	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

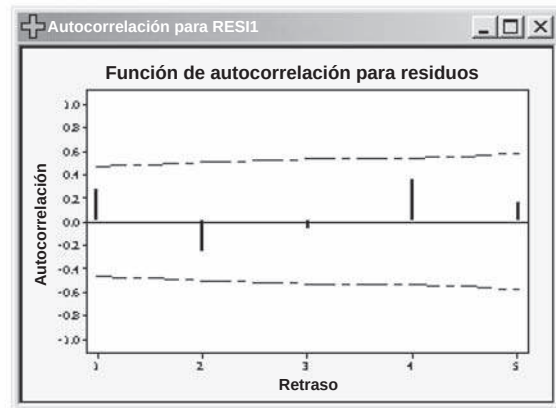


FIGURA 8-6 Autocorrelaciones residuales para el análisis de regresión en la tabla 8-7; ejemplo 8.4

de incremento aproximado en las ventas anuales en ambos casos), pero su error estándar actual ($s_{b_1} = .093$) es alrededor de cuatro veces más grande que el error estándar anterior ($s_{b_1} = .023$). El error estándar anterior probablemente está subestimando el error estándar verdadero a causa de la correlación serial.

Al verificar el estadístico de Durbin-Watson para $n = 20$, $k = 1$, y un nivel de significancia de .05, Fred encontró que $d_L = 1.20 < DW = 1.28 < d_U = 1.41$, de manera que la prueba de la correlación serial positiva no es concluyente. Sin embargo, una revisión de las autocorrelaciones residuales, presentada en la figura 8-6, indica que están comprendidas dentro de sus dos límites de error estándar (las líneas punteadas en la figura) para unos cuantos de los primeros retrasos. Fred concluyó que se había eliminado la correlación serial y utilizó la ecuación ajustada para pronosticar.

Con el fin de aplicar el modelo final para el pronóstico, Fred escribió

$$\hat{Y}'_t = b_1 X'_t \quad \text{donde} \quad \begin{aligned} \hat{Y}'_t &= \text{Ln } \hat{Y}_t - \text{Ln } \hat{Y}_{t-1} \\ X'_t &= \text{Ln } X_t - \text{Ln } X_{t-1} \end{aligned}$$

Al sustituir $\hat{Y}'_t$ y X'_t y al reacomodar los términos,

$$\text{Ln } \hat{Y}_t = \text{Ln } \hat{Y}_{t-1} + b_1(\text{Ln } X_t - \text{Ln } X_{t-1}) \quad (8.9)$$

El pronóstico de las ventas de Sears para 1997 se obtuvo estableciendo $b_1 = 1.01$ y $t = 22$:

$$\text{Ln } \hat{Y}_{22} = \text{Ln } \hat{Y}_{21} + 1.01(\text{Ln } X_{22} - \text{Ln } X_{21})$$

Se conocían las ventas de 1996, de manera que $\hat{Y}_{21} = Y_{21} = 13,640$. Se conocía el ingreso disponible de 1996, así que $X_{21} = 1,076.7$. Para continuar, Fred necesitaba el ingreso disponible de 1997. Un economista familiarizado con la región del oeste envió a Fred una estimación de \$1,185 millones para el ingreso disponible en 1997. Fred usó esta estimación del experto y estableció $X_{22} = 1,185$. La ecuación del pronóstico se convirtió en

$$\begin{aligned} \text{Ln } \hat{Y}_{22} &= \text{Ln}(13,640) + 1.01(\text{Ln}(1,185) - \text{Ln}(1,076.7)) \\ &= 9.5208 + 1.01(7.0775 - 6.9817) = 9.6176 \end{aligned}$$

o, tomando los antilogaritmos,

$$\hat{Y}_{22} = e^{9.6176} = 15,027$$

El pronóstico de Fred para las ventas de Sears para 1997 en la región oeste fue de \$15,027 en miles de dólares. Fred podía usar la ecuación 8.9 y el procedimiento descrito anteriormente para generar los pronósticos de los años 1998, 1999 y en adelante, pero para hacerlo, necesitaba estimar los ingresos personales disponibles de estos años.

Errores autocorrelacionados y diferencias generalizadas

El objetivo es describir adecuadamente la naturaleza de la relación entre las variables Y y X cuando la correlación serial está presente.

Considere otra vez el modelo de regresión con errores correlacionados serialmente (véanse las ecuaciones 8.1 y 8.2):

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_t + \varepsilon_t \\ \varepsilon_t &= \rho \varepsilon_{t-1} + v_t \end{aligned}$$

Se dice que los errores, ε_t , siguen un modelo autorregresivo de primer orden o modelo AR(1).⁹

Recuerde que, después de alguna manipulación algebraica, el anterior sistema de ecuaciones se puede escribir como un modelo de regresión lineal simple que implica las diferencias generalizadas $Y'_t = Y_t - \rho Y_{t-1}$ y $X'_t = X_t - \rho X_{t-1}$ (véase la ecuación 8.5):

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t$$

El modelo de regresión que implica las diferencias generalizadas está diseñado específicamente para eliminar la correlación serial en los errores. Se supone que los errores, v_t , son independientes y están distribuidos de manera normal con media cero y varianza constante, σ_v^2 . Por consiguiente, parece razonable ajustar este modelo directamente, usando el principio de mínimos cuadrados. Sin embargo, se desconoce el coeficiente de autocorrelación, ρ , así que no es posible determinar Y'_t y X'_t . Por lo tanto, el modelo no puede ajustarse usando los mínimos cuadrados ordinarios.

Existen dos métodos generales para estimar β_0 y β_1 , los parámetros de interés primordial. Un método consiste en usar el modelo corregido de la correlación serial (ecuación 8.5) y esti-

⁹Los modelos autorregresivos de primer orden se presentan formalmente en la siguiente sección de este capítulo y se estudiarán de nuevo en el capítulo 9.

Equation: LSVAR Workfile: SEARS DATA HW				
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: SALES				
Method: Least Squares				
Date: 07/08/02 Time: 12:41				
Sample: 1976 1996				
Included observations: 21				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-524.3323	188.4470	-2.782386	0.0119
INCOME	14.04963	0.318500	44.11184	0.0000
R-squared	0.990330	Mean dependent var	7102.476	
Adjusted R-squared	0.989821	S.D. dependent var	3404.744	
S.E. of regression	343.5060	Akaike info criterion	14.60668	
Sum squared resid	2241931.	Schwarz criterion	14.70616	
Log likelihood	-151.3701	F-statistic	1945.854	
Durbin-Watson stat	0.630087	Prob(F-statistic)	0.000000	

FIGURA 8-7 Resultados de E-Views para la regresión de ventas de Sears con base en el ingreso disponible para el ejemplo 8.5

mar los parámetros β_0 , β_1 , y ρ directamente, usando una técnica numérica llamada *mínimos cuadrados no lineales*. Este método emplea una rutina de búsqueda para encontrar los valores de los parámetros que minimicen el error en la suma de los cuadrados, $\sum v_t^2$. El otro método consiste en estimar ρ y usar la estimación $\hat{\rho}$, para obtener las diferencias generalizadas, y luego ajustar el modelo incluye estas diferencias generalizadas usando el método de los mínimos cuadrados ordinarios.¹⁰

El siguiente ejemplo ilustra el método de los mínimos cuadrados no lineales con el uso de E-Views, un paquete de software de uso común para la elaboración de modelos econométricos.

Ejemplo 8.5

Los datos de Sears se encuentran en la tabla 8-5. Los mínimos cuadrados ordinarios se usan para ajustar un modelo de regresión lineal simple que relaciona las ventas con el ingreso personal disponible. El resultado de E-Views se presenta en la figura 8-7. De acuerdo con esta última, la ecuación de regresión ajustada es

$$\hat{Y}_t = -524.33 + 14.05X_t$$

donde

$$\begin{aligned} b_1 &= 14.05 \\ s_{b_1} &= .319 \\ t = b_1/s_{b_1} &= 44.11 \\ r^2 &= .99 \\ DW &= .63 \end{aligned}$$

¹⁰Las técnicas para estimar ρ y explicar la correlación serial se analizan en Pyndick y Rubinfeld (1998).

Equation: LSARVAR Workfile: SEARS DATA HW

View Procs Objects Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: SALES
 Method: Least Squares
 Date: 07/08/02 Time: 13:38
 Sample(adjusted): 1977 1996
 Included observations: 20 after adjusting endpoints
 Convergence achieved after 151 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	54483.00	3344949.	0.016288	0.9872
INCOME	9.262387	7.240644	1.279221	0.2180
AR(1)	0.997245	0.170084	5.863238	0.0000

R-squared	0.995227	Mean dependent var	7292.250
Adjusted R-squared	0.994666	S.D. dependent var	3377.322
S.E. of regression	246.6659	Akaike info criterion	13.99143
Sum squared resid	1034349.	Schwarz criterion	14.14079
Log likelihood	-136.9143	F-statistic	1772.444
Durbin-Watson stat	1.121974	Prob(F-statistic)	0.000000

FIGURA 8-6 Resultados de E-Views para la regresión de diferencias generalizadas de las ventas de Sears sobre las diferencias generalizadas del ingreso disponible para el ejemplo 8.5

Puesto que el estadístico DW es cercano a cero, se trata de una autocorrelación positiva.¹¹ En este punto, tomamos en cuenta los errores correlacionados seriales (AR(1)) y ajustamos el modelo que implica 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 $X'_t = X_t - \rho X_{t-1}$. E-Views se utiliza para estimar los parámetros en este modelo directamente. El resultado de E-Views se presenta 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 del estadístico de Durbin-Watson, 1.12, está en la región “no concluyente” en el nivel .01. Note que las estimaciones del coeficiente de la pendiente en las dos regresiones son similares (14.05

¹¹Para $n = 21$, $k = 1$, y $\alpha = .01$, $d_L = .97$. Como $DW = .63 < d_L = .97$, rechazamos $H_0: \rho = 0$ a favor de $H_1: \rho > 0$.

y 9.26). Sin embargo, el error estándar asociado con b_1 en la segunda regresión es mucho más grande que el error estándar correspondiente en la primera regresión (7.241 contra .319). Por lo tanto, el estadístico t de la prueba de significancia del coeficiente de la pendiente en la segunda regresión es mucho menor que el de la primera regresión (1.28 contra 44.11). De hecho, el valor p asociado con el estadístico t en la segunda regresión es de .218, lo cual indica que el coeficiente de la pendiente no es significativamente diferente de cero. La fuerte correlación serial tiene poco efecto en la estimación del coeficiente de la pendiente en la relación entre Y y X . Sin embargo, la correlación serial (positiva) fuerte da como resultado una subestimación severa del error estándar del coeficiente estimado de la pendiente. De hecho, una razón para el ajuste de la correlación serial es evitar caer en errores de inferencia a causa de los valores de t que son muy grandes.

Finalmente, $\hat{\rho} = .997$ está muy cerca de 1. Esto sugiere que la relación entre Y y X podría representarse mediante un modelo de regresión lineal simple con las diferencias $Y'_t = Y_t - Y_{t-1}$ y $X'_t = X_t - X_{t-1}$. Este tema se explora en el problema 17.

Modelos autorregresivos

La autocorrelación implica que los valores de la variable dependiente en un periodo están linealmente relacionados con los valores de la variable dependiente en otro periodo. Por lo tanto, una manera de resolver el problema de la correlación serial es modelar directamente la asociación en diferentes periodos. Esto puede hacerse en un esquema de regresión, usando la variable dependiente retrasada uno o más periodos como la variable explicativa o independiente. Los modelos de regresión formulados de esta manera se llaman *modelos autorregresivos*. El modelo autorregresivo de primer orden se escribe como

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \quad (8.10)$$

donde se supone que los errores, ε_t , tienen las propiedades del modelo de regresión común. Una vez que este modelo se ha ajustado a los datos mediante los mínimos cuadrados, la ecuación del pronóstico se convierte en

$$\hat{Y}_t = b_0 + b_1 Y_{t-1} \quad (8.11)$$

Un *modelo autorregresivo* expresa un pronóstico como una función de valores previos de la serie de tiempo.

Los modelos autorregresivos son un subconjunto de los modelos de promedio móvil integrado autorregresivo (ARIMA), los cuales estudiaremos con más detalle en el capítulo 9.

Ejemplo 8.6

Los datos de las ventas de Novak Corporation del ejemplo 8.3 y contenidos en la tabla 8-2 se utilizarán para ilustrar el desarrollo de un modelo autorregresivo. Advierta que, después de que las ventas fueron retrasadas como se indica en tabla 8-2, se pierde un año de datos, puesto que se desconocen las ventas de Novak para el año 1989. El tamaño muestral es $n = 16$ en vez de $n = 17$. El modelo autorregresivo de primer orden se desarrolla con las ventas de Novak retrasadas un año (atraso de Y) como la variable explicativa.

El resultado de Minitab que se obtiene de ajustar el modelo autorregresivo para los datos de las ventas de Novak se presenta en la tabla 8-8. También se muestran en la tabla un pronóstico de ventas para 2007 y un intervalo de predicción del 95%.

El pronóstico de las ventas de Novak para 2007 es de \$31.722 millones. Esta cifra se calcula, considerando el error de redondeo, como sigue:

$$\begin{aligned} \hat{Y} &= b_0 + b_1 Y_{t-1} = -.109 + 1.094 Y_{t-1} \\ \hat{Y}_{18} &= -.109 + 1.094 Y_{18-1} = -.109 + 1.094 Y_{17} \\ \hat{Y}_{18} &= -.109 + 1.094(29.1) = 31.726 \end{aligned}$$

TABLA 8-8 Resultado de Minitab para las ventas de Novak: modelo autorregresivo ajustado para el ejemplo 8.6**Regression Analysis: Sales (Y) versus Y-Lagged**

The regression equation is

$$\text{Sales}(Y) = -0.109 + 1.09 Y - \text{Lagged}$$

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 = 0.487455 R-Sq = 99.5% R-Sq(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 usarse en este ejemplo. Cuando se incluye en la regresión una variable dependiente retrasada como variable explicativa, el estadístico de Durbin-Watson tiene un sesgo hacia el 2. En su lugar, la prueba de correlación serial se puede basar en el estadístico h de Durbin-Watson.¹²

El coeficiente de intercepción en esta regresión es pequeño y no es significativamente diferente de cero. El hecho de correr otra vez la regresión sin el término de intercepción deja a la estimación del coeficiente de la pendiente y a los pronósticos subsiguientes sin cambios en lo esencial. El hecho de que el coeficiente estimado de la pendiente sea alrededor de 1 sugiere que el pronóstico de ventas del año siguiente es casi igual a las ventas del presente año.

Resumen

Cuando se aplica el análisis de regresión a las series de tiempo, con frecuencia los residuos están autocorrelacionados. Puesto que el análisis de regresión supone que los errores son independientes, puede haber problemas. La R^2 de una regresión con datos que tienen una autocorrelación tal vez parezca artificialmente alta. Además, los errores estándar de los coeficientes de la regresión podrían estar seriamente subestimados, en tanto que los estadísticos t correspondientes tal vez estén inflados.

Una causa de los residuos autocorrelacionados es la omisión de una o más variables explicativas. Esta omisión generalmente significa que una parte importante de la variación de la variable dependiente no se ha explicado adecuadamente. Una solución para este problema es buscar la(s) variable(s) faltante(s) para incluirla(s) en el modelo. Otras soluciones a los problemas causados por las autocorrelaciones son considerar los modelos de regresión con datos diferenciados o los modelos autorregresivos.

DATOS DE SERIES DE TIEMPO Y EL PROBLEMA DE HETEROSCEDASTICIDAD

La variabilidad en algunas series de tiempo tiende a incrementarse con el nivel de las series. La serie de tiempo de las ventas de Cavanaugh Company que se presentó en la figura 5-1 (parte inferior) y que se analizó en el ejemplo 5.2 es un caso de este tipo. La variabilidad puede incre-

¹²Pindyck y Rubinfeld (1998) hacen una descripción de la prueba h de la correlación serial.

mentarse si una variable está creciendo a un ritmo constante y no en una cantidad constante en el tiempo. La variabilidad inconstante se llama *heteroscedasticidad*.

En un esquema de regresión, la heteroscedasticidad ocurre si la varianza del término de error, ε , no es constante. Si la variabilidad de periodos recientes es mayor de lo que fue en periodos pasados, entonces el error estándar de la estimación, $s_{y \cdot x}$, subestima la desviación estándar actual del término de error. Si la desviación estándar de la estimación se usa luego para establecer los límites de pronóstico para observaciones futuras, estos límites podrían ser demasiado estrechos para el nivel de confianza establecido.

Algunas veces el problema de la heteroscedasticidad se resuelve con una simple transformación de los datos. Por ejemplo, en el caso de dos variables, el modelo logarítmico lineal descrito en la ecuación 8.8 podría usarse para reducir la heteroscedasticidad. Por otra parte, si las variables están expresadas como cantidades en dólares, la conversión de dólares actuales a dólares constantes (véase el análisis de la deflación del precio en el capítulo 5) permite superar el problema del incremento de error en la variabilidad.

Ejemplo 8.7

Considere de nuevo los datos de las ventas de Reynolds Metals presentados en el ejemplo 8.2 y en la tabla 8.1. El resultado de una regresión lineal simple de las ventas sobre el ingreso personal disponible se indica en la figura 8-5. Una gráfica de la secuencia de tiempo de los residuos de esta regresión se encuentra en la figura 8-9.

Además de la autocorrelación positiva en los residuos (sucesión de residuos negativos seguida de una sucesión de residuos positivos; véase la explicación del ejemplo 8.2), es claro en la figura 8-9 que el tamaño de los residuos aumenta con el tiempo. Un enfoque para este problema es intentar un modelo logarítmico lineal (ecuación 8.8) para los datos de Reynolds Metals.

Los resultados del ajuste de un modelo logarítmico lineal se presentan en la figura 8-10. Al comparar la figura 8-10 con la figura 8-5, se observa que los residuos (desviaciones de la línea ajustada) de la regresión lineal logarítmica tienen un tamaño más uniforme a lo largo del periodo bajo estudio,

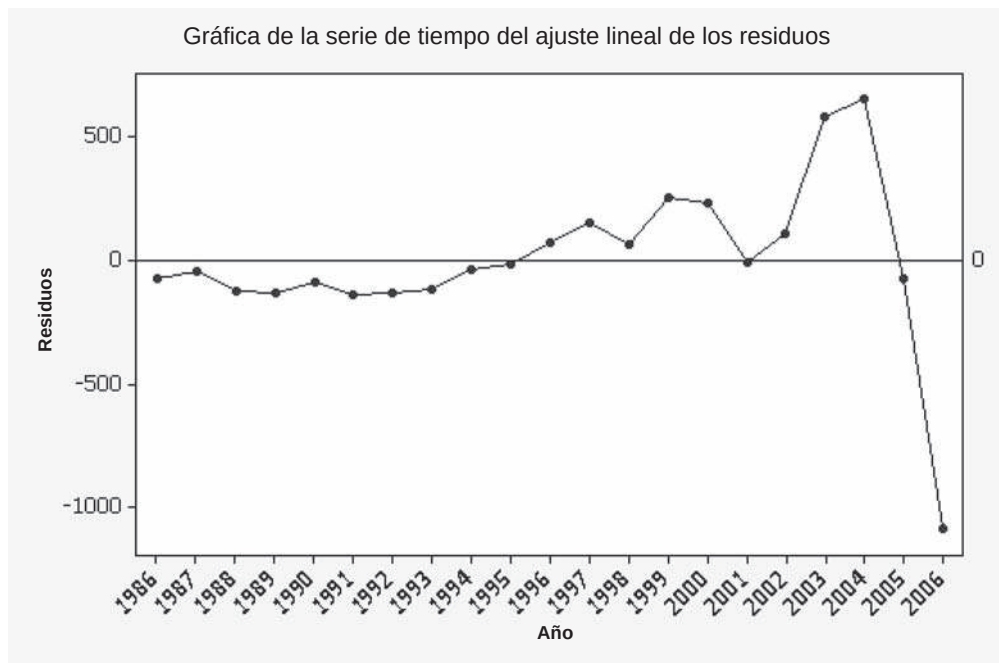


FIGURA 8-9 Gráfica de la secuencia de tiempo de los residuos de la regresión de los datos de ventas de Reynolds Metals sobre el ingreso disponible; ejemplo 8.7

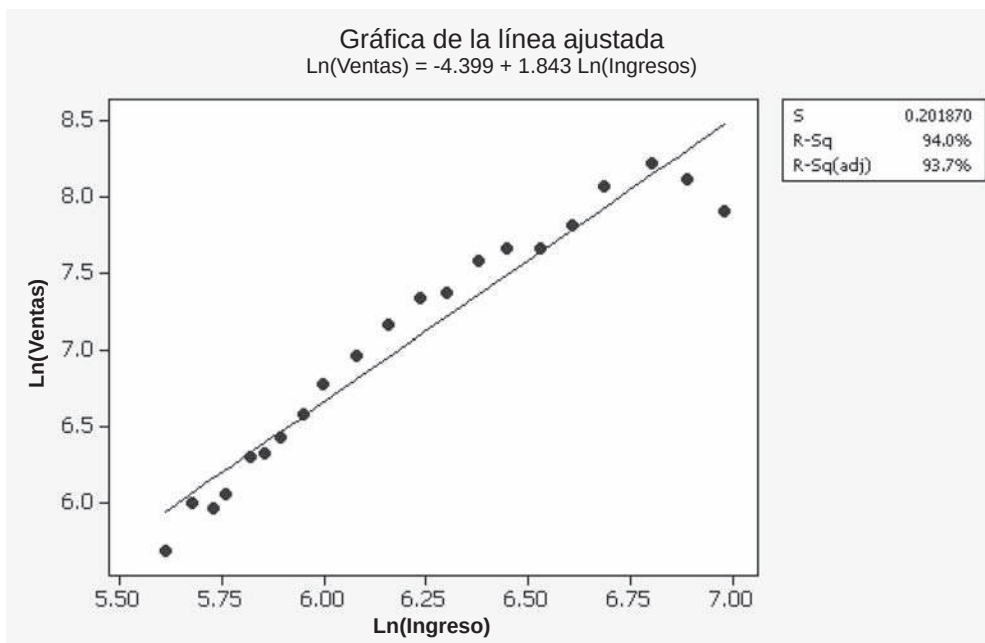


FIGURA 8-10 Gráfica de regresión ajustada del modelo logarítmico lineal para los datos de Reynolds Metals; ejemplo 8.7



FIGURA 8-11 Gráfica de la secuencia de tiempo de los residuos del modelo logarítmico lineal ajustado con un término cuadrático para los datos de Reynolds Metals; ejemplo 8.7

pero la línea recta ajustada no refleja la curvatura de los datos. Se agregó una variable explicativa adicional, $X_2 = X_1^2 = (\text{Ln Ingreso})^2$, y el modelo¹³

$$\text{Ln Ventas} = \beta_0 + \beta_1 \text{Ln Ingreso} + \beta_2 (\text{Ln Ingreso})^2 + \varepsilon$$

se ajustó a los datos. Una gráfica de secuencia del tiempo para los residuos de esta regresión se despliega en la figura 8-11.

Los residuos de la figura 8-11 parecen estar distribuidos aleatoriamente alrededor de cero con variabilidad constante. Parece que la regresión final representa adecuadamente los datos de Reynolds Metals. En este modelo no hay razón para dudar de los supuestos del término de error.

USO DE LA REGRESIÓN PARA PRONOSTICAR DATOS ESTACIONALES

Los modelos de descomposición para series de tiempo con patrones estacionales se analizaron en el capítulo 5. A continuación analizaremos un modelo de regresión que representa datos estacionales que están estrechamente alineados con una descomposición aditiva. En este modelo, la estacionalidad se maneja usando *variables ficticias* en la función de regresión.

Un modelo estacional para datos trimestrales con una tendencia en el 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 = variable que se va a pronosticar

t = índice de tiempo

S_2 = variable ficticia igual a 1 para el segundo trimestre del año y 0 para cualquier otro

S_3 = variable ficticia igual a 1 para el tercer trimestre del año y 0 para cualquier otro

S_4 = variable ficticia igual a 1 para el cuarto trimestre del año y 0 para cualquier otro

ε_t = errores, suponiendo que son independientes y están distribuidos de manera normal con media cero y varianza constante

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ = coeficientes que se van a estimar

Observe que los cuatro niveles de la variable cualitativa (trimestre o temporada) se describen sólo con tres variables ficticias. El último trimestre (el primer trimestre en este caso) se maneja mediante el término de intercepción β_0 . Para ver esto, para 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 del segundo trimestre, $S_2 = 1$, $S_3 = S_4 = 0$, y el nivel esperado es

$$E(Y_t) = \beta_0 + \beta_1 t + \beta_2 = (\beta_0 + \beta_2) + \beta_1 t$$

Surgen patrones similares en el tercero y cuarto trimestres.

En la ecuación 8.12, trimestres diferentes arrojan términos de intercepción diferentes. La tendencia y el patrón estacional se modelan como una serie de líneas rectas, una por cada trimestre. Se postula que las cuatro líneas tienen la misma pendiente, pero diferentes intercepciones. Los coeficientes de regresión para las variables ficticias representan cambios en la inter-

¹³Un modelo de regresión con variables explicativas $X, X^2, X^3, \dots$ se llama *modelo de regresión polinomial*.

cepción en relación con la intercepción del primer trimestre β_0 . Si existe un patrón estacional pero no hay tendencia, entonces se aplica la ecuación 8.12 con $\beta_1 = 0$.

Ejemplo 8.8

Una vez, James Brown, analista de la Washington Water Power Company, quería pronosticar el consumo doméstico de electricidad para el tercero y cuarto trimestres de 1996. Él sabía que los datos eran estacionales y decidió emplear la ecuación 8.12 para desarrollar la ecuación del pronóstico. Utilizó datos trimestrales desde 1980 hasta el segundo trimestre de 1996. Los datos del consumo de electricidad en millones de kilowatts-hora aparecen en la tabla 8-9.

James creó las variables ficticias S_2 , S_3 , y S_4 , para representar el segundo, tercero y cuarto trimestres, respectivamente. Los datos de los cuatro trimestres de 1980 se proporcionan en la tabla 8-10.

TABLA 8-9 Datos del consumo de electricidad para la Washington Water Power de 1980 a 1996; ejemplo 8.8

<i>Año</i>	<i>Trimestre</i>	<i>Kilowatts-hora (millones)</i>	<i>Año</i>	<i>Trimestre</i>	<i>Kilowatts-hora (millones)</i>
1980	1	1,071	1989	1	1,036
	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	1,065	1991	1	1,085
	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	1,047	1993	1	1,018
	2	667		2	670
	3	495		3	566
	4	794		4	811
1985	1	1,068	1994	1	962
	2	625		2	647
	3	499		3	630
	4	850		4	803
1986	1	975	1995	1	1,002
	2	623		2	887
	3	496		3	615
	4	728		4	828
1987	1	933	1996	1	1,003
	2	582		2	706
	3	490			
	4	708			
1988	1	953			
	2	604			
	3	508			
	4	708			

Fuente: Con base en el reporte anual de la Washington Water Power, varios

TABLA 8-10 Datos del consumo de electricidad de la Washington Water Power para cuatro trimestres de 1980; ejemplo 8.8

Y_t	S_2	S_3	S_4
1,071	0	0	0
648	1	0	0
480	0	1	0
746	0	0	1

TABLA 8-11 Resultado de computadora de la Washington Water Power para el ejemplo 8.8

Regression Analysis: Hours versus Time, 2nd Qt., 3rd Qt., 4th Qt.

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

$$S = 52.2488 \text{ R-Sq} = 92.4\% \text{ R-Sq(adj)} = 91.9\%$$

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

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% 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 de Minitab para correr el análisis estacional se describen en la sección de aplicaciones de Minitab al final del capítulo. Los resultados se presentan en la tabla 8-11.

El modelo de regresión estacional ajustado es

$$\hat{Y}_t = 968 + .938t - 342S_2 - 472S_3 - 230S_4$$

donde $S_2 = 2$ do. trim., $S_3 = 3$ er. trim. y $S_4 = 4$ to. trim. James notó que el modelo explicaba el 92.4% de la variabilidad en el consumo doméstico de electricidad medido en kilowatts-hora. El pronóstico para el tercer trimestre de 1996 se calculó como

$$\text{Horas} = 968 + .938(67) - 342(0) - 472(1) - 230(0) = 559$$

El pronóstico para el cuarto trimestre de 1996 fue

$$\text{Horas} = 968 + .938(68) - 342(0) - 472(0) - 230(1) = 802$$

James vio que los pronósticos de los diferentes trimestres se ubicaban a lo largo de cuatro líneas rectas. Todas las líneas tienen la misma pendiente (.938), pero las intercepciones cambian dependiendo del trimestre. Los pronósticos del primer trimestre caen a lo largo de la línea con intercepción en 968. La intercepción de los pronósticos del segundo trimestre caen a lo largo de la línea con intercepción $968 - 342 = 626$. La intercepción de los pronósticos del tercer trimestre es $968 - 472 = 496$, y la intercepción de los pronósticos del cuarto trimestre es $968 - 230 = 738$. James estaba contento de que el modelo del pronóstico reflejaba el patrón estacional y una ligera tendencia ascendente observada en las series. Dentro de un año determinado, los pronósticos del consumo de electricidad llegan a un máximo para el primer trimestre, son más bajos para el segundo trimestre, llegan a un mínimo para el tercer trimestre, y son la segunda cifra más alta para el cuarto trimestre.

PRONÓSTICOS ECONOMETRÍCOS

Cuando se aplica el análisis de regresión a datos económicos, las predicciones desarrolladas con estos modelos se llaman *pronósticos económicos*. Sin embargo, puesto que la teoría económica a menudo sugiere que los valores adquiridos por las cifras de interés están determinados por la interacción simultánea de diferentes fuerzas económicas, es necesario modelar esta interacción con un conjunto de ecuaciones simultáneas. Esta idea conduce a la elaboración de *modelos econométricos de ecuaciones simultáneas*. Estos modelos implican ecuaciones individuales que parecen ecuaciones de regresión. No obstante, en un sistema simultáneo, las ecuaciones individuales están relacionadas, y el modelo econométrico permite la *determinación conjunta de un grupo de variables dependientes* en términos de diversas variables independientes. Esto contrasta con la situación de regresión común, en la cual una sola ecuación determina el valor esperado de una variable dependiente en términos de las variables independientes.

Un modelo econométrico de ecuaciones simultáneas determina colectivamente los valores de un conjunto de variables dependientes, llamadas *variables endógenas* por los econométricos, en términos de valores de las variables independientes llamadas *variables exógenas*. Se supone que los valores de las *variables exógenas* influyen en las variables endógenas, pero no al revés. Un modelo de ecuaciones simultáneas completo tendrá tantas ecuaciones como variables endógenas contenga.

La simultaneidad en un sistema econométrico genera algunos problemas que requieren un tratamiento estadístico especial. Un tratamiento completo de modelos econométricos está más allá del alcance de este libro.¹⁴ Sin embargo, un modelo de dos ecuaciones ilustrará algunos conceptos.

La teoría económica sostiene que, en equilibrio, la cantidad suministrada es igual a la cantidad demandada a un precio particular. Es decir, la cantidad demandada, la cantidad suministrada y el precio se determinan de manera simultánea. En un estudio de elasticidad de la demanda por cambio en el precio, 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 de mano de obra

ε_t, v_t = términos de error independientes o que no están correlacionados entre sí.

¹⁴Pindyck y Rubinfeld (1998) ofrecen una explicación introductoria de los modelos econométricos de ecuaciones simultáneas.

Observe que en este modelo las variables de precio y cantidad, P_t y Q_t , aparecen en ambas ecuaciones. En la primera ecuación, la cantidad vendida está parcialmente determinada por el precio, y en la segunda ecuación, el precio está parcialmente determinado por la cantidad vendida. El precio y la cantidad son variables endógenas cuyos valores se determinan dentro del sistema. Las variables restantes, el ingreso y la temperatura en la primera ecuación, así como el costo de mano de obra en la segunda ecuación, son variables exógenas, cuyos valores se determinan fuera del sistema. Dadas las estimaciones adecuadas de los coeficientes en el modelo (el problema de la identificación), es posible determinar los pronósticos de, digamos, la demanda futura (ventas). Desde luego, para estimar la demanda futura, los valores futuros de las variables exógenas deben especificarse o estimarse por fuera del sistema. Además, debe determinarse los valores futuros de la variable precio.

Actualmente, los modelos econométricos a gran escala se usan para modelar el comportamiento de empresas específicas de una industria, sectores industriales seleccionados dentro de la economía, y la economía total. Los modelos econométricos pueden incluir cualquier número de ecuaciones simultáneas parecidas a una regresión. Los modelos econométricos permiten entender cómo funciona la economía y sirven para generar pronósticos de variables económicas clave. Los modelos econométricos son una ayuda importante en la formulación de políticas.

SERIES DE TIEMPO COINTEGRADAS

En el ejemplo 8.1 señalamos que un análisis de regresión que implica dos series de tiempo *no estacionarias* (nivel de la media no natural) altamente autocorrelacionadas puede generar regresiones falsas. Para protegerse contra las regresiones falsas, las regresiones entre variables no estacionarias deben realizarse como regresiones entre cambios en las variables (véase Granger y Newbold 1974).¹⁵ Si la regresión original es falsa, no debe haber relación entre los cambios. Sin embargo, existen situaciones especiales en que dos variables no estacionarias están relacionadas en el sentido de que tienden a adherirse una a la otra, y una regresión que sólo implica cambios en las variables pierde el mecanismo que conecta las variables adheridas.

Para ilustrar el caso de variables adheridas, consideremos que tanto el nivel de inventario como las ventas unitarias de un producto pueden ser no estacionarias (digamos, a causa del crecimiento); sin embargo, la diferencia entre las ventas y el nivel de inventario puede ser *estacionaria* (varía alrededor de un nivel fijo), un subproducto natural de los esquemas que ajustan los inventarios a las ventas inciertas. Es decir, como resultado de los ajustes continuos, las diferencias entre el inventario y las ventas pueden brincar de un periodo al siguiente, pero tienden a ser relativamente constantes. Las ventas y el inventario se adhieren una a la otra.

Para proceder más formalmente, dejemos que Y_t y X_t sean dos series de tiempo no estacionarias, cada una de las cuales se puede convertir en una serie de tiempo estacionaria por diferenciación simple. Se dice que las series no estacionarias que se convierten en estacionarias, al ser diferenciadas una vez, son integradas de orden 1.¹⁶ Para simplificar, suponga que las diferencias estacionarias de cada serie son los errores aleatorios. Entonces podemos escribir

$$Y_t - Y_{t-1} = \varepsilon_t$$

y

$$X_t - X_{t-1} = \nu_t$$

¹⁵También, una regresión falsa puede identificarse generalmente con un examen cuidadoso de los residuos y las autocorrelaciones residuales de la regresión que implica a las variables originales.

¹⁶Se dice que las series no estacionarias que se convierten en estacionarias cuando se diferencian n veces son integradas de orden n .

donde ε_t y v_t son términos de error independientes, con distribución normal, los dos con medias cero y varianzas σ_ε^2 y σ_v^2 , respectivamente.¹⁷ Tal como se plantea, Y_t y X_t no están vinculadas. Pero como en el ejemplo anterior de ventas de inventario, suponga que la diferencia $Y_t - X_t$ es estacionaria de la siguiente forma:

$$Y_t - X_t = \eta_t$$

donde η_t es el error independiente y normalmente distribuido con media cero y varianza σ_η^2 . La combinación lineal $Y_t - X_t$ que vincula Y_t y X_t se llama *relación cointegradora* de orden 0.¹⁸ En general, no hay manera de obtener una combinación lineal (media ponderada) de dos series no estacionarias, Y_t y X_t , para producir una serie estacionaria. Sin embargo, en el caso especial donde existe una relación tal, decimos que Y_t y X_t están cointegradas, con la relación de cointegración dada por la combinación lineal.¹⁹

Se dice que un conjunto de series de tiempo no estacionarias para las cuales la diferenciación simple produce una serie estacionaria en cada caso está *cointegrado* si y sólo si alguna combinación lineal de la serie es estacionaria. La combinación lineal estacionaria de la serie de tiempo cointegrada se llama *relación de cointegración*.

La cointegración es un concepto probabilístico. En nuestro ejemplo, la diferencia entre las ventas y el inventario no es fija, sino que fluctúa como una variable aleatoria estacionaria alrededor de algún nivel.

Ahora, la desviación $\eta = Y - X$ en cualquier tiempo puede influir en la evolución futura de las variables Y y X . Por ejemplo, muy poco inventario en relación con las ventas podría traer consigo escasez y ventas reducidas en el futuro y también podría tener como resultado un incremento en los inventarios futuros. Nosotros podemos adaptar la noción del comportamiento del vínculo de reciente que influye en la evolución futura modelando $Y_t - Y_{t-1}$ y $X_t - X_{t-1}$ como una función de valores retrasados de $\eta_t = Y_t - X_t$, el *término de corrección del error*.

$$Y_t - Y_{t-1} = \beta_0 + \beta_1(Y_{t-1} - X_{t-1}) + \varepsilon_t \quad (8.13a)$$

$$X_t - X_{t-1} = \beta'_0 + \beta'_1(Y_{t-1} - X_{t-1}) + v_t \quad (8.13b)$$

donde β_0 , β'_0 , β_1 , y β'_1 son parámetros que se van a estimar.

Los modelos de las ecuaciones 8.13a y 8.13b son los más sencillos de los modelos para series de tiempo cointegradas y pretenden ilustrar los conceptos. En general, la identificación, la estimación y la verificación de modelos de variables cointegradas plantean retos especiales. Más aún, los pronósticos de series cointegradas y las medidas de exactitud de los pronósticos deben preservar las relaciones cointegradoras. Si, en el intento de modelar la relación entre dos o más series de tiempo cointegradas, uno pierde el mecanismo de corrección del error que conecta las variables cointegradas, se pueden observar los efectos falsos.

La cointegración y los correspondientes modelos de corrección del error son de interés para los especialistas en elaborar pronósticos en los negocios, las finanzas, la economía y en las oficinas gubernamentales porque los vínculos entre las variables inherentemente no estaciona-

¹⁷En este caso se dice que las series Y_t y X_t siguen *caminos aleatorios*; es decir, es igualmente probable que un cambio en la posición en cada caso sea positivo o negativo. Las dos series de tiempo graficadas en la figura 8-2 siguen caminos aleatorios

¹⁸La relación de cointegración es de orden 0 porque implica las variables originales, no las variables diferenciadas.

¹⁹La cointegración no requiere que la diferencia en las variables sea estacionaria. Dos variables no estacionarias están cointegradas si existe *cualquier* combinación lineal de las variables que sea estacionaria.

rias a menudo son resultado de una acción de la administración o del gobierno, o bien, derivan de la aplicación de una política regulatoria, y son los subproductos naturales de variables teóricas de largo plazo.

Al lector interesado en aprender más de las series de tiempo cointegradas se le recomienda consultar los textos de Murray (1994), Diebold (2004) y otras referencias citadas al final de este capítulo.

APLICACIÓN A LA ADMINISTRACIÓN

Las aplicaciones descritas en el capítulo 5 también son apropiadas para este capítulo. Las técnicas descritas en el presente capítulo permiten al analista detectar y corregir el problema de correlación serial y, por lo tanto, desarrollar los mejores modelos de pronósticos. El resultado neto es que la administración y/o los economistas pueden tratar con una enorme variedad de datos dependientes del tiempo y sentir confianza de que las predicciones tienen fundamento. Las áreas en las que estas técnicas son particularmente útiles incluyen las siguientes:

- Pronóstico de ventas
- Proyecciones de precio de acciones y bonos
- Proyecciones del costo de materias primas
- Proyecciones de penetración de un producto nuevo
- Estimación de necesidades de personal
- Estudios de relación publicidad-ventas
- Control de inventario

Como estas aplicaciones implican variables que evolucionan con el tiempo, es probable que las variables estén autocorrelacionadas. Los modelos de pronóstico basados en las técnicas presentadas en este capítulo deben dar pronósticos más confiables que algunas de las técnicas consideradas anteriormente y que ignoran la autocorrelación.

Glosario

Autocorrelación (correlación serial). Existe cuando las observaciones sucesivas en el tiempo están relacionadas una con otra.

Modelo autorregresivo. Expresa un pronóstico como una función de valores previos de las series de tiempo.

Series de tiempo cointegradas. Se dice que un conjunto de series de tiempo no estacionarias para las cuales la diferenciación simple produce

una serie estacionaria en cada caso está cointegrado si y sólo si alguna combinación lineal de las series es estacionaria.

Relación de cointegración. La combinación lineal estacionaria de series de tiempo cointegradas se llama relación de cointegración.

Prueba de Durbin-Watson. Prueba usada para determinar si la autocorrelación positiva está presente.

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

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \quad (8.2)$$

Estadístico 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 del estadístico de Durbin-Watson con la autocorrelación de los residuos de retraso 1 (tamaño n)

$$DW = 2(1 - r_1(e)) \quad (8.4)$$

Modelo de regresión lineal simple transformado

$$Y'_t = \beta_0(1 - \rho) + \beta_1 X'_t + v_t \quad (8.5)$$

Diferencias generalizadas

$$Y'_t = Y_t - \rho Y_{t-1} \quad (8.6)$$

$$X'_t = X_t - \rho X_{t-1}$$

Primeras diferencias o diferencias simples

$$Y'_t = Y_t - Y_{t-1} \quad (8.7)$$

$$X'_t = X_t - X_{t-1}$$

Modelo logarítmico de regresión lineal

$$\text{Ln } Y_t = \beta_0 + \beta_1 \text{Ln } X_t + \varepsilon_t \quad (8.8)$$

Ecuación del pronóstico para la forma diferenciada del modelo logarítmico de regresión lineal

$$\text{Ln } \hat{Y}_t = \text{Ln } \hat{Y}_{t-1} + b_1(\text{Ln } X_t - \text{Ln } X_{t-1}) \quad (8.9)$$

Modelo autorregresivo de primer orden

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \quad (8.10)$$

Ecuación del 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 datos trimestrales

$$Y_t = \beta_0 + \beta_1 t + \beta_2 S_2 + \beta_3 S_3 + \beta_4 S_4 + \varepsilon_t \quad (8.12)$$

Modelos de corrección de errores

$$Y_t - Y_{t-1} = \beta_0 + \beta_1(Y_{t-1} - X_{t-1}) + \varepsilon_t \quad (8.13a)$$

$$X_t - X_{t-1} = \beta'_0 + \beta'_1(Y_{t-1} - X_{t-1}) + v_t \quad (8.13b)$$

Coefficientes estandarizados (véase el caso 8-2)

$$B = b \left(\frac{s_x}{s_y} \right) \quad (8.14)$$

Problemas

1. ¿Qué es la correlación serial y por qué se puede convertir en un problema cuando se analizan los datos de una serie de tiempo?
2. ¿Cuál es la causa principal de la correlación serial?
3. ¿Qué supuesto implícito de la regresión se viola con frecuencia cuando se analizan las variables de las series de tiempo?
4. Mencione el estadístico que se usa comúnmente para detectar la correlación serial.
5. Usted hace una prueba de correlación serial, en el nivel de .01, con 32 residuos de una regresión con dos variables independientes. Si el estadístico de Durbin-Watson es igual a 1.0, ¿cuál es su conclusión?
6. Usted hace una prueba de correlación serial, en el nivel de .05, con 61 residuos de una regresión con una variable independiente. Si el estadístico de Durbin-Watson es igual a 1.6, ¿cuál es su conclusión?
7. Sugiera alternativas para resolver el problema de la correlación serial.
8. ¿Cuáles son las variables explicativas en un modelo autorregresivo?
9. Tamson Russell, una economista que trabaja para el gobierno, estaba tratando de determinar la función de la demanda de combustible para vehículos de pasajeros en Estados Unidos. Tamson desarrolló un modelo que se basaba en el precio real de un galón de gasolina regular para predecir el consumo anual de gasolina. Después de agregar al modelo una variable que representaba la población de Estados Unidos, pudo explicar el 76.6% de la variación en el consumo de combustible. ¿Tenía Tamson un problema de correlación serial? Los datos aparecen en la tabla P-9.
10. Se pidió a Decision Science Association hacer un estudio de factibilidad de una propuesta para construir un centro vacacional ubicado a menos de media milla de la presa Grand Coulee. Mark Craze no estaba contento con el modelo de regresión que se basaba en el precio de un galón de gasolina para predecir el número de visitantes al Grand Coulee Dam Visitors Center. Después de graficar los datos en un diagrama de dispersión, Mark decidió usar una variable ficticia para representar las celebraciones importantes en el área. Mark usó un 1 para representar una celebración y un 0 si no había celebraciones. Note que el 1 en 1974 representa la feria mundial Expo 74 celebrada en Spokane, Washington, el 1 en 1983 representa el 50 aniversario de la construcción de la presa Grand Coulee, y el 1 en 1986 representa la feria mundial de Vancouver, Canadá. Mark también decidió considerar el tiempo como variable explicativa. Los datos se presentan en la tabla P-10. Suponga que a usted se le pidió escribir un reporte para que Mark lo presente a su jefe. Indique si la correlación serial es un problema. También indique qué información adicional se necesitaría conocer para determinar si es recomendable construir el centro vacacional.
11. Jim Jackson, un analista de tarifas de la Washington Water Power, se está preparando para una audiencia ante la comisión de tarifas necesaria para pronosticar los ingresos por consumo doméstico de electricidad para 1996. Jim decidió investigar tres variables explicativas potenciales: uso doméstico en kilowatts-hora (kWh), cargo doméstico por kWh (centavos/kWh), y número de clientes residenciales. Él recopiló datos de 1968 a 1995. Los datos se presentan en la tabla P-11. Jim testificó ante la Comisión de tarifas de Idaho, donde le preguntaron si la correlación serial era un problema. Él no supo qué contestar y le pidió a usted escribir una respuesta a la pregunta de la comisión.

TABLA P-9

<i>Año</i>	<i>Combustible consumido por automotores (miles de millones de galones) Y</i>	<i>Precio de gasolina (\$/galón) X₁</i>	<i>Población de Estados Unidos X₂</i>
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 Abstracto of the United States, varios años.

TABLA P-10

<i>Año</i>	<i>Número de visitantes Y</i>	<i>Tiempo X₁</i>	<i>Precio de gasolina (\$/galón) X₂</i>	<i>Celebración X₃</i>
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: Basado en el Gran Coulee Dam Visitors Center y el Statistical Abstract of the United States, 1988.

TABLA P-11

<i>Año</i>	<i>Ingreso (millones de \$) Y</i>	<i>Uso por kWh X₁</i>	<i>Cargo (centavos/kWh) X₂</i>	<i>Número de clientes X₃</i>
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: Basado en "Financial and Operating Supplement", Washington Water Power Annual Report, varios años.

12. Paul Raymond, presidente de la Washington Water Power, estaba preocupado por la posibilidad de un intento de compra y por el hecho de que el número de poseedores de acciones comunes ha ido disminuyendo desde 1983. Suponga que le da instrucciones a usted para que estudie el número de accionistas comunes desde 1968 y se prepare para calcular un pronóstico para 1996. Usted decide investigar tres variables explicativas potenciales: ganancias por acción (común), dividendos por acción (común) y dividendos pagados. Usted recopila los datos de 1968 a 1995, como se indica en la tabla P-12.
 - a) Ingrese estos datos en la computadora y determine el mejor modelo usando sus variables explicativas seleccionadas.
 - b) ¿Es un problema la correlación serial en este modelo?
 - c) Si la correlación es un problema, redacte un escrito para Paul donde analice varias soluciones para el problema de correlación e incluya su recomendación final.
13. La Thompson Airlines determinó que el 5% del número total de pasajeros nacionales estadounidenses vuela en los aviones de la compañía. Se le asigna a usted la tarea de pronos-

TABLA P-12

Año	Accionistas comunes Y	Ganancias por acción (\$) X_1	Dividendos por acción (\$) X_2	Pago de dividendos X_3
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: Basado en el "Financial and Operating Supplement", del Washington Water Power Annual Report, varios años.

ticar el número de pasajeros que volarán en la Thompson Airlines en 2007. Los datos se presentan en la tabla P-13.

- Desarrolle un modelo de regresión de series de tiempo, usando el tiempo como la variable independiente y el número de pasajeros como la variable dependiente. Ajuste este modelo.
- ¿Es viable el supuesto de errores independientes para este modelo?
- Ajuste el modelo del inciso a) con los logaritmos del número de pasajeros como la variable dependiente.
- Repita el inciso a) con el tiempo representado por una tendencia exponencial (véase la ecuación 5.6).
- ¿Cuál modelo prefiere usted, el del inciso c) o d)? ¿Por qué?
- ¿Los errores de los modelos de los incisos c) y d) parecen ser independientes? Si no es así, ¿qué problema(s) podría(n) surgir cuando se use uno (o ambos) de estos modelos ajustados para pronosticar?
- Con base en su modelo preferido, pronostique el número de pasajeros para 2007 de la Thompson Airlines.

TABLA P-13

<i>Año</i>	<i>Número de pasajeros (miles)</i>	<i>Año</i>	<i>Número de pasajeros (miles)</i>	<i>Año</i>	<i>Número de pasajeros (miles)</i>	<i>Año</i>	<i>Número de pasajeros (miles)</i>
1982	22.8	1989	45.4	1996	79.9	2003	144.8
1983	26.1	1990	46.3	1997	96.3	2004	147.9
1984	29.4	1991	45.8	1998	109.0	2005	150.1
1985	34.5	1992	48.0	1999	116.0	2006	151.9
1986	37.6	1993	54.6	2000	117.2		
1987	40.3	1994	61.9	2001	124.9		
1988	39.5	1995	69.9	2002	136.6		

14. La Thomas Furniture Company concluyó que su programa de producción puede mejorarse desarrollando un método exacto para pronosticar las ventas trimestrales. El analista de la compañía, el señor Estes, decide investigar la relación entre los permisos para construcción de viviendas y las ventas de muebles en el área de Springfield. Estes siente que los permisos irán por delante de las ventas uno o dos trimestres. Además, se pregunta si las ventas de muebles son estacionales. Estes decide considerar otra variable independiente:

$$X_2 = \begin{cases} 0 & \text{para ventas del primero o segundo trimestres} \\ 1 & \text{para ventas del tercero o cuarto trimestres} \end{cases}$$

Los datos se presentan en la tabla P-14.

- Desarrolle un modelo de regresión que use los permisos para construcción de casas como variable explicativa.
- Efectúe la prueba de autocorrelación para este modelo.
- Desarrolle un modelo de regresión que use los permisos y la variable ficticia de la estacionalidad como variables explicativas.
- ¿Existe un patrón estacional significativo en estos datos? (Haga una prueba con un nivel de .05).
- ¿Existe un problema de autocorrelación con el modelo de regresión múltiple desarrollado en el inciso c)? Si es así, ¿cómo podría corregirse?

TABLA P-14

<i>Año</i>	<i>Trimestre</i>	<i>Ventas (en miles)</i> <i>Y</i>	<i>Permisos</i> <i>X₁</i>	<i>Año</i>	<i>Trimestre</i>	<i>Ventas (1,000s)</i> <i>Y</i>	<i>Permisos</i> <i>X₁</i>
2002	3		19	2005	1	120	72
	4		3		2	150	31
2003	1	120	35		3	660	19
	2	80	11		4	270	14
	3	400	11	2006	1	200	75
	4	200	16		2	280	41
2004	1	75	32		3	800	17
	2	120	10		4	320	10
	3	270	12	2007			
	4	155	21				

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), p. 132.

- f) Con base en el modelo del inciso a), pronostique las ventas de la Thomas Furniture Company para los cuatro trimestres de 2007. (Observe que usted necesitará algunos valores adicionales de los permisos de construcción para desarrollar pronósticos de los cuatro trimestres de 2007).
15. National Presto es un fabricante de pequeños aparatos electrodomésticos y artículos para el hogar que incluyen ollas de presión, calentadores, envasadores, sartenes, parrillas, hornos asadores, freidoras, máquinas para hacer rosetas de maíz, abrelatas, cafeteras, rebanadoras, mezcladoras y hornillas portátiles. Sus ventas trimestrales en millones de dólares por varios años se presentan en la tabla P-15. Presto hace la mayoría de sus ventas en Navidad, de manera que existe un fuerte efecto estacional. Desarrolle un modelo de regresión múltiple usando variables ficticias para pronosticar las ventas para el tercero y cuarto trimestres de 1996. Escriba un reporte que resuma sus resultados.
16. Los datos de la tabla P-16 presentan las ventas trimestrales ajustadas estacionalmente de la Dickson Corporation y de la industria completa durante 20 trimestres.
- Ajuste un modelo de regresión lineal y guarde los residuos. Grafique los residuos contra el tiempo y obtenga las autocorrelaciones residuales. ¿Qué encontró usted?
 - Calcule el estadístico de Durbin-Watson y determine si existe autocorrelación.
 - Estime el coeficiente de regresión, β_1 , usando diferencias generalizadas. (Estime ρ con el coeficiente de autocorrelación residual con un retraso de 1).
 - Compare los errores estándar de las dos estimaciones de β_1 obtenidas usando los datos originales y las diferencias generalizadas. ¿Cuál estimación es más exacta? Explique.
17. Remítase al ejemplo 8.5. Usando los datos de Sears de la tabla 8-5, convierta las ventas y los valores del ingreso disponible a diferencias simples. Es decir, construya los números $Y'_t = Y_t - Y_{t-1}$ y $X'_t = X_t - X_{t-1}$. Ajuste un modelo de regresión lineal simple a los datos diferenciados. Compare sus resultados con los resultados obtenidos por el método de diferencias generalizadas del ejemplo 8.5. ¿Esperaba que fueran muy diferentes? Explique.
18. Si bien los datos de la serie de tiempo de la tabla P-18 son antiguos, proporcionan las bases para realizar algún modelo de regresión de interés. Con base en los datos de la tabla P-17,

TABLA P-16

Año	Trimestre	Ventas de Dickson (en miles)	Ventas de la industria (en millones)	Año	Trimestre	Ventas de Dickson (en miles)	Ventas de la industria (en millones)
		Y	X_1			Y	X_1
2002	1	83.8	31.8	2004	3	98.2	37.1
	2	85.6	32.5		4	97.2	36.6
	3	87.8	33.2	2005	1	100.1	37.6
	4	86.1	32.4		2	102.6	38.3
2003	1	89.6	33.8		3	105.4	39.3
	2	91.0	34.3	2006	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
2004	1	96.4	36.4		3	110.1	41.1
	2	96.0	36.3		4	111.1	41.4

TABLA P-17

Año	Ahorros personales Y	Ingresos personales X	Año	Ahorros personales Y	Ingresos personales X	Año	Ahorros personales Y	Ingresos personales 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			

intente relacionar los ahorros personales con los ingresos personales (en miles de millones de dólares) para el periodo de 1935 a 1954.

- Ajuste un modelo de regresión lineal simple a los datos de la tabla P-17, usando el ingreso personal para predecir los ahorros personales. **1.** Haga una prueba de significancia de la regresión con un nivel $\alpha = .01$; **2.** calcule r^2 e interprete esta cantidad; y **3.** haga una prueba de autocorrelación usando la prueba de Durbin-Watson con $\alpha = .05$. ¿Debe usted modificar su conclusión del inciso 1? ¿Cómo puede mejorarse el modelo?
- Construya una variable ficticia, X_2 , para los años de la Segunda Guerra Mundial. Sea $X_2 = 0$ para tiempos de paz, y $X_2 = 1$ para tiempos de guerra. La guerra tuvo lugar de 1941 a 1945. Ajuste un modelo de regresión lineal múltiple usando el ingreso personal y la variable ficticia de los años de guerra como variables explicativas de los ahorros personales. Evalúe los resultados. Específicamente, **1.** elabore una prueba para determinar si el conocimiento de los años de guerra hace una contribución significativa al pronóstico de los ahorros personales más allá de la proporcionada por los ingresos personales (establezca $\alpha = .01$), y **2.** haga una prueba de autocorrelación. ¿El modelo de regresión múltiple es mejor que el modelo de regresión lineal simple del inciso a)? Argumente.

- Circuit City Inc., es un minorista de equipos de audio y video y otros productos electrónicos para el consumidor y la oficina. Recientemente, las ventas han disminuido; en diciembre

TABLA P-18

Año	31 de mayo	31 de agosto	30 de nov.	Feb 28
1996	1,742	1,806	1,999	2,897
1997	1,851	1,948	2,039	3,156
1998	2,078	2,369	2,222	3,289
1999	2,205	2,423	2,495	3,476
2000	2,449	2,566	2,336	3,177
2001	1,882	2,037	2,280	3,391
2002	2,118	2,221	2,422	3,239
2003	2,150	2,350	2,600	3,400

Fuente: *The Value Investment Survey* (Nueva York: Value Line, 2002), p.1725.

disminuyeron un total de 5%. Entre las razones estuvieron una competencia feroz en el ambiente minorista entre los vendedores de productos electrónicos, la deflación del precio en muchos productos, una baja repentina en el tráfico de la tienda y una escasa demanda de muchos artículos de hardware.²⁰ Jim Lowe tiene la tarea de pronosticar las ventas para 2003. Jim tiene acceso a las estimaciones de *The Value Line Investment Survey* (véase la tabla P-18); sin embargo, teme que las estimaciones sean optimistas.

- Grafique los datos de 1996 a 2002 como una serie de tiempo. ¿Estos datos parecen ser estacionales? Calcule la función de autocorrelación.
 - A partir de los resultados del inciso a), un modelo autorregresivo con la variable explicativa ventas retrasada cuatro periodos podría ser adecuado. ¿Por qué?
 - Dejemos que Y_t sean las ventas de Circuit City. Con base en los datos de 2002, ajuste un modelo autorregresivo de la forma $Y_t = \beta_0 + \beta_1 Y_{t-4} + \varepsilon_t$, y guarde los residuos. Calcule las autocorrelaciones residuales.
 - Use el modelo ajustado del inciso c) para generar pronósticos de ventas para los cuatro trimestres de 2003. Compare sus pronósticos con las estimaciones de *Value Line*.
 - ¿Estaba Jim en lo correcto al pensar que las estimaciones de *Value Line* para el 2003 eran optimistas?
 - ¿Piensa usted que el modelo autorregresivo del inciso c) puede mejorarse? Discútalos.
20. La demanda de una mercancía depende generalmente del ingreso del consumidor, el precio real de la mercancía y el precio real de los accesorios o productos similares. La tabla P-19 indica el consumo per cápita de pollo en Estados Unidos (en libras); el ingreso disponible per cápita (en dólares); y los precios al menudeo del pollo, puerco y carne de res (en centavos por libra) de varios años.
- Calcule la matriz de correlación para todas las variables, usando tanto las unidades originales como las unidades transformadas en logaritmos. Comente sobre la fortaleza implicada de la asociación lineal entre el pollo consumido y cada una de las variables restantes. ¿Puede usted pensar una razón por la que se debe tener cuidado al interpretar las magnitudes de los coeficientes de correlación obtenidos a partir de los datos de la serie de tiempo?
 - Usando los datos originales, corra un programa de regresión por pasos con el consumo de pollo como la variable dependiente y las variables restantes como variables explicativas. Establezca α para enter = alpha to remove = .05.

²⁰A finales de 2006, Wal-Mart decidió rebajar drásticamente el precio de sus televisores de pantalla plana de 42 pulgadas a menos de \$1,000. Este movimiento disparó la fusión financiera de algunos minoristas de productos electrónicos de consumo durante varios meses siguientes. Circuit City cerró 70 tiendas en Estados Unidos, despidió a 3,400 empleados y puso en subasta pública sus 800 tiendas canadienses. Fuente: MSN Money, 23 de abril, 2007.

TABLA P-19

<i>Año</i>	<i>Consumo per cápita de pollo (libras)</i>	<i>Ingreso disponible (\$)</i>	<i>Precio del pollo (¢/libra)</i>	<i>Precio del puerco (¢/libra)</i>	<i>Precio de la res (¢/libra)</i>
1	28.0	397.5	42.2	50.7	78.3
2	29.9	413.3	38.1	52.0	79.2
3	30.1	439.2	40.3	54.0	79.2
4	30.9	459.7	39.5	55.3	79.2
5	31.4	492.9	37.3	54.7	77.4
6	33.7	528.6	38.1	63.7	80.2
7	35.6	560.3	39.3	69.8	80.4
8	35.6	624.6	37.8	65.9	83.9
9	37.1	666.4	38.4	64.5	85.5
10	38.5	717.8	40.1	70.0	93.7
11	40.3	768.2	38.6	73.2	106.1
12	40.2	843.3	39.8	67.8	104.8
13	41.7	911.6	39.7	79.1	114.0
14	39.9	931.1	52.1	95.4	124.1
15	39.7	1,021.5	48.9	94.2	127.6
16	39.0	1,165.9	58.3	123.5	142.9
17	39.1	1,349.6	57.9	129.9	143.6
18	42.8	1,449.4	56.5	117.6	139.2
19	44.9	1,575.5	63.7	130.9	165.5
20	48.3	1,759.1	61.6	129.8	203.3
21	49.0	1,994.2	58.9	128.0	219.6
22	49.4	2,258.1	66.4	141.0	221.6
23	49.6	2,478.7	70.4	168.2	232.6

- c) Corra una regresión completa del consumo de pollo sobre las variables restantes. Asegúrese de eliminar una por una las variables que considere que no son significativas hasta que usted esté satisfecho con su modelo final. ¿Es congruente su resultado con el resultado del procedimiento por pasos del inciso b)? ¿Es probable que la correlación serial sea un problema en este análisis de regresión?
21. Repita los incisos b) y c) del problema 20 con los datos transformados logarítmicamente. Haga una interpretación de los coeficientes del ingreso y del precio del pollo en términos de elasticidades. Con base en su función de regresión ajustada final, indique cómo se obtendría un pronóstico del consumo de pollo para los años siguientes.
22. Remítase al problema 20. Considere sólo las variables de consumo de pollo, ingreso y precio del pollo en las unidades originales. Calcule las diferencias simples de cada una de las variables. Con base en los datos diferenciados haga una regresión del cambio en el consumo de pollo sobre el cambio en el ingreso y el cambio en el precio del pollo. Examine los residuos. Interprete los resultados de su análisis de regresión. ¿Alguna de sus variables explicativas tiene mucho poder explicativo? Explique.
23. Remítase al problema 20. Corra una regresión lineal simple del consumo de pollo sobre el consumo de pollo retrasado un periodo de tiempo. Examine los residuos. Interprete los

resultados de su análisis de regresión. Interprete los resultados de su análisis de regresión. ¿Es probable que el consumo de pollo de ese año sea una buena variable explicativa del consumo de pollo para el año siguiente? Explique. ¿Podemos inferir el efecto de un cambio en el precio del pollo sobre el consumo de pollo con este modelo?

24. Considere el sistema de dos variables

$$X_t = X_{t-1} + v_t$$

$$Y_t = X_t + \varepsilon_t$$

donde v_t y ε_t están, cada una, distribuidas independientemente con media cero y varianza σ^2 . Desarrolle una expresión para $Y_t - Y_{t-1}$, y demuestre que X y Y están cointegradas. ¿Cuál es en este caso la combinación lineal de cointegración?

CASOS

CASO 8-1 COMPAÑÍA DE SU ELECCIÓN

La salud de una compañía puede revisarse cada mes, cada trimestre y/o anualmente con mediciones de un arreglo de variables. Para cualquiera de estas variables existirán muchas otras que ayudan a comprender su comportamiento y que pueden servir como variables explicativas en una ecuación de pronóstico.

El propósito de este caso es estimular la identificación de una variable importante en la serie de tiempo

para una compañía de su elección y luego analizar los patrones en los datos usando el análisis de autocorrelación. Además, usted puede usar un programa de computación adecuado para desarrollar una ecuación que resulte útil para pronosticar valores futuros de la variable de su serie de tiempo.

TAREA

1. Identifique una compañía u organización que le interese. La compañía puede ser una compañía local o nacional que haya publicado información, incluyendo las mediciones de variables en series de tiempo. Identifique una variable clave para la compañía que seleccionó y registre los valores de varios años, trimestres o meses.
2. Calcule varios coeficientes de autocorrelación y grafique la función de autocorrelación.
3. Con base en el patrón de la función de autocorrelación, describa los patrones de su serie de tiempo.
4. Calcule las primeras diferencias para sus datos y elabore la función de autocorrelación para los datos diferenciados. Describa los patrones resultantes en las series de tiempo de las primeras diferencias.
5. Identifique diversas variables explicativas potenciales para su variable dependiente. En este proceso usted puede usar los registros de su compañía y otras fuentes de datos.
6. Desarrolle una ecuación de pronóstico para su variable dependiente usando una o más de las variables explicativas identificadas.
7. Examine los residuos de su modelo ajustado. En particular, verifique la autocorrelación. Una vez que usted está satisfecho con su ecuación de pronóstico, genere los pronósticos para los siguientes seis periodos. Si es posible, compare sus pronósticos con los valores reales.

CASO 8-2 ÍNDICE DE ACTIVIDAD DE NEGOCIOS PARA EL CONDADO DE SPOKANE

Antes de 1973, el condado de Spokane, Washington, no tenía una medición actualizada de la actividad general de negocios. Sin embargo, lo que pasa en esa zona, como un todo, afecta a los negocios locales, las oficinas de gobierno y a los individuos. Los planes y las políticas elaborados por una unidad económica estarían incompletos sin algún conocimiento confiable acerca del funcionamiento reciente de la economía de la cual la unidad es un componente. Un índice de la actividad de negocios de Spokane debería servir como un insumo vital en la formulación de estrategias y decisiones de las organizaciones públicas y privadas.

Un índice de actividad de negocios es un indicador de los cambios en relación con las condiciones globales de los negocios dentro de una región específica. A nivel nacional, el producto interno bruto (PIB, calculado por el Departamento de Comercio) y el índice de producción industrial (compilado por la Reserva Federal) se consideran, en general, como excelentes indicadores. Cada una de estas series está basada en miles de piezas de información: la recopilación, revisión y cálculo es una tarea que consume tiempo y dinero. Para un área local como el condado de Spokane, Washington, una versión simplificada del índice, capaz de proporcionar una exactitud razonable e información actual a un costo moderado, es muy deseable.

La regresión múltiple se puede usar para elaborar un índice de actividad de negocios. Existen tres preguntas esenciales que deben contestarse para elaborar tal índice:

- ¿Cuáles son los componentes del índice?
- ¿Estos componentes representan adecuadamente los cambios en las condiciones generales de los negocios?
- ¿Qué peso se debe asignar a cada uno de los componentes seleccionados?

Las respuestas a estas preguntas se obtienen a través de un análisis de regresión.

El doctor Shik Chun Young, profesor de economía en la Universidad del Este de Washington, está intentando desarrollar un índice de actividad de negocios para el condado de Spokane. Young seleccionó el ingreso personal como la variable dependiente. A nivel del condado, el ingreso personal se considera como el mejor indicador de las condiciones locales de los negocios. El

ingreso personal mide el ingreso total recibido por las familias antes de pagar impuestos. Puesto que las actividades productivas por lo general se remuneran monetariamente, el ingreso personal, de hecho, se considera como un dato representativo razonable del desempeño económico en general. ¿Por qué entonces es necesario elaborar otro índice si el ingreso personal puede servir como un buen indicador de la actividad de los negocios? Por desgracia, los datos del ingreso personal al nivel del condado son estimados por el Departamento de Comercio de los Estados Unidos sobre una base anual y están disponibles 16 meses más tarde. Por consiguiente, estos datos se usan poco para la planeación de corto plazo. La tarea de Young es establecer un índice de la actividad de los negocios actualizado.

Las variables independientes se obtienen de aquellos datos locales que están fácilmente disponibles sobre una base mensual. Actualmente, cerca de 50 series de datos mensuales están disponibles, desde empleo, actividades bancarias y transacciones inmobiliarias hasta consumo eléctrico. Si todas las series se incluyeran en el análisis de regresión, el trabajo no sería muy productivo porque quizá sólo unas cuantas de estas series serían significativas desde el punto de vista estadístico. Por lo tanto, es necesario conocer la relación entre el ingreso personal y los datos disponibles para determinar cuáles variables independientes van a incluirse en la ecuación de regresión. Con el conocimiento de la economía de Spokane, Young seleccionó las 10 series siguientes:

- X_1 , empleo total
- X_2 , empleo en la manufactura
- X_3 , empleo en la construcción
- X_4 , empleo en el comercio al mayoreo y al menudeo
- X_5 , empleo en servicios
- X_6 , débitos bancarios
- X_7 , depósitos bancarios a la vista
- X_8 , permisos de construcción expedidos
- X_9 , préstamos hipotecarios
- X_{10} , consumo total de electricidad

El primer paso en el análisis es ajustar el modelo

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_{10} X_{10}$$

donde

$$\begin{aligned}
 Y &= \text{ingreso personal} \\
 \beta_0 &= \text{la intercepción con el eje} \\
 \beta_1, \beta_2, \dots, \beta_{10} &= \text{los coeficientes de las respectivas variables independientes}
 \end{aligned}$$

Cuando el modelo anterior se ajusta a los datos, la R^2 ajustada es .96, lo cual significa que las 10 variables en conjunto explican el 96% de la variabilidad en la variable dependiente, el ingreso personal. Sin embargo, otros estadísticos de regresión indican problemas. Primero, de estas 10 variables independientes, sólo tres —a saber, empleo total, empleo en el área de servicios y débitos bancarios— tienen un valor t calculado significativo en el nivel .05. Segundo, la matriz de correlación indica un alto grado de interdependencia entre varias de las variables independientes (multicolinealidad).²¹ Por ejemplo, el empleo total y los débitos bancarios tienen un coeficiente de correlación de .88; el consumo eléctrico total y los depósitos bancarios a la vista .76; y los permisos de construcción expedidos y préstamos hipotecarios, .68. Tercero, una prueba de autocorrelación usando el estadístico de Durbin-Watson de .91 indica que los valores sucesivos de la variable dependiente están correlacionados positivamente. Desde luego, la autocorrelación es algo común en los datos de las series de tiempo; en general, las observaciones en la misma serie tienden a estar relacionadas unas con otras.

Puesto que uno de los supuestos básicos en el análisis de regresión es que las observaciones de la variable dependiente son aleatorias, Young elige tratar primero el problema de autocorrelación. Decide calcular las primeras diferencias, o cambios, en un intento por minimizar la interdependencia entre las observaciones de cada una de las series de tiempo. Las 10 variables independientes ahora se miden por la diferencia entre los periodos y no por el valor absoluto de cada periodo. Con la finalidad de que los conjuntos de datos puedan distinguirse, se usa una nueva designación de las variables independientes:

$$\begin{aligned}
 \Delta X_1 &= \text{cambio en el empleo total} \\
 \Delta X_2 &= \text{cambio en el empleo de la manufactura} \\
 \Delta X_3 &= \text{cambio en el empleo de la construcción} \\
 \Delta X_4 &= \text{cambio en el empleo de comercio al mayoreo y menudeo} \\
 \Delta X_5 &= \text{cambio en el empleo de servicios}
 \end{aligned}$$

$$\begin{aligned}
 \Delta X_6 &= \text{cambio en débitos bancarios} \\
 \Delta X_7 &= \text{cambio en depósitos bancarios a la vista} \\
 \Delta X_8 &= \text{cambio en los permisos de construcción expedidos} \\
 \Delta X_9 &= \text{cambio en préstamos hipotecarios} \\
 \Delta X_{10} &= \text{cambio en el consumo total de electricidad}
 \end{aligned}$$

El modelo de regresión se convierte en

$$E(\Delta Y) = \beta_0 + \beta_1 \Delta X_1 + \beta_2 \Delta X_2 + \dots + \beta_{10} \Delta X_{10}$$

donde

$$\begin{aligned}
 \Delta Y &= \text{el cambio en el ingreso personal} \\
 \beta_0 &= \text{la intercepción } \Delta Y \\
 \beta_1, \beta_2, \dots, \beta_{10} &= \text{los coeficientes de las respectivas variables independientes}
 \end{aligned}$$

Al correr la regresión usando este modelo, con base en los datos de la primera diferencia, se produce un estadístico de Durbin-Watson de 1.71. Esto indica que ya no existe autocorrelación seria.

El siguiente paso es determinar cuáles de las 10 variables son variables explicativas significativas de la variable dependiente. Con la variable dependiente, ΔY , se hace una regresión contra varias combinaciones posibles de las 10 variables explicativas potenciales para seleccionar la mejor ecuación. Los criterios empleados en la selección son

- Una $\bar{R}^2$ satisfactoriamente alta
- Coeficientes de correlación bajos entre las variables independientes
- Coeficientes significativos (en el nivel .05) para cada una de las variables independientes

Después de un escrutinio cuidadoso de los resultados de la regresión, Young encuentra que la ecuación que contiene ΔX_4 , ΔX_5 , y ΔX_{10} como las variables independientes es la que mejor cumple con los criterios de selección.

Sin embargo, Young razona que (además de los usos comerciales e industriales) el consumo total de electricidad incluye el consumo doméstico, el cual no debe tener una relación significativa con la actividad de negocios en el corto plazo. Para probar esta hipótesis, Young subdivide el consumo total de electricidad en cuatro variables:

$$\begin{aligned}
 \Delta X_{11} &= \text{cambio en el uso doméstico de electricidad} \\
 \Delta X_{12} &= \text{cambio en el uso comercial de electricidad}
 \end{aligned}$$

²¹Debe tenerse cierto cuidado al interpretar una correlación entre dos variables de una serie de tiempo, ya que la autocorrelación en la serie individual puede generar correlación falsa; véase el ejemplo 8.1.

TABLA 8-12 Variables de regresión de Young

Ecuación	Variables independientes	Variable dependiente
A	$\Delta X_4, \Delta X_5, \Delta X_{11}$	ΔY
B	$\Delta X_4, \Delta X_5, \Delta X_{12}$	ΔY
C	$\Delta X_4, \Delta X_5, \Delta X_{13}$	ΔY
D	$\Delta X_4, \Delta X_5, \Delta X_{14}$	ΔY

ΔX_{13} , cambio en el uso industrial de electricidad
 ΔX_{14} , cambio en el uso comercial e industrial de electricidad

Las cuatro variables, combinadas con ΔX_4 y ΔX_5 , se utilizan para generar cuatro nuevas ecuaciones de regresión (véase la tabla 8-12).

El análisis estadístico indica que la ecuación D en la tabla 8-12 es la mejor. Comparada con la ecuación anterior, la cual contiene ΔX_4 , ΔX_5 , y ΔX_{10} como variables independientes, la ecuación A es la única que muestra un deterioro en la significancia estadística. Este resultado confirma la noción de Young de que los usos industriales y comerciales de electricidad son mejores variables explicativas del ingreso personal que el consumo total de electricidad, el cual incluye el uso doméstico de electricidad.

Por lo tanto, se selecciona la ecuación D como la ecuación de regresión final, y los resultados son

$$\Delta \hat{Y} = -1.86 + 17.10\Delta X_4 + 23.01\Delta X_5 + .007\Delta X_{14}$$

(4.07) (5.61) (.002)

$$n = 15 \quad F = 26.26$$

$$DW = 1.77 \quad \bar{R}^2 = .835$$

Las cifras entre paréntesis debajo de los coeficientes de regresión son los errores estándar de la estimación. Los valores t de los coeficientes son 4.20, 4.10 y 3.50 para ΔX_4 , ΔX_5 , y ΔX_{14} , respectivamente. La $\bar{R}^2$ indica que aproximadamente el 84% de la varianza en el cambio del ingreso personal se explica por las tres variables independientes. El estadístico de Durbin-Watson (DW) indica que la autocorrelación no es un problema. Además, la matriz del coeficiente de correlación de la figura 8-13 indica un grado bajo de interdependencia entre las tres variables independientes.

Para efectos de la elaboración del índice, las variables independientes en la ecuación de regresión final se convierten en los componentes del índice. Los pesos de los componentes se pueden determinar a partir de los

TABLA 8-13 Matriz del coeficiente de correlación

	ΔX_4	ΔX_5	ΔX_{14}
ΔX_4	1.00		
ΔX_5	.45	1.00	
ΔX_{14}	.11	.12	1.00

coeficientes de regresión. (Recuerde que el coeficiente de regresión representa el cambio promedio en la variable dependiente por una unidad de incremento en la variable independiente). Sin embargo, puesto que las variables en la ecuación de regresión no se miden en las mismas unidades (por ejemplo, ΔY se mide en miles de dólares, y ΔY_{14} en miles de kilowatts-hora), los coeficientes de regresión deben transformarse en valores relativos. Esta transformación se lleva a cabo calculando sus *coeficientes estandarizados* o *coeficientes B*.

$$B = b \left(\frac{s_x}{s_y} \right) \quad (8.14)$$

donde

b = coeficiente de regresión de la variable independiente

s_x = desviación estándar de la variable independiente

s_y = desviación estándar de la variable dependiente

Por lo general, los valores de todos estos estadísticos están disponibles en el resultado de la regresión que arroja la computadora. Así, los coeficientes estandarizados de las tres variables independientes son

$$B_4 = .4959$$

$$B_5 = .4833$$

$$B_{14} = .3019$$

$$\text{Total} = 1.2811$$

Como la suma de los pesos en un índice debe dar 100%, los coeficientes estandarizados se normalizan como se indica en la tabla 8-14.

Después de determinar los componentes y sus pesos respectivos, los siguientes pasos nos dan el índice:

1. Calcule el cambio porcentual de cada componente a partir del periodo base.

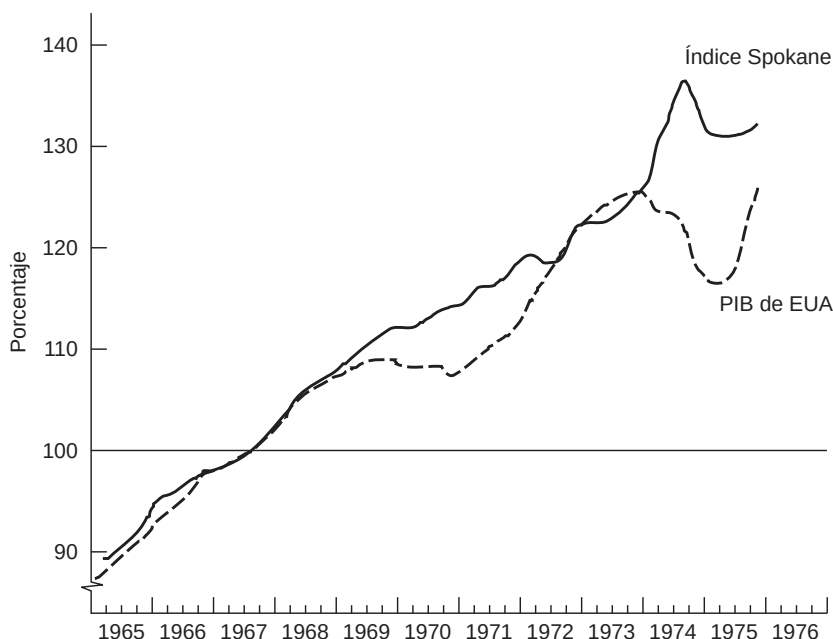


FIGURA 8-12 Índice de actividad en los negocios para el condado de Spokane y el PIB de Estados Unidos, en dólares constantes (1967 = 100)

TABLA 8-14 Coeficientes estandarizados

Componente	Peso
ΔX_4	$\frac{.4959}{1.2811} = .3871$
ΔX_5	$\frac{.4833}{1.2811} = .3772$
ΔX_{14}	$\frac{.3019}{1.2811} = .2357$
	Total = 1.000

- Multiplique el cambio porcentual por el peso apropiado.
- Suma los cambios porcentuales ponderados que se obtuvieron en el paso 2.

El índice completo de actividad del área de Spokane, para el periodo considerado, se compara con el PIB de Estados Unidos, en dólares constantes (1967 = 100), en la figura 8-12.

PREGUNTAS

- ¿Por qué decidió Young resolver primero el problema de la autocorrelación?
- ¿Hubiera sido mejor eliminar primero la multicolinealidad y luego enfrentar la autocorrelación?
- ¿Cómo afecta el tamaño pequeño de la muestra al análisis?
- ¿La regresión realizada sobre las primeras diferencias debió pasar por el origen?
- ¿El uso de los datos retrasados tiene algún potencial?
- ¿Qué conclusiones se pueden extraer de la comparación entre el índice de la actividad de negocios del condado de Spokane y el PIB?

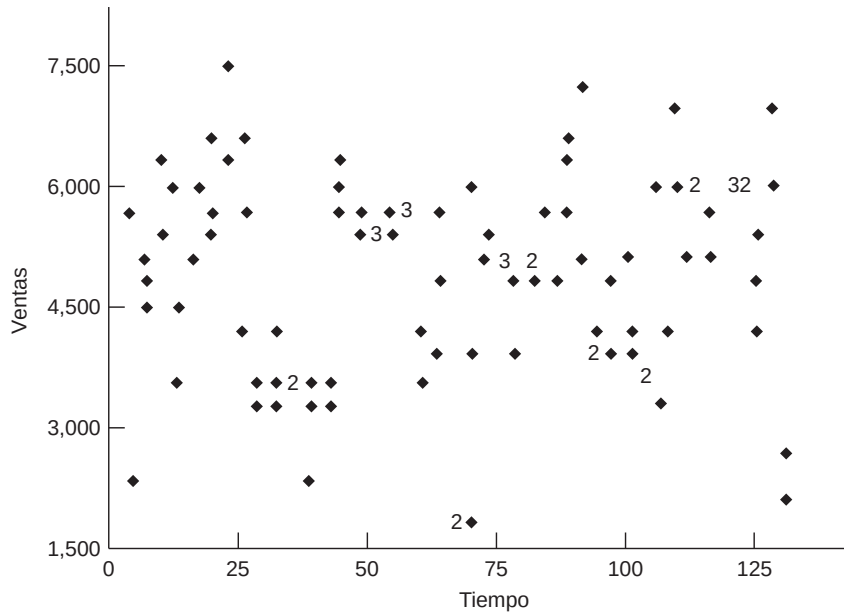


FIGURA 8-13 Ventas de un restaurante, enero de 1981 a diciembre de 1982

CASO 8-3 VENTAS DE UN RESTAURANTE²²

Jim Price, en el momento en que estaba terminando sus estudios de maestría en administración, trabajaba en un pequeño restaurante cerca de la Universidad Marquette en Milwaukee, Wisconsin. Un día la gerente del restaurante le pidió a Jim que se presentara en su oficina. Le dijo que estaba muy interesada en el pronóstico semanal de las ventas y quería saber si Jim podía ayudarla. Puesto que Jim había tomado justamente un curso de estadística avanzada, le dijo que le encantaría el reto.

Jim pidió a la gerente del restaurante que le proporcionara todos los registros históricos que tuviera disponibles. Ella le indicó que el restaurante compilaba las ventas de la semana previa todos los lunes por la mañana. Jim inició su análisis obteniendo los datos de las ventas semanales desde el domingo de la semana del 1 de enero de 1981, al último domingo de la semana del 29 de diciembre de 1982: un total de 104 observaciones.

Las ventas promedio semanales de las 104 semanas fueron de \$4,862. La figura 8-13 es una gráfica de las

ventas semanales en el tiempo. La gráfica indica que las ventas semanales fueron bastante volátiles, abarcando desde \$1,870 a \$7,548, con muy poca tendencia. Como Jim había completado recientemente un curso de análisis de regresión, decidió usar las ventas semanales como la variable dependiente y ver si podía encontrar algunas variables independientes o variables explicativas.

Jim probó tres variables explicativas. La primera era el tiempo. La segunda era una variable ficticia que indicaba si la Universidad Marquette tenía cupo completo esa semana (0 = cupo incompleto, 1 = cupo completo). El examen de los datos de ventas de la figura 8-13 reveló que las ventas semanales siempre bajaban cuando Marquette no tenía el cupo completo, es decir, durante las vacaciones de Navidad, las vacaciones de primavera y las vacaciones de verano. Jim no estaba sorprendido, puesto que el restaurante está localizado cerca del campus de la universidad Marquette y la mayoría de sus clientes son miembros de la comunidad universitaria. Jim trató de que la tercera variable

²²Frank G. Forst, de la Marquette University, Milwaukee, Wisconsin, contribuyó con este caso.

TABLA 8-15 Matriz de correlación de las ventas del restaurante

	<i>Ventas actuales</i>	<i>Tiempo</i>	<i>Variable ficticia</i>	<i>Ventas retrasadas una semana</i>
Ventas actuales	1.000	.049	.772	.580
Tiempo		1.000	-.048	.120
Variable ficticia			1.000	.490
Ventas retrasadas una semana				1.000

TABLA 8-16 Modelos de regresión del restaurante

<i>Variable(s) explicativa(s) del modelo</i>	<i>R²</i>	<i>Estadístico de Durbin-Watson</i>	<i>¿Autocorrelación significativa al nivel .05?</i>	<i>Cantidad de colinealidad</i>
(1) Tiempo	.0024	.81	Sí	Ninguna
(2) Ficticia	.5960	1.30	Sí	Ninguna
(3) Ventas retrasadas una semana	.3360	N.A.*	No	Ninguna
(4) Tiempo y variable ficticia	.6030	1.32	Yes	Muy poca
(5) Tiempo y ventas retrasadas una semana	.3360	N.A.*	No	Poca
(6) Variable ficticia y ventas retrasadas una semana	.6490	N.A.*	No	Poca
(7) Tiempo, variable ficticia y ventas retrasadas una semana	.6510	N.A.*	No	Moderada

*N.A. indica que el estadístico de Durbin-Watson no es aplicable porque no se puede usar con una variable dependiente retrasada.

explicativa fuera la de las ventas retrasadas una semana, puesto que el examen de la figura 8-13 indicaba que las ventas de dos semanas consecutivas a menudo eran similares.

Jim calculó las correlaciones simples entre las tres variables explicativas potenciales y la variable dependiente, las ventas semanales. Los resultados se presentan en la matriz de correlación de la tabla 8-15. Como Jim esperaba, casi no había tendencia en las ventas semanales, según indicaba el coeficiente de correlación de .049. Sin embargo, la variable ficticia estaba fuertemente correlacionada con las ventas actuales, $r = .772$; en otras palabras, si la Universidad Marquette tenía o no cupo completo era una buena variable explicativa potencial de las ventas semanales actuales. Las ventas semanales actuales estaban moderadamente relacionadas, $r = .580$, con las ventas retrasadas una

semana. Sin embargo, Jim también notó que la variable ficticia estaba moderadamente relacionada con, $r = .490$, con las ventas retrasadas una semana.

Jim experimentó con varios modelos de regresión. Los resultados de los diferentes modelos de regresión se presentan en la tabla 8-16. Puesto que los datos de ventas casi no tienen tendencia, el factor “tiempo” agrega muy poco poder de predicción al modelo de regresión. Observe que el modelo 4 tiene una R^2 sólo 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 sólo ligeramente mayor que el modelo 6. Por otro lado, el factor “ventas retrasadas una semana” agrega una buena cantidad de poder de explicación al modelo de regresión. Finalmente, el

modelo 6 tiene una R^2 más alta que el modelo 2, sin una cantidad significativa de autocorrelación residual.

Jim seleccionó el modelo de regresión 6 para pronosticar las ventas semanales por las siguientes razones:

1. El modelo 6 tenía la segunda R^2 , más grande, sólo .002 más abajo del modelo 7.
2. Los parámetros del modelo 6 eran, cada uno, significativamente diferentes de cero al nivel .01.
3. La autocorrelación no es un problema para el modelo 6.
4. El modelo 6 es más sencillo que el modelo 7 y no tiene tanta multicolinealidad.

La función de regresión ajustada que Jim usó fue

$$\hat{Y} = 2,614.3 + 1,610.7(\text{variable ficticia}) + .2605(\text{ventas retrasadas una semana})$$

La R^2 significa que el 64.9% de la variación en las ventas semanales se puede explicar por el hecho de si la Universidad Marquette tenía o no cupo completo y también a partir de las ventas de la semana anterior. La ecuación de regresión implica que el promedio de ventas semanales es de alrededor de \$1,611 más alto cuando la universidad tiene cupo completo, manteniendo constantes las ventas de la semana anterior.

Jim estaba contento con su trabajo, pero se preguntaba si otro tipo de modelo de pronóstico no sería más efectivo. Por esta razón, decidió tomar un curso de elaboración de pronósticos.

PREGUNTAS

1. ¿Era correcto el uso de una variable ficticia por parte de Jim?
2. ¿Es correcto que Jim use ventas retrasadas una semana como variable explicativa?
3. ¿Está de acuerdo con las conclusiones de Jim?
4. ¿Habría otro tipo de modelo de pronóstico más efectivo para pronosticar las ventas semanales?

CASO 8-4 MR. TUX

John Mosby corrió un análisis de regresión simple usando el tiempo como la variable explicativa y observó una decepcionante r^2 de .563 (véase el caso 6-4). Como sabía que sus datos de ventas tenían un fuerte componente estacional, decidió ajustar un modelo de regresión múltiple estacional.

John creó 11 variables ficticias para el efecto estacional. Puesto que los datos de Mr. Tux son mensuales, codificó S1 como 1 si los datos eran de enero, y como 0 si no lo eran. John hizo esto para cada mes, terminando con S11, donde asignaba 1 para noviembre y 0 si no lo era.

John corrió el modelo usando 12 variables explicativas: 1 para el tiempo y las otras 11 representando el efecto estacional mensual. Los resultados se presentan en la tabla 8-17.

John calculó manualmente el *MAPE* para los últimos 12 meses de datos y encontró que era del 21.25%. Él no está seguro si tiene un buen modelo. El modelo explica

el 87.7% de la varianza de la variable de ventas. Sin embargo, un *MAPE* por encima del 20% parece alto. En seguida, John generó las autocorrelaciones de los residuos del modelo, las cuales se presentan en la figura 8-14.

Las autocorrelaciones residuales tienen un pico en el retraso 12, el retraso estacional. Un tanto insatisfecho con los resultados de este modelo de regresión estacional, John decidió intentar con un modelo autorregresivo para pronosticar sus ventas mensuales (los datos aparecen en el caso 2-2).

En virtud del fuerte componente estacional de sus datos, John decidió modelar este componente usando un modelo autorregresivo con los valores de Y retrasados 12 meses. El hecho de retrasar los datos 12 periodos deja a John con un tamaño muestral de 84. El resultado de Minitab se presenta en la tabla 8-18. La función de autocorrelación de los residuos del modelo autorregresivo se presenta en la figura 8-15.

TABLA 8-17 Resultado de Minitab para el modelo de regresión estacional de Mr. Tux

Regression Analysis: Sales versus Time, S1, ...

The regression equation is

$$\begin{aligned} \text{Sales} = & -35023 + 2752 \text{ Time} - 48459 \text{ S1} - 29808 \text{ S2} + 21681 \text{ S3} + 119019 \text{ S4} \\ & + 139212 \text{ S5} + 57713 \text{ S6} + 21689 \text{ S7} + 74014 \text{ S8} + 7872 \text{ S9} \\ & - 9009 \text{ S10} - 25050 \text{ S11} \end{aligned}$$

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

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	0.260241	2.55	6.71
2	-0.039905	-0.37	6.87
3	-0.091961	-0.84	7.72
4	-0.030178	-0.28	7.81
5	-0.266888	-2.43	15.18
6	-0.241385	-2.07	21.27
7	-0.258015	-2.12	28.31
8	0.074975	0.59	28.91
9	0.034480	0.27	29.04
10	0.051103	0.40	29.32
11	0.229282	1.79	35.14
12	0.443214	3.36	57.14
13	0.040709	0.28	57.33
14	-0.060195	-0.41	57.75
15	-0.129158	-0.88	59.68
16	-0.006461	-0.04	59.69
17	-0.044247	-0.30	59.92
18	-0.150021	-1.01	62.64
19	-0.136353	-0.91	64.91
20	0.048747	0.32	65.20
21	-0.006104	-0.04	65.21
22	-0.022500	-0.15	65.27
23	0.208753	1.38	70.89
24	0.079236	0.51	71.71

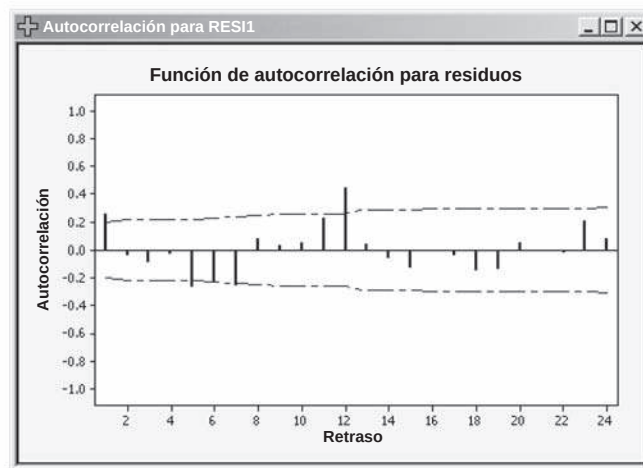


FIGURA 8-14 Función de autocorrelación para los residuos del modelo de regresión estacional de Mr. Tux

TABLA 8-18 Resultado de Minitab para el modelo autoregresivo estacional de Mr. Tux**Regression Analysis: Sales versus Lag12Sales**

The regression equation is

$$\text{Sales} = 24786 + 1.07 \text{ Lag12Sales}$$

84 cases used, 12 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	24786	5322	4.66	0.000
Lag12Sales	1.06807	0.03803	28.08	0.000

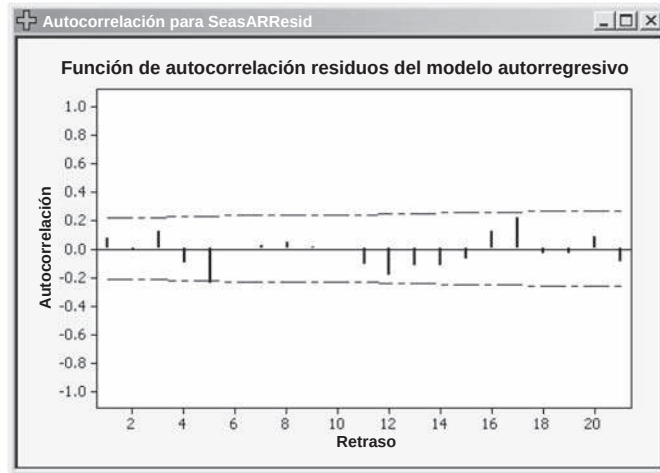
S = 30784.9 R - Sq = 90.6% R - Sq(adj) = 90.5%

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			

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	0.075468	0.69	0.50
2	-0.011101	-0.10	0.51
3	0.117958	1.07	1.75
4	-0.099816	-0.90	2.65
5	-0.245680	-2.19	8.17
6	0.000820	0.01	8.17
7	0.020569	0.17	8.21
8	0.046401	0.39	8.41
9	0.018403	0.15	8.44
10	0.005474	0.05	8.45
11	-0.110622	-0.93	9.66
12	-0.188329	-1.57	13.22
13	-0.120674	-0.98	14.70
14	-0.118386	-0.95	16.14
15	-0.068672	-0.54	16.64
16	0.115044	0.91	18.04
17	0.214570	1.68	23.01
18	-0.034505	-0.26	23.14
19	-0.034158	-0.26	23.27
20	0.079940	0.60	23.99
21	-0.092343	-0.69	24.97

**FIGURA 8-15 Función de autocorrelación para los residuos del modelo autorregresivo de Mr. Tux****TAREA**

1. Redacte un escrito para John con un análisis cuidadoso de los resultados de sus dos intentos para desarrollar un modelo de pronóstico estacional. ¿Cuál modelo es mejor? Asegúrese de que su explicación incluya una evaluación del ajuste del modelo, la exactitud potencial del pronóstico y cualquier problema remanente (por ejemplo, la autocorrelación).

CASO 8-5 CONSUMER CREDIT COUNSELING

La operación del Consumer Credit Counseling (CCC) se describió en los casos 1-2 y 3-3.

El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que necesitaba CCC para pronosticar era el número de clientes nuevos que serían atendidos durante el resto de 1993. Marv le proporcionó a Dorothy Mercer datos mensuales del número de clientes nuevos atendidos por CCC en el periodo de enero de 1985 a marzo de 1993 (véase el caso 3-3). En el caso 3-3, Dorothy usó el análisis de autocorrelación para explorar el patrón de datos. En el caso 6-5, utilizó tanto el número de personas con cupones canjeables por alimentos como un índice de actividad de los negocios para desarrollar un modelo de regresión que le permitiera pronosticar el resto de 1993.

Dorothy no estaba contenta con los resultados de su modelo de regresión. Decidió intentar con la regresión múltiple y le pidió a Marv que pensara en otras variables que podrían estar relacionadas con el número de clientes atendidos. Marv le indicó que podría intentar con el número de bancarrotas registradas y el número de permisos de construcción expedidos.

Los datos de las bancarrotas registradas de enero de 1986 a diciembre de 1992 se presentan en la tabla

8-19, y los datos de los permisos de construcción expedidos de enero de 1986 a diciembre de 1992 aparecen en la tabla 8-20.

Dorothy desarrolló un modelo de regresión múltiple que usaba el número de gente con cupones canjeables por alimentos, el índice de actividad en los negocios, el número de bancarrotas registradas, y el número de permisos de construcción expedidos. También elaboró un modelo con base únicamente en el supuesto de que los datos eran estacionales.

Dorothy se enteró recientemente de que la correlación serial a menudo es un problema cuando la regresión se ejecuta con datos de series de tiempo. Se preocupó de que alguno de los modelos desarrollados para predecir el número de clientes atendidos estuviera afectado por esta condición. Como le gustó el reporte que usted le entregó en el que usaba el modelo de análisis de la descomposición de las series de tiempo, ella le asignó a usted la tarea de verificar esta situación.

Usted recordaba vagamente haber desarrollado un modelo autorregresivo en algún momento de su experiencia estudiantil y le preguntó a Dorothy si le gustaría que usted analizara esa posibilidad. Ella determinó que sería un buen modo de que usted empleara su tiempo.

TABLA 8-19 Bancarrotas registradas, enero de 1986 a diciembre de 1992

	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>Mayo</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
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 Permisos de construcción expedidos, enero de 1986 a diciembre de 1992

	<i>Ene.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Abr.</i>	<i>Mayo</i>	<i>Jun.</i>	<i>Jul.</i>	<i>Ago.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dic.</i>
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	352	387	243	140

TAREA

1. 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 es necesario), y utilícelo para pronosticar el número de clientes nuevos para los primeros tres meses de 1993. Compare sus pronósticos con las observaciones reales.
2. Desarrolle un modelo autorregresivo y genere pronósticos para los primeros tres meses de 1993. ¿Cuál

modelo (regresión múltiple o autorregresión) cree usted que es el mejor candidato para generar pronósticos para el resto de 1993? Prepare un escrito a Dorothy en el que le dé la información solicitada en relación con el problema de correlación serial. Incluya un análisis de los resultados de sus esfuerzos para desarrollar un modelo adecuado para pronosticar el número de clientes nuevos durante el resto de 1993.

CASO 8-6 AAA WASHINGTON²³

Se presentó una visión general de AAA Washington en el caso 5-5, cuando se pidió a los estudiantes que prepararan la descomposición de una serie de tiempo de las llamadas de emergencia en el camino recibidas durante cinco años. La descomposición de la serie de tiempo ejecutada en el caso 5-5 mostró que el patrón que Michael DeCoria había observado en el volumen de llamadas del servicio de emergencia en el camino probablemente era cíclico en su naturaleza. Michael quería predecir el efecto cíclico sobre el volumen de llamadas de emergencia al servicio para años futuros.

Otra investigación realizada por el club identificó varios factores que tenían un efecto sobre el volumen de llamadas de emergencia al servicio. Entre estos factores están la temperatura promedio diaria y la cantidad de lluvia registrada en un día. Esta investigación reveló que las llamadas al servicio de emergencia aumentan conforme se incrementa la lluvia y conforme el promedio de la temperatura diaria disminuye. El club también cree que el número total de llamadas que recibe el servicio de emergencia es dependiente del número de miembros en el club.

Michael había observado que la tendencia cíclica de la serie de tiempo parecía estar atrasada en relación con el ciclo general de la economía. Él sugirió que la tasa de desempleo para el estado de Washington sería una buena medición sustituta para el estado general de la economía de Washington. Los datos sobre la temperatura promedio mensual, la precipitación mensual y la

tasa de desempleo en el estado de Washington se reunieron en la tabla 8-21. Una conversación con el gerente del centro del servicio de emergencia trajo consigo dos observaciones importantes: **1.** Los fabricantes de automóviles parecen diseñar los vehículos para que operen mejor a 65 grados Fahrenheit y **2.** el volumen de llamadas parece incrementarse más drásticamente cuando la temperatura promedio cae unos cuantos grados a partir de una temperatura promedio en el rango de los 30 grados de lo que aumenta cuando ocurre una caída similar a partir de una temperatura promedio en el rango de los 60 grados. Esta información sugería que el efecto de la temperatura sobre el servicio de emergencia era no lineal.

En el caso 6-6, se investigaron cuatro modelos de regresión lineal que consideraban el número total de llamadas al servicio como la variable dependiente, y la tasa de desempleo, la temperatura, la precipitación y el número de miembros como las cuatro variables independientes. La variable de la temperatura se transformó sustrayendo 65 grados de los valores mensuales promedio de temperatura. Luego se investigó una relación no lineal.

En el caso 7-2 se desarrolló un modelo de regresión múltiple. Se sometieron a prueba variables como la precipitación, el número de miembros, la temperatura mensual promedio transformada exponencialmente y la tasa de desempleo retrasada 11 meses.

²³Este caso fue proporcionado por Steve Branton, ex alumno y graduado de la maestría en administración, en la Universidad del Este de Washington.

TABLA 8-21 Datos de AAA Washington						
Año	Mes	Llamadas	Tasa de desempleo	Temperatura	Precipitación	Miembros
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	—
	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
1989	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
	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
1990	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
	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
1991	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

TABLA 8-21 (continuación)

<i>Año</i>	<i>Mes</i>	<i>Llamadas</i>	<i>Tasa de desempleo</i>	<i>Temperatura</i>	<i>Precipitación</i>	<i>Miembros</i>
1992	Mayo	19,077	5.8146	54.3	2.12	450,001
	Junio	19,141	5.9480	58.2	1.61	452,303
	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
	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

PREGUNTAS

- Con base en los datos de la tabla 8-21 y las variables independientes descritas en el caso 7-2, ¿cuál modelo de regresión es el mejor para predecir? ¿Los signos de los coeficientes de las variables independientes son los que usted esperaría que fueran? ¿Los coeficientes de las variables independientes son significativamente diferentes de cero?
- ¿La correlación serial es un problema en este caso? Si algunos coeficientes no son significativamente diferentes de cero, intente correr una regresión sin estas variables independientes. Experimente con diferentes potencias para la transformación exponencial.
- Redacte un escrito para Michael recomendando el modelo de regresión que usted cree es el más apropiado para predecir la naturaleza cíclica del volumen de llamadas del servicio de emergencia en el camino.

CASO 8-7 ALOMEGA FOOD STORES

En el ejemplo 1.1 Julie Ruth recopiló los datos de las ventas mensuales de su compañía junto con muchas otras variables que pensó podrían estar relacionadas con las ventas. Sus esfuerzos para pronosticar usando estos datos con Minitab se detallaron en los casos 2-3, 3-4 y 5-6. Su deseo de mejorar la exactitud de sus pronósticos sugería el uso de la regresión múltiple.

Julie reconoció en su análisis de descomposición de las ventas que sus datos tenían un fuerte componente estacional (véase el caso 5-6). Como también pensaba que las ventas estaban influidas por las cifras gastadas en publicidad en el periódico y en televisión (véase el caso 2-3), decidió correr una regresión relacionando las ventas con la publicidad en el periódico

(Papers), publicidad en TV (TV) y 11 variables ficticias 0-1, utilizando diciembre como el mes base, para explicar los efectos estacionales mensuales. Una parte del resultado de Minitab de la regresión de Julie se presenta en la tabla 8-22.

Julie estaba muy contenta de ver una R^2 con un valor de aproximadamente 91%, una mejora significativa en relación con el valor de la r^2 del modelo de regresión con una sola variable explicativa TV (revise los resultados de la regresión en el caso 2-3). Más aún, el factor de inflación de la varianza (VIF) de cada una de las variables explicativas era pequeño, y el estadístico de Durbin-Watson era cercano a 2.

TABLA 8-22 Resultado de Minitab para el modelo de regresión de Julie Ruth de 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 = 2.27

Julie se dio cuenta de que era necesario un examen de los residuos antes de estar completamente satisfecha con sus resultados e intentó hacer un análisis residual completo antes de proponer su modelo de regresión como una herramienta útil para pronosticar las ventas.

A Julie le gustaba el hecho de que las variables explicativas Papers y TV estuvieran bajo el control de

la compañía. Los valores de estas variables podrían entonces seleccionarse por adelantado para generar pronósticos de las ventas futuras.

Julie estaba casi lista para confrontar a su torturador, Jackson Tilson, pero aún tenía que unir algunos cabos sueltos.

PREGUNTAS

1. Julie recopiló datos sobre otras variables que no estaban incluidas en su modelo de regresión múltiple. ¿Debería incluirse una o más de estas variables en su modelo? Más generalmente, ¿cómo puede asegurarse Julie de que tiene el conjunto “correcto” de variables explicativas.
2. Suponiendo que no existen variables explicativas importantes, ¿está usted satisfecho con el modelo de pronóstico de Julie? ¿Cómo podría usted “vender” el modelo a la alta dirección (y a Jackson Tilson)?
3. ¿Cómo podría usarse el modelo de Julie para determinar cifras futuras de gasto en publicidad en periódicos y televisión?
4. ¿Qué condiciones podrían motivar a Julie a reexaminar su modelo de regresión o, quizás, a buscar otro método para pronosticar las ventas?

CASO 8-8 SURTIDO COOKIES

Los esfuerzos de Jaime Luna por comprender la tendencia y estacionalidad en las ventas mensuales de Surtido Cookies se examinaron en los casos 3-5, 4-8 y 5-7. El hecho de que las ventas de galletas sean estacionales es ahora bien conocido. Jaime y su equipo han intentado un procedimiento de suavización y descomposición, y ahora están interesados en tratar con métodos de regresión en su intento permanente por obtener un procedimiento que genere los “mejores” pronósticos de las ventas en el futuro. Jaime no está seguro cómo formular el modelo (o modelos) de regresión. No ha identificado las variables explicativas potenciales para las ventas, pero piensa que no todo está perdido porque el patrón estacional está bien identificado, y un modelo con variables ficticias que representen los diferentes meses podría funcionar. Además, las ventas anteriores a enero pueden ser útiles para pronosticar las ventas actuales de enero, las ventas anteriores a

febrero servirán para pronosticar las ventas actuales de febrero, etcétera.

Al codificar el primer mes como 1, el segundo mes como 2, y así sucesivamente, Jaime define una variable con tendencia en el tiempo. Luego construye un conjunto de variables ficticias con diciembre como el mes base. Finalmente, Jaime ajusta un modelo de regresión con tendencia en el tiempo y variables ficticias mensuales para sus datos de ventas (véase la tabla 3-12). El resultado de Minitab se presenta en la tabla 8-23. Jan es la variable ficticia que representa las ventas de enero, con una definición similar para los meses restantes. Después de examinar los residuos, Jaime está bastante contento con los resultados de su análisis de regresión. Está listo para considerar otro modelo de regresión, usando las ventas retrasadas un número adecuado de periodos como una variable explicativa, pero necesita ayuda.

PREGUNTAS

1. Todos los coeficientes de las variables ficticias en la regresión de Jaime son negativos excepto el de Nov. ¿Tiene esto sentido? Explique.
2. ¿Está usted satisfecho con el modelo de regresión de Jaime? ¿Qué cambios haría usted, en caso necesario?
3. Con base en el modelo ajustado de Jaime de la tabla 8-23, genere pronósticos para los siete meses restantes de 2003.
4. Ajuste un modelo autorregresivo para los datos de Jaime con ventas retrasadas 12 meses como variable explicativa. ¿Este modelo es razonable? Genere

pronósticos para los siete meses restantes de 2003 usando su modelo autorregresivo.

5. ¿Cuál modelo prefiere, el de la regresión con variable ficticia o el de autorregresión? ¿Por qué?

TABLA 8-23 Resultado de Minitab para el modelo de regresión de Jaime Luna para las ventas de Surtido Cookies

Regression Analysis: SurtidoSales versus Time, Jan, ...

The regression equation is

$$\text{SurtidoSales} = 1703762 + 4759 \text{ Time} - 1015200 \text{ Jan} - 1259183 \text{ Feb} - 1286031 \text{ Mar} - 1126296 \text{ Apr} - 1188322 \text{ May} - 1170518 \text{ Jun} - 1108576 \text{ Jul} - 1176156 \text{ Aug} - 355847 \text{ Sep} - 178199 \text{ Oct} + 225028 \text{ Nov}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	1703762	107455	15.86	0.000	
Time	4759	2172	2.19	0.037	1.0
Jan	-1015200	124780	-8.14	0.000	2.1
Feb	-1259183	124610	-10.11	0.000	2.1
Mar	-1286031	124477	-10.33	0.000	2.1
Apr	-1126296	124383	-9.06	0.000	2.1
May	-1188322	124326	-9.56	0.000	2.1
Jun	-1170518	133527	-8.77	0.000	1.9
Jul	-1108576	133332	-8.31	0.000	1.9
Aug	-1176156	133173	-8.83	0.000	1.9
Sep	-355847	133049	-2.67	0.012	1.9
Oct	-178199	132960	-1.34	0.191	1.9
Nov	225028	132907	1.69	0.102	1.9

S = 162756 R - Sq = 93.8% R - Sq(adj) = 91.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	1.11783E + 13	9.31528E + 11	35.17	0.000
Residual Error	28	7.41703E + 11	26489395582		
Total	40	1.19200E + 13			

Durbin-Watson statistic = 1.50

CASO 8-9 SOUTHWEST MEDICAL CENTER

Mary Beasley había cubierto una gran parte del análisis de regresión en su programa ejecutivo de maestría en administración y pensó que un modelo de regresión podría representar bien sus datos. Ella sabía que el total de visitas facturables al Medical Oncology tenía una tendencia ascendente y quizá un componente estacional. Mary estaba familiarizada con el uso de variables ficticias para representar la estacionalidad en un contexto de regresión. Como los datos mensuales de Mary estaban organizados sobre la base del año fiscal y

comenzaban en septiembre FY 1994-1995 (véase la tabla 4-13), creó 11 variables ficticias, usando agosto como el mes base. Luego corrió una regresión con las visitas facturables como la variable dependiente y las 11 variables ficticias más el tiempo (codificadas como 1, 2, 3(coded 1, 2, 3, ...)) como las variables independientes. Después de examinar su resultado de Minitab, incluyendo un análisis de los residuos, Mary estaba decepcionada. Quizás la regresión no era el camino a seguir.

PREGUNTAS

1. Repita el análisis de regresión de Mary. ¿Tiene razón Mary para estar decepcionada? Explique.
2. Escriba un breve reporte para Mary describiendo cómo puede mejorar su modelo de regresión original y/o sugiérale otro modelo de regresión que podría intentar.

Aplicaciones de Minitab

El problema. En el ejemplo 8.8 James Brown intenta pronosticar el consumo doméstico de electricidad para el tercero y cuarto trimestres de 1996 para la Washington Water Power.

Solución de Minitab

1. Introduzca la variable Hours de la tabla 8-9 en la columna C1. Introduzca Time en la columna C2 (1, 2, 3, ..., 66). Introduzca las variables ficticias mostradas en la tabla 8-10 en las columnas C3, C4, y C5.
2. Para correr el modelo de regresión estacional, haga clic en los siguientes menús:
Stat>Regression>Regression
3. Aparecerá una ventana de diálogo Regression como la que se presenta en la figura 6-18
 - a) Se selecciona Hours como la respuesta (Response) o variable dependiente.
 - b) Se registra Time y 4th Qt (esto es, tiempo y 4° trimestre) como variables explicativas (Predictors) o independientes.
 - c) Haga clic en Options para obtener la ventana de diálogo Regression-Options.
4. Aparecerá la ventana de diálogo Regression-Options mostrada en la figura 8-16.

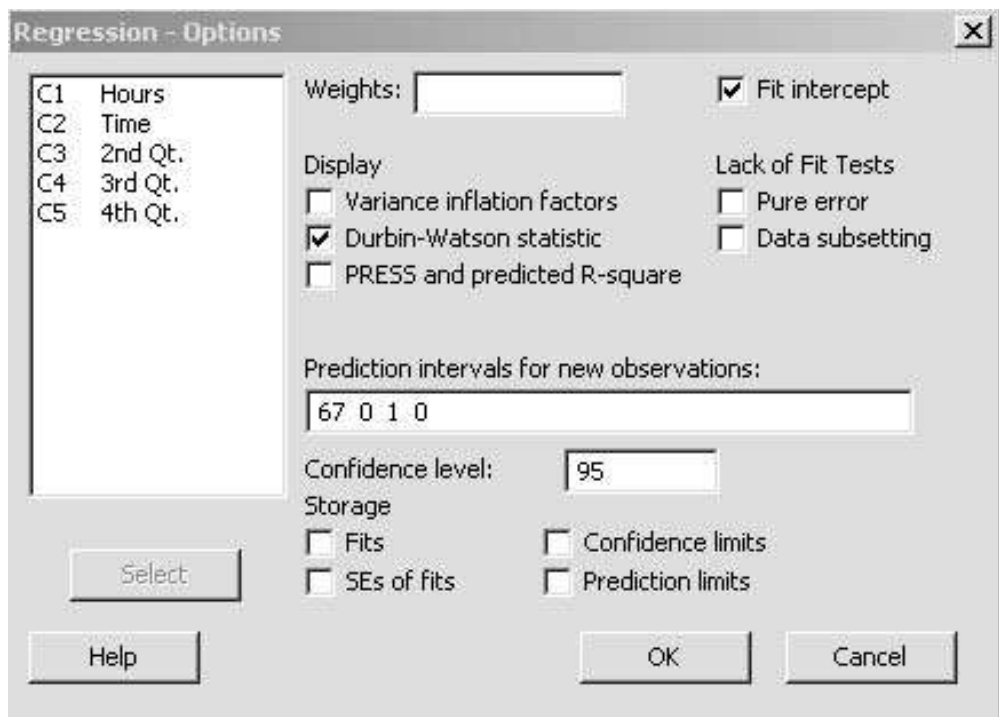


FIGURA 8-16 Ventana de diálogo de Regression-Options de Minitab

- a) Haga clic en la ventana del estadístico de Durbin-Watson.
- b) En el espacio debajo de Prediction intervals for new observations:, teclee las nuevas observaciones del tercer trimestre: 67 0 1 0.
- c) Haga clic en OK y luego en OK en la ventana 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 hay intercepción en Y). Para correr este tipo de modelo, haga clic para borrar la marca de verificación en la ventana Fit intercept mostrada en la figura 8-16.

Aplicaciones de Excel

El problema. En el ejemplo 8.6 se desarrolló un modelo autorregresivo de primer orden para los datos de venta de Novak Corporation.

Solución de Excel

1. Ingrese el archivo Minitab que contiene los datos presentados en la tabla 8-2 y resalte la columna de Sales. Haga clic en los siguientes menús:

Edit>Copy Cells

Ahora introduzca su hoja de cálculo Excel, resalte la celda A3 y haga clic en los siguientes menús:

Edit>Paste

Los datos de Sales aparecen en la columna A. Después de que el encabezado de esta columna se introduce en la primera fila, la hoja de cálculo se ve como la de la figura 8-17 sin la variable retrasada.

2. Para crear la variable Sales retrasada un periodo, posicione el ratón en A3 y resalte hasta A18. Haga clic en los siguientes menús:

Edit>Copy

Ahora resalte B4 y haga clic en

Edit>Paste

El resultado aparece en la figura 8-17.

3. Haga clic en los siguientes menús para correr el modelo autorregresivo:

Tools>Data Analysis

Aparecerá la ventana de diálogo Data Analysis. Seleccione Regression debajo de Analysis Tools. La ventana de diálogo Regression se presentó en la figura 6-22.

- a) Introduzca A4:A19 en el Input Y Range.
- b) Introduzca B4:B19 en el Input X Range
- c) Haga clic en el botón junto a New Worksheet Ply.
- d) Haga clic en OK, y aparecerá el resultado desplegado en la figura 8-18.

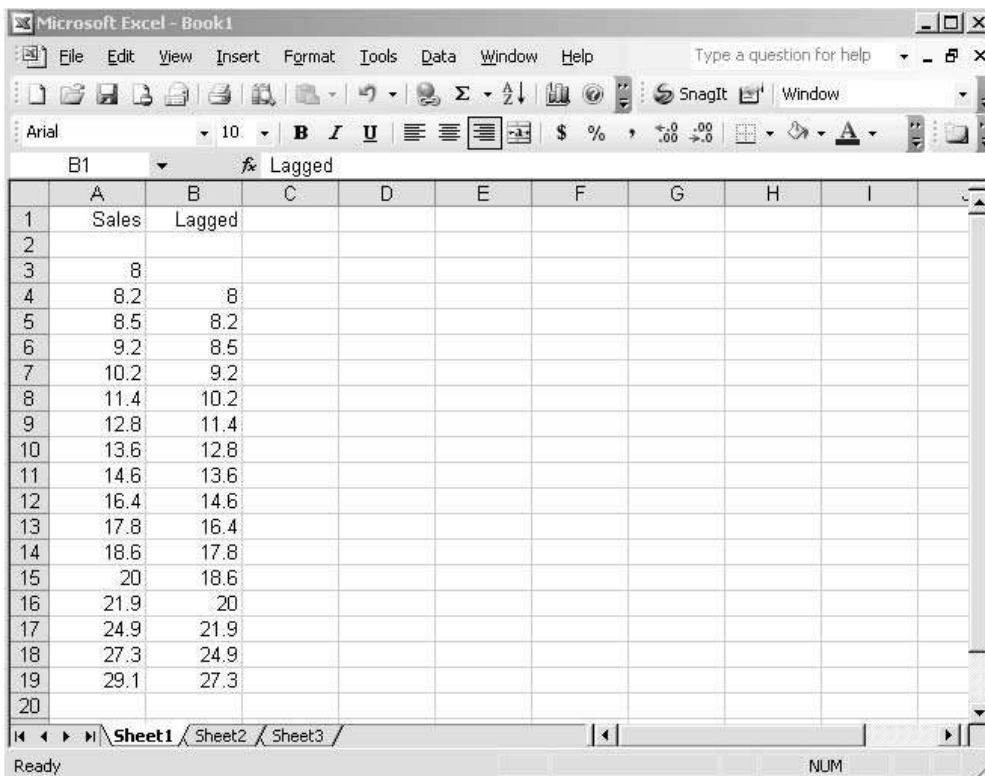


FIGURA 8-17 Hoja de cálculo de Excel después de retrasar la variable de ventas un periodo

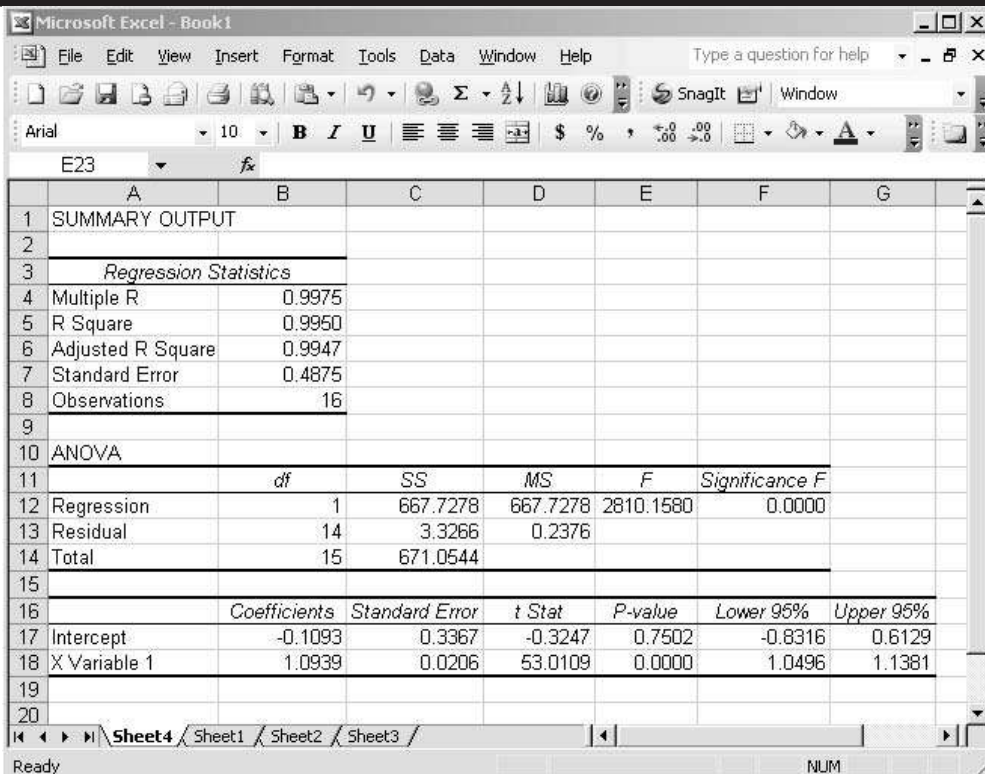


FIGURA 8-18 Resultado de Excel para un modelo de autorregresión

Referencias

- Diebold, F. X. *Elements of Forecasting*, 3a. ed. Cincinnati, Ohio: South-Western, 2004.
- Durbin, J. y G. S. Watson. "Testing for Serial Correlation in Least Squares Regression II." *Biometrika* 38 (1951): 159-178.
- Engle, R. F., y C. W. J. Granger. "Co-Integration and Error Correction: Representation, Estimation and Testing". *Econometrica* 55 (1987): 251-276.
- Granger, C. W. J., y P. Newbold. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2 (1974): 111-120.
- Levenbach, H. y J.P. Cleary. *Forecasting: Practice and Process for Demand Management*. Belmont, Calif.: Thomson Brooks/Cole, 2006.
- Murray, M.P. "A Drunk and Her Dog: An Illustration of Cointegration and Error Correction". *American Statistician* 48 (1) (1994): 37-39
- Newbold, P., y T. Bos. *Introductory Business and Economic Forecasting*, 2a. ed. Cincinnati, Ohio: South-Western, 1994.
- Pindyck, R.S. y D.L. Rubinfeld. *Econometric Models and Economic Forecasts*, 4a. ed. 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.

LA METODOLOGÍA
BOX-JENKINS (ARIMA)

Se han presentado varios métodos para el análisis y pronóstico de las series de tiempo. En el capítulo 4 se explicó la suavización, y los pronósticos de las series de tiempo se generaban por un mecanismo de suavización particular (promedio). En el capítulo 5 se descompusieron las series de tiempo en componentes de tendencia, estacional e irregular, y los pronósticos se generaron extrapolando la tendencia estimada y la estacionalidad. En el capítulo 8 se consideraron los modelos de regresión apropiados para las series de tiempo. Los pronósticos de la variable dependiente, Y , generados a partir de estos modelos por lo general requieren pronósticos de valores futuros de las variables independientes, las X .

Este capítulo presenta una clase de modelos que generan pronósticos exactos con base en una descripción de patrones históricos de los datos. Los modelos autorregresivos integrados de promedio móvil (ARIMA) son una clase de modelos lineales que son capaces de representar tanto series de tiempo *estacionarias* como *no estacionarias*. Recuerde que los procesos estacionarios varían alrededor de un nivel fijo, en tanto que los procesos no estacionarios carecen de un nivel promedio constante natural. Los modelos autorregresivos analizados en el capítulo 8 en realidad son un subconjunto especial de los modelos ARIMA que resultan útiles para modelar series de tiempo estacionarias.

Los modelos ARIMA no implican variables independientes en su construcción. En vez de ello, utilizan la información de la serie misma para generar los pronósticos. Por ejemplo, un modelo ARIMA para ventas mensuales proyectaría el patrón de ventas históricas para generar un pronóstico para las ventas del siguiente mes.

Los modelos ARIMA dependen mucho de los patrones de autocorrelación en los datos. La metodología para identificar, ajustar y verificar los modelos ARIMA adecuados fue impulsada fuertemente por dos especialistas en estadística, G. E. P. Box y G. M. Jenkins. Por esta razón, los modelos ARIMA para la generación de pronósticos se conocen a menudo como metodología Box-Jenkins.

METODOLOGÍA BOX-JENKINS

La metodología *Box-Jenkins* para generar pronósticos es diferente de la mayoría de los métodos porque no *supone* ningún patrón particular en los datos históricos de las series que se van a pronosticar. Se basa en un enfoque iterativo para identificar un modelo posible a partir de una clase general de modelos. Luego, el modelo seleccionado se coteja con los datos históricos para ver si describe la serie con exactitud. El modelo está bien ajustado si los residuos son generalmente pequeños, están distribuidos aleatoriamente y no contienen información útil. Si el modelo especificado *no* es satisfactorio, el proceso se repite usando un nuevo modelo diseñado para mejorar el original. Este procedimiento iterativo continúa hasta que se encuentra un modelo satisfactorio. En ese momento, el modelo se considera útil para pronosticar. La figura 9.1 ilustra la estrategia de construcción de un modelo Box-Jenkins.

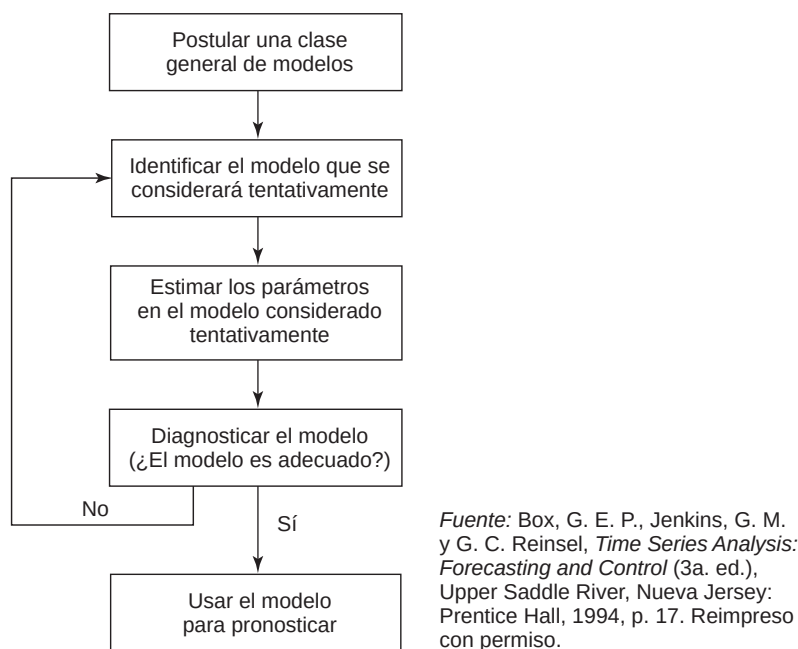


FIGURA 9-1 Diagrama de flujo de la estrategia de construcción de un modelo Box-Jenkins

La selección inicial de un modelo ARIMA se basa en examinar una gráfica de la serie de tiempo (para observar su carácter general) y en analizar sus autocorrelaciones para varios retrasos de tiempo. Específicamente, el patrón de las autocorrelaciones muestrales calculado a partir de la serie de tiempo se coteja con el patrón conocido de autocorrelación asociado con un modelo ARIMA particular. Este acoplamiento se hace tanto para las autocorrelaciones como para las autocorrelaciones parciales.¹ Los coeficientes de autocorrelación teórica de algunos de los modelos ARIMA más comunes se presentan en las figuras 9-2, 9-3 y 9-4. La importancia de los patrones presentados en estas figuras se hará evidente conforme avanza el capítulo.

La *metodología Box-Jenkins* se refiere a un conjunto de procedimientos para identificar, ajustar y verificar modelos ARIMA con los datos de la serie de tiempo. Los pronósticos se derivan directamente de la forma de un modelo ajustado.

Al seleccionar un modelo, recuerde que las autocorrelaciones calculadas de los datos no serán exactamente iguales a las autocorrelaciones teóricas asociadas con un modelo ARIMA. Las autocorrelaciones calculadas de los datos están sujetas a la variación de la muestra. Sin embargo, usted debe ser capaz de igualar adecuadamente la mayoría de los datos de la serie de tiempo con un modelo ARIMA. Si la selección inicial no es totalmente correcta, se presentarán deficiencias en el análisis de los residuos (verificación del modelo), y el modelo original podrá

¹Una autocorrelación parcial en el retraso de tiempo k es la correlación entre Y_t y Y_{t-k} (las respuestas para los periodos t y $t-k$, respectivamente) después de ajustar los efectos de los valores que intervienen, $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$. Newbold y Bos (1994), así como Box, Jenkins y Reinsel (1994) se encargaron de derivar y estudiar las autocorrelaciones parciales.

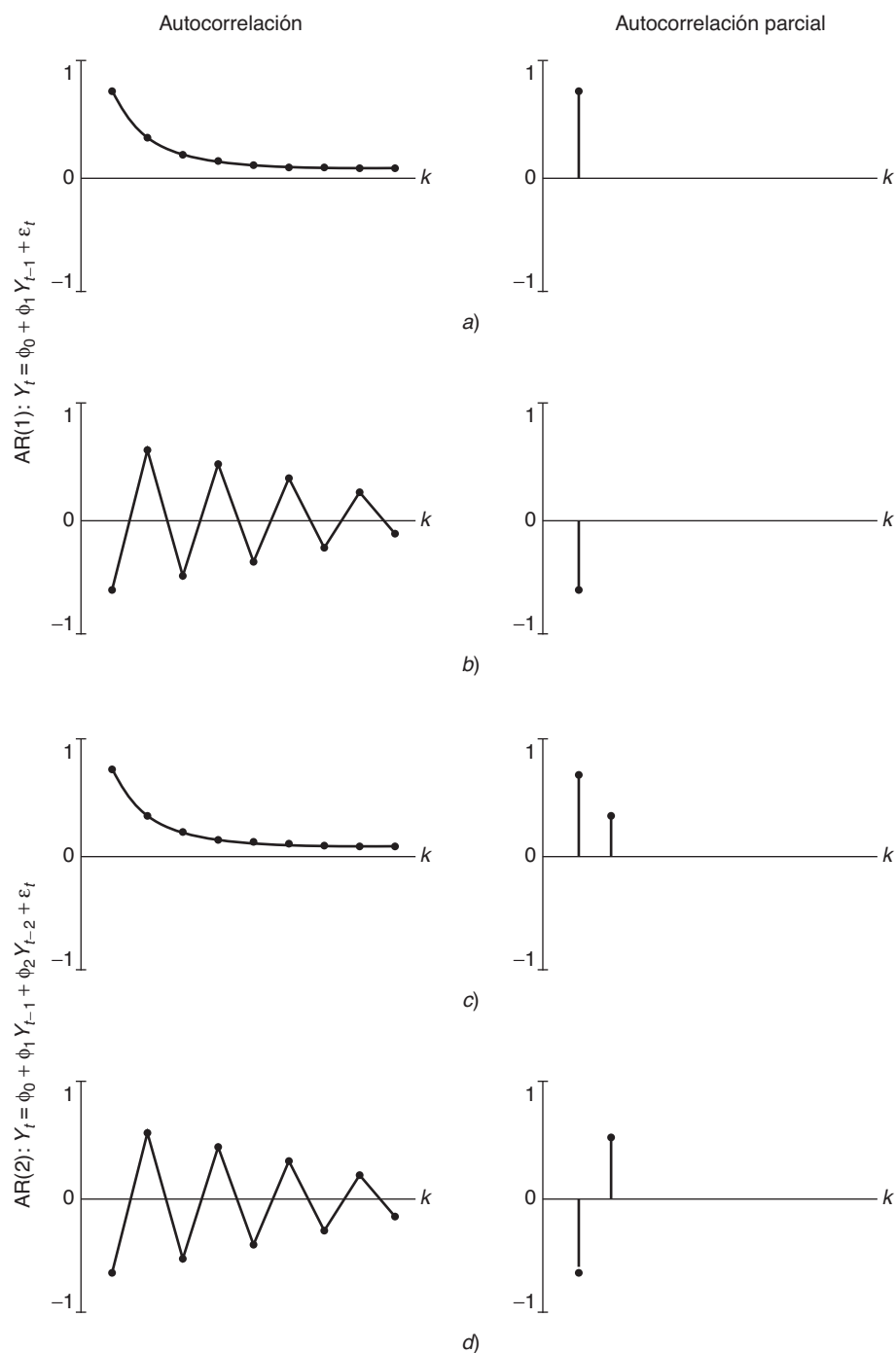


FIGURA 9-2 Coeficientes de autocorrelación y autocorrelación parcial de los modelos AR(1) y AR(2)

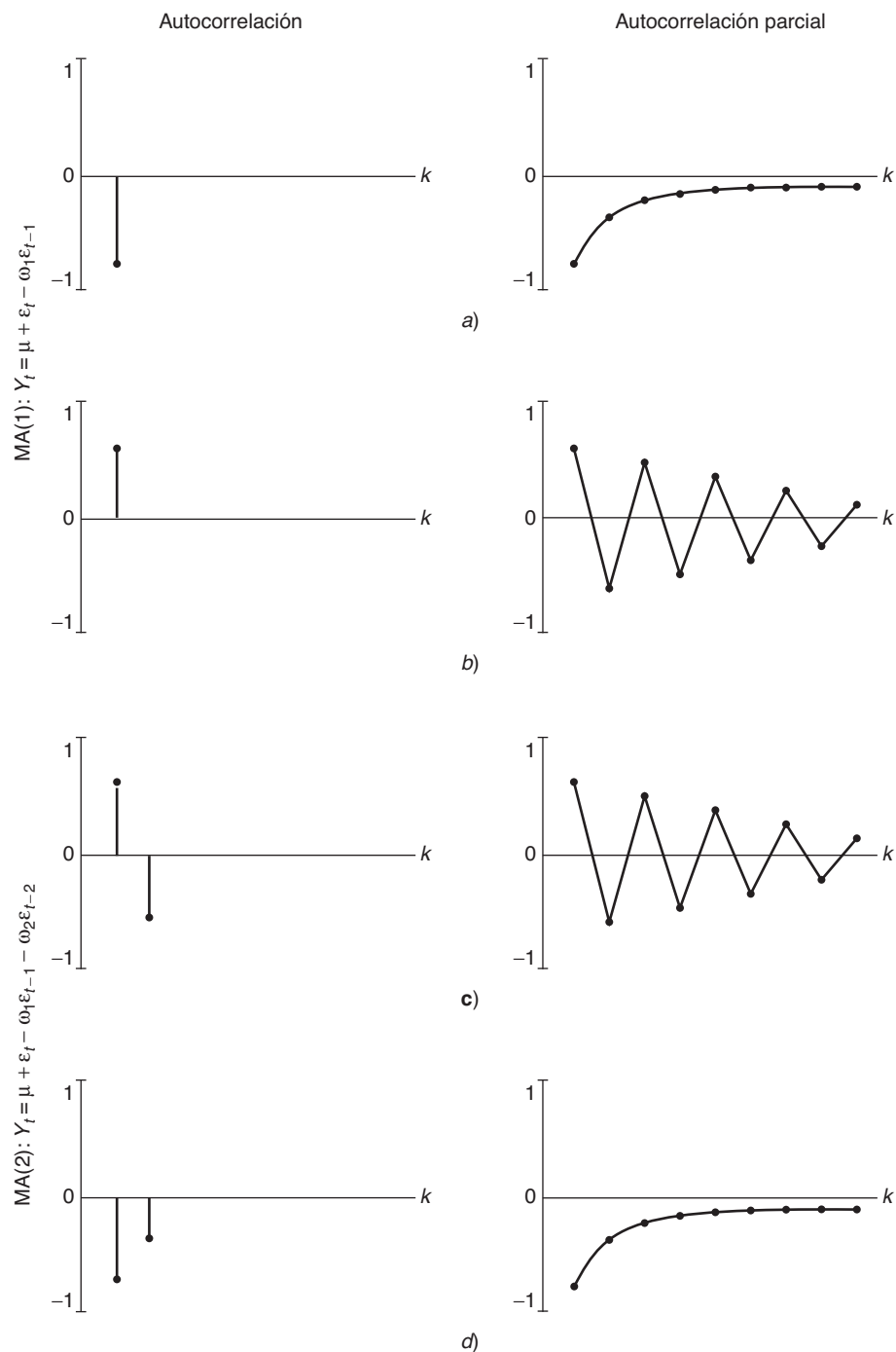


FIGURA 9-3 Coeficientes de autocorrelación y autocorrelación parcial de los modelos MA(1) y MA(2)

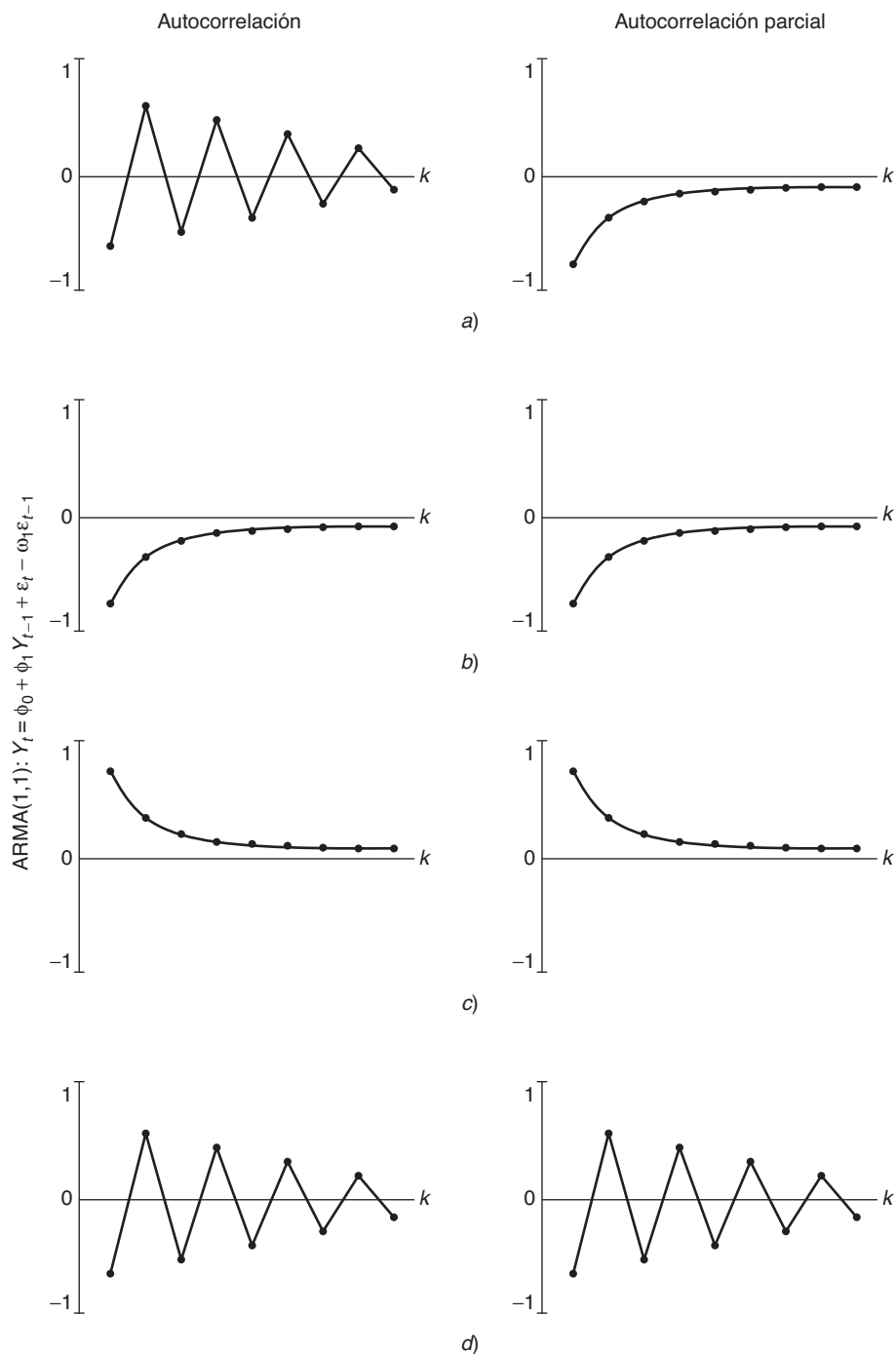


FIGURA 9-4 Coeficientes de autocorrelación y autocorrelación parcial de un modelo mixto ARMA(1, 1)

modificarse. Conforme usted adquiera experiencia, esta tarea iterativa de construcción de modelos se hará mucho más fácil.

Modelos autorregresivos

En el capítulo 8 se presentó un modelo autorregresivo de primer orden. Un modelo autorregresivo de orden p tiene la 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 respuesta (o dependiente) en el tiempo t

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ = variable de respuesta en los retrasos de tiempo $t-1, t-2, \dots, t-p$, respectivamente; estas Y desempeñan el papel de variables independientes

$\phi_0, \phi_1, \phi_2, \dots, \phi_p$ = coeficientes que serán estimados²

ε_t = término de error en el tiempo t , el cual representa los efectos de variables no explicadas por el modelo; los supuestos acerca del término de error son los mismos que los del modelo de regresión estándar

El modelo de la ecuación 9.1 tiene la apariencia de un modelo de regresión con valores retrasados de la variable dependiente en las posiciones de la variable independiente, de aquí el nombre de *modelo autorregresivo*. Los modelos autorregresivos son modelos adecuados para series de tiempo estacionarias, y el coeficiente ϕ_0 está relacionado 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 ϕ_0 .

La ecuación de un modelo AR de orden 1, o modelo AR(1), y de un modelo AR de orden 2, o modelo AR(2), se presenta en la figura 9-2. Las figuras 9-2a) y b) ilustran el comportamiento teórico de las funciones de autocorrelación y de autocorrelación parcial, respectivamente, de un modelo AR(1). Observe las diferencias en el comportamiento de las funciones de autocorrelación y autocorrelación parcial. Los coeficientes de autocorrelación se aproximan gradualmente hacia cero, mientras que los coeficientes de autocorrelación parcial caen a cero después del primer retraso de tiempo. Las figuras 9-2c) y d) muestran las autocorrelaciones y las autocorrelaciones parciales para un modelo AR(2). Otra vez, los coeficientes de autocorrelación tienden a cero, mientras que los coeficientes de autocorrelación parcial caen a cero después del segundo retraso de tiempo. Este tipo de patrón generalmente se mantendrá para cualquier modelo AR(p). Debe recordarse que las funciones de autocorrelación muestrales van a diferir de estas funciones teóricas a causa de las variaciones de la muestra.

Ejemplo 9.1

A continuación se realiza un pronóstico con un modelo AR(2) empleando un conjunto de datos de 75 lecturas de un proceso que se indican en la tabla 9-5. Sólo se usarán las últimas cinco observaciones de la tabla 9-5 para elaborar los pronósticos que se presentan en la tabla 9-1.

Se selecciona un modelo AR(2) para procesar los datos, y el programa Minitab calcula los mínimos cuadrados estimados $\hat{\phi}_0 = 115.2$, $\hat{\phi}_1 = -.535$ y $\hat{\phi}_2 = .0055$. Suponga que para el tiempo $t-1 = 75$ se requiere un pronóstico de la observación del siguiente periodo, $t = 76$. Puesto que el mejor pronóstico para el término de error es su valor promedio igual a cero, el pronóstico para el periodo $t = 76$ es

$$\begin{aligned} \hat{Y}_t &= \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2} \\ \hat{Y}_{76} &= 115.2 - .535 Y_{75} + .0055 Y_{74} \\ \hat{Y}_{76} &= 115.2 - .535(72) + .0055(99) = 77.2 \end{aligned}$$

²El coeficiente ϕ_0 está relacionado con la media del proceso, μ , mediante $\phi_0 = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$.

TABLA 9-1 Pronóstico con un modelo autorregresivo AR(2) para el ejemplo 9.1

<i>Periodo</i>	<i>Tiempo</i> <i>t</i>	<i>Valores</i> <i>Y_t</i>	<i>Pronósticos</i> <i>Ŷ_t</i>	<i>Residuos</i> <i>e_t</i>
<i>t</i> -5	71	90	—	—
<i>t</i> -4	72	78	—	—
<i>t</i> -3	73	87	74.0	13.0
<i>t</i> -2	74	99	69.1	29.9
<i>t</i> -1	75	72	62.7	9.3
<i>t</i>	76		77.2	

Para modelos autorregresivos, los pronósticos dependen de los valores observados en periodos anteriores. Para modelos AR(2), los pronósticos del siguiente valor dependen de los valores observados en los dos periodos anteriores. Para modelos AR(3), los pronósticos del siguiente valor dependen de los valores observados en los tres periodos anteriores, etcétera.

Modelos de promedios móviles

Un modelo de promedio móvil de *q*-ésimo orden tiene 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 el tiempo *t*

μ = promedio constante en el proceso

$\omega_1, \omega_2, \dots, \omega_q$ = coeficientes que se estimarán

ε_t = término de error, el cual representa los efectos de las variables no explicadas por el modelo; los supuestos acerca del término de error son los mismos que aquellos del modelo de regresión estándar

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ = errores en periodos anteriores que, para el tiempo *t*, se incorporan a la respuesta, Y_t

La ecuación 9.2 es similar a la ecuación 9.1, excepto que la variable dependiente, Y_t , depende de los valores anteriores de los errores y no de los valores de la variable misma. Los modelos de promedio móvil (MA) permiten hacer pronósticos de Y_t con base en una combinación lineal de un número finito de errores pasados, mientras que los modelos autorregresivos (AR) pronostican Y_t como una función lineal de un número finito de valores pasados de Y_t .

El término *promedio móvil* para el modelo de la ecuación 9.2 es histórico y no debe confundirse con los procedimientos de promedio móvil estudiados en el capítulo 4. Aquí, *promedio móvil* se refiere al hecho de que la desviación de la respuesta de su media, $Y_t - \mu$, es una combinación lineal de los errores actuales y pasados y que, conforme el tiempo avanza, los errores implicados en esta combinación lineal también se mueven hacia delante.

$$\begin{aligned} Y_t - \mu &= \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \\ Y_{t+1} - \mu &= \varepsilon_{t+1} - \omega_1 \varepsilon_t - \omega_2 \varepsilon_{t-1} - \cdots - \omega_q \varepsilon_{t-q+1} \end{aligned}$$

Los pesos $\omega_1, \omega_2, \dots, \omega_q$ no necesariamente suman 1 y pueden ser positivos o negativos, aun cuando estén precedidos de un signo negativo en la especificación del modelo.

La figura 9-3 presenta las ecuaciones de un modelo de promedio móvil de orden 1 o modelo MA(1) y un modelo de promedio móvil de orden 2, o modelo MA(2). Los términos de error se pueden agregar en secuencia para obtener un MA(*q*), donde *q* es el número de térmi-

nos de error pasados que se incluirán en el pronóstico de la siguiente observación. Las figuras 9-3a) y b) también ilustran el comportamiento de los coeficientes teóricos de la autocorrelación y la autocorrelación parcial del modelo MA(1). Observe que las funciones de autocorrelación y autocorrelación parcial de los modelos AR y MA se comportan de formas muy diferentes. Esto es una ventaja, porque así los modelos AR y MA pueden distinguirse por sus patrones de autocorrelación y autocorrelación parcial.

Los coeficientes de autocorrelación para el modelo MA(1) caen a cero después del primer retraso, mientras que los coeficientes de autocorrelación parcial tienden a cero gradualmente. Además, los coeficientes de autocorrelación del modelo MA(2) se vuelven cero después del segundo retraso de tiempo, mientras que las autocorrelaciones parciales tienden gradualmente a cero [véase las figuras 9-3c) y d)]. Otra vez, debemos mencionar que las funciones de autocorrelación muestrales van a diferir de estas funciones teóricas a causa de la variación de la muestra.

Ejemplo 9.2

Las 75 lecturas del proceso incluidas en la tabla 9-5 nos permitirán ilustrar el método de pronóstico con el modelo MA(2). La demostración se hace usando sólo las cinco últimas observaciones y los residuos contenidos en la tabla 9-2.

Se elige un modelo MA(2), y el programa Minitab calcula los mínimos cuadrados³ de los coeficientes del modelo, obteniendo $\hat{\mu} = 75.4$, $\hat{\omega}_1 = .5667$ y $\hat{\omega}_2 = -.3560$. Otra vez, suponga que para el tiempo $t - 1 = 75$ se requiere un pronóstico de la observación para el periodo $t = 76$. Como en el tiempo $t - 1$ la mejor alternativa para los términos de error en el periodo siguiente es su valor promedio, cero, y las mejores alternativas para los errores en los periodos de tiempo actual y previo son los residuos correspondientes, el pronóstico del 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 los dos residuos, e_{75} y e_{74} , se sustituyen por los errores ε_{75} y ε_{74} en el cálculo del pronóstico para el periodo 76. Para calcular pronósticos a partir de modelos de promedio móvil, los errores correspondientes a los periodos que ya han ocurrido se sustituyen por los residuos de esos periodos. El número de residuos implicados en el pronóstico de la siguiente observación es igual al orden del modelo de promedio móvil.

TABLA 9-2 Elaboración de un pronóstico con un modelo de promedio móvil MA(2), para el ejemplo 9.2

Periodo	Tiempo t	Valores Y_t	Pronósticos $\hat{Y}_t$	Residuos e_t
$t-5$	71	90	—	13.9
$t-4$	72	78	—	8.9
$t-3$	73	87	75.3	11.7
$t-2$	74	99	72.0	27.0
$t-1$	75	72	64.3	7.7
t	76		80.6	

³Las estimaciones de los mínimos cuadrados para los coeficientes en los modelos de promedio móvil, o modelos que implican términos de promedio móvil, deben obtenerse de forma iterativa, usando un algoritmo no lineal de mínimos cuadrados. A partir de un punto de arranque inicial, un algoritmo no lineal de mínimos cuadrados genera estimaciones de coeficientes mejorados que tienen una suma más pequeña de errores al cuadrado. Las estimaciones se mejoran continuamente hasta que la suma de los errores al cuadrado ya no puede hacerse apreciablemente más pequeña. Si bien los modelos autorregresivos se pueden ajustar con paquetes de regresión estándar, las estimaciones de los mínimos cuadrados de los coeficientes en modelos autorregresivos a menudo se obtienen usando un procedimiento no lineal de mínimos cuadrados.

Modelos de promedio móvil autorregresivos

Un modelo con términos autorregresivos se puede combinar con un modelo que tenga términos de promedio móvil para obtener un modelo “mixto” de promedio móvil autorregresivo. Para representar estos modelos, es conveniente utilizar la notación $\text{ARMA}(p, q)$, donde p es el orden de la parte autorregresiva y q es el orden de la parte del promedio móvil. Un modelo $\text{ARMA}(p, q)$ 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 $\text{ARMA}(p, q)$ pueden describir una gran variedad de comportamientos de las series de tiempo estacionarias.⁴ Los pronósticos generados por un modelo $\text{ARMA}(p, q)$ dependen de los valores actuales y pasados de la respuesta, Y , así como de los valores actuales y pasados de los errores (residuos), e .

La figura 9-4 presenta la ecuación de un modelo $\text{ARMA}(1, 1)$ y los posibles comportamientos de las autocorrelaciones y autocorrelaciones parciales teóricas. En este caso, tanto las autocorrelaciones como las autocorrelaciones parciales se desvanecen; ninguna termina de manera abrupta.

Resumen

Los patrones de autocorrelación y autocorrelación parcial de los procesos de promedio móvil autorregresivos se resumen como sigue:

	<i>Autocorrelaciones</i>	<i>Autocorrelaciones parciales</i>
$\text{MA}(q)$	Terminan después del orden q del proceso	Se desvanecen
$\text{AR}(p)$	Se desvanecen	Terminan después del orden p del proceso
$\text{ARMA}(p, q)$	Se desvanecen	Se desvanecen

Los números de términos autorregresivos y de promedio móvil (orden p y orden q) en un modelo ARMA están determinados por los patrones de las autocorrelaciones y autocorrelaciones parciales de la muestra y los valores de los criterios de selección del modelo que se estudiarán más adelante en este capítulo. En la práctica, los valores de p y q rara vez exceden de 2.

ESTRATEGIA DE IMPLEMENTACIÓN DE LA CONSTRUCCIÓN DEL MODELO

Como se muestra en la figura 9-1, el método Box-Jenkins se basa en una estrategia iterativa para la construcción del modelo que consiste en la selección de un modelo inicial (identificación del modelo), la estimación de los coeficientes del modelo (estimación de parámetros) y el análisis de los residuos (verificación del modelo). Si es necesario, el modelo inicial se modifica y el proceso se repite hasta que los residuos indiquen que ya no es necesaria otra modificación. En este punto, estamos en condiciones de utilizar el modelo ajustado para pronosticar.

A continuación se estudian con detalle los pasos de la estrategia para la construcción del modelo.

Paso 1: Identificación del modelo

1. El primer paso en la identificación del modelo es determinar si la serie es estacionaria, es decir, si la serie de tiempo parece variar alrededor de un nivel fijo. Es útil observar una gráfica de la serie junto con la función de autocorrelación de la muestra. Se recomienda una

⁴Observe que cuando $q = 0$, el modelo $\text{ARMA}(p, 0)$ se reduce a un modelo autorregresivo puro de orden p . De forma similar, cuando $p = 0$, el modelo $\text{ARMA}(0, q)$ es un modelo de promedio móvil puro de orden q .

serie de tiempo no estacionaria si la serie parece crecer o declinar en el tiempo y las autocorrelaciones de la muestra no se desvanecen rápidamente. Las series de tiempo representadas en la figura 8-2 son no estacionarias, y el patrón de las autocorrelaciones muestrales mostradas en la figura 8-3 es característico de una serie no estacionaria.

Si la serie es no estacionaria, a menudo se puede convertir en una serie estacionaria por diferenciación. Es decir, la serie original se sustituye por una serie de las diferencias. Se especifica entonces un modelo ARMA para la serie diferenciada. En efecto, el analista está modelando cambios en vez de niveles.

Como ejemplo, suponga que la serie original, Y_t , se incrementa por lo general con el tiempo, pero las primeras diferencias, $\Delta Y_t = Y_t - Y_{t-1}$, varían alrededor de un nivel fijo. Sería recomendable modelar las diferencias estacionarias empleando un modelo ARMA de, digamos, orden $p = 1$ y $q = 1$. En este caso, el modelo es⁵

$$\Delta Y_t = \phi_1 \Delta Y_{t-1} + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

o

$$(Y_t - Y_{t-1}) = \phi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

En algunos casos, es necesario tomar las diferencias de las diferencias antes de obtener los datos estacionarios. Si la diferenciación simple se realiza dos veces, y los datos estacionarios son

$$\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

La diferenciación se hace hasta que una 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 requeridas para alcanzar la estacionalidad se denota por d .

Los modelos de las series no estacionarias se llaman modelos de promedio móvil *integrados* autorregresivos y se denotan por ARIMA(p, d, q).⁶ Aquí, p indica el orden de la parte autorregresiva, d indica el número de diferenciaciones, y q indica 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. Por consiguiente, de ahora en adelante, la notación ARIMA(p, d, q) se usará para identificar modelos tanto de series estacionarias ($d = 0$) como no estacionarias ($d > 0$).

Si bien los modelos ARIMA implican diferencias, los pronósticos de 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 se utilizará.

La identificación de la forma del modelo se lleva a cabo comparando las autocorrelaciones y las autocorrelaciones parciales calculadas con los datos de las autocorrelaciones y las autocorrelaciones parciales teóricas de los diferentes modelos ARIMA. Las correlaciones teóricas de algunos de los modelos ARIMA más comunes se presentan en las figuras 9-2, 9-3 y 9-4 para ayudar en la selección de un modelo adecuado.

Cada modelo ARIMA tiene un conjunto único de autocorrelaciones y autocorrelaciones parciales, y el analista debe tener la capacidad de identificar los valores que se obtienen a partir de la muestra a uno de los patrones teóricos.

Es posible que exista alguna ambigüedad en la identificación de un modelo ARIMA adecuado a partir de los patrones de las autocorrelaciones y las autocorrelaciones parciales. Por ello, la selección del modelo inicial debería considerarse como tentativa. Se deben realizar análisis durante los pasos 2 y 3 para determinar si el modelo es adecuado. Si no lo es,

⁵Cuando se usa un modelo ARMA para una serie diferenciada, tal vez no se requiera el término constante, ϕ_0 .

⁶El término *integrado* significa que las diferencias se deben sumar (o integrar) para obtener la serie original.

se debe intentar un modelo alternativo. Con un poco de práctica, el analista se volverá más apto para identificar el modelo adecuado.

Recuerde que, si las autocorrelaciones de la muestra se desvanecen exponencialmente hacia cero y las autocorrelaciones parciales de la muestra se cortan de forma abrupta, el modelo requerirá términos autorregresivos. Si las autocorrelaciones de la muestra se cortan manera abrupta y las autocorrelaciones parciales de la muestra se desvanecen, el modelo requerirá términos de promedio móvil. Si tanto las autocorrelaciones de la muestra como las autocorrelaciones parciales de la muestra se desvanecen, tanto los términos autorregresivos como los términos de promedio móvil están indicados. Al contar el número de autocorrelaciones y autocorrelaciones parciales significativas de la muestra, pueden determinarse los órdenes de las partes MA y AR. Para juzgar su significancia, tanto las autocorrelaciones de la muestra como las autocorrelaciones parciales de la muestra se comparan generalmente con $\pm 2/\sqrt{n}$ donde n es el número de observaciones en la serie de tiempo. Estos límites funcionan bien cuando n es grande.

En iguales condiciones, son preferibles los modelos simples en vez de los modelos complejos. Esto se conoce como el *principio de parsimonia*. Con una cantidad limitada de datos es relativamente fácil encontrar un modelo con un número grande de parámetros que se ajusten bien a los datos. Sin embargo, los pronósticos de un modelo tal probablemente van a ser deficientes porque buena parte de la variación en los datos se debe al error aleatorio que se está modelando. El objetivo es desarrollar el modelo más sencillo que ofrezca una descripción adecuada de las principales características de los datos.

El *principio de parsimonia* se refiere a la preferencia por los modelos sencillos por encima de los modelos complejos.

Paso 2: Estimación del modelo

1. Una vez que se ha seleccionado un modelo tentativo, se deben estimar los parámetros para ese modelo.

Los parámetros de los modelos ARIMA se estiman minimizando la suma de cuadrados de los errores de ajuste. Estos mínimos cuadrados estimados deben obtenerse, en general, usando un procedimiento no lineal de mínimos cuadrados. Un procedimiento no lineal de mínimos cuadrados es simplemente un algoritmo que obtiene el mínimo de la función de la suma de los errores al cuadrado. Una vez que se determinan las estimaciones de los mínimos cuadrados y sus errores estándar, se pueden construir e interpretar los valores t del modo habitual. Los parámetros que son considerados significativamente diferentes de cero se conservan en el modelo ajustado; los parámetros que no son significativos se eliminan del modelo.

Por ejemplo, suponga que un modelo ARIMA(1, 0, 1) se ha ajustado a una serie de tiempo de 100 observaciones y la ecuación ajustada es

$$\hat{Y}_t = 33.4 + .25Y_{t-1} - .5\hat{\varepsilon}_{t-1}$$

(7.02) (.17) (.21)

donde los números entre paréntesis debajo de los coeficientes estimados son sus errores estándar. Puesto que la razón t del coeficiente del término autorregresivo es $t = .25/.17 = 1.47$ con un valor de p de .14, la hipótesis $H_0: \phi_1 = 0$ no se rechaza, y este término puede eliminarse del modelo. Un modelo ARIMA(0, 0, 1), es decir, un modelo MA(1), puede entonces ajustarse a los datos.

2. Se calcula el error cuadrático medio de los residuos, una estimación de la varianza del error, ε_t .

El error cuadrático medio de los residuos se define como⁷

$$s^2 = \frac{\sum_{t=1}^n e_t^2}{n - r} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n - r} \quad (9.4)$$

donde

$$\begin{aligned} e_t &= Y_t - \hat{Y}_t = \text{residuo para el tiempo } t \\ n &= \text{número de residuos} \\ r &= \text{número total de parámetros estimados} \end{aligned}$$

3. El error cuadrático medio de los residuos es útil para evaluar el ajuste y comparar modelos diferentes. También es útil para calcular los límites de error de pronóstico.

Paso 3: Verificación del modelo

Antes de utilizar el modelo para pronosticar, debe verificarse qué tan adecuado es. En esencia, un modelo es adecuado si los residuos no se pueden usar para mejorar los pronósticos. Es decir, los residuos deben ser aleatorios.

1. Muchas de las gráficas residuales que son útiles en el análisis de regresión se pueden desarrollar para los residuos de un modelo ARIMA. Un histograma y una gráfica de probabilidad normal (para verificar la normalidad), así como una gráfica de la secuencia del tiempo (para verificar los valores atípicos) son particularmente útiles.
2. Las autocorrelaciones residuales individuales deben ser pequeñas y generalmente dentro de $\pm 2/\sqrt{n}$ de cero. Las autocorrelaciones residuales significativas en retrasos cortos o retrasos estacionales sugieren que el modelo es inadecuado y que debe seleccionarse un nuevo modelo o modificarse.
3. Las autocorrelaciones residuales como un grupo deben ser congruentes con aquellas producidas por los errores aleatorios.

Una verificación general de la idoneidad del modelo se realiza mediante una prueba de distribución chi cuadrada (χ^2) con base en el estadístico Q de Ljung-Box. Esta prueba considera los tamaños de las autocorrelaciones residuales como un grupo. El estadístico de prueba Q es

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2(e)}{n - k} \quad (9.5)$$

el cual está distribuido aproximadamente como una variable aleatoria chi cuadrada con $m - r$ grados de libertad, donde r es el número total de parámetros estimados en el modelo ARIMA. En la ecuación 9.5,

$$\begin{aligned} r_k(e) &= \text{autocorrelación residual para el retraso } k \\ n &= \text{número de residuos} \\ k &= \text{retraso de tiempo} \\ m &= \text{número de retrasos de tiempo que van a ser evaluados} \end{aligned}$$

Si el valor p asociado con el estadístico Q es pequeño (digamos, el valor $p < .05$), el modelo se considera inadecuado. El analista debe considerar un nuevo modelo o uno modificado y continuar el análisis hasta encontrar un modelo satisfactorio.

El criterio desempeña un papel importante en la tarea de construcción del modelo. Dos modelos sencillos en competencia pueden describir adecuadamente los datos, y se debe hacer una selección con base en la naturaleza de los pronósticos. Además, se pueden ignorar algunos

⁷La raíz cuadrada de s^2 , $s = \sqrt{s^2}$, es análoga al error estándar de la estimación (véase la ecuación 7.4).

residuos grandes si éstos se pueden explicar por circunstancias inusuales y el modelo es adecuado para el resto de las observaciones.

Paso 4: Elaboración de pronósticos con el modelo

1. Una vez que se ha encontrado un modelo adecuado, es factible elaborar los pronósticos de uno o varios periodos futuros.

Con base en los pronósticos también se pueden construir intervalos de predicción. En general, para un nivel de confianza dado, cuanto más largo sea el tiempo guía del pronóstico, mayor será el intervalo de predicción. Esto es razonable, puesto que se espera que la incertidumbre sea mayor para el pronóstico de un valor distante que para el pronóstico de, digamos, la siguiente observación. El cálculo de los pronósticos e intervalos de predicción es una labor tediosa y es preferible dejarla a la computadora. Los programas de computadora que ajustan modelos ARIMA generan pronósticos e intervalos de predicción a requerimiento del analista.

2. Conforme más datos están disponibles, se puede usar el mismo modelo ARIMA para generar pronósticos modificados de otro origen de tiempo.
3. Si el patrón de la serie parece estar cambiando con el tiempo, los datos nuevos se pueden usar para recalcular los parámetros del modelo o, si es necesario, desarrollar un modelo completamente nuevo.

Es una buena idea hacer un seguimiento de los errores de pronóstico. Si los errores más recientes tienden a ser sistemáticamente más grandes que los errores anteriores, tal vez sea el momento de reevaluar el modelo. En este punto, es posible que se requiera otra estrategia de iteración para construir el modelo. Lo mismo es verdadero si los errores recientes del pronóstico tienden a ser positivos de manera sistemática (predicen una respuesta muy baja) o negativos (predicen una respuesta muy alta).

Ejemplo 9.3

Cameron Consulting Corporation se especializa en ofrecer portafolios de inversión. A Lynn Stephens, la analista de la compañía se le asignó la tarea de desarrollar técnicas más complejas para pronosticar los promedios del Dow Jones. Lynn asistió recientemente a un taller sobre metodología Box-Jenkins y decidió intentar esta técnica con el Índice de transportación del Dow Jones. La tabla 9-3 representa 65 promedios diarios al cierre de operaciones del Índice de transportación durante los meses de verano.

TABLA 9-3 Promedios diarios al cierre de operaciones del Índice de transportación para el ejemplo 9.3

<i>Periodo</i>	<i>Promedio al cierre</i>	<i>Periodo</i>	<i>Promedio al cierre</i>	<i>Periodo</i>	<i>Promedio al cierre</i>	<i>Periodo</i>	<i>Promedio al cierre</i>	<i>Periodo</i>	<i>Promedio al cierre</i>
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		

Lynn empezó el análisis examinando la gráfica de datos presentada en la figura 9-5, donde parece haber una tendencia ascendente en la serie. Su siguiente paso en la identificación de un modelo tentativo fue examinar las autocorrelaciones de la muestra de los datos presentados en la figura 9-6. Cuando Lynn observó que varias de las primeras autocorrelaciones eran persistentemente grandes y tendían a cero con lentitud, se dio cuenta de que su observación original era correcta: esta serie de tiempo era no 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. La gráfica de los datos diferenciados (no mostrada) parece variar alrededor de un nivel fijo. De hecho, la media muestral de las diferencias era 1.035. Las autocorrelaciones de la muestra para las



FIGURA 9-5 Promedios diarios al cierre de operaciones del Índice de transporte del Dow Jones

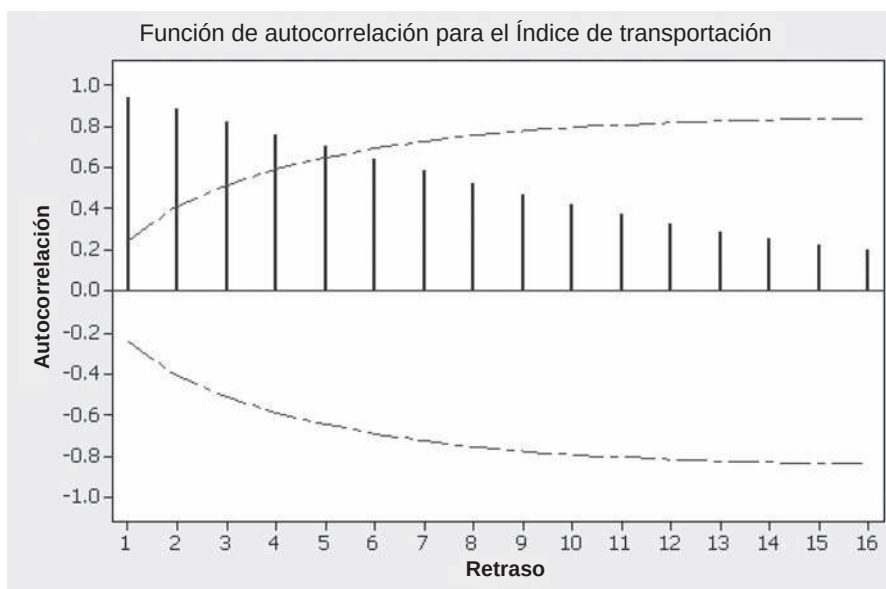


FIGURA 9-6 Función de autocorrelación de la muestra del Índice de transporte para el ejemplo 9.3

diferencias se presentan en la figura 9-7. Las autocorrelaciones parciales de la muestra para las diferencias aparecen en la figura 9-8.

Lynn estaba perpleja. Al comparar las autocorrelaciones con sus límites de error, la única autocorrelación significativa estaba en el retraso 1. De manera análoga, sólo la autocorrelación parcial del retraso 1 era significativa. Las autocorrelaciones aparentan cortarse después del retraso 1, mostrando un comportamiento tipo MA(1). Al mismo tiempo, las autocorrelaciones parciales parecían cortarse

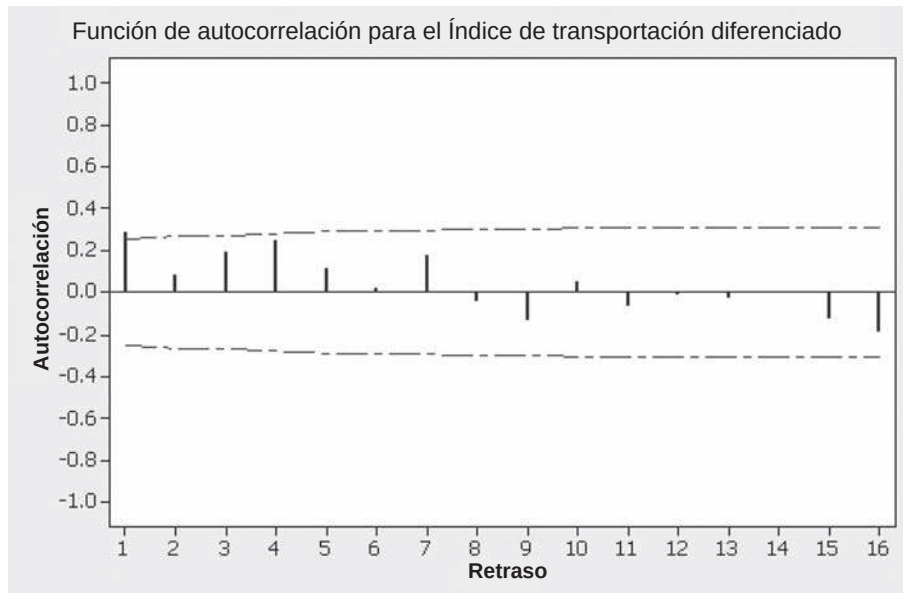


FIGURA 9-7 Función de autocorrelación de la muestra para las primeras diferencias del Índice de transportación; ejemplo 9.3

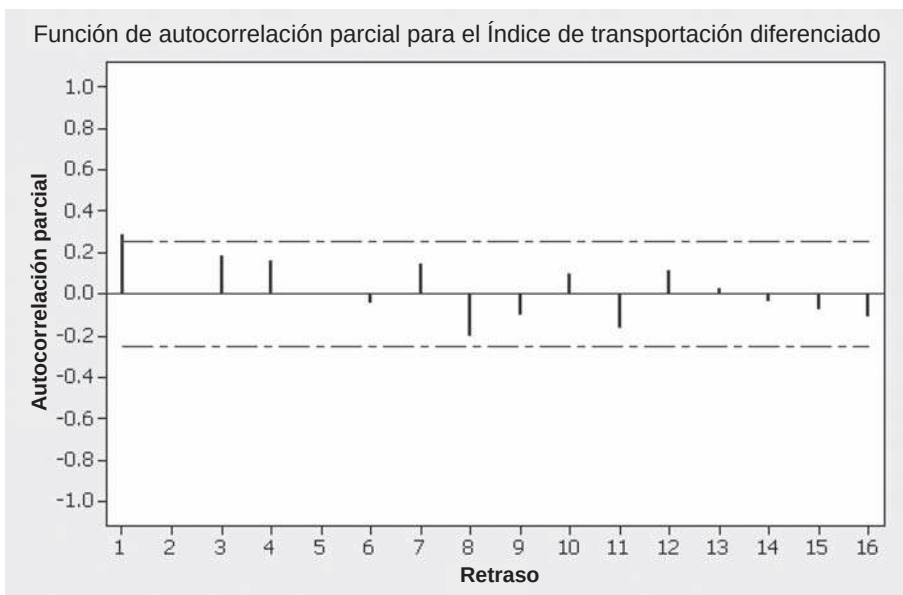


FIGURA 9-8 Autocorrelaciones parciales de la muestra de las primeras diferencias del Índice de transportación; ejemplo 9.3

después del retraso 1, mostrando un comportamiento tipo AR(1). Al parecer, ninguno de los patrones se desvanece de una manera suave en los retrasos cortos. Lynn decidió ajustar tanto el modelo ARIMA(1, 1, 0) como el modelo ARIMA(0, 1, 1) para el Índice de transportación. Decidió incluir un término constante en cada modelo para tener en cuenta el hecho de que la serie de las diferencias parece variar alrededor de un nivel arriba de cero. Si Y_t representa el Índice de transportación, entonces la serie diferenciada es $\Delta Y_t = Y_t - Y_{t-1}$, y los modelos de Lynn son

$$\text{ARIMA}(1, 1, 0): \Delta Y_t = \phi_0 + \phi_1 \Delta Y_{t-1} + \varepsilon_t$$

$$\text{ARIMA}(0, 1, 1): \Delta Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1}$$

Los resultados de Minitab para los modelos de Lynn se presentan en la tabla 9-4. Las autocorrelaciones residuales del modelo ARIMA(1, 1, 0) ajustado se aprecian en la figura 9-9.

Ambos modelos se ajustan a los datos igualmente bien. Los cuadrados medios residuales (MS) son

$$\text{ARIMA}(1, 1, 0): s^2 = 3.536$$

$$\text{ARIMA}(0, 1, 1): s^2 = 3.538$$

Lynn también notó que la constante en el modelo ARIMA(0, 1, 1) se estima como $\hat{\mu} = 1.0381$, (esencialmente) la media muestral de las diferencias.

La figura 9-9 indica que no existe autocorrelación residual significativa para el modelo ARIMA(1, 1, 0). Aunque no se muestra la función de autocorrelación residual para el modelo ARIMA(0, 1, 1), los resultados son similares. Los estadísticos Q de Ljung-Box calculados para los grupos de retrasos $m = 12, 24, 36$ y 48 no son significativos, como lo indicaron los valores grandes de p de cada modelo. Lynn determinó que cualquier modelo era adecuado. Más aún, los pronósticos de un paso anticipado que ofrecen los dos modelos son aproximadamente los mismos.

Para resolver su dilema, Lynn optó por el modelo ARIMA(1, 1, 0) con base en su ajuste ligeramente mejor. Verificó el pronóstico para el periodo 66 de este modelo como sigue:

$$Y_t - Y_{t-1} = \phi_0 + \phi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t$$

o

$$Y_t = Y_{t-1} + \phi_0 + \phi_1(Y_{t-1} - Y_{t-2}) + \varepsilon_t$$

Función de autocorrelación: Residuos ARIMA(1, 1, 0)

Retraso	ACF	T	LBQ
1	-0.002164	-0.02	0.00
2	-0.054880	-0.44	0.21
3	0.126703	1.01	1.32
4	0.192990	1.52	3.94
5	0.053912	0.41	4.15
6	-0.069976	-0.53	4.50
7	0.215477	1.62	7.94
8	-0.062525	-0.45	8.24
9	-0.155310	-1.12	10.09
10	0.119656	0.85	11.21
11	-0.087686	-0.61	11.82
12	0.005717	0.04	11.83
13	-0.024945	-0.17	11.88
14	0.052536	0.37	12.11
15	-0.083232	-0.58	12.71
16	-0.170800	-1.18	15.28

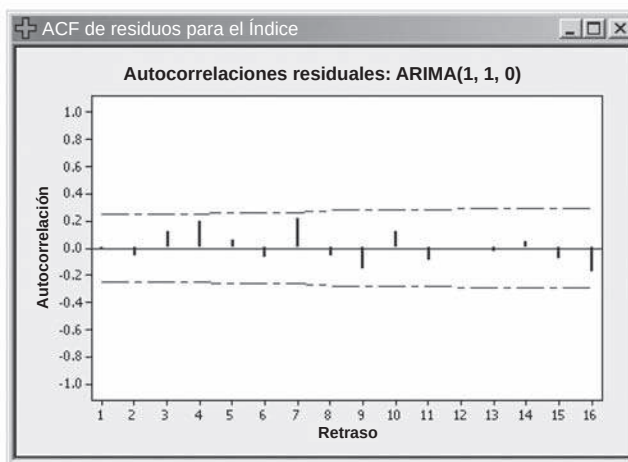


FIGURA 9-9 Autocorrelaciones residuales: modelo ARIMA (1, 1, 0) ajustado al Índice de transportación para el ejemplo 9.3

Con $\hat{\phi}_0 = .741$ y $\hat{\phi}_1 = .284$, la ecuación del pronóstico se convierte 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 concuerda con el resultado de la tabla 9-4. El intervalo de predicción para el valor real en el periodo 66 calculado por Minitab es (286.3, 293.6).

TABLA 9-4 Resultado de Minitab para los modelos ARIMA(1, 1, 0) y ARIMA(0, 1, 1) ajustados al Índice de transportación para el ejemplo 9.3

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

Period	Forecast	95% Limits		Actual
		Lower	Upper	
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

Period	Forecast	95% Limits		Actual
		Lower	Upper	
66	290.053	286.366	293.740	

TABLA 9-5 Lecturas del proceso de 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	

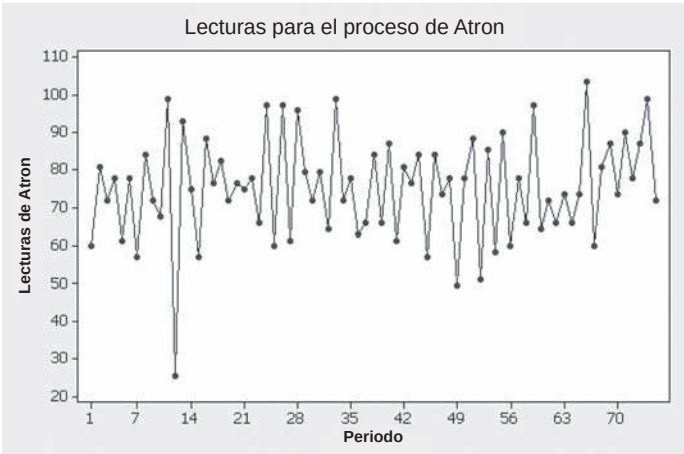


FIGURA 9-10 Lecturas del proceso de Atron

Ejemplo 9.4
El analista de la Atron Corporation, Jim White, tenía una serie de tiempo de las lecturas de un proceso industrial que requería de un pronóstico. Los datos están contenidos en la tabla 9-5. Las lecturas se encuentran graficadas en la figura 9-10. Jim cree que la metodología Box-Jenkins podría trabajar mejor con estos datos.
Jim empezó por identificar un modelo tentativo observando la gráfica de los datos y las autocorrelaciones de la muestra presentadas en la figura 9-11. La serie de tiempo de las lecturas parece variar alrededor de un nivel fijo de 80, y las autocorrelaciones se desvanecen rápidamente hacia cero. Jim concluyó que la serie de tiempo era estacionaria.
El primer coeficiente de autocorrelación de la muestra (−.53) es significativamente diferente de cero al nivel del 5%, puesto que permanece fuera del rango

$$0 \pm 2 \frac{1}{\sqrt{n}} = 0 \pm 2 \frac{1}{\sqrt{75}} = 0 \pm 2(.115) = 0 \pm .23$$

La autocorrelación en el retraso 2 es casi significativa al nivel del 5% y es de signo opuesto a la autocorrelación del retraso 1, r_1 . Las autocorrelaciones restantes son pequeñas y están comprendidas dentro de sus límites de error individuales. El patrón de las autocorrelaciones sugiere un modelo AR(1) [véase la figura 9-2b)] o quizás un modelo MA(2) si las autocorrelaciones se cortan (se vuelven cero) después del retraso 2. Jim decidió examinar las autocorrelaciones parciales de la muestra que se reproducen en la figura 9-12.

Autocorrelaciones: Lecturas de Atron

Retraso	ACF	T	LBQ
1	-0.528475	-4.58	21.80
2	0.281574	1.95	28.07
3	-0.038337	-0.25	28.19
4	0.007939	0.05	28.19

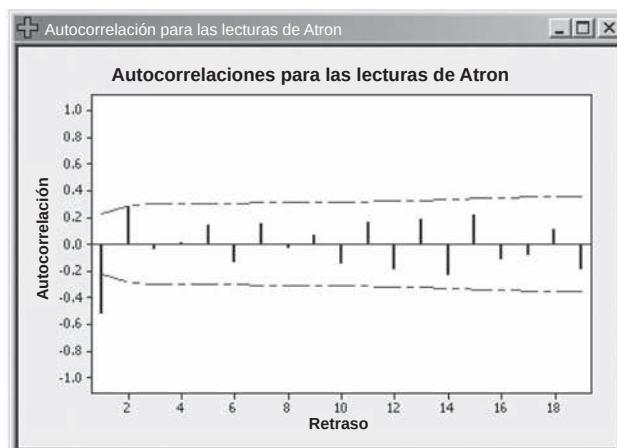


FIGURA 9-11 Función de autocorrelación de la muestra de las lecturas de Atron para el ejemplo 9.4

Autocorrelaciones parciales: Lecturas de Atron

Retraso	PACF	T
1	-0.528475	-4.58
2	0.003175	0.03
3	0.154951	1.34
4	0.065018	0.56

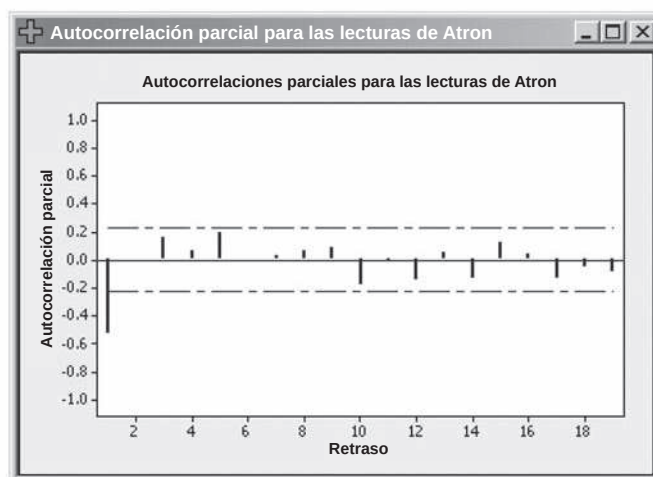


FIGURA 9-12 Autocorrelaciones parciales de la muestra de las lecturas de Atron para el ejemplo 9.4

Jim notó que la primera autocorrelación parcial ($-.53$) era significativamente diferente de cero, pero ninguna de las otras autocorrelaciones parciales se acercaban a la significancia.⁸ Jim pensó que las autocorrelaciones y las autocorrelaciones parciales sugerían un AR(1) [o, de manera equivalente, un modelo ARIMA(1, 0, 0)], pero para ir a lo seguro, decidió ajustar también un modelo MA(2) [ARIMA(0, 0, 2)]. Si ambos modelos son adecuados, él decidirá el asunto usando el principio de parsimonia.

La tabla 9-6 presenta los resultados de Minitab de ajustar los modelos AR(1) y MA(2) a las lecturas del proceso de Atron. Se incluye un término constante en ambos modelos para tomar en cuenta el hecho de que las lecturas varían alrededor de algún otro nivel diferente de cero.⁹

⁸El valor de la primera autocorrelación parcial siempre es igual al valor de la primera autocorrelación.

⁹Si los datos se hubieran expresado como desviaciones de la media muestral, no se requeriría un término constante en ningún modelo.

TABLA 9-6 Resultado de Minitab para los modelos MA(2) y AR(1) de las lecturas de Atron para el ejemplo 9.4**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

Period	Forecast	95% Limits		Actual
		Lower	Upper	
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

Period	Forecast	95% Limits		Actual
		Lower	Upper	
76	77.116	54.097	100.136	
77	74.364	48.227	100.502	
78	75.845	48.872	102.817	

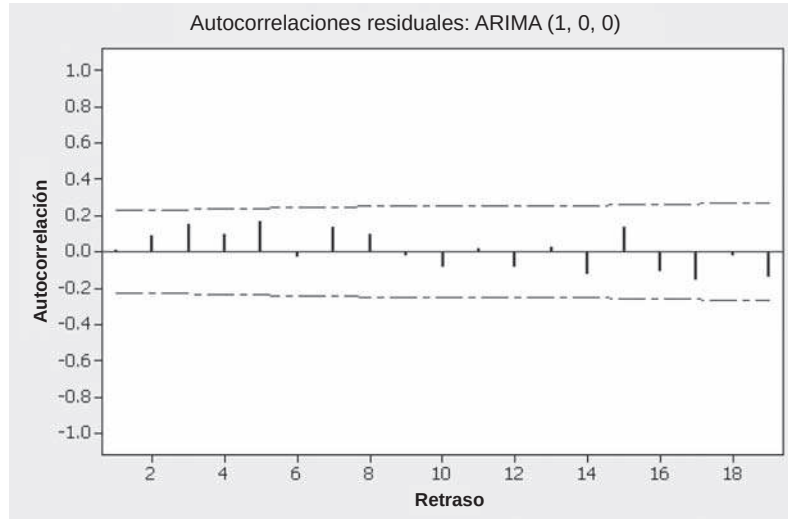


FIGURA 9-13 Función de autocorrelación de los residuos del modelo AR(1) para el ejemplo 9.4

Ambos modelos parecen ajustarse bien a los datos. Los coeficientes estimados son significativamente diferentes 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 uno o dos pasos de los dos modelos son algo diferentes, pero los pronósticos anticipados tres pasos (el periodo 78) son casi los mismos. Para un origen de pronóstico fijo, los pronósticos de procesos estacionarios tarde o temprano igualarán el nivel promedio estimado. En este caso, se estima que el nivel promedio es aproximadamente $\mu = 75$ para ambos modelos.

El estadístico Q de Ljung-Box no es significativo para las colecciones de retrasos $m = 12, 24, 36$ y 48 para cualquier modelo. Las autocorrelaciones residuales del modelo AR(1) se presentan en la figura 9-13. Las autocorrelaciones residuales individuales son pequeñas y están comprendidas dentro de los límites de error. La función de autocorrelación residual del modelo MA(2) es similar. En los dos modelos, no hay razón para creer que los errores no son aleatorios.

Puesto 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 apeló al principio de parsimonia y decidió utilizar el modelo más sencillo AR(1) para pronosticar lecturas futuras.

La ecuación del pronóstico AR(1) es¹⁰

$$\hat{Y}_t = 115.842 + (-.538)Y_{t-1} = 115.842 - .538Y_{t-1}$$

así que, para el periodo 76,

$$\hat{Y}_{76} = 115.842 - .538Y_{75} = 115.842 - .538(72) = 77.11$$

Más aún, el pronóstico anticipado dos pasos es

$$\hat{Y}_{77} = 115.842 - .538\hat{Y}_{76} = 115.842 - .538(77.11) = 74.36$$

Jim estaba contento de ver que estos resultados concordaban con los de Minitab.

¹⁰El término de error, ε_t , se elimina porque, para el pronóstico $\hat{Y}_t$, la mejor opción de ε_t es equivalente a cero.

Ejemplo 9.5

Jim White estaba contento con los resultados de sus pronósticos para la serie de tiempo de las lecturas presentadas en la tabla 9-5. Decidió usar la metodología Box-Jenkins para tratar de pronosticar los errores (desviaciones del objetivo) resultantes del control de calidad de un proceso de manufactura bajo su supervisión. Los datos aparecen en la tabla 9-7, y la serie de tiempo de los errores está graficada en la figura 9-14.

Jim inició el proceso de identificación examinando la gráfica de la serie de tiempo de los errores y verificando las autocorrelaciones y las autocorrelaciones parciales presentadas 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. Existe una autocorrelación significativa de $-.50$ en el retraso 1 y las autocorrelaciones en

TABLA 9-7 Errores de control de calidad de Atron para el ejemplo 9.5 (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

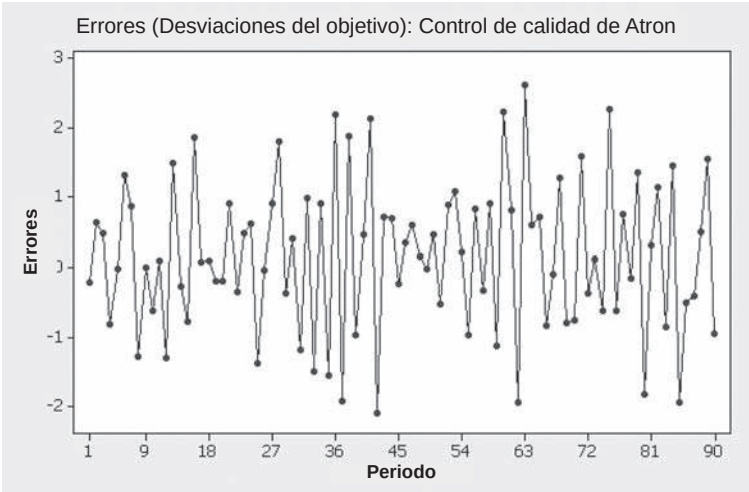


FIGURA 9-14 Errores (desviaciones del objetivo) del control de calidad de Atron para el ejemplo 9.5

Función de autocorrelación: Errores

Retraso	ACF	T	LBQ
1	-0.495654	-4.70	22.86
2	0.088540	0.69	23.59
3	-0.055353	-0.43	23.89
4	0.073234	0.56	24.40

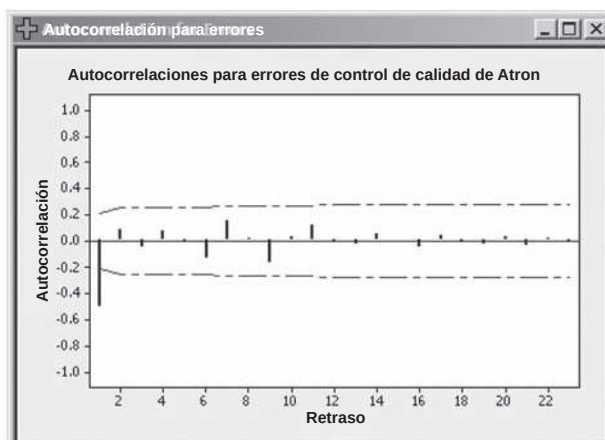


FIGURA 9-15 Autocorrelaciones de la muestra de los errores de control de calidad de Atron para el ejemplo 9.5

Función de autocorrelación parcial: Errores

Retraso	PACF	T
1	-0.495654	-4.70
2	-0.208309	-1.98
3	-0.146308	-1.39
4	-0.011216	-0.11

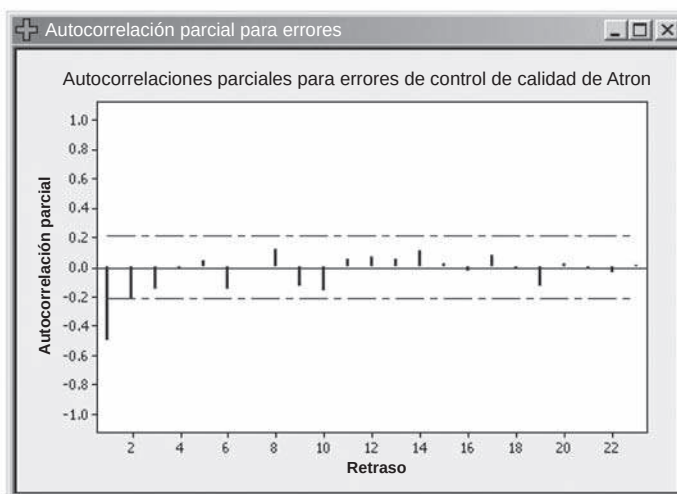


FIGURA 9-16 Autocorrelaciones parciales de la muestra de los errores de control de calidad de Atron para el ejemplo 9.5

los retrasos restantes son pequeñas y están comprendidas dentro de sus límites de error; las autocorrelaciones de la muestra parecen desvanecerse después del retraso 1. Iniciando con una autocorrelación parcial significativa en el retraso 1, las primeras tres autocorrelaciones parciales de la muestra son todas negativas y decaen hacia cero. Los comportamientos de las autocorrelaciones de la muestra y de las autocorrelaciones parciales de la muestra se parecen mucho a las cifras teóricas del proceso MA(1) [ARIMA(0, 0, 1)] presentado en la figura 9-3a). Jim estaba convencido de que su serie podía representarse con un modelo MA(1).

Jiim empleó Minitab para ajustar el modelo MA(1) a sus datos. Los resultados se despliegan en la tabla 9-8. Los parámetros estimados del modelo MA(1) son $\hat{\mu} = .1513$ y $\hat{\omega}_1 = .5875$. Los dos son significativamente diferentes de cero. Las autocorrelaciones residuales presentadas en la figura 9-17, y los estadísticos Q de Ljung-Box indican que los errores son aleatorios.

La ecuación del pronóstico MA(1) es

$$\hat{Y}_t = \hat{\mu} - \hat{\omega}_1 \varepsilon_{t-1} = .1513 - .5875 \varepsilon_{t-1}$$

TABLA 9-8 Resultados de Minitab para el modelo MA(1) del control de calidad de Atron para el ejemplo 9.5

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% Limits		Actual
		Lower	Upper	
91	0.43350	-1.37018	2.23719	
92	0.15129	-1.94064	2.24322	

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	-0.049643	-0.47	0.23
2	0.076340	0.72	0.78
3	0.003799	0.04	0.78
4	0.063375	0.60	1.17
5	-0.029768	-0.28	1.25
6	-0.112420	-1.05	2.50
7	0.101193	0.94	3.52
8	-0.017494	-0.16	3.55
9	-0.169055	-1.55	6.47
10	0.015437	0.14	6.50
11	0.146365	1.31	8.74
12	0.056395	0.49	9.08

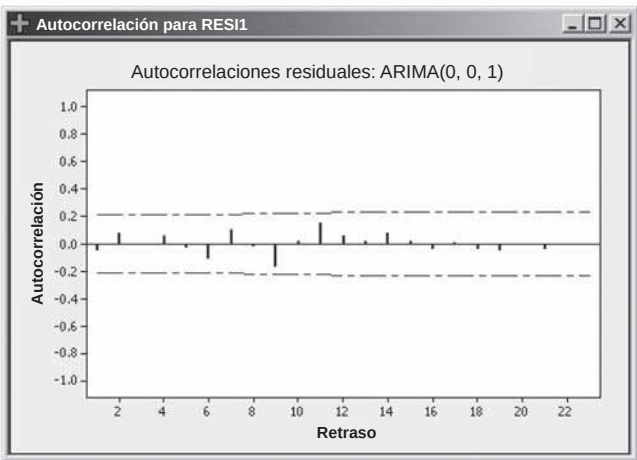


FIGURA 9-17 Función de autocorrelación de los residuos para el modelo MA(1); ejemplo 9.5

donde ε_{t-1} es estimado por el residuo correspondiente e_{t-1} . Para pronosticar el error (desviación del objetivo) del periodo 91, se requiere el residuo del periodo 90, $e_{90} = -.4804$. Jim calculó

$$\hat{Y}_{91} = .1513 - .5875(-.4804) = .4335$$

El pronóstico de Jim para el error de control de calidad en el periodo 92 es simplemente la media estimada de la serie porque, en el origen del pronóstico, $t = 90$, la mejor opción para el término de error en el periodo 91, ε_{91} , es cero. Por lo tanto,

$$\hat{Y}_{92} = .1513 - .5875(0) = .1513$$

Jim estaba satisfecho de que sus pronósticos calculados coincidieran con los de Minitab.

Ejemplo 9.6

El analista de Atron Corporation, Jim White, estaba encantado con los pronósticos de los errores de control de calidad que desarrolló usando la metodología Box-Jenkins tal como se describió en el ejemplo 9.5. Jim se encontró a su amigo Ed Jones y le contó acerca de su éxito. Ed había estado lidiando con una aplicación similar y decidió intentar con la metodología Box-Jenkins. Los datos de Ed se presentan en la tabla 9-9, y la figura 9-18 contiene la gráfica de una serie de tiempo de los datos.

La gráfica de la serie y las autocorrelaciones de la muestra, contenidas en la figura 9-19, sugieren que la serie de los errores de control de calidad es estacionaria. Los errores parecen variar alrededor de un nivel fijo de cero, y las autocorrelaciones se desvanecen muy rápidamente.

Ed notó que las primeras dos autocorrelaciones eran significativamente diferentes de cero y, quizás lo más importante, las autocorrelaciones de unos cuantos de los primeros retrasos disminuían hacia cero de una forma muy parecida al patrón teórico de un proceso AR(1) [véase la figura 9-2a]. Ed también examinó las autocorrelaciones parciales de la muestra mostradas en la figura 9-20. Como lo sospechaba, las autocorrelaciones parciales parecían cortarse después de la autocorrelación parcial significativa del retraso 1.¹¹ Considerados en conjunto, los patrones de las autocorrelaciones de la muestra y de las autocorrelaciones parciales de la muestra eran congruentes con un proceso AR(1), y

TABLA 9-9 Errores de 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

¹¹Ed atribuyó la autocorrelación parcial significativa del retraso 5 a un error de muestreo, puesto que no podía encontrar una razón física para que los errores de control de calidad separados por cinco periodos estuvieran correlacionados.

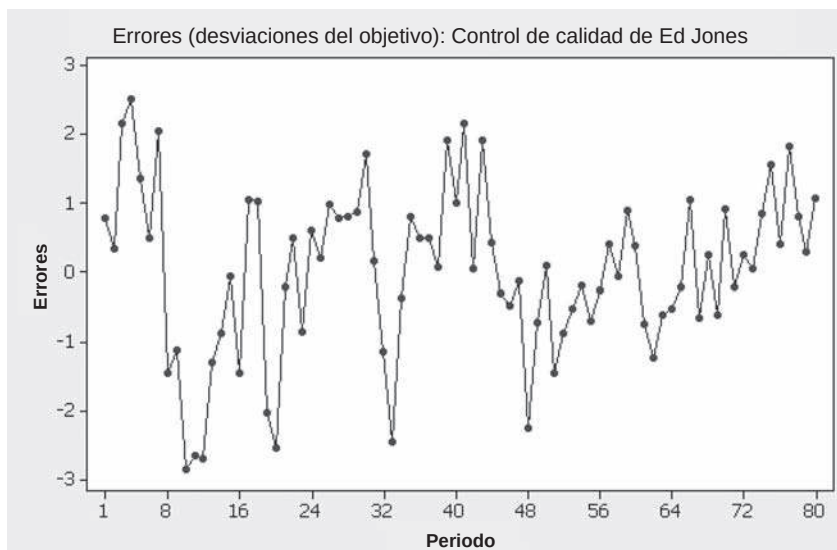


FIGURA 9-18 Errores (desviaciones del objetivo) de control de calidad de Ed Jones para el ejemplo 9.6

Función de autocorrelación: Errores

Retraso	ACF	T	LBQ
1	0.493582	4.41	20.23
2	0.346020	2.54	30.30
3	0.193314	1.32	33.48
4	0.128236	0.85	34.90
5	-0.126079	-0.83	36.29
6	-0.163715	-1.07	38.67

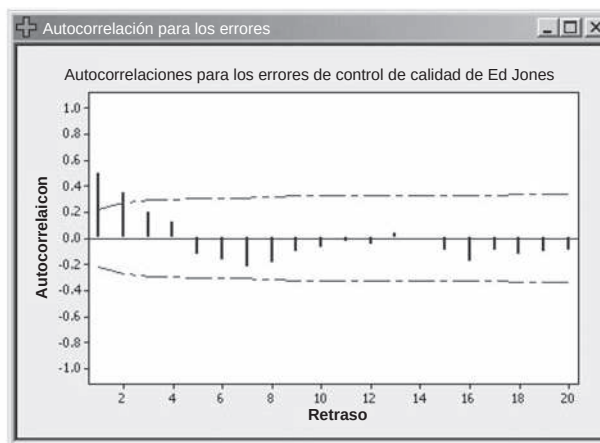


FIGURA 9-19 Autocorrelaciones de la muestra para los errores de control de calidad de Ed Jones para el ejemplo 9.6

Ed se sintió cómodo ajustando un modelo $AR(1)$ [ARIMA(1, 0, 0)] a los errores de su serie de tiempo (desviaciones del objetivo).

El resultado del intento inicial de Ed en Minitab con la metodología Box-Jenkins se presenta en la tabla 9-10. Puesto que la media muestral de la serie de error es extremadamente pequeña (cerca de cero) en comparación con su desviación estándar, Ed no incluyó un término constante en su modelo $AR(1)$.

Se estimó que el parámetro en el modelo $AR(1)$ es $\hat{\phi}_1 = .501$, y que es significativamente diferente de cero ($t = 5.11$). El error cuadrático medio de los residuos es $s^2 = 1.0998$. Las gráficas de los residuos (no mostradas), los estadísticos chi cuadrada de Ljung-Box y las autocorrelaciones residuales desplegadas en la figura 9-21 sugieren que el modelo $AR(1)$ es adecuado. No hay razón para dudar de los supuestos usuales del término de error.

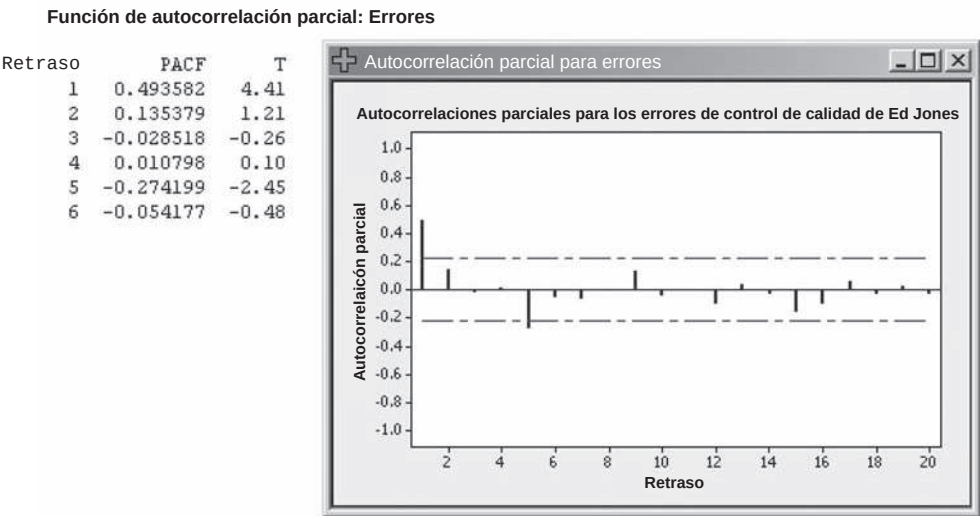


FIGURA 9-20 Autocorrelaciones parciales de la muestra de los errores de control de calidad de Ed Jones para el ejemplo 9.6

TABLA 9-10 Resultado de Minitab para los errores de control de calidad del modelo AR(1) de Ed Jones para el ejemplo 9.6

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

Period	Forecast	95% Limits Lower	Upper	Actual
81	0.53088	-1.52498	2.58673	
82	0.26588	-2.03340	2.56515	

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	-0.069795	-0.62	0.40
2	0.113142	1.01	1.48
3	0.002175	0.02	1.48
4	0.178268	1.57	4.22
5	-0.187692	-1.60	7.31
6	-0.039199	-0.32	7.44
7	-0.117368	-0.97	8.68
8	-0.103656	-0.85	9.66
9	0.013443	0.11	9.68
10	-0.037854	-0.31	9.81
11	0.044251	0.36	10.00
12	-0.085789	-0.69	10.71

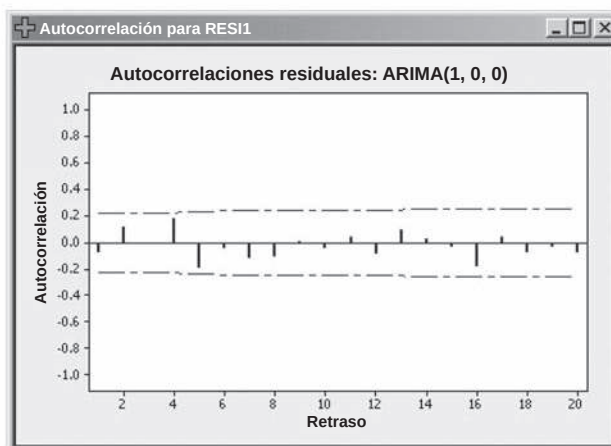


FIGURA 9-21 Función de autocorrelación de los residuos del modelo AR(1) para el ejemplo 9.6

La ecuación del pronóstico AR(1) es

$$\hat{Y}_t = .501Y_{t-1}$$

de manera que los pronósticos para los periodos 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 calculados por Ed concuerdan con los pronósticos anticipados uno y dos pasos generados por Minitab.

Como era su primera experiencia con la metodología Box-Jenkins, Ed decidió ajustar un modelo ligeramente más complejo para ver si los resultados confirmaban la elección de un modelo AR(1) para sus datos. Decidió agregar un parámetro más y ajustar un modelo ARIMA(1, 1)[ARIMA(1, 0, 1)] a los errores de control de calidad. Pensó que, si su modelo AR(1) era correcto, el parámetro de promedio móvil en el modelo ARIMA(1, 1) no debería ser significativo.

La tabla 9-11 presenta el resultado del modelo ARIMA(1, 1) ajustado a la serie de Ed. Él estaba contento de ver que el parámetro MA(1) no era significativamente diferente de cero ($t = 1.04$), lo que implicaba que este término no era necesario. Desde luego, como se trata de un modelo ligeramente más general que el modelo AR(1), aún representa de manera adecuada los datos, como lo indican $s^2 = 1.0958$ y el comportamiento aleatorio de los residuos.

Ejemplo 9.7

A Jill Blake, analista de ISC Corporation, se le pidió desarrollar pronósticos para los precios de las acciones de la compañía al cierre de operaciones. Las acciones se habían debilitado con poco crecimiento durante algún tiempo, y la gerencia quería algunas proyecciones para discutir las con el consejo de administración. Los precios de las acciones de ISC se encuentran graficados en la figura 9-22 y listados en la tabla 9-12.

Un tanto sorprendentemente, ya que los precios de las acciones tienden a ser no estacionarios, la gráfica de los precios de las acciones de ISC tiende a variar alrededor de un nivel fijo de aproximadamente 250 (véase la figura 9-22). Jill concluyó que su serie de precios era estacionaria.

Jill cree que la metodología Box-Jenkins podría venirle bien a su propósito e inmediatamente generó las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra de la serie de precios de las acciones. Los resultados se despliegan en las figuras 9-23 y 9-24.

Jill notó que las autocorrelaciones de la muestra alternan en signo y se desvanecen hacia cero en los retrasos cortos. Las autocorrelaciones parciales de la muestra son similares, pero se cortan clara-

TABLA 9-11 Resultado de Minitab del modelo ARIMA(1, 0, 1) para los errores de control de calidad de Ed Jones; ejemplo 9.6

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% Limits		Actual
		Lower	Upper	
81	0.52814	-1.52399	2.58027	
82	0.35230	-1.89526	2.59985	

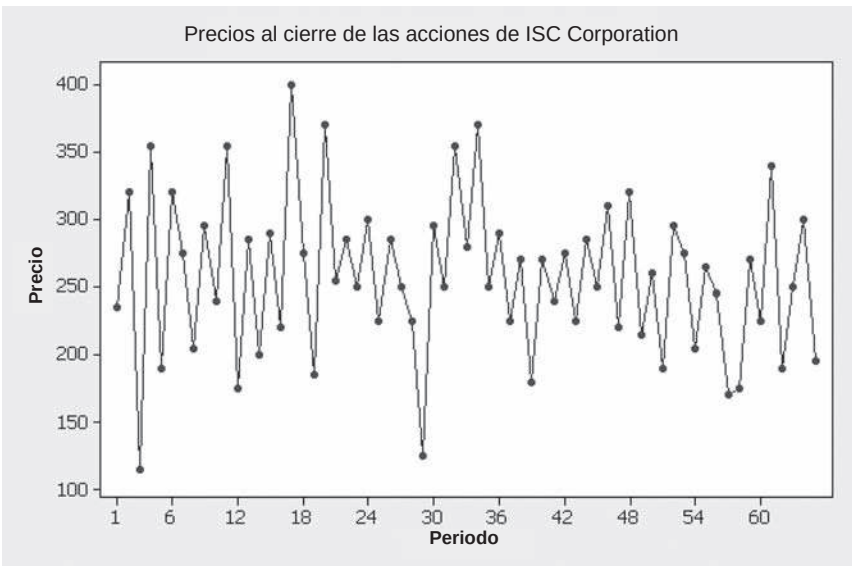


FIGURA 9-22 Precios al cierre de operaciones de las acciones de ISC Corporation para el ejemplo 9.7

mente (se vuelven cero) después del retraso 2. Jill recordó los patrones posibles de las autocorrelaciones teóricas y las autocorrelaciones parciales de un proceso AR(2) [véase la figura 9-2d)] y determinó que las cantidades muestrales correspondientes son congruentes con uno de esos patrones. Ella identificó un modelo AR(2) [ARIMA(2, 0, 0)] para sus datos.

TABLA 9-12 Precios al cierre de operaciones de las acciones de ISC Corporation para el 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

Función de autocorrelación: Precios de las acciones ISC

Retraso	ACF	T	LBQ
1	-0.401525	-3.24	10.97
2	0.333244	2.34	18.65
3	-0.204772	-1.33	21.59
4	0.111059	0.70	22.47
5	-0.182873	-1.15	24.90
6	0.010786	0.07	24.91

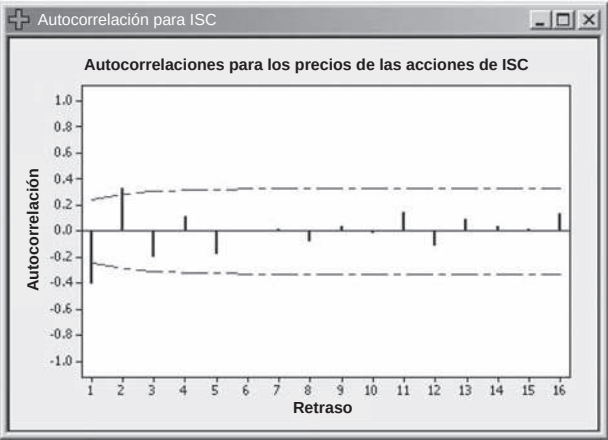


FIGURA 9-23 Autocorrelaciones de la muestra para los precios de las acciones de la ISC; ejemplo 9.7

Jill usó Minitab para estimar los parámetros de su modelo. Incluyó un término constante para tomar en cuenta un nivel diferente de cero. El resultado se presenta en la tabla 9-13.

Los parámetros estimados son $\hat{\phi}_0 = 284.9$, $\hat{\phi}_1 = -.324$ y $\hat{\phi}_2 = .219$. El coeficiente estimado $\hat{\phi}_2$ no es significativo ($t = 1.75$) en el nivel del 5%, pero es significativo en el nivel del 10%. Jill decidió mantener este parámetro en su modelo. El cuadrado medio residual es $s^2 = 2,808$, y las autocorrelaciones residuales indicadas en la figura 9-25 están todas comprendidas dentro de sus dos límites de error estándar. Además, los valores de los estadísticos Q de Ljung-Box para los retrasos $m = 12, 24, 36$ y 48 son todos pequeños. Jill concluyó que su modelo era adecuado.

Jill utilizó su modelo para generar los pronósticos de los periodos 66 y 67 como sigue.

A partir del origen del pronóstico $t = 65$, el pronóstico del periodo 66 se genera mediante la ecuación

$$\hat{Y}_t = \hat{\phi}_0 + \hat{\phi}_1 Y_{t-1} + \hat{\phi}_2 Y_{t-2}$$

Función de autocorrelación parcial: Precios de acciones de ISC

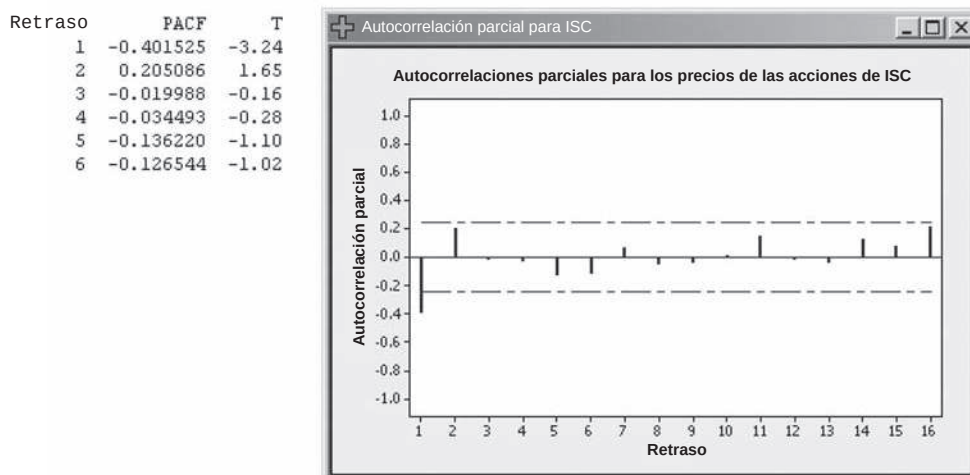


FIGURA 9-24 Autocorrelaciones parciales de la muestra de los precios de las acciones de ISC para el ejemplo 9.7

TABLA 9-13 Resultado de Minitab del modelo AR(2) para los precios al cierre de las acciones para el ejemplo 9.7

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

Period	Forecast	95% Limits		Actual
		Lower	Upper	
66	287.446	183.565	391.328	
67	234.450	125.244	343.656	

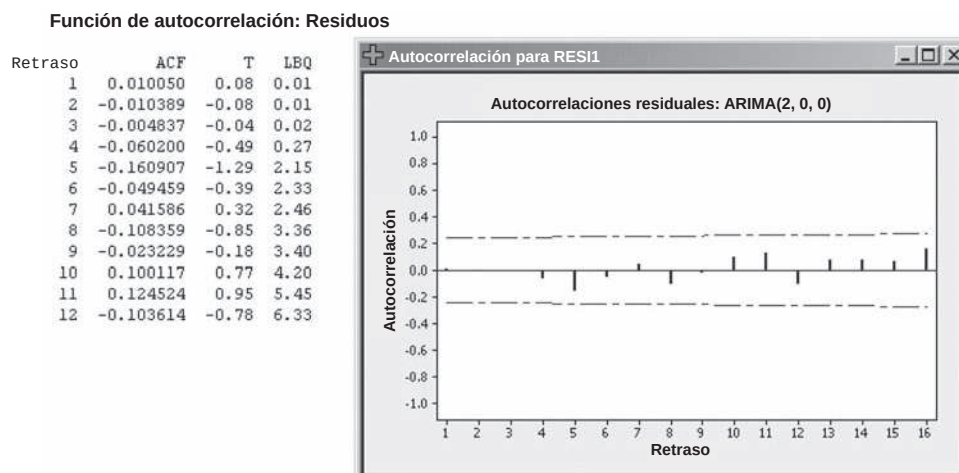


FIGURA 9-25 Función de autocorrelación para los residuos del modelo AR(2); ejemplo 9.7

de manera que

$$\begin{aligned}\hat{Y}_{66} &= 284.9 + (-.324)Y_{65} + .219Y_{64} \\ &= 284.9 - .324(195) + .219(300) = 287.4\end{aligned}$$

De forma similar,

$$\begin{aligned}\hat{Y}_{67} &= 284.9 + (-.324)\hat{Y}_{66} + .219Y_{65} \\ &= 284.9 - .324(287.4) + .219(195) = 234.5\end{aligned}$$

Los resultados concuerdan con los mostrados en el resultado de Minitab de la tabla 9-13. En las series estacionarias, los límites de predicción del 95% son aproximadamente

$$\hat{Y} \pm 2s \quad (9.6)$$

donde $\hat{Y}$ es el pronóstico y s es la raíz cuadrada del error cuadrático medio. Por ejemplo, los límites de predicción aproximadamente del 95% para el periodo 66 son

$$287.4 \pm 2\sqrt{2808} = 287.4 \pm 106 \text{ or } (181.4, 393.4)$$

Este intervalo está cerca del intervalo de 95% dado por Minitab para el periodo 66 en la tabla 9-13.

Advertencias sobre la construcción de modelos

En los modelos ARIMA, *no* es 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 existe duda, comience con un modelo que contenga pocos parámetros en vez de muchos. La necesidad de parámetros adicionales se hará evidente a partir de un examen de las autocorrelaciones residuales y de las autocorrelaciones parciales. Si es evidente un comportamiento MA en las autocorrelaciones residuales y las autocorrelaciones parciales, agregue un parámetro MA y ajuste el modelo en revisión. Si las autocorrelaciones residuales se parecen a las del proceso AR, agregue un término AR y reajuste el modelo.

Los mínimos cuadrados estimados de parámetros autorregresivos y de promedios móviles en los modelos ARIMA tienden a estar altamente correlacionados. Cuando existen más parámetros de los necesarios, esto trae consigo “compensaciones” entre los parámetros y modelos inestables que pueden generar pronósticos defectuosos.

Para resumir, es buena práctica iniciar con un pequeño número de parámetros claramente justificados y agregar parámetros de uno en uno conforme sea necesario. Por otro lado, si los parámetros en el modelo ARIMA ajustado no son significativos (a juzgar por sus razones t), elimine un parámetro cada vez y reajuste el modelo. En virtud de la alta correlación entre los parámetros estimados, se puede dar el caso de que un parámetro que anteriormente no era significativo se convierta en significativo.

Criterios de selección del modelo

Es posible identificar (seleccionar) los modelos ARIMA examinando una gráfica de la serie y al establecer una correspondencia entre los patrones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra con los patrones teóricos conocidos de los procesos ARIMA. Sin embargo, existe cierta subjetividad implicada en este procedimiento, y es posible que dos (o más) modelos iniciales sean congruentes con los patrones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Más aún, después de la estimación y verificación, ambos modelos pueden representar adecuadamente los datos. Si los modelos contienen el mismo número de parámetros, el modelo con menor error cuadrático medio, s^2 , es por lo general el preferido. (Este fue el caso de Lynn Stephens en el ejemplo 9.3). Si los modelos contienen diferentes números de parámetros, el principio de parsimonia conduce a la selección del modelo más sencillo. Sin embargo, el modelo con más parámetros puede tener un error cuadrático medio considerablemente menor.

Se ha desarrollado un método para la selección del modelo que considere tanto el ajuste del modelo como el número de parámetros. El criterio de información de Akaike (1974), o *AIC*, selecciona el mejor modelo de un grupo de modelos candidatos que minimiza

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n} r \quad (9.7)$$

donde

$\ln$ = logaritmo natural

$\hat{\sigma}^2$ = suma residual de cuadrados dividida entre el número de observaciones

n = número de observaciones (residuos)

r = número total de parámetros (incluyendo el término constante) en el modelo ARIMA

El criterio de información bayesiano que desarrolló Schwarz (1978), o *BIC*, selecciona el modelo que minimiza

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n} r \quad (9.8)$$

El segundo término tanto en el criterio *AIC* como en el criterio *BIC* es un factor que “penaliza” la inclusión de parámetros adicionales en el modelo. Como el criterio *BIC* impone una penalización mayor para el número de parámetros que el criterio *AIC*, el uso del *BIC* mínimo para la selección del modelo tendrá como resultado un modelo cuyo número de parámetros no será mayor que el seleccionado por el *AIC*. A menudo los dos criterios producen el mismo resultado.

El *AIC* y el *BIC* deben considerarse como procedimientos adicionales para ayudar en la selección del modelo. No se deben usar como sustitutos de una revisión cuidadosa de las autocorrelaciones de la muestra y las autocorrelaciones parciales.

Ejemplo 9.8

En el ejemplo 9.4 Jim White encontró que dos modelos ARIMA parecían dar una descripción adecuada de las lecturas del proceso de la Atron Corporation. Uno era un modelo AR(1) con $r = 2$ parámetros estimados (incluyendo un término constante) y $\hat{\sigma}^2 = 10,065/75 = 134.2$. El segundo

modelo era un modelo MA(2) con $r = 3$ parámetros (incluyendo un término constante) y $\hat{\sigma}^2 = 9,725/75 = 129.7$. Jim calculó los criterios AIC y BIC como sigue.

$$AIC = \ln(134.2) + \frac{2}{75}(2) = 4.8993 + .0533 = 4.9526$$

AR(1):

$$BIC = \ln(134.2) + \frac{\ln(75)}{75}(2) = 4.8993 + .1151 = 5.0144$$

$$AIC = \ln(129.7) + \frac{2}{75}(3) = 4.8652 + .0800 = 4.9452$$

MA(2):

$$BIC = \ln(129.7) + \frac{\ln(75)}{75}(3) = 4.8652 + .1727 = 5.0379$$

El AIC y el BIC arrojan resultados contrastantes. El AIC es menor para el modelo MA(2), mientras que el BIC es menor para el modelo AR(1). Jim recordaba que el BIC no seleccionaría un modelo con más parámetros que el AIC a causa de la mayor penalización por incluir parámetros adicionales. Jim estaba contento de ver que su selección inicial del modelo AR(1), con base en el principio de parsimonia, estaba apoyada por el criterio BIC .

Modelos ARIMA para datos estacionales

Los datos estacionales tienen un patrón distintivo que se repite cada año. Las ventas mensuales de Cavanaugh Company presentadas en la figura 5.1 registran un máximo al final del otoño (octubre y noviembre). Para datos mensuales con un patrón estacional anual, las observaciones para los mismos meses en años diferentes deben estar correlacionadas. Es decir, enero en un año debe ser similar a enero del año siguiente, febrero de un año debe ser similar a febrero del año siguiente, y así sucesivamente. De esta manera, no sólo hay observaciones dentro del año relacionadas entre sí (correlacionadas), sino que también hay observaciones entre los años que están relacionadas unas con otras (correlacionadas). Si la duración del periodo estacional es S , de manera que $S = 12$ para datos mensuales y $S = 4$ para datos trimestrales, las autocorrelaciones y autocorrelaciones parciales para procesos estacionales son diferentes de cero en retrasos cortos (asociación dentro de un año) y en retrasos que son múltiplos del periodo estacional S (asociación entre años). Las interpretaciones de las autocorrelaciones y de las autocorrelaciones parciales en retrasos estacionales son las mismas que las interpretaciones de las autocorrelaciones y autocorrelaciones parciales en retrasos cortos.

Los modelos estacionales ARIMA contienen términos autorregresivos regulares y de promedios móviles que explican la correlación en retrasos cortos y términos autorregresivos estacionales y de promedio móvil que explican la correlación en retrasos estacionales. Además, para series estacionales no estacionarias, se requiere una diferencia estacional adicional para especificar completamente el modelo.

Los siguientes ejemplos ilustran el desarrollo de modelos estacionales ARIMA. La misma estrategia empleada para la construcción de modelos no estacionales se utiliza para modelos estacionales. Una vez que se ha establecido un modelo adecuado, éste se puede usar para pronosticar.

Ejemplo 9.9

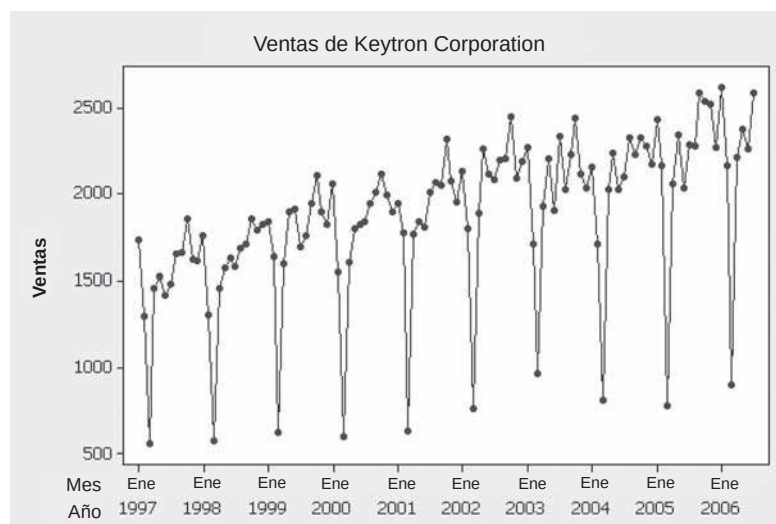
Se le dio la responsabilidad a Kathy Fisher de pronosticar las ventas de Keytron Corporation. Kathy pudo recolectar datos de ventas de 115 meses consecutivos. Los datos iniciaban en enero de 1997 y se presentan en la tabla 9-14.

Kathy examinó la gráfica de la serie de tiempo de la figura 9-26 y notó el pronunciado patrón estacional junto con una tendencia general ascendente. Ella sentía que la serie era no estacionaria y decidió desarrollar un modelo estacional ARIMA para sus datos.

Kathy empezó con el paso de la identificación del modelo examinando la autocorrelación muestral de la figura 9-27. Kathy vio que las autocorrelaciones en los retrasos pequeños parecían cortarse

TABLA 9-14 Ventas de Keytron para el ejemplo 9.9 (léase por columnas)

1,736.8	1,627.6	1,895.4	1,768.0	2,202.2	1,708.2	2,288.0
1,297.4	1,575.6	1,822.6	1,840.8	2,449.2	806.0	2,275.0
559.0	1,682.2	2,054.0	1,804.4	2,090.4	2,028.0	2,581.8
1,455.6	1,710.8	1,544.4	2,007.2	2,184.0	2,236.0	2,540.2
1,526.2	1,853.8	600.6	2,067.0	2,267.2	2,028.0	2,519.4
1,419.6	1,788.8	1,604.2	2,048.8	1,705.6	2,100.8	2,267.2
1,484.6	1,822.4	1,796.6	2,314.0	962.0	2,327.0	2,615.6
1,651.0	1,838.2	1,822.6	2,072.2	1,929.2	2,225.6	2,163.2
1,661.4	1,635.4	1,835.6	1,952.6	2,202.2	2,321.8	899.6
1,851.2	618.8	1,944.8	2,134.6	1,903.2	2,275.0	2,210.0
1,617.2	1,593.8	2,009.8	1,799.2	2,337.4	2,171.0	2,376.4
1,614.6	1,898.0	2,116.4	756.6	2,022.8	2,431.0	2,259.4
1,757.6	1,911.0	1,994.2	1,890.2	2,225.6	2,165.8	2,584.4
1,302.6	1,695.0	1,895.4	2,256.8	2,441.4	780.0	
572.0	1,757.6	1,947.4	2,111.2	2,113.8	2,056.6	
1,458.6	1,944.8	1,770.6	2,080.0	2,035.8	2,340.0	
1,567.8	2,108.6	626.6	2,191.8	2,152.8	2,033.2	

**FIGURA 9-26 Ventas de Keytron para el ejemplo 9.9**

después del retraso 1, a pesar de que había un pequeño margen en el retraso 3. Katy también vio que las autocorrelaciones en los retrasos estacionales $-12, 24$, y 36 (que no se muestran) — eran grandes y no se desvanecían rápidamente. Esto sugería que la serie era no estacionaria y confirmaba la impresión que Kathy tenía de la serie de tiempo. Antes de continuar, Kathy decidió diferenciar la serie con respecto al retraso estacional para ver si podía convertir la serie no estacionaria en una estacionaria.

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 se puede calcular es

$$Y_{13} - Y_1 = 1757.6 - 1736.8 = 20.8$$

Se calculan las diferencias estacionales restantes, y la serie de las diferencias estacionales de las ventas se presentan en la figura 9-28.

Función de autocorrelación: Ventas

Retraso	ACF	T	LBQ
1	0.431371	4.63	21.96
2	0.153911	1.41	24.78
3	0.237841	2.14	31.58
4	0.128583	1.11	33.58
5	0.041549	0.36	33.79
6	0.061352	0.53	34.26
7	0.028549	0.24	34.36
8	0.136014	1.16	36.69
9	0.212966	1.80	42.44
10	0.105905	0.87	43.88
11	0.332651	2.71	58.20
12	0.791934	6.08	140.12
13	0.324223	1.94	153.99
14	0.067091	0.39	154.59
15	0.127275	0.74	156.77
16	0.039851	0.23	156.99
17	-0.031030	-0.18	157.12
18	-0.012252	-0.07	157.14
19	-0.049556	-0.29	157.48
20	0.041444	0.24	157.73
21	0.107932	0.62	159.39
22	0.026250	0.15	159.49
23	0.227137	1.30	167.04
24	0.635500	3.59	226.75
25	0.238127	1.22	235.23
26	0.000420	0.00	235.23

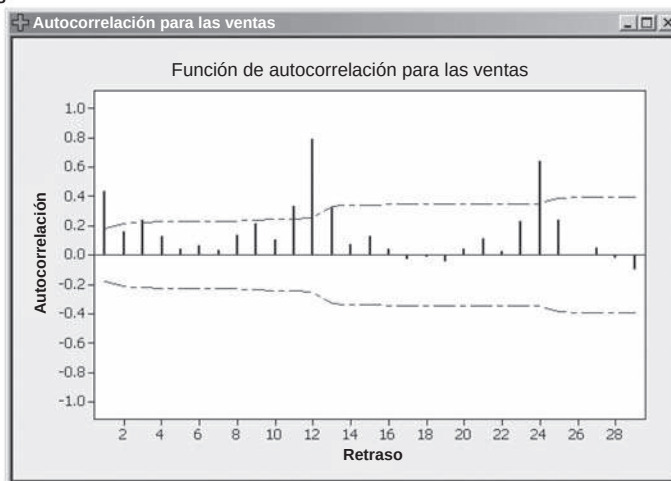


FIGURA 9-27 Autocorrelaciones muestrales de las ventas de Keytron para el ejemplo 9.9

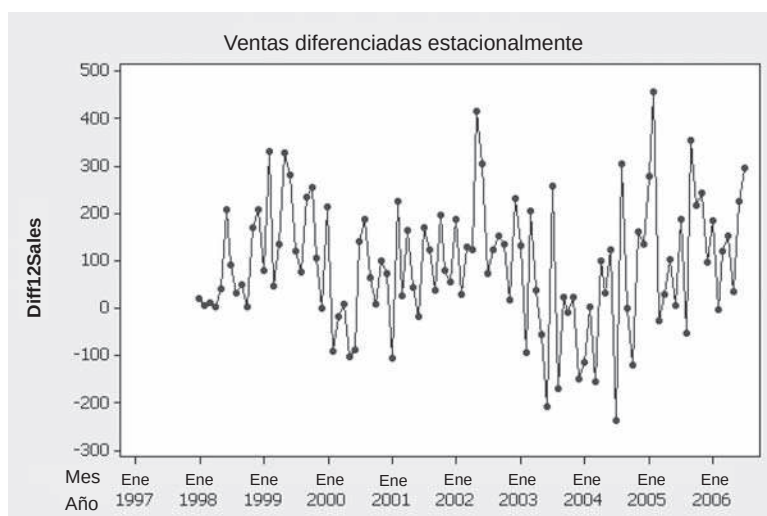


FIGURA 9-28 Ventas de Keytron diferenciadas estacionalmente para el ejemplo 9.9

Las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra de la serie diferenciada se presentan en las figuras 9-29 y 9-30, respectivamente. La figura 9-28 indica que los datos diferenciados de forma estacional son estacionarios y parecen variar aproximadamente alrededor del nivel 100. Las autocorrelaciones (figura 9-29) tienen un punto significativo en el retraso 12 (se corta), y las autocorrelaciones parciales de la muestra (figura 9-30) tienen puntos significativos en los retrasos 12 y 24 que se hacen progresivamente más pequeños (se desvanecen). Este comportamiento sugiere un término MA(1) en el retraso estacional 12.

Kathy identificó un modelo ARIMA(0, 0, 0)(0, 1, 1)₁₂ para sus datos de ventas. Esta notación significa que el modelo tiene

- $p = 0$ términos autorregresivos regulares
- $d = 0$ diferencias regulares
- $q = 0$ términos regulares de promedio móvil
- $P = 0$ términos autorregresivos estacionales

Función de autocorrelación: Diff12Sales

Retraso	ACF	T	LBQ
1	0.082290	0.84	0.72
2	0.164373	1.66	3.61
3	0.206938	2.03	8.24
4	0.128010	1.21	10.03
5	0.018390	0.17	10.07
6	0.117004	1.09	11.60
7	0.156824	1.44	14.37
8	0.064878	0.59	14.85
9	0.074454	0.67	15.48
10	0.028748	0.26	15.58
11	0.023747	0.21	15.65
12	-0.379941	-3.40	32.80
13	-0.009173	-0.07	32.81
14	-0.111889	-0.90	34.33
15	-0.159221	-1.28	37.45
16	-0.146828	-1.16	40.13
17	0.000677	0.01	40.13
18	-0.161582	-1.26	43.45
19	-0.144594	-1.11	46.14
20	-0.178727	-1.36	50.31
21	-0.173436	-1.29	54.27
22	-0.068707	-0.50	54.90
23	-0.116692	-0.85	56.74
24	-0.109328	-0.79	58.38
25	-0.112515	-0.81	60.14
26	-0.000583	-0.00	60.14

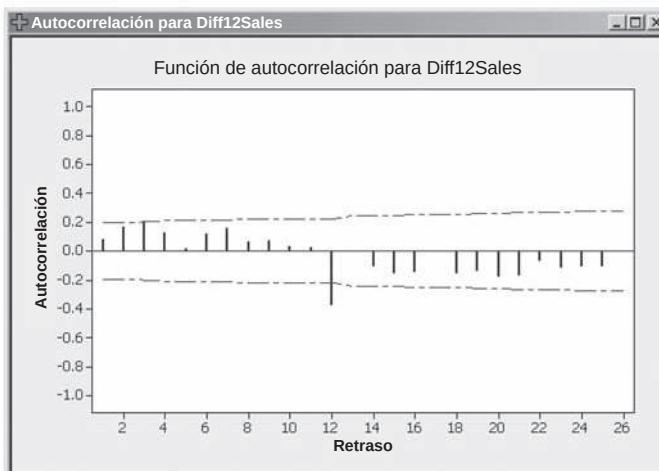


FIGURA 9-29 Autocorrelaciones muestrales de las ventas diferenciadas estacionalmente de Keytron para el ejemplo 9.9

Función de autocorrelación parcial: Diff12Sales

Retraso	PACF	T
1	0.082290	0.84
2	0.158675	1.61
3	0.188491	1.91
4	0.086124	0.87
5	-0.053533	-0.54
6	0.052428	0.53
7	0.125409	1.27
8	0.033422	0.34
9	0.007631	0.08
10	-0.056774	-0.58
11	-0.024356	-0.25
12	-0.436449	-4.43
13	-0.022329	-0.23
14	-0.036679	-0.37
15	-0.036390	-0.37
16	-0.091175	-0.93
17	0.048016	0.49
18	-0.036175	-0.37
19	0.012596	0.13
20	-0.113143	-1.15
21	-0.037869	-0.38
22	0.089451	0.91
23	0.024185	0.25
24	-0.248846	-2.53
25	-0.064072	-0.65
26	0.063984	0.65

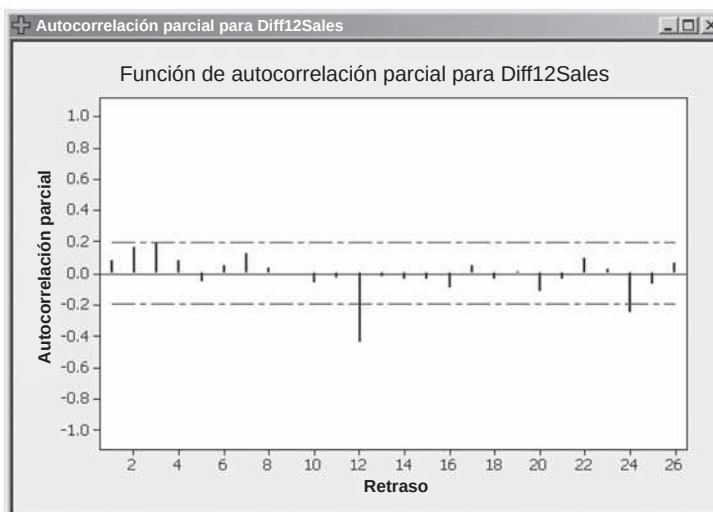


FIGURA 9-30 Autocorrelaciones parciales de la muestra de las ventas diferenciadas estacionalmente de Keytron para el ejemplo 9.9

$D = 1$ diferencia estacional en el retraso 12

$Q = 1$ término de promedio móvil estacional en el retraso 12

Puesto que la serie diferenciada estacionalmente varía alrededor de un nivel diferente de cero, Kathy incluyó un término constante en su modelo. El modelo identificado tiene la forma

$$Y_t - Y_{t-12} = \mu + \varepsilon_t - \Omega_1 \varepsilon_{t-12} \quad (9.9)$$

donde μ es el nivel promedio del proceso diferenciado estacionalmente y Ω_1 es el parámetro de promedio móvil estacional.

TABLA 9-15 Resultado de Minitab del modelo ARIMA(0, 0, 0)(0, 1, 1)₁₂ de las ventas de Keytron para el ejemplo 9.9

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

Period	Forecast	95% Limits		Actual
		Lower	Upper	
116	2419.69	2208.46	2630.93	
117	2504.31	2293.07	2715.54	

Kathy empleó Minitab para estimar los parámetros de su modelo, calcular las autocorrelaciones residuales y generar pronósticos. El resultado se presenta en la tabla 9-15. Las autocorrelaciones residuales aparecen en la figura 9-31, y los pronósticos para los siguientes 12 meses están incorporados al final de la serie de ventas en la figura 9-32.

Kathy estaba contenta de ver que su modelo inicial se ajustaba muy bien a los datos. Los estadísticos chi cuadrada de Ljung-Box, para los grupos de retrasos $m = 12, 24, 36$, y 48 , eran claramente no significativos, como lo evidenciaban los grandes valores de p . Las autocorrelaciones residuales eran uniformemente pequeñas sin un patrón aparente. Kathy estaba lista para usar su modelo ajustado para verificar los pronósticos generados por Minitab.

Función de autocorrelación: Residuos

Retraso	ACF	T	LBQ
1	0.038808	0.39	0.16
2	0.144138	1.46	2.38
3	0.232836	2.31	8.25
4	0.042162	0.40	8.44
5	0.040750	0.38	8.62
6	0.058868	0.55	9.01
7	0.083296	0.78	9.79
8	0.030700	0.29	9.90
9	0.146424	1.37	12.37
10	0.038860	0.36	12.54
11	0.000928	0.01	12.54
12	0.001198	0.01	12.54

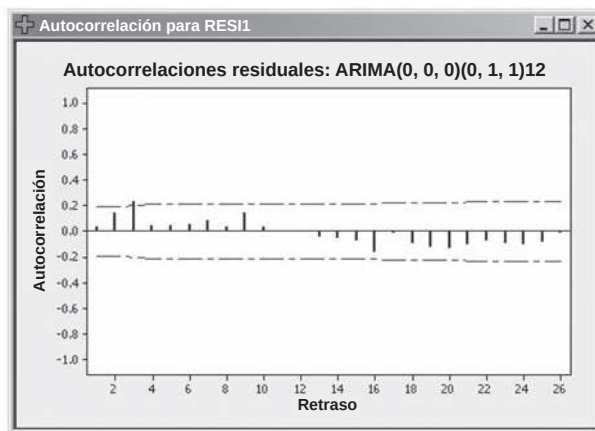


FIGURA 9-31 Función de autocorrelación para los residuos del modelo ARIMA(0, 0, 0)(0, 1, 1)₁₂ para las ventas de Keytron; ejemplo 9.9

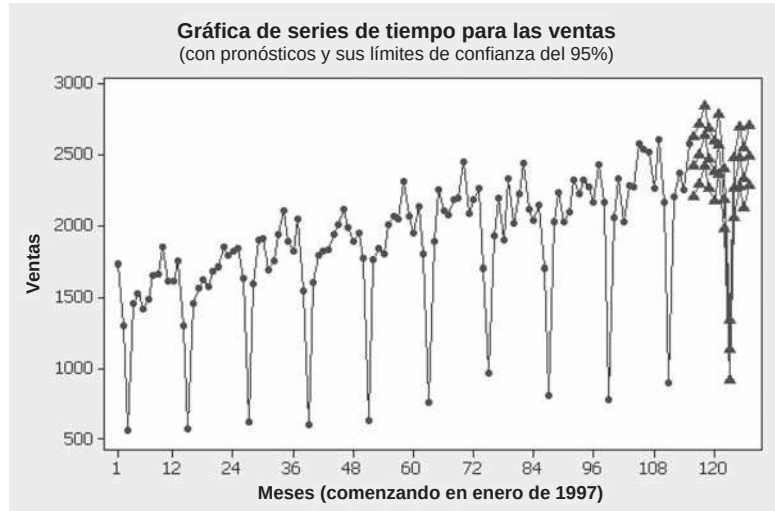


FIGURA 9-32 Ventas y pronósticos para los 12 meses siguientes para Keytron; ejemplo 9.9

Los parámetros estimados fueron $\hat{\mu} = 85.457$ y $\hat{\Omega}_1 = .818$. Usando estos valores, es posible despejar Y_t en la ecuación 9.9 para obtener

$$Y_t = Y_{t-12} + 85.457 + \varepsilon_t - .818\varepsilon_{t-12}$$

Para el pronóstico de ventas del periodo 116, Kathy estableció que $t = 116$ y notó que, para el tiempo que estaba haciendo el pronóstico, su mejor opción de ε_{16} (el error del siguiente periodo) era cero. Por lo tanto, la ecuación del pronóstico es

$$\hat{Y}_{116} = Y_{104} + 85.457 - .818e_{104}$$

donde $e_{104} = -72.418$ es el residuo (el error estimado) del periodo 104. Así,

$$\hat{Y}_{116} = 2,275 + 85.457 - .818(-72.418) = 2,419.7$$

De forma similar,

$$\hat{Y}_{117} = Y_{105} + 85.457 - .818e_{105}$$

$$\hat{Y}_{117} = 2,581.8 + 85.457 - .818(199.214) = 2,504.3$$

Los pronósticos de Kathy concordaron con los resultados de Minitab. Ella estaba convencida de que comprendía cómo usar el modelo ajustado para generar pronósticos. Kathy estaba satisfecha con los pronósticos de ventas para los siguientes 12 meses mostrados en la figura 9-32. Los pronósticos parecían ser completamente congruentes con el comportamiento de la serie. Ella pensaba que había capturado el patrón estacional y el crecimiento de las ventas bastante bien.

Ejemplo 9.10

En el ejemplo 3.5, Perkin Kendell, analista de Coastal Marine Corporation, utilizó el análisis de autocorrelación para concluir que las ventas trimestrales eran estacionales. Decidió pronosticar las ventas para 2007 usando la metodología Box-Jenkins. Los datos se graficaron en la figura 3.14. La gráfica de la serie de tiempo tiene un patrón estacional muy claro con quizás una ligera tendencia ascendente.

Perkin inició su análisis examinando las autocorrelaciones de la muestra de su serie original y las diversas diferencias de la serie. Un amigo de Perkin, un experto en la metodología Box-Jenkins, le

dijo que era una buena idea siempre que hubiera una posibilidad de que la serie fuera no estacionaria. Una vez que el modelo más simple sugerido por un patrón de autocorrelación se ha identificado tentativamente, la identificación se puede confirmar examinando las autocorrelaciones parciales de la muestra. Perkin había calculado con Minitab las funciones de autocorrelación para las siguientes series:

$$\begin{aligned}
 Y_t &= \text{serie de ventas originales} \\
 \Delta Y_t &= Y_t - Y_{t-1} = \text{serie de las primeras (regulares) diferencias} \\
 \Delta_4 Y_t &= Y_t - Y_{t-4} = \text{serie de las diferencias estacionales de orden } S = 4 \\
 \left[\begin{aligned} \Delta \Delta_4 Y_t &= \Delta(\Delta_4 Y_t) = \Delta(Y_t - Y_{t-4}) \\ &= Y_t - Y_{t-1} - Y_{t-4} + Y_{t-5} \end{aligned} \right] &= \text{serie integrada con una diferencia estacional} \\
 &\quad \text{seguida por una diferencia regular}
 \end{aligned}$$

A veces, es necesario tener una diferencia estacional y una diferencia regular antes de obtener como resultado una serie estacionaria.

Las autocorrelaciones de la muestra para la serie original, para la serie diferenciada estacionalmente, y para la serie con una diferencia regular y otra estacional se presentan en las figuras 9-33 (también en la figura 3-15), 9-34 y 9-35.

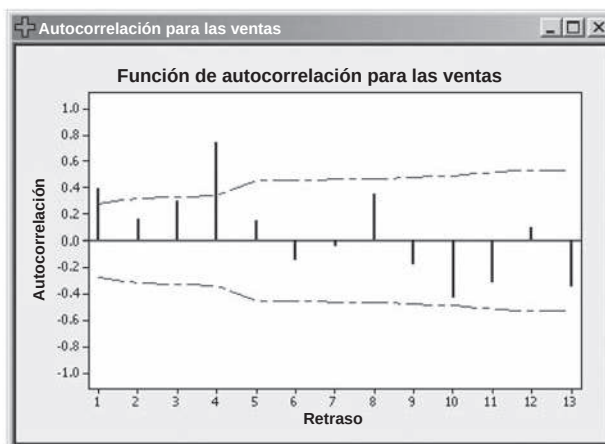
La función de autocorrelación de la serie original tiene picos (significativamente) grandes en los retrasos 1 y 4. Sin embargo, las autocorrelaciones en los retrasos estacionales 4, 8 y 12 decaen hacia cero. Esto puede indicar una serie no estacionaria y la necesidad de una diferencia estacional. El intento de Perkin por ajustar un modelo ARIMA a la serie original con un término constante, un término regular de promedio móvil y un término autorregresivo estacional no tuvo éxito. El parámetro autorregresivo estacional estimado resultó cercano a uno (lo que implicaba que era necesaria una diferencia estacional), y las autocorrelaciones residuales eran generalmente grandes e incongruentes con errores aleatorios. Perkin volvió en seguida a la función de autocorrelación de la serie diferenciada estacionalmente de la figura 9-34.

Las autocorrelaciones de los datos diferenciados estacionalmente son grandes en los retrasos cortos y declinan más bien lentamente en la forma de un patrón parecido a una onda. Perkins concluyó que la serie aún podía ser no estacionaria y pensó que además podría requerirse una diferencia regular además de la diferencia estacional antes de alcanzar la estacionalidad. Por consiguiente, generó en Minitab las autocorrelaciones muestrales para las ventas con primeras diferencias y las ventas diferenciadas estacionalmente, que aparecen en la figura 9-35.

FIGURA 9-33 Autocorrelaciones muestrales de las ventas de Coastal Marine para el ejemplo 9.10

Función de autocorrelación: Ventas

Retraso	ACF	T	LBQ
1	0.392843	2.83	8.50
2	0.153948	0.97	9.83
3	0.293739	1.82	14.77
4	0.743533	4.34	47.11
5	0.150619	0.67	48.47
6	-0.153014	-0.67	49.90
7	-0.047049	-0.21	50.04
8	0.346979	1.51	57.72
9	-0.182597	-0.76	59.90
10	-0.434737	-1.80	72.53
11	-0.315090	-1.23	79.33
12	0.091175	0.35	79.92
13	-0.353272	-1.34	88.90



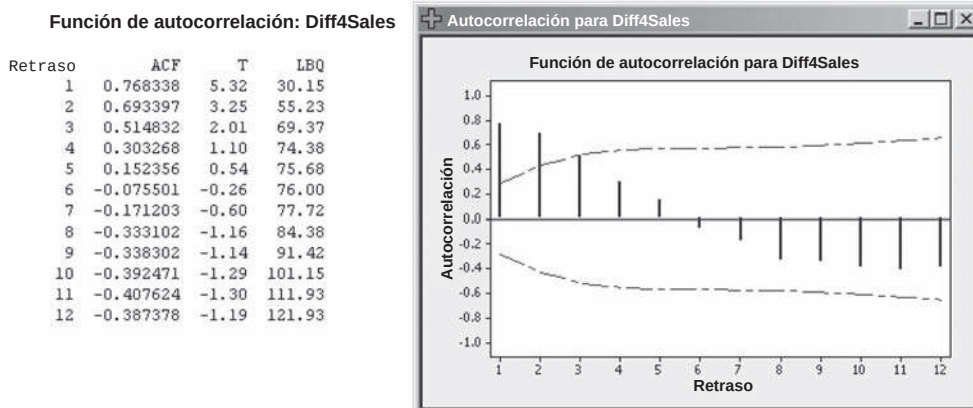


FIGURA 9-34 Autocorrelaciones muestrales de las ventas diferenciadas estacionalmente de Coastal Marine para el ejemplo 9.10

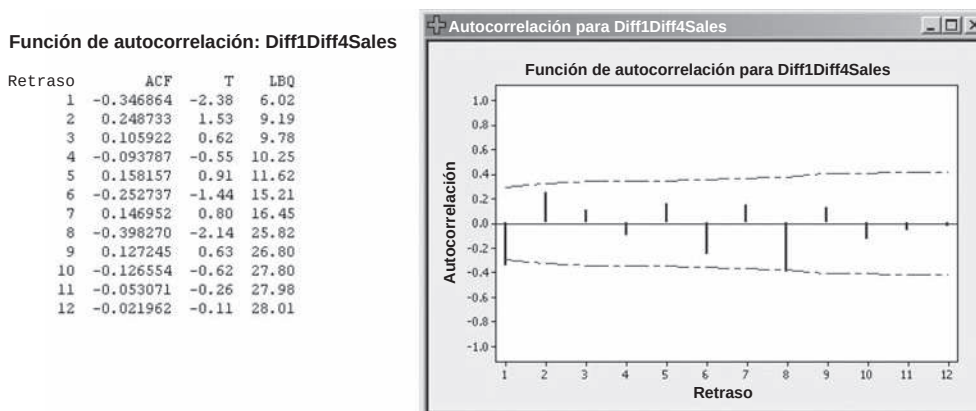


FIGURA 9-35 Autocorrelaciones muestrales de las primeras diferencias de las ventas y para las ventas diferenciadas estacionalmente de Coastal Marine; ejemplo 9.10

Al examinar las autocorrelaciones de la figura 9-35, Perkin observó sólo dos valores significativos, los de los retrasos 1 y 8. Además, vio que las autocorrelaciones de los primeros dos retrasos alternaban en signo. Perkin pensó que un modelo ARIMA con un término autorregresivo regular y quizás un término de promedio móvil estacional en el retraso 8 podría ser adecuado. Antes de continuar, Perkin quería confirmar su elección inicial examinando las autocorrelaciones parciales para la serie de primeras diferencias y las diferencias estacionales.

Las autocorrelaciones parciales de la muestra de las primeras diferencias y para las diferencias estacionales se presentan en la figura 9-36. Las autocorrelaciones parciales parecen cortarse después del primer retraso, en consonancia con el comportamiento AR(1) que Perkin observó en las autocorrelaciones, y existe una autocorrelación parcial significativa en el retraso 8.

Perkin no estaba seguro de cómo manejar los parámetros estacionales, de manera que decidió mantenerlos simples y ajustar un modelo ARIMA(1, 1, 0)(0, 1, 0)₄. Este modelo permite una diferencia regular, una diferencia estacional y un término autorregresivo regular, pero no los coeficientes autorregresivos estacionales o de promedios móviles. Una gráfica de la serie después de la diferenciación indica que varía alrededor de cero, de manera que no se incluye en el modelo el término cons-

Función de autocorrelación parcial: Diff1Diff4Sales

Retraso	PACF	T
1	-0.346864	-2.38
2	0.145982	1.00
3	0.267429	1.83
4	-0.026932	-0.18
5	0.036653	0.25
6	-0.238502	-1.64
7	-0.020543	-0.14
8	-0.376340	-2.58
9	-0.043510	-0.30
10	0.012796	0.09
11	0.124917	0.86
12	-0.111998	-0.77

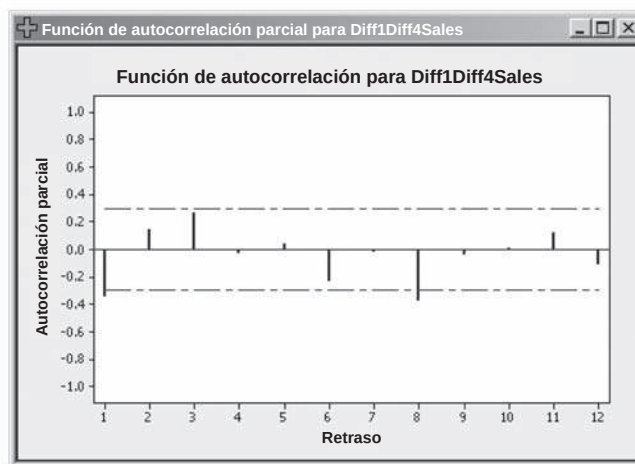


FIGURA 9-36 Autocorrelaciones parciales muestrales de las primeras diferencias de ventas y para las ventas diferenciadas estacionalmente de Coastal Marine; ejemplo 9.10

tante. Perkin razonó que, si fueran necesarios los parámetros estacionales, lo pondría de manifiesto la evidencia en las autocorrelaciones residuales de su modelo inicial.

El resultado del modelo $ARIMA(1, 1, 0)(0, 1, 0)_4$ se presenta en la parte superior de la tabla 9-16, y las autocorrelaciones residuales se muestran en la parte superior de la figura 9-37. El modelo parece ajustar bastante bien con un coeficiente $AR(1)$ estimado de $\phi_1 = .352$ y un error cuadrático medio de los residuos $s^2 = 1,040.7$. Sin embargo, se presenta un pico significativo en el retraso 8 en las autocorrelaciones residuales, y el estadístico de Ljung-Box para los primeros $m = 12$ retrasos es $Q = 19.7$ con un valor p de .05. Esto último sugiere que, como grupo, las primeras 12 autocorrelaciones residuales son más grandes de lo que se esperaba para errores aleatorios.

Perkin decidió modificar su modelo inicial e incluyó términos de promedio móvil correspondientes a los retrasos estacionales 4 y 8. La evidencia sugiere que sólo el término del retraso 8 es necesario; sin embargo, el programa Minitab requiere la inclusión de los parámetros estacionales AR o MA en el modelo hasta, e incluyendo, el múltiplo más alto del retraso estacional que sea necesario. En este caso, el retraso estacional es $S = 4$, y se hace necesario un parámetro de promedio móvil en el retraso $2 \times 4 = 8$.

Perkin usó Minitab para ajustar un modelo $ARIMA(1, 1, 0)(0, 1, 2)_4$ a sus datos.¹² El resultado se presenta en la tabla 9-16 (abajo) y las autocorrelaciones residuales aparecen en la figura 9-37 (abajo). Perkin estaba contento con los resultados. El error cuadrático medio de los residuos se redujo a $s^2 = 716.5$, y las autocorrelaciones residuales, a juzgar por los valores individuales y los estadísticos chi cuadrada de Ljung-Box, son congruentes con los errores aleatorios. Las gráficas residuales (no se muestran) sugieren que los supuestos del otro término de error son apropiados.

Para obtener experiencia adicional con la metodología Box-Jenkins, Perkin quería verificar los pronósticos de ventas para los dos trimestres siguientes que se presentan en el resultado de Minitab. Usando parámetros estimados en vez de los valores verdaderos, el modelo final es

$$W_t = -.351W_{t-1} + \varepsilon_t - .239\varepsilon_{t-4} - .672\varepsilon_{t-8}$$

donde

$$W_t = \Delta\Delta_4 Y_t = Y_t - Y_{t-1} - Y_{t-4} + Y_{t-5}$$

¹²Los comandos de Minitab para correr el modelo $ARIMA(1, 1, 0)(0, 1, 2)_4$ se describen al final de este capítulo.

TABLA 9-16 Resultado de Minitab para el modelo ARIMA(1, 1, 0)(0, 1, 0)₄ (arriba) y el modelo ARIMA(1, 1, 0)(0, 1, 2)₄ (abajo) de las ventas de Coastal Marine; ejemplo 9.10

ARIMA (1,1,0)(0,1,0)₄ Model: Coastal Marine Sales

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.3517	0.1384	-2.54	0.015

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 52, after differencing 47

Residuals: SS = 47874.1 (backforecasts excluded)

MS = 1040.7 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	0.421	0.610	*

ARIMA (1, 1, 0)(0,1,2)₄ Model: Coastal Marine Sales

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.3505	0.1423	-2.46	0.018
SMA 4	0.2392	0.1340	1.78	0.081
SMA 8	0.6716	0.1404	4.78	0.000

Differencing: 1 regular, 1 seasonal of order 4

Number of observations: Original series 52, after differencing 47

Residuals: SS = 31525.4 (backforecasts excluded)

MS = 716.5 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	0.928	*

Forecasts from period 52

Period	Forecast	95% Limits		Actual
		Lower	Upper	
53	169.401	116.926	221.875	
54	230.617	168.047	293.188	

es la serie diferenciada adecuadamente. En el origen del pronóstico $t - 1$, la ecuación del pronóstico, escrita en términos de las Y , toma 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 residuos (errores estimados) para los periodos $t - 4$ y $t - 8$, respectivamente.

El pronóstico de ventas para el periodo 53 es

$$\begin{aligned}\hat{Y}_{53} &= Y_{52} + Y_{49} - Y_{48} - .351(Y_{52} - Y_{51} - Y_{48} + Y_{47}) - .239e_{49} - .672e_{45} \\ \hat{Y}_{53} &= 281.4 + 232.1 - 338.2 - .351(281.4 - 291.0 - 338.2 + 329.6) \\ &\quad - .239(-1.726) - .672(18.898) \\ &= 169.4\end{aligned}$$

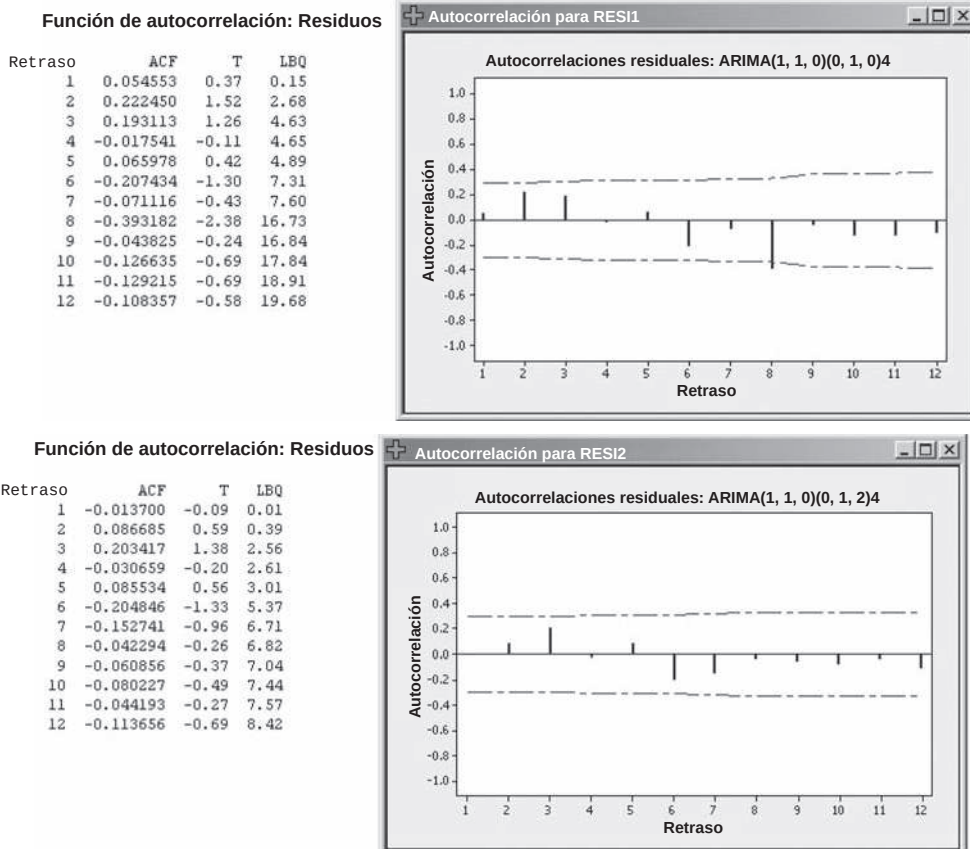


FIGURA 9-37 Autocorrelaciones de los residuos para el modelo $ARIMA(1, 1, 0)(0, 1, 0)_4$ (arriba) y autocorrelaciones de los residuos del modelo $ARIMA(1, 1, 0)(0, 1, 2)_4$ (abajo) de las ventas de Coastal Marine; ejemplo 9.10

y el pronóstico del periodo 54 es

$$\begin{aligned}\hat{Y}_{54} &= \hat{Y}_{53} + Y_{50} - Y_{49} - .351(\hat{Y}_{53} - Y_{52} - Y_{49} + Y_{48}) - .239e_{50} - .672e_{46} \\ \hat{Y}_{54} &= 169.4 + 285.6 - 232.1 - .351(169.4 - 281.4 - 232.1 + 338.2) \\ &\quad - .239(-37.040) - .672(4.781) \\ &= 230.6\end{aligned}$$

Perkin estaba contento de ver que sus pronósticos calculados coincidían con los generados por Minitab. Los pronósticos para los siguientes cuatro trimestres están graficados en la figura 9-38. Los pronósticos parecen razonables. El patrón de los pronósticos es muy parecido al patrón histórico. Perkin sintió que su esfuerzo para elaborar el pronóstico tuvo éxito y hacía planes para recomendar la metodología Box-Jenkins a sus colegas.

Suavización exponencial simple y el modelo ARIMA

Varios modelos ARIMA generan pronósticos iguales o aproximadamente iguales a los que resultaron de los métodos de suavización estudiados 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 origen del pronóstico es t y que se requiere un pronóstico Y_{t+1} . Al sustituir t por $t + 1$ en la ecuación 9.10 y al despejar Y_{t+1} tenemos

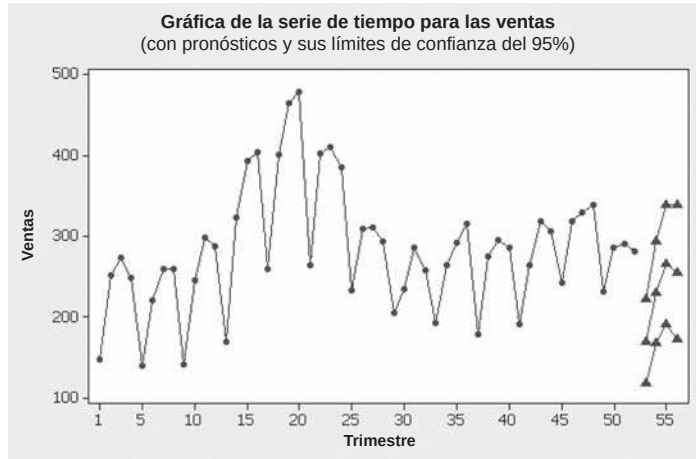


FIGURA 9-38 Pronósticos de los siguientes cuatro trimestres del modelo $ARIMA(1, 1, 0)(0, 1, 2)_4$ de las ventas de Coastal Marine para el ejemplo 9-38

$$Y_{t+1} = Y_t + \varepsilon_{t+1} - \omega_1 \varepsilon_t$$

Debido a que, para el tiempo t , la mejor opción de ε_{t+1} es cero y ε_t es estimada por el residuo $\varepsilon_t = Y_t - \hat{Y}_t$, la ecuación del pronóstico es

$$\hat{Y}_{t+1} = Y_t - \omega_1(Y_t - \hat{Y}_t) = (1 - \omega_1)Y_t + \omega_1\hat{Y}_t \quad (9.11)$$

Dejemos que $\alpha = 1 - \omega_1$, y la ecuación 9.11 es idéntica a la ecuación de suavización exponencial simple 4.13:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t$$

La elaboración de un pronóstico usando la suavización exponencial simple es equivalente a generar pronósticos usando un 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. La suavización exponencial simple debería trabajar bien en las series de tiempo que pueden describirse adecuadamente por un modelo $ARIMA(0, 1, 1)$. Para aquellas series de tiempo que no pueden describirse de manera adecuada mediante este modelo, los pronósticos generados por la suavización exponencial no son particularmente buenos.

En la suavización exponencial simple, el parámetro α generalmente está restringido al rango $0 < \alpha < 1$. El parámetro de promedio móvil ω_1 en el modelo $ARIMA(0, 1, 1)$ está restringido al rango $-1 < \omega_1 < 1$, entonces, en sentido estricto, los dos métodos de pronósticos son equivalentes para valores positivos de los parámetros α y ω_1 .

Ventajas y desventajas de los modelos ARIMA

La metodología Box-Jenkins para el análisis de las series de tiempo es una herramienta muy poderosa para generar pronósticos exactos de corto alcance. Los modelos ARIMA son bastante flexibles y pueden representar un amplio rango de características de las series de tiempo que se presentan en la práctica. Existen modelos formales para probar la idoneidad de un modelo. Más aún, los pronósticos y los intervalos de predicción provienen directamente del modelo ajustado.

Sin embargo, los modelos ARIMA tienen ciertas desventajas.

1. Se requiere una cantidad relativamente grande de datos. Se debe reconocer que, si los datos son estacionales con, digamos, un periodo estacional $S = 12$, entonces las observaciones de

un año constituyen esencialmente un punto de datos (véase el patrón estacional), y no 12. En términos generales, para datos no estacionales, se requieren 40 observaciones o más para desarrollar un modelo ARIMA. Para datos estacionales se requieren alrededor de 6 a 10 años de datos —dependiendo de la duración del periodo estacional— para construir un modelo ARIMA.

2. No existen caminos fáciles para actualizar los parámetros de un modelo ARIMA conforme se dispone de nuevos datos, como sucede con algunos métodos de suavización. Periódicamente el modelo tiene que reajustarse por completo, y algunas veces se debe desarrollar un modelo nuevo.
3. La elaboración de un modelo ARIMA satisfactorio requiere una gran inversión de tiempo y otros recursos. Los costos de la manipulación de datos y el desarrollo del modelo son sustancialmente más altos para los modelos ARIMA que para técnicas más tradicionales de elaboración de pronósticos, como la suavización.

APLICACIÓN A LA ADMINISTRACIÓN

De acuerdo con Bernstein (1996), los pronósticos son uno de los más importantes insumos que desarrollan los gerentes para ayudarse en el proceso de toma de decisiones. Prácticamente, cada decisión operativa importante depende en algún grado de un pronóstico. La acumulación de inventario está relacionada con el pronóstico de la demanda esperada; el departamento de producción tiene que programar las necesidades de personal y pedidos de materia prima para el próximo mes o los dos siguientes; el departamento de finanzas debe preparar el financiamiento de corto plazo para el siguiente trimestre; el departamento de personal debe definir los requerimientos de contratación y despidos. La lista de aplicaciones de los pronósticos es bastante larga.

Los ejecutivos están muy conscientes de la importancia de los pronósticos. Desde luego, pasan una gran parte de tiempo estudiando las tendencias económicas, los asuntos políticos y la forma como algunos sucesos pueden afectar la demanda de productos y/o servicios. Un tema de interés es la importancia que dan los ejecutivos a los métodos de pronósticos cuantitativos contra los pronósticos basados sólo en sus criterios. Este asunto es esencialmente sensible cuando están implicados acontecimientos que tienen un efecto significativo en la demanda. Un problema con los métodos cuantitativos de pronósticos es que dependen de los datos históricos. Por esta razón, probablemente son menos efectivos al considerar que con frecuencia los resultados implican una demanda notoriamente más alta o más baja.

Un problema que encaran con frecuencia los gerentes es la necesidad de elaborar pronósticos de corto plazo de un buen número de diferentes artículos. Un ejemplo común es el gerente que debe programar la producción con base en algún pronóstico de demanda de varios cientos de artículos diferentes en la misma línea de producción. Las técnicas que se usan con más frecuencia en esta situación son los métodos de suavización.

Las principales ventajas de la suavización exponencial son su bajo costo y su simplicidad, aunque tal vez no sea tan exacta como los métodos más complejos, como los modelos ARIMA en general. Sin embargo, cuando se necesitan pronósticos para sistemas de inventario que tienen miles de artículos, los métodos de suavización a menudo son la única opción razonable.

El pronóstico exacto desde la perspectiva de una serie de tiempo depende de la suposición de que el futuro probablemente será como el pasado y de que los patrones pasados se puedan describir adecuadamente. Las técnicas de las series de tiempo son más útiles para pronosticar variables que probablemente seguirán fluctuando de acuerdo con ciertos patrones estables en el futuro.

La metodología Box-Jenkins es una herramienta muy poderosa para elaborar pronósticos exactos de corto plazo. Los gerentes deben estar conscientes de que la construcción satisfactoria de un modelo ARIMA con la técnica Box-Jenkins requiere una gran cantidad de datos históricos y mucho tiempo por parte del analista.

Son muchas las aplicaciones de la metodología Box-Jenkins. Los modelos ARIMA se usan para

- Calcular un cambio en la estructura de precios de la industria telefónica de Estados Unidos
- Investigar la relación entre la concentración de amonio, la tasa de flujo y la temperatura de los ríos
- Pronosticar el volumen accionario anual
- Pronosticar la cantidad de pozos petroleros en actividad
- Analizar el número de unidades de viviendas privadas
- Analizar las observaciones diarias de avances porcentuales del número de acciones negociadas
- Analizar la competencia entre rutas de ferrocarril y aerolíneas
- Pronosticar el empleo
- Analizar un gran número de series de tiempo de energía de una compañía de servicio público
- Analizar el efecto de las promociones sobre las ventas de productos de consumo
- Pronosticar diferentes categorías de aseguramiento de calidad en productos

Glosario

Metodología Box-Jenkins. Se refiere a un conjunto de procedimientos para identificar, ajustar y verificar modelos ARIMA con datos de las series de tiempo. Los pronósticos se derivan directamente de la forma del modelo ajustado.

Principio de parsimonia. Se refiere a la preferencia por los modelos simples por encima de 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 ARIMA(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 cuadrático medio de los residuos

$$s^2 = \frac{\sum_{t=1}^n e_t^2}{n-r} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n-r} \quad (9.4)$$

Estadístico Q de Ljung-Box

$$Q = n(n+2) \sum_{k=1}^m \frac{r_k^2(e)}{n-k} \quad (9.5)$$

Intervalos aproximados de predicción del 95% para la serie estacionaria

$$\hat{Y} \pm 2s \quad (9.6)$$

Criterio de información de Akaike

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n}r \quad (9.7)$$

Criterio de información bayesiano (Schwarz) (BIC)

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n}r \quad (9.8)$$

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 pronóstico del modelo ARIMA(0, 1, 1) en forma de suavización 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

- Para una muestra de 100 observaciones de datos aleatorios, calcule un intervalo de confianza del 95% del coeficiente de autocorrelación en cualquier retraso.
 - Si todos los coeficientes de autocorrelación están dentro de sus intervalos de confianza del 95% y no muestran un patrón particular, ¿qué conclusión se puede extraer del proceso?
 - Si las primeras tres observaciones son positivas y significativamente diferentes de cero y, en general, el patrón de autocorrelación declina hacia cero, ¿qué conclusiones se pueden extraer del proceso?
 - Un proceso se revisa trimestralmente. Si las autocorrelaciones r_4 , r_8 y r_{12} son significativamente diferentes de cero, ¿qué conclusión se puede extraer?
- Suponga que el siguiente modelo de serie de tiempo se ajustó a datos históricos y resulta ser 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 $Y_4 = 31.9$. Suponiendo que $\hat{Y}_1 = 35$ y $\varepsilon_0 = 0$, calcule los pronósticos para los periodos 5, 6 y 7 si el periodo 4 es el origen del pronóstico.

- El modelo de una serie de tiempo se ajustó y verificó con datos históricos de producción, para dar

$$Y_t = 50 + .45Y_{t-1} + \varepsilon_t$$

Suponga que para el tiempo $t = 60$ la observación es $Y_{60} = 57$.

- Determine los pronósticos para los periodos 61, 62 y 63 a partir del origen 60.
 - Suponga que el valor observado Y_{61} es 59. Actualice los pronósticos de los periodos 62 y 63.
 - Suponga que la estimación de la varianza del término de error es $s^2 = 3.2$. Calcule un intervalo de predicción del 95% para el pronóstico del periodo 61.
- Complete la información faltante en la tabla P-4, indicando si las autocorrelaciones teóricas y las autocorrelaciones parciales se desvanecen o se cortan en estos modelos.

TABLA P-4

<i>Modelo</i>	<i>Autocorrelaciones</i>	<i>Autocorrelaciones parciales</i>
MA		
AR		
ARIMA		

5. Dadas las gráficas de la figura P-5 de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra, identifique tentativamente un modelo ARIMA para cada par de gráficas.
6. Un modelo ARIMA(1, 1, 0) [modelo AR(1) para las primeras diferencias] es ajustado a las observaciones de una serie de tiempo. Las primeras 12 autocorrelaciones residuales se presentan en la figura P-6. El modelo se ajustó con una constante.
 - a) Con base en una revisión de las autocorrelaciones residuales, ¿el modelo parece adecuado? ¿Por qué?
 - b) Si usted concluye que el modelo no es adecuado, indique cómo podría modificarse para eliminar cualquier insuficiencia.
7. Chips Bakery ha tenido problemas para pronosticar la demanda de su pan especial con alto contenido de fibra y le gustaría que usted le ayudara. Los datos de la demanda semanal, las autocorrelaciones de los datos originales y varias diferencias de los datos se presentan en las tablas P-7A-D.
 - a) Revise las gráficas de autocorrelación y sugiera un modelo tentativo para estos datos. ¿Cómo se decidió por este modelo? (*Nota:* No diferencie más de lo necesario. Se recomienda mucha diferenciación si las autocorrelaciones de la muestra con retrasos cortos tienden a crecer en magnitud).
 - b) Use un programa de computadora para modelar con ARIMA, ajuste y verifique su modelo identificado con los datos de demanda del pan.
 - c) Escriba la ecuación del pronóstico de la demanda para el pan de alto contenido de fibra para el periodo 53.
 - d) Usando un programa de computadora para modelar con ARIMA, pronostique la demanda para el pan de alto contenido de fibra para los siguientes cuatro periodos a partir del origen del pronóstico 52. Además, construya intervalos de predicción del 95%.
8. La tabla P-8 contiene una serie de tiempo con 126 observaciones. Use un programa de computadora para modelar con ARIMA, obtenga una gráfica de los datos, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA adecuado y genere los pronósticos para los siguientes tres periodos a partir del origen del pronóstico $t = 126$.
9. La tabla P-9 contiene una serie de tiempo de 80 observaciones. Usando un programa de computadora para modelar con ARIMA, obtenga una gráfica de los datos, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA adecuado y genere los pronósticos de los siguientes tres periodos a partir del origen del pronóstico $t = 80$.
10. La tabla P-10 contiene una serie de tiempo de 80 observaciones. Usando un programa de computadora para modelar con ARIMA, obtenga una gráfica de los datos, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Desarrolle un modelo ARIMA adecuado y genere los pronósticos de los siguientes tres periodos a partir del origen del pronóstico $t = 80$.
11. La tabla P-11 contiene una serie de tiempo de 96 observaciones mensuales. Usando un programa de computadora para modelar con ARIMA, obtenga una gráfica de los datos, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Desarrolle

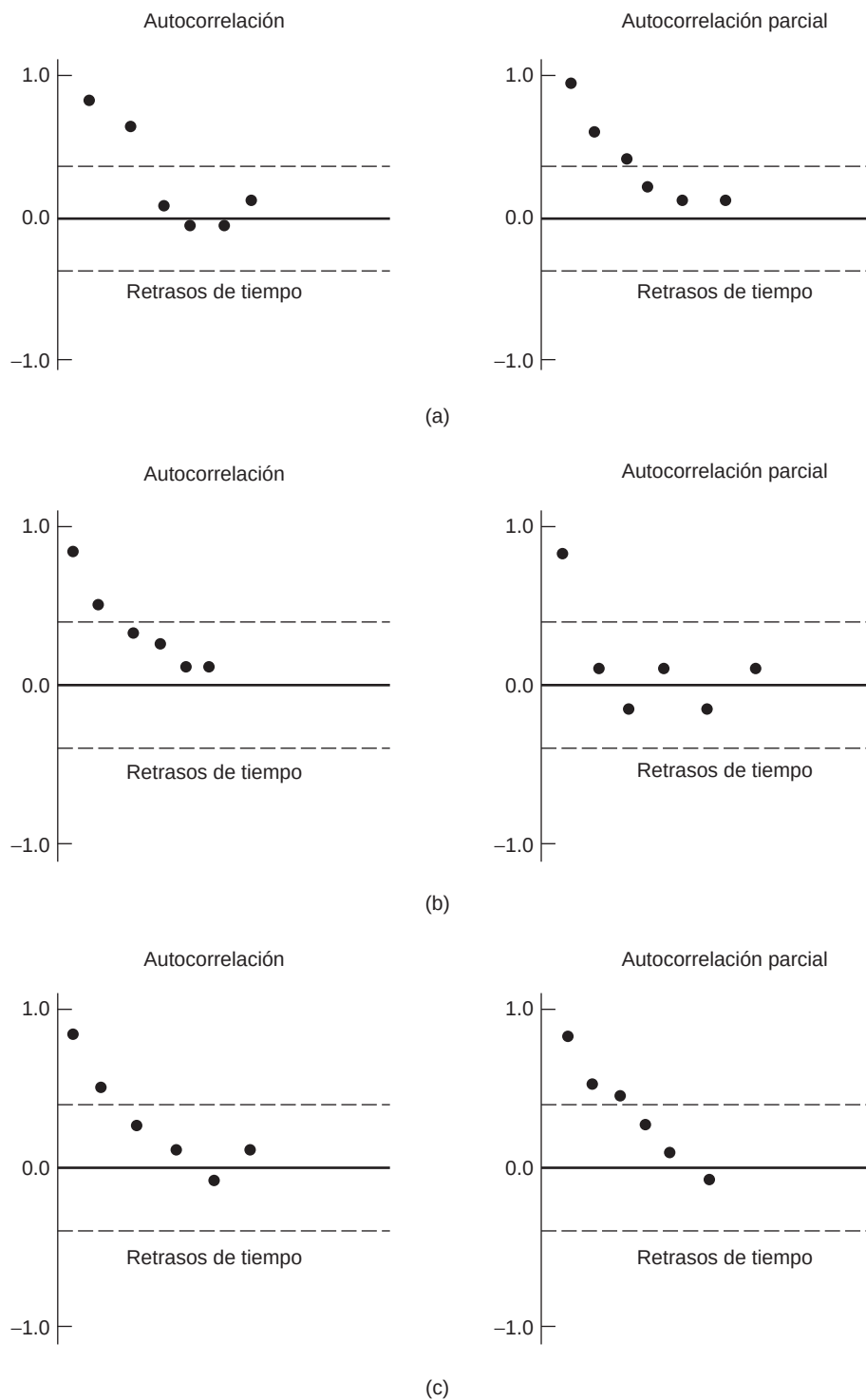


FIGURA P-5 Autocorrelaciones de la muestra y autocorrelaciones parciales de la muestra

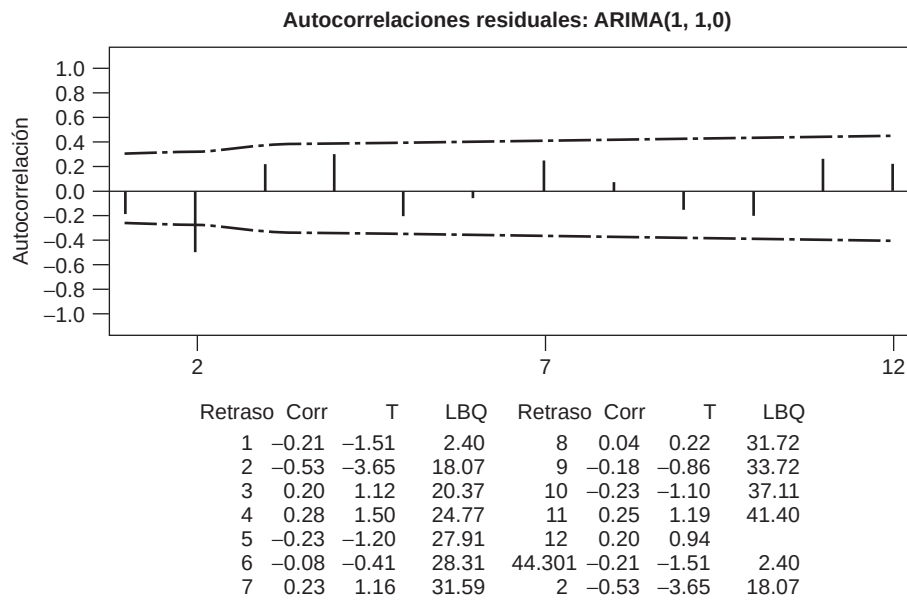


FIGURA P-6 Autocorrelaciones para los residuos de un modelo ARIMA(1, 1, 0)

un modelo ARIMA adecuado y genere los pronósticos de los siguientes 12 periodos a partir del origen del pronóstico $t = 96$.

12. Los datos de la tabla P-12 son los precios semanales de las acciones de IBM.
- a) Usando un programa de computadora para modelar con ARIMA, obtenga una gráfica de los datos, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Use esta información para identificar tentativamente un modelo ARIMA adecuado para esta serie.

TABLA P-7A Demanda semanal de ventas del pan con alto contenido de fibra (en miles)

Semana	Demanda	Semana	Demanda	Semana	Demanda	Semana	Demanda
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 de los datos originales

<i>Retraso</i>	<i>Autocorrelación</i>	<i>Retraso</i>	<i>Autocorrelación</i>
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

<i>Retraso</i>	<i>Autocorrelación</i>	<i>Retraso</i>	<i>Autocorrelación</i>
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

<i>Retraso</i>	<i>Autocorrelación</i>	<i>Retraso</i>	<i>Autocorrelación</i>
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>	<i>Y_t</i>	<i>t</i>	<i>Y_t</i>	<i>t</i>	<i>Y_t</i>	<i>t</i>	<i>Y_t</i>	<i>t</i>	<i>Y_t</i>
1	60.9	7	54.2	13	55.6	19	59.6	25	50.9
2	64.2	8	56.9	14	52.2	20	55.6	26	59.6
3	64.2	9	58.2	15	58.2	21	60.2	27	54.2
4	59.6	10	56.2	16	62.9	22	62.2	28	50.9
5	62.2	11	60.9	17	57.6	23	58.2	29	52.2
6	60.9	12	56.9	18	58.9	24	51.6	30	57.6

TABLA P-8 (Continuación)

t	Y_t	t	Y_t	t	Y_t	t	Y_t
31	51.6	55	58.2	79	51.6	103	64.2
32	46.9	56	60.9	80	48.2	104	58.2
33	51.6	57	51.6	81	47.6	105	52.9
34	57.6	58	54.9	82	50.2	106	56.9
35	60.2	59	66.2	83	58.2	107	51.6
36	64.2	60	57.6	84	65.6	108	48.2
37	62.2	61	48.9	85	53.6	109	47.6
38	53.6	62	46.2	86	55.6	110	50.2
39	50.9	63	50.9	87	61.6	111	58.2
40	54.2	64	57.6	88	57.6	112	65.6
41	56.2	65	54.9	89	56.2	113	53.6
42	59.6	66	51.6	90	60.9	114	55.6
43	66.2	67	50.2	91	57.6	115	61.6
44	57.6	68	50.9	92	51.6	116	57.6
45	48.9	69	56.9	93	56.2	117	56.2
46	50.9	70	50.2	94	52.2	118	60.9
47	60.2	71	54.2	95	50.2	119	57.6
48	64.2	72	58.2	96	56.9	120	51.6
49	56.9	73	56.9	97	56.9	121	56.2
50	56.9	74	55.6	98	50.2	122	52.2
51	63.6	75	64.2	99	54.2	123	50.2
52	58.2	76	58.2	100	58.2	124	56.9
53	56.9	77	52.9	101	56.9	125	56.9
54	58.9	78	56.9	102	55.6	126	50.2

TABLA P-9

t	Y_t	t	Y_t	t	Y_t	t	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							
t	Y_t	t	Y_t	t	Y_t	t	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							
t	Y_t	t	Y_t	t	Y_t	t	Y_t
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	Mayo 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. 3	252	Jun. 2	257	Sept. 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

- b) ¿Es estacionaria la serie de IBM? ¿Qué corrección recomendaría usted si la serie fuera no estacionaria?
- c) Ajuste un modelo ARIMA a la serie de IBM. Interprete el resultado. ¿Los cambios sucesivos son aleatorios?
- d) Ejecute la verificación del diagnóstico para determinar la idoneidad de su modelo ajustado.
- e) Después de que se ha encontrado un modelo satisfactorio, pronostique el precio de las acciones de IBM para la primera semana de enero del año siguiente. ¿Cómo difiere su pronóstico del pronóstico informal, el cual dice que el pronóstico de la primera semana de enero es el precio de la última semana de diciembre (precio actual)?
13. Los datos de la tabla P-13 son las cotizaciones de 150 días de las acciones de DEF Corporation al cierre de operaciones. Determine el modelo ARIMA adecuado y pronos-

TABLA P-13

<i>Periodo</i>	<i>DEF</i>	<i>Periodo</i>	<i>DEF</i>	<i>Periodo</i>	<i>DEF</i>
1	136.0	16	136.0	31	140.0
2	132.8	17	138.4	32	144.8
3	130.4	18	137.6	33	140.0
4	128.8	19	139.2	34	139.2
5	136.8	20	134.4	35	139.2
6	135.2	21	136.8	36	136.8
7	134.4	22	139.2	37	140.8
8	139.2	23	139.2	38	141.6
9	136.8	24	140.0	39	139.2
10	136.0	25	139.2	40	142.4
11	133.6	26	140.8	41	140.8
12	139.2	27	139.2	42	140.0
13	137.6	28	138.4	43	132.0
14	139.2	29	136.0	44	142.4
15	139.2	30	142.4	45	138.4

TABLA P-13 (Continuación)

<i>Periodo</i>	<i>DEF</i>	<i>Periodo</i>	<i>DEF</i>	<i>Periodo</i>	<i>DEF</i>
46	138.4	81	134.4	116	133.6
47	136.8	82	133.6	117	134.4
48	139.2	83	131.2	118	130.4
49	135.2	84	132.0	119	132.8
50	138.4	85	131.2	120	134.4
51	140.8	86	132.8	121	135.2
52	135.2	87	132.0	122	136.8
53	133.6	88	133.6	123	134.4
54	134.4	89	131.2	124	136.0
55	134.4	90	131.2	125	137.6
56	137.6	91	129.6	126	138.4
57	134.4	92	131.2	127	137.6
58	140.8	93	130.4	128	138.4
59	137.6	94	131.2	129	137.6
60	132.8	95	136.0	130	137.6
61	136.8	96	135.2	131	140.0
62	135.2	97	136.8	132	135.2
63	132.8	98	136.8	133	135.2
64	144.0	99	133.6	134	135.2
65	137.6	100	135.2	135	136.0
66	138.4	101	136.0	136	132.0
67	136.0	102	137.6	137	133.6
68	135.2	103	131.2	138	134.4
69	138.4	104	136.0	139	133.6
70	134.4	105	136.0	140	133.6
71	138.4	106	133.6	141	132.8
72	139.2	107	129.6	142	132.0
73	141.6	108	132.8	143	136.0
74	134.4	109	135.2	144	133.6
75	135.2	110	132.0	145	133.6
76	136.0	111	132.8	146	135.2
77	135.2	112	132.8	147	139.2
78	136.0	113	136.0	148	136.8
79	132.8	114	136.8	149	136.0
80	133.6	115	136.8	150	134.4

tique el precio de la acción cinco días adelante del origen del pronóstico $t = 145$. Compare sus pronósticos con los precios reales usando el *MAPE*. ¿Qué tan exactos son sus pronósticos?

14. Los datos de la tabla P-14 son las cifras de accidentes automovilísticos semanales de los años 2006 y 2007 en el condado de Havana. Determine el modelo ARIMA adecuado y pronostique los accidentes para la semana 91. Haga comentarios sobre la exactitud de su pronóstico.
15. La tabla P-15 incluye 120 observaciones mensuales sobre el precio en centavos de la fanega de maíz en Omaha, Nebraska. Identifique el mejor modelo ARIMA para estos datos. Genere pronósticos del precio de maíz para los siguientes 12 meses. Haga comentarios sobre el patrón de los pronósticos.
16. Use la metodología Box-Jenkins para modelar y pronosticar las ventas mensuales de Cavanaugh Company dadas en la tabla P-14 del capítulo 5. (*Sugerencia*: Considere una transformación logarítmica antes de modelar estos datos).

TABLA P-14

<i>Periodo</i>	<i>Observación</i>	<i>Periodo</i>	<i>Observación</i>	<i>Periodo</i>	<i>Observación</i>
1	101	31	16	61	79
2	84	32	17	62	53
3	54	33	21	63	40
4	39	34	28	64	27
5	26	35	30	65	31
6	40	36	51	66	56
7	99	37	62	67	78
8	148	38	57	68	114
9	147	39	46	69	140
10	134	40	40	70	112
11	106	41	32	71	82
12	83	42	23	72	80
13	76	43	20	73	70
14	63	44	18	74	55
15	57	45	24	75	37
16	37	46	33	76	23
17	32	47	52	77	20
18	22	48	66	78	39
19	20	49	78	79	71
20	23	50	83	80	110
21	30	51	87	81	112
22	50	52	64	82	93
23	61	53	44	83	75
24	59	54	24	84	60
25	64	55	29	85	63
26	58	56	73	86	46
27	44	57	138	87	32
28	26	58	154	88	23
29	24	59	119	89	53
30	18	60	102	90	90

17. Con base en la metodología Box-Jenkins, modele y pronostique las ventas trimestrales de Disney Company que se presentan en la tabla P-16 del capítulo 5. (*Sugerencia:* Considere una transformación logarítmica antes de modelar los datos).
18. Con base en la metodología Box-Jenkins, modele y pronostique la demanda mensual de gasolina de Yukong Oil Company presentada en la tabla P-17 del capítulo 5.
19. Con base en la metodología Box-Jenkins, modele y pronostique los números mensuales de hombres con empleo de 16 años de edad en adelante en Estados Unidos de acuerdo con la tabla P-25 del capítulo 5.
20. Con base en la metodología Box-Jenkins, modele y pronostique las ventas trimestrales de las tiendas Wal-Mart de acuerdo con la tabla P-27 del capítulo 5.
21. Con base en la metodología Box-Jenkins, modele y pronostique los números anuales de terremotos severos de acuerdo con la tabla P-18 del capítulo 4.
22. Con base en la metodología Box-Jenkins, modele y pronostique las ventas trimestrales de The Gap de acuerdo con la tabla P-20 del capítulo 4.

TABLA P-15

t	Y_t	t	Y_t	t	Y_t
1	125	41	206	81	282
2	126	42	252	82	250
3	126	43	215	83	231
4	129	44	238	84	247
5	132	45	249	85	252
6	136	46	250	86	250
7	139	47	250	87	253
8	137	48	275	88	255
9	155	49	292	89	236
10	144	50	304	90	221
11	142	51	301	91	204
12	153	52	265	92	187
13	159	53	271	93	187
14	160	54	289	94	182
15	156	55	319	95	210
16	153	56	377	96	221
17	151	57	350	97	222
18	158	58	378	98	236
19	150	59	358	99	256
20	129	60	359	100	253
21	117	61	333	101	250
22	112	62	310	102	232
23	109	63	293	103	221
24	126	64	285	104	214
25	123	65	284	105	221
26	121	66	283	106	225
27	123	67	291	107	224
28	125	68	312	108	228
29	128	69	288	109	233
30	128	70	280	110	239
31	130	71	255	111	251
32	129	72	258	112	261
33	139	73	266	113	268
34	131	74	271	114	304
35	131	75	272	115	285
36	166	76	265	116	274
37	159	77	281	117	281
38	163	78	294	118	259
39	158	79	300	119	270
40	165	80	284	120	251

23. La figura P-23 es la gráfica de una serie de tiempo del número semanal de casos de influenza A positiva en una región de Texas de la semana del 2 de septiembre de 2003 a la semana del 10 de abril de 2007. El pico del número de casos durante el periodo se registró de noviembre a abril y el número de casos es cero durante las semanas restantes de finales de la primavera a principios de otoño. Aunque son datos médicos, esta serie comparte muchas de las características de la demanda de ciertas piezas de refacción. El número de casos de influenza y la demanda de refacciones son datos “amontonados”, con periodos de actividad seguidos de periodos inactivos largos y variables. ¿Puede usted pensar en los problemas que una serie

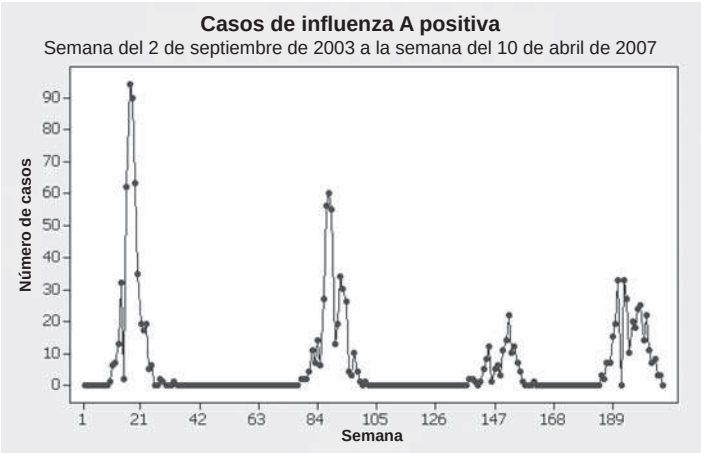


FIGURA P-23 Gráfica de la serie de tiempo de los casos semanales de influenza A positiva

como la de la influenza podría plantear para los modelos ARIMA? ¿Puede usted pensar en un modelo ARIMA simple no estacional que generara un pronóstico razonable del número de casos de influenza A positiva con una semana de anticipación?

CASOS

CASO 9-1 VENTAS EN UN RESTAURANTE

Este caso se refiere a los datos de ventas y a la situación del restaurante descrito en el caso 8-3. Jim Price concluyó un curso de pronósticos y está ansioso por aplicar la metodología Box-Jenkins a las ventas del restaurante. Estos datos, presentados en la tabla 9-17A, inician en la

semana que termina el domingo 4 de enero de 1981 y continúa hasta la semana que termina el domingo 26 de diciembre de 1982. La tabla 9-17B contiene datos nuevos de la semana que termina el 2 de enero de 1983, hasta la semana que termina el 30 de octubre de 1983.

TABLA 9-17A Ventas en restaurante: Datos antiguos

<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	9/6/81	5,739	5/9/82	5,837
1/11/81	2,514	9/13/81	5,560	5/16/82	7,162
1/18/81	5,843	9/20/81	5,335	5/23/82	4,997
1/25/81	4,912	9/27/81	5,305	5/30/82	4,063
2/1/81	5,133	10/4/81	5,364	6/6/82	3,942
2/8/81	4,563	10/11/81	5,511	6/13/82	4,011
2/15/81	5,416	10/18/81	5,698	6/20/82	3,999
2/22/81	6,416	10/25/81	5,382	6/27/82	4,794
3/1/81	5,879	11/1/81	5,629	7/4/82	4,956

TABLA 9-17A (Continación)

<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>
3/8/81	3,460	11/8/81	5,617	7/11/82	3,885
3/15/81	4,517	11/15/81	5,742	7/18/82	4,209
3/22/81	5,188	11/22/81	3,747	7/25/82	3,614
3/29/81	5,944	11/29/81	4,159	8/1/82	3,722
4/5/81	5,842	12/6/81	4,853	8/8/82	4,307
4/12/81	6,589	12/13/81	5,607	8/15/82	3,322
4/19/81	5,447	12/20/81	3,946	8/22/82	5,962
4/26/81	7,548	12/27/81	1,919	8/29/82	6,784
5/3/81	6,403	1/3/82	1,898	9/5/82	6,069
5/10/81	4,103	1/10/82	1,870	9/12/82	5,897
5/17/81	6,594	1/17/82	3,962	9/19/82	5,916
5/24/81	5,742	1/24/82	5,973	9/26/82	4,998
5/31/81	3,714	1/31/82	5,009	10/3/82	5,111
6/7/81	3,399	2/7/82	5,328	10/10/82	5,612
6/14/81	3,376	2/14/82	5,014	10/17/82	5,906
6/21/81	3,627	2/21/82	4,986	10/24/82	6,010
6/28/81	4,201	2/28/82	5,213	10/31/82	5,937
7/5/81	3,515	3/7/82	4,807	11/7/82	6,004
7/12/81	3,645	3/14/82	3,964	11/14/82	5,959
7/19/81	3,416	3/21/82	5,201	11/21/82	4,223
7/26/81	3,565	3/28/82	4,863	11/28/82	4,679
8/2/81	2,428	4/4/82	5,019	12/5/82	5,307
8/9/81	3,292	4/11/82	4,868	12/12/82	6,101
8/16/81	3,460	4/18/82	5,777	12/19/82	6,896
8/23/81	6,212	4/25/82	6,543	12/26/82	2,214
8/30/81	6,057	5/2/82	6,352		

TABLA 9-17B Ventas del Restaurante: Datos Nuevos

<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 utilizarse con los datos originales?
2. ¿Cuáles son sus pronósticos para las primeras cuatro semanas de enero de 1983?
3. ¿Cómo se comparan estos pronósticos con las ventas reales?
4. ¿Cómo se compara su modelo Box-Jenkins con el modelo de regresión empleado en el capítulo 8?
5. ¿Utilizaría usted el mismo modelo Box-Jenkins si los datos nuevos estuvieran combinados con los datos antiguos?

CASO 9-2 MR. TUX

John Mosby decidió aplicar el método Box-Jenkins para pronosticar sus datos de ventas mensuales. Él entiende que este procedimiento es más complejo que los métodos más sencillos que ha ensayado. También sabe que es posible obtener pronósticos más exactos con este método avanzado. Él tiene acceso al programa de computadora Minitab que permite ajustar modelos ARIMA. Y como ya tiene los datos recopilados y almacenados en su computadora, decidió hacer un intento.

John decide usar todo el conjunto de datos, que consiste en 96 meses de cifras de ventas. John sabe que un modelo ARIMA puede manejar la estructura estacional de su serie de tiempo así como modelar la correlación de un mes a otro.

John calculó la función de autocorrelación de la muestra de sus datos en el caso 3-2 (véase la figura 3-25) y determinó que sus datos tenían una tendencia (no estacionaria). Luego calculó la función de autocorrelación de las primeras diferencias de sus ventas. El resultado se aprecia en la figura 3-26. Al observar la gráfica, John inmediatamente notó los picos significativos de los retrasos 12 y 24, que indicaban estacionalidad y quizá la necesidad de una diferencia estacional. La función de autocorrelación de la muestra de los datos después de una (primera) diferencia regular y una diferencia estacional se despliegan en la figura 9-39. Las autocorrelaciones parciales de la muestra de los datos diferenciados se presentan en la figura 9-40.

Función de autocorrelación: Diff1Diff2Sales

Retraso	ACF	T	LBQ
1	-0.434972	-3.96	16.28
2	-0.119297	-0.93	17.52
3	0.177448	1.36	20.29
4	-0.013968	-0.10	20.31
5	-0.223625	-1.68	24.84
6	0.126474	0.92	26.30
7	-0.053425	-0.38	26.57
8	0.054957	0.39	26.85
9	-0.006380	-0.05	26.85
10	0.032094	0.23	26.95
11	-0.006051	-0.04	26.96
12	-0.037127	-0.27	27.09
13	0.024916	0.18	27.16
14	-0.025145	-0.18	27.22
15	-0.081508	-0.58	27.91
16	0.028684	0.20	28.00
17	0.184183	1.31	31.62
18	-0.148161	-1.03	34.01
19	-0.059354	-0.41	34.39
20	0.165452	1.14	37.46
21	-0.044315	-0.30	37.68

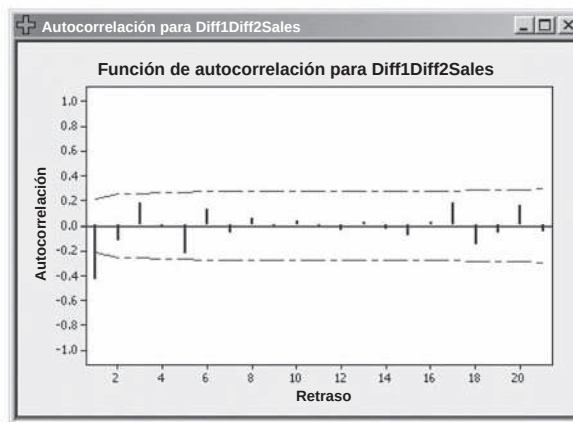


FIGURA 9-39 Autocorrelaciones de Mr. Tux: Series regulares y diferenciadas estacionalmente

Función de autocorrelación parcial: Diff1Diff2Sales

Retraso	PACF	T
1	-0.434972	-3.96
2	-0.380486	-3.47
3	-0.086077	-0.78
4	0.013655	0.12
5	-0.232666	-2.12
6	-0.163070	-1.49
7	-0.237648	-2.17
8	-0.061466	-0.56
9	-0.042868	-0.39
10	-0.010757	-0.10
11	-0.009411	-0.09
12	-0.091044	-0.83
13	-0.021219	-0.19
14	-0.051663	-0.47
15	-0.144650	-1.32
16	-0.191300	-1.74
17	0.089113	0.81
18	0.032755	0.30
19	-0.142079	-1.29
20	-0.075125	-0.68
21	-0.016020	-0.15

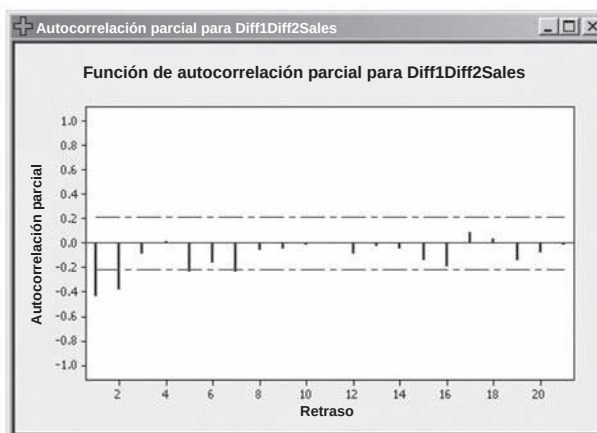


FIGURA 9-40 Autocorrelaciones parciales de Mr. Tux: Series regulares y diferenciadas estacionalmente

PREGUNTAS

1. Discuta los problemas, si existen, de explicar el método Box-Jenkins al ejecutivo financiero de John y a otros de su equipo de administración.
2. Dadas las autocorrelaciones de la figura 9-39 y las autocorrelaciones parciales de la figura 9-40, ¿qué términos regulares (no estacionales) podría John incluir en un modelo ARIMA de las ventas de Mr. Tux? ¿Qué términos estacionales podría incluir en el modelo?
3. Usando un programa para modelar ARIMA, ajuste y verifique un modelo ARIMA de las ventas de Mr. Tux. Genere los pronósticos para los 12 meses siguientes.

CASO 9-3 CONSUMER CREDIT COUNSELING

En los casos 1-2 y 3-3 se describió la operación del Consumer Credit Counseling (CCC).

El director ejecutivo, Marv Harnishfeger, determinó que la variable más importante que CCC necesitaba para pronosticar era el número de clientes nuevos que serían atendidos durante el resto de 1993. Marv le dio a Dorothy Mercer los datos mensuales del número de clientes nuevos atendidos por CCC durante el periodo de enero de 1985 a marzo de 1993 (véase el caso 3-3).

Dorothy estaba muy contenta con usted por su aportación a la definición del pronóstico. Sin embargo, usted no estaba completamente satisfecho y decidió que ahora era tiempo de impresionarla realmente. Usted le dice que ella podría intentar una de las técnicas más poderosas usadas para pronosticar: la metodología Box-Jenkins y los modelos ARIMA. Dorothy nunca había oído hablar de la metodología Box-Jenkins, pero estaba dispuesta a que usted hiciera el intento.

TAREA

1. Redacte un escrito para Dorothy donde explique en qué consiste la metodología Box-Jenkins.
2. Desarrolle un modelo ARIMA usando la metodología Box-Jenkins y pronostique el número mensual de clientes nuevos para el resto de 1993.
3. Redacte un segundo escrito para Dorothy que resuma los resultados de este análisis.

CASO 9-4 THE LYDIA E. PINKHAM MEDICINE COMPANY¹³

La compañía de Lydia E. Pinkham era un negocio familiar cuyos ingresos, en su mayor parte, provenían del compuesto vegetal de Lydia Pinkham. Quizás los estudiantes de ahora podrían tomar ese compuesto para liberar el estrés; por desgracia, ya no se vende más. La fotografía de Lydia Pinkham aparecía en la etiqueta, y el compuesto era comercializado en particular para las mujeres. Los anuncios del compuesto incluían esta invitación: “Escriba con total libertad a la señora Pinkham, en Lynn, Massachussets, y siga el consejo que ofrece gratis a todas las mujeres. Éste es el consejo que ha llevado la felicidad a muchos hogares, los cuales estaban al borde del fracaso por el nerviosismo y la irritabilidad”. De hecho, la compañía aseguraba que una empleada contestaría cada carta. Las mujeres *escribían* a la señora Pinkham. Una de ellas afirmaba: “Sin [el compuesto vegetal de Lydia Pinkham] yo ahora estaría muerta, o aún peor, loca... Ya me había dado por vencida; conforme intentaba una cosa y otra, me convencía de que no existía nada que me hiciera bien. Pero, gracias a su medicamento, ahora estoy bien y muy fuerte; de hecho, soy una persona completamente diferente”. Éste y otros testimonios acerca del compuesto se reprodujeron en anuncios impresos.

La naturaleza única de la compañía —un producto dominante que explicaba la mayor parte de las ventas de la compañía, sin equipo de ventas, y una gran proporción de las ganancias de las ventas invertida en publicidad— y la disponibilidad de los datos tanto de las ventas como de la publicidad llevaron al Comité Sobre Determinación de Precios de la Oficina Nacional de Investigación Económica (NBER) en 1943 a recomendar que los datos se sometieran a un análisis. La investigación no se realizó en varios años (Palda, 1964). Los analistas estudiaron los datos usando modelos causales que incluyen los datos de publicidad y otras variables económicas (similares a las presentadas en el capítulo 8). Sin embargo, varios investigadores sugirieron que usando los enfoques Box-Jenkins sólo con los datos de las ventas, se obtienen predicciones comparables, o incluso superiores, que las de los enfoques causales (véase, por ejem-

plo, Kyle, 1978). Resulta atractivo estudiar las ventas por dos razones:

1. El producto mismo no cambió durante el periodo de los datos; es decir, no existen cambios en la serie causados por cambios en el producto.
2. La fuerza de ventas no cambió durante el periodo de los datos, y la proporción de ingresos gastados en publicidad fue constante. Por lo tanto, no existían cambios en los datos causados por promociones especiales u otros fenómenos de marketing.

Por lo general, los datos reales no son tan “limpios” en términos de la continuidad de marketing y del producto.

La tarea, entonces, consiste en determinar cuál modelo Box-Jenkins (ARIMA) es el “mejor” para estos datos. El modelo se desarrollará usando los datos de 1907 a 1948 y se probará con los datos de 1949 a 1960 contenidos en la tabla 9-18.

IDENTIFICACIÓN DEL MODELO

Se utilizó un programa de computadora capaz de manejar modelos ARIMA para examinar los datos de 1907 a 1948; los datos de 1949 a 1960 se utilizaron para examinar la eficiencia del modelo de pronóstico seleccionado. Las pruebas preliminares sugieren que los datos son estacionarios (es decir, aparentemente no había tendencia), de manera que no se aplicó la diferenciación. Después de revisar las autocorrelaciones y las autocorrelaciones parciales, se determinó que un modelo AR era el más adecuado para los datos. [Las autocorrelaciones (ACF) y las autocorrelaciones parciales (PACF) de 10 periodos están contenidas en la tabla 9-19]. Las autocorrelaciones y autocorrelaciones parciales parecían ser congruentes con las de un proceso AR(2). Para verificar el orden p del componente AR, se aplicó el criterio de información de Akaike (AIC) (véase la ecuación 9.7) para modelos autorregresivos de órdenes $p = 1, 2$ y 3. El AIC conduce a la selección de un modelo AR(2) para los datos de Lydia Pinkham.

¹³Este caso fue contribución de la doctora Susan C. White, de la George Washington University, en Washington, D.C. Para más información, véase el reporte de Susan C. White, “Predicting Time Series with Neural Networks Versus Statistical Models: The Lydia Pinkham Data”, *Proceedings of the 24th Annual Conference of the Decision Sciences Institute, Southwest Region*, 1993, 108-110.

TABLA 9-18 Datos de Lydia E. Pinkham Medicine Company

Año	Ventas (miles de dólares corrientes)	Año	Ventas (miles de dólares corrientes)	Año	Ventas (miles de dólares corrientes)
1907	1,016	1925	3,438	1943	2,602
1908	921	1926	2,917	1944	2,518
1909	934	1927	2,359	1945	2,637
1910	976	1928	2,240	1946	2,177
1911	930	1929	2,196	1947	1,920
1912	1,052	1930	2,111	1948	1,910
1913	1,184	1931	1,806	1949	1,984
1914	1,089	1932	1,644	1950	1,787
1915	1,087	1933	1,814	1951	1,689
1916	1,154	1934	1,770	1952	1,866
1917	1,330	1935	1,518	1953	1,896
1918	1,980	1936	1,103	1954	1,684
1919	2,223	1937	1,266	1955	1,633
1920	2,203	1938	1,473	1956	1,657
1921	2,514	1939	1,423	1957	1,569
1922	2,726	1940	1,767	1958	1,390
1923	3,185	1941	2,161	1959	1,397
1924	3,352	1942	2,336	1960	1,289

TABLA 9-19

Retraso k	ACF	PACF
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

ESTIMACIÓN DEL MODELO Y PRUEBA DE SU ADECUACIÓN

Se usó un programa de computadora para estimar los parámetros (incluyendo un término constante) del modelo AR(2) utilizando los datos de 1907 a 1948. Al utilizar los parámetros estimados, el modelo resultante es

$$Y_t = 178.6 + 1.423Y_{t-1} - .521Y_{t-2} + \varepsilon_t$$

(.137) (.136)

donde los números entre paréntesis debajo de los coeficientes autorregresivos son sus desviaciones estándar estimadas. Cada coeficiente autorregresivo es significativamente diferente de cero para cualquier nivel de significancia razonable. Todas las autocorrelaciones resi-

duales resultaron pequeñas y todas estuvieron dentro de sus límites de error del 95%. Los estadísticos chi cuadrada de Ljung-Box tenían valores p de .63, .21 y .64 para los grupos de retrasos $m = 12, 24$ y 36 , respectivamente. El modelo AR(2) parece adecuado para los datos de ventas de Lydia Pinkham.

ELABORACIÓN DEL PRONÓSTICO 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

TABLA 9-20

Año	Real	Pronóstico	Error
1949	1,984	1,896.4	87.6
1950	1,787	2,006.7	-219.7
1951	1,689	1,687.8	1.2
1952	1,866	1,651.0	215.0
1953	1,896	1,953.9	-57.9
1954	1,684	1,904.4	-220.4
1955	1,633	1,587.1	45.9
1956	1,657	1,625.0	32.0
1957	1,569	1,685.7	-116.7
1958	1,390	1,548.0	-158.0
1959	1,387	1,339.1	47.9
1960	1,289	1,428.1	-139.1

1958 se utilizan para calcular el pronóstico de 1959). La ecuación del pronóstico es

$$\hat{Y}_t = 178.6 + 1.423Y_{t-1} - .521Y_{t-2}$$

y los pronósticos un paso adelante, así como los errores de pronóstico se presentan en la tabla 9-20.

Además de los pronósticos un paso adelante se calcularon también algunas medidas de precisión. Los pronósticos del modelo AR(2) tienen un error porcentual absoluto medio (MAPE) del 6.9% y una desviación media absoluta (MAD) de \$112 (miles de

dólares actuales). Estas cifras se comparan favorablemente con las medidas de precisión de los modelos causales desarrollados por otros investigadores.

RESUMEN Y CONCLUSIONES

Un modelo parsimonioso (el número más pequeño de parámetros) AR(2) se ajustó a los datos de Lydia Pinkham para los años de 1907 a 1948. Este modelo generó pronósticos un paso adelante bastante exactos para los años de 1949 a 1960.

PREGUNTAS

1. Después de concluir este análisis, se obtiene la cifra de ventas para 1961: \$1,426. ¿Cuál es el pronóstico del modelo para 1961? Si este año se agregara al conjunto de datos de prueba, ¿cuál sería el MAPE y la MAD?
2. Existe alguna evidencia de que los datos de Lydia Pinkham pueden ser no estacionarios. Por ejemplo, las autocorrelaciones de la muestra tienden a ser grandes (persisten) para varios retrasos. Diferencie los datos. Elabore una gráfica de series de tiempo de las diferencias. Usando Minitab o un programa similar, ajuste un modelo ARIMA(1, 1, 0) para las ventas anuales de 1907 a 1948. Genere pronósticos un paso adelante de los años 1949 a 1960. ¿Cuál

modelo AR(2) o ARIMA(1, 1, 0), cree usted que es mejor para los datos de Lydia Pinkham?

3. Los datos de Lydia Pinkham son interesantes por su naturaleza única (sin cambios) en el producto y el marketing para el periodo de 54 años representado. ¿Qué factores podrían afectar las ventas anuales de automóviles y de cobre durante ese mismo periodo? ¿Por qué?

Para información adicional sobre la compañía Lydia E. Pinkham Medicine y Lydia Pinkham, véase Sarah Stage, *Female Complaints: Lydia Pinkham and the Business of Women's Medicine* (Nueva York: Norton, 1979).

CASO 9-5 CIUDAD COLLEGE STATION

Para financiar sus operaciones, la ciudad de College Station, Texas, depende de los ingresos por impuestos sobre ventas mensuales. Los ingresos por impuestos sobre ventas generados justifican aproximadamente el 44% del total del presupuesto del fondo general. Es importante pronosticar la cantidad de los ingresos por impuestos sobre ventas que recibirá esta ciudad del estado de Texas mensualmente, de manera que se puedan hacer coincidir los gastos anticipados con el ingreso proyectado.

A Charles Lemore, un analista de la ciudad, se le pidió desarrollar un modelo de pronóstico para los ingresos mensuales de impuestos sobre ventas. Charles recientemente terminó un curso de pronósticos en la universidad local y decidió poner en práctica la metodología Box-Jenkins. Comenzó por examinar los datos y las funciones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra.

Los ingresos por impuestos sobre ventas (en miles de dólares) para el periodo de enero de 1991 a noviembre de 1999 están contenidos en la tabla 9-21. En la figura 9-41 se presenta una gráfica de una serie de tiempo de los ingresos.

Un fuerte componente estacional se hace evidente en una gráfica de la serie y en las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra contenidas en las figuras 9-42 y 9-43, respectivamente. Las autocorrelaciones en los retrasos múltiples de 3 sugieren que existe un patrón estacional trimestral a lo largo de un año (el segundo pago mensual a la ciudad dentro de cualquier trimestre es relativamente grande), así como un patrón anual. Además, la serie contiene una clara tendencia ascendente.

Charles decidió diferenciar la serie de ingresos en relación con el periodo estacional $S = 12$. Las autocorrelaciones de la muestra de la serie diferenciada estacionalmente se presentan en la figura 9-44. Una gráfica de la serie diferenciada estacionalmente (no se muestra) parece variar alrededor de un nivel constante de 50.

Charles está listo para ajustar y verificar un modelo ARIMA para la serie de los ingresos por impuestos sobre ventas, pero necesita la ayuda de usted. A él también le gustaría generar los pronósticos para los siguientes 12 meses, una vez que tenga el modelo adecuado.

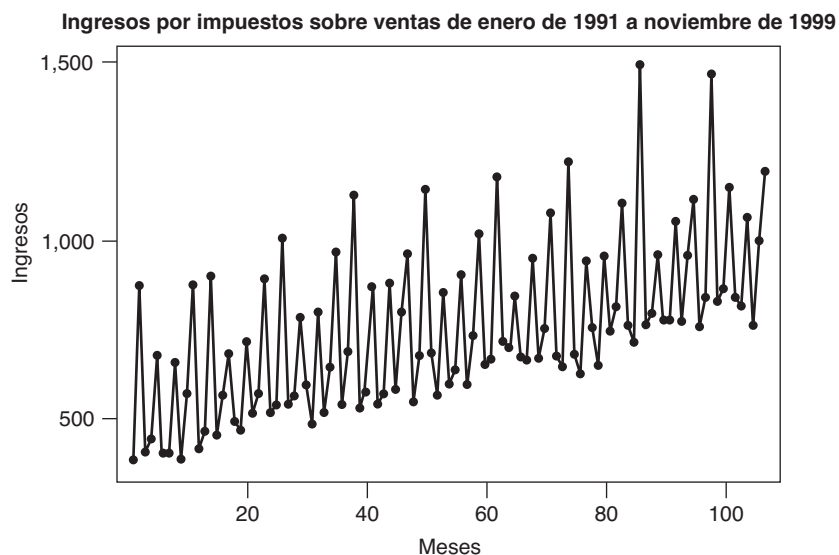
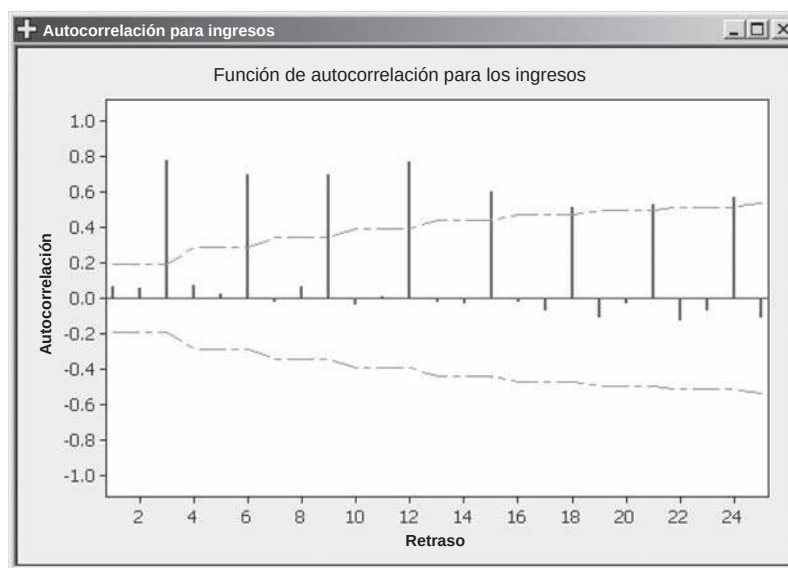


FIGURA 9-41 Ingresos por impuestos sobre ventas de la Ciudad College Station



Función de autocorrelación: Ingresos

Retraso	ACF	T	LBQ	Retraso	ACF	T	LBQ
1	0.065295	0.68	0.47	14	-0.028273	-0.13	255.29
2	0.052000	0.54	0.77	15	0.598940	2.68	300.77
3	0.778231	7.99	68.69	16	-0.024427	-0.10	300.85
4	0.069311	0.48	69.23	17	-0.073904	-0.31	301.56
5	0.027452	0.19	69.32	18	0.515622	2.16	336.40
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	375.81
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	-0.111329	-0.41	426.51
13	-0.027366	-0.12	255.19				

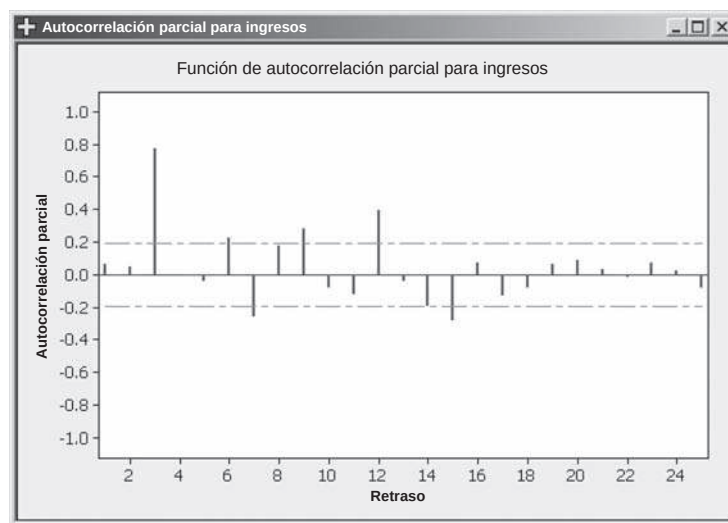
FIGURA 9-42 Autocorrelaciones de la muestra de los ingresos por impuestos sobre ventas

TAREA

1. Desarrolle un modelo ARIMA para los ingresos por impuestos sobre ventas usando la metodología Box-Jenkins.
2. A partir de su modelo, genere los pronósticos de ingreso para los siguientes 12 meses. Agregue estos pronósticos al final de la serie y grafique los resultados. ¿Está usted contento con el patrón de los pronósticos?
3. Redacte un breve escrito resumiendo sus descubrimientos.

TABLA 9-21

Año	Ene.	Feb.	Mar.	Abr.	Mayo	Jun.	Jul.	Ago.	Sept.	Oct.	Nov.	Dic.
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	1,005	543	568	785	596	488	799	518	644	968	544
1994	691	1,125	531	576	868	541	571	876	579	801	960	546
1995	675	1,139	684	564	851	595	641	902	596	735	1,014	652
1996	663	1,174	714	698	842	675	665	948	670	753	1,073	676
1997	646	1,219	680	628	943	756	649	951	742	814	1,100	759
1998	715	1,487	763	793	955	780	777	1,052	775	957	1,117	757
1999	839	1,459	828	862	1,147	840	814	1,064	766	997	1,187	



Función de autocorrelación parcial: Ingresos

Retraso	PACF	T	Retraso	PACF	T
1	0.065295	0.68	14	-0.189410	-1.96
2	0.047941	0.50	15	-0.285778	-2.96
3	0.776959	8.04	16	0.072095	0.75
4	-0.002013	-0.02	17	-0.130604	-1.35
5	-0.036516	-0.38	18	-0.082646	-0.85
6	0.224379	2.32	19	0.065410	0.68
7	-0.256006	-2.65	20	0.091950	0.95
8	0.177547	1.84	21	0.028949	0.30
9	0.278791	2.88	22	-0.017353	-0.18
10	-0.078668	-0.81	23	0.072479	0.75
11	-0.123788	-1.28	24	0.022790	0.24
12	0.396276	4.10	25	-0.079367	-0.82
13	-0.042746	-0.44			

FIGURA 9-43 Autocorrelaciones parciales de la muestra de los ingresos por impuestos sobre ventas

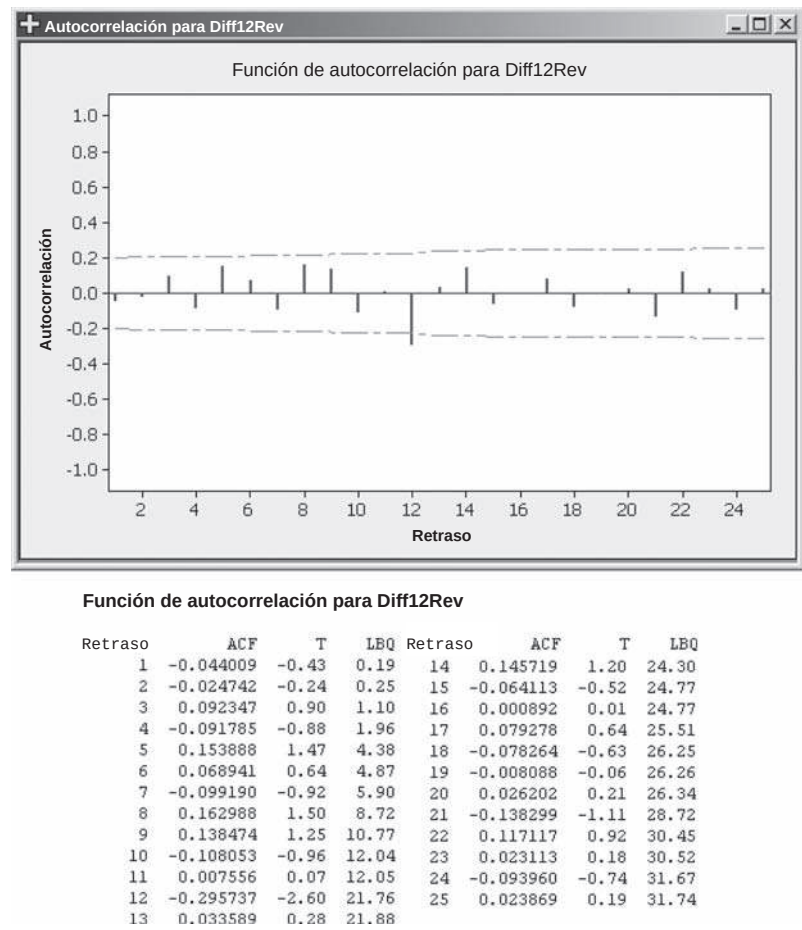


FIGURA 9-44 Autocorrelaciones de la muestra de los ingresos por impuestos sobre ventas diferenciados estacionalmente

CASO 9-6 UPS AIR FINANCE DIVISION

Air Finance Division es una de las diferentes divisiones dentro de la United Parcel Service Financial Corporation, una subsidiaria propiedad de la United Parcel Service (UPS). Air Finance Division tiene la responsabilidad de brindar servicios financieros a la compañía como un todo en relación con todas las adquisiciones de aviones. Además, esta división brinda financiamiento a entidades externas independientes para la compra de aviones como una operación independiente y diferente. Históricamente, el segmento financiero ajeno a UPS de Air Finance Division ha dado una tasa de retorno más alta y mayores ingresos que el segmento UPS; sin embargo, su seguimiento es más costoso.

Los pronósticos de financiamiento del segmento de mercado ajeno a UPS de Air Financial Division han sido muy subjetivos y no son tan confiables como los pronósticos de otros segmentos. La tabla 9-22 lista los datos correspondientes a 10 años de requerimientos financieros mensuales para el segmento ajeno a UPS de Air Financial Division. Estos datos fueron recopilados de reportes administrativos durante el periodo de enero de 1989 a diciembre de 1998 y son las cifras de fin de mes en millones de dólares.

TABLA 9-22

Año	Ene.	Feb.	Mar.	Abr.	Mayo	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

Los datos de la tabla 9-22 están graficados en la figura 9-45. La serie de tiempo de los requerimientos de financiamiento tiene un fuerte componente estacional junto con una tendencia general ascendente. Las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra se presentan en las figuras 9-46 y 9-47, respectivamente.

Se aplicará la metodología Box-Jenkins para desarrollar un modelo ARIMA para pronosticar los requerimientos futuros de financiamiento del segmento ajeno a UPS de Air Finance Division.

En virtud del carácter no estacionario de la serie de tiempo, se examinan las autocorrelaciones y autocor-

relaciones parciales de varias diferencias de la serie. La función de autocorrelación de la serie con una diferencia regular y una diferencia estacional de orden $S = 12$ se presenta en la figura 9-48. Es decir, estas autocorrelaciones se calcularon para las diferencias

$$\begin{aligned}
 W_t &= \Delta \Delta_{12} Y_t = \Delta(Y_t - Y_{t-12}) \\
 &= Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}
 \end{aligned}$$

Existe claramente un pico significativo en el retraso 12 en las autocorrelaciones de la muestra de la figura 9-48. Esto sugiere un modelo ARIMA para los W_t (el pro-

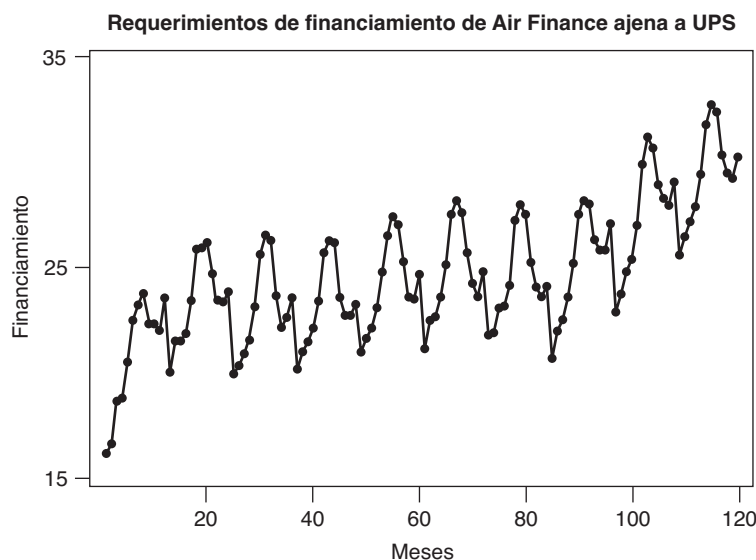
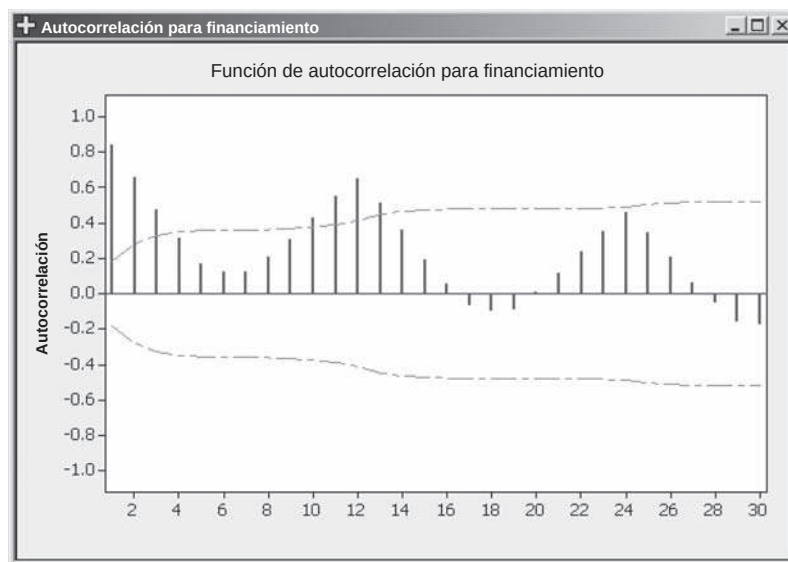


FIGURA 9-45 Requerimientos de financiamiento para el segmento ajeno a UPS de Air Finance Division



Función de autocorrelación: Financiamiento

Retraso	ACF	T	LBQ	Retraso	ACF	T	LBQ
1	0.837546	9.17	86.30	16	0.051635	0.21	385.19
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	387.38
4	0.309926	1.76	179.24	19	-0.090741	-0.38	388.58
5	0.167326	0.93	182.80	20	0.006750	0.03	388.59
6	0.119600	0.66	184.64	21	0.114121	0.47	390.51
7	0.118763	0.65	186.47	22	0.233914	0.97	398.68
8	0.208453	1.14	192.15	23	0.353911	1.45	417.59
9	0.308387	1.67	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	0.202990	0.78	473.83
12	0.649303	3.10	326.37	27	0.061054	0.23	474.41
13	0.511713	2.27	362.20	28	-0.056059	-0.22	474.91
14	0.355561	1.51	379.66	29	-0.156711	-0.60	478.86
15	0.192224	0.80	384.81	30	-0.175532	-0.67	483.88

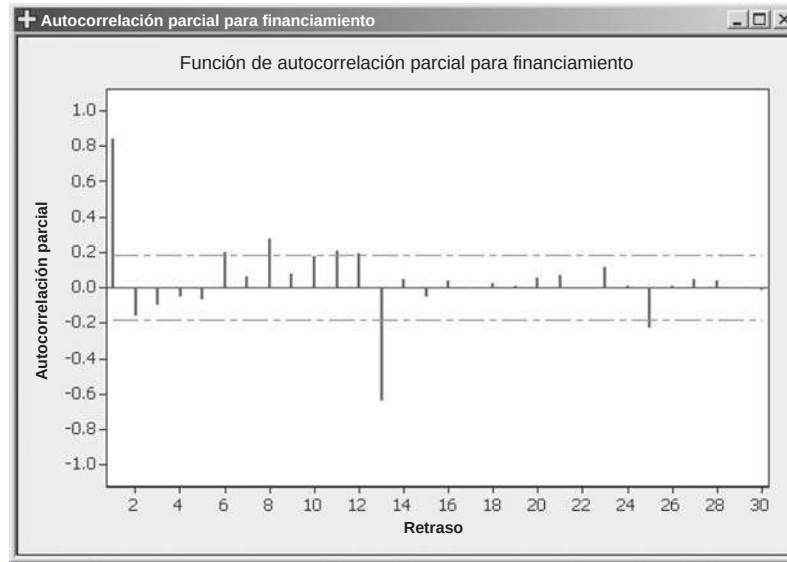
FIGURA 9-46 Autocorrelaciones de la muestra para los requerimientos de financiamiento

ceso con una diferencia regular y una diferencia estacional) que tiene un término de promedio móvil en el retraso estacional 12. Por consiguiente, un modelo

$ARIMA(0, 1, 0)(0, 1, 1)_{12}$ podría ser una buena selección inicial para los datos de requerimientos de financiamiento.

PREGUNTAS

1. Empleando Minitab o un software equivalente, ajuste un modelo $ARIMA(0, 1, 0)(0, 1, 1)_{12}$ para los datos de la tabla 9-22. ¿Cree usted que se requiere un término constante en el modelo? Explique.
2. ¿El modelo sugerido en la pregunta 1 es el adecuado? Discuta en relación con las gráficas residuales, las autocorrelaciones residuales y los estadísticos chi cuadrada de Ljung-Box. Si el modelo no es adecuado, modifique y reajuste el modelo inicial hasta que usted sienta que ha logrado un modelo satisfactorio.
3. Usando el modelo que desarrolló en la pregunta 2, genere los pronósticos de los requerimientos de financiamiento para los 12 meses siguientes.
4. Escriba un reporte que resuma sus descubrimientos. Incluya en su reporte una gráfica de la serie original y los pronósticos.



Función de autocorrelación parcial: Financiamiento

Retraso	PACF	T	Retraso	PACF	T
1	0.837546	9.17	16	0.040984	0.45
2	-0.160317	-1.76	17	-0.010412	-0.11
3	-0.101375	-1.11	18	0.025745	0.28
4	-0.053070	-0.58	19	0.008608	0.09
5	-0.065903	-0.72	20	0.056303	0.62
6	0.200352	2.19	21	0.068118	0.75
7	0.064260	0.70	22	-0.000821	-0.01
8	0.277640	3.04	23	0.112046	1.23
9	0.072994	0.80	24	0.006096	0.07
10	0.175353	1.92	25	-0.231823	-2.54
11	0.207235	2.27	26	0.010987	0.12
12	0.192003	2.10	27	0.042028	0.46
13	-0.642453	-7.04	28	0.039898	0.44
14	0.047584	0.52	29	-0.004358	-0.05
15	-0.053889	-0.59	30	-0.018842	-0.21

FIGURA 9-47 Autocorrelaciones parciales de la muestra de los requerimientos de financiamiento

CASO 9-7 AAA WASHINGTON

En el caso 5-5 se presentó una visión general de AAA Washington cuando se pidió a los estudiantes que realizaran la descomposición de una serie de tiempo de las llamadas recibidas en el club de servicio de emergencias en el camino durante cinco años. La descomposición de la serie de tiempo realizada en el caso 5-5 indicaba que el patrón que Michael DeCoria observó en el volumen de llamadas del servicio de emergencias en el camino era probablemente de naturaleza cíclica. A Michael le gustaría predecir el efecto cíclico sobre el volumen de llamadas en años futuros.

En el caso 6-6 se investigaron cuatro modelos de regresión lineal usando el número total de llamadas al

servicio como la variable dependiente y la tasa de desempleo, la temperatura, la precipitación pluvial y el número de socios del club como las cuatro variables independientes. La variable de temperatura se transformó restando 65 grados del promedio de los valores mensuales de temperatura. Luego se investigó una relación no lineal.

En el caso 7-2 se desarrolló un modelo de regresión múltiple. Se probaron variables retrasadas 11 meses, tales como la precipitación, el número de socios del club, la temperatura mensual promedio transformada exponencialmente y la tasa de desempleo.

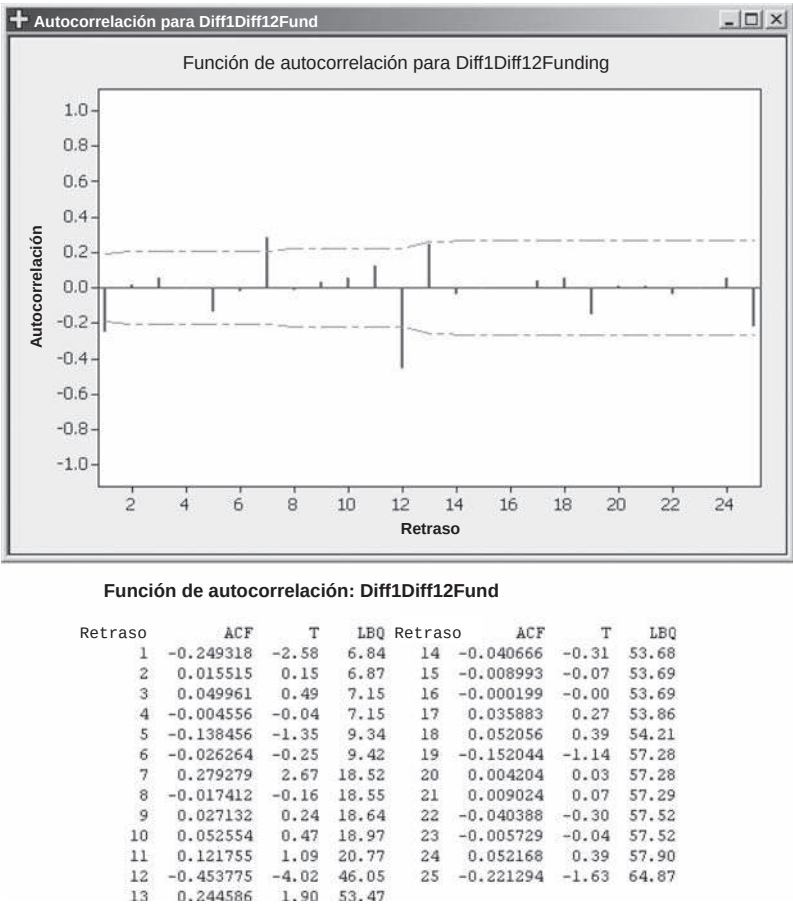


FIGURA 9-48 Autocorrelaciones de la muestra de los requerimientos de financiamiento diferenciados regularmente y diferenciados estacionalmente

En el caso 8-6 se verificó la correlación serial del modelo de regresión múltiple desarrollado en el caso 7-2, y se recomendó a Michael un modelo que se creyó era el más idóneo para predecir la naturaleza cíclica del volumen de llamadas al servicio de emergencias en el camino. Michael todavía no está satisfecho. El modelo presentado en el caso 8-6 explicaba sólo alrededor del 71% de la varianza en la variable del volumen de lla-

madas del servicio de emergencias en el camino. Él le pidió a usted intentar con la metodología Box-Jenkins para pronosticar el volumen de llamadas con los datos presentados en la tabla 9-23. Usted decide dividir el conjunto 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

1. Desarrolle un modelo ARIMA para el volumen de llamadas al servicio de emergencias en el camino. Elabore un escrito para Michael resumiendo sus descubrimientos.

TABLA 9-23 Volumen de llamadas del servicio de emergencia en el camino, por mes

<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
1989	Enero	23,327	1992	Septiembre	19,647
	Febrero	24,050		Octubre	22,013
	Marzo	24,010		Noviembre	22,375
	Abril	19,735		Diciembre	22,727
	Mayo	20,153		Enero	22,367
	Junio	19,512		Febrero	21,155
	Julio	19,892		Marzo	21,209
	Agosto	20,326		Abril	19,286
1990	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
	Abril	20,155		Diciembre	26,039
	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			

CASO 9-8 MINORISTA EN LA WEB

En el caso 4-6, Pat Niebuhr y su equipo estaban interesados en desarrollar pronósticos alternativos del total de pedidos y contactos por pedido (CPP) como parte del proyecto de un plan de dotación de personal para los centros de contacto de un minorista grande en la Web. El departamento de finanzas entregó a Pat los pronósticos del total de pedidos y CPP, pero como él tenía algunos datos históricos disponibles, decidió intentar la generación de los pronósticos alternativos usando métodos de suavización. Luego, podría comparar sus pronósticos con los que le dio el departamento de

finanzas. Sin embargo, la variable adecuada para propósitos de planeación es

$$\text{Contactos} = \text{Pedidos} \times \text{CPP}$$

y Pat estaba convencido, puesto que los datos estaban disponibles, que pronosticar directamente los contactos sería mejor que generar un pronóstico de contactos multiplicando un pronóstico de pedidos por un pronóstico de CPP. Pat tenía alguna capacitación en los méto-

TABLA 9-24 Contactos del minorista en la Web, de junio del 2001 a junio del 2003

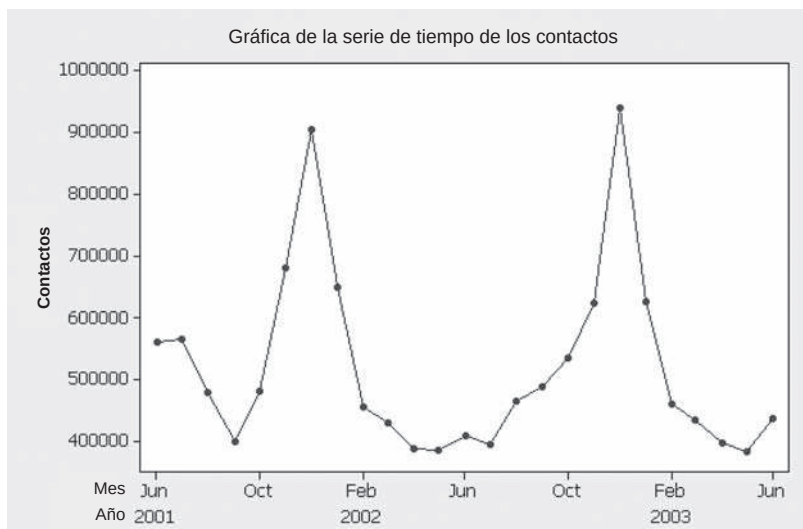
<i>Mes</i>	<i>Contactos</i>
Jun-01	561,516
Jul-01	566,056
Ago-01	478,739
Sept-01	398,672
Oct-01	482,174
Nov-01	681,482
Dic-01	906,251
Ene-02	649,727
Feb-02	454,580
Mar-02	430,809
Abr-02	387,877
Mayo-02	384,972
Jun-02	409,512
Jul-02	394,388
Ago-02	465,686
Sept-02	489,443
Oct-02	534,183
Nov-02	625,216
Dic-02	940,097
Ene-03	627,356
Feb-03	459,306
Mar-03	434,151
Abr-03	397,658
Mayo-03	383,449
Jun-03	437,019

dos Box-Jenkins y pensó que debía intentar con un modelo ARIMA. Pero Pat admitía que él no tenía muchas observaciones y se preguntaba si un modelo ARIMA era factible.

Los datos de contactos mensuales se presentan en la tabla 9-24. Una gráfica de la serie de tiempo de con-

tactos se presenta en la figura 9-49. Aun cuando él sólo tenía alrededor de dos años de datos, a partir de la gráfica de la serie de tiempo, Pat concluyó que el número de contactos era estacional.

Pat estaba inseguro acerca de la estacionalidad. Las autocorrelaciones de la muestra para la serie de

**FIGURA 9-49** Número mensual de contactos de un minorista en la Web

contactos se presentan en la figura 9-50. Existe una autocorrelación significativa en el retraso 1, pero los coeficientes de autocorrelación no se desvanecen lentamente como los de las series no estacionarias. Puesto que la gráfica de la serie de tiempo parece presentar una ligera tendencia descendente, Pat decidió diferenciar la serie de los contactos y examinar una gráfica y las autocorrelaciones de las diferencias. Los resultados se presentan en las figuras 9-51 y 9-52.

Las diferencias de los contactos parecen variar alrededor de cero, y la función de autocorrelación no indica una autocorrelación significativa, si bien las bandas de error son bastante amplias a causa del pequeño tamaño de la muestra. Existe una autocorrelación razonablemente grande (casi significativa) en el retraso 12, y como la gráfica de la serie de tiempo de los contactos presenta un patrón estacional, Pat decidió intentar un modelo estacional para los contactos.

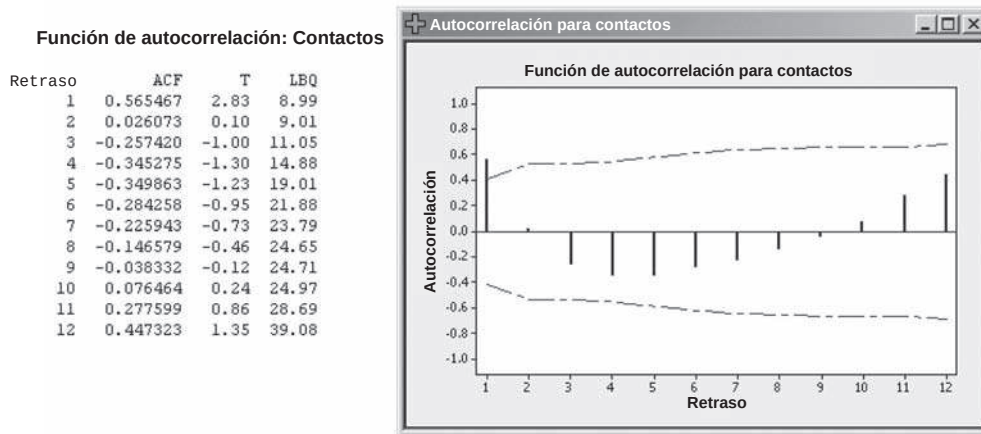


FIGURA 9-50 Función de autocorrelación de la muestra para los contactos

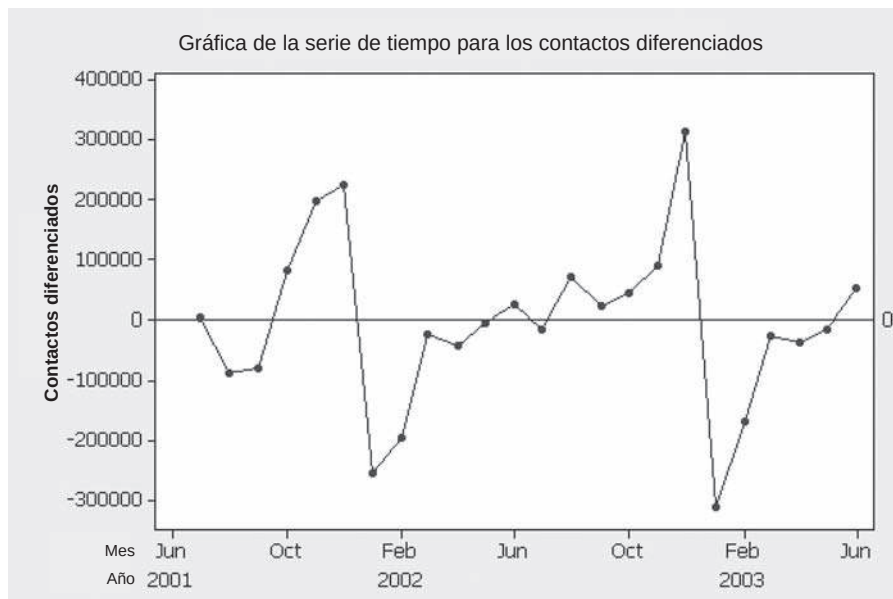


FIGURA 9-51 Gráfica de la serie de tiempo de los contactos diferenciados

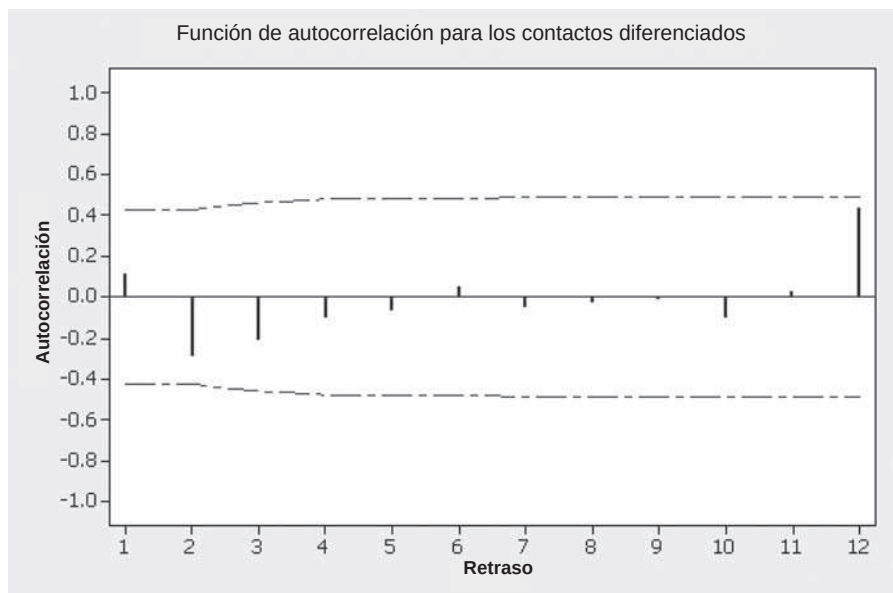


FIGURA 9-52 Función de autocorrelación de la muestra de los contactos diferenciados

PREGUNTAS

1. Usando Minitab o un software equivalente, ajuste un modelo $ARIMA(0, 1, 0)(0, 0, 1)_{12}$ a los datos de la tabla 9-24. ¿Por qué parece que este modelo es una opción inicial razonable? ¿Existe otro modelo candidato que usted pudiera considerar?
2. ¿El modelo sugerido en la pregunta 1 es adecuado? Discuta en relación con las gráficas residuales, las autocorrelaciones residuales y los estadísticos chi cuadrada de Ljung-Box. Si el modelo no es adecuado, modifique y reajuste el modelo inicial hasta que usted sienta que ha logrado un modelo satisfactorio.
3. A partir del modelo que usted desarrolló en la pregunta 2, genere los pronósticos de los contactos para los 12 meses siguientes. Incluya los pronósticos en una gráfica de la serie original de contactos. Elabore un breve reporte comentando sobre la sensatez de los pronósticos.
4. ¿Tiene usted alguna reserva acerca del uso del modelo ARIMA en esta situación? Discuta.

CASO 9-9 SURTIDO COOKIES

Los esfuerzos de Jaime Luna para modelar y pronosticar las ventas mensuales de Surtido Cookies se han documentado en los casos 3-5, 4-8, 5-7 y 8-8. En el caso 4-8 se analizó una gráfica de la serie de tiempo de las ventas y las autocorrelaciones correspondientes. En el caso 5-7 un análisis de descomposición de las ventas confirmó la existencia de un componente de tendencia y un componente estacional. Un modelo de regresión múltiple con variables ficticias de tendencia en el tiempo y estacionales se consideró en el caso 8-8. En un reciente seminario sobre pronósticos, Jaime conoció

los modelos ARIMA. Aprendió que para identificar un modelo ARIMA adecuado, por lo general es una buena idea examinar gráficas de las series de tiempo y varias de sus diferencias. Más aún, las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra de las series y sus diferencias brindan información útil para la identificación del modelo. Jaime decidió intentar ajustar un modelo ARIMA a sus datos y generar los pronósticos. Como las ventas de Surtido Cookies son estacionales, $S = 12$, Jaime construyó gráficas y funciones de autocorrelación de la serie origi-

nal, las diferencias de la serie original (DiffSales), las diferencias estacionales de la serie (Diff12Sales) y la serie con una diferencia regular seguida por una diferencia estacional (DiffDiff12Sales). Una gráfica de la serie de tiempo y la función de autocorrelación de

la muestra para Diff12Sales se presentan en las figuras 9-53 y 9-54, respectivamente. Luego de examinar esta información, Jaime seleccionó un modelo inicial ARIMA, pero necesita su ayuda.

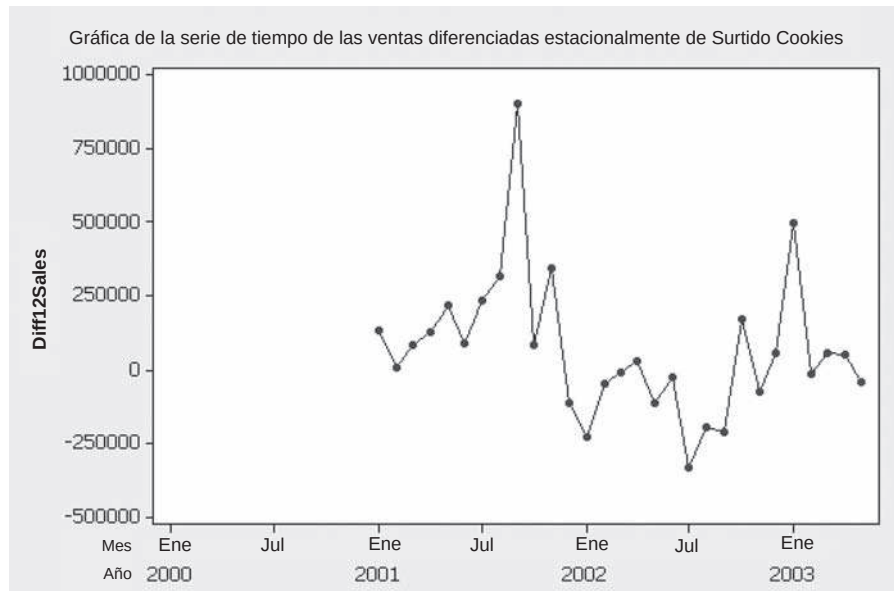


FIGURA 9-53 Gráfica de la serie de tiempo de las ventas diferenciadas estacionalmente de Surtido Cookies

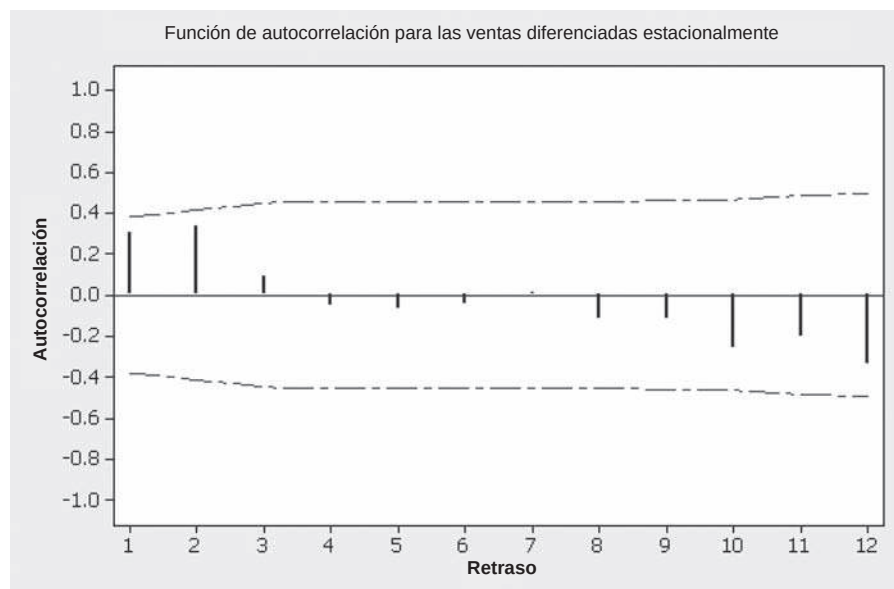


FIGURA 9-54 Autocorrelaciones de la muestra de las ventas diferenciadas estacionalmente de Surtido Cookies

PREGUNTAS

1. Usando Minitab o un software equivalente, ajuste un modelo $ARIMA(0, 0, 0)(0, 1, 1)_{12}$ a los datos de la tabla 3-12. ¿Por qué parece que este modelo es una opción inicial razonable para Jaime? Luego de examinar las gráficas y las autocorrelaciones de la serie original y las diferencias aparte de $Diff12Sales$, ¿existe otro modelo candidato que usted pudiera considerar?
2. ¿Es adecuado el modelo sugerido en la pregunta 1? Discuta acerca de las gráficas residuales, las autocorrelaciones residuales y los estadísticos chi cuadrada de Ljung-Box. Si el modelo no es adecuado, modifique y reajuste el modelo inicial hasta que usted sienta que ha logrado un modelo satisfactorio.
3. A partir del modelo que desarrolló en la pregunta 2, genere pronósticos de las ventas de Surtido Cookies para los 12 meses siguientes. Incluya los pronósticos en una gráfica de la serie original de las ventas de Surtido Cookies. Escriba un reporte breve describiendo la naturaleza de los pronósticos y si parecen razonables.

CASO 9-10 SOUTHWEST MEDICAL CENTER

Mary Beasley estaba muy decepcionada de sus esfuerzos para pronosticar. Ella intentó la suavización de Winters (véase el caso 4-7), la descomposición (véase el caso 5-8) y la regresión (véase el caso 8-9) con resultados menos que satisfactorios. Mary sabe que su serie de tiempo de visitas facturables totales mensuales al Medical Oncology es no estacionaria y estacional. Ella sabe, por uno de sus cursos de la maestría en administración, que los modelos ARIMA pueden describir las series de tiempo que son no estacionarias y estacionales,

y decide hacer un intento con un modelo ARIMA. Llegado a este punto, no tenía nada que perder.

Siguiendo el procedimiento establecido para examinar las gráficas, las autocorrelaciones y las autocorrelaciones parciales para la serie original y varias diferencias, Mary generó estas cantidades para la serie original (TotVisits), las diferencias de la serie (DiffTotVisits), las diferencias estacionales de la serie (Diff12TotVisits) y la serie con una diferencia regular seguida por una diferencia estacional (DiffDiff12TotVisits). En la figura

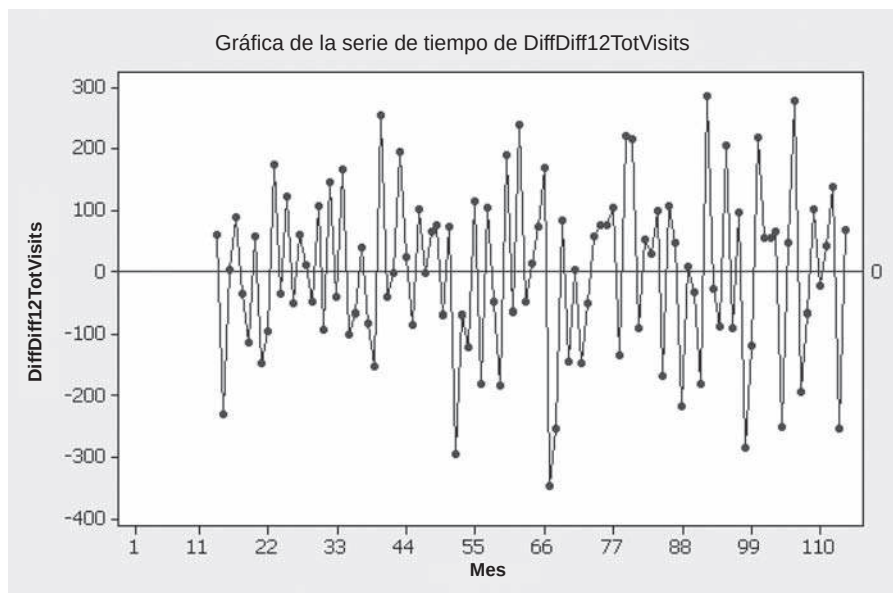


FIGURA 9-55 Gráfica de la serie de tiempo para DiffDiff12TotVisits

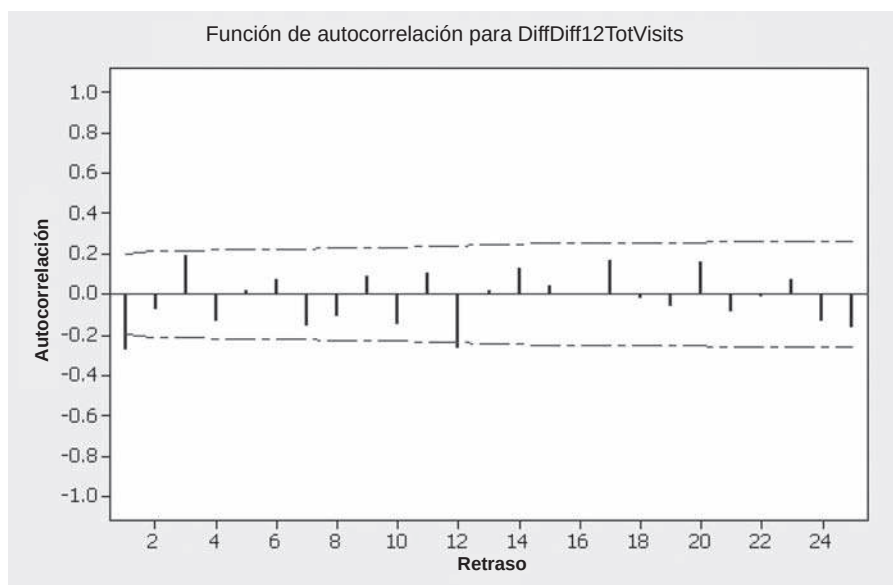


FIGURA 9-56 Autocorrelaciones de la muestra de DiffDiff12TotVisits

9-55 se presenta la gráfica de DiffDiff12TotVisits. La función de autocorrelación correspondiente aparece en la figura 9-56.

Al examinar la gráfica de la serie de tiempo, Mary ve que las diferencias parecen ser estacionarias y variar alrededor de cero. La función de autocorrelación en la figura 9-56 parece tener dos autocorrelaciones negati-

vas significativas en los retrasos 1 y 12. Esto sugiere un parámetro de promedio móvil en el retraso 1 y un parámetro de promedio móvil estacional en el retraso 12 en un modelo ARIMA con una diferencia regular y una diferencia estacional. Mary está lista para ajustar un modelo $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ para sus datos.

PREGUNTAS

1. Con base en los datos de Mary Beasley de la tabla 4-13, construya la gráfica de una serie de tiempo y calcule las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra de las visitas facturables totales, Y_t ; la serie diferenciada, ΔY_t ; y la serie diferenciada estacionalmente, $\Delta_{12} Y_t$. Dada esta información y la de las figuras 9-55 y 9-56, ¿piensa usted que el modelo inicial de Mary es razonable? Discútalalo.
2. Usando Minitab o un software equivalente, ajuste un modelo $ARIMA(0, 1, 1)(0, 1, 1)_{12}$ para los datos de

Mary Beasley y genere los pronósticos para los 12 meses siguientes. Examine los residuos para verificar la idoneidad del ajuste. ¿Debería Mary estar contenta ahora con los pronósticos de las visitas facturables totales?

3. Al examinar la gráfica de la serie de tiempo de las visitas facturables totales a Medical Oncology, ¿puede usted ver alguna(s) característica(s) de los datos que pudiera(n) dificultar la modelación?

Aplicaciones de Minitab

El problema. En el ejemplo 9.10 Perkin Kendell, un analista de Coastal Marine Corporation, quiere pronosticar las ventas para 2007.

Solución de Minitab

1. Si los datos están almacenados en un archivo, ábralo usando los siguientes menús:

File>Open Worksheet

Si los datos no están en un archivo, introdúzcalos en una hoja de cálculo.

2. Para calcular las autocorrelaciones de la variable de ventas, haga clic en

Stat>Time Series>Autocorrelation

3. Aparecerá la ventana de diálogo de la función de autocorrelación de la figura 3-28.

a) Haga doble clic en la variable Sales y aparecerá a la derecha de Series:

b) Haga clic en OK y aparecerá la figura 9-33.

4. Para diferenciar estacionalmente los datos, haga clic en los siguientes menús:

Stat>Time Series>Differences

5. Aparecerá la ventana de diálogo Differences de la figura 3-29.

a) Haga doble clic en la variable Sales y aparecerá a la derecha de Series:

b) Posiciónese en Store differences in: e introduzca C2.

c) Posiciónese en Lag: e introduzca 4. Haga clic en OK y aparecerán las cuartas diferencias en la columna 2 iniciando en la fila 5.

6. En C2 titule la variable Diff4Sales. Para calcular las autocorrelaciones para esta variable, repita el paso 2 usando Diff4Sales como la variable a la derecha de Series:

7. Para calcular las primeras diferencias de la variable Diff4Sales, repita el paso 5, guardando las diferencias en C3 y usando un retraso (Lag:) de 1.

8. Etiquete en C3 la variable Diff1Diff4Sales. Para calcular las autocorrelaciones de esta variable, repita el paso 2 usando Diff1Diff4Sales como la variable a la derecha de Series:

9. Para calcular las autocorrelaciones parciales de la variable Diff1Diff4Sales, haga clic en

Stat>Time Series>Partial Autocorrelation

10. Aparecerá la ventana de diálogo Partial Autocorrelation Function, que es similar a la de la figura 3-28.

a) Haga doble clic en la variable Diff1Diff4Sales, y aparecerá a la derecha de Series:

b) Haga clic en OK y aparecerá la figura 9-36.

11. Para correr un modelo ARIMA(1, 1, 0)(0, 1, 2), haga clic en los siguientes menús:

Stat>Time Series>ARIMA

12. Aparecerá la ventana de diálogo de la figura 9-57.

a) Haga doble clic en la variable Sales y aparecerá a la derecha de Series:

b) Haga clic en la casilla de verificación a la izquierda de Fit seasonal model e indique 4 a la derecha de Period:.

c) Debajo de Nonseasonal y a la derecha de Autorregresive: ponga un 1 a la derecha de Difference: y un 0 a la derecha de Moving average:

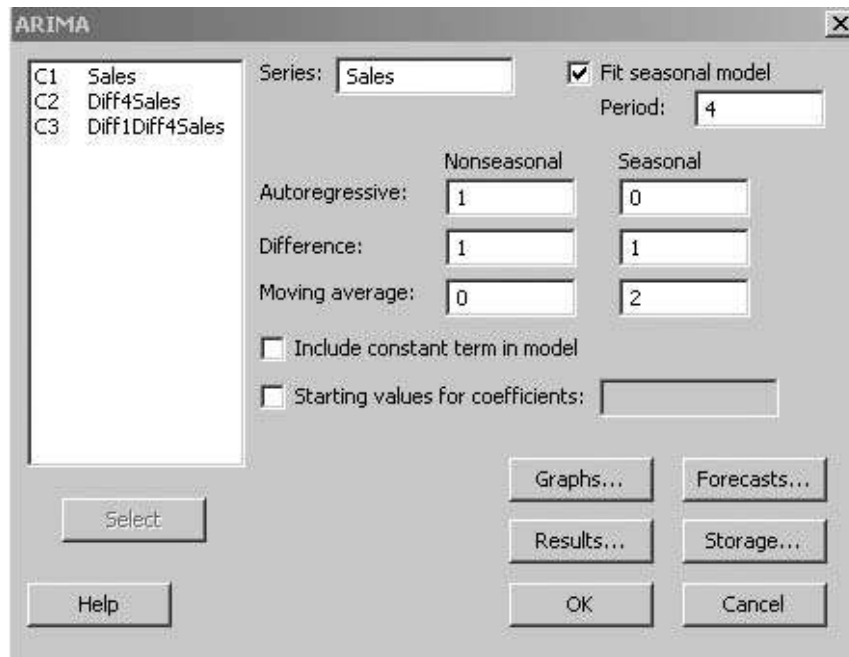


FIGURA 9-57 Ventana de diálogo ARIMA de Minitab

- d) Debajo de Seasonal y a la derecha de Autorregresive: ponga un 0, un 1 a la derecha de Difference: y un 2 a la derecha de Moving average:
 - e) Como los datos se han diferenciado, desactive la casilla de verificación de Include constant term model.
 - f) Haga clic en Forecasts y aparecerá la ventana de diálogo ARIMA-Forecasts. Para pronosticar dos periodos del futuro, ponga un 2 a la derecha de Lead:. Haga clic en OK.
 - g) Haga clic en Storage y aparecerá la ventana de diálogo ARIMA-Storage. Haga clic en el cuadro siguiente a Residuals y luego haga clic en OK. Haga clic en OK sobre la ventana de diálogo de ARIMA y aparecerá la parte inferior de la tabla 9-16.
 - h) Para calcular las autocorrelaciones de los residuos, repita el paso 2 usando RESI1 como la variable a la derecha de Series:
13. Para desarrollar la gráfica de una serie de tiempo que incluye un pronóstico, haga clic en Graphs en la ventana de diálogo ARIMA y aparecerá la ventana de diálogo ARIMA-Graphs mostrada en la figura 9-58.
- a) Haga clic en la casilla de verificación a la izquierda de Time series plot.
 - b) Un método alternativo para obtener las autocorrelaciones de los residuos es hacer clic en la casilla de verificación a la izquierda de ACF of residuals.
 - c) Haga clic en OK. Haga clic en OK de la ventana de diálogo ARIMA y aparecerá la figura 9-38.

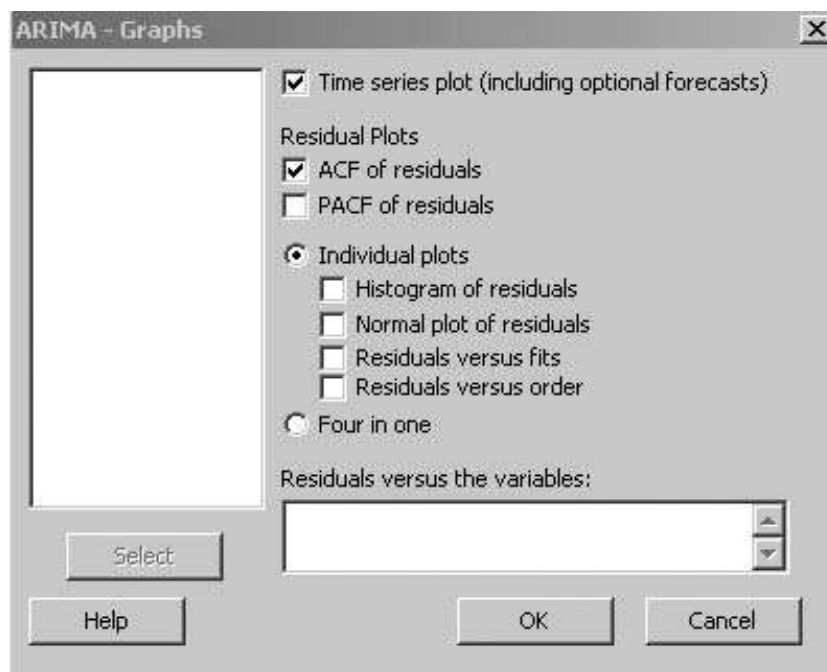


FIGURA 9-58 Minitab ARIMA-Graphs Dialog Box

Referencias

- Akaike, H. "A new Look at the Statistical Model Identification". *IEEE Transactions Automatic Control* 19 (1974): 716-723.
- Bernstein, P. *Against the Gods: The Remarkable Story of Risk*. Nueva York: Wiley, 1996.
- Bowerman, B. L., R. T. O'Connell y A.B. Koehler. *Forecasting, Time Series and Regression*, 4a. ed. Belmont, Calif.: Thomson Brooks/Cole, 2005.
- Box, G. E. P., G. M. Jenkins y G. C. Reinsel. *Time Series Analysis: Forecasting and Control*, 3a. ed. Upper Saddle River, N.J.: Prentice Hall, 1994.
- Diebold, F. X. *Elements of Forecasting*, 3a. ed. Cincinnati, Ohio: South-Western, 2004.
- Kyle, P. W. "Lydia Pinkham Revisited: A Box-Jenkins Approach". *Journal of Advertising Research* 18(2)(1978): 31-39.
- Levenbach, H. y J. P. Cleary. *Forecasting Practice and Process for Demand Management*. Belmont, Calif.: Thomson Brooks/Cole, 2006.
- Newbold, P. y T. Bos. *Introductory Business and Economic Forecasting*, 2a. ed. Cincinnati, Ohio: South-Western, 1994.
- Ord, K. y S. Lowe. "Automatic Forecasting". *American Statistician* 50(1996): 88-94.
- 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*. Englewood Cliffs, N. J.: Prentice Hall, 1964.
- Pindyck R. S. y D. L. Rubinfeld. *Econometric Models and Economic Forecasts*, 4a. ed. Nueva York: McGraw-Hill, 1998.
- Quenouille, M. H. "The Joint Distribution of Serial Correlation Coefficients". *Annals of Mathematical Statistics* 20(1994): 561-571.
- Schwarz, G. "Estimating the Dimension of a Model". *Annals of Statistics* 6(1978): 461-464.

CAPÍTULO 10

PRONÓSTICOS DE JUICIO Y AJUSTE DE PRONÓSTICO

Las técnicas de elaboración de pronósticos en este libro implican la manipulación de datos históricos para generar predicciones o pronósticos de importantes variables de interés. Los temas en capítulos anteriores estaban relacionados con análisis complejos de datos y quizá implicaban que el criterio del analista permanecía al margen. El hecho de descartar el criterio en la interpretación del método cuantitativo para pronosticar es definitivamente una visión errónea. De hecho, como se enfatizó en el capítulo 1, el buen juicio es un componente esencial de todas las técnicas de pronóstico. Se requiere buen juicio al determinar los datos que son relevantes para el problema, al seleccionar un modelo adecuado y al interpretar los resultados del proceso de análisis. Algunas veces el criterio participa en buena parte del análisis.

Este capítulo analiza algunos de los principales elementos del pronóstico que se adjuntan o que complementan el manejo sistemático de datos históricos. En el capítulo 11 se describen algunas consideraciones sobre el manejo del proceso de pronósticos.

En muchas situaciones de pronóstico, sólo se emplea el análisis de datos históricos para generar el pronóstico final; el juicio o la opinión del analista no se considera¹ en el proceso. Este libro se enfoca principalmente en esas técnicas para pronosticar y, por consiguiente, en los pronósticos de corto y mediano plazos. Tales pronósticos son un elemento esencial en la mayoría de los niveles administrativos de una organización y están asociados a la mayoría de las decisiones cruciales que se deben tomar.

Estos procedimientos dependen de la manipulación de datos históricos y suponen que el pasado y el futuro son iguales, excepto por las variables específicas identificadas que afectan la probabilidad de resultados futuros. Esta suposición excluye un cambio importante en el ámbito tecnológico de la sociedad, una suposición que se considera errónea a la luz de varios hechos recientes. Considere, por ejemplo, la introducción y proliferación de la alta velocidad, las computadoras personales, los teléfonos celulares y los asistentes personales manuales de bajo costo, entre otros avances.

En algunas situaciones que requieren de pronósticos, el analista complementa el proceso de análisis después de considerar circunstancias inusuales de la situación o después de reconocer que la historia pasada no es un pronóstico exacto del futuro. El criterio aportado al proceso del pronóstico tiende a aumentar cuando los datos históricos son escasos o cuando se cree que éstos son parcialmente irrelevantes. En casos extremos, tal vez el analista piense que los datos históricos no son directamente relevantes para el proceso de pronóstico. En estas condiciones, al formular el pronóstico o escenario para el futuro, se utilizan pronósticos basados por completo en las opiniones de “expertos”.

¹Usamos aquí el término *analista* para designar a cualquier persona que elabore pronósticos de juicio o ajustes a los pronósticos basados en modelos. En este contexto, el analista puede o no ser la persona que elaboró el pronóstico con base en un modelo.

Es conveniente pensar en el pronóstico en términos de la información subyacente. Para métodos de suavización, descomposición y procedimientos ARIMA, el conjunto de información subyacente consiste en la historia de la serie de tiempo que se está pronosticando. Para los métodos de regresión, el conjunto de información también está integrado por datos históricos, sólo que se trata de datos históricos de más de una variable. Los pronósticos de juicio o ajustes de pronósticos se basan generalmente en el dominio del conocimiento. El dominio del conocimiento es cualquier información relevante diferente de la serie de tiempo que se usa para la elaboración del pronóstico. Es decir, el dominio del conocimiento no es la información de la serie de tiempo. Ahora, en principio, más información debería traer consigo mejores pronósticos. Así, los pronósticos basados en datos históricos y el dominio del conocimiento deben ser mejores que los pronósticos que se basan solamente en un grupo u otro de información. Sin embargo, la evidencia disponible está mezclada.

En ocasiones, los pronósticos de juicio son la única alternativa, pero pueden ser mejores o peores que los pronósticos cuantitativos cuando ambos están disponibles. Si los pronósticos de juicio son el resultado del presupuesto organizacional o de incentivos (objetivos a lograr), pueden ser peores que los pronósticos cuantitativos. Si los pronósticos de juicio se generan a partir de conocimiento actualizado, mientras que los métodos estadísticos tienen un atraso en la disponibilidad de los datos, y/o si el dominio del conocimiento representa un componente que no puede modelarse (por ejemplo, las acciones de un competidor), los pronósticos de juicio serán más precisos.

Cuando se emplea el juicio para modificar un pronóstico cuantitativo, la investigación sugiere que, en general, las modificaciones de juicio tienden a reducir la exactitud del pronóstico. Sin embargo, las modificaciones de juicio pueden tener como resultado una mayor precisión si el dominio del conocimiento es particularmente relevante y si los ajustes tienden a ser “grandes” en vez de “pequeños”.

Como hemos mencionado, cuando no se dispone de información histórica o cuando los datos históricos son escasos, se debe confiar en el juicio si se desean pronósticos o predicciones acerca del futuro. Como tales situaciones se presentan con frecuencia, especialmente en la alta dirección, se han desarrollado técnicas para mejorar la precisión de los pronósticos aprovechando el juicio de los ejecutivos de la mejor manera. El uso de estas técnicas es digno de consideración, en vista de que los ejecutivos consideran a menudo que su propio juicio es superior a otros métodos para predecir el futuro. Como afirma Makridakis (1986), “la gente prefiere hacer pronósticos usando el criterio. Ellos creen que su conocimiento del producto, del mercado y de los clientes, en conjunción con su entendimiento y la información interna, les confiere la capacidad para pronosticar juiciosamente” (p. 63).

A continuación se presentan algunas preguntas, cada una de las cuales sugiere el uso de la imaginación y de la lluvia de ideas con la finalidad de no depender completamente de la recopilación y la manipulación de datos históricos. Para contestar cada una de estas preguntas, alguna de las técnicas para pronosticar examinadas en este capítulo ayudará al equipo de dirección de la compañía a obtener una comprensión valiosa del ambiente operativo futuro de la empresa.

- ¿Cuál será la distribución de la edad en Estados Unidos para el año 2030?
- ¿Cuántos estadounidenses trabajarán en casa dentro de 25 años a partir de ahora?
- ¿Qué ciudades tendrán la mayor población y serán centros de negocios dentro de 20 años?
- ¿Hasta qué punto Estados Unidos dependerá de otros países en cuanto a la manufactura de artículos clave para el consumidor al término de 10 años?
- ¿Qué tan populares serán en 20 años las compras desde casa usando la televisión y las computadoras?
- ¿Qué tipos de diversión disfrutarán los ciudadanos de Estados Unidos en el año 2025?
- ¿Estados Unidos comenzará a retirarse de sus negocios alrededor del mundo durante los próximos 25 años? Si es así, ¿de qué manera afectará esto a los hombres de negocios?

Ahora describiremos algunas de las técnicas para elaborar *pronósticos de juicio*, llamados así porque el juicio es el componente único o más importante del proceso.² Algunas técnicas se conocen también como métodos tecnológicos de pronóstico, puesto que con frecuencia suponen la proyección de los efectos de cambios tecnológicos hacia un futuro incierto.

PRONÓSTICOS DE JUICIO

El método Delphi

Cuando los expertos están reunidos en una sala de juntas y se preguntan acerca del futuro, la dinámica de grupo puede distorsionar algunas veces el proceso y dar como resultado un consenso que no es pensado cuidadosamente por todos los participantes. El método Delphi se usó primero en la Fuerza Aérea —en un proyecto fundado por la RAND Corporation en la década de 1950, que pretendía eliminar el aspecto dinámico de las deliberaciones de los pronosticadores. En la primera ronda del método, los expertos respondieron por escrito las preguntas planteadas por un equipo de investigación de la organización. Luego, el equipo hizo un resumen de los comentarios de los participantes y lo entregó a éstos por correo. Los participantes estaban entonces en posibilidad de leer las reacciones de los demás y de defender su punto de vista original o de modificarlo con base en las opiniones de los otros.

Este proceso continúa durante dos o tres rondas más hasta que los diversos puntos de vista desarrollados cuidadosamente satisfacen a los investigadores. Entonces se invita a los participantes a reunirse para compartir y debatir sus puntos de vista. Al final de este proceso, el equipo investigador tiene una buena comprensión del futuro y, por consiguiente, está en condiciones de empezar a planear la postura de la organización.

Como Rowe y Wright (1999) apuntaron:

Delphi no es un procedimiento diseñado para desafiar a la estadística o a los procedimientos basados en un modelo, contra los cuales el juicio humano ha demostrado ser inferior: es una propuesta para usarse en situaciones que requieren de juicio y pronósticos en las que los métodos estadísticos puros basados en un modelo resultan imprácticos o no son posibles por la falta de datos históricos, económicos o técnicos apropiados... (p. 354)

Cualquier procedimiento Delphi tiene cuatro características fundamentales: anonimato, iteración, retroalimentación controlada y agregación de respuestas del grupo. El anonimato se obtiene a través del uso de cuestionarios administrados manualmente, por correo, por teléfono o a través de una red de computadoras. Cuando los expertos expresan su opinión de manera privada, están libres de presiones sociales del grupo y pueden concentrarse sólo en los méritos de cada idea, propuesta o situación. Más aún, con la iteración de los cuestionarios en varias rondas, los expertos participantes tienen la oportunidad de cambiar sus opiniones sin perder el prestigio ante los ojos del resto de los participantes (gracias al anonimato).

Después de cada iteración de los cuestionarios, las opiniones de cada experto se resumen brevemente, y cada miembro del grupo es informado de las posiciones de sus colegas anónimos. En consecuencia, la retroalimentación contiene las opiniones y los juicios de todos los miembros del grupo y no sólo las más elocuentes. Después de la ronda final, se resume el juicio del grupo (usando a menudo técnicas de estadística descriptiva, como promedios o cuartiles para describir rangos) de manera que cada juicio recibe el mismo peso.

Por lo general, la primera ronda del método Delphi no está estructurada, lo que da a los expertos completa libertad para identificar y comentar sobre aquellos asuntos que consideren

²Para conocer un esquema organizado de diversos métodos para pronosticar incluyendo los pronósticos de juicio, consulte a Georgoff y Murdick (1986).

importantes. El equipo de monitoreo consolida luego estos temas en un grupo individual de declaraciones estructuradas (el cuestionario) a partir del cual las opiniones y los juicios de los participantes en Delphi se pueden reunir en términos cuantitativos en rondas posteriores. Los resultados de cada ronda se resumen y se entregan a cada experto para una consideración ulterior. Después de la segunda ronda, se da la oportunidad a los participantes de cambiar sus opiniones anteriores con base en la retroalimentación proporcionada. En la mayoría de las aplicaciones de la técnica Delphi, el número de rondas rara vez excede de tres. Estudios empíricos del método Delphi sugieren que la precisión tiende a aumentar durante las rondas Delphi y los participantes tienden a ser más precisos que los grupos que interactúan de una forma no estructurada.

Ejemplo 10.1

Applied Biosystems suministra a las empresas de ciencias de la vida y a las instituciones de investigación una diversidad de productos y servicios para investigar genes y proteínas, estudiar la interacción de medicamentos con los sistemas del cuerpo y la composición genética, para efectuar pruebas de contaminación en alimentos y el ambiente, así como estudios de identificación del DNA. En los primeros años de la compañía, los sistemas de instrumentación y equipo generaban la mayoría de los ingresos por ventas. En los años recientes, el área del negocio que vende consumibles (reactivos, servicios, material para análisis) ha crecido rápidamente, y ahora los consumibles representan alrededor de la mitad de los ingresos de la compañía. Applied Biosystems hace la mayoría de sus negocios en Estados Unidos, pero está considerando expandir su presencia en Europa, Japón y Australia, donde actualmente tiene porcentajes relativamente pequeños del mercado de equipo para biosistemas. Antes de aventurar gastos de capital asociados con la expansión global, la compañía está interesada en desarrollar pronósticos del crecimiento de las ventas en Europa, Japón y Australia para los próximos 10 años. Con pocos datos históricos sobre ventas en estas regiones que sirvan de guía, Applied Biosystems decidió contratar a tres expertos familiarizados con la investigación en ciencias de la vida y en la situación económica internacional.

El experto A es un genetista que tiene amplia experiencia en la industria farmacéutica y pasó algún tiempo en Europa estudiando el crecimiento de la investigación vinculada con la genética. El experto B es un economista notable y tiene gran experiencia en el estudio de la economía japonesa, en particular. La experta C, originaria de Australia, trabajó algún tiempo en las oficinas centrales de la Organización Mundial de la Salud en tecnologías de la salud y el área farmacéutica. Actualmente trabaja como consultora para el gobierno y la industria de Australia.

A Applied Biosystems le gustaría pronosticar el crecimiento de las ventas tanto en equipos como en consumibles durante los próximos 10 años en cada región: Europa, Japón y Australia. Para empezar, se le pidió a cada experto que proporcionara una estimación del crecimiento de las ventas de instrumentación, equipo y de los consumibles en cada región para los próximos 10 años. A cada uno se le informó el nivel actual de las ventas anuales y una estimación de la participación en el mercado por grupo para Europa, Japón y Australia. Con el dinero presupuestado, Applied Biosystems podrá obtener el compromiso de los tres expertos para dos rondas de reportes. Los expertos, quienes no se conocen entre sí y que están ubicados en diferentes partes del mundo, están comunicados por computadora con el gerente del proyecto de Applied Biosystems.

Los resultados después de la primera ronda se resumen en la tabla 10-1.

En la primera ronda, el experto A no ve mucho crecimiento en las ventas de productos y servicios de Applied Biosystems en Europa para los próximos 10 años. Si bien la investigación genética en Europa probablemente va a crecer de forma considerable, la competencia entre las firmas europeas y no europeas con una fuerte presencia en el mercado biológico probablemente limitará las oportunidades de Applied Biosystems a pesar de su excelente reputación. Asimismo, el experto A ve un crecimiento relativamente limitado de las ventas de instrumental y equipo de Applied Biosystems en Japón debido a la proclividad de los investigadores japoneses a usar el excelente equipo japonés. Sin embargo, considera que existe una oportunidad de crecimiento apreciable en la venta de consumibles en Japón. El experto A no tiene una opinión acerca del crecimiento de las ventas de Applied Biosystems en Australia.

El experto B prevé un crecimiento limitado para las ventas del equipo e instrumental de Applied Biosystems en Europa y Japón, pero una oportunidad sustancial de crecimiento en Australia, aun cuando el mercado australiano, como un todo, es relativamente pequeño. El experto B es un poco más optimista acerca del crecimiento de las ventas en Europa que el experto A, pero ambos están de acuerdo en que el crecimiento de las ventas, particularmente de consumibles, probablemente va a ser pequeño en los próximos 10 años. El experto B piensa que hay probabilidades de que haya un enorme crecimiento en las ventas tanto de equipo e instrumental como de consumibles en Australia, y que la

TABLA 10-1 Resultados de la primera ronda del método Delphi para el ejemplo 10.1

	<i>Experto A</i>	<i>Experto B</i>	<i>Experto C</i>
Equipo/Instrumental Europa	+10% a +40%	+30% a +70%	+0% a +150%
Consumibles Europa	+5% a +15%	+10% a +25%	+20% a +60%
Equipo/Instrumental Japón	+0% a +50%	−10% a +40%	—
Consumibles Japón	+40% a +100%	+50% a +200%	—
Equipo/Instrumental Australia	—	+100% a +200%	+150% a +300%
Consumibles Australia	—	+200% a +400%	+200% a +400%

venta de consumibles es particularmente atractiva. El experto B considera que Applied Biosystems no tiene gran competencia en Australia y que el compromiso en ese país por el cuidado de la salud y el mantenimiento de un ambiente de calidad seguirá siendo fuerte.

La experta C no está tan segura acerca del crecimiento de las ventas de equipo e instrumental en Europa, como los expertos A y B, y sugiere un amplio rango de 0 al 150%. Su pronóstico del crecimiento de las ventas de consumibles en Europa, si bien es más optimista que los de los expertos A y B, es aún relativamente pequeño. La experta C no tiene opinión acerca del crecimiento de las ventas en Japón de Applied Biosystems. Pero es muy optimista sobre el crecimiento potencial de las ventas de los productos y servicios de Applied Biosystems en su nativa Australia. Sus estimaciones del crecimiento de las ventas son grandes tanto para equipo e instrumental como para consumibles, y son del mismo orden en magnitud que las del experto B.

La información de la tabla 10-1 se entrega a los expertos, a quienes se les solicita una segunda ronda de pronósticos sobre el crecimiento en las ventas. Cada experto tiene la oportunidad de ajustar su pronóstico original después de revisar los resultados de la primera ronda. Las opiniones del experto A permanecen esencialmente sin cambio. El experto B ajusta su rango del crecimiento de las ventas del equipo e instrumental en el mercado europeo de +20% a +60%, un ligero ajuste hacia abajo. También ajusta hacia abajo el límite superior de su rango para el crecimiento de la venta de consumibles en el mercado japonés a +150%. Deja sin cambio sus pronósticos de crecimiento de las ventas en el mercado australiano. La experta C ajusta los rangos de sus pronósticos de crecimiento de las ventas para el mercado europeo a

Equipo/Instrumental Europa	+10% a +90%
Consumibles Europa	+15% a +40%

Ella deja sin cambio sus pronósticos para Australia.

Una llamada telefónica final a los participantes en el método Delphi con los resultados de la segunda ronda no produce algún cambio adicional. Con los resultados del ejercicio de Delphi a la mano, Applied Biosystems decide mantener su presencia actual en Europa, con mayor énfasis en las ventas de equipo e instrumental. La compañía decide desplegar un mayor esfuerzo de marketing en Japón para incrementar las ventas de productos en sus grupos de consumibles, con la posibilidad de asociarse con una empresa japonesa en los esfuerzos de venta y distribución. Australia parece representar una mejor oportunidad, pero el mercado total es pequeño, y Applied Biosystems decide hacer un pequeño análisis más de costo-beneficio antes de comprometerse en el mercado australiano.

La ventaja del método Delphi es que a los expertos notables se les puede pedir que consideren cuidadosamente el tema de interés y que contesten reflexivamente a los puntos de vista de otros sin la interferencia de la dinámica de grupo. El resultado, si el proceso se maneja con cuidado, puede ser un buen consenso acerca del futuro, junto con varios escenarios alternativos.³

Formulación de escenarios

La formulación de escenarios implica la definición detallada de un futuro incierto a través de la redacción de un “guion” para el ambiente de una organización durante varios años del futuro. La nueva tecnología, los cambios demográficos y el cambio en las demandas del consumidor

³Para una descripción detallada del método Delphi, véase Parente y Anderson-Parente (1987).

son, entre otros, los factores que se consideran y entretienen en la especulación para provocar la reflexión de los altos directivos.

Se elabora un escrito sobre el escenario más probable, junto con uno o más escenarios menos probables, pero posibles. Considerando la postura de la compañía para cada uno de estos ambientes futuros posibles, la alta dirección está en una mejor posición para reaccionar cuando ocurran los cambios reales del ambiente del negocio y para reconocer las implicaciones a largo plazo de cambios sutiles que, de otro modo, pasarían inadvertidos. De este modo, la organización está en una mejor posición para mantener su rentabilidad a largo plazo en vez de concentrarse en las utilidades de corto plazo e ignorar el ambiente tecnológico cambiante en el cual opera.

Al proceso de describir por escrito el escenario le sigue una fase de debate, algunas veces realizado por un grupo diferente del que desarrolló los escenarios. El debate entre los grupos se emplea entonces para defender y modificar los puntos de vista, de manera que se pueda desarrollar un sólido consenso y escenarios alternativos. Por ejemplo, tal vez el personal de planeación de la compañía desarrolle los escenarios, y luego sea el equipo de la alta dirección quien los discuta. Incluso si ninguno de los escenarios prueba más adelante ser totalmente verdadero, este proceso estimula el pensamiento a largo plazo del equipo de alta dirección y lo prepara mejor para reconocer y reaccionar ante los cambios importantes del ambiente.

Ejemplo 10.2

Una compañía que se dedica a fabricar cables para teléfonos y televisión decide realizar un ejercicio de formulación de escenarios antes de su reunión anual durante un fin de semana. Se pide a cada miembro del grupo que describa por escrito tres escenarios que podría enfrentar la compañía dentro de cinco años: el peor de los escenarios, el más probable y el mejor escenario. Después de que se han realizado las tareas de redacción, y justo antes del fin de semana de la reunión, el presidente y su vicepresidente ejecutivo resumen las contribuciones en los tres escenarios siguientes sobre los cuales intentan que se enfoque la discusión del grupo durante la reunión de dos días:

1. El uso de Internet continúa creciendo rápidamente pero poco a poco abandona el cable en favor del acceso satelital. Incluso el servicio telefónico depende cada vez más de medios diferentes al cable como lo hace la televisión casera. La compañía mira sus ventas y utilidades en una declinación sin fin y pronto tiene que salir del mercado.
2. Internet y el servicio de televisión doméstica continúan creciendo rápidamente, pero son suministrados por varios proveedores. Los servicios satelitales se utilizan ampliamente, pero los cables ocultos continúan siendo parte integral de los servicios de alta tecnología, especialmente en ciudades grandes, tanto en las aplicaciones domésticas como industriales. El liderazgo de la compañía en el desarrollo del cable y su distribución tiene como resultado ventas y utilidades crecientes.
3. A causa de los problemas técnicos y los detalles de seguridad, el uso del satélite para Internet y servicio telefónico declina hasta restringir su uso principalmente en zonas rurales. El uso del cable crece rápidamente tanto en aplicaciones residenciales como industriales, y la compañía prospera conforme sus productos bien posicionados y su liderazgo se traducen en un predominio en la industria.

El presidente de la compañía y el vicepresidente ejecutivo intentan discutir ampliamente acerca de cada uno de estos tres escenarios. Ellos quieren desarrollar estrategias de largo alcance que se ajusten a todas las posibilidades futuras y creen que al enfocarse en estos tres casos se fortalecerán ellos mismos junto con su equipo administrativo.

Combinación de pronósticos

Una rama en desarrollo del estudio de los pronósticos implica la combinación de dos o más métodos para generar los pronósticos finales. Un número del *International Journal of Forecasting* contenía una sección especial sobre esta nueva técnica. Partes de los resúmenes de tres artículos sobre este asunto ilustran la naturaleza en desarrollo de la combinación de pronósticos.

1. De acuerdo con Armstrong (1989), la investigación proveniente de unos 200 estudios demuestra que la combinación de pronósticos da por resultado una precisión modesta pero consistente. Sin embargo, esta investigación no define bien las condiciones en las cuales la combinación es más efectiva o cómo se deben combinar los métodos en cada situación.

2. Mahoud (1989) indica que la cantidad de investigación sobre la combinación de pronósticos es considerable. Sin embargo, se conoce relativamente poco acerca de cuándo y cómo los gerentes combinan los pronósticos. Los asuntos administrativos importantes que requieren de mayor estudio son el ajuste administrativo de los pronósticos cuantitativos, el uso de sistemas expertos en la combinación de pronósticos y el análisis de los costos de la combinación de pronósticos.
3. Existe mucha bibliografía acumulada durante años en lo que concierne a la combinación de pronósticos. La principal conclusión de esta línea de investigación, de acuerdo con Clemen (1989), es que la exactitud del pronóstico puede mejorarse sustancialmente mediante la combinación de muchos pronósticos individuales. Este documento ofrece un repaso y bibliografía sobre el tema.

En un mundo en el que los conjuntos de información pueden combinarse de forma rápida y a bajo costo, siempre es mejor combinar conjuntos de información en vez de pronósticos. Sin embargo, cuando la mezcla de los conjuntos de información es imposible (información cualitativa contra información cuantitativa) o el costo es prohibitivo (se deben cumplir plazos para obtener los pronósticos de manera oportuna), la combinación de pronósticos es la única alternativa. La combinación de pronósticos se puede ver como un vínculo entre el proceso de generación de pronósticos en tiempo real y en el corto plazo, y el proceso continuo en el largo plazo de desarrollo del modelo basado en toda la información disponible.

Ejemplo 10.3

Una compañía que fabrica piezas para tractores quiere pronosticar el número de unidades que se venderán durante los próximos 10 años. A partir de este pronóstico, se desarrollarán estrategias para seguir siendo competitivos en el negocio. Específicamente, la compañía está preocupada por la capacidad de su planta. Un pronóstico de los negocios futuros sería de gran ayuda tanto en el desarrollo de planes de expansión como en los tratos con sus fuentes de apoyo financiero.

Luego de una recopilación extensiva de datos y después de los esfuerzos por realizar el pronóstico, la compañía debe tomar la decisión de cuál de los dos pronósticos aceptar. Si bien no están muy alejados para la gran mayoría de los años, existen algunas diferencias. El primer pronóstico fue elaborado por una compañía con una sólida trayectoria especializada en la elaboración de pronósticos. Los métodos que usó esta compañía eran desconocidos para la administración de la empresa, pero a esta última se le informó que el proceso era “matemáticamente complejo”. El segundo pronóstico tuvo su origen en una reunión de ejecutivos en que participó la alta dirección junto con personal de marketing familiarizado con las expectativas de los clientes.

Después de los debates con los ejecutivos de la compañía, el presidente decidió combinar los dos pronósticos. Estos debates tendían a favorecer el pronóstico profesional por encima del pronóstico hecho en casa, de manera que el presidente decidió ponderar el primero con el 75% y el más segundo con un 25%. La tabla 10-2 presenta los dos pronósticos para cada uno de los 10 años siguientes en

TABLA 10-2 **Combinación de pronósticos de unidades vendidas (miles)**

<i>Año futuro</i>	<i>Pronóstico profesional</i>	<i>Pronóstico de casa</i>	<i>Pronóstico final</i>
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

unidades vendidas, en miles, seguidos por el pronóstico combinado. Cada pronóstico final se calculó multiplicando el primer pronóstico por .75 y el segundo por .25 y luego sumándolos. Por ejemplo, para el año 1, el pronóstico combinado está dado por

$$(.75)328 + (.25)335 = 329.75$$

Observe que los pronósticos finales se encuentran entre el pronóstico profesional y el pronóstico de casa, pero están más cerca de los pronósticos profesionales, puesto que estos últimos recibieron una mayor ponderación.

Uno de los beneficios de la combinación de pronósticos es que minimiza los efectos de sesgo, el cual resulta de la asignación indebida de peso a un método de pronóstico particular. Una estrategia consiste en combinar diferentes pronósticos promediando simplemente los pronósticos individuales generados por métodos diferentes. Si $\hat{Y}_{11}, \hat{Y}_{12}, \dots, \hat{Y}_{1m}$ son los pronósticos anticipados un paso producidos por m métodos diferentes, entonces el pronóstico anticipado un paso combinado, $\hat{Y}_{1C}$, obtenido por promedio simple es

$$\hat{Y}_{1C} = \frac{\hat{Y}_{11} + \hat{Y}_{12} + \dots + \hat{Y}_{1m}}{m} \quad (10.1)$$

Se obtienen resultados similares para otros tiempos principales de pronóstico.

Como en el ejemplo 10.2, los pronósticos se pueden combinar con ponderaciones desiguales. En este caso, cada pronóstico anticipado un paso, $\hat{Y}_{1i}$, recibe un peso $0 < w_i < 1$ donde $\sum w_i = 1$. De manera que con los m pronósticos anticipados un paso producidos por m métodos diferentes, 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

$$\sum_{i=1}^m w_i = 1$$

No hay cambio en la ponderación conforme el tiempo principal del pronóstico cambia. Por ejemplo, el pronóstico combinado dos pasos adelante, $\hat{Y}_{2C}$, se calcularía como el pronóstico combinado en la ecuación 10.2 con los pronósticos anticipados dos pasos para cada método que sustituye el pronóstico anticipado un paso. El hecho de que las ponderaciones sumen 1 garantiza que el pronóstico combinado estará en algún lado entre los valores más pequeños y los más grandes de los pronósticos individuales.

Existen varios procedimientos para determinar los pesos, w_i . Si se encuentra disponible un registro del desempeño de cada método de pronóstico, se pueden tomar las ponderaciones como inversamente proporcionales a la suma de los errores del pronóstico al cuadrado. Alternativamente, las ponderaciones se pueden obtener por métodos de regresión. Newbold y Bos (1994) ofrecen un estudio adicional de los procedimientos para la combinación de pronósticos.

En los años venideros, probablemente se hará más investigación sobre las ventajas de los pronósticos combinados, así como sobre las técnicas para obtenerlos. El objetivo de las combinaciones será desarrollar pronósticos precisos que sean efectivos en términos de costos.

Los pronósticos y las redes neurales

Los métodos de pronósticos convencionales, como los analizados en este libro, dependen de datos históricos para desarrollar un modelo y utilizar este último para proyectar las variables de interés en el futuro. Estos valores proyectados se convierten en los pronósticos que se usan luego para desarrollar planes para los negocios. En estos modelos se supone que el futuro será muy parecido al pasado, excepto por aquellas variables específicamente reconocidas por el modelo.

Los modelos convencionales hacen algunas veces suposiciones acerca de la forma de las distribuciones de la población, las cuales pueden o no estar sujetas a verificación. La estimación de intervalos empleando un modelo de regresión, por ejemplo, supone que la población subyacente sigue una distribución normal.

El campo en desarrollo de la inteligencia artificial intenta duplicar los procesos del cerebro y el sistema nervioso humanos usando la computadora. Si bien este campo se originó en la biología y la psicología, está avanzando rápidamente en otras áreas, lo que incluye los negocios y la economía. Los tres principales impulsores de la inteligencia artificial son el procesamiento del lenguaje, la robótica y las redes neurales artificiales. Este último campo es el que tiene más aplicaciones comerciales, incluyendo la de los pronósticos.

En las redes neurales, muchos ejemplos que reflejan el rango completo de relaciones pasadas entre todas las variables que podrían afectar el resultado de las variables dependientes son programados en la computadora. El programa de la red neural asimila entonces estos ejemplos e intenta desarrollar las relaciones subyacentes a través del “aprendizaje” conforme los asimila. Este proceso de aprendizaje, llamado también *entrenamiento*, es análogo al entrenamiento para el aprendizaje en el trabajo.

Unos cuantos investigadores de los pronósticos han notado la similitud entre las técnicas de las redes neurales y los métodos convencionales de pronósticos que intentan encontrar variables para predecir con éxito la variable dependiente. En teoría, la ventaja de las redes neurales como una herramienta para pronosticar es que las relaciones no necesitan especificarse con anticipación, puesto que el método implica el aprendizaje de las relaciones usando los ejemplos suministrados. Las redes neurales no requieren alguna suposición acerca de la distribución subyacente de la población y, a diferencia de muchos métodos convencionales de pronósticos, pueden operar con información incompleta.⁴

En ocasiones, la aplicación exitosa de una red neural se conoce como *plug-compatible*. Esto significa que el programa de la red neural puede sustituir rápida y fácilmente y sin ningún problema el modelo actual, como un análisis de regresión. Algunas veces es posible obtener una mejora en el desempeño, como mayor exactitud en los pronósticos, por ejemplo, con un efecto mínimo en las operaciones en curso. Las redes neurales son especialmente valiosas cuando los datos están fuertemente correlacionados, o no existen, y cuando los sistemas son notoriamente no lineales.

A continuación se mencionan algunos ejemplos de aplicaciones exitosas de las redes neurales en escenarios prácticos.⁵

- Una planta de Kodak en Texas redujo sus costos en \$3 millones por año manteniendo el mismo rendimiento y calidad del producto. Todo empezó recopilando datos históricos de operación, los cuales se usaron para entrenar a una red neural para pronosticar la calidad del producto como una función del proceso con varios parámetros.
- Uno de los problemas clásicos en la interpretación de las señales sísmicas en las labores de exploración en busca de petróleo es la obtención de la primera evidencia de la onda de choque en los registros del sensor. Los métodos tradicionales de procesamiento de señales no han tenido mucho éxito en la identificación correcta de esta primera evidencia: sólo el 30% para cada algoritmo. Varias compañías petroleras han usado un programa de redes para identificar correctamente la primera onda, con un 95% de éxito en la petrolera Amoco. Esto ha tenido un efecto importante en su capacidad para mejorar el procesamiento de datos sísmicos para modelar sus depósitos.

⁴Si bien las redes neurales no generan pronósticos de juicio en el sentido habitual, los pronósticos de circuitos neurales son distintos de los procedimientos basados en datos analizados en los capítulos 4 a 9, y por eso están incluidos en este capítulo. Una red neural es un intento por emular el proceso de pensamiento humano en una computadora. Más aún, se requiere de un sólido juicio para utilizar una red neural de manera efectiva.

⁵Estos ejemplos fueron proporcionados por NeuralWare, un proveedor comercial de software de redes neurales ubicado en 230 Main Street, Suite 200, Carnegie, PA 15106.

- Un fabricante alemán de motores eléctricos para ventiladores de automóviles utilizó maestros inspectores para determinar si un motor terminado era demasiado ruidoso o tenía sonidos “extraños”. A una red neural le tomó una semana para aprender del maestro inspector y actualmente la computadora está haciendo la prueba de ruido. El resultado financiero es un producto uniforme de mayor calidad que despierta pocos reclamos de los clientes. Los maestros inspectores se han liberado de su estresante trabajo para concentrarse en otras tareas de mejoramiento de la calidad en la fábrica.
- En una compañía de préstamos se ha utilizado desde 1989 una red neural para determinar si los clientes son idóneos para el otorgamiento de créditos al instante.
- Varias compañías han descubierto que las redes neurales se pueden usar para desarrollar indicadores útiles en el comercio. Los resultados de los sistemas señalan que los indicadores comerciales de las redes neurales, en conjunto con las estrategias comerciales adecuadas, permiten generar utilidades de manera consistente.
- La fuerza aérea está usando redes neurales para pronosticar fallas en los componentes de los aviones de su flota. Se recopiló una gran variedad de datos de cada avión, y la red fue entrenada para predecir la probabilidad de fallas específicas en aviones específicos. Estos resultados se usaron para el inventario de refacciones y el mantenimiento preventivo.
- Se desarrolló un producto comercial para pronosticar el resultado de las carreras de caballos. El sistema requiere que el usuario introduzca información específica acerca de los caballos en una carrera. Ahorra un 80% de tiempo en lo que decide cuál es la mejor apuesta

Como lo indican los ejemplos anteriores, las redes neurales han tenido muchos éxitos prácticos, pero el desempeño de las redes con los pronósticos ha sido mixto. Si bien están libres de las restricciones limitantes de especificar las relaciones por adelantado y el supuesto de las distribuciones de la población, la construcción de una red neural efectiva no es una tarea fácil. Existe una tendencia para hacer las redes más complicadas de lo que se requiere, desde luego, incorporando ruido en la señal. Como resultado, la eficacia de las redes neurales para pronosticar es reducida. Para ver un ejemplo de la aplicación de una red neural en los pronósticos, véase el caso 10-3.

Resumen de pronósticos de juicio

Un peligro potencial al usar la mayoría de los métodos de pronóstico analizados en este libro es que implican la manipulación de datos históricos para la generación de pronósticos. Esta práctica es un procedimiento válido sólo si el pronóstico es para un futuro cercano, pero se hace menos confiable conforme se aleja el horizonte del pronóstico. Los métodos estudiados en este capítulo son aportaciones valiosas al arsenal de los pronosticadores cuando la preocupación particular de la alta dirección, el pronóstico a largo plazo del entorno de la compañía, se está considerando.

Muchos autores han fascinado al público en general con sus especulaciones acerca de las tendencias a largo plazo que actualmente se desarrollan en nuestra sociedad y en el futuro que ellos pretenden, además de las habilidades administrativas que serán necesarias para lidiar con ellas. Entre las más provocadoras, sin un orden particular, se encuentran las siguientes:

- 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 21st Century Manager: Future-Focused Skills for the Next Millennium* (Kogan Page, 1999)
- Chuck Martin, *Net Future: The 7 Cyberrends That Will Drive Your Business, Create New Wealth, and Define Your Future* (McGraw-Hill, 1998)

- Alvin Toffler, *Future Shock* (Random House, 1970)
- Alvin Toffler, *The Third Wave* (Bantam Books, 1980)
- John Naisbitt, *Megatrends* (Warner Books, 1982)
- John Naisbitt y Patricia Aburdene, *Megatrends 2000* (William Morrow, 1990)
- John Naisbitt, *Eight Asian Megatrends That Are Reshaping the World* (audiotape) (Simon & Schuster, 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, aun cuando algunas de ellas son de hace 25 o 30 años, porque estos autores tienen una habilidad única para especular acerca del futuro de un modo provocativo. Un enfoque más serio acerca de la clase de cambios de los que ellos hablan se llama *pronóstico tecnológico* y se está convirtiendo de modo creciente en un tema importante para varias empresas. Esto no es sorprendente, puesto que se estima que el 25% de la tecnología es reemplazada cada año. Obviamente, hasta la manipulación más compleja de datos históricos para generar pronósticos puede fallar por un gran margen en estas circunstancias. Considere, por ejemplo, los siguientes campos en desarrollo. Conforme estas tecnologías se desarrollen, el impacto en muchas empresas será considerable.

- Inteligencia artificial
- Ingeniería genética
- Bioelectricidad
- Robótica multisensorial
- Láser
- Fibra óptica
- Microondas
- Satélites avanzados
- Energía solar
- Superconductores

La Pricewaterhouse Coopers Company ha puesto a la disposición del público en general una publicación interna, *Technology Forecast*. Este documento anual analiza las tecnologías de información clave y pronostica su entrada en escena para un periodo que comprende de un año a tres. Éste es un ejemplo de una gran preocupación de la compañía por los avances tecnológicos y sus efectos en la sociedad y en la compañía.

La tasa de desarrollo tecnológico significa que los gerentes deben estar cada vez más alerta al cambio. Los patrones históricos pueden ofrecer por lo menos algunas pistas del futuro en casi todas las empresas. Como se enfatizó en el capítulo 1, para generar pronósticos útiles, es necesario hacer una mezcla prudente de la manipulación histórica de datos con el uso sensato del criterio y la experiencia.

OTRAS HERRAMIENTAS ÚTILES PARA HACER JUICIOS ACERCA DEL FUTURO

Se han desarrollado varias herramientas para ayudar a los encargados de tomar decisiones en el proceso de ponderar información acerca del futuro y optar por las mejores decisiones posibles. Con frecuencia se analiza un grupo de herramientas en los libros de texto y artículos en revistas bajo el título general de “toma de decisiones” o “toma de decisiones bajo incertidumbre”. Cuando los directores se enfrentan a la tarea de tomar decisiones a la luz de la incertidumbre, su habilidad para pronosticar se convierte en un elemento crucial en el proceso de toma de decisiones

Cuando el análisis de los datos históricos concluye, el encargado de tomar decisiones debe hacer un juicio considerando las alteraciones en el curso de acción de la empresa. En otras palabras, el analista debe entrelazar los resultados del proceso del pronóstico con los procedimientos de toma de decisiones existentes en la compañía. A continuación se analizan brevemente algunos elementos de la teoría de decisiones que son relevantes en esta etapa del proceso.⁶

El concepto de *valor esperado* se describió en el capítulo 2. Los encargados de tomar decisiones lo usan con frecuencia de manera explícita o implícita. Recuerde que este concepto implica el cálculo del valor de la media que una variable aleatoria numérica asumirá durante muchas pruebas. En la tabla 10-3, una variable aleatoria discreta, X , es desplegada en una distribución de probabilidad; cada valor futuro posible que X puede asumir se presenta junto con la probabilidad de cada uno.

Observe que la suma de las probabilidades de la tabla 10-3 es 1.00, o 100%, lo cual significa que se ha identificado cada valor posible que X puede tomar. En la tabla 10-3 hacemos que X represente el número de los principales contratos nuevos que una empresa firmará durante el próximo año fiscal. La pregunta que contesta el valor esperado es: ¿Cuántos contratos nuevos se pueden esperar, en promedio, si la distribución de probabilidad de la tabla 10-3 es válida? La ecuación 10.3 se emplea para calcular el valor esperado de una distribución de probabilidad

$$E(X) = \sum X[P(X)] \quad (10.3)$$

donde

$E(X)$ = el valor esperado

X = los valores que puede adoptar la variable aleatoria

$P(X)$ = la probabilidad de que ocurra cada X

El valor esperado de la tabla 10-3 se calcula usando 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 presentado en la tabla 10-3 es 3.2. Si en este ejemplo X representa el número de los principales contratos nuevos para el próximo año fiscal, y si las probabilidades de los contratos nuevos son las mismas cada año, en promedio, se firmarían 3.2 nuevos contratos. Observe que el valor de 3.2 no es posible en cualquier año, sólo es posible para los valores enteros: 1, 2, 3, 4, 5. Sin embargo, el valor 3.2 representa el resultado promedio durante muchas pruebas. Los encargados de tomar decisiones a menudo están interesados en los valores esperados y los usan como sus mejores pronósticos para variables numéricas críticas en la planeación del futuro incierto.

**TABLA 10-3 Distribución
de probabilidad**

X	$P(X)$
1	.10
2	.20
3	.25
4	.30
5	.15
	1.00

⁶Hammon, Keeney y Raiffa (1999) escribieron un excelente libro (y muy claro) sobre la toma de decisiones.

La teoría de decisiones dirige los elementos que integran la función de toma de decisiones de los líderes de negocios. A menudo los valores esperados se entrelazan en esta consideración más general de toma de decisiones. El *diagrama de árbol de decisión* se utiliza para ayudar a que el encargado de tomar decisiones visualice una situación compleja y tome decisiones racionales. En la figura 10-1 se presenta un diagrama de árbol de decisión

La figura 10-1 refleja la incertidumbre que existe acerca de la naturaleza de las ventas futuras y conlleva la decisión de si se construye una nueva planta o se repara la planta actual. El problema es que, si la compañía supiera que habrá mucha demanda, sería mejor en términos económicos construir la nueva planta; por otro lado, si supiera que la demanda será baja, las utilidades serían mayores si se repara la antigua planta. Aun en este ejemplo tan simple, es posible ver el beneficio del diagrama de árbol de decisión: ayuda al encargado de tomar decisiones a visualizar las diferentes opciones disponibles para identificar las incertidumbres más allá del control de la empresa y determinar explícitamente costos, utilidades y las probabilidades de eventos futuros. En situaciones más complicadas, los beneficios del diagrama del árbol y la teoría formal de decisiones son más evidentes.

El *teorema de Bayes* es un concepto estadístico diseñado para revisar de manera preliminar las probabilidades con base en la evidencia de la muestra. Este concepto a menudo se aplica en situaciones en que primero se determinan las estimaciones de las probabilidades de un evento futuro y éstas se modifican después de recolectar evidencia de la muestra. Un ejemplo es el concepto de prueba de mercado que utilizan muchos fabricantes de productos para el consumidor. Una compañía podría estimar la probabilidad de aceptación del público de un nuevo producto como muy alta. Sin embargo, antes de arriesgar los millones de dólares que una campaña nacional requiere, podría realizarse una prueba de mercado en regiones que la compañía considera como buenas representantes de los mercados. Los resultados de la prueba de mercado se

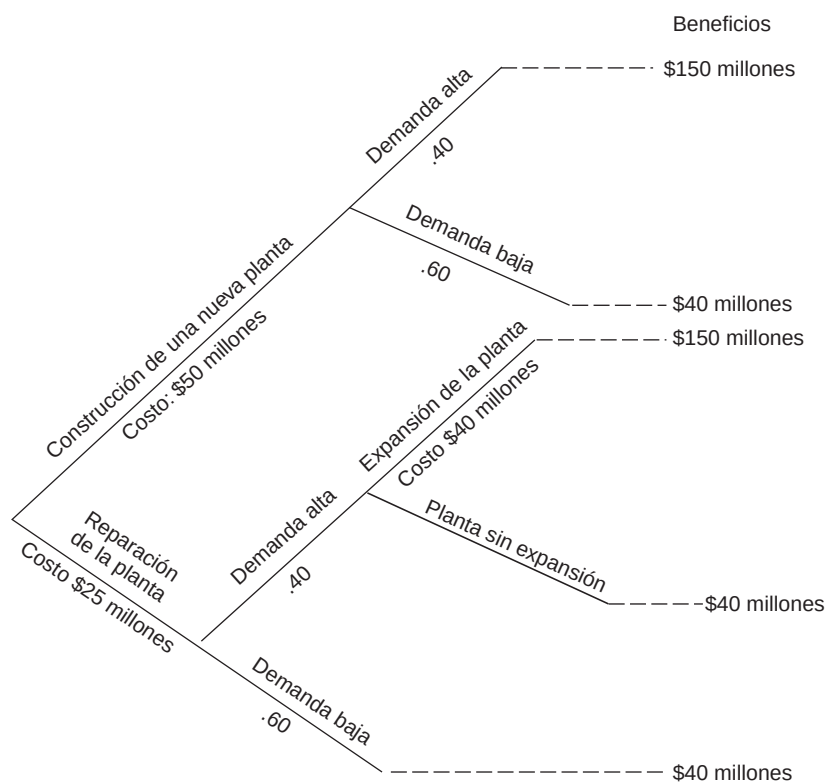


FIGURA 10-1 Diagrama de árbol de decisión

usan entonces para modificar las estimaciones originales del éxito del producto, y luego se toma la decisión acerca del lanzamiento nacional del nuevo producto. Una versión simplificada del teorema de Bayes se presenta en la ecuación 10.4.⁷

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (10.4)$$

donde

$P(A|B)$ = La probabilidad de que ocurra el evento A , habiendo ocurrido el evento B

$P(B)$ = La probabilidad de que ocurra el evento B

$P(A)$ = La probabilidad de que ocurra el evento A

$P(B|A)$ = La probabilidad de que ocurra el evento B , habiendo ocurrido el evento A

Ejemplo 10.4

La figura 10-2 representa una aplicación específica del teorema de Bayes en un escenario de prueba de mercado. Los directivos de una gran compañía de productos al consumidor necesitan decidir si lanzan un nuevo producto a nivel nacional. Ellos estiman que su nuevo producto tiene un 50% de probabilidad de tener grandes ventas en el mercado nacional; es decir, $P(H) = .50$. También están considerando el uso de una prueba de mercado para determinar si pueden hacer un mejor pronóstico de ventas altas o bajas del nuevo producto. La figura 10-2 muestra un diagrama de árbol de decisión de los resultados de la prueba de mercado.

Experiencias anteriores indican que, cuando se lanza al mercado un nuevo producto y se hacen grandes ventas, la prueba de mercado fue exitosa en un 80% de los casos, es decir, $P(S|H) = .80$. Las

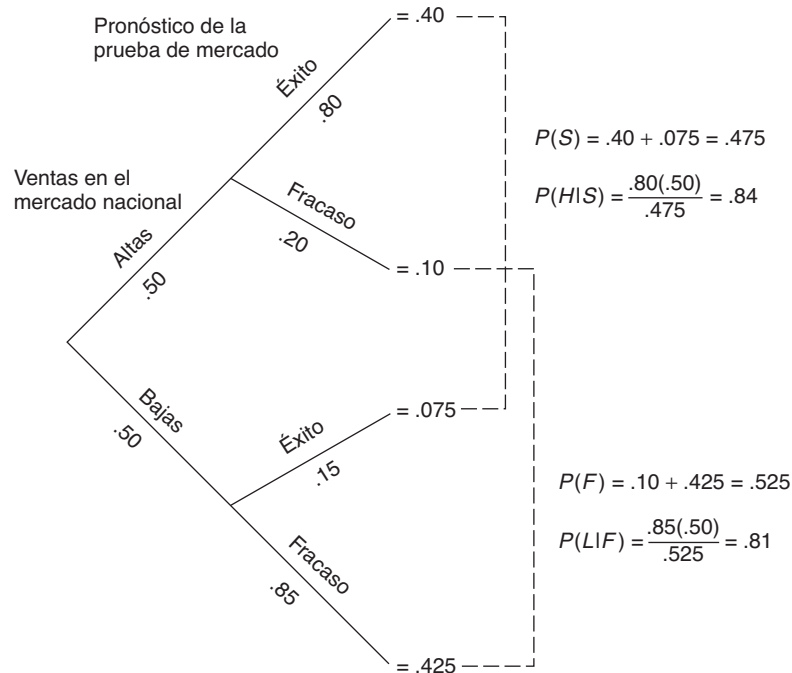


FIGURA 10-2 Decisión de Bayes para la introducción de un nuevo producto

⁷El enunciado general del teorema de Bayes para n eventos, $A_1, A_2, \dots, A_n$, cada uno de los cuales puede ocurrir conjuntamente con otro evento, B , is

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)} \quad i = 1, 2, \dots, n$$

experiencias anteriores también indican que, cuando se introducía un nuevo producto y se registraban pocas ventas, la prueba de mercado fallaba (ventas bajas) el 85% del tiempo, o $P(F|L) = .85$.

El teorema de Bayes sirve para calcular la probabilidad de que el nuevo producto tenga éxito en el mercado nacional en condiciones diferentes de prueba. La figura 10-2 presenta los cálculos de las dos probabilidades de interés particular:

- Si la prueba de mercado pronostica el éxito, existe un 84% de probabilidad de que el producto tenga ventas altas a nivel nacional. O lo que es lo mismo,

$$P(H|S) = .84 = .80(.50)/.475 = P(S|H)P(H)/P(S)$$

- Si la prueba de mercado pronostica el fracaso, existe un 81% de probabilidad de que el producto tenga ventas bajas a nivel 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 dirección a decidir si introducir o no el nuevo producto a nivel nacional. La exactitud de la prueba de mercado es suficiente para cambiar la probabilidad de ventas altas (o ventas bajas) del valor preestablecido de la prueba del 50%. Si la prueba de mercado es exitosa, la probabilidad de ventas altas crece al 84%. Si la prueba de mercado falla, la probabilidad de ventas bajas aumenta al 81%. Las probabilidades fueron calculadas usando el teorema de Bayes. La decisión acerca de la introducción del producto será mucho más fácil de tomar que si se hiciera sin la prueba de mercado.

Otra herramienta para la elaboración de pronósticos es la simulación por computadora. La *simulación* es un conjunto de técnicas numéricas y de programación para modelar escenarios sujetos a la incertidumbre y para realizar experimentos muestrales en esos modelos usando una computadora. Cada corrida de simulación (replicación) produce un resultado posible (“pronóstico”) del problema que se somete a estudio. Muchas corridas (replicaciones) dan la oportunidad al encargado de tomar decisiones de observar el conjunto completo de resultados posibles, así como la probabilidad de que ocurra. Estos escenarios generados por computadora se pueden resumir y utilizar para tomar la mejor decisión.

Existen muchas razones para usar la simulación, en vez del análisis matemático, para reunir información en un ambiente de incertidumbre.

- Muchas representaciones de los sistemas de negocios reales son demasiado complejas para analizarlas matemáticamente.
- El interés principal podría ser la experimentación con el sistema o encontrar los mejores niveles de las variables que influyen en el sistema o, simplemente, estudiar el sistema. Experimentar con el sistema real puede ser imposible (situaciones de combate) o extremadamente caro (diseño de instalaciones nucleares) o consumirían tanto tiempo que sólo es posible una replicación.
- La simulación es útil por sí misma ya que conlleva una mejor comprensión del sistema.
- Como herramienta, la simulación tiene cierto grado de credibilidad en la dirección. Es relativamente fácil explicar a la dirección los resultados obtenidos en un estudio de simulación.

Desde una perspectiva de negocios, la mayor ventaja de un ejercicio de simulación en computadora es que los pronósticos que son resultado de varias decisiones de la administración pueden obtenerse antes de tomar dichas decisiones. La simulación se utiliza en una gran variedad de escenarios, incluyendo los modelos de operaciones de manufactura, programación de la fuerza de trabajo, modelos de alternativas de inversión, análisis y precio de instrumentos de inversión complejos, y modelos del flujo de pasajeros en un aeropuerto.⁸ Los diagramas de árbol de decisión como el de la figura 10-1 con frecuencia forman parte de los experimentos de simulación por computadora.

⁸Law y Kelton (2000); Ross (2006); Seila, Ceric y Tadikamalla (2003) ofrecen buenos análisis de la simulación por computadora

Fórmulas clave

Pronóstico combinado obtenido por promedio simple

$$\hat{Y}_{1C} = \frac{\hat{Y}_{11} + \hat{Y}_{12} + \cdots + \hat{Y}_{1m}}{m} \quad (10.1)$$

Pronóstico combinado obtenido por 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

$$E(X) = \sum X[P(X)] \quad (10.3)$$

Teorema de Bayes

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (10.4)$$

Problemas

1. Identifique dos escenarios de negocio donde el método Delphi podría usarse para generar pronósticos. ¿Puede pensar en las dificultades y los riesgos asociados con el uso del método Delphi?
2. Considere las ventas reales presentadas en la tabla P-2 junto con los pronósticos anticipados un paso generados por el método de Winters y por un modelo de regresión.
 - a) Elabore los pronósticos combinados de las ventas realizadas tomando un promedio simple de los pronósticos generados por el método de Winters y el modelo de regresión.
 - b) Elabore los pronósticos combinados de las ventas realizadas tomando un promedio ponderado de los pronósticos de Winters y de los pronósticos de regresión con pesos $w_1 = .8$ y $w_2 = 1 - w_1 = .2$.
 - c) Con base en las ventas reales, determine los MAPE para los pronósticos de Winters y los pronósticos de regresión.
 - d) Repita el inciso c) usando los pronósticos combinados de los incisos a) y b). Con base en la medida del MAPE, ¿cuál grupo de pronósticos prefiere usted?

TABLA P-2

<i>Mes</i>	<i>Ventas reales</i>	<i>Pronósticos de Winters</i>	<i>Pronósticos de regresión</i>
1	5,502	4,586	5,263
2	6,535	5,944	6,008
3	7,390	6,706	6,832
4	4,388	4,530	4,886
5	4,521	5,010	4,918
6	5,679	6,574	5,630
7	8,578	8,462	7,963
8	7,088	6,415	5,942
9	4,839	4,457	5,156
10	4,050	4,638	3,819

CASOS

CASO 10-1 RESTAURANTE GOLDEN GARDENS

Sue y Bill Golden decidieron abrir un restaurante en una ciudad del medio oeste de Estados Unidos. Ellos han pasado un año investigando el área y visitando restaurantes de precios medios a altos. Definitivamente ellos creen que hay espacio para otro restaurante y han encontrado un buen sitio que está disponible a un buen precio.

Además, tienen contacto con varios chefs de primera clase y creen que pueden atraer a uno de ellos a su nuevo restaurante. Sus investigaciones en los bancos locales los han convencido de que pueden conseguir financiamiento fácilmente, dados sus propios recursos y su experiencia en el negocio de restaurantes.

El único asunto problemático para los Golden es la atmósfera que darían a su restaurante. Ya han iden-

tificado tres grupos de personas residentes en el área que normalmente comen fuera de casa, y no ha habido consenso al respecto. Han hablado mucho del asunto entre ellos mismos pero ahora creen que otras opiniones podrían ser valiosas.

Después de leer algo acerca de las técnicas que se usan en los pronósticos de juicio, creen que algunas de éstas podrían ayudarlos a decidir sobre la atmósfera de su nuevo restaurante. Ellos han identificado un buen número de amigos y asociados en otras ciudades que estarían dispuestos a ayudarlos, pero no están seguros cómo utilizar sus talentos.

PREGUNTAS

1. ¿Qué método sugeriría usted a los Golden para aprovechar la experiencia de sus amigos al decidir acerca de la atmósfera de su nuevo restaurante?
2. ¿Existen otros métodos que hayan pasado por alto al tratar de investigar este asunto?

CASO 10-2 ALOMEGA FOOD STORES

En el ejemplo 1.1 se describió cómo Julie Ruth, la presidente de Alomega Food Stores, recopiló las ventas mensuales de su compañía junto con otras muchas variables que, en su opinión, podrían estar relacionadas con las ventas. Los casos de Alomega (casos 2-3, 3-4, 5-6 y 8-7) describieron sus intentos para usar varios procedimientos disponibles en Minitab en un intento por generar pronósticos de las ventas mensuales.

En el caso 8-7, Julie desarrolló un modelo de regresión múltiple que explicaba casi el 91% de la varianza de la variable de ventas. Ella se sintió cómoda con este modelo, pero estaba especialmente sensible por los comentarios negativos expresados por Jackson Tilson, su gerente de producción, durante una junta reciente

(véase el ejemplo 1.1). Tilson dijo: “He estado tratando de mantener la boca cerrada durante esta reunión, pero esto es realmente demasiado. Creo que estamos gastando el tiempo de mucha gente con toda esta recopilación de datos y estamos jugando con las computadoras. Todo lo que tienen que hacer es hablar con el personal de piso y con los gerentes de las tiendas para entender qué está pasando. He visto pasar esto aquí antes, y aquí vamos otra vez. Todo lo que su gente necesita es apagar sus computadoras, salir de sus lujosas oficinas y hablar un poco con gente real”.

Julie pensó que las políticas de la oficina dictaban que ella debía poner atención a la opinión de Jackson. Consultó con varias personas, incluyendo a Tilson, para

**TABLA 10-4 Ventas mensuales de 27 tiendas
Alomega Food Stores, 2003 a 2006**

<i>Mes</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
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
Mayo	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
Sept.	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

preguntarles sus opiniones acerca de cómo pronosticar las ventas de enero de 2007. La gran mayoría dijo que la mejor predicción la daría la cifra de las ventas de enero del año anterior. Asimismo, el pronóstico para febrero del 2007 estaría basado en la cifra de las ventas de febrero de 2006.

Con base en esta información, Julie desarrolló un modelo de pronóstico informal:

$$\hat{Y}_{t+1} = Y_{t-11}$$

que usaba el valor mensual del último año para predecir este valor mensual del año actual.

PREGUNTAS

1. ¿Cuánta precisión tiene el pronóstico informal de Julie?

2. ¿Cómo se compara el modelo informal con el modelo de regresión múltiple desarrollado en el caso 8-7?

3. Hasta que Julie pueda experimentar estos dos métodos en acción, ella está considerando la combinación de pronósticos. Cree que este enfoque
- cumpliría con las políticas de la empresa y que le permitiría usar un enfoque más eficiente. ¿Sería esto una buena idea?

4. ¿Debe Julie usar un método de promedio simple o un promedio ponderado para combinar los pronósticos?

**CASO 10-3 LYDIA E. PINKHAM MEDICINE
COMPANY**

Este caso presenta una aplicación real del uso de redes neurales para pronosticar series de tiempo. Los autores entienden que los estudiantes no tienen los antecedentes para comprender plenamente este caso. Sin embargo, se obtendrán algunos beneficios a través de la experiencia de este caso real.

La compañía Lydia E. Pinkham Medicine Company y Lydia Pinkham's Vegetable Compound fueron presentadas en el caso 9-4.

Ha habido muchos intentos por usar redes neurales para pronosticar series de tiempo. La mayoría de los intentos han tenido lugar en el campo de la utilización

de la energía, puesto que las compañías de energía necesitan pronósticos precisos del horario de demanda de su producto. Sin embargo, algunas investigaciones se han enfocado en series de tiempo de negocios más tradicionales, como las series micro y macroeconómicas, datos demográficos y datos de compañías específicas. Prácticamente todo este trabajo ha utilizado una red de conexiones hacia adelante entrenada mediante la propagación hacia atrás. Este caso de estudio empleará este tipo de red para pronosticar las ventas de Lydia E. Pinkham Company. Los pronósticos resultantes se compararán con los pronósticos del modelo AR(2) presentado en el caso 9-4.

La figura 10-3 describe 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 el número de entradas a la red. En este caso, las dos entradas son Y_{t-1} y Y_{t-2} . (Usar los dos periodos previos para predecir el

periodo actual es congruente con el modelo AR(2); por lo tanto, 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 llama *oculta* porque no está directamente conectada con el “mundo exterior”, como lo están las capas de entrada y de salida). El número de nodos en la capa oculta se selecciona arbitrariamente en un sentido: muy pocos nodos ocultos restringen la capacidad de la red para “ajustar” los datos, y demasiados nodos ocultos causan que la red memorice los datos de entrenamiento (o de estimación). La memorización trae consigo un rendimiento muy deficiente sobre la muestra de prueba. En este caso, el número de nodos en la capa oculta es simplemente dos veces el número de entradas. Por último, el 1 indica un nodo de salida que brinda el pronóstico de un paso anticipado, o Y_t .

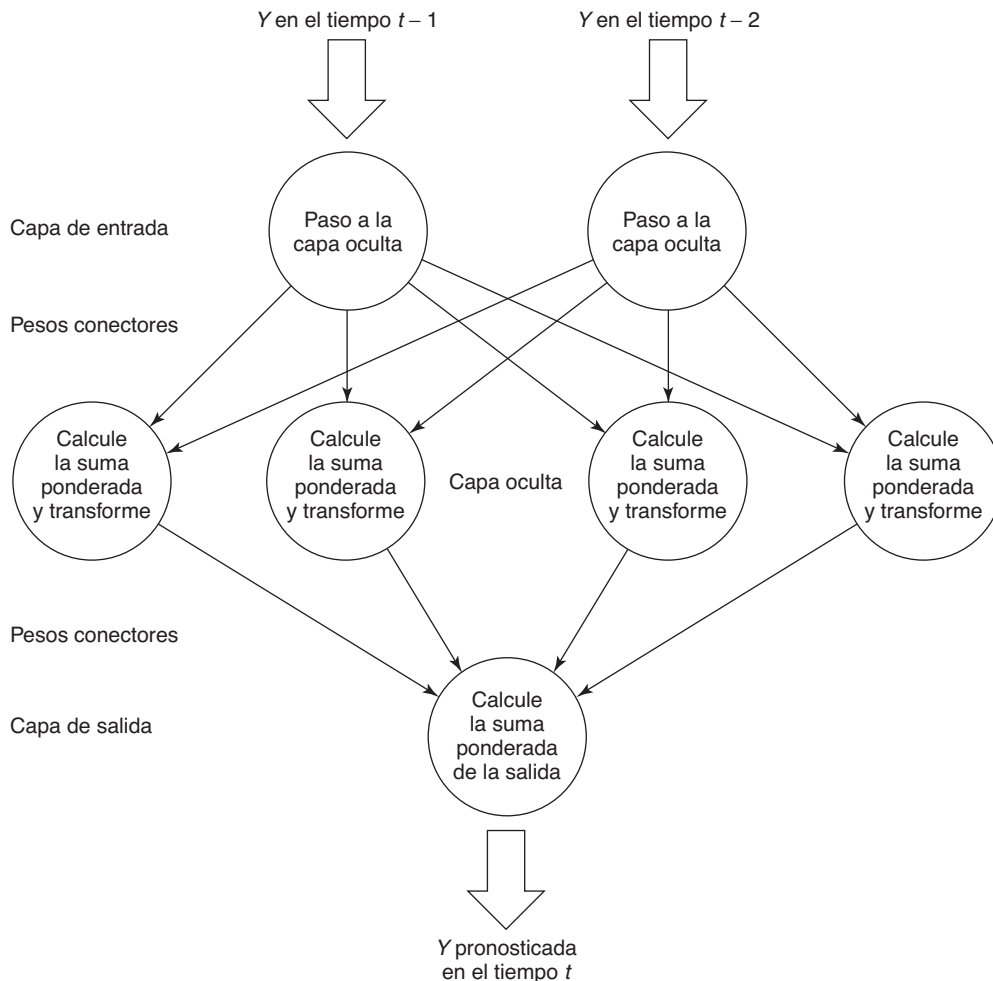


FIGURA 10-3 Red con conexiones hacia adelante 2-4-1

TABLA 10-5 Resultados de la red neural para Lydia E. Pinkham Company

<i>Iteraciones</i>	<i>Datos de estimación</i>			<i>Datos de entrenamiento</i>		
	<i>MAD</i>	<i>MAPE</i>	<i>MSE</i>	<i>MAD</i>	<i>MAPE</i>	<i>MSE</i>
10,000	189.7	11.06%	51,923.0	182.2	11.26%	43,441.1
20,000	189.8	10.80%	53,926.3	169.6	10.18%	37,737.4
50,000	195.0	11.22%	54,311.3	176.3	10.51%	41,529.7
75,000	196.0	11.40%	54,424.2	178.1	10.66%	42,758.1
100,000	196.5	11.53%	54,725.5	181.9	11.02%	43,985.4

La red neural calcula su salida de la siguiente manera. Cada uno de los arcos conectores entre los nodos de dos capas adyacentes tiene un peso asociado. Cada nodo en la capa oculta calcula la suma ponderada de sus entradas. (Los nodos de la capa de entrada simplemente pasan las entradas a la capa oculta). Luego, esta suma ponderada es “transformada” de algún modo, tal como $Y = 1/(1 + e^{-x})$ donde Y es el dato “transformado” y X es la suma ponderada. La Y se pasa entonces a la capa de salida, donde cada nodo calcula otra vez una suma ponderada. Esta suma ponderada final es la salida de la red. La red está entrenada para ajustar todos pesos conectores de una manera iterativa.

EL MODELO DE LA RED NEURAL

La red neural fue entrenada usando BrainMaker.⁹ Para este estudio el tamaño del paso se estableció en 0.500 y la tolerancia al entrenamiento en .001. (Otros paquetes pueden pedirle al usuario la especificación de una tasa de aprendizaje y un término de impulso). Como en el caso 9-4, las primeras 42 observaciones se usan para entrenar la red; las últimas 12 se usan para evaluar el rendimiento de la predicción de la red un paso anticipado. Un problema en el uso de redes neurales para pronosticar series de tiempo radica en la determina-

ción de cuánto tiempo entrenar la red; una red sobreentrenada tiende a memorizar los datos de entrenamiento y realizar una prueba deficiente. Así, algunos investigadores han sugerido simplemente detener el entrenamiento “pronto”, antes de que la red haya memorizado los datos. Para determinar el efecto de entrenamiento, la red fue entrenada para 10,000, 20,000, 50,000, 75,000, y 100,00 iteraciones. (Una iteración es la presentación de una observación; las iteraciones listadas representan 250, 500, 1,250, 1,875 y 2,500 pases a través del conjunto completo de entrenamiento, respectivamente). Esto permite al analista evaluar el posible efecto del sobreentrenamiento. (El trabajo se hizo originalmente en una PC con un procesador Intel 386SX-20, y el “tiempo” de entrenamiento para 100,000 iteraciones fue aproximadamente de 20 minutos).

RESULTADOS

El *MAD*, *MAPE* y el *MSE* para los diferentes modelos de redes neurales se presentan en la tabla 10-5, y no se comparan favorablemente con el modelo AR(2). El autor de este caso está experimentando actualmente con un tipo diferente de red neural, el cual produce resultados comparables a los del modelo AR(2).¹⁰

PREGUNTAS

1. Encuentre un artículo que describa una aplicación de redes neurales de series de tiempo. ¿Qué método usaron los autores y qué tan exitoso fue éste?
2. Si usted tiene acceso a un programa de una red neural, trate de obtener la red que elabore mejor los pronósticos que la red 2-4-1 presentada aquí.
3. ¿Por qué las redes neurales se consideran como una alternativa viable para los métodos de pronosticar estudiados en este texto?

⁹BrainMaker, un paquete de simulación para PC con base en redes neurales, disponible comercialmente, fue desarrollado por California Scientific Software en 1990.

¹⁰Para información adicional sobre redes neurales, véanse Haykin (1998) y Khanna (1990).

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.
- 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) (May 1991): 501–516.
- 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.
- Haykin, S. *Neural Networks: A Comprehensive Foundation*, 2a. ed. Upper Saddle River, N.J.: Prentice Hall, 1998.
- Jain, C. L. *A Managerial Guide to Judgmental Forecasting*. Nueva York: Graceway, 1987.
- Jain, C. L. "Myths and Realities of Forecasting." *Journal of Business Forecasting* 9 (otoño de 1990): 18–22.
- Khanna, T. *Foundations of Neural Networks*. Reading, Mass.: 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*, 3a. ed. Boston: McGraw-Hill, 2000.
- Lawrence, M., P. Goodwin, M. O'Conner y D. Onkal. "Judgmental Forecasting: A Review of Progress over the Last 25 Years." *International Journal of Forecasting* 22 (3) (2006): 493–518.
- Levenbach, H. y J. P. Cleary. *Forecasting Practice and Process for Demand Management*. Belmont, Calif.: Thomson Brooks/Cole, 2006.
- 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., and T. Bos. *Introductory Business and Economic Forecasting*, 2a. ed. Cincinnati, Ohio: South-Western, 1994.
- Parente, F. J. y J. K. Anderson-Parente. "Delphi Inquiry Systems." In *Judgmental Forecasting*, G. Wright y P. Ayton, (eds.) Nueva York: Wiley, 1987, 129–156.
- Price Waterhouse World Technology Centre. *Technology Forecast*. Menlo Park, Calif.: Price Waterhouse World Technology Centre, 2006.
- Ross, S. *Simulation*, 4th ed. New York: Academic Press, 2006.
- 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, Calif.: Thomson 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.

ADMINISTRACIÓN DEL PROCESO DE PRONÓSTICOS

El capítulo 1 presentó el tema de generación de pronóstico con sugerencias acerca de su utilidad dentro de la organización moderna. Actualmente las organizaciones deben elaborar pronósticos para tomar decisiones oportunas frente a la incertidumbre. Este proceso de hacer conjeturas con cierto fundamento acerca del futuro (pronóstico) por lo general implica un proceso racional de extender los datos y las experiencias históricas hacia el futuro. Ahora nos es posible repasar las nociones básicas y los puntos importantes presentados en el capítulo 1 aprovechando el conocimiento de las numerosas técnicas para pronosticar analizadas en los capítulos anteriores.

EL PROCESO DEL PRONÓSTICO

Este proceso se compone de dos fases. Una fase tiene lugar a nivel estratégico, y la otra tiene lugar a nivel operativo. En el nivel estratégico, las decisiones incluyen qué pronosticar, cómo usar los pronósticos y quién será responsable de la generación de los pronósticos. La fase operativa del pronóstico consiste en la recopilación de datos, la generación del pronóstico y la evaluación de los pronósticos en curso. El proceso del pronóstico es como cualquier otro proceso: si no se controla y verifica, es como una espiral fuera de control.

Un punto clave discutido en el capítulo 1 es la importancia de utilizar un buen juicio administrativo junto con técnicas cuantitativas para desarrollar buenos pronósticos. Por supuesto que el buen juicio es importante y ahora lo podemos ilustrar usando varias de las técnicas de pronosticar estudiadas en capítulos anteriores. Siempre es necesaria una mezcla juiciosa de técnicas cuantitativas con el sentido común si queremos que los pronósticos sean precisos y comprensibles, y sean de utilidad para los responsables de tomar decisiones en la empresa.

La descomposición de una serie de tiempo es un buen ejemplo de una técnica que implica la necesidad de un juicio confiable junto con el análisis de la historia pasada. Una compañía que utiliza un programa de análisis de una serie de tiempo con, digamos, datos de las ventas mensuales que produce un análisis histórico de la variación tendencia-ciclo, variación estacional y movimientos irregulares podría recombinar los dos primeros factores para generar un pronóstico de las unidades vendidas. La selección adecuada de la función de tendencia y la naturaleza de la recombinación requieren mucho criterio acerca del comportamiento futuro. Al aplicar el proceso de descomposición, también es importante decidir cuántos periodos históricos se deben incluir en el análisis.

El análisis de regresión requiere criterio junto con análisis estadístico siempre que se elaboran los pronósticos. Si una regresión múltiple fue elaborada usando la clasificación del desempeño en el trabajo como la variable dependiente y dos variables —la calificación en la prueba y edad— como las variables explicativas, se podría obtener un valor de R^2 del 60%. Además, los valores t de las variables explicativas podrían ser significativos junto con el valor

de regresión F . Entonces el pronosticador tal vez se vea tentado a medir las dos variables explicativas de cada candidato al puesto y usarlas para predecir el desempeño en el trabajo. Sin embargo, es necesario considerar tres preguntas adicionales. Primero, ¿es suficiente la explicación del 60% de la variación para efectos del pronóstico? Quizás un “criterio intuitivo” para la contratación de una persona es un mejor método, o quizás se necesita mayor precisión en el pronóstico y deben considerarse otras variables explicativas. Segundo, ¿se puede suponer que los candidatos al trabajo son iguales, en esencia, a los de la muestra en el estudio de regresión? Si de alguna manera difieren en forma sustancial, el modelo del pronóstico tal vez no sea válido. Finalmente, ¿está justificado el costo del proceso del pronóstico en términos de los beneficios recibidos? La prueba puede ser costosa para la compañía, especialmente si se le paga a una agencia externa de pruebas, y se debe justificar por los beneficios del pronóstico.

La regresión de los datos de las series de tiempo es común en las organizaciones que dan seguimiento a medidas importantes de desempeño de manera semanal, mensual o trimestral. Como la autocorrelación es un acontecimiento común en tales estudios, es de vital importancia la comprensión y el tratamiento de esta condición para que los resultados de esos análisis sean válidos en el proceso de toma de decisiones. Por desgracia, esta comprensión a menudo no existe; esta deficiencia se ha convertido en un problema creciente con el advenimiento del software del análisis de regresión de bajo costo.

Las técnicas Box-Jenkins estudiadas en el capítulo 9 ilustran un problema común en los pronósticos. Estos procedimientos a menudo son métodos de pronóstico de gran calidad, que generan errores menores de pronóstico en muchas situaciones complejas. Su desventaja es que se requiere mucha preparación por parte del usuario. Si el responsable de tomar decisiones no comprende el proceso utilizado para la generación de pronósticos, éstos pueden ser ignorados a pesar de su precisión.

Los aspectos de corto, mediano y largo plazos de las técnicas de pronóstico en los diferentes niveles de la administración en una compañía se pueden ilustrar con el análisis de las series de tiempo y los pronósticos tecnológicos. La primera y segunda líneas de administración en la organización podrían estar interesadas en el análisis de una serie de tiempo de las ventas mensuales, usando la suavización y modelos ARIMA, con los datos recopilados durante los cuatro últimos años. Los pronósticos generados con estos procedimientos, quizás modificados por el juicio, podrían usarse para planear la producción mensual de unidades para el siguiente año fiscal. Los gerentes de nivel medio podrían usar los mismos procedimientos de las series de tiempo, complementados con un análisis de tendencia por separado, para analizar las ventas anuales de unidades durante los pasados ocho años y pronosticar las ventas anuales para los próximos cinco años. En este caso, el objetivo podría ser planear los gastos de capital necesarios para el negocio durante el periodo de los siguientes cinco años. Al mismo tiempo, la alta dirección podría estar comprometida en el pronóstico tecnológico usando el método Delphi, junto con una formulación o descripción de escenarios. Su propósito sería evaluar la posición actual de la compañía en el mercado y detectar los cambios tecnológicos o sociales que amenacen su nicho de mercado durante los próximos 20 años, o las oportunidades que no son evidentes en sus operaciones diarias.

Las técnicas de análisis de datos presentadas en este libro se resumen en la tabla 11-1. Esta tabla ofrece descripciones, señala aplicaciones y niveles de costo estimados e indica si son necesarios recursos informáticos para la aplicación de cada técnica. Cada una de estas últimas hace referencia a los capítulos en los cuales se analiza. Los resúmenes como el de la tabla 11-1 deben considerarse como guías y no como enunciados definitivos que no se pueden cuestionar.

MONITOREO DE PRONÓSTICOS

Después de que se han establecido los objetivos del pronóstico, la recolección de datos y la selección de una técnica de pronóstico adecuada están entre los pasos siguientes en la aplicación de un esfuerzo efectivo del pronóstico en curso. En este libro se han descrito varios

TABLA 11-1 Modelos de pronósticos

Método	Descripción	Aplicaciones	Computadora		
			Costo	¿Necesaria?	Capítulo
Modelos causales de pronósticos					
Análisis de regresión	Pronóstico explicativo; supone una relación de causa-efecto entre la información de entrada y la información resultante	Pronóstico de corto y mediano alcance de productos y servicios existentes; estrategias de marketing, producción, contratación de personal y planeación de instalaciones	De bajo a medio	Por lo general	6
Regresión múltiple	Pronóstico explicativo: supone una relación de causa-efecto entre la información introducida en el sistema y la información que resulta de éste	Igual que en el caso anterior	De bajo a medio	Sí	7
Modelos de pronóstico de series de tiempo					
Método de descomposición	Pronóstico explicativo: supone una relación de causa-efecto entre el tiempo y la información resultante del sistema; el sistema se descompone en sus elementos	Pronóstico de medio alcance para una planta nueva y planeación del equipo, financiamiento, desarrollo del nuevo producto y nuevos métodos de ensamblaje; pronóstico de corto alcance para personal, publicidad, inventario, financiamiento y planeación de la producción	De bajo a medio	Sí	5
Promedios móviles	Se usan para eliminar la aleatoriedad en una serie de tiempo; el pronóstico se basa en los datos de la serie de tiempo suavizados por un promedio móvil	Pronóstico de corto alcance para operaciones como inventario, programación, control, fijación de precios y calendarización de promociones especiales; se emplea para calcular los componentes cíclico y estacional del método de descomposición de corto plazo	Bajo	No	4 y 5

TABLA 11-1 (continuación)

<i>Método</i>	<i>Descripción</i>	<i>Aplicaciones</i>	<i>Computadora</i>		
			<i>Costo</i>	<i>¿Necesaria?</i>	<i>Capítulo</i>
Suavización exponencial	Similar a los promedios móviles, pero los valores son ponderados exponencialmente, otorgando mayor peso a los datos más recientes	Pronósticos de corto alcance para operaciones como inventario, programación, control, fijación de precios y calendarización de promociones especiales	Bajo	Sí	4
Modelos autorregresivos	Se emplean con variables económicas para explicar las relaciones entre observaciones adyacentes en una serie de tiempo	Pronósticos de corto y mediano plazos para datos económicos —incluyendo precio,— inventario, producción, acciones y ventas—ordenados en una serie de tiempo	Medio	Sí	8 y 9
Técnicas Box-Jenkins	No supone ningún patrón particular en los datos históricos de la serie que va a ser pronosticada; usa un método iterativo de identificación y ajuste de un modelo posiblemente útil tomado de un grupo general de modelos	Igual que en el caso anterior	Alto	Sí	9
Redes neurales	Usa un programa complejo de computadora para asimilar datos relevantes y reconocer patrones mediante “aprendizaje” como lo hacen los humanos	Uso creciente de una gran variedad de aplicaciones de pronóstico; está en la fase de desarrollo	Alto	Sí	10

pasos adicionales de la fase operativa del pronóstico, haciendo énfasis en el aprendizaje de las técnicas comúnmente utilizadas para generar los pronósticos reales. Los pasos operativos clave en el proceso de pronóstico se resumen en la figura 11-1.

La recolección y el examen de los datos históricos adecuados se describieron con anterioridad en este libro (capítulo 3), al igual que las consideraciones para la selección de una técnica o un modelo de pronóstico. Como sugiere la figura 11-1, el siguiente paso a menudo es el de pronosticar varios periodos históricos donde se conocen los valores reales. Los errores resultantes se resumen de varias maneras, como se explicó en el capítulo 3, y este proceso se continúa hasta que se encuentra una técnica con una proporción suficiente entre costo y beneficio. El modelo se usa entonces para pronosticar los periodos futuros, y los resultados se incorporan al proceso de toma de decisiones de la empresa.

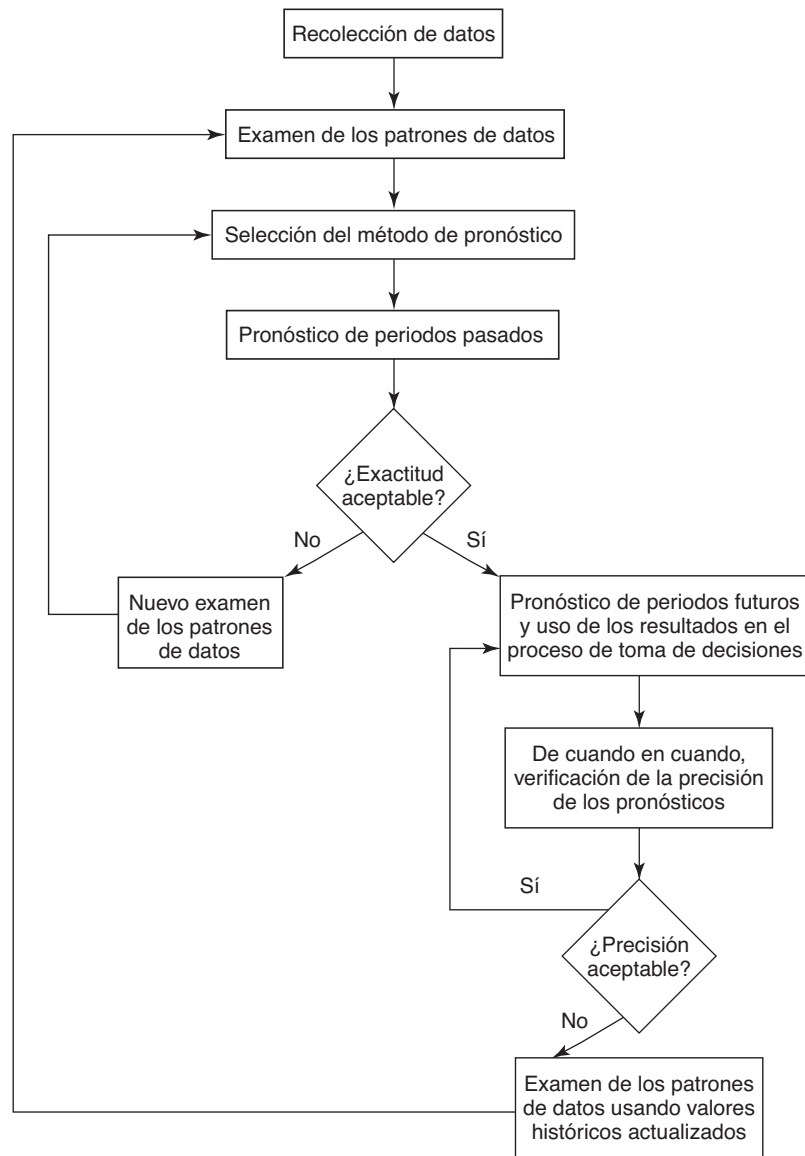


FIGURA 11-1 Fase operativa del proceso del pronóstico

De vez en cuando es necesario hacer una pausa en el proceso de pronosticar y reconsiderar los procedimientos utilizados. Los pasos usuales son los siguientes:

1. Se descartan los valores históricos más antiguos de los datos que se están usando en la técnica del pronóstico, y los valores reales más recientes se agregan al banco de datos.
2. Después de esta actualización de datos, se vuelven a estimar los parámetros en el modelo de pronóstico. Por ejemplo, es posible que el mejor valor (o valores) de la constante (o constantes) de ponderación en la suavización exponencial cambie, quizás de manera notable, cuando se agreguen datos más recientes. O tal vez cambien los coeficientes en un análisis de regresión cuando se usan datos diferentes para ajustar la función de regresión.

- Se examina la precisión del modelo de pronóstico con los nuevos parámetros. Si se considera que esta precisión es suficiente, el modelo se utiliza como antes hasta el siguiente periodo de actualización. Si la precisión del modelo se considera inadecuada o marginal, se revisan los patrones de los nuevos datos con la posibilidad de seleccionar un nuevo procedimiento para pronosticar. Este proceso continúa hasta que la precisión del modelo seleccionado, de acuerdo con el juicio que resulte de la comparación entre el pronóstico y los valores reales, se considere adecuada.

El proceso anterior se resume en el diagrama de flujo de la figura 11-1 y representa el tipo de retroalimentación que se encuentra en el diseño de sistemas de todo tipo.

En ocasiones, los pronósticos se monitorean usando constantemente una señal de control, un concepto estudiado en el capítulo 4 después de la suavización exponencial simple. La idea es establecer límites dentro de los cuales se espera que caigan los errores generados por los pronósticos si el proceso de pronóstico es el adecuado. Mientras los errores caigan dentro de estos límites aceptables, el proceso de pronóstico continúa. Tan pronto como un error cae fuera del rango aceptable, la atención de la gerencia se enfocará en el proceso del pronóstico y se aplicarán los pasos de actualización y revisión esbozados anteriormente. Este concepto se ilustra en el ejemplo 11.1.

Ejemplo 11.1

Sue Bradley es la responsable del pronóstico mensual de las ventas en dólares de su compañía. Sue seleccionó un modelo de pronóstico que tiene una tasa de error aceptable para sus gerentes. Específicamente, el error estándar de este proceso para pronosticar es de \$935; es decir, el pronóstico y los valores reales de las ventas mensuales en dólares por lo regular están alejados de \$935.

Sue supone que los errores del pronóstico tienen distribución normal con una media de cero y una desviación estándar de \$935. Ella hace esta suposición después de examinar una gráfica de los errores de pronóstico anteriores y encuentra que tienen una forma de campana alrededor de cero. Usando un nivel de confianza del 95%, determina los siguientes límites dentro de los cuales se espera que caiga el error del pronóstico de cada mes.

$$0 \pm (1.96)(935)$$

$$0 \pm 1,833$$

Por lo tanto, Sue espera que cada pronóstico mensual esté a menos de \$1,833 del valor real del mes, con un 95% de confianza. Si esto es así, el procedimiento para pronosticar continuará intacto. Pero si el error fuera mayor que \$1,833, ella examinará ambos parámetros en su técnica seleccionada para pronosticar e incluso consideraría el uso de otra técnica.

Para supervisar más fácilmente los errores del pronóstico, Sue diseñó una gráfica para seguirles la pista. Durante el transcurso de varios meses, Sue encontró dos gráficas que motivaron la revisión cuidadosa de sus procedimientos para pronosticar. La primera (figura 11-2) muestra los errores de

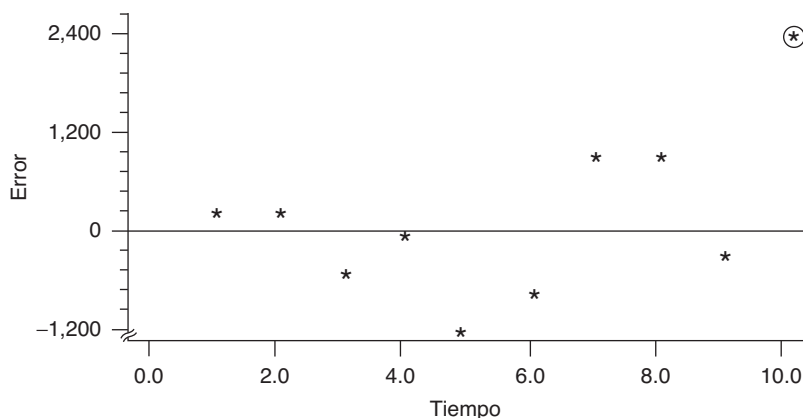


FIGURA 11-2 Primera gráfica de errores del ejemplo 11.1

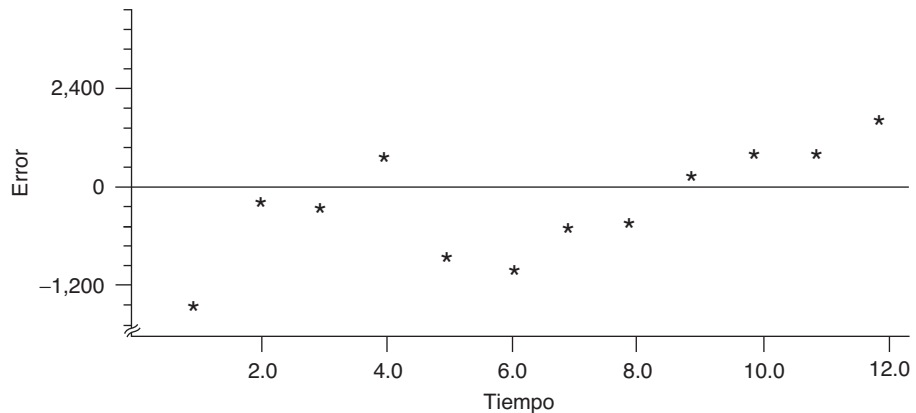


FIGURA 11-3 Segunda gráfica de errores del ejemplo 11.1

pronóstico que parecen estar distribuidos de manera aleatoria hasta los periodos más recientes. Este error fuera de la tolerancia condujo a Sue a estimar nuevamente los parámetros en su modelo de pronóstico después de actualizar su base de datos agregando los valores recientes y eliminando el mismo número de valores antiguos. Algún tiempo más tarde, una segunda gráfica de error (figura 11-3) hizo que Sue revisara una vez más su proceso de pronóstico. Si bien ninguno de los errores ha excedido el límite de tolerancia, Sue notó que los errores recientes no parecían estar distribuidos de manera aleatoria. De hecho, los errores se incrementan en la dirección positiva, y parece obvio que pronto estarán fuera de control. Sue actualizó su base de datos y, después de un examen minucioso de los patrones de datos, seleccionó una nueva técnica para pronosticar.

REVISIÓN DE LOS PASOS DEL PRONÓSTICO

En el capítulo 1 se hizo varias veces el planteamiento de que los juicios sensatos deben aplicarse constantemente junto con el análisis cuantitativo si se desea obtener pronósticos útiles y precisos. Se plantearon varias preguntas para saber si el proceso del pronóstico se administró correctamente. Aquí se presentan una vez más:

- ¿Por qué son necesarios los pronósticos?
- ¿Quién usará el pronóstico y cuáles son sus requerimientos específicos?
- ¿Qué nivel de detalle o de generalización se requiere, y cuál es el horizonte de tiempo adecuado?
- ¿Qué datos están disponibles y serán éstos suficientes para generar el pronóstico que se necesita?
- ¿Cuánto costará el pronóstico?
- ¿Qué tan exacto se espera que sea el pronóstico?
- ¿El pronóstico estará a tiempo para ayudar en el proceso de toma de decisiones?
- ¿El pronosticador entiende claramente cómo se usará el pronóstico en la organización?
- ¿Está disponible un proceso de retroalimentación para evaluar el pronóstico y ajustar el proceso como corresponde?

Cuando las preguntas anteriores se analizan y contestan adecuadamente, estamos en condiciones de iniciar el proceso real de pronóstico. En el capítulo 1 se listaron los pasos que se siguen en dicho proceso:

Planteamiento del problema y recopilación de datos
 Manipulación y depuración de datos
 Elaboración y evaluación del modelo

Implementación del modelo (el pronóstico real)
Evaluación del pronóstico

A esta lista se agrega otro paso: la retroalimentación después del proceso de pronóstico debe ayudar a determinar si se logró la precisión suficiente y si la gerencia está satisfecha con la utilidad y el costo del pronóstico en el proceso de toma de decisiones.

Si bien el interés principal de este libro ha sido la elaboración de modelos o la selección del método adecuado para pronosticar, también se ha enfatizado la importancia de la administración del proceso para la elaboración de pronósticos. Las preguntas listadas anteriormente son importantes en todos los escenarios de pronósticos y deben analizarse si queremos obtener resultados útiles. Especialmente recomendamos estudiar los casos al final de este capítulo, los cuales están diseñados para enfatizar los aspectos administrativos de los pronósticos.

RESPONSABILIDAD AL ELABORAR EL PRONÓSTICO

La ubicación del proceso de elaboración de pronósticos en una empresa varía dependiendo del tamaño de la firma, la importancia asociada al pronóstico formal y la naturaleza del estilo de administración de la compañía. La responsabilidad de efectuar pronósticos se ubica en algún punto entre un departamento separado dedicado a pronosticar y varias unidades administrativas pequeñas que realizan pronósticos, las cuales no tienen conexión directa con otros esfuerzos en el mismo sentido dentro de la empresa.

Los equipos dedicados a realizar pronósticos son más comunes en las empresas grandes que en las pequeñas. Las empresas grandes se pueden dar el lujo de contratar los expertos necesarios para la elaboración de pronósticos complejos y pueden suministrarles modernos elementos informáticos (software y computadoras). La ventaja de este trabajo centralizado es que esta especialización está a la disposición de todas las unidades de la organización. La desventaja es que la coordinación entre el personal dedicado al pronóstico y los gerentes de línea es muy difícil de alcanzar. El personal que elabora pronósticos tal vez tenga que dedicar más tiempo a la negociación con los usuarios y a la explicación de sus funciones que a la elaboración misma del pronóstico.

En el otro extremo se encuentra el caso de las empresas en las que los pronósticos se realizan en cada una de las áreas que la integran, las cuales no tienen coordinación o cooperación entre sí. La ventaja de este proceso es que se eliminan los malos entendidos entre quienes elaboran el pronóstico y quienes lo usan: son las mismas personas. Los pronósticos generados en estas condiciones tienden a ser aceptados y considerados en el proceso de toma de decisiones. La desventaja es que la complejidad y, por lo tanto, la precisión del pronóstico suelen ser difíciles de alcanzar porque la disponibilidad informática y la especialización técnica no están distribuidas de manera uniforme entre los usuarios. En general, es difícil persuadir a la alta dirección de adquirir el hardware y software apropiados, por ejemplo, cuando saben que éstos existen en otras unidades de la compañía.

Más aún, la falta de coordinación y cooperación a menudo causa problemas cuando una unidad utiliza pronósticos generados por otra unidad de la compañía. Si las unidades no son responsables de la calidad de los pronósticos, es posible que se presenten sesgos que traen consigo dificultades operativas.

En cierto caso, el departamento de marketing era responsable de los pronósticos. El marketing se midió por la disponibilidad del producto, la cual se definió como el porcentaje de tiempo que un producto particular estaba disponible para ser comprado por un cliente. Sin embargo, no se penalizaba al área de marketing por algún exceso en el inventario que pudiera darse como resultado de la confianza del área de producción en los pronósticos de marketing. Si bien esos pronósticos se basaban en las condiciones del mercado, al mismo tiempo estaban influidos fuertemente por las expectativas de cuotas e ingresos del mercado. En consecuencia, el equipo de marketing generó pronósticos de demanda optimistas para garantizar la disponibilidad del producto que facilitaría cumplir sus objetivos de ingresos. Por otro lado, la

planeación de producción de unidades se midió tanto por la disponibilidad del producto como por el exceso de inventario. Este grupo usó métodos estadísticos para amortiguar la incertidumbre y estuvo muy motivado para mantener los pronósticos precisos y sin sesgos.

Muchas organizaciones intentan ubicar la responsabilidad de elaboración de pronósticos en el punto medio entre los dos extremos que acabamos de mencionar. Se puede asignar un pequeño equipo de pronosticadores a las subunidades dentro de la firma para prestar servicio a las diferentes áreas funcionales, de acuerdo con las necesidades de estas últimas. La tarea de este equipo de pronósticos implica la coordinación adecuada con clientes, así como la generación de pronósticos confiables. Algunas veces, la responsabilidad de elaboración de los pronósticos se combina con otras funciones del personal, tales como apoyo estadístico o computacional.

La disponibilidad de computadoras pequeñas y de bajo costo y el software de pronósticos ha restringido la función de los pronósticos dentro de la organización. Ahora es posible para los gerentes tener acceso a complejas herramientas de pronóstico por una fracción de lo que costaban hace algunos años. Sin embargo, el conocimiento requerido para usar adecuadamente el hardware y el software no viene incluido en el mismo paquete; la necesidad de comprender el uso adecuado de las técnicas de pronóstico se ha incrementado conforme el manejo del equipo de cómputo ha dejado de ser una función exclusiva de los “expertos” para distribuirse entre los usuarios de una organización.

COSTOS DE LOS PRONÓSTICOS

El hardware y software computacional, más el personal especializado, son los costos más obvios implicados en la creación de los pronósticos. Pero existen costos adicionales que no son obvios, como el consumo de tiempo del personal de la compañía y el dinero que esto implica. El tiempo que destina el personal asalariado (en distintos departamentos) a recopilar datos para elaborar pronósticos, supervisar el proceso e interpretar los resultados debe considerarse como parte de los costos del pronóstico. Estos costos deben equilibrarse con los beneficios recibidos si se desea tomar decisiones racionales en relación con la utilidad de los pronósticos resultantes.

Una alternativa para la elaboración interna de pronósticos es contratar consultores con esa finalidad. Esta práctica es especialmente atractiva si la necesidad de un pronóstico es un requerimiento ocasional y no una necesidad continua. Por otra parte, el requerimiento de un pronóstico más allá de la capacidad técnica de la compañía sugiere la contratación de consultores profesionales. La contratación externa de asistencia para pronosticar hace más fácil la identificación de los costos.

SISTEMAS DE INFORMACIÓN PARA ADMINISTRAR Y PRONOSTICAR

Los sistemas de información gerencial (SIG) de las empresas modernas han incrementado en años recientes su complejidad y utilidad. El principal beneficio en el proceso de elaboración de pronósticos implica su enorme capacidad para recolectar y registrar datos a través de la organización. El pronosticador debe resistir la tentación de recopilar datos para los modelos de pronóstico que se están usando si es que ya han sido recolectados y registrados en el sistema de información gerencial de la compañía.

El rápido desarrollo de las redes de computadoras dentro de muchas firmas también ha incrementado el acceso de los usuarios a los datos almacenados. Además, programas de software integrador como SAP¹ tienen módulos de pronóstico que utilizan bases de datos a las que se puede tener acceso con el programa.

¹SAP, con sus oficinas generales en Walldorf, Alemania, es, en el momento de escribir este libro, la compañía de software interempresarial más grande del mundo.

Es indispensable confiar en las bases de datos existentes, incluso si los datos disponibles no están exactamente en el formato o la secuencia de tiempo que desea el pronosticador. Se debe evaluar la modificación del modelo de pronóstico o de los datos disponibles antes de descartar los datos del SIG y efectuar la recopilación de nuevos datos. Este consejo supone, desde luego, que la recopilación de datos nuevos implica un gasto considerable de tiempo y dinero. Si los datos necesarios para un modelo de pronóstico fueran fáciles de obtener en el formato correcto, esto sería preferible al uso de datos preexistentes en una base de datos que no tienen la forma correcta o son anacrónicos.

Una ventaja adicional para usar los datos disponibles en el SIG de la compañía o red interna es que el proceso de elaboración del pronóstico se convierte en un componente del sistema. Como tal, entra a la red de distribución y toma de decisiones ya establecida por el sistema y puede favorecer la incorporación al proceso de toma de decisiones de la compañía. Esta situación contrasta con un procedimiento de elaboración de pronósticos que intenta infiltrar los procedimientos de toma de decisiones que ya están utilizando los gerentes de la compañía.

IMPORTANCIA DE LA GERENCIA EN LA ADMINISTRACIÓN DE LOS PRONÓSTICOS

Hay que considerar varios factores importantes si los pronósticos generados en una empresa se van a convertir en parte importante del proceso de toma de decisiones. Primero, debe reconocerse que los gerentes eficaces están interesados en resultados útiles y prácticos. En general, los pronósticos deben satisfacer las necesidades de esos gerentes.

Segundo, los pronósticos deben ser lo suficientemente precisos para ser útiles. Los buenos gerentes no serán fieles a un proceso de pronóstico, sin importar la complejidad de éste, si no se obtienen resultados precisos.

Por último, en el proceso de pronóstico debe considerarse la intuición sobre el costo-beneficio de los buenos gerentes. La habilidad para evaluar situaciones en términos del costo en el que se incurre contra el beneficio recibido es la clave de un proceso efectivo de administración y debe ser considerado por el equipo encargado de efectuar los pronósticos. Esta situación a menudo crea dificultades entre el pronosticador y el usuario. El pronosticador siempre debe tener en mente que el objetivo del pronóstico es mejorar la eficiencia del negocio, lo cual incluye la maximización de las utilidades.

EL FUTURO DE LOS PRONÓSTICOS

Como se mencionó en el capítulo 1, los pronósticos continúan ganando importancia en las organizaciones modernas debido a la complejidad creciente del mundo de los negocios, junto con la disponibilidad de software y equipo de cómputo menos caros y más poderosos. La competencia permanente en las áreas de computadoras pequeñas y del software es obvia para todos; menos obvias son las tendencias a largo plazo que lentamente cambian el escenario de los negocios y que ejercen presiones sutiles pero poderosas sobre el modo de operar de la gente de negocios.

Considere algunas de las “megatendencias” que identificó Naisbitt (1982) en su libro con ese título (para referencias adicionales de esta clase, véase el capítulo 10). Estas importantes tendencias ya están aquí en su mayoría y refuerzan la necesidad de contar con pronósticos efectivos:

- De sociedad industrial a sociedad de la información
- De economía nacional a economía mundial
- De pensamiento de corto plazo a pensamiento de largo plazo
- De centralización a descentralización
- De dos opciones (esto o aquello) a múltiples opciones

La economía actual es una economía de feroz competencia global. Gracias a Internet, una gran cantidad de información está disponible a cualquier hora del día. Los cambios en el mercado son rápidos y repentinos, y probablemente permanentes. El ciclo de vida de un producto continúa acortándose. Los negocios por vía electrónica (e-business) son un modo de vida, y las transacciones a través de la Web continuarán incrementándose.

Estas tendencias enfatizan la necesidad creciente por parte de la gerencia de familiarizarse con asuntos complicados y, en particular, con la necesidad de desarrollar métodos complejos para manejar la incertidumbre futura. En otras palabras, hacen énfasis en la importancia creciente de combinar el buen juicio con métodos complejos de manipulación de datos para el pronóstico de negocios sólidos. Conforme las tendencias de Naisbitt y la dinámica creciente del ambiente de negocios continúan desarrollándose, cobra más importancia la habilidad de los líderes de negocios para reaccionar de forma rápida, y sin perder de vista la rentabilidad, ante los sucesos cambiantes. La pregunta básica en los negocios “¿Qué pasará en seguida?” siempre tendrá gran importancia; los pronósticos tienen relación directa con esta pregunta. Nuevos retos y oportunidades esperan a los pronosticadores.

Problemas

1. Escriba una respuesta breve para cada una de las siguientes afirmaciones.
 - a) Los pronósticos siempre son incorrectos, entonces, ¿por qué conceder tanta importancia a la planeación de la demanda?
 - b) Un buen pronóstico es muy caro.
 - c) Todo lo que necesitamos hacer es contratar una “persona cuantitativa” para realizar nuestro pronóstico.
2. ¿Se le ocurren algunas tendencias futuras en los negocios, no discutidas en este capítulo y que puedan influir en el modo de hacer pronósticos en el futuro? Explique brevemente.

CASOS

CASO 11-1 BOUNDARY ELECTRONICS

Boundary Electronics es un gran proveedor de productos electrónicos para el hogar. Entre sus productos de mayor venta están las videograbadoras y los sistemas de televisión por satélite. Como la compañía ha crecido muy rápidamente, Guy Preston, presidente de Boundary, está preocupado de que un cambio en las condiciones del mercado pueda alterar su patrón de ventas.

Al preguntar a sus gerentes acerca del futuro de la compañía, Guy descubrió dos cosas. Primero, la mayoría de los gerentes están muy ocupados pensando en los problemas diarios relacionados con la satisfacción de la demanda creciente como para pensar en el futuro a largo plazo. Segundo, sus opiniones varían considerablemente, desde muy optimistas a muy pesimistas. Como presidente de la compañía, Guy piensa que tiene una responsabilidad muy seria para evaluar el ambiente de la compañía en el futuro.

Después de pensar en este asunto, Guy planea una reunión sabatina con seis miembros de su equipo de alta dirección. Rentó una sala de juntas en un hotel local y organizó los almuerzos y los descansos para tomar café durante el día. Cuando el equipo se reunió el sábado por la mañana, presentó el tema del día y luego instruyó a cada persona para que describieran en una o dos páginas el ambiente operativo de la compañía para los siguientes 20 años en cada una de las siguientes situaciones:

1. El ambiente de la compañía esencialmente continúa siendo el mismo. Los productos demandados por el mercado serán modificaciones de los productos actuales, y no intervendrá nueva tecnología.
2. Los cambios tecnológicos mayores volverán obsoletos los productos actuales de la compañía.

Se tendrán que desarrollar productos nuevos para satisfacer las demandas recreativas de la población de Estados Unidos.

3. Entre estos dos extremos, ¿cuál es el escenario más probable del ambiente operativo de la compañía?

Guy les da una hora a todos los miembros del equipo para desarrollar los escenarios de cada una de estas tres situaciones. Durante esta hora, Guy piensa

acerca de lo que pasará el resto del día. Espera que haya algunas ideas novedosas desarrolladas por sus gerentes y que las discusiones posteriores sean animadas e interesantes. Además de obtener ideas para él mismo, Guy espera que el ejercicio del día ayude a sus gerentes a mirar más allá de los problemas y las oportunidades inmediatos de la compañía y les dé una visión de más largo plazo de la compañía.

PREGUNTAS

1. ¿Qué proceso cree usted que debe usar Guy después de concluir la actividad de formulación de escenarios por escrito?
2. ¿Existe algún otro enfoque que Guy podría intentar, considerando sus objetivos?
3. ¿Cree usted que Guy alcanzará sus objetivos con la reunión del sábado?

CASO 11-2 BUSBY ASSOCIATES

Jill Tilley estaba recién graduada de la escuela de negocios de una universidad cuando se empleó con Busby Associates, un exportador grande de equipo agrícola. El presidente de Busby sabía, por el currículum de Jill, que esta última había tomado un curso de pronósticos y decidió asignarle una tarea de pronóstico que habían analizado varias veces los directores de Busby.

El presidente de Busby creía que había una fuerte relación entre las ventas de exportación de la compañía y las cifras nacionales de exportaciones. Estas últimas estaban disponibles en las oficinas gubernamentales, de manera que el proyecto de Jill era pronosticar una buena y representativa variable de las exportaciones. Si este trabajo tenía éxito, el presidente de Busby creía que la compañía tendría una poderosa herramienta para pronosticar sus ventas de exportación.

Por el momento, Jill localizó el número más reciente del *Survey of Current Business* en una biblioteca local y registró las cifras trimestrales de bienes de consumo exportados en miles de millones de dólares. Ella creía que esta variable era un buen representante del total de exportaciones nacionales. Anticipando la posibilidad de usar un análisis de regresión para el pronóstico, también registró los valores de otras variables que creyó que podrían correlacionarse bien con esta varia-

ble dependiente. Terminó con los valores de cuatro variables de 14 trimestres.

Luego calculó tres variables adicionales a partir de los valores de su variable dependiente: el cambio en Y , el cambio porcentual en Y , y Y retrasada un periodo. Por el tiempo en que empezó a pensar en varias maneras de pronosticar su variable, había recopilado los datos presentados en la tabla 11-2.

Jill tecleó sus datos en un programa de computadora que ejecutó el análisis de regresión y calculó la matriz de correlación de sus siete variables. Después de examinar esta matriz, seleccionó tres regresiones con una variable explicativa y seis regresiones con dos variables explicativas. Luego corrió estas regresiones y seleccionó la que consideraba la mejor: utilizó una variable explicativa (Y retrasada un periodo) con los siguientes resultados:

$$r^2 = .98$$

$$t = 25.9$$

$$F = 671.6$$

Jill examinó las autocorrelaciones residuales y encontró que eran pequeñas. Concluyó que no necesitaba corregir ninguna autocorrelación en su regre-

TABLA 11-2 Serie de tiempo trimestral

Periodo	Variable						
	1	2	3	4	5	6	7
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 ^a	112.1	1.3	3.19	40.7

^aEstimado

Clave de variable:

1: Bienes de consumo, exportados, miles de millones de dólares

2: Ahorros personales brutos, miles de millones de dólares

3: Ingreso nacional por comercio al menudeo, miles de millones de dólares

4: Índices de precio ponderado fijo para compras de la defensa nacional, 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 para las variables 1 a 4: *Survey of Current Business* (Departamento de Comercio de Estados Unidos) 70(7) (julio de 1990).

sión. Jill creyó que había encontrado una buena variable explicativa (*Y* retrasada un periodo).

Por otra parte, se dio cuenta de que el tamaño de su muestra era bastante pequeño: 13 trimestres. Volvió a consultar el *Survey of Current Business* para recopilar más datos y se decepcionó al descubrir que, durante los años en los que ella estaba interesada, la definición de su variable dependiente cambió, teniendo como resultado una serie de tiempo inconsistente. Es decir, la serie dio un salto hacia arriba a la mitad del camino del periodo que estaba estudiando.

Jill comentó este problema con su jefe y estuvo de acuerdo en usar las exportaciones totales de mercancías como la variable dependiente en vez de las exportaciones de bienes de consumo. Jill descubrió que esta variable era consistente a lo largo de varios números del *Survey of Current Business* y que se podían recopilar varios años de datos. Reunió los datos que se pre-

sentan en la tabla 11-3, con los datos retrasados un periodo, y corrió otra vez el análisis de regresión usando *Y* retrasada un periodo como variable explicativa.

Esta vez encontró de nuevo buenos datos estadísticos en el reporte de la regresión, excepto por la autocorrelación residual del retraso 1. Ese valor, .47, era significativamente diferente de cero en el nivel del 5%. Jill no estaba segura de qué hacer. Ejecutó corridas adicionales, incluyendo el número de periodo y el cambio de *Y* como variables explicativas adicionales. Pero después de examinar los residuales, no podía concluir que la autocorrelación hubiera sido eliminada. Jill decidió considerar otras técnicas para pronosticar su nueva variable dependiente: las exportaciones totales de mercancías. Usó los datos de la serie de tiempo de la columna *Y* de la tabla 11-3. La gráfica de la serie de tiempo de las exportaciones totales de mercancías para los años de 1984 al segundo trimestre de 1990 se presenta en la figura 11-4.

TABLA 11-3 Datos trimestrales de la serie de tiempo: Total de exportaciones de mercancías (en 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, varias ediciones.

Después de estudiar la figura 11-4, Jill decidió usar sólo los últimos 16 datos en su pronóstico. Observó que, iniciando con el periodo 9, la serie había tenido un incremento relativamente constante, mientras que antes de ese periodo mostraba un incremento y una declinación.

El examen de sus datos reveló que no había estacionalidad, así que decidió intentar dos métodos de suavización:

Suavización exponencial simple

Suavización exponencial lineal de Holt, la cual puede manejar una tendencia en los datos

Para la suavización exponencial simple, Jill restringió la constante de suavización, α , para dejarla entre 0 y 1, y encontró que la mejor α estaba muy cerca de 1. El programa Minitab seleccionó los valores óptimos de las constantes de suavización de Holt, α y β . Las medidas de error y las constantes de suavización seleccionadas se presentan en la tabla 11-4. A partir de los resultados que se presentan en la tabla 11-4, Jill determinó que la suavización lineal de Holt era la más prometedora y usó este procedimiento para generar los pronósticos para los cuatro trimestres siguientes.

Jill observó que la constante de suavización óptima empleando la suavización exponencial simple era casi 1.00 (.99). Aparentemente, al revisar los datos en esta modalidad, el programa estaba usando básicamente cada dato para predecir el siguiente. Esto equivale a usar un método del pronóstico informal simple para pronosticar, es decir, un modelo que dice que las diferencias consecutivas son aleatorias.



FIGURA 11-4 Gráfica de valores de datos trimestrales: Total de exportaciones de mercancías, del primer trimestre de 1984 al segundo trimestre de 1990 (miles de millones de \$)

TABLA 11-4

<i>Técnica</i>	<i>MSE</i>	<i>MAPE</i>	<i>Constante(s) óptima(s) de suavización</i>
Exponencial simple	203.7	4.2	$\alpha = .99$
Lineal de Holt	35.8	1.5	$\alpha = 1.77, \beta = .14$

Usando el método de suavización de Holt, los pronósticos para los siguientes cuatro trimestres futuros son

<i>Periodo</i>	<i>Pronóstico</i>
17	400.2
18	408.8
19	417.4
20	425.9

Jill se dio cuenta de que al terminar cada trimestre un nuevo valor real de las exportaciones totales de mercancías estaría disponible y que los pronósticos de periodos futuros podían actualizarse.

Jill se reunió entonces con su jefe para discutir los resultados. Ella le dijo que pensaba que tenía un buen modo para pronosticar la variable nacional, las exportaciones totales de mercancías, usando suavización exponencial con ajustes de tendencia. Su jefe le pidió que le explicara este método, y ella así lo hizo. Su siguiente tarea era usar los datos reales para verificar la sospecha del presidente de Busby: que las exportaciones de Busby estaban bien correlacionadas con las exportaciones nacionales. Si ella podía establecer este vínculo, Busby tendría un buen método para pronosticar sus exportaciones y podría utilizar los pronósticos para planear las operaciones futuras.

PREGUNTAS

1. Jill no consideró la combinación de los pronósticos generados por los dos métodos que analizó. ¿Qué sucedería si lo hiciera? ¿Cuáles serían las ventajas y desventajas de tal acción?
2. Las constantes óptimas de suavización usadas por la suavización exponencial lineal de Holt fueron $\alpha = 1.77$ y $\beta = .14$. Conforme surgen datos nuevos en los siguientes trimestres, Jill probablemente debería correr otra vez los datos para ver si estos valores cambian. ¿Con qué frecuencia cree usted que debería hacer esto?
3. ¿Es posible que la selección del método de pronóstico pueda cambiar a otra técnica conforme se agrega un nuevo trimestre de datos a la base de datos? ¿Debe Jill correr nuevamente su análisis completo de cuando en cuando para verificar esto?
4. Un colega de Jill le sugirió que hiciera un intento con la metodología Box-Jenkins ARIMA. ¿Qué recomendación haría usted a Jill si decide intentar con Box-Jenkins? ¿Puede usted identificar un modelo ARIMA tentativo?

CASO 11-3 CONSUMER CREDIT COUNSELING

Las operaciones de Consumer Credit Counseling se describieron en los casos 1-2 y 3-3.

El director ejecutivo, Marv Harnishfeger, concluyó que la variable más importante que CCC necesitaba pronosticar era el número de clientes nuevos que serían atendidos durante el resto de 1993. Marv le dio a Dorothy Mercer los datos mensuales del número de clientes atendidos por CCC en el periodo de enero de 1985 a marzo de 1993 (véase el caso 3-3).

Dorothy, con la ayuda de usted, ha intentado varios modos de pronosticar la variable más importante. Estos trabajos se explicaron en los casos 4-3, 5-3, 6-5, 8-5 y 9-3. Al concluir tales intentos de pronóstico, Dorothy decidió que era tiempo de resumir estos trabajos e intentar determinar el método para pronosticar el resto del año.

TAREA

1. Suponga que Dorothy le pide a usted ayudarla con este problema. Escriba un reporte recomendando un curso de acción. Tenga en mente que Marv debe desarrollar los pronósticos del número de clientes atendidos tan precisos como sea posible y que él

pueda usar en la toma de decisiones diaria. Sea específico acerca de sus recomendaciones a Dorothy y Marv. Recuerde tomar en cuenta los puntos discutidos en este capítulo, como los costos.

CASO 11-4 MR. TUX

El dueño de varias tiendas de alquiler Mr. Tux, John Mosby, ha intentado varios modos de pronosticar su variable más importante, las ventas mensuales en dólares. Sus trabajos se explicaron en los casos 1-1, 2-2, 3-2, 4-2, 5-2, 6-4, 8-4 y 9-2. Luego de concluir, John decidió que era tiempo de resumir estos trabajos e intentar llegar al método de pronóstico del futuro. Él se dio cuenta de que debía actualizar tanto sus datos como su método de pronóstico en algún momento del futuro, pero necesitaba seleccionar una manera de pronosticar los siguientes meses.

Para empezar, John hizo un resumen de los resultados de los métodos que había intentado hasta el momento:

- Caso 2-2: Usar el promedio anual para pronosticar las futuras ventas anuales podría funcionar, pero como John observó una tendencia ascendente, necesita un buen modo de extender estos promedios hacia el futuro. John también está bastante preocupado por el efecto estacional, puesto que sabe que sus ventas mensuales varían considerablemente. Sus trabajos usando promedios mensuales no rindieron frutos.
- Caso 3-2: Utilizando el programa de Minitab, John estableció que en estos datos existía tanto una tendencia como un efecto estacional. Aun cuando él sabía que estos elementos existían antes de iniciar, estaba satisfecho de ver que los resultados de su programa de computación los confirmaban estadísticamente. El programa también indicaba que varios coeficientes de autocorrelación estaban fuera de los límites de error de muestreo, lo que le indicaba a John que tanto los efectos de tendencia como los estacionales necesitaban reflejarse en su modelo final de pronóstico.
- Caso 4-2: Cuando John utilizó la suavización exponencial, incluyendo métodos que tomaban en cuenta factores de tendencia y estacionales,

las mediciones del error resultante eran insatisfactorias. Se dio cuenta de que estas mediciones, como el error promedio y el error porcentual promedio, procedían de la predicción de valores pasados de su variable. Pero como eran muy grandes, no quería usar estas técnicas para pronosticar el futuro.

- Caso 5-2: Finalmente John obtuvo algunos resultados alentadores usando el método de descomposición para construir una línea de tendencia, índices estacionales y un componente cíclico para sus datos. Estaba listo para mostrar a su ejecutivo bancario los índices estacionales y realizar los acuerdos que deseaba en los pagos de su préstamo. Además, generó pronósticos para unos cuantos meses siguientes volviendo a reunir los componentes estimados. Sin embargo, John estaba algo intranquilo por el amplio rango de sus proyecciones.
- Caso 6-4: El análisis de regresión simple fue la siguiente técnica que John intentó, usando el periodo de tiempo como la variable independiente o explicativa. Pensó que esta variable explicaría el factor de tendencia que sabía que existía en los datos. Este método no explicaba la estacionalidad en las ventas, y el valor de 56% de la r^2 era insatisfactorio.
- Caso 8-4: Después, John usó una regresión múltiple usando tanto el número de periodo para reflejar la tendencia como una serie de variables ficticias para explicar el efecto estacional (meses). El valor de su R^2 de 88% era una mejora considerable sobre su regresión simple, pero el error del pronóstico del 21% para los últimos 12 meses de sus datos, de acuerdo con las mediciones del error porcentual absoluto medio (MAPE), era inaceptable. Con esta clase de error, John decidió no usar la regresión múltiple. También intentó un modelo autorregresivo estacional, el cual tuvo un resultado del 91% para la R^2 . John estaba muy contento con este resultado.

- Caso 9-2: La metodología Box-JenKins ARIMA incomodaba a John desde el principio porque no la entendía plenamente. Sin embargo, admitía que este modelo autorregresivo estacional era un modelo ARIMA particular. Quería saber si podía mejorar su modelo. Sabía que tenía que explicar sus pronósticos a los inversionistas y ejecutivos bancarios en su intento por obtener capital para la expansión, así que quería un método de pronóstico que fuera preciso y comprensible.

Pensando en estas cosas, John se dio cuenta de que el tiempo transcurría y que debía generar cuanto antes los pronósticos de sus ingresos mensuales. Tenía el tiempo limitado para realizar modificaciones a los métodos que usó y pensó en combinar dos o más métodos. Pero no tenía tiempo para adquirir software nuevo e intentar métodos completamente diferentes. Cuando se preguntó qué debía hacer, miró uno de sus refranes favoritos escrito en un cuadro colgado en la pared de su oficina: “Hagamos algo, incluso si no es del todo correcto”.

TAREA

1. Suponga que usted ha sido contratado para ayudar a John Mosby con su pronóstico. Redacte un escrito que contenga un resumen de sus esfuerzos a la fecha y recomiéndele un curso de acción. Tenga en mente que John debe desarrollar rápidamente

los pronósticos para las ventas mensuales que sean tan precisos como sea posible y que él pueda usar en sus pláticas con los inversionistas. Sea muy específico acerca de lo que usted recomienda al dueño de Mr. Tux.

CASO 11-5 ALOMEGA FOOD STORES

El ejemplo 1.1 describía cómo Julie Ruth, presidente de Alomega Food Stores, recopiló los datos de las ventas mensuales de su compañía junto con otras variables que, en su opinión, podrían estar relacionadas con las ventas. Los casos 2-3, 3-4, 5-6, 8-7 y 10-2 describían los esfuerzos de Julie para usar varios procedimientos de Minitab para hacer pronósticos significativos de las ventas mensuales.

Julie sabía que su equipo técnico estaba usando los mismos datos para generar buenos pronósticos, pero no sabía cómo iban. Además, ella realmente quería obtener por su cuenta un buen método de pronóstico. Sabía que, al ser la primera mujer que ocupaba la presi-

dencia de Alomega, había saltado por encima de algunos candidatos al puesto y que podría haber algún resentimiento entre su equipo administrativo. Estaba especialmente sensible a los comentarios negativos de Jackson Tilson, su gerente de producción, durante una junta reciente (véase el final del ejemplo 1.1).

Revisando su trabajo, Julie decidió descartar el análisis de regresión simple que realizó y que se resumió en el caso 2-3. Tenía facultad para elegir entre el análisis de descomposición descrito en el caso 5-6, la regresión múltiple descrita en el caso 8-7, y la combinación de métodos descrita en el caso 10-2.

PREGUNTAS

1. Suponga que usted fue contratado recientemente por Alomega y se le asigna la tarea de asistir a Julie en el desarrollo de un método efectivo para pronosticar las ventas mensuales. Después de revisar los trabajos realizados por Julie hasta la fecha, ¿cuál de los métodos que ha intentado le recomendaría?
2. Con base en su respuesta a la pregunta 1, elabore un escrito detallado para Julie en el que describa las razones de su selección e indique hasta qué grado será efectivo ese método de pronóstico.

3. Además de los métodos que Julie estaba considerando, ¿qué otros procedimientos sugeriría usted? Utilice Minitab o algún otro paquete de software para pronosticar e intente otros métodos; luego, compárelos con su respuesta a la pregunta 1.
4. ¿Debe Julie combinar pronósticos?

CASO 11-6 SOUTHWEST MEDICAL CENTER

Mary Beasley no está muy feliz. Ha intentado varios métodos con series de tiempo para pronosticar el total de visitas mensuales facturables al Medical Oncology (véanse los casos 4-7, 5-8, 8-9 y 9-10). Ella ha tenido grandes éxitos con un modelo ARIMA, aun cuando las autocorrelaciones residuales evaluadas por el estadístico Q de Ljung-Box (véase la ecuación 9.5) parecen ser grandes. Al examinar una gráfica de su serie (véase la figura 4-23), Mary nota que el comportamiento de las visitas totales de los años fiscales 1997-1998, 1998-1999 y 1999-2000 —tres años en la mitad de los datos— es

diferente al comportamiento registrado en los años restantes. El total mensual de visitas generalmente es más grande y más volátil para estos años fiscales intermedios. Es como si esa parte de la serie de tiempo se hubiera “levantado” del resto de las series. Mary no está segura de por qué ése debía ser el caso y se pregunta si quizás hubo un cambio en el procedimiento, o alguna otra intervención, que sea responsable del cambio de carácter de la serie de tiempo en estos años intermedios.

TAREA

1. Usted ha sido contratado para ayudar a Mary con el pronóstico. Elabore un escrito que resuma su trabajo hasta la fecha y en el que le recomiende un curso de acción. Asegúrese de que su escrito incluya los siguientes asuntos: ¿Debe Mary aceptar los pronósticos de uno de sus procedimientos actuales? Si es así, ¿cómo debería “vender” los pronósticos a la dirección? Dado el carácter algo inconsistente de su serie de tiempo, ¿debe Mary buscar otra técnica de pronóstico? ¿Debería usar uno de sus métodos actuales, pero con un conjunto de datos modificados?

Referencias

Armstrong, J. S., ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, Mass.: Kluwer, 2001.

Naisbitt, J. *Megatrends*. Nueva York: Warner Books, 1982.

Apéndice A

Datos del caso 7-1

Y	X_1	X_2	X_3	X_4	X_5	X_6	Y	X_1	X_2	X_3	X_4	X_5	X_6
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 B

Tablas

Tabla B-1	Términos individuales de la distribución binomial
Tabla B-2	Áreas de distribución de probabilidad normal estándar
Tabla B-3	Valores críticos de t
Tabla B-4	Valores críticos de chi cuadrada
Tabla B-5	Distribución F
Tabla B-6	Límites de la prueba de Durbin-Watson

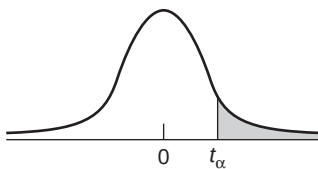
TABLA B-1 Términos individuales de la distribución binomial

		p																		
n	x	.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
2	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000	.4950	.4800	.4550	.4200	.3750	.3200	.2550	.1800	.0950
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500	.3025	.3600	.4225	.4900	.5625	.6400	.7225	.8100	.9025
	3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250	.0911	.0640	.0429	.0270	.0156	.0900	.0034	.0010
3	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750	.3341	.2880	.2389	.1890	.1406	.0960	.0574	.0270	.0071
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750	.4084	.4320	.4436	.4410	.4219	.3840	.3251	.2430	.1354
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250	.1664	.2160	.2746	.3430	.4219	.5120	.6141	.7290	.8574
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625	.0410	.0256	.0150	.0081	.0039	.0016	.0005	.0001	.0000
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500	.2005	.1536	.1115	.0756	.0469	.0256	.0115	.0036	.0005
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750	.3675	.3456	.3105	.2646	.2109	.1536	.0975	.0486	.0135
5	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500	.2995	.3456	.3845	.4116	.4219	.4096	.3685	.2916	.1715
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625	.0915	.1296	.1785	.2401	.3164	.4096	.5220	.6561	.8145
	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313	.0185	.0102	.0053	.0024	.0010	.0003	.0001	.0000	.0000
6	1	.2036	.3281	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1563	.1128	.0768	.0488	.0284	.0146	.0064	.0022	.0004	.0000
	2	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125	.2757	.2304	.1811	.1323	.0879	.0512	.0244	.0081	.0011
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125	.3369	.3456	.3364	.3087	.2637	.2048	.1382	.0729	.0214
7	4	.0000	.0004	.0022	.0064	.0146	.0283	.0488	.0768	.1128	.1562	.2059	.2592	.3124	.3601	.3955	.4096	.3915	.3281	.2036
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312	.0503	.0778	.1160	.1681	.2373	.3277	.4437	.5905	.7738
	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156	.0083	.0041	.0018	.0007	.0002	.0001	.0000	.0000	.0000
8	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938	.0609	.0369	.0205	.0102	.0044	.0015	.0004	.0001	.0000
	2	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344	.1861	.1382	.0951	.0595	.0330	.0154	.0055	.0012	.0001
	3	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125	.3032	.2765	.2355	.1852	.1318	.0819	.0415	.0146	.0021
9	4	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344	.2780	.3110	.3280	.3241	.2966	.2458	.1762	.0984	.0305
	5	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0937	.1359	.1866	.2437	.3025	.3560	.3932	.3993	.3543	.2321
	6	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156	.0277	.0467	.0754	.1176	.1780	.2621	.3771	.5314	.7351
10	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078	.0037	.0016	.0006	.0002	.0001	.0000	.0000	.0000	.0000
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547	.0320	.0172	.0084	.0036	.0013	.0004	.0001	.0000	.0000
	2	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641	.1172	.0774	.0466	.0250	.0115	.0043	.0012	.0002	.0000
11	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734	.2388	.1935	.1442	.0972	.0577	.0287	.0109	.0026	.0002
	4	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734	.2918	.2903	.2679	.2269	.1730	.1147	.0617	.0230	.0036
	5	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641	.2140	.2613	.2985	.3177	.3115	.2753	.2097	.1240	.0406
12	6	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547	.0872	.1306	.1848	.2471	.3115	.3670	.3960	.3720	.2573
	7	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078	.0152	.0280	.0490	.0824	.1335	.2097	.3206	.4783	.6983
	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039	.0017	.0007	.0002	.0001	.0000	.0000	.0000	.0000	.0000
13	1	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0313	.0164	.0079	.0033	.0012	.0004	.0001	.0000	.0000	.0000
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094	.0703	.0413	.0217	.0100	.0038	.0011	.0002	.0000	.0000

TABLA B-1 (Continuación)

		<i>p</i>																		
<i>n</i>	<i>x</i>	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
	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	.8996	.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	.0000
	1	.2986	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176	.0083	.0035	.0013	.0004	.0001	.0000	.0000	.0000	.0000
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703	.0407	.0212	.0098	.0039	.0012	.0003	.0000	.0000	.0000
	3	.0077	.0446	.1069	.1762	.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	.0000
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098	.0042	.0016	.0005	.0001	.0000	.0000	.0000	.0000	.0000
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439	.0229	.0106	.0043	.0014	.0004	.0001	.0000	.0000	.0000
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172	.0746	.0425	.0212	.0090	.0031	.0008	.0001	.0000	.0000
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051	.1596	.1115	.0689	.0368	.0162	.0055	.0012	.0001	.0000
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461	.2340	.2007	.1536	.1029	.0584	.0264	.0085	.0015	.0001
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051	.2384	.2508	.2377	.2001	.1460	.0881	.0401	.0112	.0010
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172	.1665	.2150	.2522	.2668	.2503	.2013	.1298	.0574	.0105
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439	.0763	.1209	.1757	.2335	.2816	.3020	.2759	.1937	.0746
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098	.0207	.0403	.0725	.1211	.1877	.2684	.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	.0000
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054	.0021	.0007	.0002	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269	.0126	.0052	.0018	.0005	.0001	.0000	.0000	.0000	.0000
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806	.0462	.0234	.0102	.0037	.0011	.0002	.0000	.0000	.0000
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611	.1128	.0701	.0379	.0173	.0064	.0017	.0003	.0000	.0000
	5	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256	.1931	.1471	.0985	.0566	.0268	.0097	.0023	.0003	.0000
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256	.2360	.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	.0000	.0001	.0005	.0018	.0052	.0126	.0269	.0513	.0887	.1395	.1998	.2581	.2953	.2866	.2131	.0867
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054	.0125	.0266	.0518	.0932	.1549	.2362	.3248	.3835	.3293
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005	.0014	.0036	.0088	.0198	.0422	.0859	.1673	.3138	.5688
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	1	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029	.0010	.0003	.0001	.0000	.0000	.0000	.0000	.0000	.0000
	2	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161	.0068	.0025	.0008	.0002	.0000	.0000	.0000	.0000	.0000
	3	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537	.0277	.0125	.0048	.0015	.0004	.0001	.0000	.0000	.0000
	4	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208	.0762	.0420	.0199	.0078	.0024	.0005	.0001	.0000	.0000
	5	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934	.1489	.1009	.0591	.0291	.0115	.0033	.0006	.0000	.0000
	6	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256	.2124	.1766	.1281	.0792	.0401	.0155	.0040	.0005	.0000
	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	.0000	.0002	.0008	.0025	.0068	.0161	.0339	.0639	.1088	.1678	.2323	.2835	.2924	.2301	.0988
	11	.0000	.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	.0000	.0001	.0002	.0008	.0022	.0057	.0138	.0317	.0687	.1422	.2824	.5404

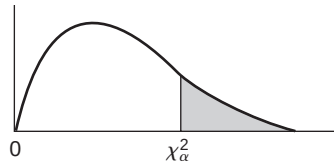
Fuente: Tabla A, páginas 464-466 de *Business Statistics: Concepts and Applications*, de William J. Stevenson. Derechos reservados © 1978 por William J. Stevenson. Reimpreso con permiso de Harper & Row, Publishers, Inc.

TABLA B-3 Valores críticos de t 

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
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	2.529	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", Maxine Merrington, *Biometrika*, vol. 32(1941), p. 300.
Reproducido con permiso del profesor D. V. Lindley

TABLA B-4 Valores críticos de chi cuadrada

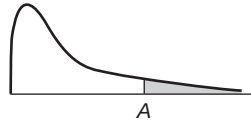


df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$
1	0.0000393	0.0001571	0.0009821	0.0039321	0.0157908
2	0.0100251	0.0201007	0.0506356	0.102587	0.210720
3	0.0717212	0.114832	0.215795	0.351846	0.584375
4	0.206990	0.297110	0.484419	0.710721	1.063623
5	0.411740	0.554300	0.831211	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 B-4 (Continuación)

<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 χ^2 -Distribution", *Biometrika*, vol. 32(1941), pp. 188-189, de Catherine M. Thompson. Reproducido con permiso del profesor D. V. Lindley.

TABLA B-5 Distribución F 

Por ejemplo, el valor F para $\delta_1 = 3$, $\delta_2 = 10$, correspondiente al área .01 en la cola derecha es de 6.55.

El valor F correspondiente al área .05 en la cola derecha aparece en caracteres normales.

El valor F correspondiente al área .01 en la cola derecha aparece en negritas.

Δ_2 , Grados de libertad del denominador	Δ_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
	4,052	4,999	5,403	5,625	5,764	5,859	5,928	5,981	6,022	6,056
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39
	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78
	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74
	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.29	10.15	10.05
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63
	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34
	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13
	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97
	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86
	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76
	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60
	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55
	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	4.49	3.36	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45
	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
	8.28	6.01	5.09	4.58	4.25	4.01	3.85	3.71	3.60	3.51

TABLA B-5 (Continuación)

Δ_2 , Grados de libertad del denominador	Δ_1 , Grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	10
19	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.55 3.77	2.48 3.63	2.43 3.52	2.38 3.43
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.52 3.71	2.45 3.56	2.40 3.45	2.35 3.37
21	4.32 8.02	3.47 5.78	3.07 4.87	2.84 4.37	2.68 4.04	2.57 3.81	2.49 3.65	2.42 3.51	2.37 3.40	2.32 3.31
22	4.30 7.94	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.47 3.59	2.40 3.45	2.35 3.35	2.30 3.26
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.45 3.54	2.38 3.41	2.32 3.30	2.28 3.21
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.43 3.50	2.36 3.36	2.30 3.25	2.26 3.17
25	4.24 7.77	3.38 5.57	2.99 4.68	2.76 4.18	2.60 3.86	2.49 3.63	2.41 3.46	2.34 3.32	2.28 3.21	2.24 3.13

Fuente: Abreviado con permiso de *Statistical Methods*, 7a. ed., de George W. Snedecor y William C. Cochran. Derechos reservados © 1980 por Iowa State University Press, Ames, Iowa.

TABLA B-6 Límite de la prueba de Durbin-Watson

n	Nivel de significancia $\alpha = .05$									
	k = 1		k = 2		k = 3		k = 4		k = 5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
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
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79

TABLA B-6 (Continuación)

<i>n</i>	<i>Nivel de significancia $\alpha = .05$</i>									
	<i>k = 1</i>		<i>k = 2</i>		<i>k = 3</i>		<i>k = 4</i>		<i>k = 5</i>	
	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>
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>n</i>	<i>Nivel de significancia $\alpha = .01$</i>									
	<i>k = 1</i>		<i>k = 2</i>		<i>k = 3</i>		<i>k = 4</i>		<i>k = 5</i>	
	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</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 B-6 (Continuación)

<i>n</i>	<i>Nivel de significancia $\alpha = 01$</i>									
	<i>k = 1</i>		<i>k = 2</i>		<i>k = 3</i>		<i>k = 4</i>		<i>k = 5</i>	
	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>	<i>d_L</i>	<i>d_U</i>
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: Reimpreso con permiso de J. Durbin y G. S. Watson, "Testing for Serial Correlation in Least Squares Regresión-II", *Biometrika*, vol. 38(1951), pp. 159-178.

Apéndice C

Conjunto de datos y bases de datos

Este apéndice contiene conjuntos de datos de 12 series de tiempo y tres bases de datos de regresión múltiple. La primera base de datos incluye 7 variables para 25 sitios para acampar, la segunda base de datos contiene 9 variables financieras medidas en cada una de 266 corporaciones estadounidenses en un año reciente,¹ y la tercera base de datos incluye 13 variables relacionadas con las remuneraciones de directores ejecutivos en 50 compañías estadounidenses en un año reciente.

Los conjuntos de datos de la mayoría de los ejemplos, problemas y casos, están disponibles en Internet. Para tener acceso a estos datos en Internet, entre al sitio de Prentice Hall, www.prenhall.com/hanke, y siga las instrucciones de descarga.

Conjuntos de datos

<i>Año</i>	<i>Población total</i>	<i>Poder de compra del dólar, precios al consumidor</i>	<i>Tarifa federal del salario mínimo, dólares corrientes</i>
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

Conjuntos de datos (Cont.)

<i>Año</i>	<i>Población total</i>	<i>Poder de compra del dólar, precios al consumidor</i>	<i>Tarifa federal del salario mínimo, dólares corrientes</i>
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	

<i>Año</i>	<i>Ingresos de Kodak Eastman</i>	<i>Depósitos en mercado de dinero</i>	<i>Suscriptores de televisión por cable</i>
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

¹Estamos en deuda con el doctor Lynn Stephens de Eastern Washington University por brindarnos estos datos.

Principales electrodomésticos, embarques de la industria, congeladores (miles)

<i>Año</i>	<i>1993</i>	<i>1994</i>	<i>1995</i>	<i>1996</i>	<i>1997</i>	<i>1998</i>	<i>1999</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</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>1993</i>	<i>1994</i>	<i>1995</i>	<i>1996</i>	<i>1997</i>	<i>1998</i>	<i>1999</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</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	

Suscripciones nuevas (miles)

<i>Año</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</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

Pasajeros en autobuses cada día, Seattle, Washington, durante 146 días en la ruta de Pike Street (léase horizontalmente)

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														

Estadísticas de ocupación mensual del motel Nine

<i>Año</i>	<i>1997</i>	<i>1998</i>	<i>1999</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
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)

<i>Año</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
Enero	5,276	5,466	5,480	5,579	5,681
Febrero	4,919	4,821	4,988	5,033	5,254
Marzo	5,227	5,307	5,272	5,319	5,525
Abril	5,003	5,193	5,254	5,285	5,489
Mayo	5,004	5,087	5,156	5,188	5,320
Junio	4,949	5,102	5,175	5,197	5,417
Julio	5,219	5,399	5,490	5,604	5,711
Agosto	5,265	5,301	5,366	5,357	5,419
Septiembre	5,007	5,063	5,200	5,133	5,318
Octubre	5,079	5,295	5,410	5,483	5,409
Noviembre	4,974	5,021	5,123	5,186	5,310
Diciembre	5,282	4,991	5,428	5,472	5,458

<i>Año</i>	<i>Índice de precios de educación superior, 1983 = 100</i>	<i>Número de granjas en Estados Unidos (en miles)</i>
1970	39.5	
	42.1	
	44.3	
	46.7	
	49.9	
1975	54.3	2,521
	57.9	2,497
	61.7	2,456
	65.8	2,436
	70.6	2,437
1980	77.5	2,440
	85.9	2,440
	94.0	2,407
	100.0	2,379
	104.7	2,334
1985	110.5	2,293
	115.6	2,250
	120.3	2,213
	125.8	2,197
	133.1	2,171
1990	140.8	2,140
	148.3	2,105
	153.1	2,094
	158.2	2,040

<i>Año</i>	<i>Precio promedio del gas natural (centavos por millón de pies cúbicos)</i>	<i>Precio promedio anual de la acciones ordinarias de aparatos electrodomésticos según la empresa calificadora Moody's</i>
1972	18.6	80.20
	21.6	71.21
	30.4	48.26
1975	44.5	51.25
	58.0	60.10
	79.0	67.55
	90.5	63.54
	117.8	60.28
1980	158.8	54.80
	198.2	55.41
	245.7	63.56
	259.3	74.04
	265.5	71.16
1985	248.0	87.24
	194.2	111.11
	166.5	105.90
	169.0	97.99
	169.0	110.45
1990	172.0	112.61
	164.0	126.97
	174.0	
	203.0	

Miles de cargas de un vagón de fletes, productos forestales

Año	2002	2003	2004	2005	2006
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

Industrial Price/Earnings Ratio

Año	Trimestre	P/G	Año	Trimestre	P/G
1999	1	17.56	2003	1	15.80
	2	18.09		2	17.00
	3	16.47		3	14.80
	4	18.73		4	15.90
2000	1	22.16	2004	1	18.80
	2	21.80		2	20.40
	3	20.82		3	23.70
	4	14.09		4	29.20
2001	1	13.61	2005	1	28.40
	2	13.29		2	26.80
	3	12.50		3	26.00
	4	12.35		4	26.00
2002	1	12.46			
	2	13.30			
	3	14.73			
	4	15.40			

BASES DE DATOS

Sitios para acampar

Un grupo de empresarios de Spokane está planeando desarrollar varios sitios para acampar en el estado. Uno de sus problemas es determinar la tarifa que se cobrará diariamente. De acuerdo con sus observaciones, la tarifa depende de muchas variables, por ejemplo, el tamaño del espacio para acampar o si hay alberca. En un intento por ser objetivos, se recopiló la siguiente información de 25 campamentos de Washington (según *Rand McNally Campground Guide for Washington*) y se realizó un análisis en computadora. Las variables que se analizaron son

Y = Tarifa diaria (FEE)

X_1 = Número de acres (ACRES)

X_2 = Número de sitios (SITES)

X_3 = Si existen o no baños con agua (TOILET)

X_4 = Si existe o no alberca (POOL)

X_5 = Si existen o no instalaciones eléctricas (HOOKUP)

X_6 = Número adicional de instalaciones recreativas (ADD)

<i>Sitio</i>	<i>Y</i>	<i>X₁</i>	<i>X₂</i>	<i>X₃</i>	<i>X₄</i>	<i>X₅</i>	<i>X₆</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

Datos financieros de las compañías estadounidenses seleccionadas

<i>Número de compañía</i>	<i>Ventas (millones de \$)</i>	<i>Empleados (en miles)</i>	<i>Gastos de capital (millones de \$)</i>	<i>Gastos intangibles (millones de \$)</i>	<i>Costo de bienes vendidos (millones de \$)</i>	<i>Gastos de mano de obra y rela- cionados (millones de \$)</i>	<i>Gastos de publicidad (millones de \$)</i>	<i>Gastos de investigación y desarrollo (millones de \$)</i>
1	3,221.800842.0000		147.9000	30.6000	2,285.2007	599.7998	118.3000	28.0000
2	1,690.600120.9050		93.0000	29.1000	1,057.2002	343.2000	114.9000	8.9000
3	2,197.276439.0000		66.8670	55.8600	1,387.0679	661.3997	95.5680	11.1820
4	2,357.820623.3000		59.5560	69.6080	1,743.7952	25.6320	51.9170	8.5000
5	8,129.000035.0000		297.0000	29.0000	7,423.0000	1,178.0000	12.8000	9.2530
6	11,851.000023.0000		394.0000	20.0000	10,942.0000	2,556.0000	11.6530	14.6000
7	323.86063.9000		2.5900	4.2880	233.5300	22.8350	3.5290	30.7320
8	660.48568.3780		10.9840	3.3720	582.2649	25.6250	44.9990	64.8730
9	4,351.160150.9120		102.7080	217.0920	4,156.8671	12.8360	66.2640	8.7790
10	985.83575.5000		16.6010	29.5900	874.1287	19.5000	112.3860	18.3650
11	3,802.558139.6000		206.1020	157.3520	2,997.2703	518.0000	139.7290	16.4130
12	2,576.046422.6000		50.6690	47.0790	1,885.9053	349.4910	48.8170	9.5000
13	106.016028.0000		1.3120	42.0000	84.6590	35.5550	22.9370	8.7330
14	5,669.894546.8810		103.0000	31.1000	4,424.3007	785.0000	141.3000	18.5000
15	319.65702.8940		4.5770	2.2090	246.6980	42.8370	87.0000	1.1000
16	511.721710.1000		19.5600	27.0000	286.2288	48.9990	1.8700	23.6520
17	884.618922.8010		58.0940	33.0000	467.4436	36.5000	16.0350	29.6320
18	166.37502.3000		3.9510	5.2890	111.0310	31.0000	4.0230	38.5420
19	59.131018.0000		1.1400	14.5000	43.7430	26.3210	90.3250	56.9820
20	136.69703.1000		2.0090	18.4930	105.3300	15.8880	46.3000	8.6330
21	767.87998.1000		37.4250	18.0560	519.3948	112.1350	21.8470	2.7860
22	61.32801.1390		1.3880	26.3250	35.2020	17.3140	2.4270	88.5230
23	445.63875.8000		18.9780	12.6000	213.2880	12.1000	62.8060	1.4600

Datos financieros de las compañías estadounidenses seleccionadas (Cont.)

<i>Número de compañía</i>	<i>Ventas (millones de \$)</i>	<i>Empleados (en miles)</i>	<i>Gastos de capital (millones de \$)</i>	<i>Gastos intangibles (millones de \$)</i>	<i>Costo de bienes vendidos (millones de \$)</i>	<i>Gastos de mano de obra y rela- cionados (millones de \$)</i>	<i>Gastos de publicidad (millones de \$)</i>	<i>Gastos de investigación y desarrollo (millones de \$)</i>
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	271.0080	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

Datos financieros de las compañías estadounidenses seleccionadas (Cont.)

<i>Número de compañía</i>	<i>Ventas (millones de \$)</i>	<i>Empleados (en miles)</i>	<i>Gastos de capital (millones de \$)</i>	<i>Gastos intangibles (millones de \$)</i>	<i>Costo de bienes vendidos (millones de \$)</i>	<i>Gastos de mano de obra y rela- cionados (millones de \$)</i>	<i>Gastos de publicidad (millones de \$)</i>	<i>Gastos de investigación y desarrollo (millones de \$)</i>
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

Datos financieros de las compañías estadounidenses seleccionadas (Cont.)

<i>Número de compañía</i>	<i>Ventas (millones de \$)</i>	<i>Empleados (en miles)</i>	<i>Gastos de capital (millones de \$)</i>	<i>Gastos intangibles (millones de \$)</i>	<i>Costo de bienes vendidos (millones de \$)</i>	<i>Gastos de mano de obra y rela- cionados (millones de \$)</i>	<i>Gastos de publicidad (millones de \$)</i>	<i>Gastos de investigación y desarrollo (millones de \$)</i>
124	1,167.83402.8100		48.6920	8.4580	1,017.60382.3500		7.0620	59.2350
125	12,298.398050.7000		1,221.8008	10.4000	9,285.71091,016.5000		13.1160	64.6000
126	439.47271.9020		65.1100	39.8650	263.810851.1480		27.3460	31.2730
127	29,127.0030108.7000		1,897.0005	9.9040	20,032.000078.7700		16.5000	86.0000
128	1,993.66248.0000		43.4190	45.7820	1,755.56623.5730		31.1370	43.0750
129	4,660.894518.1000		636.1238	28.4900	3,675.6895440.7996		3.4000	11.6000
130	976.45788.8280		14.8590	55.2940	879.351691.8000		15.3440	90.2000
131	3,834.93246.6610		316.7156	68.2690	3,557.47347.4050		2.8250	69.8820
132	9,535.738242.7800		1,107.3838	75.7000	7,075.1875971.0000		6.5480	29.7730
133	657.77761.2640		56.1460	36.8860	565.017614.4700		27.0030	4.2120
134	100.457043.0750		44.0680	7.1610	72.783022.0310		49.6110	83.1580
135	60,334.5110130.0000		4,186.9296	40.6150	45,999.00703,405.0000		32.6000	290.0000
136	2,150.000090.2110		311.7000	91.2150	1,460.799657.4030		1.3040	25.1000
137	18,069.000058.3000		1,680.0000	74.5900	13,442.00001,345.0000		25.4000	88.8250
138	109.738069.8870		32.2560	36.9560	97.01302.5200		4.9030	6.7300
139	592.77103.2520		123.7680	3.8770	420.320667.3300		28.7000	1.4590
140	4,642.394514.3280		353.5999	33.5620	4,085.0989324.0000		40.7000	25.0000
141	2,072.441211.1480		270.1846	42.1160	1,640.81181.2400		5.5500	4.9810
142	4,509.382813.3540		502.2720	1.6500	2,838.0845236.4540		2.0370	12.8000
143	34,736.0030207.7000		1,760.7100	2.7200	26,053.906020.9400		3.7000	16.7160
144	1,191.03374.2070		255.6150	1.9000	865.647782.6730		.2670	99.8430
145	312.73004.2120		76.5000	6.4200	452.413017.0050		17.9000	52.1780
146	1,553.10779.1500		343.9539	23.6410	988.8760185.6600		12.5840	9.7580
147	6,997.773430.0080		956.1719	11.2330	4,886.8125720.5000		17.1000	58.2460
148	513.18805.1420		41.9800	41.9800	375.359925.0200		11.3330	11.9490
149	28,085.003094.8000		2,913.0000	32.5600	20,632.00002,344.0000		89.0000	231.0000
150	11,062.898034.9740		1,774.3904	43.0250	8,259.76561,051.0000		16.5890	114.0000
151	23,232.406037.5750		1,049.6729	90.2110	19,964.6050994.0000		37.0000	89.7370
152	14,961.500047.0110		1,744.0364	69.8870	10,046.00001,126.7310		19.9200	80.3250
153	5,197.707024.1450		762.2510	4.2120	3,336.7566431.9976		7.7130	15.0520
154	7,428.234333.7210		601.1216	6.7310	5,714.30859.7320		58.0000	21.0000
155	28,607.503067.8410		1,344.3777	10.4000	24,787.60501,572.7996		56.6000	52.0000
156	87.61006.7310		12.7120	39.8650	74.551031.5580		31.0000	44.6610
157	1,165.67363.5310		26.6780	9.9040	1,035.71296.6000		18.5170	2.4490
158	567.36501.5420		97.4910	28.9400	480.511023.5230		3.300	52.3290
159	5,954.941416.2970		732.0000	55.9240	4,540.4609444.8997		31.7000	18.5000
160	368.09402.3150		15.0860	2.7160	319.493910.6050		3.4800	6.2850
161	751.73276.2550		51.1060	13.5380	606.83183.5230		6.8000	9.9000
162	895.408710.9000		145.5140	9.3840	681.965626.3250		39.0000	30.6000
163	1,063.290816.1790		51.1480	25.7670	746.282012.6000		16.6980	14.6320
164	1,306.086719.3970		78.7700	2.7490	1,021.4856435.2998		23.3000	13.2830
165	140.44401.9190		3.5730	55.8600	122.321027.3350		35.0000	29.1000
166	4,357.281252.1400		110.4470	12.0830	3,540.96121,235.0000		3.9000	55.8600
167	263.90483.7000		7.4050	27.2080	203.34402.8080		8.3780	3.2500
168	6,184.894594.5000		398.2000	69.6080	5,224.00002,550.0000		50.9120	37.1000
169	257.65093.3640		14.4730	7.5700	190.41908.3710		5.5000	69.6080
170	50.515052.5350		29.1000	29.0000	18.056043.5920		39.6000	29.0000
171	419.64704.3020		22.0310	20.0000	341.5906135.6000		22.6000	20.0000
172	1,227.449020.0000		57.4030	4.2880	999.752027.8920		28.0000	9.0000
173	779.34508.8000		22.0670	3.3700	678.4258229.1270		46.8810	4.2880

Datos financieros de las compañías estadounidenses seleccionadas (Cont.)

<i>Número de compañía</i>	<i>Ventas (millones de \$)</i>	<i>Empleados (en miles)</i>	<i>Gastos de capital (millones de \$)</i>	<i>Gastos intangibles (millones de \$)</i>	<i>Costo de bienes vendidos (millones de \$)</i>	<i>Gastos de mano de obra y rela- cionados (millones de \$)</i>	<i>Gastos de publicidad (millones de \$)</i>	<i>Gastos de investigación y desarrollo (millones de \$)</i>
174	72.1760	1.30002.5210		29.5900	50.9650	24.8290	2.8940	3.3700
175	3,248.0076	36.0620263.6167		19.4460	2,710.3455	974.3379	10.1000	29.5900
176	921.1270	12.659067.3340		10.5250	771.0059	23.5420	22.8010	157.3520
177	711.9827	12.5120133.3850		45.0790	653.8069	351.4700	2.3000	45.0790
178	72.4110	1.02501.2400		42.0000	60.0820	93.0000	18.0000	42.0000
179	297.5686	4.152020.9420		.8990	248.7160	123.1000	3.1000	31.1000
180	677.8489	6.070017.0050		31.1000	613.3047	169.2570	8.1000	2.2090
181	582.6238	1.400025.0290		2.2090	474.3450	66.8670	1.1390	27.0000
182	3,750.4109	38.1700120.8280		27.0000	3,240.7886	1,132.6216	5.8000	33.0000
183	88.8070	2.33309.7320		7.2110	66.6540	59.5500	16.0270	5.2890
184	306.9397	2.800031.5880		33.0000	220.4980	2.5900	8.7000	14.5000
185	331.7366	5.20006.6000		11.0250	295.3848	10.9840	4.0000	18.4930
186	546.9500	8.900023.5230		5.2890	439.0479	16.0010	2.1000	18.0560
187	7.5910	30.60007.8900		14.5000	5.0480	50.6690	5.0000	26.3250
188	3,479.4573	41.3940170.3720		18.4930	3,100.5391	1,177.5999	1.5760	15.0500
189	485.6138	6.658058.6750		3.5250	335.3318	42.0000	93.0000	1.4320
190	123.2280	2.045010.6050		5.2550	96.6630	20.9000	66.8670	12.6000
191	488.2327	4.650020.4800		1.1111	402.8457	402.8400	77.0101	22.2426
192	100.7820	1.70302.4430		1.6800	88.7960	4.0000	21.0000	28.3032
193	165.7970	4.76603.2790		88.0003	120.1080	2.0000	4.0008	18.2022
194	274.8440	3.550021.7900		2.9530	213.1860	3.0000	3.7521	24.2628
195	11,049.5000	166.8480667.7998		55.5000	9,955.3984	4,485.1953	22.0007	52.5000
196	1,154.8477	14.419032.2360		4.0800	1,037.4727	424.4556	21.1234	30.3234
197	578.7107	11.492026.3000		8.0141	433.8230	1.0111	12.3456	5.3300
198	124.5440	1.80004.6280		1.9850	101.5300	23.6630	78.9101	36.3840
199	3,711.2029	63.4000303.3838		4.5720	2,729.9280	22.0222	91.0111	33.0000
200	124.8600	2.00005.2240		2.3200	79.7770	51.0000	21.3141	2.6500
201	2,466.0000	26.8650161.7000		2.0202	2,028.7996	18.4021	3.2000	14.9000
202	2,829.2991	36.2000156.8000		27.1000	2,261.0000	930.2000	51.1617	25.2000
203	814.8196	14.800048.5520		16.0111	622.9507	204.9000	18.1920	1.4150
204	4,051.7996	46.0000349.5999		2.6000	3,036.5999	1,215.2996	21.2223	56.6000
205	67.0390	28.00003.5010		2.5170	54.9070	66.5620	24.2526	42.4446
206	240.5670	4.00005.5670		1.3220	184.1350	61.6900	2.5860	3.0470
207	45.2140	2.00001.4110		18.1010	38.0970	62.3201	27.2829	48.5052
208	69.9520	81.000033.3333		8.0033	65.4570	52.3302	30.3132	54.5658
209	54.5490	1.12701.7720		17.7200	42.5990	42.4444	33.3435	60.6264
210	317.4480	5.784012.6650		11.0330	254.1990	80.1010	36.3738	66.6870
211	847.9927	24.000085.0240		19.7930	664.9578	34.1021	39.4041	10.4000
212	467.9546	4.845013.1650		2.3810	400.5806	4.0999	42.4344	1.0011
213	126.6750	14.00077.7490		14.1441	109.6830	50.6410	45.4647	1.0022
214	85.7290	49.00002.1610		49.4949	72.8400	9.9901	48.4950	1.0033
215	680.7666	8.220019.2340		77.7878	578.8528	9.8175	51.5253	1.3090
216	211.3230	1.56704.8350		15.6180	171.4130	65.0000	54.5556	1.8201
217	254.3030	3.10002.7620		2.3570	205.8410	42.4381	57.5859	2.0880
218	1,396.8108	29.416079.9820		28.2626	1,000.2886	3.8107	16.1580	3.4510
219	3,981.0000	52.9000188.3000		70.3000	3,120.5999	1,085.7996	75.8000	37.5000
220	3,943.0990	56.5320259.5000		49.9000	3,352.3008	1,275.7002	60.6162	42.3000

Datos financieros de las compañías estadounidenses seleccionadas (Cont.)

<i>Número de compañía</i>	<i>Ventas (millones de \$)</i>	<i>Empleados (en miles)</i>	<i>Gastos de capital (millones de \$)</i>	<i>Gastos intangibles (millones de \$)</i>	<i>Costo de bienes vendidos (millones de \$)</i>	<i>Gastos de mano de obra y rela- cionados (millones de \$)</i>	<i>Gastos de publicidad (millones de \$)</i>	<i>Gastos de investigación y desarrollo (millones de \$)</i>
221	1,260.2349	17.2880	103.0320	11.4810	1,055.943612.0000		63.6465	2.1133
222	973.2527	9.8850	25.4530	5.5580	848.72274.0877		66.6768	3.3210
223	19.9060	18.0002	5.6666	1.4100	16.51703.3876		69.7071	4.2242
224	66.8260	1.3200	6.1110	88.1388	48.94804.5222		72.7374	5.6210
225	178.7460	2.1980	5.5430	138.0000	138.569043.4350		75.7677	6.2155
226	26.7510	1.0560	8.8888	211.0113	17.993018.1111		78.7980	7.2102
227	20.5750	43.1111	7.7777	82.1003	13.972014.2222		81.8283	8.9712
228	51.5960	18.5216	1.6940	1.1620	38.819088.9922		81.0077	24.2601
229	106.1150	2.6000	4.6850	9.9210	64.050012.4243		77.0222	23.2810
230	8.5160	14.2421	12.0818	12.1402	5.95007.8246		22.4443	24.8588
231	308.8916	5.7000	15.8370	13.1402	144.734042.4444		47.8877	2.7060
232	753.8069	16.8750	37.4620	3.6210	491.1160210.0050		16.4370	4.9340
233	41.2960	1.1080	2.5820	12.1213	28.132081.8118		12.5456	24.5293
234	145.6840	3.4190	13.3250	1.0087	105.163051.7100		51.8196	1.8480
235	51.3130	1.0000	1.5700	8.0025	35.973043.4400		21.4231	59.6085
236	21.4070	12.5358	18.7842	5.5554	12.955012.8888		37.8286	64.8821
237	585.6597	8.2000	56.0530	80.9960	359.835077.9999		13.6920	8.9610
238	516.7239	10.3000	17.9320	9.3610	376.41701.1007		5.6670	5.6000
239	316.8147	7.0000	3.9360	12.1314	267.24562.0008		86.8686	76.7686
240	509.7000	10.0000	27.0360	15.1617	375.3457179.9240		85.8686	3.6080
241	341.3887	7.1270	7.1570	8.1819	287.69079.0007		86.8888	86.7795
242	33.0660	1.0872	1.9540	9.2021	24.072012.7210		83.1111	95.9594
243	200.5920	4.0000	5.3320	20.0290	153.54807.6660		82.2222	94.9388
244	184.5810	4.0500	7.2780	10.3570	142.71608.7770		22.6665	1.0790
245	217.7520	4.0880	7.3840	10.1146	179.102078.3910		44.6621	89.9012
246	386.8118	7.4040	18.4880	47.1213	302.55862.9990		18.1716	3.8620
247	69.1530	12.1212	1.6190	48.1415	54.431011.3410		15.1413	13.8125
248	81.4650	1.6220	4.1890	16.4950	70.50804.4555		12.1110	47.8552
249	329.5518	6.0040	12.2520	8.0540	269.637712.1417		9.8765	51.9185
250	36.3870	133.0000	12.7246	51.5355	27.769021.8283		4.3210	54.3321
251	344.7937	7.5000	24.7400	57.5982	205.061092.9395		8.1234	4.8200
252	22.8030	84.1000	2.1060	83.4952	10.683096.9899		5.6788	43.8388
253	196.3030	5.4660	5.9730	99.9242	142.152097.9294		12.4582	2.2710
254	31.5660	13.7124	8.1264	10.1115	22.375095.0092		14.5220	66.7274
255	108.8580	1.7000	1.2870	92.4445	45.913092.6666		1.4330	53.5422
256	83.6260	1.2320	4.1220	55.6677	45.095092.5555		13.5620	22.5673
257	390.8726	6.1660	17.3310	40.5880	296.857758.2130		18.0000	10.0000
258	363.9839	7.0160	11.2700	11.5610	234.63202.7310		6.4860	86.0000
259	52.2620	.4420	5.1030	1.1500	43.511012.1440		44.0700	16.0000
260	228.6110	5.6500	1.8370	41.5600	161.470020.5400		14.9460	3.0730
261	60.8250	1.5000	1.4910	45.3100	41.68201.8970		1.6590	30.7300
262	16.6890	40.5000	57.6000	9.8540	9.845018.3700		35.2020	63.5300
263	39.8290	62.1000	3.9900	1.5610	32.65804.9080		9.2530	30.0000
264	28.9020	93.4000	1.1040	36.5000	23.14102.5900		18.3650	10.0000
265	8.7410	27.0000	55.6000	32.1400	6.370059.3100		27.6600	56.6660
266	61.9446	7.0000	35.0470	43.2000	432.3777160.6660		19.2020	6.3530

muneración de directores ejecutivos (CEO)

bonifica- ciones (miles de \$)	Otras remune- raciones (miles de \$)	Total de remune- raciones (miles de \$)	Edad	Educación*	Experiencia profesional	Titula- ridad (años)	Experiencia (años)	Valoración de mercado (millones de \$)	Porcentaje de partici- pación en la propiedad	Utilidades de la empresa (millones de \$)	Ventas de la empresa (millones de \$s)
275	5	453	64	2	3	40	26	54.5	3.14	91	872
429	78	1,948	55	1	1	23	23	7.6	0.55	145	1,237
0	89	1,735	47	2	7	5	5	21.7	0.52	-47	1,712
325	24	643	65	1	3	29	23	8.9	0.89	44	1,681
105	102	1,461	63	1	6	23	8	3.6	0.05	201	5,673
25	7	357	54	2	5	20	1	0.5	0.06	71	1,117
0	11	669	61	2	5	2	2	0.7	0.05	-187	1,475
289	82	2,094	63	1	3	41	8	5.9	0.04	1,166	10,818
69	24	597	57	2	5	27	13	1.4	0.03	377	2,686
38	29	889	56	2	1	5	5	0.7	0.03	224	2,201
129	11	514	57	2	4	3	3	4.1	0.17	79	661
11	8	466	48	2	1	17	1	0.2	0.01	189	1,539
0	52	2,833	50	1	3	4	4	11.7	0.39	-332	11,663
282	17	427	54	1	8	31	15	71.4	10.09	55	2,366
0	74	1,856	60	2	7	33	3	0.4	0.03	-507	4,864
423	92	1,652	60	2	1	34	14	11.5	0.06	856	14,949
20	1	782	49	2	9	18	9	1.3	0.07	14	5,061
0	108	613	56	2	9	8	1	0.1	0.02	-29	1,929
448	64	1,488	58	1	3	9	8	9.4	0.56	126	2,643
12	1	447	50	2	9	5	1	0.3	0.03	54	1,084
687	55	1,752	63	1	7	14	14	534.2	3.99	249	5,137
1,452	89	2,497	64	0	1	28	28	221.1	3.98	91	844
37	31	768	60	2	9	30	8	0.7	0.02	322	2,097
489	40	2,342	71	1	4	46	34	9.6	0.83	99	835
0	13	3,409	64	0	9	30	30	29.4	0.31	-99	12,021

2,108	38	98	2,244	64	2	5	41	5	4.0	0.04	30	4,451
597	0	4	601	59	2	4	35	5	0.1	0.12	-85	1,911
616	862	76	1,554	61	1	8	41	17	30.6	2.23	82	1,435
237	221	4	462	61	2	5	25	11	16.8	1.03	27	1,314
571	0	16	587	55	2	4	5	5	1.6	0.17	-76	2,301
269	391	28	688	54	2	9	28	28	1,689.0	34.04	317	3,277
721	101	71	893	60	2	5	36	15	2.0	0.04	417	4,444
328	238	34	600	60	0	7	42	1	85.6	17.66	43	1,214
538	25	7	570	60	1	4	3	3	0.2	0.21	49	804
741	104	9	854	62	1	7	30	3	2.6	0.17	81	669
607	380	47	1,034	51	1	6	23	3	7.0	0.83	82	578
1,044	107	36	1,187	55	2	8	2	1	3.1	1.21	10	1,214
2,409	1,487	143	4,039	55	1	7	32	17	35.2	0.29	715	12,651
287	198	32	517	59	1	8	37	21	181.0	6.70	136	3,180
567	15	34	616	51	2	1	16	9	0.3	0.01	237	2,754
682	0	2	684	62	2	1	36	2	0.8	0.01	-1,086	12,794
1,226	174	56	1,456	52	1	4	30	10	6.0	0.17	98	4,439
952	80	56	1,088	45	2	7	11	11	2.6	0.17	48	415
432	0	12	444	50	2	7	25	3	0.1	0.01	-50	1,569
1,085	440	97	1,622	57	1	8	29	5	9.7	0.19	347	9,886
1,009	117	0	1,126	64	1	3	28	14	3.4	0.21	63	2,545
1,711	182	134	2,027	62	2	2	25	5	3.4	0.04	806	8,379
408	183	9	600	52	2	7	12	2	4.2	0.26	10	21,351
543	13	2	558	62	2	5	34	12	0.2	0.01	265	2,359
278	209	74	561	61	2	6	24	24	15.4	0.95	52	695

**0 = sin título universitario; 1 = universitario no graduado; 2 = con título universitario

Índice

A

AAA Washington, 210-212, 271-274,
328-330, 389-391, 469-471
Ace Manufacturing, 269
Ajuste
de la curva S, 80
estacional, 167, 179-180, 193
Alcam Electronics, 53-54
Alomega Food Stores, 8-9, 56, 99, 212-213,
392-393, 497-498
Alpha. *Véase* Nivel de significancia
American Economic Association, 7
American Statistical Association, 7
Análisis
(en una serie de tiempo), 185
de las series de tiempo, 19, 62-64, 79,
80, 339-368, 399-446
cointegración en el, 365-367
correlación serial de la regresión y
correlación serial de primer
orden, 340-343
descomposición del, 166-185
diferencias generalizadas en el,
354-357
error de especificación en el modelo,
348-349
errores autocorrelacionados y,
354-358
modelo autorregresivo de primer
orden, 357-358, 404
modelo estacional, 361, 368, 332-342
modelo logarítmico de regresión
lineal y, 368
pronósticos de, 174-175, 185-187,
361-364, 411
pronósticos económicos y, 364-365
regresión con diferencias, 350-357
soluciones para la autocorrelación,
347-358
tendencia en el, 168-175
de regresión, 39-42, 79, 80, 221-254,
281-313, 339-368
aplicación a la administración, 251,
252, 310-311
autocorrelación y, 240, 254, 339-338
coeficiente de correlación y, 235, 289,
313
coeficiente de determinación en el,
234-236, 239, 254, 289-290, 312,
313
correlación serial y, 230

datos de series de tiempo, con,
339-368
descomposición de la varianza,
230-234, 253, 287, 312
ecuación de pronóstico puntual, 227,
291
ecuación de regresión de la
población, 224-225, 283, 312
error estándar del coeficiente de
regresión del, 237, 254, 291, 297
error estándar del pronóstico del,
228, 252, 253, 291
error estándar para la estimación
del, 226-227, 252, 253,
287-288, 312
estadístico F, 238, 239, 254, 289, 312
estadístico t, 237, 254, 291, 313
función de regresión ajustada, 46,
223, 252, 253, 284, 312
intervalo de predicción, 228, 253,
291, 313
línea de regresión de la muestra en,
226
método de mínimos cuadrados, 39,
46, 223, 253, 284
modelo de regresión lineal, 226
procedimiento de todas las
regresiones posibles en,
302-304
pronóstico Y, 227-230, 291, 313
prueba de hipótesis en, 236-238, 254,
289, 291, 312, 313
regresión lineal simple, 221-254
regresión lineal simple versus
regresión múltiple, 283
regresión por pasos y, 304-307
residuos en el, 239-241, 307-309
suma de cuadrados de los errores
(SSE), 231, 232, 234, 235, 253,
287, 289, 312
suma de los cuadrados de la
regresión (SSR), 231, 234, 235,
253, 287, 289, 312
suma total de cuadrados (SST), 231,
232, 234, 235, 253, 287, 289, 312
supuestos en el, 230, 239, 339, 489
transformaciones de variables y, 239,
243-246
variables explicativas en, 243, 281,
339
Anderson-Parente, 485
ANOVA (Análisis de varianza), 231

Aplicación(es)
a la administración, 44, 86, 135-136,
189-190, 250, 252, 310-311, 367,
444-445
del análisis de regresión simple, 250,
252
Armstrong, J. Scott, 486
Autocorrelación(es), 64-67, 87, 339-358,
367
correlación serial de primer orden y,
340-343
correlogramas y, 22, 67, 87
de k-ésimo orden, 65, 87
de segundo orden, 66
definición de, 64, 87
diferencias generalizadas y, 347,
354-357, 368
enfoque no lineal de mínimos
cuadrados y, 355-357
error estándar del coeficiente de
autocorrelación, 68, 87
fórmula del coeficiente de, 65, 87
fórmulas de, 65, 87
parcial, 400-403, 407
prueba de Durbin-Watson, 343-347,
367
regresión de series de tiempo y, 339-358
soluciones de, 347-358
técnica Box-Jenkins y, 400-407

B

“Basura- entra, basura sale”, 61
B&E Data Links, 7
Béisbol de fantasía (A), 330-334, (B),
334-336
Bell y Hillmer, 179
Bernstein, Peter, 1, 2, 444
Beta. *Véase* Prueba de hipótesis
Boletín de la Reserva Federal, 165
Boundary Electronics, 513-514
Buen ajuste, 37, 39
Burns, Arthur F., 184
Busby Associates, 514-517
Business Week, 310
Butcher Products, Inc., 268-269

C

Casos de estudio
AAA Washington, 210-212, 271-274,
328-330, 389-391, 469-471
Ace Manufacturing, 269

- Alcam Electronics, 53-54
 - Alomega Food Stores, 56, 99, 212-213, 392-393, 497-498, 519-520
 - Béisbol de fantasía (A), 330-334
 - Béisbol de fantasía (B), 334-336
 - Boundary Electronics, 513-514
 - Busby Associates, 514-517
 - Butcher Products Inc., 268-269
 - Ciudad College Station, 463-466
 - Compañía de su elección, 378
 - Consumer Credit Counseling (CCC), 10-11, 98-99, 148, 206-207, 270-271, 388-389, 460, 517-518
 - Índice de actividad en los negocios del condado de Spokane, 379-382
 - Lydia E. Pinkham Medicine Company, 461-463, 498-500
 - Mercado de bonos, 324-327
 - Minorista en la red, 154-157, 471-474
 - Mr. Tux, 10, 54-55, 97-98, 147, 202-206, 270, 385-387, 459-460, 518-519
 - Murphy Brothers Furniture, 94-96, 148, 207-209
 - Proyección de ingresos a cinco años del Downtown Radiology, 149-154
 - Restaurante Golden Garden, 497
 - Solar Alternative Company, 145-146
 - Southwest Medical Center, 158-159, 214, 394-395, 476-477, 520
 - Surtido Cookies, 100, 159, 213-214, 393-394, 474-476
 - The Small Engine Doctor, 201-202
 - Tiger Transport, 266-267
 - UPS Air Finance Division, 466-469
 - Ventas en un restaurante, 383-385, 457-459
 - Causalidad contra correlación, 38-39
 - Censo II, 187-189
 - Chi cuadrada, 69, 527-528
 - Cíclico-irregular, 180, 193
 - Ciclo de vida del producto, 171-172
 - Ciudad de College Station, 463-466
 - Cleman, Robert, 487
 - Coefficiente
 - de autocorrelación. *Véase* Coeficiente de autocorrelación, 37-39, 40, 44, 46, 65, 87, 235, 289
 - coeficiente de determinación y, 235
 - fórmula del, 38, 46
 - de determinación (en la regresión lineal simple), 234-236, 239, 252
 - coeficiente de correlación versus, 235
 - fórmula del, 235, 254
 - relación con el estadístico F, 239, 254
 - parciales o globales de regresión, 285, 286, 312
 - Colinealidad, 281-282, 297-300, 312
 - Combinación de pronósticos, 486-488
 - beneficios de la, 488
 - promedio simple, 488, 496
 - promedio ponderado, 488, 496
 - Compatibilidad, 489
 - Competencia M3-IJF, 81
 - componente(s)
 - cíclico, 63, 79, 87
 - estacional de las series de tiempo, 76, 87, 166-167, 175-180
 - análisis de las series de tiempo y, 63-64, 110, 166-167
 - promedio móvil y, 116
 - suavización exponencial ajustada del, 130-135
 - técnica Box-Jenkins y el, 432-442
 - modelo aditivo de los, 167, 176, 179, 187, 192
 - irregular (en el análisis de las series de tiempo), 167, 180
 - Computadoras
 - acumulación de datos y, 61
 - cambio tecnológico, 1, 80, 495, 511-512
 - consideraciones de costo, 80, 511
 - organización de la función de pronóstico y, 80, 511
 - Confiabilidad (de los datos), 61
 - Construcción del modelo, 5, 407-411
 - Continuo micro-macro, 3
 - Control, 124-126
 - Consultores, pronósticos de, 511
 - Consumer Credit Counseling (CCC), 10-11, 98-99, 148, 206-207, 270-271, 388-389, 460, 517-518
 - Correlación, 34-39, 44, 221
 - aplicación a la administración, 44
 - causalidad versus, 39
 - coeficiente de, 37-39, 40, 44, 46, 235, 289
 - diagramas de dispersión, 34-37
 - matriz de, 282
 - relación lineal en la, 39
 - serial, 339-358
 - de primer orden, 340, 367
 - modelos autorregresivos y, 357-358
 - prueba de Durbin-Watson, 343, 347
 - soluciones de la, 347-358
 - Correlograma, 22, 67, 87
 - Costos del pronóstico, 6, 80, 504, 511
 - Criterio
 - de información
 - bayesiano (BIC) (Schwars), 431, 446
 - de Akaike (AIC), 431, 446
 - de los mínimos cuadrados, 223. *Véase* también Método de los mínimos cuadrados
 - Cuantitativos versus cualitativos 2, 3, 4, 79-80, 135, 503
 - Cuartiles, 18
 - Curvas de crecimiento, 80, 174, 246-250
- ## D
- Datos
 - acumulación de, 5, 61
 - aleatorios, 69-71
 - cíclicos, 63, 79
 - confiables, 61
 - confianza en bases de datos existentes, 512
 - consistencia en los, 61
 - datos de corte transversal, 19, 61, 62, 87
 - de corte transversal, 19, 61, 62, 87
 - estacionales, 63-64, 76, 79, 166-167
 - estacionarios, 72, 78, 365, 407-408
 - exactitud de los, 61
 - horizontales, 62
 - no estacionarios, 72, 109 365, 407-408
 - oportunos, 61
 - relevantes, 61
 - series de tiempo, 62-64, 87, 165
 - tendencia en los, 63, 72, 75, 87
 - tipos de, 61-62
 - Deflación del precio, 191, 192
 - Descomposición
 - clásica, 79, 80, 166-185
 - de la variabilidad, 230-234, 287
 - en el análisis de las series de tiempo, 166-185
 - Censo II, método de, 187-189
 - clásica, 79, 80, 166-185
 - problemas con, 166
 - Desviación
 - cuadrática media (MSD), 173
 - estándar, 16
 - media absoluta (MAD), 79, 88, 173
 - Diagrama
 - de árbol de decisiones, 493
 - de dispersión, 22, 34-37, 45, 61
 - Diferenciación, 72, 350-354, 354-357, 368, 408
 - Directorio de la Reserva Federal, 187
 - Distribución(es)
 - binomial, 24, 44, 45, 523-524
 - chi cuadrada, 527-528
 - continua, 25
 - de la muestra, 26-29, 45
 - de probabilidad, 22-26
 - F, 529-530
 - muestreo, 26-29
 - normal, 25, 45, 525
 - t, 28, 526
 - Downtown Radiology, cinco años de proyección de ingresos, 149-154
- ## E
- Ecuación
 - binomial, 24, 45
 - de la pendiente de la regresión alternativa, 40, 46
 - de regresión ajustada (estimada), 224, 253, 284, 312
 - Eight Asian Megatrends That Are Reshaping the World* (Nasbitt), 491
 - Ejemplos
 - de Excel
 - coeficientes de autocorrelación, 103-104
 - de computadora. *Véase* Excel; Minitab
 - desviación media y estándar, 58-59
 - modelo autorregresivo de primer orden, 296-297
 - modelo de tendencia exponencial, 217-219
 - regresión lineal simple, 277-278
 - regresión múltiple, 337-338
 - suavización exponencial, 161-163
 - de Minitab,
 - análisis de regresión, 274-277
 - ARIMA, 478-480
 - autocorrelación y diferenciación, 101-102

- descomposición de series de tiempo, 214-217
- estadística descriptiva y, 56-58
- promedio móvil y suavización exponencial, 159-161
- regresión en series de tiempo, 395-396
- regresión por pasos, 336-337
- El shock del futuro* (Toffler), 491
- Eliminación de la estacionalidad de los datos (desestacionalizar), 179-184
- Error(es)
 - cuadrático medio (MSE), 82, 88, 231, 232, 287
- de pronósticos, 81-84, 87-88, 107, 124-126, 225, 226-227, 307-309, 410
- estándar
 - de la estimación, 226-227, 252, 253, 287-288, 312
 - del coeficiente de autocorrelación, 68, 87
 - del pronóstico, 228, 252, 253, 291
 - porcentual
 - absoluto medio (MAPE), 83, 88, 173
 - medio (MPE), 83, 88
- Estadística descriptiva, 15-22, 44
- Estadístico
 - F, 238, 239, 254, 288-290, 312
 - regresión múltiple, de, 288-289
 - relación con el coeficiente de determinación, 239, 254, 290, 313
- Q, 69, 87
- t*, 28, 46, 237, 254, 291, 313
- Estimación, 29-30, 44
- puntual, 29, 30, 45
- F**
 - Fildes, R., 81
 - Filtración adaptable, 80
 - Formulación de escenarios, 485-486
 - Fortune*, 310, 500, 19
- G**
 - Gates, Bill, 490
 - Georgoff y Murdick, 483
 - Gibson, Rowan, 490
 - Global Paradox* (Naisbitt), 491
 - Grados de libertad, 17, 28-29, 44
 - Gráfica(s)
 - Box, 19-21
 - de probabilidad normal, 42
 - de puntos, 19
- H**
 - Heteroscedasticidad, 230, 358-361
 - Hibon, M., 81
 - Hipótesis nula, 30-31
 - Histograma, 21-22, 239
 - Horizonte de tiempo, 79, 80
- I**
 - Implementación del modelo, 5
 - Indicadores
 - adelantados, 80, 184
 - coincidentes, 184
 - de negocios, 184-185
 - de predicción, 228
 - del 95% para una muestra grande, 228, 253, 291, 313
 - del ciclo de negocios, 184-185, 192
 - económicos (negocios), 184-185, 192
 - estimados, 29, 30, 45, 46
 - retrasados, 184
 - Índice
 - de la actividad de negocios en el condado de Spokane, 379-382
 - de precios, 190-191
 - Inferencia estadística, 15, 29-34
 - Intercepción en Y, 39, 46, 223, 253
 - International Journal of Forecasting*, 486
 - Intervalo
 - de confianza (estimación del intervalo), 29, 46
 - en la regresión, 228-229, 253, 291, 313
 - fórmula para la media en el, 30, 46
- J**
 - Juicio
 - intuitivo. Véase Pronósticos en los negocios, aspectos de juicio
 - personal (subjetivo, intuitivo). Véase Pronósticos en los negocios, juicios, aspectos de
- K**
 - Kamp, Di, 490
 - 21st Century Manager: Future-Focused Skills for the Next Millennium, 490
- L**
 - La tercera ola* (Toffler), 491
 - Línea
 - de regresión, 222-226
 - ajustada, 223, 253
 - de los mínimos cuadrados, 223
 - Ljung (y Box), 69, 84-85, 87, 130
 - Lydia E. Pinkham Medicine Company, 461-463, 498-500
- M**
 - MAD. Véase Desviación media absoluta
 - Mahoud, E., 487
 - Makridakis, Spyros, 6, 81, 187, 482
 - Martin, Chuck, 490
 - Matriz de correlación, 282
 - Media (estadística), 15
 - de una muestra, 15
 - estacionaria, 62
 - Medidas de tendencia central. Véase Mediana
 - Megatendencias 2000* (Naisbitt y Aburdene), 491
 - Mercado de bonos, 324-327
 - Método(s)
 - de la proporción del promedio móvil, 176-179
 - de mínimos cuadrados, 39, 46, 169
 - análisis de regresión en, 223, 253, 284
 - de promedio para pronosticar, 78, 107, 111-119, 134, 135-136
 - de pronósticos en los negocios autorregresivo. Véase Modelos autorregresivos de promedio móvil (ARIMA)
 - construcción y evaluación del modelo, 5
 - criterio de selección del modelo, 431-432
 - métodos de suavización de. Véase Métodos de suavización exponencial
 - métodos exponenciales de, 80, 107, 111, 119-135, 136
 - métodos informales de, 80, 108-111, 136
 - modelos estacionales y, 130-134, 166-168, 361-364, 432-442
 - resumen de, 80, 505-506
 - tipos de, 2-3
 - de suavización exponencial, 79, 80, 107, 119-135, 136, 444, 506
 - ajustado para tendencia (método de Holt), 126-130, 132, 137
 - ajustado para tendencia y variación estacional (método de Winters), 130-135, 137
 - aplicación a la administración, 135-136
 - pronóstico de corto plazo y, 134
 - seguimiento de, 124-126
 - Delphi, 483-485
 - características clave del, 483
 - ventajas del, 485
 - elaboración de pronósticos. Véase Pronósticos en los negocios
 - informales, 78, 80, 108-111, 136
 - fórmulas de los, 108-110, 136
 - para seleccionar la ecuación de regresión múltiple, 300-307
- Microcomputadoras. Véase Computadoras
- Minorista en la red, 154-157, 471-474
- Mitchel, Wesley, 184
- Modelo(s)
 - aleatorio, 69, 87
 - autorregresivos (AR), 357-358, 367, 404-405, 445
 - aplicación a la administración, 444-445
 - autocorrelación, 400-405, 407
 - técnica Box-Jenkins y, 404-405
 - autorregresivos de promedio móvil (ARIMA), 78, 79, 399, 407, 445
 - análisis estacional, 432-442
 - aplicación a la administración, 444-445
 - características de los, 399
 - criterio de información bayesiano (BIC), 431-432
 - criterio de información de Akaike, 431-432
 - desventajas de los, 443-444
 - diferenciación y, 408
 - elaboración de pronósticos con, 411
 - estimación del modelo, 409-410

identificación del modelo, 407-409
 implementación del modelo, 407-430
 mejores prácticas de selección del parámetro, 430
 ventajas, 443-444
 verificación del modelo, 410-411
 de componentes no observados, 167
 de crecimiento de Gompertz, 80, 174
 de promedio móvil (MA), 78, 79, 80, 113-119, 136, 137, 405-406
 desventajas de los, 113
 de regresión lineal, 226
 de ruido blanco, 69
 de suavización de Holt, 126-130, 132, 137
 de Winters, 130-135, 137
 econométricos, 79, 80, 364-365
 logístico Pearl-Reed, 174
 multiplicativo, 167, 176, 192
Monthly Labor Review, 165
 Moore, Geoffrey H., 185
 Morgenthau, Henry, 184
 Mr. Tux, 10, 54-55, 97-98, 147, 202-206, 270, 385-387, 459-460, 518-519
 MSE. *Véase* Error cuadrático medio
 Muestra, 15, 29, 37, 44
 coeficiente de correlación de la, 38, 46
 desviación estándar de la, 16
 media de la, 15
 varianza de la, 16
 Multicolinealidad, 281, 297-300
 Murphy Brothers Furniture, 94-96, 148, 207-209

N

Naisbitt, John
 Eight Asian Megatrends That Are Reshaping the World (cinta de audio), 491
 Global Paradox, 491
 Megatendencias 2000 (con Patricia Aburdene), 491
 Megatendencias, 491
Net Future: The 7 Cyber trends That Will Drive Your Business, Create New Wealth, and Define Your Future (Martin), 490
 NeuralWare Inc., 489
 Nivel
 de significancia, 31
 de confianza, 29
 Normalidad, evaluación de, 42-43
 Número índice (para variación estacional), 176, 192

O

Oficina Nacional de Investigación Económica (NBER), 184, 185

P

Paquetes de software de pronósticos, 6-7
 Parente, 485
 Pendiente, en la relación lineal, 39, 223, 253
 Población (estadística), 15, 37

Políticas del pronóstico, 9
 Predicción
 intervalo de, 228, 253, 291, 313
 pronóstico puntual o, 227, 291
 Pricewaterhouse Coopers, 491
 Primera diferenciación, 351, 368
 Principio de parsimonia, 409, 445
 Probabilidad de la significancia, 33, 45
 Procedimiento
 de regresión por pasos, 304-307
 de todas las regresiones posibles, 290-292
 Promedio
 móvil simple, 113-116, 137
 simple, 111-113, 136, 137
 (media), 18. *Véase* Media.
 variable doble, 116-119, 137
 Pronóstico(s). *Véase* Pronósticos en los negocios
 cualitativos. *Véase* Pronósticos en los negocios, aspectos del juicio sobre
 de corto plazo, 79, 80, 185, 504
 de juicio. *Véase* Pronóstico en los negocios, aspectos de juicio
 de largo plazo, 80, 189, 504, 505-506
 econométrico, 364-365
 en los negocios
 a largo plazo, 80, 189-190, 504
 aceptación por la administración, 9, 80
 adecuación de las técnicas de pronóstico, 84-86
 administración de, 6, 503-513
 análisis de las series de tiempo y, 19, 62-64, 80, 165-192, 339-367, 381
 análisis de regresión y, 34, 44, 80, 221-252, 281-311, 339-367, 503-504, 505
 aplicaciones a la administración, 44, 86, 135-136, 189-190, 250, 252, 310-311, 367, 444-445
 aspectos de juicio de, 1, 3, 5, 7, 9-10, 115, 165-166, 246-247, 306-307, 430-431, 481-495, 503
 combinación de pronósticos, 486-488
 correlación y, 34, 44, 221
 costo de los, 6, 80, 504, 509, 511
 cuantitativo contra cualitativo, 2, 3, 6, 79-80, 135-136, 503
 datos en los, 5, 61-86
 de corto plazo, 79-80, 185, 504
 escenario de escritura en, 485-486
 evaluación de los, 5-6
 futuro de los, 512-513
 historia de los, 1-2
 horizontes de tiempo, 79, 80
 intermedio, 79-80, 504
 intuitivo, *Véase* Pronósticos en los negocios, aspectos de juicio de juicio personal en los. *Véase* Pronósticos en los negocios, aspectos de juicio de medición del error en, 81-84, 88, 107, 122-126, 226-227, 309-310, 410-411
 método Delphi en, 483-485

métodos de promedio para, 78, 107, 111-119, 134, 135-136
 micropronósticos, 4
 necesidad de, 1, 2, 6, 80, 135-136, 165, 367, 444, 503
 pasos en, 4-6, 509-510
 planteamiento del problema y recopilación de datos en los, 4-6
 políticas de, 9
 procedimiento de la curva de crecimiento, 79, 80, 174, 246-250
 proceso de, 4-6, 503-513
 pronósticos macroeconómicos, 3-4
 redes neurales y, 488-490
 reglas de, 9
 regresión múltiple y, 281-311
 residuos en los, 82, 87, 223, 239-241, 410, 430
 responsabilidad de los, 510-511
 selección y evaluación del método de, 4, 76-81, 107, 124-125, 126, 136, 175, 180, 503-504, 505-506
 sistemas de información gerencial (SIG) y, 511-512
 subjetivos. *Véase* Pronósticos en los negocios aspectos de tecnológicos, 481, 491
 teoría de las decisiones y, 493-495
 valor esperado y, 23, 45, 492, 496
 variable independiente contra variable dependiente, 221, 281-282, *Véase también* Autocorrelación, colinealidad; correlación serial
 intermedios (mediano plazo) 70, 80, 504, 505-506
 macro, 3, 3-4
 macroeconómicos, 3-4, 63, 180
 necesidad de, 1, 2, 6, 80, 135-136, 165, 367, 444, 503
 puntual, 227
 y los gerentes. *Véase* Pronósticos en los negocios, aplicaciones a la administración

Prueba

Anderson-Darling, 43
 de hipótesis, 30-32, 42-43, 236-239, 289, 291
 Durbin-Watson, 343-347, 367, 368, 530-532
 Publicaciones sobre cambio tecnológico (libros), 490-491

Q

Quenouille, M. H., 68

R

Raíz cuadrada del error cuadrático medio (RMSE), 83, 88
 Rand Corporation y el método Delphi, 483-485
 Rango, 18
 intercuartilar, 18
 Redes Neuronales (y pronósticos), 488, 490

Regresión

- lineal simple, 221-254
- múltiple, 79, 80, 281-313, 312, 505
 - aplicación a la administración, 310-311
 - coeficiente ajustado de correlación de, 290, 313
 - coeficiente de correlación, 289, 313
 - coeficiente de determinación, 289, 312
 - coeficiente de regresión parciales o netos en la, 285, 286, 312
 - coeficiente de regresión y, 285-286
 - definición de, 281
 - ecuación, selección de “la mejor”, 300-307
 - error estándar de la estimación en la, 287-288, 312
 - estadístico F y, 289, 312
 - estadístico *t* de, 291, 313
 - factor de inflación de la varianza, 297, 313
 - función de regresión múltiple de la población, 283, 312
 - función de, 283-284, 312
 - grados de libertad en la, 287, 312
 - inferencia de modelos, 286-291
 - interpretación de la, 285-286
 - matriz de correlación y, 282
 - modelos de, 283-285
 - multicolinealidad y, 281-282, 297-300
 - procedimiento de regresión por pasos, 304-307, 312
 - procedimiento de todas las regresiones posibles, 302-304
 - pronósticos econométricos y, 310-311
 - regresión simple versus, 283
 - residuos y, 307-309
 - selección de la ecuación en, 300-307
 - sobreajustes en la, 309-310, 312
 - tabla ANOVA, 288
 - variables explicativas en la, 281-282, 290-291
 - variables ficticias en la, 293-297, 312

Relación

- curva, 35, 36
 - de cointegración, 366, 367
 - lineal, 37-39, 221
 - negativa imperfecta, 35
 - negativa perfecta, 35
 - positiva perfecta, 35
 - negativa, 35, 37
 - positiva, 35
 - X-Y, 224, 225 *Véase también* Análisis de regresión
- Residuos, 82, 87, 180, 223, 239-241, 253, 307-309, 343-344, 410
- regresión múltiple y, 307-309
 - series de tiempo y, 343-344, 410
- Restaurante Golden Garden, 497
- Rethinking the Future: Rethinking Business, Principles, Competition, Control and Complexity, Leadership,*

Markets and the World (Gibson, ed., Toffler, y Toffler), 490

S

- “Señales de advertencia tempranas de la economía” (Moore y Shiskin), 185
- Señales de control, 124, 125, 136
- Series
 - aleatorias, 68
 - de tiempo
 - cointegradas, 365, 367
 - estacionarias, 72, 78, 365, 407-408
- Shiskin Julius, 185, 187
- Similitud pasado-futuro, supuestos de, 4
- Simulación en computadora, 495
- Small Engine Doctor, 201-202
- Sobreajustes, 309-310, 312
- Southwest Medical Center, 158-159, 214, 394-395, 476-477, 520
- Statistical Abstract of the United States*, 165
- Statistical Applications for the Sciences (SAS), 7
- Suavización. *Véase* Suavización
 - exponencial de Solar Alternative Company, 145-146
 - cuadrática exponencial, 80
 - exponencial simple, 119-124, 137
 - exponencial lineal, 80
- Supuestos
 - de similitud pasado-futuro, 4
 - en el análisis de regresión, 230, 239, 339, 489
- Surtido Cookies, 100, 159, 213-214, 393-394, 474-476
- Survey of Current Business, The*, 165, 184

T

- Taylor, Jim, 491
- Technology Forecast*, 491
- Técnica Box-Jenkins, 78, 399-445
 - análisis estacional y, 432-442
 - aplicación a la administración, 444-445
 - criterio de información
 - bayesiano (BIC) y, 431-432
 - de Akaike (AIC) y, 431-432
 - estadístico de Ljung-Box (modificado de Box-Pierce) y, 410
 - estrategia de construcción del modelo y, 407-411
 - modelo autorregresivo (AR)
 - de promedio móvil (ARIMA), (MA) en, 404, 405-406, 407
 - ventajas y desventajas de la, 443-444
- Tendencia, 62, 63, 72-75, 87, 109, 168-175, 192
 - central, medidas de, 18, 44
 - ciclo, 166, 169
 - cuadrática, 172, 192
 - definición de, 63, 87
 - exponencial, 173, 192
 - lineal, 169, 192

Teorema

- central del límite, 27
 - de Bayes, 493-495, 496
- Teoría de decisiones, 491-495
- The 500-Year Delta: What Happens After What Comes Next* (Taylor), 491
- The Road Ahead* (Gates), 490
- Tiger Transport, 266-267
- Toffler, Alvin
 - El shock del futuro, 491
 - La tercera ola, 491
- Tolerancia de error, 29
- Transformaciones de variables, 239, 243, 246

U

- UPS Air Finance Division, 466-469

V

Valor(es)

- atípicos, 20
 - esperado, 23, 45, 492, 496
 - p, 32-33, 45
 - probable. *Véase* Valor p.
- Variable(s)
- aleatoria, 22, 23
 - continua, 23, 44
 - discreta, 22-23, 44-45
 - valor esperado de la, 23
 - dependiente, 221, 226, 252, 281
 - endógenas, 364, 365
 - explicativas, 243, 281, 339
 - ficticias, 293-297, 312, 361
 - independiente, 221, 226, 252, 281
 - indicadoras. *Véase* Variables ficticias
 - retrasadas, 357. *Véase también* Modelos autorregresivos
- Variaciones cíclicas, 166, 180, 185
- Varianza, 16
- heteroscedasticidad y, 230, 358-361
- Ventas en un restaurante, 383-385, 457-459

W

- Wall Street Journal*, 310
- World Wide Web, 7, 165, 513

X

- X-12-ARIMA, 187-189

Y

- Y, pronóstico de, 227-230, 291, 313

Z

- Z como número de varianza, 25, 38

