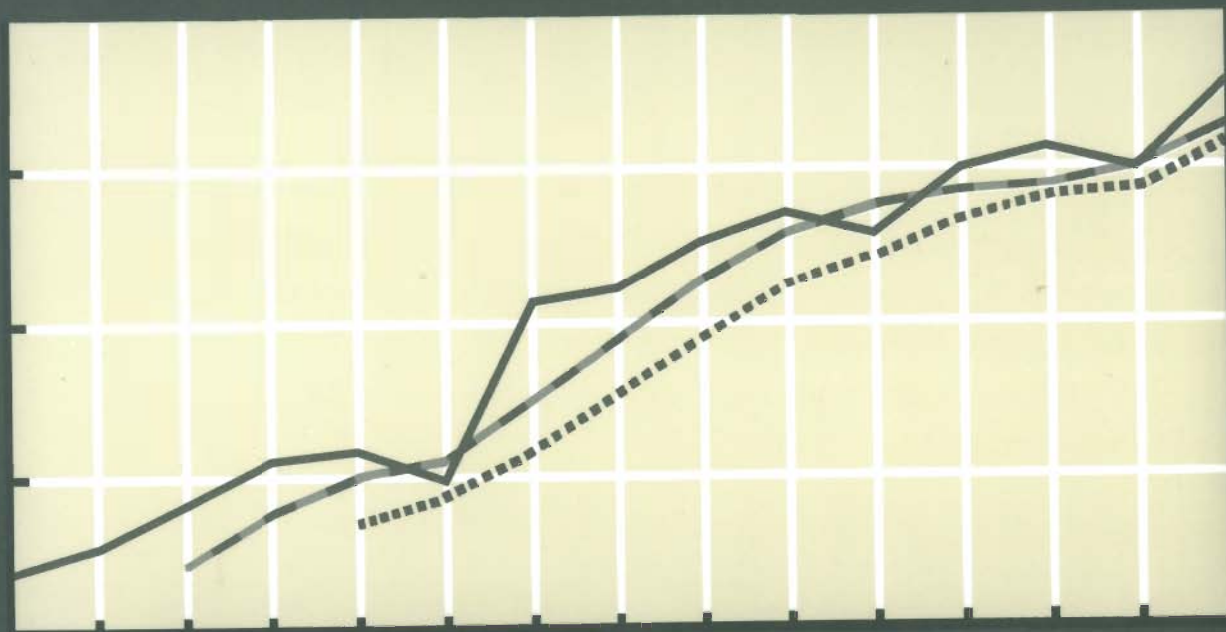


PRONOSTICOS EN LOS NEGOCIOS

QUINTA EDICION



Prentice
Hall

®

JOHN E. HANKE ARTHUR G. REITSCH

Pronósticos en los negocios

Pronósticos en los negocios

Quinta edición

JOHN. E. HANKE

ARTHUR G. REITSH

Eastern Washington University

Traducción:

SERGIO KOURCHENKO BARRENA
Traductor

Revisor técnico:

JUAN ANTONIO TORRE MARINA
Ingeniero Mecánico Electricista
Universidad Anáhuac



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

Edición en inglés:

Editorial / Production Supervision: Maryland Composition
Acquisition Editor: Richard Wohl
Cover Designer: Wendy Alling Judy
Production Coordinator: Marie McNamara

HANKE: PRONÓSTICOS EN LOS NEGOCIOS 5/E

Traducido del inglés de la obra: Business Forecasting

All Rights Reserved. Authorized translation from English language edition
published by Prentice Hall Inc.

Todos los Derechos reservados. Traducción autorizada de la edición en inglés
publicada por Prentice Hall Inc.

All rights reserved. No part of this book may be reproduced or transmitted in
any form or by any means, electronic or mechanical, including photocopying
recording or by any information storage retrieval system, without permission in
writing from the publisher.

Prohibida la reproducción total o parcial de esta obra, por cualquier medio ó
método sin autorización por escrito del editor.

Derechos reservados © 1996 respecto a la primera edición en español publicada

Atacomulco Núm. 500-5° Piso
Col. Industrial Atoto
53519, Naucalpan de Juárez, Edo. de México

ISBN 968-880-681-1

Original English Language Edition Published by Prentice Hall Inc.
Copyright © MCMXCV
All Rights Reserved
ISBN 0-205-16005-0

Impreso en México / Printed in Mexico

Dedicado a:

*Harry, Gerry, Jack e Irene
(quienes no necesitan leerlo);*

*Judy y Judy
(quien lo leerá y probablemente
no deseará leerlo, respectivamente);*

*Jill, Amy, Julie, Katrina y Kevin
(quienes podrían leerlo);*

*y a todos los estudiantes
(que más vale que lo lean).*

Contenido

Prefacio xv

1 INTRODUCCION A LOS PRONÓSTICOS 1

Actividad	1
Historia de los pronósticos	1
La necesidad de pronosticar	2
Tipos de pronóstico	4
Pronóstico macroeconómico	4
Selección del método de pronóstico	5
Pasos a seguir en el pronóstico	6
Administración del proceso de pronóstico	7
Paquetes de cómputo para pronóstico	8
Resumen	11
Caso de estudio 1.1: Mr. Tux	11
Caso de estudio 1.2: Consumer Credit Counseling	12
Versión estudiantil de Forecast Plus	13
Pronóstico mediante hojas de cálculo	14
Introducción al paquete de cómputo MINITAB	16
Introducción al paquete de cómputo SAS	18
Bibliografía	20

2 REVISIÓN DE CONCEPTOS ESTADÍSTICOS BÁSICOS 22

Actividad	22
Estadística descriptiva	22
Distribuciones de probabilidad	26
Distribuciones muestrales	30
Estimación	31
Prueba de hipótesis	33
Prueba de la bondad de ajuste	37
Análisis de correlación	39

Diagramas de dispersión	39
Coeficiente de correlación	43
Aplicación en la administración	47
Glosario	48
Fórmulas clave	49
Problemas	50
Caso de estudio 2.1: Alcam Electronics	55
Caso de estudio 2.2: Mr. Tux	57
Paquete de cómputo MINITAB	58
Bibliografía	60

3 FUENTES DE DATOS 61

Introducción	61
Tipos de datos	62
Fuentes de datos	63
Fuentes secundarias de datos	64
Fuentes externas	64
Datos propios	66
Fuentes primarias de datos	67
Aplicación en la administración	73
Glosario	75
Problemas	75
Caso de estudio 3.1: Argyl Food Products	76
Caso de estudio 3.2: Solomon's Jewelry	79
Caso de estudio 3.3: Mr. Tux	80
Caso de estudio 3.4: Consumer Credit Counseling	80
Caso de estudio 3.5: Estudio de factibilidad de mercado	82
Versión estudiantil de Forecast Plus	89
Pronóstico mediante hojas de cálculo	94
Paquete de cómputo MINITAB	95
Bibliografía	96

4 EXPLORACIÓN DE LOS PATRONES DE DATOS Y SELECCIÓN DE LA TÉCNICA DE PRONÓSTICO 97

Actividad	97
Componentes de series de tiempo	98
Exploración de patrones de datos mediante análisis de autocorrelación	100
Selección de una técnica de pronóstico	114
Técnicas de pronóstico para datos estacionarios	115
Técnicas de pronóstico para datos con una tendencia	116
Técnicas de pronóstico para datos con estacionalidad	116

Técnicas de pronóstico para series cíclicas	117
Otros factores por considerar al elegir una técnica de pronóstico	117
Medición del error en el pronóstico	119
Determinación de lo adecuado de una técnica de pronóstico	122
Aplicación en la administración	124
Glosario	125
Fórmulas clave	125
Problemas	126
Caso de estudio 4.1: Murphy Brothers Furniture	131
Caso de estudio 4.2: Mr. Tux	134
Caso de estudio 4.3: Consumer Credit Counseling	136
Versión estudiantil de Forecast Plus	136
Pronóstico mediante hojas de cálculo	139
El problema	139
La solución mediante hoja de cálculo	139
Paquete de cómputo MINITAB	141
Paquetes de cómputo SAS	143

5 PROMEDIOS MÓVILES Y MÉTODOS DE ATENUACIÓN 146

Actividad	146
Modelos no formales	151
Modelos de promedio	151
Promedios simples	151
Promedios móviles	152
Promedios móvil doble	154
Métodos de atenuación exponencial	158
Rastreo	161
Atenuación exponencial doble	163
Atenuación exponencial ajustada a la tendencia: Método de Holt	168
Atenuación exponencial ajustada a la tendencia y a la variación estacional: Modelo de Winter	172
Aplicación en la administración	176
Glosario	177
Fórmulas clave	178
Problemas	179
Caso de estudio 5.1: The Solar Alternative Company	184
Caso de estudio 5.2: Mr. Tux	185
Caso de estudio 5.3: Consumer Credit Counseling	186
Caso de estudio 5.4: Proyección de ingreso quinquenal para Downtown Radiology	187
Versión estudiantil de Forecast Plus	196
Pronóstico mediante hojas de cálculo	199
Bibliografía	202

6 ANÁLISIS DE REGRESIÓN 203

Línea de regresión	205	
Error estándar de la estimación		208
Predicción de Y	209	
Coefficiente de determinación		213
Residuos	217	
Prueba de hipótesis	220	
Salida de cómputo	223	
Transformación de variables		226
Aplicación en la administración		230
Glosario	232	
Fórmulas clave	232	
Problemas	233	
Caso de estudio 6.1: Tiger Transports		239
Caso de estudio 6.2: Butcher Products, Inc.		241
Caso de estudio 6.3: Departamento de Personal de Ace		243
Caso de estudio 6.4: Mr Tux		244
Caso de estudio 6.5: Consumer Credit Counseling		245
Versión estudiantil de Forecast Plus		246
Pronóstico mediante hojas de cálculo		249
Paquete de cómputo MINITAB		250
Paquete de cómputo SAS		252
Bibliografía	254	

7 REGRESIÓN MÚLTIPLE 255

Actividades	255	
Variables de predicción	255	
Matriz de correlación	256	
Ecuación de regresión múltiple	257	
Coefficientes de regresión	260	
Inferencia estadística en la regresión múltiple		261
Residuos	261	
Error estándar de la estimación		263
Salida de cómputo	264	
Variables ficticias	266	
Validación del modelo	271	
Heteroscedasticidad	271	
Colinealidad	272	
Selección de la mejor ecuación de regresión		274
Todas las regresiones posibles	277	
Regresión por pasos	279	
Notas finales sobre la regresión por pasos		283

Uso de la regresión para pronosticar datos estacionales	284
Pronósticos econométricos	287
Sobreajuste	289
Aplicación en la administración	289
Glosario	291
Fórmulas clave	291
Problemas	292
Caso de estudio 7.1: El Mercado de Valores	299
Caso de estudio 7.2: Ventas de restaurante	302
Caso de estudio 7.3: Mr. Tux	305
Caso de estudio 7.4: Consumer Credit Counseling	307
Versión estudiantil de Forecast Plus	308
Pronóstico mediante hojas de cálculo	312
Paquete de cómputo MINITAB	314
Paquete de cómputo SAS	315
Bibliografía	317

8 ANÁLISIS DE SERIES DE TIEMPO

318

Actividades	318
Descomposición	319
Índice de precios	320
Tendencia	322
Tendencia no lineal	327
Variación cíclica	330
Variación estacional	338
Tendencia estacional	342
Datos ajustados en forma estacional	346
Variaciones de corto plazo cíclica e irregular	347
Pronóstico estacional	349
El método de descomposición Census II	350
Aplicación en la administración	351
Glosario	353
Fórmulas clave	353
Problemas	353
Caso de estudio 8.1: The Small Engine Doctor	358
Caso de estudio 8.2: Mr. Tux	360
Caso de estudio 8.3: Consumer Credit Counseling	365
Versión estudiantil de Forecast Plus	366
Pronóstico mediante hojas de cálculo	367
Paquete de cómputo MINITAB	371
Bibliografía	373

9 REGRESION DE DATOS DE SERIES DE TIEMPO 374

Actividad	374	
El problema de heterosedasticidad durante la regresión de series de tiempo	374	
El problema de correlación serial durante la regresión de datos de series de tiempo	377	
Prueba de Durbin-Watson para correlación serial	380	
Soluciones a problemas de correlación serial	383	
Error de especificación en el modelo (omisión de una variable)		384
Regresión de cambios porcentuales	387	
Modelos autorregresivos	392	
Mínimos cuadrados generalizados	393	
Primera diferenciación	396	
El enfoque iterativo	398	
Aplicación en la administración	400	
Glosario	401	
Fórmulas clave	401	
Problemas	402	
Caso de estudio 9.1: La compañía de su elección	408	
Caso de estudio 9.2: Índice de actividad empresarial del condado de Spokane		409
Caso de estudio 9.3: Mr. Tux	415	
Caso de estudio 9.4: Consumer Credit Counseling	416	
Versión estudiantil de Forecast Plus	417	
Pronóstico mediante hoja de cálculo	421	
Paquete de cómputo MINITAB	424	
Paquete de cómputo SAS	426	
Ejemplo de TSP	428	
Bibliografía	430	

10 LA METODOLOGÍA BOX-JENKINS (ARIMA) 431

Actividad	431	
Técnica de Box-Jenkins	431	
Autocorrelaciones parciales	436	
Modelo autorregresivo	436	
Modelos de promedio móvil	439	
Modelos autorregresivos de promedio móvil	438	
Aplicación de la metodología	439	
Etapas 1: Identificación del modelo	439	
Etapas 2: Estimación del modelo y prueba de su adecuación		440
Etapas 3: Pronóstico con el modelo	441	
Un análisis estacional	465	

Aplicación en la administración	478
Glosario	480
Fórmulas clave	480
Problemas	480
Caso de estudio 10.1: Ventas de restaurante	487
Caso de estudio 10.2: Mr. Tux	488
Caso de estudio 10.3: Consumer Credit Counseling	490
Caso de estudio 10.4: Lydia E. Pinkham Medicine Company	491
Caso de estudio 10.5: Demanda diaria de permisos para pesca de trucha	495
Versión estudiantil de Forecast Plus	507
Paquete de cómputo MINITAB	511
Paquete de cómputo SAS	514
Bibliografía	519

11 ELEMENTOS DE JUICIO EN LOS PRONÓSTICOS 520

Actividad	520
Pronóstico de juicio	520
Curvas de crecimiento	522
El método Delphi	523
Formulación de escenarios	523
Combinación de pronósticos	524
Los pronósticos y las redes neurales	525
Resumen de la pronóstico de juicio	527
Otros elementos de juicio en los pronósticos	528
Administración del proceso de pronóstico	533
El proceso de pronóstico	533
Monitoreo de los pronósticos	536
Revisión de los pasos de pronóstico	539
La responsabilidad en los pronósticos	540
Costo de los pronósticos	541
Administración en los pronósticos	542
Los pronósticos y el sistema de información administrativa	542
El futuro de los pronósticos	543
Glosario	544
Fórmula clave	544
Caso de estudio 11.1: Boundary Electronics	544
Caso de estudio 11.2: Golden Gardens Restaurant	545
Caso de estudio 11.3: Busby Associates	545
Caso de estudio 11.4: Mr. Tux	550
Caso de estudio 11.5: Consumer Credit Counseling	552
Caso de estudio 11.6: Nueva visita a Lydia E. Pinkham Medicine Company	552
Bibliografía	555

APÉNDICES 557

A	DERIVACIONES	559	
B	GRÁFICAS DE PROPORCIÓN O SEMILOGARÍTMICAS		562
C	TABLAS	567	
D	DATOS PARA EL CASO DE ESTUDIO 7.1	580	
E	CONJUNTOS DE DATOS Y BASE DE DATOS		583

ÍNDICE 598

Prefacio

El objetivo de la quinta edición de *Pronóstico Empresarial*, continúa siendo el mismo que en las ediciones anteriores: presentar las técnicas estadísticas básicas para que las empresas, de manera individual, elaboren pronósticos y planes a largo plazo. El libro se escribió en un estilo directo y utiliza en forma amplia ejemplos prácticos de algunas empresas. Con el fin de proporcionar al estudiante el enlace necesario entre los conceptos teóricos y las aplicaciones en la vida real, al final de cada uno de los capítulos, se incluyen casos de estudio, los cuales forman un total de cuarenta y tres. Se hace énfasis en la aplicación gerencial de las técnicas, para la toma de decisiones. Se supone que los estudiantes poseen información adquirida en cursos introductorios tanto de estadística como de computación.

ORGANIZACIÓN

Todos los capítulos se revisaron para mejorar la claridad del texto e incrementar la efectividad de la enseñanza-aprendizaje. El contenido está organizado en seis secciones.

La primera sección (capítulos 1 y 2) presenta material que se necesita para estudiar pronósticos. La naturaleza de los pronósticos y un breve repaso de los conceptos estadísticos básicos establecen la etapa para la cobertura de las técnicas, que comienzan en la segunda sección.

En la segunda sección (capítulos 3 y 4) se enfatizan los métodos de recopilación de datos, la exploración de patrones de datos y la selección de una técnica de pronóstico.

La tercera sección (capítulo 5) cubre las técnicas de promedio y atenuación.

En la cuarta sección (capítulos 6 y 7) se abarcan las técnicas de pronóstico causal, tales como la correlación, regresión y análisis de regresión múltiple.

La quinta sección (capítulos 8 a 10) comprende las técnicas que involucran al análisis de series de tiempo.

El capítulo final del libro (capítulo 11) se estudian las técnicas de pronóstico y de juicio, junto con una exposición de la administración y monitoreo del proceso de pronóstico.

CAMBIOS EN LA QUINTA EDICIÓN

En la presente edición, se incorporaron los siguientes aspectos:

Como parte del paquete del libro de texto, se puede adquirir un paquete de pronóstico para microcomputadora denominado Forecast Plus, de Walonick Associates. Al final

de cada capítulo se incluye una demostración de cómo utilizar este paquete para resolver problemas de pronóstico.

Al final de cada capítulo, se incluye también un ejemplo de cómo usar el paquete MINITAB en situaciones de pronóstico.

Al final de los capítulos 1, 4, 6, 7, 9 y 10 se presentan ejemplos de la utilización del paquete SAS para resolver problemas de pronóstico.

También, se muestran ejemplos del uso de Lotus 1-2-3 para la solución de problemas de pronóstico, al final de cada capítulo.

Mediante una solicitud, se encuentra disponible, un nuevo disco de datos, el cual contiene numerosos conjuntos de datos para su utilización en prácticas estudiantiles.

Se incorporaron además diversos ejercicios y casos nuevos de estudio.

Se agregaron dos casos globales que se desarrollan en los capítulos 1 al 3 y se resuelven al final de cada uno de los capítulos sucesivos. En el capítulo 11 se incluyó una breve exposición de las redes neuronales.

EL PAPEL DE LA COMPUTADORA

En las primeras cuatro ediciones, se reconoció que la computadora constituye una poderosa herramienta en los pronósticos. Hoy en día, con gerentes modernos que aprovechan la facilidad y disponibilidad para efectuare pronósticos complejos a través de microcomputadoras de escritorio, la computadora es aún más importante.

Como resultado de las muchas horas de insomnio que pasaron tratando de decidir qué hacer con respecto a la computadora, los autores condujeron un estudio a nivel nacional (en EUA) de todas las instituciones miembros de la AACSB, para determinar qué era lo que hacían los cuerpos docentes respecto al uso de las computadoras en la enseñanza de los pronósticos. Los resultados mostraron lo siguiente:

1. La mayoría de los cuerpos docentes asignados a pronóstico (94.2%) intenta proporcionar experiencia al estudiante en el uso de la computadora.
2. En la encuesta se mencionan diversos paquetes estadísticos para macrocomputadoras y paquetes específicos de pronóstico para computadoras personales. Los paquetes mencionados con mayor frecuencia fueron: MINITAB, SAS, micro TSP y LOTUS 1-2-3.

Los autores decidieron emplear los siguientes enfoques para ayudar a cuerpos docentes y estudiantes a utilizar la computadora con fines de pronóstico:

1. Es posible adquirir, junto con el libro de texto, el paquete de pronóstico para microcomputadora Forecast Plus, de Walonick Associates.
2. Al final de cada capítulo y a lo largo de los capítulos relevantes del libro, se presentan instrucciones para utilizar el paquete MINITAB.
3. Al final de cada capítulo, se incluyen instrucciones de Lotus 1-2-3.

4. Casi al final del capítulo 1, se incluyó una breve lista general de los paquetes de pronóstico.
5. A lo largo del texto se insertaron diferentes ejemplos de salidas de computadora.

AGRADECIMIENTOS

Los autores están en deuda con los diversos instructores alrededor del mundo que utilizaron las cuatro primeras ediciones y proporcionaron sugerencias invaluable para mejorar el libro. Agradecemos en especial al Profr. Frank Frost (Marquette University); al Profr. William Darrow (Towson State University); al Profr. William C. Struning (Seton Hall University); a Susan Winters (Northwestern State University); al Profr. Mark Craze (Eastern Washington University); al Profr. Shik Chun Young (Eastern Washington University); a Judy Johnson y a Dorothy Mercer por su aportación de casos de estudio y a Jennifer Dahl por elaborar el índice.

Los autores aprecian sinceramente el tiempo dedicado por J. Scott Armstrong, Judy Johnson y Essam Muhoud, quienes aportaron valiosos materiales. También, partes del texto, en particular varios conjuntos de datos, son adaptaciones de los que aparecieron en la segunda edición de nuestro libro *Understanding Business Statistics*, publicado por Richard D. Irwin, Inc., a quien damos aquí crédito por su reutilización.

Agradecemos además a los revisores Arthur B. Jeffery (University South Alabama); Barry P. Laiss (California State University—Fresno); Leo P. Mahoney (Bryant College); Ray Nelson (Brigham Young University); William C. Struning (Seton Hall University); Marietta Tretter (Texas A & M University) y a Theodore Tsukahara (Saint Mary's College of California), por sus muy constructivos comentarios en la revisión del libro. Si hubiésemos tenido el talento suficiente para concretar todo lo que sugirieron nuestros revisores, el libro se habría mejorado en un 100%. Tal como está, realizamos nuestro mejor esfuerzo.

Por último, agradecemos a nuestras computadoras y nos preguntamos cómo fue que escribimos un libro de texto sin una. Cualquier error es atribuible a nosotros, no a las computadoras.

Introducción a los pronósticos

ACTIVIDAD

El interés de este libro se centra en el proceso del pronóstico empresarial. Lo cual incluye el estudio de datos históricos para descubrir sus patrones y tendencias fundamentales. Este conocimiento se utiliza para proyectar los datos a periodos futuros como pronósticos. La modificación de resultados numéricos, mediante un juicio acertado también representa un componente clave de pronóstico efectivo. Al hacerse más complejo el mundo de los negocios, ha aumentado la necesidad de asegurar, sobre cierta base racional, el futuro; por lo que el proceso de pronóstico ha adquirido una posición prominente en el proceso de administración de empresas.

HISTORIA DE LOS PRONÓSTICOS

Muchas de las técnicas de pronóstico que se utilizan actualmente y que se exponen en este libro se desarrollaron en el siglo XIX; un ejemplo de ello son los análisis de regresión. En contraste, sólo recientemente fue que algunos de los temas de este libro se desarrollaron y recibieron atención. En esta categoría se encuentra el procedimiento de Box-Jenkins.

Con el desarrollo de técnicas de pronóstico más complejas, junto con el advenimiento de las computadoras, los pronósticos recibieron más y más atención durante los años recientes. Este desarrollo es en especial cierto desde la proliferación de la pequeña computadora personal. Ahora, todos los administradores poseen la capacidad de utilizar técnicas de análisis de datos muy complejas para fines de pronóstico, y una comprensión de dichas técnicas es esencial hoy en día para los administradores de empresas.

Al crecer la preocupación de los administradores por el proceso de pronóstico, se continúan desarrollando nuevas técnicas de pronóstico. Esta atención se enfoca de manera particular en los errores, que son parte inherente de cualquier procedimiento de pronóstico. Es raro que los pronósticos coincidan al pie de la letra con el futuro, una vez llegado éste; quienes pronostican sólo pueden intentar que los inevitables errores sean tan pequeños como sea posible.

LA NECESIDAD DE PRONOSTICAR

En vista de las imprecisiones inherentes al proceso, ¿por qué es necesario pronosticar? La respuesta es que todas las organizaciones operan en una atmósfera de incertidumbre y que, a pesar de este hecho, se deben tomar decisiones que afectan el futuro de la organización. Para los gerentes de organizaciones, las conjeturas académicas son más valiosas que las no académicas. Este libro expone varias formas de realizar pronósticos, que se basan en métodos lógicos de manipulación de datos que se generaron por sucesos históricos.

Lo anterior no quiere decir que el pronóstico intuitivo sea malo. Al contrario, los sentimientos instintivos de quienes administran organizaciones, con frecuencia proporcionan el mejor pronóstico disponible. Este libro expone las técnicas de pronóstico que pueden emplearse para complementar el sentido común y la capacidad administrativa de los que toman decisiones; en el capítulo 11 se abordan los elementos de juicio en el proceso de pronóstico. En nuestra opinión, quienes toman decisiones lo harán mejor si a partir de la comprensión de las técnicas de pronóstico, tanto cualitativas como cuantitativas, las utilizan de manera adecuada, en vez de que se vean forzados a planear el futuro sin el beneficio de esta valiosa información complementaria.

En los últimos años, el papel del pronóstico con base en el juicio ha cambiado. Antes de la llegada de las técnicas modernas de pronóstico y del poder de las computadoras, el juicio del administrador era la única herramienta de pronóstico disponible. No existe evidencia de que los pronósticos basados sólo en juicios no sean tan precisos como aquellos que emplean la cabal aplicación de técnicas cuantitativas:

El ser humano posee un conocimiento único e información interior que no están disponibles en los métodos cuantitativos. Sin embargo, de manera sorprendente, estudios empíricos y experimentos de laboratorio han demostrado que sus pronósticos no son más precisos que los de los métodos cuantitativos. El ser humano tiende a ser optimista y subestimar la incertidumbre del futuro. Además, el costo del pronóstico con métodos de juicio es a menudo considerablemente más alto que cuando se utilizan métodos cuantitativos.¹

Creemos que quien pronostica de manera más efectiva es capaz de formular una hábil mezcla de buen juicio y técnicas de pronóstico cuantitativas, así como el evitar el caso extremo de la total dependencia en alguno de ellos. En un extremo encontramos al ejecutivo que por ignorancia y miedo a las técnicas cuantitativas y a las computadoras, se basa sólo en la intuición y el sentir. En el otro extremo está el pronosticador capacitado en las últimas técnicas de manipulación de datos, que no es capaz o no desea relacionar el proceso de pronóstico con las necesidades de la organización y con quienes toman decisiones en ésta. Opinamos que las técnicas de pronóstico cuantitativo que se exponen en este libro son sólo el

¹ S. Makridakis, "The Art and Science of Forecasting" *International Journal of Forecasting*, Vol. 2 (1986), p. 17.

punto de partida en el pronóstico efectivo de resultados importantes para la organización: el análisis, los juicios, el sentido común y la experiencia empresarial deben aplicarse en el punto donde estas técnicas importantes han generado sus resultados.

Debido a que siempre ha sido cambiante el mundo en el que operan las organizaciones, siempre ha existido la necesidad de hacer pronósticos. Sin embargo, en los últimos años se ha incrementado la confianza en las técnicas que abarcan una compleja manipulación de datos. Una nueva tecnología y nuevas disciplinas aparecieron de la noche a la mañana; la actividad gubernamental se intensificó en todos los niveles; la competencia se hizo más cerrada en muchas áreas; en casi todas las industrias se implantó el comercio internacional; crecieron y se crearon nuevas agencias de ayuda y servicios. Estos factores se combinaron para crear un clima organizacional que es más complejo, con una dinámica más rápida y más competitiva que nunca. Las organizaciones que no puedan reaccionar con rapidez a las condiciones cambiantes y prever el futuro con algún grado de precisión, están condenadas a la extinción.

Las computadoras, junto con las técnicas cuantitativas que hacen posible, se han vuelto más que recomendables en las organizaciones modernas: se han vuelto esenciales. Las dificultades antes expuestas generan una enorme cantidad de datos y una tremenda necesidad de extraer información útil de estos datos. Las herramientas modernas de pronóstico, junto con la capacidad de la computadora se han hecho indispensables para las organizaciones que operan en el mundo moderno.

¿Quién requiere hacer pronósticos? Casi cualquier organización, grande y pequeña, pública y privada, utiliza el pronóstico ya sea explícito o implícito, debido a que casi todas las organizaciones deben planear cómo enfrentar las condiciones futuras de las cuales tiene un conocimiento imperfecto. Además, la necesidad de hacer pronósticos cruza todas las líneas funcionales lo mismo que todo tipo de organizaciones. Se requiere hacer pronósticos en las áreas de finanzas, comercialización, personal y de producción, tanto en organizaciones gubernamentales y de búsqueda de ganancias, como en pequeños clubes sociales y en los partidos políticos nacionales. Considere las siguientes preguntas que sugieren la necesidad de algunos procedimientos de pronóstico:

- Si incrementamos nuestro presupuesto de publicidad en un 10%, ¿cómo afectará a las ventas?
- ¿Qué ingresos en rentas públicas puede esperar el gobierno del estado durante los próximos dos años?
- ¿Cuántas unidades deberemos vender para recuperar nuestra inversión fija en equipo de producción?
- ¿Qué factores podemos identificar que ayuden a explicar la variabilidad de las ventas mensuales unitarias?
- ¿Cuál es el pronóstico año por año del saldo total de préstamos de nuestro banco en los próximos diez años?
- ¿Habrá una recesión? Si es así, ¿cuándo empezará?, ¿qué tan severa será? y ¿cuándo terminará?

TIPOS DE PRONÓSTICO

Cuando los gerentes de organizaciones se enfrentan con la necesidad de tomar decisiones en una atmósfera de incertidumbre, ¿qué tipos de pronóstico tienen disponibles? En primer término, se deben clasificar los procedimientos de pronóstico de largo o corto plazos. Los *pronósticos a largo plazo* son necesarios para establecer el curso general de la organización para un largo periodo; de ahí que se conviertan en el enfoque particular de la alta dirección. Los *pronósticos a corto plazo* se utilizan para diseñar estrategias inmediatas y que usan los administradores de rango medio y de primera línea para enfrentar las necesidades del futuro inmediato.

También se podría clasificar a los pronósticos en términos de su posición en el entorno micro-macro, es decir, según el grado en que intervienen pequeños detalles vs. grandes valores resumidos. Por ejemplo, el gerente de una planta pudiera estar interesado en pronosticar el número de trabajadores que requerirá en los próximos meses (un *micropronóstico*), mientras que el gobierno federal está pronosticando el número total de personas empleadas en toda la nación (un *macropronóstico*). De nuevo, los diferentes niveles de administración en una organización tienden a enfocar diferentes niveles del entorno micro-macro. Por ejemplo, la alta dirección estaría interesada en pronosticar las ventas de toda la compañía, en tanto que los vendedores individuales estarían mucho más interesados en pronosticar su propio volumen de ventas.

Los procedimientos de pronóstico pueden también clasificarse de acuerdo con su tendencia a ser más cuantitativos o cualitativos. En uno de los extremos, *una técnica puramente cualitativa* es aquella que no requiere de una abierta manipulación de datos, sólo se utiliza el “juicio” de quien pronostica. Desde luego, incluso aquí, el “juicio” del pronosticador es en realidad el resultado de la manipulación mental de datos históricos pasados. En el otro extremo, *las técnicas puramente cuantitativas* no requieren de elementos de juicio; son procedimientos mecánicos que producen resultados cuantitativos. Por supuesto, ciertos procesos cuantitativos requieren de una manipulación de datos mucho más compleja que otros. Este libro hace hincapié en las técnicas de pronóstico cuantitativas, ya que es necesario un mayor entendimiento de estos útiles procedimientos en la administración efectiva de las organizaciones modernas. No obstante, debemos enfatizar de nuevo que junto con los procedimientos mecánicos y de manipulación de datos, se deben emplear elementos de juicio y sentido común. Sólo en esta forma se puede llevar a cabo un pronóstico inteligente.

PRONÓSTICO MACROECONÓMICO

Por lo regular pensamos en los pronósticos en términos de pronosticar variables importantes para una compañía individual o quizá para una parte de una compañía. Ejemplos de ello son las ventas mensuales de la empresa, las ventas unitarias de una de las tiendas de la compañía y las horas de ausencia por empleado y mes en una fábrica.

En contraste, existe un creciente interés en el pronóstico de importantes variables para la economía de una nación. Se ha realizado un gran trabajo en la evaluación de este tipo de pronóstico económico global, denominado *pronóstico macroeconómico*. Ejemplos que interesan al gobierno federal de EUA son el índice de desempleo, el producto nacional bruto y la tasa líder de interés.

En parte, la política económica se basa en proyecciones de importantes indicadores económicos, como los tres mencionados anteriormente. Por este motivo, hay un gran interés en mejorar los métodos de pronóstico que enfocan tales mediciones globales del comportamiento económico de un país. En la actualidad, los métodos de pronóstico se pueden dividir en forma genérica en dos enfoques; 1) métodos que usan los enfoques tradicionales de análisis de series de tiempo y 2) métodos menos estructurados que se enfocan en las propiedades estadísticas de las mediciones históricas. Ambos enfoques se describen en este libro.

Una de las principales dificultades para el desarrollo de pronósticos precisos de la actividad económica global, consiste en cambios significativos en algún factor económico clave. Entre tales factores se encuentran los cambios significativos en los precios del petróleo, variaciones súbitas de la inflación, y los cambios de política global en el gobierno de otro país que afectan la economía propia.

La posibilidad de dichos cambios significativos en el escenario económico ha generado una pregunta clave en el pronóstico macroeconómico: ¿deberían modificarse los pronósticos generados mediante el procedimiento de pronóstico utilizando el juicio de quien los produce? Esta pregunta aparece a menudo en el trabajo que se realiza en la actualidad sobre la metodología de pronóstico.

En materia de pronóstico macroeconómico se sigue desarrollando un gran esfuerzo tanto teórico como práctico. Un ejemplar del *International Journal of Forecasting* (Vol. 6, núm. 3, octubre de 1990), está dedicado a este tema. Considerando la importancia de un pronóstico económico preciso para la formulación de la política nacional, se puede esperar una continua atención sobre este tipo de pronóstico en el futuro.

SELECCIÓN DEL MÉTODO DE PRONÓSTICO

La exposición anterior sugiere varios factores a considerar en la selección de un método de pronóstico. Se debe contemplar el nivel de detalle. ¿Se requiere de un pronóstico de detalles específicos (un micropronóstico)? ¿O se necesita conocer el estado futuro de algún factor global o general (un macropronóstico)? ¿Se precisa el pronóstico de algún punto en el futuro cercano (un pronóstico a corto plazo), o para un punto en el futuro distante (un pronóstico a largo plazo)? y, ¿hasta qué grado son apropiados los métodos cualitativos (de juicio) y cuantitativos (de manipulación de datos)?

La consideración que se impone en la selección de un método de pronóstico es la de que los resultados deben facilitar el proceso de toma de decisiones de los administradores de la organización. Por lo tanto, el requerimiento esencial no es que el método de pronóstico comprenda un proceso matemático complicado o que sea lo último en complejidad. En vez de ello, el método elegido deberá producir un pronóstico que sea preciso y comprensible para los administradores, de modo que pueda ayudar a producir mejores decisiones.

Además, la utilización del procedimiento de pronóstico debe producir un beneficio que exceda al costo asociado con su uso.

PASOS A SEGUIR EN EL PRONÓSTICO

Todos los procedimientos formales de pronóstico comprenden la extensión de las experiencias del pasado al futuro incierto. De ahí la suposición de que las condiciones que generaron los datos anteriores son indistinguibles de las condiciones futuras, con excepción de aquellas variables reconocidas de manera explícita por el modelo de pronóstico. Por ejemplo, si se está pronosticando el índice de desempeño de los empleados en el trabajo, usando sólo como pronóstico la calificación del examen de admisión, se asume que el índice de desempeño en el trabajo de cada persona se afecta sólo por dicho examen. Considerando que la suposición de pasado y futuro indistinguibles no se cumple, resultarán pronósticos imprecisos, a menos que se modifiquen a juicio de quien pronostica.

La aceptación de que las técnicas de pronóstico funcionan sobre datos generados en sucesos históricos pasados conduce a la identificación de cuatro pasos en el proceso del pronóstico:

1. Recopilación de datos.
2. Reducción o condensación de datos.
3. Construcción del modelo.
4. Extrapolación del modelo (el pronóstico en sí).

El paso 1 sugiere la importancia de obtener los datos adecuados y asegurarse que son correctos. Con frecuencia este paso es el mayor reto de todo el proceso de pronóstico y el más difícil de controlar, ya que los pasos siguientes se efectúan sobre los datos, sean o no relevantes para el problema en cuestión. Siempre que se hace necesario obtener datos pertinentes en una organización, abundan los problemas de recopilación y control de calidad.

El paso 2, la reducción de datos con frecuencia es necesaria ya que en el proceso de pronóstico es posible tener muchos o muy pocos datos. Algunos datos pueden no ser pertinentes al problema, por lo que reducirían la precisión del pronóstico. Otros datos pueden ser los adecuados, pero sólo en ciertos periodos históricos. Por ejemplo, en el pronóstico de las ventas de automóviles compactos podría desearse emplear sólo datos de ventas de automóviles a partir del embargo petrolero de la década de 1970, en vez de datos de los últimos 50 años.

El paso 3, la construcción del modelo, implica el ajustar los datos reunidos en un modelo de pronóstico que sea el adecuado para minimizar el error en el pronóstico. Entre más sencillo sea el modelo, será mejor para lograr la aceptación del proceso por parte de los administradores que toman las decisiones en la empresa. Con frecuencia se debe establecer un balance entre un enfoque de pronóstico complejo que ofrezca ligeramente más precisión y un enfoque sencillo que sea fácil de entender y ganar el apoyo de quienes toman las decisiones, de manera que lo utilicen efectivamente. Es obvio que los elementos de juicio forman parte de este proceso de selección. Ya que el libro expone diversos modelos

de pronóstico y su aplicabilidad, confiamos en la capacidad del lector para ejercitar el buen juicio en la elección y uso apropiado de los modelos de pronóstico, capacidad que se incrementará después de estudiar este material.

El paso 4 consiste en la extrapolación en sí del modelo de pronóstico, lo cual ocurre una vez que se recolectaron y tal vez redujeron, los datos adecuados y que se seleccionó un modelo de pronóstico apropiado. Es común que quien realizó el pronóstico revise la precisión del proceso mediante el pronóstico de periodos recientes de los que se conocen los valores históricos reales. Es entonces cuando se observan los errores de pronóstico y se resumen de algún modo. Ciertos procedimientos de pronóstico suman los valores absolutos de los errores y pueden reportar esta suma, o dividirla entre el número de intentos de pronóstico para obtener el error de pronóstico promedio. Otros procedimientos obtienen la suma de cuadrados de los errores, que se compara luego con cifras similares de métodos de pronóstico alternativos. Algunos procedimientos también rastrean y reportan la magnitud de los términos de error sobre el periodo de pronóstico. El examen de los patrones de error conduce con frecuencia al analista a la modificación del procedimiento de pronóstico, el cual genera después pronósticos más precisos. Los métodos específicos de medición de error en el pronóstico se abordan al final del capítulo 4.

ADMINISTRACIÓN DEL PROCESO DE PRONÓSTICO

La exposición en este capítulo tiene como objeto subrayar nuestra creencia de que la capacidad administrativa y el sentido común deben formar parte del proceso de pronóstico. Se debe pensar en quien pronostica como en un asesor de la dirección, en vez del monitor de un dispositivo automático de toma de decisiones. Por desgracia, este último es el caso en la práctica, en especial en el entorno de la computadora. De nueva cuenta, las técnicas en el proceso de pronóstico deben ser vistas como lo que en realidad son, herramientas que utilizarían los administradores para llegar a mejores decisiones.

Se puede mejorar la utilidad de los pronósticos si los administradores adoptan una actitud más realista. No se debe ver al proceso como un sustituto de la profecía, sino como la mejor forma de identificar y extrapolar patrones o relaciones establecidos con el fin de pronosticar. Si se admite tal actitud, se deben considerar inevitables los errores de pronóstico e investigar las circunstancias que los generan.²

Dicho lo anterior, si el proceso de pronóstico se va a conducir de la manera adecuada, deben siempre surgir diversas preguntas clave.

- ¿Por qué se requiere de los pronósticos?
- ¿Quién utilizará los pronósticos y cuáles son sus requerimientos específicos?
- ¿Qué nivel de detalle o agregación se requiere y cual es el horizonte adecuado en el tiempo?

² Ibid., p. 33.

- ¿Qué datos hay disponibles? ¿serán suficientes para generar el pronóstico que se requiere?
- ¿Cuál será el costo del pronóstico?
- ¿Qué tan preciso podemos esperar que sea el pronóstico?
- ¿Se hará a tiempo el pronóstico para ayudar al proceso de toma de decisiones?
- Quien pronostica, ¿tiene un claro entendimiento de cómo se usará el pronóstico en la organización?
- ¿Hay disponible un proceso de retroalimentación para evaluar el pronóstico una vez hecho y ajustar el proceso de acuerdo con ello?

PAQUETES DE CÓMPUTO PARA PRONÓSTICO

En la década pasada el desarrollo que ha tenido el mayor impacto en el pronóstico es el de los paquetes de programas de cómputo diseñados específicamente para tratar en forma directa diferentes métodos de pronóstico. Hay dos tipos de paquetes de cómputo de interés para los pronosticadores: 1) paquetes estadísticos que incluyen análisis de regresión y otras técnicas que se utilizan con frecuencia en los pronósticos; y 2) paquetes de pronóstico diseñados específicamente para aplicaciones de este tipo.

Se han desarrollado cientos de paquetes estadísticos y de pronóstico tanto para macro como microcomputadoras (o computadoras personales, a las que con frecuencia se les llama PCs). Los administradores con PCs sobre sus escritorios y el conocimiento de técnicas de pronóstico, ya no dependen de un equipo de trabajo para realizar sus pronósticos. Los administradores modernos están aprovechando la ventaja de la facilidad y disponibilidad de métodos complejos de pronóstico que proporcionan las computadoras personales.

En esta sección, se mencionan algunos de los paquetes de cómputo estadísticos y de pronóstico más utilizados. Con frecuencia se utilizan paquetes estadísticos que funcionan en macrocomputadoras e incluyen partes que tratan directamente varios métodos de pronóstico. Tres de los paquetes más populares son:

1. Minitab: Véanse ejemplos de instrucciones y salidas a lo largo del texto y al final de cada capítulo.
2. Statistical Package for the Social Sciences (SPSS). Paquete estadístico para las ciencias sociales.
3. Statistical Analysis System (SAS). Sistema de análisis estadístico. Véanse ejemplos de salidas al final de la mayoría de los capítulos.

Un útil resumen de 48 paquetes estadísticos para macrocomputadoras fue compilado por Mahmoud, Rice, McGee y Beaumont (1986).³ Desafortunadamente, estas subrutinas de pronóstico requieren, para su aplicación, de un conocimiento total del paquete y de un analista altamente calificado.

³ E. Mahmoud, G. Rice, V. E. McGee y C. Beaumont, "Mainframe Multipurpose Forecasting Software: A Survey", *Journal of Forecasting*, Vol. 5 (1986), pp. 127-137.

En años recientes, se desarrolló un nuevo tipo de paquete de cómputo para macrocomputadoras, dirigido específicamente a las necesidades de los administradores. Sibyl/Runner, propiedad de Applied Decisions Systems, que también proporciona asesoría, ha ganado una amplia aceptación al principio de la década de 1990. Ha sido utilizado con éxito tanto por compañías como por instituciones académicas. Un paquete más reciente denominado FuturCast: The Total Forecasting System desarrollado por Spyros Makridakis y Robert Carbone, es otro buen ejemplo de un paquete interactivo de pronóstico diseñado para administradores. Este paquete lo comercializa Futurion Associates, Inc. Otros dos paquetes que se usan ampliamente en el pronóstico económico son TSP (véase el ejemplo de salida al final del capítulo 9) y Shazam. Otro útil resumen de 37 paquetes de pronóstico para macrocomputadora fue compilado por Mahmoud, Rice, McGee y Beaumont (1986).⁴

Minitab, SPSS, TSP (MicroTSP) y Sybil/Runner han creado recientemente versiones que funcionan en computadoras personales. Statpro, StatPac, Daisy, Decision Support Modeling, VisiTrend/Plot, Graph N' Calc, Expert Choice, Nuametrics, SYSTAT y Xtrapolator son paquetes similares a Minitab, que contienen varios métodos de pronóstico y que pueden funcionar en computadoras personales. El muy utilizado paquete de ajuste estacional X11ARIMA/88, se encuentra disponible tanto para macrocomputadoras como para PCs.

Se desarrollaron diversos paquetes específicamente diseñados para funcionar en microcomputadoras. Por ejemplo, Walonick Associates ofrece Forecast Plus (véanse ejemplos de instrucciones y salidas al final de cada capítulo). Otros paquetes de programas diseñados para microcomputadoras que se comercializan en la actualidad son:

- AutoBox 3.0 de Automatic Forecasting Systems, Inc.
- Autocast II de Levenbach Associates
- Business and Economic Forecasting: Decision Support System Software de John Wiley & Sons
- Easy Forecaster Plus I o II del Institute of Business Forecasting
- Economics Software Program (ESP) de Chase Econometrics
- Express Easycast de Management Decisions Systems, Inc.
- ET (Econometrics Toolkit) de Econometric Software, Inc.
- Forecast Master de Scientific System, Inc.
- 4 Cast de Scientific Systems, Inc.
- Forecasting de Hewlett Packard
- Forecast Pro de Business Forecast Systems, Inc.
- Micranal de Gwilym Jenkins & Partners, Ltd
- Micro-BJ de Stratix

⁴ E. Mahmoud, G. Rice, V. E. McGee y C. Beaumont, "Mainframe Specific Purpose Forecasting Software: A Survey", *Journal of Forecasting*, Vol. 5 (1986), pp. 75-83.

- MICCROFIT3 de Oxford University Press
- MTS (the Multiple Time Series Program) de Automatic Forecasting Systems
- Pro*Cast de Fleming Software
- +Forecast de Computer Software Consultants, Inc.
- RATS de Var Econometrics
- Smartforecasts II de Smart Software, Inc.
- SORITEC de Sorites Group
- Timestat de Timestat, Inc.
- SCA Statistical System de Scientific Computing Associates Corp.
- The Forecaster de Wadsworth Electronic Publishing Company
- ySTAT de Ming Telecomputing, Inc.

Se puede encontrar un análisis muy útil de muchos de estos paquetes en el ejemplar del 14 de marzo de 1989 de *PC Magazine*.⁵ Otro útil resumen de 94 paquetes de pronóstico para computadoras personales fue compilado por Beaumont, Mahmoud y McGee (1985).⁶

Se debe mencionar además que los paquetes integrados Lotus 1-2-3 (véanse ejemplos e instrucciones al final de cada capítulo), Excel, Quattro Pro, Framework, Symphony, Visi On, Lisa 7/7 Office System y otros similares, se pueden usar para pronósticos. Ninguno de estos paquetes integrados (con la posible excepción del Lisa 7/7), realiza pronósticos completos por sí mismo, pero cada uno puede modificarse con facilidad para realizarlas. Algunas adaptaciones que se pueden hacer a Lotus 1-2-3 son:

- ForeCalc de Business Forecast Systems
- Forecast! de Intex Solutions, Inc.
- Forecast/DSS y Lotus-FORMAN, ambos de John Willey & Sons, Inc.
- Add a Stat y Multifit, ambos de Abacus Scientific Software Company
- STAT-PACKETS de Statpac, Inc.
- Spreadsheet Forecaster de Levenbach Associates.

Para una excelente exposición sobre las adaptaciones de pronóstico para Lotus 1-2-3, consulte la sección de software de los ejemplares de marzo y junio de 1992 del *International Journal of Forecasting*.⁷

Para otra excelente exposición sobre los programas en general de pronóstico empresarial, consulte el artículo "Choosing and Using Business Forecasting Software" en el ejemplar de enero de 1985 de *Creative Computing*.⁸

⁵ "Statistical Analysis Forecasting", *PC Magazine*, Vol. 8(5) (marzo, 1989), pp. 225-241.

⁶ C. Beaumont, E. Mahmoud y V. E. McGee, "Microcomputer Forecasting Software: A Survey", *Journal of Forecasting*, Vol. 4 (1985), pp. 305-311.

⁷ L. Tashman y R. Whitycombe, "Software Reviews", *International Journal of Forecasting*, Vol. 7 (1992), pp. 543-553 y Vol. 8 (1992), pp. 121-127.

⁸ K. Keating, "Choosing and Using Business Forecasting Software", *Creative Computing*, 11 (1) (enero, 1985), pp. 119-135.

RESUMEN

El propósito del pronóstico consiste en reducir el margen de incertidumbre dentro del que se deben efectuar los juicios de la administración. Este propósito sugiere dos reglas principales a las que debe adherirse el proceso de pronóstico:

1. El pronóstico debe ser técnicamente correcto y producir predicciones precisas.
2. El procedimiento de pronóstico y sus resultados deben ser presentados con efectividad a la administración, de modo que los pronósticos se utilicen en el proceso de toma de decisiones en beneficio de la empresa; también los resultados deben ser justificados con base en su costo-beneficio.

La última consideración a menudo se malinterpreta y puede ser frustrante para los pronosticadores profesionales. Aun cuando los pronósticos se usen en beneficio de la empresa, deben ser utilizados por aquellos que tienen la autoridad en la toma de decisiones. De este hecho surge la pregunta de lo que podría llamarse la “política” de los pronósticos. Con frecuencia, los gastos sustanciales y en ocasiones mayores y la distribución de recursos dentro de la empresa recaen sobre la visión de la administración en los sucesos futuros. Como el movimiento de recursos y el poder dentro de la organización se basa a menudo en el rumbo que se percibe del futuro (pronósticos), no es de sorprender que haya una cierta cantidad de intriga alrededor del proceso de pronóstico. Esta consideración subraya la importancia de la segunda regla primaria: los pronósticos generados dentro de la empresa deben ser entendidos y apreciados por quienes toman las decisiones, de manera que encuentren su vía en la administración de la empresa.

El resto de este libro aborda varios modelos y procedimientos de pronóstico. En primer término se presenta un repaso de los conceptos estadísticos básicos, seguido de una exposición de los procedimientos de análisis de regresión. A continuación se expone el pronóstico con series de tiempo de datos y en los siguientes capítulos, técnicas más avanzadas de predicción. El capítulo final del libro trata sobre el proceso de pronóstico y su relación con quienes toman las decisiones en la empresa.

CASO DE ESTUDIO 1.1

MR. TUX

John Mosby es propietario de varias tiendas Mr. Tux de renta de ropa de etiqueta, la mayoría de ellas se ubican en las áreas de Spokane y Seattle, Washington.⁹ Su tienda de Spokane también hace camisas de etiqueta, las cuales distribuye a las tiendas de renta a lo largo del país. Ya que la actividad de renta varía de tiempo en tiempo debido a graduaciones, reuniones y otras actividades, John sabe que su negocio es de temporada. A él le gustaría medir este efecto de temporada, tanto para ayudarlo en la administración de su negocio como para utilizarlo en la renegociación de pagos de su préstamo con el banco.

⁹ Estamos en deuda con John Mosby, propietario de las tiendas de renta de ropa Mr. Tux, por su ayuda en la preparación de este caso. John fue estudiante nuestro, y aunque años más tarde no recordó los detalles específicos de las técnicas de pronóstico, sabía que con alguna ayuda podría desarrollar diversos modelos de pronóstico que le ayudaran. Esta sensación latente de que se podía extraer información útil de datos sin procesar es, en nuestra opinión, el verdadero valor de los estudios cuantitativos en el currículo de administración.

Aún de mayor interés para John es encontrar una forma de pronosticar sus ventas mensuales. Su negocio continúa creciendo, lo que a su vez requiere de mayor capital y deuda a largo plazo. John tiene fuentes para ambos tipos de financiamiento de su crecimiento, pero los inversionistas y los bancos están interesados en una forma concreta de pronosticar ventas futuras. Aunque confían en John, su opinión de que el futuro del negocio “se ve estupendo” no les convence del todo.

Como primer paso en la elaboración de un modelo de pronóstico, John se dirige a una de sus empleadas, Virginia Perot, para recopilar los datos de ventas mensuales de los últimos años. En los capítulos que siguen, se utilizan varias técnicas de pronóstico con estos datos para Mr. Tux. Los esfuerzos se concretan en el capítulo 11 y John Mosby intenta elegir la técnica de pronóstico que mejor cubra sus necesidades.

CASO DE ESTUDIO 1.2

CONSUMER CREDIT COUNSELING

Consumer Credit Counseling (CCC) (Consultoría de crédito al consumidor), es una corporación privada sin fines de lucro, fundada en 1982.¹⁰ Los objetivos de CCC eran proporcionar asistencia a los consumidores en la planeación y seguimiento de presupuestos, asistir a los consumidores en la negociación de arreglos con los otorgantes de crédito para cubrir obligaciones morosas e impartir cursos sobre la administración de dinero.

La consultoría financiera privada se proporcionaba sin costo a familias o individuos que tienen dificultades financieras o que querían mejorar su habilidad en el manejo de dinero. Los programas educativos de administración de dinero se impartían a escuelas, grupos comunitarios y empresas. Se ofrecía un programa sobre manejo de deuda como una alternativa a la bancarrota. A través de este programa, CCC negociaba con los acreedores por parte del cliente para obtener arreglos especiales de pago. Luego el cliente hacía un pago global a CCC, y ésta a su vez distribuía este pago entre los acreedores.

CCC tenía una mezcla de personal pagado y voluntario; de hecho, el personal voluntario superaba al pagado en proporción de tres a uno. Un grupo de siete empleados pagados cubría las necesidades de administración y de oficina y cerca de la mitad de las consultorías. La otra mitad de este servicio era cubierta por veintiún consultores voluntarios.

CCC dependía principalmente de los fondos corporativos para cubrir su operación y servicios. El “Programa de distribución equitativa de fondos” permitió que los acreedores que recibieron pagos provenientes de los programas de administración de deuda, donaran a CCC una parte de los fondos que les fueron cubiertos a través de estos programas.

Una gran parte del apoyo corporativo provenía de una utilidad local, la que proporcionaba fondos para sostener el puesto de un consultor de tiempo completo, así como espacio de oficina para dar consultoría en todas las oficinas.

Además, las cuotas de los clientes eran una fuente de fondos. Los clientes que participaron en la administración de deuda pagaban una cuota mensual de 15 dólares para ayudar a cubrir el costo administrativo del programa. (Las cuotas se redujeron o se exentaron a los clientes que financieramente no podían cubrirlas.)

Estos antecedentes se emplearán en los capítulos sucesivos, al enfrentar CCC dificultades relacionadas con el pronóstico de variables estratégicas.

¹⁰ Estamos en deuda con Mark Harnishfeger, Director Ejecutivo de Consumer Credit Counseling de Spokane y con Dorothy Mercer, Presidente de la Mesa Directiva, por su ayuda en la operación de este caso. Dorothy fue estudiante nuestra en la maestría en administración de consultor de tiempo completo, así como espacio de oficina para dar consultoría en todas las oficinas.

VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Forecast Plus es un paquete completo de pronóstico estadístico que maneja todas las fases del análisis de pronósticos. La versión estudiantil contiene un disco de administración de datos (Data Management Disk) y un disco del sistema principal (Main System Disk).

El paquete se diseñó para minimizar la necesidad de intercambiar discos. El disco del sistema principal contiene todos los programas de pronóstico. El disco de administración de datos contiene todos los programas necesarios para crear, modificar y manipular archivos de datos así como un conjunto completo de programas de exploración.

Si usted utiliza una unidad de disco duro, podrá copiar en él todos los programas y no tendrá que intercambiar discos.

Los pasos para ejecutar Forecast Plus son los siguientes:

1. Cree un archivo de datos.
2. Use el paquete de exploración para desarrollar accesos a los datos.
3. Seleccione una técnica de pronóstico apropiada.
4. Utilice el paquete de pronóstico para modelar los datos.
5. Ejecute el modelo de pronóstico hasta encontrar los mejores parámetros.

El primer paso en cualquier procedimiento de pronóstico consiste en recopilar datos y capturarlos en la computadora. El disco de administración de datos le permitirá crear, modificar y transformar archivos de datos. Dichos archivos pueden contener ya sea una sola variable o múltiples variables. Los programas de administración de datos son al mismo tiempo complejos y fáciles de usar.

El segundo paso es tener una comprensión general de los datos. Diferentes técnicas de pronóstico son apropiadas para diferentes tipos de patrones en los datos (por ejemplo, estacionales vs. no estacionales). El paquete de exploración (en el disco de administración de datos) es una parte muy importante de Forecast Plus, ya que le ayuda a comprender los patrones en los datos, de modo que pueda usted elegir una técnica de pronóstico adecuada.

El último paso consiste en modelar los datos con un procedimiento de pronóstico específico. El paquete de pronóstico le da a elegir entre trece diferentes procedimientos de pronóstico. Estos varían en complejidad, técnica y aplicabilidad, dependiendo del conjunto de datos que se esté analizando. Con frecuencia, el pronóstico cuantitativo consiste en un método de prueba y error para encontrar buenas constantes de atenuación; así que es posible que requiera ejecutar muchas veces un procedimiento específico de pronóstico antes de descubrir el mejor modelo.

Todos los archivos de Forecast Plus están almacenados en formato ASCII secuencial delimitado por comas: de ahí que, si es usted un programador, podrá con facilidad leer los datos con su propio programa en BASIC. También podrá utilizar los archivos creados con otros programas. Forecast Plus tiene la capacidad de leer cualquier archivo que esté almacenado en formato ASCII secuencial delimitado por comas, como los archivos que dBase almacena en dicho formato.

Introducción

Para inicializar Forecast Plus en un sistema típico de dos unidades de disco, se requiere seguir los siguientes pasos:

1. Inserte el disco del sistema operativo (DOS) en la unidad A y reinicie la computadora.
2. Introduzca la fecha y hora. (Forecast Plus utilizará esa fecha.) En la pantalla aparece el indicador de comandos A.
3. Inserte el Data Management Disk de Forecast Plus en la unidad A.
4. Inserte un disco formateado en la unidad B.
5. Teclee FP.
6. En la pantalla aparecerá el menú principal de Forecast Plus.

El menú principal

Forecast Plus está dividido en cuatro componentes. El menú principal de selección se ve de la siguiente manera:

1. Data Management (Administración de datos)
2. Exploratory Package (Paquete de exploración)
3. Forecasting Analysis (Análisis de pronósticos)
4. End of the Program (Fin del programa)

Puede usted elegir cualquiera de las cuatro secciones tecleando el número correspondiente y presionando <Enter>. Si desea terminar el programa Forecast Plus y regresar al sistema operativo (DOS), teclee el número cuatro (4). Forecast Plus terminará y el sistema regresará a DOS.

PRONÓSTICO MEDIANTE HOJAS DE CÁLCULO

Los paquetes de hojas de cálculo como Lotus 1-2-3, Excel y Quattro Pro se emplean en la solución de diversas aplicaciones en los negocios. Debido a que se almacena una gran cantidad de datos en las hojas de cálculo, es necesario que aprendamos cómo usarlas de manera eficiente en el proceso de pronóstico.

En muchos de nuestros ejemplos usaremos Lotus 1-2-3, ya que la mayoría de los archivos de Lotus se pueden leer en otros paquetes de pronóstico. La pantalla principal de 1-2-3 (que se muestra en la fig. 1.1) contiene una hoja de cálculo tabular que se utiliza para almacenar y organizar datos.

La *hoja de cálculo* consiste en una cuadrícula compuesta de columnas y renglones que se llenan con etiquetas y números. A la intersección de cada columna y renglón se le denomina *celda*. Cada celda representa una posición única para almacenar una porción de

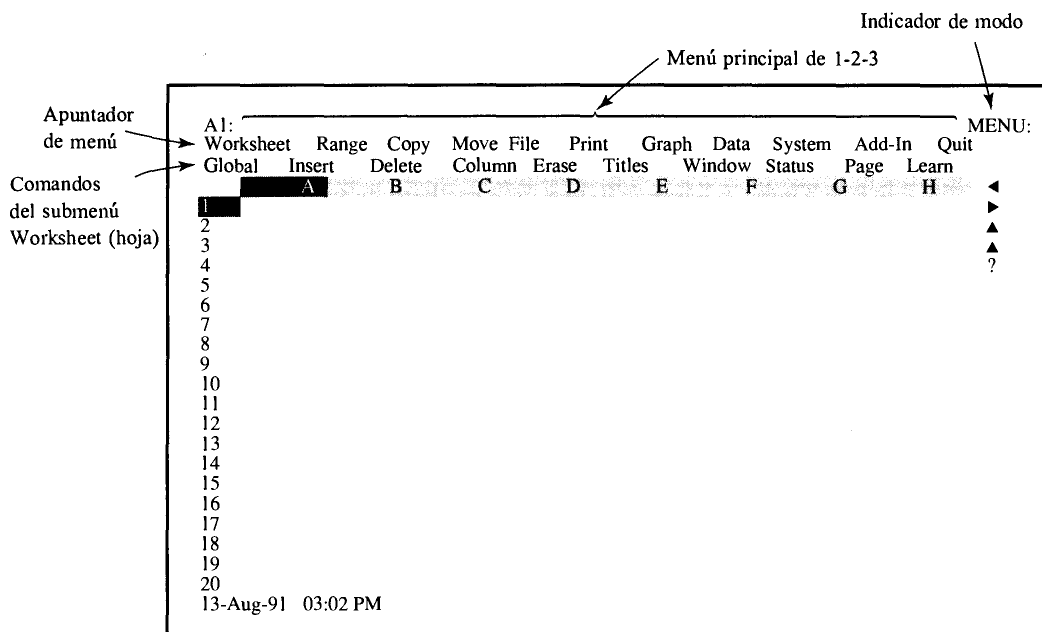


Figura 1.1 Hoja de cálculo Lotus 1-2-3

datos. Las celdas se identifican mediante la letra de la columna seguida del número del renglón. Por ejemplo, a la celda ubicada en la intersección de la columna A y el renglón 1 se le denomina A1. Observe en la figura 1.1 que la celda actual A1 está resaltada y que el apuntador de celda está ubicado ahí. Además, en la primera línea se indica que la dirección de la celda actual es A1.

El *panel de control* es un área de tres líneas en la parte superior de la pantalla, que presenta información sobre la celda actual y el menú principal.

Para activar el *menú principal*, presione la tecla / (diagonal). El menú principal de 1-2-3, a partir del cual se introducen todos los comandos, se exhibirá en la segunda línea (véase fig. 1.1). Una barra rectangular, denominada *apuntador de menú*, aparece resaltando al comando Worksheet (hoja). Los comandos posibles son:

Worksheet (Hoja)	Comandos que afectan a toda la hoja
Range (Rango)	Comandos que afectan partes específicas de la hoja
Copy (Copiar)	Copia una celda o un rango de celdas
Move (Mover)	Desplaza una celda o un rango de celdas
File (Fichero)	Comandos que afectan a archivos almacenados o por almacenar en disco
Print (Imprimir)	Imprime una hoja de cálculo
Graph (Gráfico)	Comandos que exhiben en forma gráfica datos de la hoja de cálculo
Data (Datos)	Comandos para manipulación especial de datos
System (Sistema)	Regresa temporalmente al sistema operativo (DOS) (sólo en la versión 2)
Quit (Salir)	Termina la sesión de 1-2-3 y regresa al sistema principal

La tercera línea es un *submenú* que enlista los comandos disponibles para cualquiera de los comandos del menú principal que esté resaltado. La figura 1.1 muestra los comandos de submenú disponibles a través del comando principal Worksheet (Hoja).

Lotus utiliza la tecla de función F1 como tecla de ayuda. Siempre que tenga dudas sobre un comando, oprima F1.

La sección “Pronósticos mediante hojas de cálculo” al final del capítulo 3, muestra cómo introducir datos en una hoja de cálculo. Las secciones al final de la mayoría de los capítulos muestran cómo utilizar las hojas de cálculo para pronosticar.

INTRODUCCIÓN AL PAQUETE DE CÓMPUTO MINITAB

Minitab es uno de los paquetes de software más populares que se utilizan para el análisis estadístico. Este paquete, disponible tanto para macro como para microcomputadoras, es uno de los paquetes de cómputo a los que se hace referencia a lo largo del libro. Se supone que se ejecutó Minitab para cada ejemplo de Minitab que se presenta en el texto. Al tener éxito una función, aparece en la pantalla el indicador de comandos MTB >. El tipo de información que Minitab espera que usted teclee, le será solicitado por el indicador de comandos que aparezca en el extremo izquierdo de la pantalla.

Minitab ha sido reconocido por su estructura de comandos directos y esto es lo que se presentará a lo largo del texto. No obstante, para mejorar la facilidad de uso, Minitab tiene una nueva versión para microcomputadoras denominada Release 9 (Versión 9), la cual presenta menús y cuadros de diálogo, manteniendo el lenguaje de comandos para agregar flexibilidad y velocidad.

La comprensión de la hoja de cálculo Minitab es la clave para utilizar este programa. La hoja contiene columnas (variables) y renglones (observaciones). A las columnas se les nombra C1, C2, etc., y se les puede asignar también nombres de variables como ventas, precio o peso. Los ejemplos de comandos a discutir se referirán a una hoja de cálculo típica de nombre EXAMPLE.XXX, de la manera siguiente:

		COLUMNAS O VARIABLES					
		C1	C2	C3	C4	C5	C6
	1						
Renglones	2						
u	3						
observaciones	4						
	5						
	6						
	.						
	.						
	.						

Como ejemplo, se presentan los comandos de Minitab para resolver el siguiente caso: considérense los siguientes valores de X - Y .

X	Y
5	4
7	9
3	8
0	5
10	3
2	9

1. Encontrar las siguientes sumas: ΣX , ΣY , ΣXY .
2. Mostrar que ΣX^2 no es igual a $(\Sigma X)^2$

```
MTB > OUTFILE 'EXAMPLE.PRN'
MTB > SET C1
DATA > 5 7 3 0 10 2
DATA > END
MTB > SET C2
DATA > 4 9 8 5 3 9
DATA > END
MTB > LET C3 = C1*C2
MTB > PRINT C1-C3
```

Fila	C1	C2	C3
1	5	4	20
2	7	9	63
3	3	8	24
4	0	5	0
5	10	3	30
6	2	9	18

```
MTB > LET K1 = SUM(C1)
MTB > LET K2 = SUM(C2)
MTB > LET K3 = SUM(C3)
MTB > PRINT K1-K3
K1      27.0000
K2      38.0000
K3      155.000
MTB > WRITE 'EXAMPLE.DAT' C1-C3
MTB > STOP
```

- El comando OUTFILE se utiliza para almacenar todos los comandos y resultados que se muestran en la pantalla en el archivo especificado con el nombre EXAMPLE.PRN, hasta terminar la sesión de Minitab con el comando STOP.
- El comando SET se emplea para introducir datos en la columna 1 (C1) de la hoja de cálculo.
- El subcomando DATA se usa para introducir los datos, dejando por lo menos un espacio entre los valores de los datos.
- El subcomando END se utiliza para indicar que está completo el registro de datos para la columna 1.
- El comando LET se emplea para realizar cálculos aritméticos mediante expresiones algebraicas como multiplicar C1 por C2 y almacenar el resultado en C3. También se puede utilizar para modificar un archivo de datos. Por ejemplo, cambiar el valor del renglón 6, columna 2, de 9 a 64.

```
MTB > LET C2 (6) = 64
```

- El comando PRINT se usa para imprimir el contenido de las columnas 1 a 3.
- El comando LET se emplea para sumar (SUM) la columna C1 y almacenar el resultado en el renglón K1.
- El comando WRITE guarda los datos en una hoja de cálculo de nombre EXAMPLE.DAT. Para su uso futuro, la hoja podrá siempre accesarse mediante el comando READ EXAMPLE.DAT C1-C3.

INTRODUCCIÓN AL PAQUETE DE CÓMPUTO SAS

El paquete Statistical Analysis System (SAS) (Sistema de Análisis Estadístico) es uno de los paquetes de software más populares que se emplean en la industria y los negocios para el análisis estadístico. Disponible tanto para macro como para microcomputadoras, éste es uno de los paquetes de cómputo a los que nos referiremos a lo largo del libro.

Para resolver un problema estadístico con SAS, se deben primero definir los datos y después especificar los procedimientos estadísticos (PROCS) a utilizar. Al ejecutar programas SAS se deben observar ciertas reglas. Algunas de las reglas básicas de SAS son las siguientes:

1. Las declaraciones de SAS deben terminar siempre con un punto y coma (;).
2. Los nombres deben comenzar con una letra y tener ocho o menos caracteres.
3. Las declaraciones en SAS terminan en la columna 72, las que excedan dicha columna se deben continuar en la siguiente línea.

4. Se deben definir los datos antes de ejecutar un procedimiento. Las líneas que contienen datos no terminan con un punto y coma, pero es recomendable incluir el punto y coma al final de la última línea de datos.
5. Se pueden realizar diversos procedimientos agregando declaraciones de procedimiento múltiple.
6. Se debe incluir el punto decimal al introducir datos con decimales (1.89).
7. En un programa SAS se deben incluir declaraciones de tipo JCL (lenguaje de control de trabajos, por sus siglas en inglés), las cuales son específicas para un sistema en particular, ya que permiten que el programa SAS se ejecute correctamente.

El siguiente ejemplo muestra el uso de comandos SAS. Considérense los siguientes valores de X - Y .

X	Y
5	4
7	9
3	8
0	5
10	3
2	9

1. Encontrar las siguientes sumas: ΣX , ΣY , ΣXY .
2. Demostrar que ΣX^2 no es igual a $(\Sigma X)^2$.

```

$ SAS
OPTIONS PAGESIZE=60 LINESIZE=80;
TITLE "SAS SOLUTION TO EXAMPLE 0";
DATA EXAMPLE;
  INPUT X Y;
  XY=X*Y;
CARDS;
5 4
7 9
3 8
0 5
10 3
2 9
PROC PRINT DATA=EXAMPLE;
  SUM X Y XY;
ENDSAS;

```

- La declaración OPTIONS indica que el tamaño de página de la salida será de 60 líneas y que cada línea tendrá 80 espacios.
- La declaración TITLE da nombre al programa SAS.
- La declaración DATA proporciona el nombre del archivo de datos.
- La declaración INPUT nombra las variables e indica el orden correcto de los distintos campos en las líneas de datos. Nótese que se leerán dos variables (X y Y), y que se creará una nueva variable multiplicando éstas ($XY=X*Y$).
- La declaración CARDS indica a SAS que a continuación se presentan los datos de entrada. Las siguientes seis líneas contienen los valores de los datos.
- La declaración PROC PRINT produce un listado de los datos en el archivo EXAMPLE.
- La declaración SUM especifica las variables a totalizar.

La primera línea es una declaración de tipo JCL. El resultado de la ejecución de este programa SAS se muestra en la tabla 1.1.

Tabla 1.1 Salida de SAS para el ejemplo

OBS	X	Y	XY
1	5	4	20
2	7	9	63
3	3	8	24
4	0	5	0
5	10	3	30
6	2	9	18
	<u>27</u>	<u>38</u>	<u>155</u>

BIBLIOGRAFÍA

- ADAMS, F. G. *The Business Forecasting Revolution*. Nueva York: Oxford University Press, 1986.
- BEAUMONT, C., MAHMOUD, E. y MCGEE, V. E. "Microcomputer Forecasting Software: a Survey." *Journal of Forecasting*, 4 (1985): 305-311.
- GEORGOFF, D. M. y MARDICK, R. G. "Manager's Guide of Forecasting." *Harvard Business Review*, 1 (1986): 110-120.
- HOGARTII, R. M. y MAKRIDAKIS, S. "Forecasting and Planning: An Evaluation." *Management Science*, 27 (2) (1981): 115-138.
- KEATING, K. "Choosing and Using Business Forecasting Software." *Creative Computing*, 11 (1) (1985): 119-135.
- MAKRIDAKIS, S. "The Art and Science of Forecasting". *International Journal of Forecasting*, 2 (1986): 15-39.

- MAKRIDAKIS, S. "Forecasting in the 21st Century." *International Journal of Forecasting*, 7 (2) (1991): 123-126.
- MAHMOUD, E., RICE, G., McGEE, V. E. y BEAUMONT, C. "Mainframe Multipurpose Forecasting Software: A Survey." *Journal of Forecasting*, 5 (1986): 127-137.
- MAHMOUD, E., RICE, G., McGEE, V. E. y BEAUMONT, C. "Mainframe Specific Purpose Forecasting Software: A Survey." *Journal of Forecasting*, 5 (1986): 75-83.
- REID, R. A. "The Forecasting Process: Guidelines for the MIS Manager." *Journal of Systems Management* (noviembre, 1986): 33-37.
- WRIGHT, G. y AYTON, P. eds. *Judgemental Forecasting*. Nueva York: John Wiley & Sons, 1987.

Revisión de conceptos estadísticos básicos

ACTIVIDAD

La mayoría de las técnicas de pronóstico se basa en conceptos estadísticos fundamentales que constituyen el tema de textos sobre estadística en los negocios, así como de cursos de introducción a la estadística. En este capítulo se revisan algunos de los conceptos básicos que servirán de fundamento para gran parte del material del resto del texto.

La mayoría de los procesos estadísticos hacen inferencias sobre los elementos de interés denominados *población*, después de seleccionar y medir un subgrupo de estos elementos llamado *muestra*. El cuidado en la selección de una muestra representativa y la utilización de un tamaño de muestra lo suficientemente grande, son componentes importantes de un proceso de inferencia estadística con un grado de riesgo aceptablemente bajo.

Junto con la inferencia estadística, la *estadística descriptiva* constituye la otra gran subdivisión de la estadística. Los procedimientos de estadística descriptiva se emplean para resumir o describir en forma breve conjuntos de datos, de manera que sus componentes esenciales se hagan obvios y se puedan utilizar en el proceso de toma de decisiones.

ESTADÍSTICA DESCRIPTIVA

La finalidad de los procedimientos de estadística descriptiva consiste en describir brevemente un gran conjunto de mediciones con unos cuantos valores clave resumidos. La forma más común de lograr esto es promediando los valores. En estadística, el proceso de promediar se lleva a cabo regularmente mediante el cálculo de la *media*, lo que comprende sumar todos los valores y dividir entre el número de valores. Debido a que tanto la población como la muestra que se extrae de ella poseen una media, existen dos fórmulas para calcular ésta. La fórmula apropiada depende de que los valores reunidos constituyan todos los valores de interés (la población) o un conjunto parcial de ellos (una muestra).

Las *estadística descriptiva* comprende la descripción de conjuntos de datos por medio de algunos valores clave resumidos.

Por lo regular, los valores de la población o parámetros se identifican utilizando letras griegas. El símbolo elegido para la población es la letra griega μ (mu). La fórmula para la media de la población es

$$\mu = \frac{\Sigma X}{N} \quad (2.1)$$

en donde

ΣX = la suma de todos los valores de la población

N = el tamaño de la población

El símbolo para la media de la muestra (media muestral) es $\bar{X}$ (X-barra) y se calcula utilizando la ecuación 2.2:

$$\bar{X} = \frac{\Sigma X}{n} \quad (2.2)$$

en donde

ΣX = la suma de todos los valores de la muestra

n = el tamaño de la muestra

Además de medir la tendencia central de un grupo de valores mediante el cálculo de la media, por lo regular es de interés el grado en el que los valores se encuentran dispersos alrededor de la media. Para este propósito se puede calcular la *desviación estándar* de cualquiera de ambos grupos. Se puede pensar en la desviación estándar como en la diferencia típica o estándar entre los valores de un grupo y su media. Las fórmulas siguientes corresponden a la desviación estándar poblacional (σ) y muestral (s):

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}} \quad (2.3)$$

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}} \quad (2.4)$$

en donde los numeradores representan la suma de las diferencias cuadradas entre los valores medidos y sus medias.

Por último, muchos procedimientos estadísticos hacen uso de la *varianza* poblacional o de la varianza muestral. La varianza de un conjunto de mediciones es la desviación estándar al cuadrado. De ahí que la varianza poblacional (σ^2) y de la muestral (s^2) se calculen como:

$$\sigma^2 = \frac{\sum(X - \mu)^2}{N} = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N} \quad (2.5)$$

$$s^2 = \frac{\sum(X - \bar{X})^2}{n - 1} = \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1} \quad (2.6)$$

Ejemplo 2.1

Considérese el siguiente conjunto de edades de personas, en donde se considera que los valores son una muestra de la población, en vez de la población misma:

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

Para esta muestra,

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

Como se muestra en la tabla 2.1, la media muestral es 40.2 años, la varianza muestral es 148.84 y la desviación estándar muestral es 12.2 años. Si el conjunto de datos originales se hubiera identificado como una población en vez de una muestra, los cálculos habrían sido los mismos, excepto por el denominador en el cálculo de la varianza y la desviación estándar, que sería $10(N)$ en lugar de $9(n-1)$. Al término $n-1$ se le conoce como grados de libertad.

El término *grados de libertad* se utiliza para indicar el número de elementos de datos que son libres de otros en el sentido de que no pueden deducirse de otros, y por tanto,

TABLA 2.1 CÁLCULO DE S ($\bar{X} = 40.2$)

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
		$\sum(X - \bar{X})^2 = 1,339.60$
		$s^2 = \frac{1,339.6}{10 - 1} = 148.84$
		$s = \sqrt{148.84} = 12.2$

constituyen piezas únicas de información. Por ejemplo, suponga que se hacen las siguientes declaraciones:

Pienso en el número 5. Pienso en el número 7. La suma de los dos números en que pensé es 12.

A primera vista, se presentan tres piezas de información. No obstante, si se conocen dos de estas declaraciones, la otra puede deducirse. Se podría decir que en las tres declaraciones sólo hay dos piezas únicas de información o, para usar el término estadístico, que sólo hay dos grados de libertad.

En el ejemplo que se presenta en la tabla 2.1, las edades de las diez personas constituyen una muestra con diez grados de libertad. Cualquier edad podría haberse incluido en la muestra, de ahí que cada una de las edades sea libre de variar. Al calcular la media, se usan las diez edades para contabilizar una edad media total de 40.2 años.

En la desviación estándar muestral, el cálculo difiere. Al calcular la desviación estándar muestral, se utiliza una estimación de la media de la población (la media muestral $\bar{X}$). Se introduce un sesgo debido a que el valor $\Sigma(X - \bar{X})^2$ es un valor mínimo para cualquier distribución dada. Si a cada elemento de la distribución se hubiera restado cualquier otro valor distinto de 40.2 y estas diferencias se elevaran al cuadrado y se sumaran, la suma total de cuadrados sería mayor de 1 339.6. Al utilizar en el cálculo la media muestral como una estimación de la media poblacional, por lo regular se obtendrá una desviación estándar menor que la desviación estándar poblacional. Sin embargo, este sesgo se puede corregir dividiendo el valor de $\Sigma(X - \bar{X})^2$ entre el número adecuado de grados de libertad. Ya que la media muestral se usó como una estimación de la media poblacional en el cálculo de la desviación estándar muestral, sólo nueve de las edades son libres de variar. Si se conocen nueve edades, se puede contabilizar la décima, ya que $\Sigma(X - \bar{X})$ debe ser igual a cero. Sólo se requiere de nueve edades para contabilizar la totalidad de la información. En general, siempre que se utilice en un cálculo un estadístico de la muestra como una estimación de un parámetro de la población, se pierde un grado de libertad (véase tabla 2.2.).

Los *grados de libertad* en un conjunto de datos indican el número de elementos de datos que son independientes de los otros y que se constituyen como piezas únicas de información.

TABLA 2.2 CARACTERÍSTICAS DE LA POBLACIÓN Y DE LA MUESTRA

CARACTERÍSTICA	POBLACIÓN	MUESTRA
Media	μ	$\bar{X}$
Varianza	σ^2	s^2
Desviación estándar	σ	s

DISTRIBUCIONES DE PROBABILIDAD

Se denomina *variable aleatoria* a aquella cantidad que puede tomar diferentes valores entre una prueba y otra de un experimento, el resultado exacto de una oportunidad o evento aleatorio. Si sólo son posibles ciertos valores específicos, la variable aleatoria se denomina *variable discreta*. Ejemplos de ello son el número de habitaciones de una casa, el número de personas que llegan a la caja de un supermercado a cierta hora y el número de unidades defectuosas en un lote de partes electrónicas. Si es posible que la variable aleatoria tome cualquier valor dentro de cierto rango, se le denomina *variable continua*. Ejemplos de este tipo de variable son los pesos de las personas, la longitud de una pieza manufacturada y el tiempo entre llegadas de automóviles a una caseta de cobro.

Una *variable aleatoria discreta* puede suponer sólo ciertos valores específicos, por lo regular enteros; un *variable aleatoria continua* puede suponer cualquier valor numérico dentro de algún rango.

Si se considera una variable aleatoria, su *distribución de probabilidad* enlista todos los valores posibles que puede tomar, junto con la probabilidad de cada uno.

Ejemplo 2.2 El número de días sin venta de un vendedor durante un mes se podría describir mediante la distribución de probabilidad que se muestra en la tabla 2.3. Estos valores se basan en la experiencia pasada del vendedor y se usarán para pronosticar la actividad de ventas en el futuro. La columna X enlista todos los valores (días sin venta) posibles, mientras que la columna $P(X)$ enlista las probabilidades correspondientes. Nótese que ya que se enlistan todos los valores posibles de X , las probabilidades suman 1.00 o 100%. Esto es correcto para todas las distribuciones de probabilidad, ignorando los errores de redondeo.

El *valor esperado* de una variable aleatoria es el valor promedio que supone la variable después de numerosas pruebas. El valor esperado de una variable con la distribución de

TABLA 2.3 DISTRIBUCIÓN DE LA PROBABILIDAD

X	$P(X)$
1	.10
2	.20
3	.25
4	.15
5	.30
$\Sigma P(X)_c = 1.00$	

probabilidad discreta se puede encontrar multiplicando cada posible valor de X por su probabilidad y sumando después estos productos. La ecuación 2.7 muestra este cálculo:

$$E(X) = \sum [X \cdot P(X)] \quad (2.7)$$

Para la distribución de probabilidad dada en la Tabla 2.3, el valor esperado es

$$\begin{aligned} E(X) &= (1)(.10) + (2)(.20) + (3)(.25) + (4)(.15) + (5)(.30) \\ &= 3.35 \end{aligned}$$

Por lo tanto, si se observó a este vendedor por un periodo largo de meses y se registraron los días sin ventas, el promedio sería 3.35.

El *valor esperado* de una variable aleatoria es el valor promedio de la variable en muchas pruebas u observaciones.

Para una distribución continua la probabilidad de obtener un valor específico se aproxima a cero. Por ejemplo, la probabilidad de que alguien pese 68 kg se podría considerar cero, ya que esto significaría que este peso es exactamente 68.000, sin importar que tan precisa sea la escala utilizada. Las distribuciones continuas se manejan mediante la localización de la probabilidad de que un valor quede en algún intervalo al extraerlo en forma aleatoria de la distribución. Por ejemplo, se podría calcular la probabilidad de que el peso de una persona quedara en el intervalo de 63 a 68 kg.

Algunas distribuciones teóricas suceden una y otra vez en aplicaciones estadísticas prácticas, y por esta razón es importante examinar sus propiedades y aplicaciones. Una de estas distribuciones importantes es la *distribución binomial*, que con frecuencia se usa para representar una variable aleatoria discreta. Los requisitos para un experimento binomial son:

1. Existen n pruebas idénticas, que tienen sólo uno de dos resultados posibles.
2. La probabilidad de éxito de cada resultado permanece fija entre prueba y prueba.
3. Las pruebas son independientes.

El interés está en encontrar la probabilidad de x ocurrencias exitosas en las n pruebas, en donde una ocurrencia se define de manera arbitraria como uno de dos resultados posibles. Los diversos valores de x , junto con sus probabilidades, forman la distribución binomial. Estas probabilidades se pueden obtener por medio de la siguiente fórmula binomial:

$$P(x) = C_x^n \pi^x (1 - \pi)^{n-x} \quad \text{para } x = 0, 1, 2, \dots, n \quad (2.8)$$

en donde

C_x^n = el número de combinaciones de n objetos tomados x a la vez

π = la probabilidad de éxito en cada prueba

x = el número particular de éxitos de interés

n = el número de pruebas

Una forma más fácil de obtener probabilidades binomiales consiste en hacer referencia a una tabla de distribución binomial, como la que se encuentra en el Apéndice C, tabla C.1. Las probabilidades están agrupadas en bloques que representan a n y tiene columnas con p (o π) como encabezado y renglones indicados con x .

Ejemplo 2.3 Suponga que se extraen ocho elementos de una línea de producción de la que se sabe produce partes defectuosas el 5% del tiempo. ¿Cuál es la probabilidad de obtener exactamente cero elementos defectuosos? La respuesta, a partir de la tabla binomial del Apéndice C es .6634 (aquí $n = 8$, $\pi = .05$, $x = 0$).

La *distribución normal* es una importante distribución continua, ya que por medio de ella, se pueden aproximar muchas poblaciones útiles. Para identificar la distribución normal se requiere conocer la media y la desviación estándar. Una curva normal es simétrica y en forma de campana, como la que se muestra en la figura 2.1. Esta distribución representa muchas variables de la vida real que se miden en una escala continua.

Las probabilidades de que los valores extraídos de una distribución normal queden en varios intervalos, se encuentra convirtiendo primero todos los intervalos a unidades estándar denominadas valores estandarizados. El valor estandarizado de cualquier valor de X es el número de desviaciones estándar desde el valor central de la curva (μ) a ese valor. De ahí que la fórmula sea:

$$Z = \frac{X - \mu}{\sigma} \quad (2.9)$$

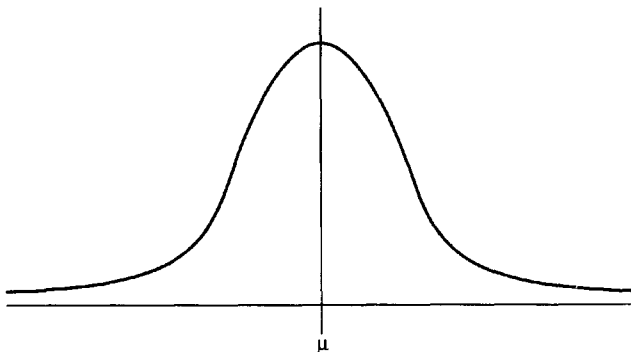


Figura 2.1 Distribución normal

en donde

X = el valor del interés

μ = la media

σ = la desviación estándar

Después de calcular el valor estandarizado, se puede consultar la tabla de la curva normal para encontrar un área bajo la curva, entre el centro de la curva (μ) y el valor de interés X .

Ejemplo 2.4

Una población formada con los pesos de piezas que elabora cierta máquina, tiene una distribución normal con una media de 10 libras y una desviación estándar de 2 libras. ¿Cuál es la probabilidad de que una pieza elegida al azar pese entre 9 y 12 libras? En la figura 2.2 se muestra la curva normal con el área correspondiente sombreada.

Ya que las tablas de curvas normales están diseñadas para dar áreas desde el centro de la curva hasta cierto punto, se deben localizar dos áreas separadas: una a cada lado de la media y sumarlas juntas. Este proceso producirá la probabilidad de que un valor quede en este intervalo. Los dos valores estandarizados son:

$$z_1 = \frac{9 - 10}{2} = -.50$$

$$z_2 = \frac{12 - 10}{2} = 1.00$$

Se ignora el signo negativo del primer valor estandarizado, ya que simplemente indica que el valor de interés de X , (9) es menor que la media de la curva (10). Estos dos valores estandarizados se llevan a la tabla de curva normal del Apéndice C, tabla C2, para encontrar las dos áreas, mismas que se suman

$$z_1 = -.50 \rightarrow .1915$$

$$z_2 = 1.00 \rightarrow .3413$$

$$.5328$$

Se concluye que existe una posibilidad de alrededor del 53%, de que una pieza elegida al azar en esta población pese entre 9 y 12 libras.

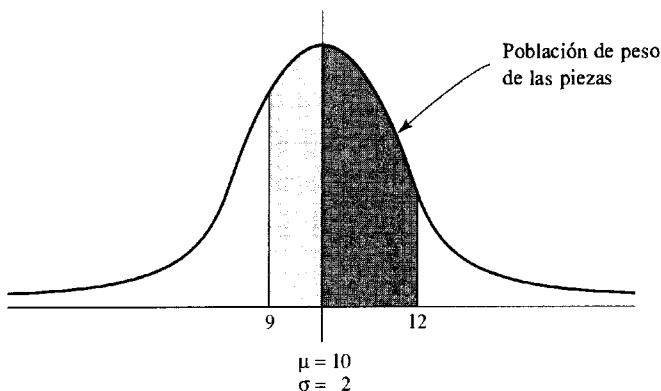


Figura 2.2 Área de la curva normal

Por último, la *distribución t* se utiliza con frecuencia en pruebas estadísticas cuando se emplea una muestra pequeña y cuando se puede suponer que la población que se investiga tiene una distribución normal. La tabla C.3 del Apéndice C, muestra valores tomados de la distribución t . Nótese que sólo se requiere especificar un valor antes de consultar la tabla, a saber, los grados de libertad [abreviado df (degrees of freedom)]. Una vez conocidos los grados de libertad, se pueden encontrar los valores de t que excluyen de la curva los porcentajes deseados. Por ejemplo, si la distribución t de interés tiene 14 grados de libertad, un valor t de 2.145 a cada lado del centro de la curva comprenderá al 95% de ésta, excluyendo el 5% restante. Más adelante en este capítulo se presentará un ejemplo del uso de la distribución t .

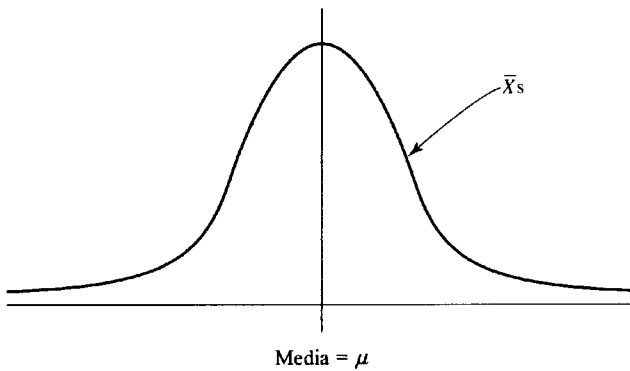
DISTRIBUCIONES MUESTRALES

En la mayoría de las aplicaciones estadísticas se toma de la población que se investiga, una muestra aleatoria, se calcula un estadístico de los datos de la muestra y se formulan conclusiones para la población con base en esta muestra. Una *distribución muestral* es la distribución de todos los posibles valores del estadístico de la muestra, que se pueden obtener de la población para un determinado tamaño de muestra. Por ejemplo, se podría tomar de una población una muestra aleatoria de 100 personas y pesarlas para calcular después la media de su peso. Se puede pensar en esta media muestral $\bar{X}$, como si se hubieran extraído de la distribución de todas las medias muestrales posibles, para muestras de tamaño 100 que pudieran tomarse de la población. En forma similar, cada estadístico muestral que pueda ser calculado, a partir de datos muestrales, puede considerarse como que ha sido extraído de una distribución muestral.

Una *distribución muestral* es el arreglo de todas los estadísticos posibles de la muestra que pueden extraerse de una población, para un tamaño de muestra dado.

El *teorema del límite central* establece que, al hacerse más grande la muestra, la distribución de muestreo de las medias de la muestra tiende hacia la distribución normal, y que la media de esta distribución normal es, la media de la población y que la desviación estándar es $\sigma\sqrt{n}$ (a este valor se le conoce como *error estándar de la distribución muestral*). Esta distribución muestral tenderá hacia la normalidad, sin importar la forma de la distribución de la población de la que se extrajeron los elementos de la muestra. La figura 2.3 muestra cómo luciría dicha distribución muestral.

El teorema del límite central es de particular importancia en las estadísticas, ya que permite al analista calcular la probabilidad de diversos resultados de muestras a través del conocimiento de las probabilidades de la curva normal.

Figura 2.3 Distribución muestral de $\bar{X}$ **Ejemplo 2.5**

¿Cuál es la probabilidad de que la media de una muestra aleatoria de 100 pescos, extraídos de una población, se encuentre en un margen de dos libras de la media real del peso de la población, si la desviación estándar de la población se estima en 15 libras? La figura 2.4 muestra la distribución apropiada.

El error estándar $\sigma/\sqrt{n} = 15/\sqrt{100} = 1.5$. Luego $z = 2/1.5 = 1.33$, con lo que se obtiene un área de .4082 en la tabla de la curva normal. Duplicando esta área para contabilizar ambos lados de la media, da por resultado un área de .8164.

Las probabilidades de que la media muestral esté en un margen de 2 libras de la media real, es de alrededor del 82%, reflejando un tamaño de muestra de 100 y la variabilidad estimada de la población $\sigma = 15$. Como se verá, esta capacidad para calcular probabilidades de resultados de muestras permitirá al analista llegar a útiles resultados en la estimación y en la prueba de hipótesis.

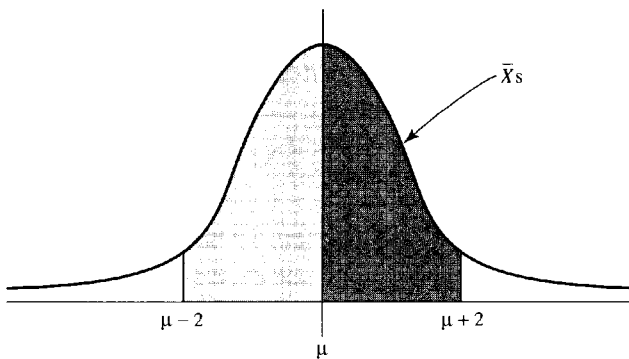


Figura 2.4 Área de la distribución muestral

ESTIMACIÓN

Un *punto estimado* de un parámetro de una población, es un valor individual calculado a partir de datos de la muestra, que estima el valor desconocido de la población. La tabla 2.4 contiene una lista de varios parámetros poblacionales y los estadísticos de la muestra que proporcionan puntos de estimación correspondientes.

Un *intervalo de estimación* o un *intervalo de confianza* es un intervalo dentro del cual es probable que se ubique el parámetro de interés de la población. Se encuentra formando un intervalo alrededor del punto de estimación y se calcula ya sea mediante la distribución normal o la distribución *t*. En ocasiones se utilizan el punto de estimación y el intervalo de

TABLA 2.4 VALORES POBLACIONALES Y MUESTRALES

PARAMÉTROS DE LA POBLACIÓN	ESTADÍSTICO DE LA MUESTRA (ESTIMACIÓN)
Media (μ)	$\bar{X}$, media muestral
Desviación estándar (σ)	s , desviación estándar muestral
Varianza (σ^2)	s^2 , varianza muestral
Proporción (π)	p , proporción de la muestra

estimación como estimaciones aproximadas de valores importantes. (Véase el caso de estudio Mr. Tux al final de este capítulo.)

Ejemplo 2.6

Se desea obtener un intervalo de estimación de la media de una población. El tamaño de la muestra es 400, la media muestral (el punto de estimación σ) es de 105 libras y la desviación estándar muestral (una estimación de σ) es de 10 libras. ¿En qué intervalo es un 95% probable que μ se ubique? Se puede encontrar formando un intervalo alrededor de $\bar{X}$, utilizando desviaciones estándar de 1.96 en la distribución de muestreo de $\bar{X}$ (a 1.96 desviaciones estándar a cada lado de la media de cualquier curva normal abarcará aproximadamente el 95% de los valores). Los cálculos son

$$\bar{X} \pm Z \frac{s}{\sqrt{n}} \quad (2.10)$$

$$105 \pm 1.96 \frac{10}{\sqrt{400}}$$

$$105 \pm .98$$

$$104.02 \text{ a } 105.98$$

Así, es 95% probable que la media muestral fuera elegida de una distribución muestral tal que la media real de la población se encuentre en el intervalo de 104.02 a 105.98.

En forma similar, se pueden formar intervalos de confianza alrededor de puntos de estimación de la proporción de la población, utilizando la siguiente fórmula

$$p \pm Z \sqrt{\frac{\pi(1-\pi)}{n}} \quad (2.11)$$

En la práctica la proporción de la población, π , se estima por lo regular con p , la proporción de la muestra, en el cálculo de la desviación estándar de la distribución muestral (error estándar de la proporción).

Ejemplo 2.7

Un intervalo de confianza del 99% para el porcentaje real π de la población, que sea familiar con un producto, se puede formar alrededor del punto de estimación $\pi = .23$ con $n = 150$.

$$.23 \pm 2.58 \sqrt{\frac{(.23)(.77)}{150}}$$

$$.23 \pm .089$$

$$.141 \text{ to } .319$$

En este ejemplo el punto de estimación de π es .23. Dado un tamaño de muestra de 150, es 99% probable que la muestra fuera extraída de tal forma que el valor real de π se encuentra en el intervalo de .141 a .319.

Un *punto de estimación* es una estimación valuada individualmente de un parámetro de la población; un *intervalo de estimación* es un intervalo numérico dentro del cual es probable que se ubique el parámetro de la población.

PRUEBA DE HIPÓTESIS

En muchas situaciones estadísticas el interés se centra en alguna afirmación sobre la población, en vez de en la estimación de algunos de sus parámetros. A este procedimiento se le denomina *prueba de hipótesis* y comprende los siguientes pasos:

- Paso 1 Establezca la hipótesis que se desea probar (llamada la hipótesis nula con el símbolo H_0) y establezca la hipótesis alterna (a la que se llega en conclusión si se descarta H_0 , el símbolo es H_1).
- Paso 2 Suponga que la hipótesis nula es cierta y consulte la distribución muestral de la cual se extrajeron las estadísticas de la muestra bajo esta suposición.
- Paso 3 Reúna una muestra aleatoria de elementos de la población, médalos y calcule las estadísticas correspondientes a la muestra.
- Paso 4 Calcule la probabilidad de que tal estadística de muestra pudiera haber sido extraída de esta distribución muestral.
- Paso 5 Si esta probabilidad es alta, no rechace la hipótesis nula; si esta probabilidad es baja, se puede rechazar la hipótesis nula con una baja probabilidad de error.

TABLA 2.5 RESULTADOS DE LA PRUEBA DE HIPÓTESIS

		No rechazar H_0	Rechazar H_0
Estado de la naturaleza	H_0 verdadera	Decisión correcta	Error tipo I, probabilidad: α
	H_0 falsa	Error tipo II, probabilidad: β	Decisión correcta

Al seguir estos pasos, pueden presentarse dos tipos de error, como se muestra en la tabla 2.5. Siempre se espera alcanzar la decisión correcta con respecto a la hipótesis nula, después de examinar la evidencia de la muestra, pero siempre existe la posibilidad de rechazar una H_0 verdadera y de no rechazar una H_0 falsa. Las probabilidades de estos eventos se conocen como alfa (α) y beta (β), respectivamente. También se conoce a alfa como el *nivel de significancia* de la prueba.

Ejemplo 2.8

Se desea probar la hipótesis de que el peso promedio de las piezas que produce cierta máquina se mantiene en 50 libras, que es el peso promedio de las piezas en los últimos años. Se toma una muestra aleatoria de 100 piezas. Se supone que la desviación estándar del peso de las piezas es de 5 libras sin importar la media del peso, ya que este valor ha permanecido constante en estudios pasados. Si se asume que la hipótesis nula es correcta (la hipótesis nula establece que la media del peso de las piezas es de 50 libras), la distribución de muestreo apropiada es la distribución normal, de acuerdo con el teorema del límite central. La prueba se demuestra en la figura 2.5.

La regla de decisión de esta prueba es:

Si la media del peso de la muestra está entre 49.02 libras y 50.98 libras, no se rechaza la hipótesis nula. Si está fuera de este intervalo, se rechaza.

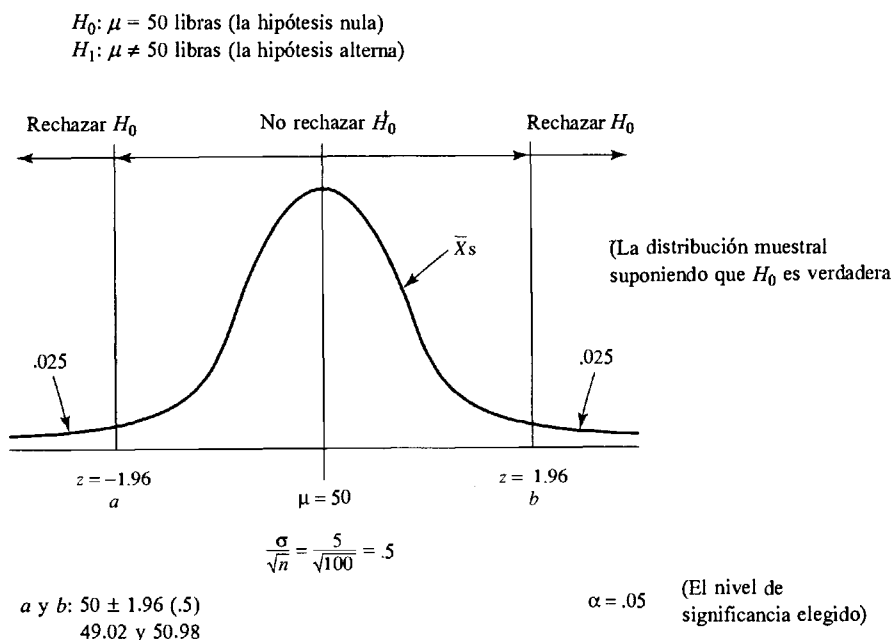


Figura 2.5 Prueba de hipótesis - ejemplo 2.8

Al establecer de esta forma la regla de decisión, la probabilidad de rechazar la hipótesis nula cuando es verdadera, es decir, cometer un error de Tipo I, es de 5% ($\alpha = .05$). Esto puede apreciarse en la curva de la figura 2.5, que muestra un área de .025 a cada extremo de la distribución de muestreo.

Para completar el ejemplo 2.8, si resulta que la media muestral es 50.4, no rechazaríamos la hipótesis nula y se concluiría que la población aún conserva una media en el peso de 50 libras. Por otra parte, si la media muestral resultara de 48.1 libras, se concluiría que la media de la población ha caído por debajo de las 50 libras.

La tabla 2.6 resume las cuatro pruebas de hipótesis más utilizadas. En cada una de estas pruebas se calcula el estadístico de la muestra que se indica, se consulta la distribución muestral apropiada después de suponer que la hipótesis nula es correcta, se forma la regla de decisión que especifica el límite donde es más probable que se ubique el estadístico y se llega a una decisión de rechazar o no la hipótesis nula.

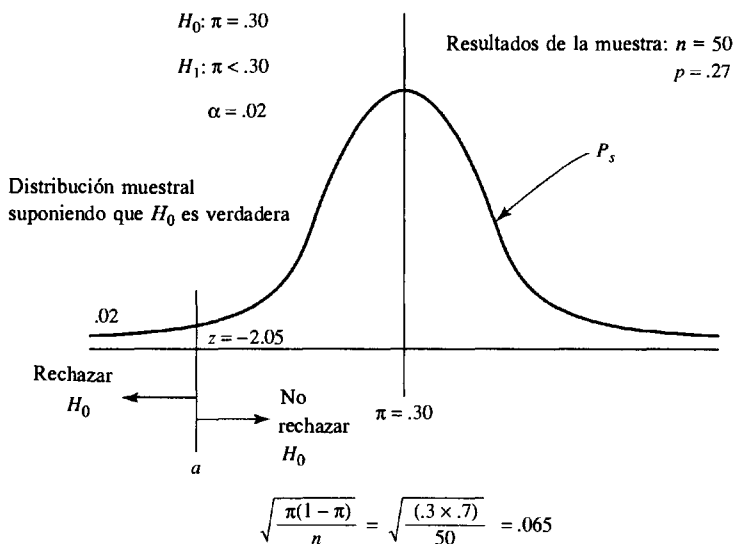
TABLA 2.6 RESUMEN DE PRUEBAS DE HIPÓTESIS

H_0	ESTADÍSTICO DE LA MUESTRA	DISTRIBUCIÓN MUESTRAL	PARÁMETROS DE LA DISTRIBUCIÓN MUESTRAL
$\mu = 50^*$	$\bar{X}$	Normal o t	μ and $\frac{\sigma}{\sqrt{n}}$
$\pi = .30^*$	p	Normal	π and $\sqrt{\frac{\pi(1-\pi)}{n}}$
$\mu_1 - \mu_2 = 0$	$\bar{X}_1 - \bar{X}_2$	Normal o t	$\mu_1 - \mu_2$ and $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\pi_1 - \pi_2 = 0$	$p_1 - p_2$	Normal	$\pi_1 - \pi_2$ and $\sqrt{\pi(1-\pi)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

*Valores de ejemplo.

Ejemplo 2.9

Por varios años, el 30% de la población ha tenido una impresión favorable de nuestro producto. Eventos recientes nos conducen a sospechar que este porcentaje pudiera haber disminuido. Para probar la hipótesis de que el porcentaje aún es del 30%, se toma una muestra de 50



$$a = .30 - 2.05(.065) = .30 - .13 = .17$$

Figura 2.6 Prueba de hipótesis - ejemplo 2.9

personas y se encuentra que el 27% de ellas tiene una buena impresión de nuestro producto. Con base en la evidencia que se muestra en la figura 2.6, ¿cuál es la conclusión?

Regla de decisión. Si $p < .17$, rechazar H_0

Conclusión. Ya que $p = .27$, no se rechaza H_0 . Con base en la evidencia de esta muestra, la hipótesis de que la proporción de la población es .30 no puede rechazarse a un nivel de significancia del 2%. No existe suficiente evidencia en la muestra para decir que ha disminuido la proporción de personas que favorecen nuestro producto.

Ejemplo 2.10

Considérese una prueba de hipótesis en donde el tamaño de la muestra es pequeño y cuya distribución muestral apropiada es la distribución t . Suponga que se desea probar la hipótesis de que la media de la calificación obtenida por estudiantes en un examen a nivel nacional es 500, contra la hipótesis alterna de que es menor de 500. Se toma de la población una muestra aleatoria de 15 estudiantes, que arroja una media de 475. La desviación estándar poblacional se estima mediante la desviación estándar de los 15 elementos de la muestra, que es 35. Se supone que la población de calificaciones de exámenes sigue una distribución normal.

La distribución t es la distribución muestral apropiada cuando n es pequeña (como regla de oro, menor de 30), se desconoce σ (y se estima mediante s), y se puede suponer que la población tiene aproximadamente una distribución normal. El número correcto de grados de libertad (df) para esta prueba es $n - 1$. La estadística de prueba es $(\bar{X} - \mu)/(s/\sqrt{n})$. La prueba se muestra en la figura 2.7, suponiendo un nivel de significancia de .05.

Regla de decisión. Si $t < -1.761$, rechazar H_0

Conclusión. Ya que $t = -2.77$, se rechaza H_0 . La evidencia de la muestra apoya el rechazo de la hipótesis de que la media de la población es 500 a un nivel de significancia del 5%. Se concluye que la media de la calificación obtenida por estudiantes en el examen nacional ha disminuido.

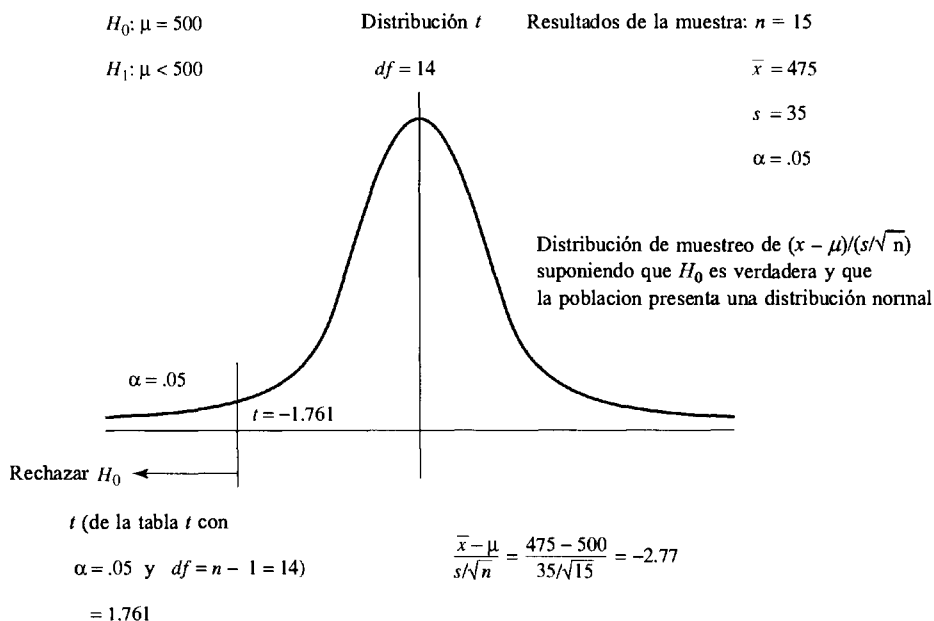


Figura 2.7 Prueba de hipótesis - ejemplo 2.10

PRUEBA DE LA BONDAD DE AJUSTE

A menudo resulta de interés determinar si los datos de la muestra se ajustan a una distribución conocida, sospechada o teórica. Se extrae una muestra de la población de interés y se observan las categorías en las que se ubican estos elementos. Si las frecuencias observadas se encuentran muy cerca de lo que teóricamente se vería bajo la distribución hipotética de la población, se acepta la formulación hipotética de la población; si se observan grandes diferencias, esta formulación se rechaza.

Ejemplo 2.11

que la gente no tiene preferencia con respecto a las cuatro marcas de leche que puede comprar en el supermercado. Si esta hipótesis es verdadera, entonces en una muestra de compradores podría esperarse que exactamente la cuarta parte de ellos, en promedio, elegiría cada una de las cuatro marcas. Si se elige al azar una muestra de 100 compradores, se esperaría que 25 de ellos comprara cada una de las cuatro marcas. Suponga que ya se tomó la muestra y que ésta revela la información de preferencia que se presenta en la tabla 2.7.

Resulta obvio que una cuarta parte de la población no eligió cada marca, sin embargo, sólo se examinó una muestra pequeña de compradores. Surge la pregunta ¿hay suficiente evidencia en la muestra para rechazar la hipótesis nula? La hipótesis bajo prueba es

H_0 : La población de la que se extrajo la muestra está distribuida de manera uniforme.

Bajo esta hipótesis, un número igual de personas de la población prefiere cada marca. La evidencia de la muestra puede o no ser lo suficientemente fuerte para rechazar este concepto. A partir de la evidencia, no resulta aparente si existe preferencia por una u otra marca en la población, o si las diferencias que se observan en la muestra se deben a un error en el muestreo aleatorio.

TABLA 2.7 PREFERENCIAS DE MARCAS – EJEMPLO 2.11

	A	B	C	D	TOTAL
Número de compradores	20	35	18	27	100

Suponga la
hipótesis de

Suponga que las observaciones reales están dispuestas de acuerdo con las frecuencias esperadas si la hipótesis nula fuera verdadera, como se muestra en la figura 2.8. Ahora se puede apreciar la bondad de ajuste entre las frecuencias observadas y esperadas. Se puede calcular un estadístico a partir de las comparaciones de la tabla 2.8; se le conoce como la distribución ji-cuadrada. La ecuación 2.12 muestra el cálculo:

$$\chi^2 = \sum_{\text{todas las categorías}} \frac{(f_o - f_e)^2}{f_e} \quad (2.12)$$

en donde

f_o = frecuencia observada

f_e = frecuencia esperada

TABLA 2.8 PREFERENCIAS DE MARCAS —HIPÓTESIS NULA CORRECTA

	A	B	C	D	TOTAL
Real (observado)	20	35	18	27	100
Teórico (esperado)	25	25	25	25	100

Se puede emplear la ecuación 2.12 para calcular el estadístico de ji-cuadrada para el ejemplo 2.11. En la tabla 2.9 se muestran los cálculos que dan por resultado una estadística de ji-cuadrada de 7.12.

Ahora, el valor de ji-cuadrada, calculado con la ecuación 2.12, se compara con el valor de la tabla de ji-cuadrada correspondiente. Si el valor calculado es menor que el de la tabla, indica un buen ajuste y no se rechaza la hipótesis nula sobre la población en estudio. Si el valor calculado es mayor que el de la tabla, indica un mal ajuste y la hipótesis nula se rechaza.

Para obtener un valor tabulado de la tabla de ji-cuadrada, es necesario conocer el número de grados de libertad (df). Para el ejemplo 2.11, se pierde un grado de libertad, ya que las frecuencias esperadas deben sumar 100, que es el número original de datos. De este modo, $df = k - 1 = 4 - 1 = 3$ (k es el número de categorías). De la tabla de ji-cuadrada,

$$\chi^2_{.05} = 7.81473$$

$$\chi^2_{.025} = 9.34840 \quad \text{para } df = 3$$

$$\chi^2_{.01} = 11.3449$$

Debido a que el valor calculado (7.12) de ji-cuadrada es menor que cualquiera de los valores tabulados, incluso al nivel de significancia de .05, no es posible rechazar la hipótesis nula a estos niveles. Como es común en las pruebas estadísticas, este resultado no prueba que la hipótesis nula sea correcta; simplemente establece que no hay evidencia suficiente para rechazarla. Se concluye que la muestra indica que no hay marca de preferencia en la población, utilizando cualquiera de los niveles de significancia indicados.

En resumen, se pueden comparar las distribuciones observadas con las distribuciones teóricas, mediante la prueba de la bondad de ajuste. Se debe calcular y comparar con los mismos valores la distribución de frecuencias esperadas de la población teórica, si existen grandes diferencias entre categorías, resultaría un valor grande de ji-cuadrada y se

TABLA 2.9 CÁLCULOS DE χ^2

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
20	25	-5	25	1.00
35	25	10	100	4.00
18	25	-7	49	1.96
27	25	2	4	.16
				$\chi^2 = 7.12$

rechazaría la hipótesis con respecto a la población. Si las frecuencias reales están muy cerca de los valores esperados entre las categorías, resultará un valor pequeño de ji-cuadrada y no podrá rechazarse la hipótesis relativa a la población.

La prueba de la bondad de ajuste determina si se puede suponer que los elementos de una muestra hayan sido extraídos de una población que sigue una distribución específica.

En una prueba de bondad de ajuste se pueden formular muy diversas hipótesis de distribución de la población. A menudo la prueba se realiza para una distribución normal de la población, ya que muchas pruebas estadísticas suponen tal distribución de la población. Las pruebas de bondad de ajuste se utilizan también para determinar si una variable es de Poisson o está distribuida en forma exponencial. En el capítulo 10 se usará una variante de esta prueba para determinar si son adecuados los modelos de Box-Jenkins. La bondad de ajuste se usa además para probar datos actuales contra la distribución registrada en periodos anteriores, contra una norma de la industria, o contra valores objetivos deseados. Por último, se utiliza la bondad de ajuste para probar la precisión de los valores de pronóstico con los valores reales (como el importe de ventas mensuales).

ANÁLISIS DE CORRELACIÓN

Un objetivo común en las aplicaciones de pronóstico consiste en examinar la relación entre dos variables. Este tema se aborda aquí, mientras que su extensión, el análisis de regresión se expone en los capítulos 6 y 7. Además, Los capítulos 9 y 10 consideran casos especiales de correlación y regresión.

El énfasis en la correlación y la regresión se justifica en virtud del amplio uso de estas técnicas en todo tipo de aplicaciones estadísticas, incluyendo los pronósticos. Además, la ampliación de los conceptos básicos de correlación y regresión en áreas más complejas se consideran en muchos otros procedimientos de pronóstico.

Se supone que las variables bajo investigación son mediciones numéricas, es decir, que se miden mediante un dispositivo o procedimiento que genera números reales en vez de categorías (se utilizan datos de intervalos o escala proporcional).

Diagramas de dispersión

Un estudio de la relación de variables comienza con el caso más sencillo, el de la relación existente entre dos variables. Suponga que se toman dos mediciones a cada uno de varios objetos. Un analista desea determinar cuál de estas variables medibles denominada Y , tiende a aumentar o disminuir mientras que la otra variable, llamada X , varía. Por ejemplo, suponga que se miden la edad y el ingreso de varios individuos, como se muestra en la figura 2.10. ¿Qué se puede decir acerca de la relación entre X y Y ?

TABLA 2.10 MEDICIONES DE INGRESO Y EDAD

PERSONA NÚMERO	Y INGRESO (\$)	X EDAD (AÑOS)
1	7,800	22
2	8,500	23
3	10,000	26
4	15,000	27
5	16,400	35

A partir de la tabla 2.10 resulta aparente que Y y X tienen una relación definida. Al aumentar X , Y tiende también a aumentar. Al observar esta muestra de cinco personas pudiera ser tentador el concluir que entre mayor sea una persona más ingresos tendrá. Desde luego, es peligroso llegar a conclusiones con base en un tamaño de muestra inadecuado, un tema a tratar más adelante. Aun así, con base en las observaciones ($n = 5$), aparentemente existe una relación definida entre X y Y .

Estos cinco puntos se pueden graficar en una escala de dos dimensiones, con los valores de X en el eje horizontal y los valores de Y en el eje vertical. A esta gráfica se le denomina *diagrama de dispersión* y aparece en la figura 2.8.

Un *diagrama de dispersión* grafica los puntos de datos X - Y en una gráfica bidimensional.

Un diagrama de dispersión ayuda a ilustrar lo que sugiere la intuición, al observar por primera vez los datos sin procesar, digamos, la apariencia de una relación lineal entre X y Y . A esta relación se le llama *positiva* ya que al incrementarse X también se incrementa Y .

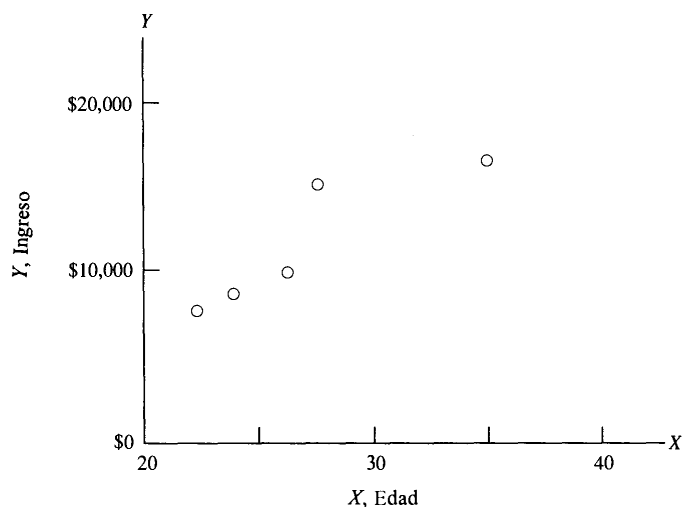


Figura 2.8 Diagrama de dispersión de edad e ingreso

En otras situaciones que comprenden a dos variables, pueden surgir diferentes patrones de diagramas de dispersión. Considérense las gráficas de la figura 2.9.

El diagrama (a) de la figura 2.9 sugiere lo que se llama una *relación lineal positiva perfecta*. Al aumentar X , aumenta también Y y en una forma perfectamente predecible. Esto es, que los puntos X y Y aparentemente se ubican sobre una línea recta. El diagrama (b) sugiere una *relación lineal negativa perfecta*. Al aumentar X , Y disminuye en una forma predecible.

Las figuras 2.9(c) y (d) ilustran *relaciones lineales imperfectas positiva y negativa*. Al aumentar X en estos diagramas de dispersión, Y aumenta (c) o disminuye (d), pero no de manera perfectamente predecible. De ahí que Y pudiera ser ligeramente mayor o menor que “lo esperado”. Esto es, los puntos X - Y no se ubican sobre una línea recta.

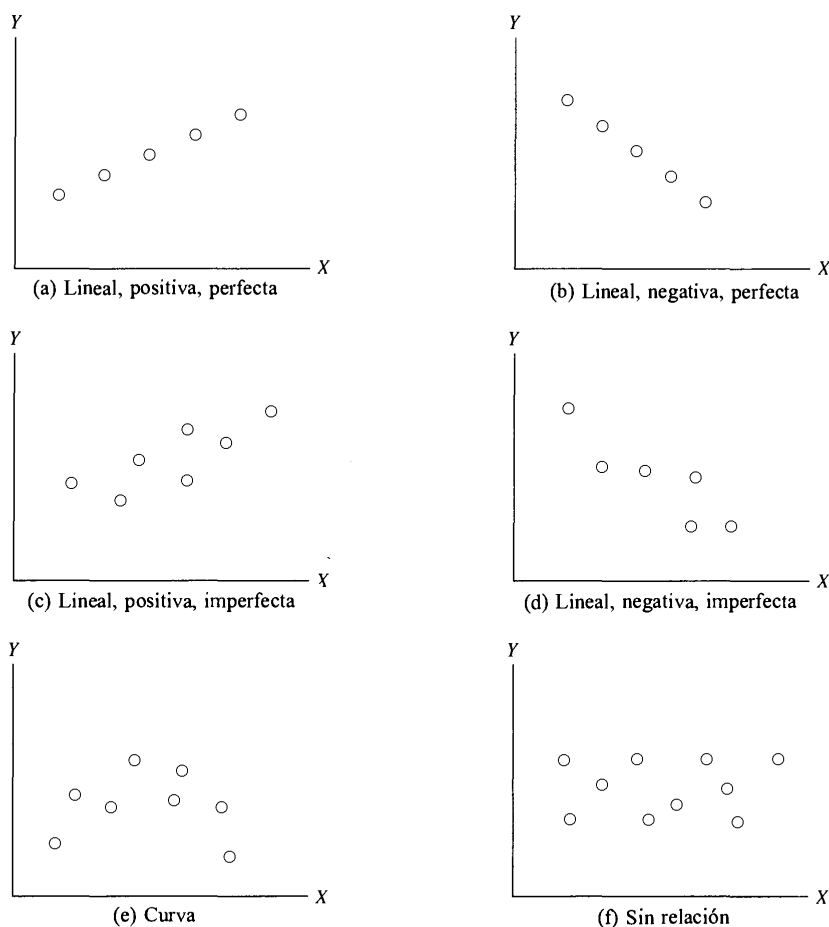


Figura 2.9 Patrones para las gráficas de datos X - Y .

Los diagramas de dispersión en las figuras 2.9(a) a la (d) ilustran lo que se conoce como *relación lineal*. La relación X - Y , sea perfecta o imperfecta, se puede resumir mediante una línea recta. En comparación, en el diagrama (e) aparece una relación *curva*.

Por último, el diagrama (f) de la figura 2.9 sugiere que no existe relación de ninguna clase entre las variables X y Y . Al aumentar X , Y no parece aumentar o disminuir en forma predecible alguna. Con base en la evidencia de la muestra que aparece en el diagrama (f), se podría concluir que en el universo que contiene *todos* los puntos X y Y , no existe relación lineal u otra, entre las variables X y Y .

Considérense ahora los dos diagramas de dispersión de la figura 2.10. Ambos diagramas de dispersión sugieren relaciones lineales positivas imperfectas entre Y y X . La diferencia consiste en que esta relación parece ser más fuerte en la figura 2.10(a) ya que todos sus puntos de datos están muy cerca de la línea recta que pasa a través de ellos. En la figura 2.10(b) se sugiere una relación más débil. Los puntos de datos están más alejados de la línea recta que los atraviesa, sugiriendo una relación lineal más débil. Más adelante en este capítulo se mostrará cómo medir la fuerza de la relación que existe entre dos variables.

Como sugieren los dos diagramas de dispersión de la figura 2.10, con frecuencia es deseable resumir la relación entre dos variables ajustando una línea recta a través de los puntos de datos. Usted aprenderá a hacerlo más adelante, pero por el momento se puede decir que se puede ajustar una línea recta a los puntos del diagrama de dispersión, de modo que resulte un buen ajuste. La cuestión que se sugiere ahora es ¿qué tan rápido aumenta o disminuye la pendiente de esta línea recta?

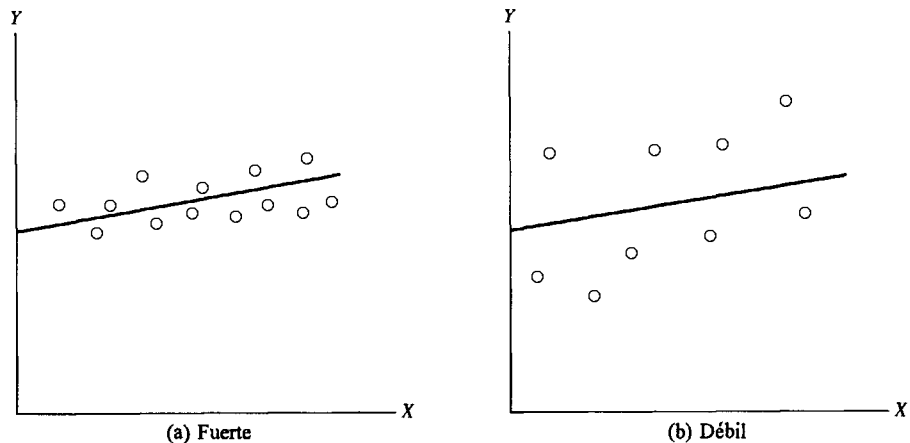


Figura 2.10 Graficación de relaciones fuerte y débil de datos X - Y .

Responder esta pregunta requiere de calcular la pendiente de la línea. La *pendiente* de cualquier línea recta se define como el cambio en Y asociado con el cambio en X . La pendiente de una línea se muestra en la figura 2.11.

En la figura 2.11(a), el cambio en Y entre los dos puntos de datos es 1, mientras que el cambio en X es 2; por lo que la pendiente de la línea es $1/2$. En la figura 2.11(b) la línea

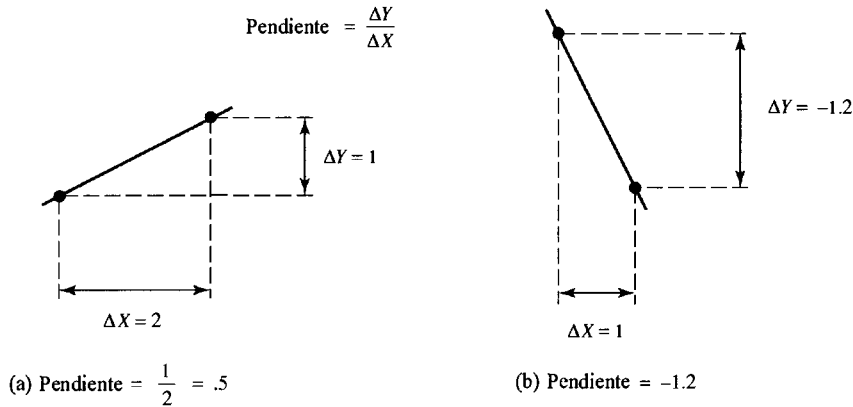


Figura 2.11 Pendiente de la recta.

tiene una pendiente negativa. Al incrementarse X en 1, Y disminuye en 1.2, por lo que la pendiente es de -1.2.

Para resumir, en la investigación de una relación entre dos variables, primero se debe saber si la relación es lineal (una línea recta) o curva. Si es lineal, se deseará saber si la relación es positiva o negativa y cuál es la inclinación de la línea que se ajusta a los puntos de datos. Por último, se necesita el *grado* de la relación, esto es, qué tan cerca están los puntos de la línea que mejor los ajusta.

Coefficiente de correlación

Se necesita una forma de medir la cantidad de relación lineal que existe entre dos variables de interés. Para usar la terminología correcta, se desea una medición de la *correlación* que existe entre dos variables. La medición que se utiliza comúnmente para esta relación es el *coeficiente de correlación*. 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 $+1$. Por lo que el coeficiente de correlación varíe entre -1 y $+1$ inclusive, dependiendo de la cantidad de correlación entre las dos variables que se midan.

El *coeficiente de correlación* mide el grado al cual se relacionan en forma lineal dos variables entre sí.

El diagrama de dispersión (a) de la figura 2.9 ilustra una situación que produciría un coeficiente de correlación de $+1$, mientras que el diagrama de dispersión (b) tiene un coeficiente de correlación de -1 . Los diagramas (e) y (f) grafican dos variables que no presentan relación lineal. Los coeficientes de correlación de estas relaciones son igual a 0, es decir, no existe relación lineal.

También es importante distinguir entre los dos grupos de puntos de datos con los que está interesado quien pronostica. En la *población* que contiene todos los puntos X - Y de interés, existe un coeficiente de correlación cuyo símbolo es ρ (la letra griega rho). Si se extrae una *muestra* de estos puntos de datos X - Y , al coeficiente de correlación de esta muestra de datos se le denomina r . En la tabla 2.11 se presenta un resumen de las características importantes de ρ y r .

TABLA 2.11 CORRELACIÓN DE LA POBLACION Y DE LA MUESTRA

POBLACIÓN	MUESTRA
Coeficiente de correlación: ρ (-1 " r " 1)	Coeficiente de correlación: r (-1 " r " 1)
$\rho = -1$: correlación negativa perfecta	$r = -1$: Correlación negativa perfecta
$\rho = 0$: sin correlación	$r = 0$: sin correlación
$\rho = 1$: correlación positiva perfecta	$r = 1$: correlación positiva perfecta

Ejemplo 2.12

Considere un conjunto específico de valores de X - Y y suponga que estos puntos constituyen la población completa de dichos puntos de datos. En la composición de los datos de la tabla 2.12, X representa el número de cursos de capacitación en ventas tomados y Y representa el número de días sin venta por mes, de los cinco vendedores del personal de la compañía.

Aparentemente existe una relación negativa en los datos ya que, mientras X se mueve en una dirección (hacia abajo), Y se mueve precisamente en otra (hacia arriba). De ahí que para esta población de puntos de datos se espera que sea aproximadamente -1 .

Con frecuencia, X y Y se miden en unidades diferentes, tales como libros y dólares, unidades vendidas y ventas en dólares, tasa de desempleo y Producto Nacional Bruto en dólares. Haciendo a un lado estos diferentes métodos para medir X y Y , sigue siendo importante el alcance de su relación entre sí. Esta medición se hace convirtiendo primero las mediciones de los datos de la muestra a unidades estándar.

Las mediciones de X - Y se convierten a valores estandarizados, los cuales se multiplican para cada medición de X - Y , obteniendo productos cruzados para cada elemento o persona.

TABLA 2.12 DATOS PARA EL EJEMPLO 2.12:
CURSOS Y DÍAS SIN VENTAS

VENDEDOR	DÍAS SIN VENTAS (Y)	CURSOS (X)
A	3	7
B	4	6
C	5	5
D	6	4
E	7	3

Estos productos cruzados son de interés ya que la media de estos valores corresponde al coeficiente de correlación. La ecuación 2.13 muestra que la media de los productos cruzados de los valores estandarizados para dos variables es el coeficiente de correlación.

$$\rho = \frac{\sum Z_y Z_x}{N} \quad (2.13)$$

El cálculo del coeficiente de correlación ρ mediante los productos cruzados de los valores estandarizados siempre producirá el valor correcto, aunque en la mayoría de los casos existe una forma más fácil de realizar los cálculos. La ecuación 2.14 es equivalente a encontrar ρ mediante el cálculo de la media de los productos cruzados de los valores estandarizados.

$$\rho = \frac{\sum Z_y Z_x}{N} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \quad (2.14)$$

(Véase el apéndice A para una prueba de esta relación.)

Para calcular ρ mediante esta ecuación, se requiere una tabla de valores. Para el ejemplo 2.12, los cálculos se ilustran en la tabla 2.13.

Ahora, tenemos todos los elementos necesarios para resolver la ecuación 2.14; la sustitución y los cálculos son

$$\begin{aligned} \rho &= \frac{(5)(115) - (25)(25)}{\sqrt{(5)(135) - 25^2} \sqrt{(5)(135) - 25^2}} = \frac{575 - 625}{\sqrt{50} \sqrt{50}} \\ &= \frac{-50}{\sqrt{50^2}} = \frac{-50}{50} = -1 \end{aligned}$$

Al examinar un conjunto de puntos de datos, se debe emplear un juicio para decidir cuándo el valor de ρ es lo suficientemente grande para considerar que las dos variables tienen una alta correlación lineal. Aunque en la mayoría de los casos comprende un conjunto de puntos X - Y de una población, se puede usar la ecuación 2.15 para encontrar el coeficiente de correlación que mida hasta qué grado las dos variables se mueven juntas en forma lineal.

Es común el caso que se haya extraído una muestra de puntos de datos de manera aleatoria, de una población bajo estudio. En este caso, se utiliza la ecuación 2.14 después de remplazar N , el tamaño de la población, por n , el tamaño de la muestra. Así, la fórmula para obtener el coeficiente de correlación de la muestra se convierte en

TABLA 2.13 CÁLCULOS NECESARIOS PARA UTILIZAR LA ECUACIÓN 2.14 PARA LOS DATOS DEL EJEMPLO 2.12

VENDEDOR	Y	X	Y^2	X^2	XY
A	3	7	9	49	21
B	4	6	16	36	24
C	5	5	25	25	25
D	6	4	36	16	24
E	7	3	49	9	21
Sumas	25	25	135	135	115

$$r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2} \sqrt{n\sum Y^2 - (\sum Y)^2}} \quad (2.15)$$

**Ejemplo
2.13**

Si se estudia la relación entre edad e ingreso (tabla 2.10), pudiera ser de interés el conocer el valor de r para estos datos. Los cálculos requeridos aparecen en la tabla 2.14.

En la tabla se puede apreciar que el coeficiente de correlación confirma lo que se observó en la tabla 2.10 y en la figura 2.8. El valor de r es positivo, sugiriendo una relación lineal positiva entre edad e ingreso. Además, en una escala del 0 al 1, el valor de r es aceptablemente alto (.89). Este resultado sugiere una fuerte relación lineal y no una débil. La pregunta que resta es saber si la combinación del tamaño de la muestra y el coeficiente de correlación es lo suficientemente fuerte para hacer declaraciones significativas respecto a la población de la que se extrajo la muestra.

Ahora, al abordar la correlación, se deben señalar dos puntos importantes.

Primero, siempre debe tenerse presente que se está midiendo la *correlación* y no la *causalidad*. Podría ser perfectamente válido que dos variables estuvieran correlacionadas con base en un coeficiente de correlación alto. Pero podría o no ser válido decir que una variable *causa* el movimiento de la otra; ésta es una cuestión para el juicio del analista. Por ejemplo, podría ser cierto que el volumen de ventas de una tienda rural en un área poco poblada estuviera altamente relacionada con el precio promedio del mercado accionario de la ciudad de Nueva York. Después de examinar una muestra grande de estas dos variables, se podría concluir que tal correlación alta existe. Probablemente sea falso que una de estas dos variables esté causando el movimiento de la otra. De hecho, es probable que el movimiento de estas dos variables esté producido por una tercera, el estado general de la economía. Es frecuente que políticos, publicistas y otras personas cometan el error de suponer causalidad con base en la correlación.

**TABLA 2.14 CÁLCULOS PARA LA CORRELACIÓN ENTRE EDAD
E INGRESO DEL EJEMPLO 2.14**

PERSONA NÚM.	Y	X	Y ²	X ²	XY
1	7,800	22	60,840,000	484	171,600
2	8,500	23	72,250,000	529	195,500
3	10,000	26	100,000,000	676	260,000
4	15,000	27	225,000,000	729	405,000
5	16,400	35	268,960,000	1,225	574,000
Sumas	57,700	133	727,050,000	3,643	1,606,100

$$\begin{aligned}
 r &= \frac{(5)(1,606,100) - (133)(57,700)}{\sqrt{(5)(3,643) - (133)^2} \sqrt{(5)(727,050,000) - (57,700)^2}} \\
 &= \frac{8,030,500 - 7,674,100}{(22.93)(17,491.71)} = \frac{356,400}{401,084.90} \\
 &= .89
 \end{aligned}$$

Segundo, nótese que la correlación mide una relación *lineal* entre dos variables. Si fuera el caso de un coeficiente de correlación bajo, se podría concluir que las dos variables no guardan una estrecha relación en forma lineal. Podría ser que si tuvieran una estrecha relación de manera curva o no lineal. De ahí que un coeficiente de correlación bajo no signifique que no hay correlación entre dos variables, solamente que no es aparente que exista una relación lineal o recta.

APLICACIÓN EN LA ADMINISTRACIÓN

Muchos de los conceptos de este capítulo de repaso se podrían considerar como material necesario para comprender las técnicas más avanzadas de pronóstico que se encuentran a lo largo del resto del texto. Sin embargo, los conceptos de este capítulo tienen también, por sí mismos, valor en muchas aplicaciones estadísticas. Mientras que algunas de estas aplicaciones pudieran no quedar bajo el título de “pronósticos”, abarcan sin embargo el uso de datos reunidos para responder preguntas sobre la incertidumbre de la operación empresarial, en especial a la incertidumbre de los resultados futuros.

Los procedimientos de estadística descriptiva mencionados con anterioridad se utilizan ampliamente dondequiera que se deben describir grandes cantidades de datos, de manera que se puedan asimilar y utilizar en el proceso de toma de decisiones. Sería casi imposible pensar en una sola área que comprenda mediciones numéricas, en la que no se resuman rutinariamente los conjuntos de datos mediante estadística descriptiva. Este hecho se aplica en particular a la media, a la que comúnmente se llama “promedio” y, en menor grado, a la desviación estándar. Cualquiera entiende el promedio de conjuntos de datos y se ha hecho durante muchos años para proporcionar una medición central de arreglos de datos. Las mediciones de dispersión como la desviación estándar están ganando terreno al hacerse más comprensibles tales mediciones.

Las distribuciones binomial y normal son buenos ejemplos de distribuciones teóricas que representan buenos modelos para muchas situaciones de la vida real. Como tales, su uso es amplio en muchas aplicaciones, incluyendo el pronóstico. Por ejemplo, el pronóstico del porcentaje de partes defectuosas en un lote, se podría encontrar mediante la utilización de una distribución binomial.

La estimación y las pruebas de hipótesis son los dos pilares de las aplicaciones estadísticas básicas. El pronóstico o estimación de los valores de una población de interés a través de la medición de una muestra aleatoria, está ampliamente difundido en donde las restricciones de tiempo y dinero fuerzan la dependencia en resultados de muestras para la información para la toma de decisiones. El muestreo se ha difundido en particular en el proceso de la auditoría contable. La prueba de hipótesis se usa ampliamente para comparar valores de la población en periodos anteriores con valores actuales; para comparar los valores del parámetro de dos distintas ramas o ubicaciones de un negocio y para detectar los cambios en mediciones clave de procesos productivos. Esta última aplicación, conocida como *control de producción* es, en muchos casos, el componente clave de los esfuerzos de una empresa para mejorar su calidad.

La prueba de la bondad de ajuste se usa con frecuencia para verificar la validez de suposiciones con respecto a distribuciones de población. Este proceso es especialmente

importante al utilizar ciertos modelos de simulación que requieren, por ejemplo, la suposición de procesos de Poisson, para que los resultados del modelo sean válidos. Dichas simulaciones se emplean para modelar situaciones reales y, después de validarlas, para pronosticar los resultados de cambios en el sistema sobre resultados de salida.

La correlación tiene un amplio uso en el examen de las relaciones de pares de variables numéricas. Como se verá en capítulos posteriores, estas relaciones son de gran importancia para los pronósticos, ya que el pronóstico de una variable a menudo comprende el intento de encontrar variables relacionadas. Tanto el análisis de regresión como el de regresión múltiple se basan en la correlación en este proceso.

GLOSARIO

Coefficiente de correlación Medición que indica hasta qué grado se relacionan linealmente dos variables numéricas entre sí.

Diagrama de dispersión Una gráfica de puntos X-Y en un espacio bidimensional.

Distribución muestral Una distribución muestral es la disposición de todas las estadísticas posibles de la muestra, que se pueden extraer de una población, para un tamaño de muestra dado.

Estadística descriptiva Métodos que se usan para describir conjuntos de datos con unos cuantos valores de resumen.

Grados de libertad Los grados de libertad en un conjunto de datos indican el número de elementos de datos que son independientes entre sí y que representan piezas únicas de información.

Intervalo de estimación Un intervalo de estimación es un intervalo numérico, dentro del cual es probable que se ubique un parámetro de la población.

Prueba de la bondad de ajuste La prueba de la bondad de ajuste determina si se puede suponer que una muestra de elementos hayan sido extraídos de una población que sigue una distribución específica.

Punto de estimación Un punto de estimación es una estimación con un solo valor, de un parámetro de la población.

Valor esperado El valor esperado de una variable aleatoria corresponde al valor promedio de la variable en muchos intentos u observaciones.

Variable aleatoria continua Una variable aleatoria continua puede tomar cualquier valor numérico dentro de cierto intervalo.

Variable aleatoria discreta Una variable aleatoria discreta sólo puede tomar ciertos valores específicos, por lo regular enteros.

FÓRMULAS CLAVE

$$\text{Media poblacional } \mu = \frac{\Sigma X}{N} \quad (2.1)$$

$$\text{Media muestral } \bar{X} = \frac{\Sigma X}{s/n} \quad (2.2)$$

Desviación estándar poblacional

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}} \quad (2.3)$$

$$\text{Desviación estándar muestral } s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1}} \quad (2.4)$$

$$\text{Varianza poblacional } \sigma^2 = \frac{\Sigma(X - \mu)^2}{N} = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N} \quad (2.5)$$

$$\text{Varianza muestral } s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n - 1} \quad (2.6)$$

$$\text{Valor esperado } E(X) = \Sigma[X \cdot P(X)] \quad (2.7)$$

$$\text{Fórmula binomial } P(x) = C_x^n \pi^x (1 - \pi)^{n-x} \quad (2.8)$$

$$\text{Fórmula para obtener valores estandarizados } Z = \frac{X - \mu}{\sigma} \quad (2.9)$$

$$\text{Intervalo de confianza para una media } \bar{X} \pm Z \frac{s}{\sqrt{n}} \quad (2.10)$$

$$\text{Intervalo de confianza para una proporción } p \pm Z \sqrt{\frac{\pi(1 - \pi)}{n}} \quad (2.11)$$

$$\text{Fórmula de ji-cuadrada } \chi^2 = \sum_{\text{todas las categorías}} \frac{(f_o - f_e)^2}{f_e} \quad (2.12)$$

Coefficiente de correlación para una población

$$\rho = \frac{\sum Z_y Z_x}{N} \quad (2.13)$$

$$\rho = \frac{\sum Z_y Z_x}{N} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \quad (2.14)$$

Coefficiente de correlación para una muestra

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{n \sum X^2 - (\sum X)^2} \sqrt{n \sum Y^2 - (\sum Y)^2}} \quad (2.15)$$

PROBLEMAS

1. Dick Hoover, propietario de Modern Office Equipment, está preocupado por los costos de transporte y de oficina en que incurren los pedidos pequeños. En un esfuerzo por reducir gastos en esta área, decide introducir una política de descuento, en beneficio de los pedidos superiores a \$40, esperando que esto motive a los clientes a consolidar varios pedidos pequeños en otros más grandes. Los datos siguientes muestran las cantidades por transacción de 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

- a. Calcule la media.
 - b. Calcule la desviación estándar.
 - c. Calcule la varianza.
 - d. Si la política tiene éxito, ¿la media de la distribución aumentará, disminuirá o permanecerá sin cambio?
 - e. Si la política tiene éxito, ¿la desviación estándar de la distribución aumentará, disminuirá o permanecerá sin cambio?
 - f. Pronostique la cantidad de la transacción del próximo cliente.
2. De acuerdo con un manual no clasificado de una oficina de impuestos, no son sólo los elementos que aparecen en su declaración los que pueden producir una auditoría. La ausencia de ciertos elementos puede elevar el número DIF (función de discriminación) y hacer que su declaración sea señalada para revisión. El número DIF indica la probabilidad de que una auditoría dé por resultado un impuesto adicional. También, el programa DIF es muy efectivo para detectar declaraciones mal llenadas. Sólo alrededor del 15% de las auditorías a contribuyentes en la categoría de ingreso de \$25 000 a \$50 000 resultan sin cambio en la cantidad de impuesto a pagar (*Tax Guide for College Teachers, 1992*).
 - a. Pronostique la probabilidad de que si la oficina de impuestos audita a un contribuyente con ingresos de \$45 000, no se produzca un cambio en la cantidad de impuesto a pagar.
 - b. Pronostique la probabilidad de que si la oficina de impuestos audita a un contribuyente con

- ingresos de \$30 000, se produzca un cambio en la cantidad de impuesto a pagar.
- c. La oficina de impuestos realiza al azar auditorías a cuatro contribuyentes con ingresos entre \$25 000 y \$50 000, pronostique la probabilidad de que dos o más auditorías den por resultado algún cambio en la cantidad de impuesto a pagar.
 3. El hotel Barcellos' Tumn Tum Resort lleva un registro del número de días que permanece cada huésped. Descubren que esta variable presenta una distribución normal con una media de siete días y una desviación estándar de dos. El pronóstico para el próximo mes indica que esperan 500 huéspedes.
 - a. Pronostique cuántos huéspedes permanecerán menos de 5 días.
 - b. Pronostique cuántos huéspedes permanecerán más de 15 días.
 - c. Pronostique cuántos huéspedes permanecerán entre 6 y 12 días.
 4. La línea aérea New Horizons Airlines desea pronosticar la media de asientos desocupados por vuelo a Alemania durante el próximo año. Para desarrollar este pronóstico, se seleccionaron en forma aleatoria de los archivos del año pasado, los registros de 49 vuelos y se tomó nota del número de asientos no ocupados en cada vuelo. La media muestral y la desviación estándar son 8.1 y 5.7, respectivamente. Desarrolle un punto e intervalo de estimación de 95% de la media de asientos no ocupados por vuelo durante el año pasado. Pronostique la media de asientos no ocupados por vuelo a Alemania en el próximo año. Explique la precisión de este pronóstico.
 5. Con base en experiencias anteriores, la California Power Company pronostica que en enero próximo, la media del uso de electricidad residencial por casa será de 700 kwh. En enero, la compañía selecciona una muestra aleatoria simple de 50 casas y calcula una media de 715 y una desviación estándar de 50. Pruebe a un nivel de significancia de .05 para determinar si el pronóstico es razonable.
 6. De acuerdo con una encuesta sobre planes familiares de viaje, realizada por el U.S. Travel Data Center para la revista *Better Homes and Gardens*, el 49% de las familias vacacionistas que pretenden viajar este año con niños, planean visitar un parque temático (*The Wall Street Journal*, abril 15, 1992). Brandon Kim, presidente de Wild Kingdom, está interesado en construir un parque temático en Madison, Wisconsin. Brandon piensa que tendrá éxito si puede pronosticar que por lo menos el 49% de las familias de los alrededores pretenden visitar un parque temático en sus vacaciones, por lo que realiza una encuesta a 700 familias y encuentra que 325 de ellas indican en sus respuestas que planean visitar un parque temático. ¿Cuál sería la conclusión de Brandon si prueba con un nivel de significancia de .01?
 7. Expertos en población indican que el tamaño de la familia se ha reducido en los últimos años. Hace diez años, el tamaño promedio de una familia era de 2.9. Considere una población con los tamaños de 200 familias que se presentan en la tabla P2.7. Seleccione de manera aleatoria una muestra de 30 elementos y pruebe la hipótesis de que el tamaño promedio de la familia no ha cambiado en los últimos diez años.
 8. La compañía Bif Peanut Butter predice que los consumidores prefieren Bif a otras marcas de mantequilla de cacahuete. La compañía contrata a una compañía de investigación independiente para efectuar una encuesta telefónica en un área de prueba de mercado. La compañía investigadora llama a 400 amas de casa y obtiene las siguientes preferencias de marca:
Redacte una breve nota para Bif, resumiendo los resultados de la investigación

MARCA A	BIF	MARCA C	MARCA D	MARCA E
75	90	82	65	88

TABLA P2.7

(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				

9. James Dobbins, supervisor de mantenimiento de Atlanta Transit Authority, quiere determinar si existe una relación positiva entre el costo anual de mantenimiento de un autobús y su edad. Si existe una relación, James piensa que puede hacer un mejor trabajo de predicción del presupuesto anual de mantenimiento de autobuses. Para ello, reúne los siguientes datos:

COSTO DE MANTENIMIENTO (\$)	EDAD (AÑOS)	
	AUTOBÚS	
	y	x
1	859	8
2	682	5
3	471	3
4	708	9
5	1094	11
6	224	2
7	320	1
8	651	8
9	1049	12

- a. Grafique un diagrama de dispersión.
 - b. ¿Qué tipo de relación existe entre estas dos variables?
 - c. Calcule el coeficiente de correlación.
10. Anna Sheehan es gerente de la cadena de supermercados Spendwise. A ella le gustaría poder predecir las ventas de los libros encuadernados en rústica (libros por semana) con base en la cantidad de espacio de repisas proporcionado (en pies). Anna reúne la información para una muestra de 11 semanas:

NÚMERO DE LIBROS VENDIDOS		ESPACIO EN REPISAS (EN PIES)
SEMANA	Y	X
1	275	6.8
2	142	3.3
3	168	4.1
4	197	4.2
5	215	4.8
6	188	3.9
7	241	4.9
8	295	7.7
9	125	3.1
10	266	5.9
11	200	5.0

- a. Grafique un diagrama de dispersión
 - b. ¿Qué tipo de relación existe entre estas dos variables?
 - c. Calcule el coeficiente de correlación.
 - d. ¿Debería Anna usar el análisis de regresión para predecir las ventas de libros en rústica?
11. Un vendedor de pasta de dientes en la parte occidental de Estados Unidos de América desea incrementar su participación en el mercado, que actualmente se ubica en el nivel del 23%. Se instituye una nueva campaña de publicidad y después de un mes de prueba con esta campaña, se hace una selección aleatoria de compradores para determinar su marca de preferencia. La agencia publicitaria señala con orgullo el hecho de que un 28% de una muestra de 1000 compradores usa el producto de la compañía. ¿Apoya la evidencia la posición de la agencia publicitaria de que su campaña incrementó la participación en el mercado?
12. Una campaña de control de calidad se ha enfocado en reducir el porcentaje de automóviles nuevos vendidos por un gran fabricante, que se regresan para reparaciones mayores. Antes de esta campaña el porcentaje era del 12%. Después de seis meses de la campaña de mejoramiento, la compañía hace una selección aleatoria entre varios de sus distribuidores y determina el número de automóviles vendidos y el número de los que regresaron para reparaciones mayores. Durante el periodo de la muestra, los distribuidores entregaron 1 542 automóviles y 123 de ellos regresaron. ¿Indica esta evidencia que la campaña de control de calidad tuvo éxito?
13. Considere una población con las 200 observaciones semanales que se presentan en la tabla P2.13. La variable independiente X es la temperatura semanal promedio de Spokane, Washington. La variable dependiente Y es el número de acciones de la empresa Sunshine Mining Stock negociadas en la Bolsa de Spokane en una semana. Seleccione aleatoriamente los datos de 16

TABLA P2.13

OBS.	Y	X	OBS.	Y	X	OBS.	Y	X	OBS.	Y	X
(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

semanas y calcule el coeficiente de correlación. (*Sugerencia:* Asegúrese que su muestra sea extraída de la población *en forma aleatoria*. Consulte a su instructor si no sabe cómo extraer una muestra aleatoria.)

CASO DE ESTUDIO 2.1

ALCAM ELECTRONICS

David Branch recibió recientemente su título en administración de empresas de una universidad pequeña e ingresó a trabajar en Alcam Electronics, un fabricante de diversos componentes electrónicos para la industria. Después de varias semanas de trabajo, la gerente y propietaria de Alcam, Alice Cameron, lo llamó a su oficina y le pidió que investigara dos aspectos con respecto a cierto transistor fabricado por Alcam, ya que una gran empresa de televisores estaba interesada en un pedido importante.

Primero, Alice quería pronosticar el porcentaje de transistores defectuosos que pudiera comprar la empresa de televisores. Debido a que la campaña de publicidad de la compañía enfatizaba la calidad del producto, naturalmente deseaba conocer la calidad de los componentes y quería documentación que apoyara las declaraciones de Alcam. Alice y David estuvieron de acuerdo en que se podría obtener una muestra de los transistores actualmente en inventario, para representar los que se enviarían en los meses siguientes a la compañía en prospecto, ya que el proceso de producción en el pasado reciente no se modificaría en el futuro inmediato.

Segundo, Alice desea saber el tiempo promedio de vida de los transistores, otro aspecto de gran importancia para la compañía de televisores. De nuevo, las unidades actualmente en existencia podrían representar a las que se producirían mientras duraba el nuevo contrato, si éste fuera aceptado.

David decidió tomar una muestra de los transistores en cuestión y formuló un plan para llevar a cabo su tarea. Numeró los compartimientos de almacenamiento que contienen los transistores, seleccionó números en forma aleatoria e hizo un muestreo de todos los transistores de los compartimientos elegidos para la muestra. Ya que cada compartimiento contenía alrededor de 20 transistores, seleccionó 10 números aleatorios, que al final le dieron un tamaño de muestra de 205 transistores. Como seleccionó 10 de los 55 compartimientos, supuso que tenía una buena muestra representativa y que podría usar los resultados de esta muestra para la totalidad de la población de transistores en inventario, así como para las unidades aún por fabricar mediante el mismo proceso.

A continuación, David sometió a pruebas de laboratorio a cada una de las 205 unidades, para determinar si funcionaban correctamente. Una vez concluidos los resultados del laboratorio, se encontró con que seis de las unidades no funcionaban de acuerdo con las especificaciones. David utilizó un nivel de confianza del 95% y realizó después los siguientes cálculos para llegar a un punto e intervalo de estimación del porcentaje defectuoso de todos los transistores fabricados en la planta:

$$p = \frac{6}{205} = .029$$

$$.029 \pm 1.96 \sqrt{\frac{(.029)(.971)}{205}}$$

$$.029 \pm .023$$

$$.006 \rightarrow .052$$

Con base en estos cálculos, David pensó en incluir la siguiente información en su reporte final para Alice:

1. Se estima que aproximadamente el 3% de los transistores fabricados mediante el proceso actual son defectuosos.
2. Existe un 95% de probabilidad de que la siguiente afirmación sea cierta: entre el .6% y el 5.2% de los transistores son defectuosos.

A continuación, David pasó a la cuestión del tiempo promedio de vida de las unidades. Debido a que las expectativas de vida se podrían extender a varios años, se dio cuenta de que ninguna de las unidades de la muestra se podría probar si se deseaba una respuesta en tiempo. De ahí que decidió contactar a diversos usuarios de este componente para determinar si había disponibles registros sobre su duración. Por fortuna, encontró tres compañías que utilizaron el transistor en el pasado y tenían registros limitados sobre la duración de componentes. En total, recibió datos sobre 38 transistores de los que se conocía su fecha de falla. Ya que dichos transistores se fabricaron con el mismo proceso que en la actualidad, razonó que los resultados de esta muestra se podrían inferir para las unidades en inventario y las aún no fabricadas.

Los resultados de los cálculos que realizó David sobre los datos de vida útil de su muestra fueron los siguientes:

$$n = 38$$

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

$$\text{Desviación estándar de vida útil} = s = 675 \text{ horas}$$

Después de encontrarse con que el tiempo promedio de vida útil de la muestra era sólo de 4 805 horas, David se preocupó pues sabía que el proveedor actual de componentes estaba garantizando un tiempo promedio de vida de 5 000 horas. Aunque el promedio de su muestra estaba sólo un poco por debajo de 5 000 horas, se dio cuenta de que el tamaño de la muestra era sólo de 38 y que esto no constituía una prueba positiva de que la calidad de Alcam fuera inferior a la del otro proveedor. Decidió entonces 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. Los siguientes son los cálculos que realizó:

$$H_0: \mu = 5000 \quad \alpha = .01$$

$$H_1: \mu < 5000 \quad s = 675$$

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

$$\text{Regla de decisión. Si } \bar{X} < 4\,744.9, \text{ rechazar } H_0$$

Considerando que la media muestral (4805) no está por debajo del punto de la regla de decisión (4744.9) para rechazar la hipótesis, David descartó el rechazar la hipótesis de que la media del tiempo

de vida de todos los componentes era igual a 5 000 horas. Pensó que estas serían buenas noticias para Alice CAMeron e incluyó un resumen de sus averiguaciones en su informe final. Pocos días después de que le presentó su informe escrito y verbal, Alice lo llamó a su oficina para felicitarle por el trabajo y compartir con él una preocupación respecto a su informe. Le dijo: “Me preocupa el bajo nivel de significancia de tu prueba de hipótesis. Tomaste sólo una probabilidad del 1% de rechazar la hipótesis nula en caso de ser correcta. Esto me parece muy conservador y me preocupa que hagamos un contrato y después descubramos que nuestro nivel de calidad no alcanza la especificación deseada de 5 000 horas”.

PREGUNTA

¿Cómo respondería al comentario de Alice Cameron?

CASO DE ESTUDIO 2.2

MR. TUX

John Mosby, propietario de varias tiendas de renta de ropa de etiqueta Mr. Tux, se interesa en pronosticar su volumen de ventas mensual (véase el caso de estudio Mr. Tux al final del capítulo 1). Como primer paso, John reúne los datos de ventas por mes para los años de 1985 a 1992.

	1985	1986	1987	1988	1989	1990	1991	1992
Ene.	6,028	16,850	15,395	27,773	31,416	51,604	58,843	71,043
Feb.	5,927	12,753	30,826	36,653	48,341	80,366	82,386	152,930
Mar.	10,515	26,901	25,589	51,157	85,651	208,938	224,803	250,559
Abr.	32,276	61,494	103,184	217,509	242,673	263,830	354,301	409,567
May.	51,920	147,862	197,608	206,229	289,554	252,216	328,263	394,747
Jun.	31,294	57,990	68,600	110,081	164,373	219,566	313,647	272,874
Jul.	23,573	51,318	39,909	102,893	160,608	149,082	214,561	230,303
Ago.	36,465	53,599	91,368	128,857	176,096	213,888	337,192	375,402
Sep.	18,959	23,038	58,781	104,776	142,363	178,947	183,482	195,409
Oct.	13,918	41,396	59,679	111,036	114,907	133,650	144,618	173,518
Nov.	17,987	19,330	33,443	63,701	113,552	116,946	139,750	181,702
Dic.	15,294	22,707	53,719	82,657	127,042	164,154	184,546	258,713

Después, John calcula el valor promedio de las ventas mensuales para cada año (por ejemplo, suma los doce valores de 1985 y divide entre doce). También calcula la desviación estándar para los doce valores mensuales de cada año. Los resultados son

AÑO	PROMEDIO (MEDIA)	DESVIACIÓN ESTÁNDAR
	\$	\$
1985	22,013	13,165
1986	44,603	35,290
1987	64,842	47,216
1988	103,610	57,197
1989	141,381	70,625
1990	169,432	63,376
1991	213,866	96,387
1992	247,231	99,153

Además, John decide construir un diagrama de dispersión como el que se muestra en la figura 2.12. Grafica los valores de la media de ventas mensuales en el eje y y los años en el eje x .

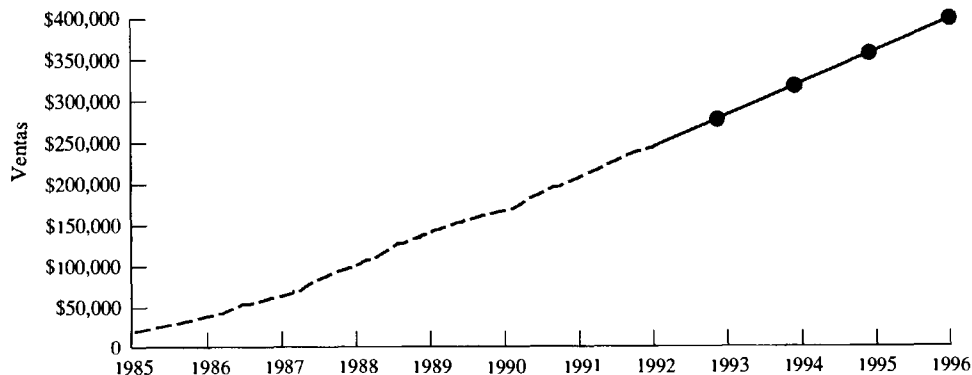


Figura 2.12 La media de ventas mensuales de Mr. Tux

PREGUNTAS

1. ¿Qué ideas de pronóstico se le ocurren al estudiar los valores de la media de ventas mensuales de John para los años de sus datos?
2. Suponga que John decide dibujar una línea recta a través de su diagrama de dispersión de modo que se “ajuste bien”, y que extiende después esta línea hacia el futuro utilizando los puntos a lo largo de la línea como su pronóstico mensual. ¿Qué tan precisos cree usted que serán estos pronósticos? Utilice los valores de desviación estándar que John calculó para responder esta pregunta.
3. Con base en su propio análisis, ¿animaría a John a seguir buscando un método de pronóstico más preciso?

PAQUETE DE CÓMPUTO MINITAB

El problema. En el problema 1, Dick Hoover, propietario de Modern Office Equipment, está preocupado por los costos de transporte y de oficina en que incurren los pedidos pequeños. Requiere calcular la media, la desviación estándar y la varianza.

Solución Minitab. Los comandos de MINITAB para resolver el problema 1 son:

```

MTB > SET C1
DATA > 10 15 20 25 15 17 41 50 5 9 12 14 35 18 19 17 28 29 11 11 43 54
DATA > 7 8 16 13 37 18
DATA > END
MTB > DESCRIBE C1

```

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
C1	28	21.32	17.00	20.69	13.37	2.53

	MIN	MAX	Q1	Q3
C1	5.00	54.00	11.25	28.75

```

MTB > BOXPLOT C1

```

```

MTB > STOP

```

- El comando SET se utiliza para introducir la cantidad por transacción de los 28 clientes.
- El comando DESCRIBE se emplea para obtener estadísticas descriptivas de la cantidad variable por transacción, ubicada en C1. Los resultados incluyen el tamaño de la muestra (N), la media (MEAN), una media de los valores restantes después de eliminar el 5% menor y el 5% mayor (TRMEAN), la mediana (MEDIAN), la desviación estándar (STDEV), el error estándar de la media (SEMEAN), el número más pequeño (MIN), el número más grande (MAX), el primer cuartil o 25o. percentil (Q1) y el tercer cuartil o 75o. percentil (Q3).
- El comando BOXPLOT genera una gráfica de caja, que proporciona otra forma de observar un conjunto de datos a fin de determinar su tendencia central, variabilidad y la existencia de valores disparados. La gráfica de caja consta de cinco mediciones de resumen, la mediana, el cuartil inferior, el cuartil superior, la observación más pequeña y la observación más grande. La caja rectangular representa la parte central de los datos, junto con líneas de guiones que se extienden a cada lado, indicando la variabilidad de los datos. El valor de la mediana, en este caso 17, aparece como una marca (+) dentro de la caja. Los extremos izquierdo y derecho de la caja representan el primer y tercer cuartiles de datos. En este ejemplo, el primer cuartil es igual a 11.25 y el tercer cuartil a 28.75.

BIBLIOGRAFÍA

- CARPENTER, J., DELORIA, D. y MORGANSTEIN, D. "Statistical Software for Microcomputers." *BYTE* (abril, 1984): 234-264.
- GROEBNER, D. y SHANNON, P. *Business Statistics. A Decision-Making Approach*. 2a ed. Columbus, Ohio: Charles E. Merrill Publishing Company, 1985.
- HANKE, J. y REITSCH, A. *Understanding Business Statistics*. Homewood, Ill.: Richard D. Irwin, Inc., 1994.
- KEATING, B. "Choosing and Using Business Forecasting Software." *Creative Computing* (enero, 1985): 119-135.
- LACHENBRUCH, P. A. "Statistical Programs for Microcomputers", *BYTE* (Noviembre, 1983): 560-570.
- MAHMOUD, E. "Accuracy in Forecasting." *Journal of Forecasting* 3 (2) (1984): 140-159.
- McCLAVE, J. y BENSON, P. *Statistics for Business and Economics*. San Francisco: Dellen Publishing Company, 1982.
- NEFFENDORF, H. "Statistical Packages for Microcomputers: A Listing." *The American Statistician* 37 (1) (1983): 83-86.
- OLSON, C. y PICCONI, M. *Statistics for Business Decision Making*. Glenview, Ill.: Scott Foresman and Company, 1983.
- SEITZ, N. *Business Forecasting Concepts and Microcomputer Applications*. Reston, Va.: Reston Publishing Company, 1984.
- SINCICH, T. *Business Statistics by Example*. San Francisco: Dellen Publishing Company, 1982.

Fuentes de datos

INTRODUCCIÓN

Una de las partes más difíciles y que ocupa mayor tiempo en los pronósticos es la recolección de datos válidos y confiables. El personal de procesamiento de datos está familiarizado con la expresión “entra basura, sale basura” (GIGO: garbage in garbage out). Esta expresión se aplica también a los pronósticos. Un pronóstico no puede ser más preciso que los datos en los que se basa. El modelo de pronóstico más complejo fallará si se aplica a datos no confiables.

El advenimiento de la computadora ha ayudado a generar una increíble acumulación de información sobre todos los temas. La difícil tarea que enfrentan quienes pronostican consiste en cómo encontrar datos pertinentes que ayuden a resolver sus problemas específicos de toma de decisiones.

Para determinar si los datos serán útiles, se pueden aplicar cuatro criterios:

1. Los datos deben ser *confiables y precisos*. Se debe tener un cuidado adecuado al recolectar los datos, que sean de una fuente confiable y con la debida atención en su precisión.
2. Los datos deben ser *pertinentes*. Deben ser representativos de las circunstancias para las cuales serán utilizados. Los datos que supongan la representación de la actividad económica deberán mostrar las alzas y bajas de acuerdo con las fluctuaciones cíclicas en el pasado histórico de la empresa.
3. Los datos deben ser *consistentes*. Cuando se modifican las definiciones relacionadas con la forma como se reúnen los datos, se deben hacer ajustes para mantener la consistencia en los patrones históricos.
4. Los datos deben ser *periódicos*. Los datos que se recolectan, y resumen y publican con base en una periodicidad serán de gran valor para el pronosticador.

TIPOS DE DATOS

En general, hay dos tipos de datos de interés para el pronosticador. Los primeros son los datos que se reúnen *en un solo punto en el tiempo*, sea éste una hora, un día, una semana, un mes o un cuatrimestre. El objetivo consiste en examinar dichos datos y después extrapolarlos o extender las relaciones reveladas hacia una población mayor. Un ejemplo sería extraer una muestra aleatoria de los archivos de personal para estudiar las circunstancias de los empleados de una compañía.

En contraste, muchos valores de datos de interés para una empresa se recolectan cada día, mes, trimestre o año. En tales casos, se han reunido *series de tiempo de datos*. Las series de tiempo se analizan para descubrir patrones pasados de crecimiento y cambio que se puedan emplear para predecir patrones futuros junto con las necesidades para el funcionamiento de la empresa. Un ejemplo sería el estudio de los montos de ventas mensuales de una empresa. El análisis de series de tiempo no proporciona la respuesta de lo que depara el futuro, pero resulta valioso en el proceso de pronósticos y ayuda a reducir errores en ellos.

Una *serie de tiempo* es una secuencia ordenada de observaciones sobre una variable en particular.

Publicaciones como: *Statistical Abstract of the United States*, *Survey of Current Business*, *Monthly Labor Review*, *Federal Reserve Bulletin* y los informes anuales de corporaciones, contienen series de tiempo de todo tipo. Por lo regular los datos que se reportan en forma mensual, trimestral o anual y que cubren aspectos de precios, producción, ventas, empleo, desempleo, horas trabajadas, combustible utilizado, energía producida, ingresos, etc., llenan páginas de éstas y otras publicaciones económico-empresariales.

Por lo general, los pronósticos a largo plazo son predicciones a futuro de 5, 10 y hasta 20 años. Estas predicciones son esenciales para permitir el tiempo suficiente a los departamentos de planeación, fabricación, ventas, finanzas y otros de una compañía, para que desarrollen planes para probables nuevas plantas, de financiamiento, de desarrollo de nuevos productos y nuevos métodos de ensamble. Tales pronósticos son de particular interés para los altos ejecutivos de una compañía. En contraste, los supervisores de primera línea se interesan principalmente en los pronósticos a corto plazo, ya que sus responsabilidades comprenden aspectos que ocurren al día, semana o mes siguientes.

En los EUA las organizaciones de negocios deben pronosticar los niveles de ventas tanto a corto como a largo plazo. La competencia por obtener los dólares de los consumidores y el deseo de alcanzar una cada vez mayor participación en el mercado, son algunas de las principales fuerzas motivadoras en los negocios. De ahí que se considere necesaria una declaración de las expectativas de la administración, denominada *pronóstico*, para hacer que las materias primas, instalaciones productivas y grupo de personal disponible alcancen la demanda proyectada.

Desde luego, existe la alternativa de no planear a futuro. Sin embargo, en un medio empresarial dinámico, esta falta de planeación pudiera ser desastrosa. Una compañía electrónica que hubiera ignorado la tendencia hacia la televisión a color y los circuitos de estado-sólido, habría perdido ya toda su participación en el mercado. De modo que resulta importante que los administradores comprendan el pasado y utilicen los datos históricos y el buen juicio para hacer planes inteligentes que cubran las demandas del futuro. Los pronósticos se elaboran para auxiliar a la administración a determinar alternativas.

Las consideraciones subjetivas son en extremo importantes en el análisis de las series de tiempo, debido a que aún no se ha encontrado un enfoque de probabilidad satisfactorio para dicho análisis. No obstante, las evaluaciones subjetivas serían necesarias en la formulación de pronósticos, incluso si hubiera disponible un enfoque de probabilidad adaptable al análisis de series de tiempo. Siempre que se examina el pasado en busca de pistas para el futuro, éstas sólo son pertinentes en la medida en que las condiciones causales que estaban en efecto con anterioridad, permanezcan en el periodo por venir. En la actividad económica y empresarial, las condiciones causales rara vez permanecen constantes, por lo que se debe reevaluar continuamente la conexión entre el pasado, el presente y el futuro.

Aunque las técnicas de análisis de series de tiempo no eliminan las evaluaciones subjetivas, hacen una útil contribución, al proporcionar un enfoque conceptual a los pronósticos. Éstos se formulan con la ayuda de un conjunto de procedimientos formales específicos y se indican, de manera explícita, los elementos de juicio.

FUENTES DE DATOS

Las fuentes de datos se pueden clasificar en primarias y secundarias. Las *fuentes secundarias de datos* son datos ya publicados, recolectados con fines diferentes de los que el pronóstico o investigación específicos necesita tener a la mano. Este tipo de datos se puede clasificar a su vez como provenientes de *fuentes internas*, originados dentro de la organización, o de *fuentes externas*, generados fuera de ella. Las publicaciones basadas en censos son buenos ejemplos de fuentes secundarias externas. Con frecuencia, se utilizan los registros de contabilidad como fuentes de datos secundarias internas.

Las *fuentes secundarias de datos*, son datos ya publicados, recolectados con fines diferentes de los que el pronóstico o investigación específicos necesita tener a la mano.

Las *fuentes primarias de datos* comprenden todos los métodos de la recolección de datos originales. Es común que este tipo de datos se reúna mediante procedimientos de muestreo, encuestas de panel o de un censo completo de elementos de interés. Aun es más común el registro semanal, mensual, trimestral o anual de las variables clave de la compañía.

Tales variables de series de tiempo con frecuencia son el centro de atención por parte de la administración.

Las fuentes primarias de datos comprenden todos los métodos de la recolección de datos originales.

Fuentes secundarias de datos

La diferencia entre las fuentes primarias y secundarias de datos es significativa en el sentido de que es más probable que una fuente primaria contenga datos más completos y precisos que los que pudieran encontrarse en una fuente secundaria. Por otra parte, los datos primarios tienden a ser más caros que los secundarios.

Fuentes externas

En años recientes, se ha incrementado en gran medida la cantidad de fuentes de datos publicados, disponibles para los pronosticadores. La proliferación de la computadora es parcialmente responsable de este aumento en la disponibilidad de datos. También ha contribuido a este incremento el hecho de que directores de empresa y de gobierno se dieran cuenta que una mayor y mejor información eleva la efectividad de la planeación y la toma de decisiones. Las bibliotecas están llenas con miles de millones de piezas de datos históricos sobre cualquier tema imaginable. Las organizaciones gubernamentales, servicios computarizados y organizaciones no lucrativas, generan una enorme cantidad de datos estadísticos que pueden emplearse como insumos de entrada para el proceso de pronósticos.

El gobierno de EUA es el más grande editor y recolector de datos del mundo. Cada ciudad grande en EUA tiene por lo menos una biblioteca designada como depósito del gobierno. Estas bibliotecas poseen una amplia colección de datos acumulados por el gobierno.

El Departamento de Análisis Económico [Bureau of Economic Analysis (BEA)], es una fuente importante de datos para las organizaciones que realizan planeación a largo plazo. Este departamento proporciona información básica en lo concerniente a inflación, crecimiento económico, desarrollo regional y el papel de la nación en la economía mundial. El BEA publica la *Survey of Current Business*, una revista mensual que proporciona estimaciones y análisis de la actividad económica de EUA. La revista consta además de dos secciones estadísticas que presentan una composición de datos económicos de fuentes públicas y privadas. La sección de Indicadores Cíclicos de los Negocios, consta de tablas de alrededor de 270 series y diagramas de aproximadamente 130 series que se emplean ampliamente en el análisis de los desarrollos cíclicos actuales. La sección de Estadísticas de los Negocios Actuales consta de tablas para más de 1900 series que cubren actividades empresariales en general e industrias específicas.

Otra fuente de datos gubernamentales que se emplea en los pronósticos, es el censo. Cada censo proporciona información detallada por área geográfica, para diversas características demográficas tales como sexo, edad, ingreso, estado civil y escolaridad. Estos datos pueden ser en extremo valiosos para aquellas compañías relacionadas con el pronóstico de series de tiempo que dependen de dichas características demográficas. Los datos del censo resultan también valiosos para los pronosticadores que realizan estudios de factibilidad. Por ejemplo ¿funcionará un centro comercial propuesto para un área geográfica?, ¿el área alrededor del centro comercial se interesará en un club atlético? Los datos del censo se utilizan con frecuencia para responder este tipo de preguntas.

Otras fuentes importantes de datos gubernamentales son:

- *Statistical Abstract of the United States*, una fuente general de estadísticas sobre muchos temas, cuya publicación es anual.
- *American Statistics Index*, que proporciona una guía para las diversas fuentes estadísticas del gobierno.
- *CIS/Index to Publications of the United States Government*, que representa la lista periódica más completa de publicaciones del Congreso de los EUA.
- *County and City Data Book*, que se publica cada cuatro años, proporciona datos estadísticos y de población divididos por ciudad y condado.
- *Economic Report of the President*, que incluye un apéndice estadístico relativo a la fuerza de trabajo, la inflación y la actividad económica.
- *Employment and Earnings*, que se publica mensualmente, representa una fuente primaria sobre fuerza de trabajo y desempleo.
- *Federal Reserve Bulletin*, que contiene series estadísticas sobre bancos y mercados financieros, así como material incidental sobre la inflación y la actividad económica.
- *Handbook of Labor Statistics*, que presenta datos históricos para muchas de las series que aparecen en *Employment and Earnings*.
- *Statistics of Income*, que publica el *Internal Revenue Service* y contiene información sobre categorías de ingreso y gasto.¹

Es frecuente que organizaciones privadas produzcan pronósticos baratos de mediciones demográficas y económicas. *Business Forecasting in the 1980s* es una bibliografía selecta comentada de los libros y artículos que proporcionan pronósticos generales, así como fuentes de datos de pronósticos específicos.² Ejemplos de dichos servicios son:

- *Forecasting Studies*, que cubre tendencias e indicadores clave.

¹ W. G. Zikmund, *Business Research Methods*, 3a. ed. (Nueva York: The Dryden Press, 1991), pp. 129-137. Bibliografía para fuentes adicionales de datos.

² Bibliografía seleccionada, anotada y compilada por Lorna Daniells of Baker Library at Harvard Business School, Boston.

- *Guide to Consumer Markets*, que proporciona estadísticas y gráficas sobre el comportamiento del consumidor en el mercado.
- *National and Regional Economic Projection Series*, que analiza y hace proyecciones de población, ingreso personal y gastos personales de consumo a nivel nacional, estatal y por áreas estadísticas.
- *Survey of Buying Power*, que contiene proyecciones a corto plazo y tasas de crecimiento para población, vivienda, poder de compra y ventas al menudeo a nivel estatal, de condado y por áreas estadísticas.³

También, organizaciones comerciales publican estadísticas que se pueden utilizar como fuente de datos en muchas aplicaciones de pronóstico. Datos financieros como rendimientos de bonos, porcentajes de utilidad de empresas, dividendos, precios de acciones y volúmenes de comercio, se pueden obtener de Dun and Bradstreet, Inc., Moody's Investors Service, Standard and Poor's Corporation Statistical Service, *The Fortune Directory* de Time, Inc., y *Value Line Investment Survey* de Arnold Berhard and Company.^{1,3}

Muchos de estos tipos de datos se pueden acceder a través de bases de datos computarizadas como Compuserve, Inc., News Retrieval Service del Dow-Jones, Compu-stat de Standard and Poor's y The Source and Value Line.

Diversas compañías de investigación de mercados venden datos con base en una suscripción. La mayor parte de esta investigación de mercado se basa en muestreos de la población. El uso de compañías privadas que proporcionen datos de pronóstico a las organizaciones, puede contribuir a que se distribuyan entre varias de estas últimas el costo de la recolección de los datos primarios, de modo que a una organización individual se le proporcionen datos muy confiables a un bajo costo. Ejemplos de compañías prominentes que proporcionan datos para auditorías de tiendas al menudeo y servicios de seguimiento de productos son A. C. Nielsen Company y Audits and Surveys, Inc. Empresas que proporcionan datos sobre paneles de consumidores y encuestas por correo son Market Facts, Inc., Market Research Corporation of America, National Family Opinion, NPD Research, Inc. y J. Walter Thompson Company. Otra fuente valiosa es Predicasts, Inc, que además de sus propios pronósticos, proporciona resúmenes y bibliografías de pronósticos elaborados por otros.⁴

Datos Propios

Además de los datos recolectados de fuentes públicas, los pronosticadores requieren utilizar datos de fuentes internas dentro de sus organizaciones. En los últimos años, ha habido un

³ R. P. Vichas, *Complete Handbook of Profitable Marketing Research Techniques* (Englewood Cliffs, N.J.: Prentice-Hall, 1982), pp. 36-45.

⁴ T.C. Kinnear y J. R. Taylor, *Marketing Research: An Applied Approach* (Nueva York: McGraw Hill, 1991), pp. 115-117, pp. 146-156.

gran incremento en el establecimiento de bases de datos en las empresas. Cuando una organización desarrolla una base de datos, se reúnen datos para diversas variables y se almacenan en un sistema computarizado de modo que estén disponibles cuando se necesiten. No obstante que los datos particulares en una base de datos varían de acuerdo a las necesidades específicas, existen lineamientos que deben seguirse para establecer cualquier base de datos para pronósticos:

- Debe haber un plan de pronóstico y sólo deben recolectarse los datos requeridos. (El seguir este lineamiento asegura que los datos necesarios estarán disponibles cuando se necesiten y evitar así el gasto y la confusión de datos excesivos.)
- Siempre que sea posible, deberán usarse los datos públicos disponibles, ya que se pueden obtener de manera más rápida y económica.
- Los datos deben almacenarse en su forma original de modo que puedan servir a diferentes propósitos.
- Antes de comenzar a usarlos, los datos deben verificarse y documentarse.
- Los sistemas de bases de datos deben diseñarse para permitir su crecimiento.

Fuentes primarias de datos

Datos de encuestas. Los encuestados son una gran fuente de información primaria para una compañía. Éstos pueden ser clientes, público en general, proveedores o empleados. La finalidad del esfuerzo de recolección de datos es saber qué es lo que las personas seleccionadas piensan sobre elementos clave de interés para la administración. La recolección de datos podría ser un esfuerzo de una sola vez o repetirse de manera periódica para generar series de tiempo de las variables, las cuales se analizarán después y se proyectarán al futuro (pronóstico).

Cuando los investigadores enfrentan la necesidad de reunir una extensa cantidad de datos de un gran número de personas, adoptan alguna forma de encuesta. La primera fase en este proceso consiste en determinar qué método específico de encuesta deberá emplearse para reunir la información requerida. Los tres métodos más comunes son las entrevistas por correo, personales y telefónicas.

En comparación con los otros dos métodos, el cuestionario por correo tiene la ventaja de llegar a un amplio número de personas dispersas, con un costo menor por contacto. Además, los analistas tienen la seguridad de que las preguntas se plantearon en la misma forma a todos los encuestados, reduciendo así la predisposición. Más aún, los encuestados pueden responder las preguntas en su tiempo libre. Por último, el cuestionario por correo puede llegar a individuos difíciles de entrevistar como doctores y otros profesionales que tienen horarios irregulares.

Por otra parte, debido a la lentitud del correo y a la tendencia de los encuestados a postergar el llenado del cuestionario, las encuestas por correo ocupan mucho más

tiempo que las encuestas telefónicas o personales. Así que si se requiere de velocidad, las encuestas por correo son las menos convenientes. Además, de hecho sólo una parte de las personas contactadas responde las encuestas por correo. Estudios pasados muestran que entre el 10 y el 50%, regresan las encuestas, dependiendo de la naturaleza del estudio. Este factor podría sesgar los resultados, si las personas que no respondieron el cuestionario difieren de los que sí lo hicieron en áreas importantes. Por último, sólo pueden incluirse pocas preguntas en las encuestas por correo, debido a que las personas no emplean mucho esfuerzo en leer las preguntas y responderlas. De ahí que aunque esta técnica sea barata por contacto, resulta cara en cuanto a formas respondidas y tiempo ocupado.

El enfoque de la entrevista personal ofrece la ventaja de asegurar buenas técnicas de muestreo, ya que los entrevistadores pueden seleccionar con cuidado a los encuestados. Este método puede arrojar la mayor cantidad de información debido a que los encuestados tienden a ser más abiertos y dispuestos a invertir tiempo respondiendo a las preguntas. Además, este enfoque ofrece la mayor flexibilidad debido a su naturaleza personal. Sin embargo, las entrevistas personales tienen el potencial más alto de desviación del entrevistador, mayores costos por contacto y mayores problemas de control del equipo de personal.

En años recientes, la entrevista telefónica ha ofrecido ventajas crecientes para el analista, como una técnica de recolección de datos. Aunque no puede lograr una muestra completamente representativa de la población, debido a que algunas personas no tienen teléfono o no se les puede localizar en determinadas horas, se acerca a ser representativa para muchos proyectos. El costo por contacto es menor que el de la entrevista personal y los resultados se pueden obtener con rapidez. Además, este método permite un gran control del entrevistador. La desventaja del teléfono es que sólo pueden emplearse preguntas cortas, fáciles de responder. También, mucha gente considera este tipo de recolección de datos como una invasión a su privacidad y no responde. Esta falta de respuesta introduce un sesgo en los resultados de la investigación.

Diseño de encuestas de muestreo. Esta sección presenta un breve repaso de las decisiones que debe hacer el pronosticador en la selección de encuestados o unidades de muestra, para incluir en el proceso de investigación. Al principio, el pronosticador debe responder las siguientes preguntas:

1. ¿Quién deseo que responda a mi encuesta y por qué?
2. ¿Cuántas personas debería incluir en la encuesta? (Tamaño de la muestra.)
3. ¿Cómo seleccionar a las personas a incluir en la encuesta? (Esta pregunta determina la confiabilidad.)

La respuesta a la primera pregunta no siempre es tan clara como podría esperarse. Por ejemplo, considérese el caso en que un fabricante medio de cosméticos contacta 150 varones con el fin de determinar la compra y hábitos de uso de lociones para después de afeitarse. El fabricante se enfrentó con la cuestión de determinar qué características demográficas

ficas deberían tener las personas a incluir en la muestra. ¿La unidad de la muestra debería ser esposos, esposas, parientes, un grupo de edad específico, o una combinación de los anteriores? El fabricante interrogó sólo a varones cuando, de hecho, debió incluir también a las mujeres, ya que en realidad ellas hacen en muchos casos las compras de lociones para después de afeitar. Como otro ejemplo de las dificultades en la selección correcta de la unidad de la muestra, considérese la situación que a menudo enfrentan los investigadores industriales. ¿Deben contactar al comprador que es responsable de las adquisiciones, al ingeniero o gerente de producción que da las especificaciones y utiliza el producto, o al ejecutivo que es con frecuencia quien toma las decisiones? El punto es que cuando los usuarios, compradores y los que toman decisiones no son la misma persona, el pronosticador debe determinar quién de ellos es más probable que tenga la información deseada.

A continuación, el pronosticador o analista debe decidir el tamaño de la muestra. Esto depende del tipo de investigación que se lleva a cabo y de qué tan confiables deben ser los resultados. Por ejemplo, en una investigación exploratoria en donde se buscan ideas más que conclusiones, por lo regular, con tan pocas como 30 entrevistas a profundidad, serán suficientes para cubrir la variedad de áreas del problema. Al mismo tiempo, al pronosticador no le interesa poder decir: “Tengo una confianza de que mis respuestas son 95% correctas con un margen de más o menos 2%”. Una investigación posterior en áreas específicas alcanzará esta meta con un tamaño de muestra mayor. Está reconocido que una mayor precisión en una encuesta significa un mayor tamaño de la muestra, y por ello un costo más elevado. Con frecuencia, el pronosticador elaborará una tabla que compare diversas combinaciones de tamaño de la muestra y costo, contra el margen de error en los resultados de la encuesta. Quien toma las decisiones podrá hacer entonces un balance sustentado entre costo y precisión en la elección del tamaño de la muestra.

Al conducir una investigación concluyente para encontrar respuestas específicas a preguntas específicas, se deben utilizar muestras grandes. Sin embargo, no es necesario muestrear a toda la población o incluso a la mayor parte de ella para lograr resultados precisos. Tal como un catador de vino puede juzgar toda una botella a partir de la primera prueba, un investigador puede analizar un gran mercado a partir de una pequeña muestra. De hecho, las muestras pequeñas que se seleccionan y miden con mucho cuidado pueden ser más confiables que muestras mayores que utilicen estándares menos exactos. Una muestra grande que tenga sesgo, como en el ejemplo de los cosméticos, pueden conducir a un extravío mayor que una muestra pequeña, ya que los resultados erróneos de la muestra mayor aparentan tener gran precisión.

En general, mientras que las muestras más grandes arrojan resultados más precisos que las muestras pequeñas, cuestan más en términos de tiempo y dinero. Por ejemplo, duplicar el tamaño de la muestra podría requerir doblar el costo, así como demorar el tiempo de decisión. De ahí que el pronosticador deba evaluar el valor de más información contra mayor costo y tiempo de recolección de datos.

Como dato interesante, el duplicar el tamaño de una muestra no duplica su precisión y en algunos casos puede agregar muy poca información. La meta del muestreo consiste en seleccionar el tamaño de muestra mínimo necesario para proporcionar la información requerida para tomar una buena decisión, dadas las restricciones fiscales, de tiempo y el grado de precisión deseado.

La tercera cuestión que debe enfrentar el pronosticador, la de determinar un método de recolección de datos, se aborda en la siguiente sección.

Muestreo aleatorio simple. La última pregunta a responder en el diseño del esfuerzo de recolección de datos es ¿cómo seleccionar personas u objetos para la muestra? De nueva cuenta, la respuesta a esta pregunta dependerá del tipo de investigación que se esté desarrollando y de sus objetivos.

En el caso de una investigación exploratoria, la selección de la muestra no es crítica. En la mayoría de los casos el pronosticador puede emplear el buen juicio para seleccionar la muestra, teniendo en cuenta las siguientes características: las personas de la muestra deberán tener un gran conocimiento respecto al problema, ser fáciles de contactar y entablar conversación y estar dispuestas a discutir el problema.

En las investigaciones concluyentes, si el pronosticador desea precisar estimaciones de características de la población y usarlas para predecir el futuro, deberá extraer una muestra aleatoria o una aproximación a ésta. Por desgracia, las muestras aleatorias son por lo regular más caras que las no aleatorias. Los gastos se derivan de 1) intentar encontrar una lista de todos los miembros de la unidad de la muestra (por ejemplo, mujeres entre 18 y 35 años), 2) seleccionar en forma aleatoria una muestra de esta lista, 3) tener que enviar entrevistadores o formas de encuestas a todas partes de una ciudad o estado y, 4) tener que contactar a los encuestados que no estén disponibles al primer intento. En las muestras no aleatorias no se incurre en la mayoría de estos costos.

Las técnicas de muestreo aleatorio aseguran que cada unidad en la población de interés tenga igual probabilidad de ser incluida en la muestra.

¿Cómo se selecciona realmente una muestra aleatoria? Una técnica que puede emplearse es el *muestreo aleatorio simple*. En este enfoque, el investigador debe tener una lista completa de todos los elementos de la población (por ejemplo, nombres de personas). Después, un generador de números aleatorios de una computadora (véase la sección Minitab al final de este capítulo) o una tabla de números aleatorios (véase tabla C.5), se utiliza para seleccionar la muestra.

Por ejemplo, suponga que un mayorista lleva 9 000 elementos y desea seleccionar 250 de ellos para monitorear la prueba de un nuevo sistema de control de inventarios. Sería posible utilizar el proceso de asignar un número a cada elemento, poner los números en tarjetas, revolverlas y extraer 250 al azar. Sin embargo, este procedimiento es caro y ocupa tiempo. Un generador o una tabla de números aleatorios simplifica este procedimiento. Primero, se asigna a cada elemento un número del 1 al 9 000. En seguida, se generan 250 números aleatorios de cinco dígitos que sean menores de 9 000, mediante un generador de números aleatorios de una computadora o se seleccionan de una tabla de números aleatorios (véase tabla C.5). Por último, se incluyen en la muestra los 250 elementos que corresponden a los 250 números aleatorios de cinco dígitos.

Por desgracia, en la mayoría de los casos resulta difícil, costoso o incluso imposible seleccionar una muestra aleatoria simple, a menudo porque no hay disponible una lista completa de la población. De modo que el investigador debe apoyarse en algún otro método. En la siguiente sección se describen varios de los métodos más comunes.

Otras técnicas de muestreo aleatorio. Debido a las dificultades que implica la selección de una muestra aleatoria simple, investigadores y pronosticadores han desarrollado otras técnicas de muestreo que se utilizan ampliamente. Las más comunes son 1) muestreo sistemático, 2) muestreo estratificado y, 3) muestreo en bloque. Como en el muestreo aleatorio, éstas son verdaderas técnicas de muestreo aleatorio, y por ello el pronosticador puede usar técnicas estadísticas establecidas en el análisis y proyección de resultados.

Una *muestra sistemática* es muy útil cuando el analista tiene una lista completa de las unidades de la muestra y éstas se encuentran enlistadas de manera aleatoria. También se puede utilizar la técnica con una población “infinita” en la que los elementos aparecen en secuencia. Un ejemplo podría ser el proceso de llenado de botellas. En esta técnica, el pronosticador selecciona un punto aleatorio de inicio en la lista. Después, se selecciona cada *enésima* unidad u objeto hasta alcanzar el número deseado.

Una variante común de esta técnica se utiliza para seleccionar nombres de un directorio telefónico. Primero, se extrae una página al azar de las primeras páginas del directorio. A continuación, se selecciona al azar una columna de esa página y, por último se elige un nombre de la columna. Suponga, por ejemplo, que el analista necesita 200 nombres y que el directorio tiene 450 páginas. Suponga además, que el analista selecciona inicialmente la página 4, columna 2 y el séptimo nombre hacia abajo. Luego, el analista podría proceder de la siguiente manera: elegir cualquier otra página y seleccionar el nombre de la columna 2 y el renglón 7, hasta llegar a 200 nombres. Este procedimiento es mucho más sencillo, rápido y barato que seleccionar una muestra aleatoria simple.

Las *técnicas de muestreo estratificado* se utilizan cuando el pronosticador sabe qué subgrupos importantes (o estratos) se pueden identificar en la población, subgrupos que son homogéneos dentro del grupo, pero heterogéneos entre grupos. En esta situación una muestra estratificada es estadísticamente más eficiente y barata que una muestra aleatoria simple. La selección de unidades de cada subestrato se hace en forma aleatoria; la proporción del total de la muestra que representa cada estrato se determina mediante el uso de técnicas estadísticas diseñadas para reflejar la importancia y variabilidad de cada subestrato.

Esta técnica se usa, por ejemplo, en la selección de diez tiendas para probar nuevos productos. Un fabricante que desea probar un nuevo producto en tiendas de abarrotes, podría utilizar el hecho de que aunque los grandes supermercados representan sólo alrededor del 20% de todas las tiendas de abarrotes al menudeo, contabilizan el 80% de las ventas totales. De ahí que la compañía podría elegir supermercados para abarcar entre el 70 y 80% de las tiendas de prueba y sólo del 20 al 30% de pequeñas tiendas.

El *muestreo en bloque* comprende el agrupamiento en bloques de la población y la selección aleatoria posterior de bloques, o grupos, en vez de elementos unitarios. Una variante particularmente útil de esta técnica es el muestreo por área. Suponga que un pronosticador desea extraer una muestra aleatoria de la gente que vive en la ciudad de Indianápolis. Ya que no existen directorios disponibles que enlisten a todas las personas

que viven en esta ciudad (los directorios telefónicos no incluyen a todas las personas que tienen teléfono y, por supuesto, tampoco incluyen a los que no lo tienen), se requiere de alguna otra técnica. De modo que se extrae una muestra por área seleccionando aleatoriamente manzanas en un mapa de la ciudad. Después, en cada manzana se entrevista a un número predeterminado de personas. Esta técnica es sencilla y económica. Se ahorran costos, ya que se puede enviar a los entrevistadores a las manzanas de las personas seleccionadas, en vez de hacerlo a localidades por completo divergentes.

Muestras no aleatorias. Cuando el analista y no la casualidad, determina qué elementos se incluirán en la muestra, el proceso de selección es no aleatorio. La finalidad de los muestreos no aleatorios consiste en obtener muestras rápidas, baratas y fáciles de seleccionar; sin embargo, éstas pueden o no ser representativas de la población en estudio. Dos de las técnicas no aleatorias más comunes que se utilizan son las muestras por cuota o conveniencia y las muestras a juicio.

El único criterio principal que se emplea en la selección de *muestras por conveniencia* es la facilidad de selección para el analista. Un ejemplo sería la utilización de estudiantes por profesores para recolectar datos para sus proyectos de investigación. En forma similar, una compañía podría probar el sabor de un nuevo producto, colocándolo en un puesto en un centro comercial y pedir a la gente que pruebe el producto. Este método se usa con frecuencia cuando se requieren estudios exploratorios rápidos y baratos, como en la prueba previa de un cuestionario o la prueba de un nuevo refresco. En ocasiones, a este método se le denomina *muestreo por cuota*, ya que el muestreo continúa hasta alcanzar el tamaño de muestra deseado.

Los *muestreos a juicio* las seleccionan los analistas que usan sus criterios acerca de los elementos a incluir en la muestra. Algunos analistas argumentan que de hecho, pueden reducir errores de muestreo y costos mediante el uso efectivo de técnicas de juicio para reducir unidades de muestra “extremas” o desviadas. Este procedimiento es aceptable en tanto el analista pueda desarrollar una muestra verdaderamente representativa de la población. No obstante, no hay forma de probar estadísticamente el grado hasta el cual la muestra es representativa. Esta técnica se utiliza también principalmente en la investigación exploratoria.

Mercados de prueba. El mercadeo de prueba se está convirtiendo en una herramienta administrativa común. El investigador de mercados o pronosticador dispone la colocación de un nuevo producto o marca en ciudades que piensa serán representativas del mercado geográfico objetivo de la compañía, y observa después, durante cierto periodo, el comportamiento de las ventas del producto en el mercado de prueba. Con frecuencia, este comportamiento se compara con las ventas en los mercados controlados, en los que se ofrece un producto o marca diferente. El objetivo consiste en observar las ventas del producto en un escenario de pequeña escala y en la suposición básica de que ese comportamiento en dicho escenario pronosticará el comportamiento en el mercado más amplio.

Los resultados de las pruebas de mercado proporcionan a la administración los datos que requiere para evaluar oportunidades potenciales de ganancia. Parte de la información que se puede obtener en una buena prueba de mercado es la siguiente:

Una participación potencial de mercado del nuevo producto
¿Quién compra el producto, con qué frecuencia y con qué fin?
Por lo regular, ¿en dónde se realizan las compras?
Los cambios de estrategia que haga la competencia
El efecto del nuevo producto sobre marcas ya establecidas

En resumen, uno de los aspectos más difíciles y que ocupan tiempo en los pronósticos es la recolección de datos válidos y confiables. En este capítulo se abordó la difícil tarea que enfrentan la mayoría de quienes pronostican, de cómo encontrar datos pertinentes que les ayuden a resolver los problemas de sus aplicaciones. Las fuentes de datos se pueden clasificar ya sea como primarias o secundarias. Las fuentes de datos primarias comprenden todos los métodos de la recolección original de datos. Las fuentes de datos secundarias son datos ya publicados, reunidos con fines distintos a los que el pronosticador o investigador necesita tener a la mano. La importancia para los pronosticadores de distinguir entre fuentes primarias y secundarias abarca el hecho de que es más probable que una fuente primaria de datos contenga datos más completos y precisos que los que pudieran encontrarse en una fuente secundaria. Por otra parte, los datos primarios son, por lo regular, más caros de obtener.

En años recientes, ha habido un incremento sustancial en la cantidad de fuentes de datos publicadas (datos secundarios) disponibles para los pronosticadores. Las bibliotecas, fuentes gubernamentales, servicios computarizados y organizaciones no lucrativas proporcionan grandes cantidades de datos estadísticos que se pueden emplear en los pronósticos.

La recolección de datos primarios requiere de la selección de los elementos que se incluirán en la muestra. En este capítulo se expusieron diversos métodos para ello: muestreo aleatorio simple, muestreo sistemático, muestreo por bloques y estratificado, muestreo a juicio y muestreo por cuota o conveniencia. Si los elementos seleccionados son personas, hay muchas formas de acercarse a ellas para determinar sus características y opiniones. Los cuestionarios por correo, encuestas telefónicas y las entrevistas personales son los principales métodos de recolección de datos que se abordaron en este capítulo.

APLICACIÓN EN LA ADMINISTRACIÓN

Las decisiones que enfrentan los administradores en todas las organizaciones requieren, por lo regular, cierta manipulación o análisis de datos. Para tales análisis es necesario que estén disponibles los datos apropiados, y este capítulo se refiere a la recolección de dichos datos pertinentes. Ya que la mayoría de las decisiones importantes en los negocios requieren de análisis, resulta en extremo importante tener a la mano el conjunto de datos pertinentes para esa decisión.

Se pueden efectuar muchas decisiones importantes en los negocios después de examinar datos ya existentes, es decir, con datos secundarios. Es importante que cada administrador esté familiarizado con las fuentes secundarias de datos, tanto dentro de la compañía como en agencias exteriores tales como gobierno, las bibliotecas y las universidades. Muchas situaciones de decisión se ven auxiliadas en gran medida por la sencilla y barata recolección de los datos existentes en las fuentes antes mencionadas.

Es aún más generalizada, en la toma de decisiones, la necesidad de reunir datos específicamente diseñados para el problema en cuestión. La recolección de dichos datos primarios se aborda en este capítulo; los administradores de empresas necesitan familiarizarse con estas técnicas, tanto para diseñar sus propios estudios de investigación, como para evaluar en forma efectiva los esfuerzos de otros. En cualquier situación de toma de decisiones es importante saber si los datos que se reunieron y analizaron son los necesarios para resolver el problema y si representan los mejores datos posibles para esta situación.

A continuación se presentan varias preguntas que podrían formularse los administradores de empresas en diversas circunstancias. La respuesta a cada pregunta implica la recolección de datos, ya sean primarios o secundarios, y la realización de un análisis adecuado de los mismos. La comprensión del material que se presenta en este capítulo es esencial para cada una de estas preguntas, si se van a identificar o recolectar datos pertinentes.

- ¿Se ha incrementado la participación en el mercado como resultado de nuestra nueva campaña publicitaria?
- ¿Cuál es la opinión de la audiencia sobre nuestro nuevo equipo ancla de noticias televisivas?
- ¿Cuál es el balance normal de efectivo para compañías similares a la nuestra?
- ¿Qué piensan de nuestros servicios los miembros de nuestra unión de crédito? ¿Qué mejoras pueden sugerir?
- ¿Cómo ven nuestros accionistas la efectividad del equipo de administración de la compañía?
- ¿Qué tendencias económicas son evidentes en nuestra industria con base en los sucesos históricos?
- ¿Cómo se relacionan nuestras ventas mensuales con los indicadores económicos regionales importantes?
- ¿Qué factores se pueden identificar que afecten el abstencionismo de nuestra fuerza laboral por tiempo?
- ¿Cuál es la opinión de los compradores del área sobre nuestra tienda, en especial, comparada con nuestros principales competidores?
- ¿Cuáles son las actitudes de varios departamentos hacia el grupo de apoyo de cómputo de la compañía?
- ¿Qué efecto tiene, para los compradores, la posición de los estantes para seleccionar los comestibles?
- ¿Cuál es el mercado de cajas de velocidades manuales diseñadas para vehículos de doble tracción?
- ¿Cuál es la opinión de los visitantes de nuestra ciudad con respecto a compras, restaurantes, hoteles, moteles y diversiones?

GLOSARIO

Fuentes primarias de datos. Las fuentes primarias de datos comprenden todos los métodos de recolección de datos originales.

Fuentes secundarias de datos. Las fuentes secundarias de datos son datos ya publicados, reunidos para fines distintos a los que el pronosticador o investigador específico necesita tener a la mano.

Muestreo aleatorio. Las técnicas de muestreo aleatorio aseguran que cada elemento en la población de interés tenga la misma posibilidad de ser incluido en la muestra.

PROBLEMAS

1. ¿Cuál es el papel de los datos secundarios en el proceso de pronóstico? ¿Cuál el de los datos primarios?
2. ¿Cuál es la ventaja de los datos primarios en relación con los secundarios?
3. Sugiera algunos ejemplos de datos secundarios que sean útiles a los pronosticadores.
4. ¿Cuáles serían las ventajas de utilizar cada uno de los siguientes métodos para llegar a los estudiantes de su universidad para realizar una encuesta?
 - a. Correo
 - b. Teléfono
 - c. Entrevistas personales
5. ¿Cuál es la diferencia entre una muestra aleatoria y una no aleatoria?
6. ¿Cuáles son las ventajas de cada una de las siguientes técnicas de muestreo?
 - a. Muestra aleatoria simple
 - b. Muestra en bloques
 - c. Muestra por conveniencia
 - d. Muestra aleatoria estratificada
 - e. Muestra sistemática
 - f. Muestra a juicio
7. A Julie Creston, administradora de la tienda departamental Bon Marche, le gustaría desarrollar un modelo para pronosticar cada mes el nivel de reducción del inventario, de modo que pudiera tomar las precauciones adecuadas para prevenir pérdidas. Julie sospecha que algunos robos han reducido el inventario. Exponga los tipos de fuentes de datos requeridos para investigar este problema.
8. La Reservación India de Spokane está ubicada en la parte oriental de Washington, aproximadamente 80 kilómetros al noroeste de la ciudad de Spokane. Las fronteras al este y al sur de la reservación están comprendidas por 58 kilómetros de la orilla del lago Roosevelt.

Durante los últimos dos años, la tribu ha investigado el potencial para un sitio de desarrollo vacacional en la confluencia que tienen en el lago Roosevelt los ríos Spokane y Columbia. Los planes tentativos incluyen la posibilidad de contruir lo siguiente: un parque recreativo vehicular, un motel (con habitaciones y foros para reuniones), una marina (un muelle y veleros para renta), una operación de renta de casas-flotantes, una tienda de abarrotes y aparejos de pesca, un

merendero, un área de playa y natación, un centro cultural con tienda de regalos y un casino. Además, existe también interés en desarrollar un campo de golf. La tribu Spokane ha solicitado propuestas para un análisis de mercado y un estudio de factibilidad para estos negocios. Antes de preparar su propuesta, necesita usted responder las siguientes cuestiones:

- a. Enliste las fuentes primarias de datos que se usarían en el desarrollo del análisis de mercado. Explique cómo se utilizaría en el estudio cada una de estas fuentes.⁵
 - b. Enliste las fuentes secundarias de datos que se usarían en el desarrollo del análisis de mercado. Explique cómo se utilizaría en el estudio cada una de estas fuentes.
9. Uptown Radiology está desarrollando un centro médico de imágenes más completo y tecnológicamente avanzado que cualquier otro actualmente ubicado en el área de New England. Uptown Radiology solicitó a la empresa PMA Research que evaluara el mercado y realizara una proyección de utilidades para cinco años. Algunos de los objetivos de este estudio serán:
- Identificar áreas de mercado para cada tipo de procedimiento ofrecido por las nuevas instalaciones.
 - Reunir y analizar los datos que se tienen sobre los ingresos existentes en el área de mercado para cada tipo de procedimiento que se ofrece en las nuevas instalaciones.
 - Identificar las tendencias en la rama de atención a la salud que afecten positiva o negativamente los ingresos por los procedimientos que ofrecen las nuevas instalaciones.
- a. ¿Qué fuentes secundarias podrían usarse para elaborar este estudio?
 - b. ¿Qué fuentes primarias podrían usarse para elaborar este estudio?
10. Localice en su biblioteca cada una de las siguientes publicaciones y registre una serie de tiempo de cada una durante varios periodos: *Survey of Current Business*, *Statistical Abstract of the United States*, *Standard and Poor's Statistical Service* y *Moody's Investors Service*.
11. Desarrolle un cuestionario para utilizar con estudiantes de su clase, para explorar sus actitudes hacia los acontecimientos actuales en la nación y en el campus. Incluya también preguntas demográficas como sexo, edad, especialidad, etc. Haga una prueba previa de su cuestionario solicitando comentarios a sus compañeros de clase sobre puntos a mejorar. Suponga que las encuestas llenadas se alimentarán en un programa de cómputo y diseñe las preguntas pensando en ello.

CASO DE ESTUDIO 3.1

ARGYL FOOD PRODUCTS⁶

Rob Arabicaf, un especialista en economía aún no graduado del Premier State College, se entusiasmó al ser aceptado en un programa interno de verano en la Argyl Food Products Company. Recientemente, Argyl absorbió a un competidor cuya línea incluía diversos productos de café. Dichos productos se

⁵ "The Spokane Indian Tribe Feasibility Study, "Decision Science Institute Proceedings, New Orleans, 1991.

⁶ This case was contributed by William C. Struning of the General Foods Corporation, Nueva York, Nueva York.

asignaron a la división de bebidas de Argyl, dirigida por Sue Maxwell. El café era nuevo para Argyl y Sue pensó que necesitaba saber hacia dónde iba el mercado del café. Pensó que esto proporcionaría a Rob, el economista neófito de Argyl, un útil proyecto.

La solicitud de Sue para que Rob prestara sus servicios fue aprobada y éste, quien ni siquiera bebía café se vio de pronto enfrentado a elaborar una predicción sobre la que Argyl basaría una decisión clave. Si se estaba extendiendo la demanda de café en los Estados Unidos, Sue colocaría fondos considerables para tener una mayor participación en el mercado. Pero, si parecía que el potencial del café se contraía, Sue consideraría colocar los productos de café de Argyl en una compañía más especializada en ese campo.

Rob pensó que podría tener una mejor oportunidad de proporcionar un pronóstico realista y confiable si aprendía un poco más acerca del café. Ya había tomado cursos que lo familiarizaron con técnicas de formulación de pronósticos. Decidió elaborar un análisis preliminar o de situación, de modo que su pronóstico final fuera más comprensible y útil a la administración.

Rob contactó diversas personas clave en el grupo de café heredado por Argyl. Le fueron de gran ayuda, pues tenían amplios antecedentes en los aspectos técnicos del café, pero en realidad no sabían mucho respecto a patrones amplios de consumo de café. Así, Rob se dio cuenta de que tendría que analizar publicaciones sobre el tema del café, en particular las relacionadas con su consumo en EUA.

Hace más o menos una década, la búsqueda de literatura habría requerido de muchas horas de examinar índices de tarjetas de publicaciones en las bibliotecas, así como el buscar en guías de publicaciones periódicas. Pero Rob recordó que en la biblioteca de la universidad podía efectuar una búsqueda por computadora de publicaciones disponibles. El bibliotecario ayudó a Rob a conectarse a un servicio denominado Dialog (Dialog Information Services, Inc.). Dialog es uno de muchos grandes servicios de información que ofrecen datos y textos vía computadora (en línea). Las ofertas de Dialog consisten en más de 100 bancos de datos, cada uno creado, actualizado y mantenido por una organización especializada en información para ese campo en particular. A Rob le satisfizo aprender que podía tener acceso a una de las bases de datos de Dialog, Coffeline, aun cuando esta base la mantiene y da servicio la Organización Internacional del Café en Londres, Inglaterra. Cuando Rob se conectó, se creó un registro de su tiempo de conexión y volumen de información obtenida, datos que conforman la base para hacerle el cargo correspondiente.

Rob se encontró con que había muchas opciones sobre consumo de café en EUA. Seleccionó una impresión que incluía un breve resumen de cada artículo, así como su referencia bibliográfica. Para leer el texto completo de un artículo dado, habría tenido que obtener el texto original. Varias compañías que trabajan a través de Dialog proporcionan impresiones de textos completos. Por desgracia, este servicio no está disponible para Coffeline. No obstante, la lectura de los resúmenes le proporcionaron muchos acercamientos a su área de interés y pudo localizar textos completos de varios artículos relacionados, contactando directamente a los editores.

A continuación, Rob decidió investigar si existía algún pronóstico previo publicado sobre el consumo de café en EUA. Si había disponible tal estudio, podría utilizarlo sin ocupar tiempo y otros recursos en elaborar su propio pronóstico. En esta ocasión seleccionó otra base de datos de Dialog, PTS Forecasts Abstracts. Los creadores de esta base de datos, Predicasts, se especializan en resumir cualquier pronóstico que puedan encontrar en literatura impresa y en poner a disposición los resúmenes a los suscriptores de Dialog. Buscando en el tema de café en EUA, Rob encontró una proyección general de consumo que preparó y publicó originalmente una agencia del gobierno. Pero tendría que hacer de todos modos su propio pronóstico, pues los datos que tenía a mano no presentaban subdivisiones (por edad, ingreso, etc.), aunque sí proporcionaban buenos parámetros de consumo agregado con los que podría comparar su propio trabajo.

A través de su uso de Dialog, Rob aprendió que la Organización Internacional del Café era la fuente mundial más importante de datos sobre café. Su solicitud a esta organización se vio recompensada al recibir datos sobre el consumo de café en EUA. Para determinar qué modelo usar para su pronóstico, usó un paquete de software estadístico para realizar un análisis exploratorio. Una simple gráfica de series de tiempo y un correlograma (exhibición de autocorrelaciones para varios periodos [se abordan en el cap. 4]), sugirieron que había en los datos una suave tendencia a la baja y un patrón estacional. Parecía también que había diversos “valores extremos”, *v.gr.* valores muy por afuera del nivel general. Rob tuvo entonces que decidir si hacía la predicción atenuando y extrapolando los datos en el futuro (con frecuencia denominado en el mundo de los negocios como *análisis técnico*), o si hacía inferencias a partir de las relaciones entre los datos de café y otras series de datos. Rob no estaba al tanto de que existen modelos que combinan estos dos enfoques.

De lo que había aprendido hasta ahora, Rob pensó que la cantidad de café consumido se podría explicar mediante influencias externas, como el precio del café al menudeo o el ingreso personal disponible. Si, por decir, el consumo de café y el ingreso estuvieran relacionados, sería posible pronosticar el consumo para varios niveles de ingreso. Recordó, de un curso básico de estadística, que podía medir y proyectar las relaciones de consumo de café y otras variables por medio de técnicas de correlación y regresión. De ahí que decidió explorar el enfoque de las relaciones.

Realizar un análisis de regresión en una computadora es relativamente rápido y fácil, siempre que se tengan disponibles y listos para utilizar todos los datos necesarios. Rob se dio cuenta de que ya que no tenía los datos a la mano, la mayor parte de su trabajo consistiría en localizar, copiar e introducir datos en su computadora. Por desgracia, Rob no tenía acceso a servicios como Data Services, Inc., I. P. Sharpe o Wharton Electronics, los cuales se especializan en mantener en línea series económicas para cargar directamente mediante un módem al espacio de trabajo de una computadora. Pero Rob localizó un servicio, Commodity Information, Inc. (CII), que pone en línea estadísticas de precio y volumen para diversos productos, incluyendo café. CII también proporciona software para permitir a sus clientes analizar los datos recibidos de ellos. Al tener disponible la información, Rob ejecutó diagramas y regresiones múltiples (véase cap. 6), para decidir qué variables independientes serían útiles para explicar el consumo de café.

Rob pensó que este análisis preliminar le había sido en extremo útil para el diseño de su proyecto formal de pronóstico. Ahora tenía un objetivo claro, un conocimiento básico del café y varias ideas sobre lo que producía los cambios en el consumo del café. Además, podía ahora presentarle a su jefa, Sue Maxwell, las estimaciones de cuánto le llevaría terminar su proyecto y cuánto costaría. De gran ayuda para estas estimaciones fue un paquete de cómputo de administración de proyectos que Sue Maxwell compró para él, de modo que pudiera afinar sus planes y estimaciones, así como mantener un control del avance del proyecto.

PREGUNTAS

1. ¿Cuál es la finalidad de realizar un análisis preliminar? ¿Es siempre necesario? ¿Puede usted justificar el tiempo y costo de realizar un análisis de este tipo?
2. ¿Qué otras alternativas pudo utilizar Rob al realizar el análisis preliminar?
3. ¿Qué otras bases de datos con información de café podría haber consultado Rob dentro de Dialog?

4. ¿Qué otros servicios de datos numéricos o estadísticos pudo utilizar Rob, si hubiera pagado las cuotas correspondientes?

CASO DE ESTUDIO 3.2

SOLOMON'S JEWELRY

Solomon's es una importante joyería ubicada en uno de los más prestigiosos centros comerciales en Houston. A lo largo de los años ha desarrollado una excelente reputación de manejar productos finos, seleccionados para atraer a la gente que desea productos de calidad. La Sra. Solomon, presidenta de la compañía, piensa que su tienda tiene una clientela leal de altos ingresos, que en su mayoría se muestran satisfechos con el servicio que reciben en Solomon's. El único problema grave que Solomon observa en el horizonte cercano es que un nuevo e importante competidor planea abrir una tienda en el mismo centro comercial dentro de dos meses. Aunque su competidor es un comerciante de una cadena de tiendas, tiende a enfatizar en mercancía de alta calidad. No obstante, la Sra. Solomon se siente segura de que su empresa puede competir efectivamente haciendo énfasis en la línea distintiva de los productos que maneja, su compromiso a largo plazo con los clientes y la alta calidad de su servicio.

Mr. Wind, contratado recientemente como vicepresidente de comercialización, es el responsable de velar porque los esfuerzos de comercialización se diseñen para complementar las prácticas de comercialización anteriores, así como contrarrestar los efectos de los esfuerzos comerciales del nuevo competidor. Wind, quien fue profesor en la universidad local, así como consultor durante mucho tiempo para la industria de joyería al menudeo, está en general complacido con lo que ve en Solomon's. El plan de comercialización está bien diseñado y coordinado, los productos son adecuados para el mercado objetivo, los precios son altos pero compatibles con la calidad del producto, los vendedores son conocedores y amables y los esfuerzos promocionales parecen efectivos.

Después de varias semanas en su nuevo empleo, Wind decide visitar las tiendas del nuevo competidor existentes en Dallas, para tratar de juzgar el impacto que podría tener en las operaciones de Solomon's. Una evaluación cuidadosa de la tienda del competidor parece indicar que está enfocada a clientela de ingresos superiores y que proporciona un servicio cortés y amistoso. Sin embargo, su línea de productos, aunque es un poco menos distintiva, también es aproximadamente un 20% más barata.

A su regreso a Houston, Wind le sugiere a la Sra. Solomon elaborar una encuesta para saber más sobre los problemas potenciales que su empresa pudiera enfrentar al abrirse la nueva tienda. Ella está de acuerdo con este plan. Así, si usted fuera Wind, ¿cómo respondería a las siguientes preguntas?

PREGUNTAS

1. ¿A quién debería contactar?
2. ¿Cómo se les debería contactar?
3. ¿Cuándo se les debería contactar?
4. ¿Qué preguntas se les deben hacer?

CASO DE ESTUDIO 3.3**MR. TUX**

John Mosby, propietario de varias tiendas Mr. Tux, está comenzando a pronosticar la variable más importante de su negocio, las ventas mensuales en dólares. (Véanse los casos de estudio Mr. Tux en los capítulos 1 y 2.) Una de sus empleadas, Virginia Perot, reunió los datos de ventas que se presentan en el caso del capítulo 2. Al pensar cómo hacer un pronóstico efectivo, siente curiosidad sobre la precisión de estos datos.

En una conversación con Virginia, él aprendió que la recolección de datos no fue tan sencilla como pensaba al asignarle esta tarea. Ella encontró la mayor parte de las cifras de ventas en los libros mayores de contabilidad, pero faltaban algunos. Con una sola excepción, pudo reconstruir los volúmenes de ventas a partir de otros registros contables y con la asistencia de la compañía de contabilidad que lleva los libros de la empresa.

El valor faltante, el de marzo de 1986, no se pudo localizar. Después de estudiar los patrones de datos, Virginia decidió estimar este valor para tener completos los ocho años de datos de ventas mensuales. Ella razonó que era probable que su estimación estuviera cerca del valor real faltante y que cualquier esfuerzo de pronóstico no sería desechado. Después de observar los valores de ventas durante este periodo se decidió por un valor de \$26 901 para marzo de 1986 (este dato se muestra en el caso del capítulo 2). Después de escuchar su historia, John estuvo de acuerdo en que su enfoque era razonable y que los pronósticos que él efectuara no se verían afectados de manera adversa por la estimación que hizo ella.

Al final de la reunión, Virginia preguntó a John qué pensaba acerca de realizar una encuesta a clientes de rentas anteriores. Esta posibilidad no se le había ocurrido a John, con lo ocupado que estaba administrando el negocio y concentrado en el desarrollo de algunos pronósticos. Cuando le preguntó a Virginia por qué sugería una encuesta, ella dijo:

Creo que sería bueno saber cómo estamos trabajando ante los ojos de nuestros clientes. A veces he oído quejas sobre diferentes aspectos, pero no sé si se trata de incidentes aislados o si hay formas en que podamos mejorar sustancialmente la forma de conducir el negocio.

PREGUNTAS

1. ¿Cómo hubiera usted manejado el problema del dato faltante?
2. ¿Cuál es su opinión respecto a la idea de Virginia de realizar una encuesta?

CASO DE ESTUDIO 3.4**CONSUMER CREDIT COUNSELING**

Marv Harnishfeger, Director Ejecutivo de Consumer Credit Counseling (CCC), empresa descrita en el capítulo 1 (caso de estudio 1.2), estaba explicando la operación a Dorothy Mercer, recientemente elegida Presidenta del Comité Ejecutivo. Marv le indicaba que le gustaría entender mejor tres mercados:

1. *El problema de administración de deudas*: Un cliente enfrenta a menudo crisis financieras (posible bancarrota) y CCC le proporciona un programa alternativo denominado programa de administración de deudas, para evitar la bancarrota.
2. *Educación*: Educación secundaria, educación sobre compras racionales para el hogar, Planeación Financiera Básica, Educación sobre Presupuestos Corporativos.
3. *Ayuda financiera a mayores*: Éste es un nuevo programa de CCC diseñado para ayudar a personas mayores a consolidar sus chequeras, pagar cuentas, negociar con acreedores y manejar problemas médicos.

Marv le explicó a Dorothy que, en su opinión, CCC tenía un problema de crecimiento y que sentía la necesidad de planear de manera estratégica su mejor colocación en los tres mercados durante los siguientes tres años. Dorothy estuvo de acuerdo y comenzó a enlistar varias preguntas sobre el mercado objetivo que ella pensaba necesitaban responderse. Marv interrumpió a Dorothy a la mitad de una oración y le pidió que escribiera su lista de preguntas. Dorothy entregó a Marv la siguiente lista:

- ¿Quiénes son nuestro mercado?
- ¿Está creciendo?, si es así, ¿qué tan rápido?
- ¿Dónde está el mejor mercado?
- ¿Hacia dónde nos deberíamos expandir?
- ¿Qué tan rápido podemos esperar la expansión?
- ¿Cómo podemos determinar nuestra clientela básica en cada mercado?
- ¿Cómo podemos identificar nuestra penetración actual?
- ¿Cómo podemos determinar nuestro punto de saturación?
- ¿Cuáles son los pronósticos para el uso de crédito?
- ¿Cuál es el pronóstico para estos pronósticos?

Para intentar responder algunas de las preguntas de Dorothy, Marv le proporcionó a Dorothy el siguiente documento, relativo a la penetración del mercado, de la National Foundation for Consumer Credit (Fundación Nacional para Crédito al Consumidor):

Desde la aparición del informe de actividades de 1992, hemos recibido varias preguntas acerca del uso e interpretación de los datos. Esta semana, queremos abordar la penetración de mercado de nuestras actividades, para examinar si estamos *ampliando* el servicio a todos aquellos que lo necesitan. Definida llanamente, la penetración de mercado es una medida de cómo una actividad cubre una necesidad.

Por ejemplo, digamos que se tienen un millón de personas que viven en su área de servicio. Esto significa que hay probablemente 400 000 familias en su área (alrededor de 2.5 personas por familia). Si efectuó las primeras 1 000 sesiones de consultoría en 1992, su penetración de mercado de las sesiones sería de 1 000 sesiones divididas entre 400 000 casas = .0025 sesiones por familia. Para trabajar con números más grandes, multiplicamos por 1 000 para obtener la penetración por cada mil casas, lo que resulta 2.5 sesiones por cada mil familias.

Después de definir cómo se calcula la penetración, las mejores preguntas son, ¿qué tan grande es nuestro mercado y cuál es una “buena” proporción de penetración de mercado? Para la primera parte de esta pregunta, los expertos opinan que las familias que actualmente requieren de nuestros servicios varían entre un 3 y un 15%. Seamos conservadores y elijamos el 5% de las familias como el grupo actual de familias que necesitan nuestra ayuda.

Usando el ejemplo de las sesiones de consultoría, esto significaría que de las 400 000 familias en nuestra área de servicio, 20 000 necesitan de consultoría de inmediato. Para cubrir la demanda actual, se requeriría de una proporción de penetración de mercado de 50.0, contra nuestra proporción de 2.5 sesiones por cada mil casas. Nuestra actividad de 2.5 sería de alrededor del 5% de lo que necesitaríamos para cubrir a todos los que necesitan ayuda.

A nivel nacional, hay 92 millones de familias en nuestra área de servicio, de las cuales 4.6 millones necesitan ayuda inmediata. Realizamos cerca de 600 000 sesiones de consultoría o cerca del 13% de ese grupo necesitado. Es obvio que tenemos por delante mucho trabajo si queremos atender a los que lo necesitan. Es necesario que avancemos intentando lograr tasas de penetración de 50 por cada mil familias si queremos ser una asociación de servicio completo de costa a costa.

Recuerde que aquellos a los que no atendemos son los principales candidatos para la quiebra, para ser atendidos por otros servicios de consultoría no lucrativos, o por clínicas sin escrúpulos de “reparación de créditos”. Si trabajamos en conjunto en un esfuerzo nacional coordinado, podemos estar seguros de que vivimos para cumplir nuestra misión de ser un “servicio” para los necesitados.

TAREA

1. Enliste las fuentes primarias de datos necesarios para desarrollar un análisis de mercado. Explique la utilidad de cada fuente primaria.
2. Enliste las fuentes secundarias de datos necesarios para desarrollar un análisis de mercado. Explique la utilidad de cada fuente secundaria.
3. Escriba un memorando para Marv que evalúe el enfoque utilizado por la Fundación para determinar la penetración en el mercado.

CASO DE ESTUDIO 3.5

ESTUDIO DE FACTIBILIDAD DE MERCADO

La McMillian Development Corporation solicitó a una compañía local de investigación que realizara un estudio de factibilidad de mercado, para determinar los factores de mercado que afectarían el éxito de un club atlético en el centro comercial de Northwood en Peyton, Wisconsin. La investigación abarca el análisis de datos demográficos, datos sobre flujo de tráfico, imagen percibida de los competidores y área potencial de crecimiento. Al concluirse el estudio, Mark Craze, analista investigador de la compañía de investigación, preparó el siguiente informe.

OBJETIVOS

Los objetivos de este estudio consisten en determinar:

- Si la población (en especial entre 18 y 54 años) del área de Northwood es lo suficientemente grande para aportar un número satisfactorio de miembros para el nuevo club.
- Si los residentes del área están interesados en afiliarse a un club atlético.
- Si los residentes del área pueden cubrir la afiliación al club atlético propuesto.
- Si un club atlético con una atmósfera moderna tiene el potencial de atraer socios de las membresías de los clubes existentes.

METODOLOGÍA

Las metodologías utilizadas para realizar este estudio son

Un análisis de información demográfica para el área principal de mercado de Northwood, incluyendo una investigación sobre crecimiento potencial.

Una comparación estadística de los datos demográficos del área principal de mercado de Northwood con el área principal de mercado del Westwood Athletic Club de Cadott, Wisconsin.

Un análisis del flujo de tráfico en el área de Northwood, utilizando fuentes secundarias de investigación y entrevistas personales.

Investigación de la imagen percibida de los competidores del área, mediante entrevistas personales con afiliados a la competencia.

Un análisis del sitio propuesto a través de una observación personal.

ANÁLISIS

La ubicación

El desarrollo propuesto se pretende ubicar en una superficie de 3 600 metros cuadrados en la parte suroeste del centro comercial de Northwood, mismo que se localiza en West Salnave Road al norte de los límites de la ciudad de Peyton, hacia el oeste de Main Street y ocupa casi una cuadra a través de Salnave Road desde la secundaria Northwest.

Accesibilidad. El centro comercial de Northwood está aproximadamente a 1.6 kilómetros de Main Street, que es la arteria de la ciudad de Peyton usada con más frecuencia en dirección norte-sur. La mayoría de la población que vive al norte del centro, pasa cerca de Northwest en sus viajes a Peyton. También, el centro comercial está bordeado por Walnut Street, que proporciona un fácil acceso para los residentes que viven al sur y suroeste.

Medio circundante. El centro comercial de Northwood consta de aproximadamente 11 200 metros cuadrados de espacio rentable, que actualmente ocupan 27 arrendatarios. Hay espacios disponibles para más de 900 automóviles.

El centro se encuentra ubicado en un vecindario suburbano de gran afluencia al norte de la ciudad de Peyton. El sitio de compras disponible más cercano al vecindario está aproximadamente a cinco kilómetros al sur de la calle Main o a 15 minutos al norte del centro de Dayton, Wisconsin.

Desarrollo del área. Todo indica que continuará un patrón continuo de crecimiento al norte de la ciudad de Peyton. Se construyeron edificios de oficinas a lo largo de la calle Main, y el desarrollo de negocios a lo largo de esta arteria principal norte-sur, continúa desplazándose hacia el norte. Un informe de investigación sobre bienes raíces en el condado de Peyton, efectuado en el otoño de 1993, indica lo siguiente:

1. El número de registros de hipotecas en el Condado de Peyton en el segundo y tercer trimestre de 1993, aumentó sustancialmente con respecto al mismo periodo de 1992.
2. El número de nuevos lotes fue alto, tanto en la parte noroeste del condado como en la ciudad de Peyton en 1992 y 1993.

3. En 1993 se registraron un gran número de unidades de condominios en la parte noreste del condado de Peyton.
4. Datos recientes muestran que el número de ventas de residencias en el condado de Peyton se incrementó en los dos últimos trimestres y que parece aumentar el precio promedio de venta en el tercer trimestre de 1993.
5. El desempleo en el condado de Peyton está disminuyendo sustancialmente.

Flujo del tráfico. El flujo del tráfico en el área de Northwood parece correr, por lo general, de norte a sur. Una encuesta de comercialización realizada por estudiantes de mercadotecnia de la Universidad Whitewater, en el centro comercial de Northwood, mostró que la mayoría de los clientes viene a comprar a Safeway, algunos a Giant-T y otros acuden a realizar actividades específicas. La mayoría de los compradores proviene de las cercanías y acude más por comodidad que por preferencia de alguna tienda. Es evidente que la gente prefiere viajar al sur, hacia el centro de Peyton, para cubrir sus necesidades de compras mayores.

La competencia

Northtown Racquet Club. Este club se ubica a un kilómetro y medio de la calle Metro sobre la calle Main. Las instalaciones tienen seis canchas de tenis, cuatro de raquetball y una piscina exterior, un salón, saunas, un gimnasio y un jacuzzi. El club tiene planeado para el próximo año convertir una de las canchas de tenis en una piscina cubierta. Tiene alrededor de 1 000 socios, 900 de los cuales provienen del área de mercado de Northwood. La cuota familiar inicial para el club de tenis es de 300 dólares*, con cuotas mensuales de \$30 dólares y una tarifa de 3 dólares por hora de uso de las canchas. La cuota individual inicial es de 200 dólares, con cuotas mensuales de 25 dólares y la misma tarifa de 3 dólares por hora de uso de las canchas. Aunque está establecida la tarifa por hora de uso de las canchas, entrevistas con socios del club indican que las tarifas han bajado. Las ventajas y desventajas de este club son las siguientes:

Ventajas

1. El club está bien establecido y tiene una buena reputación por su tenis.
2. Está convenientemente ubicado en una arteria principal.
3. Hay gran disponibilidad de las canchas durante el día.

Desventajas

1. El área de estacionamiento es inadecuada.
2. Tiene fama de estar mal administrado.
3. El club posee una atmósfera agradable, aunque no particularmente atractiva para familias o individuos de altos ingresos.

Eagles Lodge. Este club está ubicado a tres cuerdas al este de la calle Main, sobre la calle Metro. Las instalaciones incluyen cinco canchas de raquetball, una piscina cubierta de 8 por 15 metros, salón, un sauna, un jacuzzi y gimnasios para hombres y mujeres. El club tiene 10 800 socios varones, de los

* Las cantidades anotadas en esta obra están dadas en dólares de EUA y son ilustrativos. (N. del T.)

cuales, entre 5 000 y 6 000 provienen del área norte de Peyton. Se estima que de estos socios entre 2 000 y 3 500 provienen del área de mercado de Northwood. La mayoría de la gente proviene de *las áreas del censo*, al este de la calle Main. Las cuotas de membresía son de 40 dólares al año y la tarifa para las canchas de raquetball es de 2 dólares por hora.

Ventajas

1. El club está bien establecido y posee la reputación de ser el más grande de la cadena a nivel nacional.
2. Tiene una ubicación conveniente sobre una arteria principal.
3. Proporciona instalaciones tanto sociales como de salud.
4. Es en extremo económico.
5. Hay gran disponibilidad de las canchas durante el día.

Desventajas

1. Debido a lo económico de sus cuotas tiende a tener miembros de grupos de ingresos menores.
2. La atmósfera del club es agradable, pero no especialmente atractiva para familias o individuos de ingresos medios y altos.
3. En ocasiones, las instalaciones están demasiado concurridas por la tarde.

Whitewater University Aquatic Center. Otro factor que podría afectar potencialmente esta situación es el centro acuático de la Universidad Whitewater. La primera etapa (piscina cubierta), estará concluida y abrirá al público en el otoño. Las etapas subsecuentes comprenden raquetball, gimnasio para damas, etc. Las instalaciones no serán lujosas pero sí mejores que las de Eagles Lodge y es probable que compitan con las de Northtown.

Resumen de entrevistas con socios de la competencia. Los entrevistadores consideraron que el mercado real está constituido por profesionales y empresarios. Indicaron que las instalaciones de Northtown estaban diseñadas en un nivel medio. Consideraron que el Peyton Club tenía mejores instalaciones. Las instalaciones de Eagles Lodge se consideraron de nivel bajo, orientadas a un segmento de menor ingreso de la población. Se opinó que si se ubicaban las instalaciones en la parte norte, probablemente sólo atenderían a gente de los alrededores del centro comercial y que esto constituiría una base de población inadecuada.

Datos demográficos

El área principal de mercado propuesta para el club atlético comprende las *zonas del censo* 103.01, - 103.02, 105.01, 105.02, 106, 107, 108, 109, 110, 111, 112.01 y 112.02 y la *zona* 8, que está dentro de los límites de la ciudad de Peyton. En la tabla 3.1 se presenta la población reportada en el Censo de 1990 para cada *zona del censo*. Aproximadamente el 30% de la población del área principal de mercado vive al norte de Northwood. Un total de 5 337 personas o el 11% de la población vive en el área de Northwood (*zona* 105.01). El 59% restante de la población vive al sur o al sureste del área de Northwood.

TABLA 3.1 POBLACIÓN EN EL ÁREA PRINCIPAL DE MERCADO

ZONA DEL CENSO	POBLACIÓN
103.01	2,568
103.02	6,742
105.01	5,337
105.02	4,830
106	4,104
107	804
108	3,815
109	1,141
110	4,129
111	4,848
112.01	2,465
112.02	2,966
Zona 8	5,139
Total	48,888

En la tabla 3.2, se presenta un resumen de los datos demográficos del área principal de mercado. Los datos muestran que el 64.5% de las personas empleadas son trabajadores de oficina. Una cuarta parte de las personas empleadas se clasificaron como directivos o profesionales. Casi el 52% de la población tenía una edad que oscilaba entre 18 y 54 años. Más de la mitad de los hogares tuvieron ingresos superiores a \$2 5000. La mediana del ingreso por hogar fue de \$22 689. Por último, la población en el área de mercado se incrementó en promedio 2% anual desde 1980, mientras que el número de hogares aumentó en promedio un 4% anual, aproximadamente.

TABLA 3.2 DATOS DEMOGRÁFICOS PARA EL ÁREA PRINCIPAL DEL MERCADO OBJETIVO

Población	Proyección 1998	54,999
	Estimación 1993	50,990
	Censo 1990	48,888
	Censo 1980	38,862
	Porcentaje de cambio, 1980-1990	25.8%
Población por edad	Porcentaje de cambio, 1990-1993	4.3%
	0-17	31.4%
	18-34	26.5%
	35-54	25.3%
	55-64	9.3%
	65+	7.5%
	Mediana de la edad en la población total	29.8
	Mediana de la edad en la población adulta	40.7
Ocupación	Directivo/prof.	5,447
	Tec./admvo.	4,581
	Ventas	3,428
	Total empleados oficina	13,456
	Prod./mfra/repación	2,091

TABLA 3.2 (continuación)

	Operadores maq.	729	
	Obreros/trans/etc.	1,771	
	Total obreros		4,591
	Agrop/Forestal/Pesca		279
	Servicios		2,529
	Total empleados		20,855
	Total desempleados		1,706
	Fuerza total de trabajo		22,561
Educación	Graduados de secundaria (solamente)		64.0%
	Graduados universidad		22.9%
	Mediana años cursados		12.93
Hogares	Proyección 1998		18,272
	Estimación 1993		16,510
	Censo 1990		15,517
	Censo 1980		10,456
	Porcentaje de cambio, 1980-1990		48.4%
	Porcentaje de cambio, 1990-1993		6.4%
Población	Población hogares		46,242
	Hogares con menores de 18		7,436
	Hogares c/personas mayores de 65		2,543
	Población familiar		41,790
	Población no familiar		4,452
	Agrupación por cuartos		1,040
	Tamaño promedio del hogar		2.98
	Tamaño promedio de la familia		3.34
	Familiares		12,512
	No familiares		3,005
Ingreso por hogar*	\$ 0-\$7,499		8.9%
	\$ 7,500-\$ 9,999		4.0%
	\$10,000-\$14,999		7.7%
	\$15,000-\$24,999		25.0%
	\$25,000-\$34,999		24.2%
	\$35,000-\$49,999		21.0%
	\$50,000-\$74,999		6.9%
	\$75,000+		2.3%
	Mediana del ingreso por hogar (proy. 1998)	\$33,207	
	Mediana del ingreso por hogar (est. 1993)	\$26,383	
	Mediana del ingreso por hogar (censo 1990)	\$22,245	
	Ingreso per cápita (Censo 1990)	\$ 8,402	
	Ingreso promedio por hogar (Censo 1990)	\$25,036	
	Casas		17,182
	Condominios		1,097
Unidades en domicilio	1 unidad		\$13,805
	2-5 unidades		727
	+5 unidades		1,796
	Casas móviles		854
	Mediana del valor casas		\$61,536
	Mediana de la renta mensual		\$ 232
	Valor promedio condominio		\$67,711

* estimación 1993.

La tabla 3.3 pretende resumir la información demográfica más relevante.

La tabla 3.4 se diseñó para comparar el área principal de mercado con el área que atiende el Westwood Athletic Club de Cadott, Wisconsin. Después de estudiar la membresía de Westwood, se concluyó que alrededor del 95% de sus socios provienen de Cadott, Howard, Howard Lake o Rockford. De ahí que se compararan los datos demográficos de estas áreas con el área principal de mercado objetivo.

El área de mercado de Northwood tiene una gran población, aproximadamente 41.4% más habitantes que en el área de Westwood. También, el área de Northwood tiene 45.7% más población entre 18 y 54 años de edad. La mediana del ingreso por hogar resultó también ser 39.6% más alta en el área de Northwood.

Un estudio de factibilidad realizado para el Westwood Athletic Club, usó la siguiente fórmula para determinar si el área de Cadott apoyaría las nuevas instalaciones:

$$\frac{\text{Membresía activa}}{(\% \text{ Pob.}(18-54 \text{ años} \times \text{Pob. del condado)) - \text{Estudiantes universitarios}}$$

Considerando que la población del condado no es aplicable en la decisión de Northwood, una fórmula más precisa sería:

$$\frac{\text{Membresía activa}}{\text{Pob. del área de mercado (18-54 años)} - \text{Estudiantes universitarios}}$$

Se debe tener en cuenta que los valores de membresía activa son estimaciones estrictamente basadas en la mejor información. La estimación para el área de mercado de Westwood se basó en 1953 membresías para el Westwood Athletic Club y una estimación de 1 575 utilizada en 1992 para sus competidores. Es importante tomar en cuenta que muchas personas pudieran pertenecer a ambos clubes o que pudieran haberse cambiado al Westwood Club. Por este motivo, la estimación para el área de mercado de Westwood se calculó como $1\,953 + 1\,575 - 500 = 3\,028$.

TABLA 3.3 RESUMEN DE DATOS DEMOGRÁFICOS

AÑO	POBLACIÓN TOTAL	TOTAL HOGARES	POBLACIÓN HOGARES	TAMAÑO PROMEDIO HOGAR	MEDIANA INGRESO POR HOGAR
1990	48,888	15,517	46,242	2.98	\$22,245
1993	50,990	16,510	48,209	2.92	\$26,383

TABLA 3.4 ÁREA PRINCIPAL DE MERCADO COMPARADA CON EL ÁREA DE MERCADO DE WESTWOOD

	NORTHWOOD	WESTWOOD
Población	48,888	28,649
Población (18-54)	25,324	13,751
Hogares	15,517	11,227
Familias	12,512	7,609
Estudiantes universitarios	1,200	1,600
Mediana del ingreso por hogar	\$22,245	\$13,705
Membresía en clubes modernos	3,250	3,028
Membresía en clubes como % de población no universitaria entre 18 y 54 años de edad	13.5%	24.9%

La estimación para Northwood se basa en 900 socios de Northtown, 2 500 de Eagles Lodge y 100 de Peyton Club, que residen en el área de mercado.

CONCLUSIONES

1. La población base entre 18 y 54 años de edad en el área de mercado de Northwood pareciera ser suficientemente grande para apoyar al club atlético.
2. El gran número de directivos, profesionales y empleados de oficina pareciera ser un buen mercado para un club atlético moderno de clase alta.
3. El ingreso promedio por hogar pareciera ser ser buen mercado para un club atlético moderno de clase alta.
4. El único competidor directo sería el Northtown Racquet Club, que actualmente tiene 1 000 socios.
5. La mayoría de las personas que compran en el centro comercial de Northwood proviene de los alrededores inmediatos. Se requeriría de una fuerte estrategia de mercadotecnia para ampliar el área principal de mercado a las fronteras que se incluyen en este estudio. La gente tendría que viajar de ocho a trece kilómetros para participar en un club atlético moderno; no obstante, habría que venderles la idea.

PREGUNTAS

1. ¿Qué fuentes secundarias de datos usó Mark Craze para realizar este estudio?
2. ¿Qué fuentes primarias de datos usó Mark Craze para realizar este estudio?
3. ¿Qué otras fuentes de datos pudo haber utilizado para realizar esta investigación?
4. ¿Está usted de acuerdo con los resultados?

VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Al final de cada capítulo, se presentará un ejemplo del uso de Forecast Plus, utilizando los datos de Consumer Credit Counseling. Estos datos comprenderán el número de nuevos clientes atendidos por CCC en el periodo de enero de 1985 a mayo de 1993. En esta sección se mostrará cómo utilizar los programas de administración de datos para establecer un archivo de datos.

Los programas de administración de datos se ejecutan seleccionando la opción uno (1) del menú principal de Forecast Plus.

Los programas le permiten crear, modificar, transformar e imprimir archivos de datos. Una vez reunidos los datos históricos, usted utilizará los programas de administración de datos para introducir éstos en la computadora y almacenarlos en disco. De modo que es probable que los programas de administración de datos sean los primeros que utilice.

Todos los programas de administración de datos se pueden ejecutar por separado tecleando el número de la función deseada en el menú de administración de datos:

```

                                Data file name:  (NONE)

                                *** Data Management Menu ***

1.  Enter new data
2.  Edit an existing data file
3.  Edit an existing labels file
4.  Print a data file
5.  Transform a data file
6.  Trading day adjustment
7.  Restructure/merge data files
8.  Read/write DIF format files
9.  Change parameter table
10. Batch editor
11. Return to main menu

Which?
```

Programa 1: Captura de datos nuevos

Después de elegir Enter new data (Introducir datos nuevos), el programa le preguntará a qué archivo de datos desea agregarlos. Para este ejemplo, llamaremos a nuestro archivo CREDIT. Teclee CREDIT. Si oprime <Esc>, la operación se anulará y el sistema regresará al menú de administración de datos. Si escribe la palabra clave FILES. se mostrará un directorio del disco.

Si escribe el nombre de un archivo existente, el programa supondrá que desea agregar más datos al final de ese archivo. Si teclea el nombre de un archivo nuevo, el sistema le indicará que el archivo no existe mediante el mensaje:

```
The data file B:CREDIT does not exist.
```

y le preguntará si desea crearlo mediante el mensaje:

```
Do you want to create it (y/n)?
```

Oprima Y y <Enter>

Para cada archivo de datos, hay en realidad dos archivos. Uno contiene los datos reales mientras que el otro contiene información de etiquetado respecto a los datos. Por ejemplo, si crea un archivo llamado CREDIT, estará creando también un archivo de etiquetas denominado CREDIT.LBL. Los archivos de etiquetas siempre tienen la extensión .LBL.

El nombre del archivo de datos puede contener opcionalmente la designación de la unidad. El nombre sigue las mismas convenciones que en el sistema operativo IBM DOS.

Un nombre de archivo puede ser de hasta ocho caracteres; no debe haber espacios intermedios. Si no se especifica la letra de la unidad, el programa utilizará la letra especificada en la tabla de parámetros. La letra de la unidad se puede especificar agregando un prefijo al nombre del archivo (v.gr. B:). Un nombre de archivo puede también tener una extensión de uno a tres caracteres, B:CREDIT.

Si está creando un nuevo archivo de datos, será necesario introducir información de etiquetas, la cual se almacenará en el archivo .LBL. Su uso principal consiste en controlar las etiquetas en las impresiones. Ya que estamos creando un nuevo archivo de datos llamado CREDIT y nuestro disco de datos está en la unidad B:, el archivo de etiquetas se denominará B:CREDIT.LBL y la pantalla mostrará la siguiente información:

```
*** Labels File Creation ***

Creating labels file B:CREDIT.LBL

Number of variables:  1

Type of labeling:      YEAR

Starting label:
Starting observation:  1

Number of observations between each label:
Increment of the labels:  1
```

Ésta es la información que se almacenará en el archivo .LBL. Nótese que dos de los parámetros, etiqueta inicial (*Starting label*) y número de observaciones entre cada etiqueta (*Number of observations between each label*), no presentan valor por omisión. La creación de un archivo de etiquetas consiste sólo en introducir la información para cada una de las preguntas. Se pueden emplear las teclas de flecha hacia arriba y hacia abajo para mover el cursor. Si se presiona se anulará el procedimiento de registro de datos.

Número de variables (Number of variables). Este número corresponde al número de variables que habrá en el archivo de datos. El valor por omisión es una variable. Para nuestro ejemplo, tenemos una variable: número de clientes atendidos.

Tipo de etiquetas (Labeling Type). El valor por omisión de tipo de etiquetas es YEAR. Los tipos que maneja Forecast Plus son YEAR (año), MONTH (mes), DAY (día) y NUMBERED (numerada). Estos tipos comprenderán todos los requerimientos posibles de etiquetado, que será el que se use en la impresión. Por lo regular hay diversas opciones apropiadas de etiquetado para cualquier conjunto de datos dado. Por ejemplo, para los datos mensuales de nuestro ejemplo, podemos escoger tres etiquetas posibles. Ya que elegiremos YEAR, también queremos usar 12 observaciones entre cada etiqueta (cada observación representará un mes). Si elegimos MONTH, podríamos especificar una observación entre

cada etiqueta. DAY no sería una etiqueta adecuada, ya que no tenemos información diaria. NUMBERED pudiera ser una etiqueta apropiada; se podría poner el año en el número inicial y 12 en el número de observaciones entre variables.

Etiqueta inicial (Starting label). Ésta sería la etiqueta real para el primer dato reunido. Ya que seleccionamos YEAR como el tipo de etiqueta, la etiqueta inicial podría ser el año en el que se recolectaron nuestros primeros datos (*v.gr.*, 1985). Si el tipo fuera MONTH, la etiqueta inicial podría ser del 1 al 12 (*v. gr.*, Ene. = 1, Feb. = 2, Mar. = 3, etc.). Si el tipo fuera DAY, la etiqueta inicial podría ser del 1 al 7 (*v. gr.*, Dom = 1, Lun. = 2, Mar. = 3, etc.). Si el tipo fuera NUMBERED, simplemente se registra el número de la etiqueta inicial. Nótese que la etiqueta inicial para los tipos MONTH y DAY contendrá datos codificados, es decir, que no se puede registrar ENERO; en vez de ello se debe registrar el número 1 para la etiqueta inicial.

Observación inicial (Starting observation). La observación inicial se refiere a la primera pieza de información reunida. En nuestro ejemplo, el primer punto de datos es para el mes de enero, así que usamos un 1. Si la información más antigua que pudimos encontrar fue de marzo de 1985, la etiqueta inicial sería 1985 y la observación inicial sería 3.

Como no tenemos información para enero y febrero, nuestro conjunto de datos comienza con marzo.

Número de observaciones entre cada etiqueta (Number of observations between each label). No existe valor por omisión para este número, por lo que se tendrá que asignar un valor en esta pregunta. En nuestro ejemplo, hay doce observaciones entre cada etiqueta, esto es, 12 meses en cada año. Un valor de 1 significa que cada etiqueta representa un valor de datos y que no hay otros valores de datos entre cada etiqueta. Por ejemplo, si sus datos tienen la etiqueta YEAR y especifica 1 como el número de observaciones entre cada etiqueta, significa que se recolectó un elemento de información cada año.

Incremento de las etiquetas (Increment of the labels). El incremento de las etiquetas siempre tendrá un valor por omisión de 1, es decir, que cada vez que cambie una etiqueta en la impresión, se incrementará en uno.

La creación del archivo de etiquetas queda ahora de la siguiente manera:

```
Number of variables:  1

Type of labeling:     YEAR

Starting label:       1985
Starting observation: 1

Number of observations between each label: 12
Increment of the labels: 1
                        Press ENTER to accept:
```

Si presiona <Enter>, significa que está satisfecho con la información de etiquetas y ésta se almacenará en el archivo .LBL. Si presiona la tecla de flecha hacia arriba, podrá modificar de nuevo la información.

Etiquetas de variables (Variable labels). Una vez concluida la información de etiquetas, se le solicitará que introduzca la etiqueta (LABEL), para cada una de las variables en el archivo. Una etiqueta de variable puede tener hasta 20 caracteres de longitud y puede incluir cualquier carácter del teclado, *excepto la coma*, incluyendo espacios, números y caracteres especiales. Sin embargo, el primer carácter de la etiqueta de variable debe ser una letra. Se trata simplemente de una descripción de la variable que se incluirá en todas las salidas. Para nuestros ejemplos la etiqueta de la variable es:

```
NUMBER OF CLIENTS
```

Después de terminar de introducir la(s) etiqueta(s) de la(s) variable(s), estará listo para introducir los datos. Dependiendo de si se está trabajando con archivos de una o varias variables, se seleccionará automáticamente el programa apropiado de registro de datos. Para nuestro ejemplo el primer registro es:

```
Record    1:    1985  (1)  >    <
```

Una vez terminada la captura de datos para la variable cliente, la pantalla mostrará lo siguiente:

```
Record   99:    1993  (3)    199
Record  100:    1993  (4)  >    <
```

Deberá presionar <Esc>, y el programa le preguntará si desea: 1. Almacenar los datos en disco, 2. regresar a la modalidad de captura/edición, o 3. Salir sin almacenar los cambios. Y finalmente la elección, ¿cuál?, ésta sería 1:

```
Do you want to:
    1. Save the data on disk
    2. Re-enter the entry/edit mode
    3. Exit without saving the changes

Which?  1
```

El disco de datos contiene ahora dos archivos: CREDIT y CREDIT.LBL.

Pronóstico mediante hojas de cálculo

Esta sección le mostrará cómo introducir etiquetas (encabezados y títulos), así como datos en una hoja de cálculo de Lotus. Las etiquetas y datos a capturar corresponden al ejemplo 4.1.

Captura de etiquetas

1. Coloque el apuntador de celda en la celda A1 y escriba VERNON'S MUSIC STORE. Presione <Enter>.

Nótese que lo que escribió aparece en la línea de captura del panel de control y no en la celda. Si comienza con un carácter alfabético, el indicador de modalidad mostrará LABEL (etiqueta). Si comienza con un número, el indicador de modalidad mostrará VALUE (valor). Lotus tiene límites de celda flexibles que le permiten ingresar etiquetas más largas que el ancho de la celda.

2. Coloque el apuntador de celda en A2 y escriba NUMBER OF VCR'S SOLD (número de VCR [videocaseteras vendidas]). Presione <Enter>.
3. Coloque el apuntador de celda en A4 y escriba MONTH (mes). Presione <Enter>.
4. Coloque el apuntador de celda en A5 y escriba JANUARY (enero). Presione <↓>. Advierta que la tecla de flecha hacia abajo mueve el apuntador de celda en esa dirección a la siguiente celda, A6. Registre el resto de los meses terminando con DECEMBER (diciembre) en la celda A16.
5. Posicione el apuntador de celda en la celda B4 y oprima una vez la barra espaciadora y escriba después NUMBER (número). El espacio le permite centrar la etiqueta dentro de la celda.
6. Coloque el apuntador de celda en la celda B6 y escriba 123. Presione <↓>. Continúe capturando los datos, hasta escribir 160 en la columna B16. La hoja de cálculo en este momento se ve como se muestra en la fig. 3.1.
7. A continuación necesitamos establecer el directorio actual con el fin de poder grabar nuestro archivo. Seleccione /FILE DIRECTORY o /FD. Usted seleccionó FILE del menú principal y DIRECTORY del submenú. La pantalla le solicitará el directorio actual mediante:

Enter current directory: A:/

Si el directorio actual es diferente del directorio por omisión A:/, indique entonces el directorio en el que desea almacenar el archivo.

8. Para grabar su archivo teclee /FILE/SAVE o /FS, que equivale a seleccionar FILE del menú principal y SAVE del submenú. El programa le solicitará, con el siguiente mensaje, el nombre del archivo a grabar:

Enter name of file to save: A:/*.WK1

A1:										MENU:
Worksheet	Range	Copy	Move	File	Print	Graph	Data	System	Add-In	Quit
Global	Insert	Delete	Column	Erase	Titles	Window	Status	Page	Learn	
	A	B	C	D	E	F	G	H		◀
1	VERNON'S MUSIC STORE									▶
2	NUMBER OF VCERS SOLD									▶
3										▶
4	Month	Number								?
5	January	123								
6	February	130								
7	March	125								
8	April	138								
9	May	145								
10	June	142								
11	July	141								
12	August	146								
13	September	147								
14	October	157								
15	November	150								
16	December	160								
17										
18										
19										
20										
13-Aug-91 03:02 PM										

Figure 3.1

- Escriba VCR1 y presione <Enter> para almacenar su archivo como A:/VCR1.WK1.
- Para imprimir una hoja de cálculo, seleccione /PRINT RANGE o /PR. Presione <Enter> para que aparezca el apuntador de rango. El rango siempre especifica un rectángulo a imprimir, mediante las coordenadas de las celdas de las esquinas superior izquierda e inferior derecha. Cuando se le pida que introduzca un rango de celdas, escriba directamente en la celda las coordenadas A1..B16. Presione ahora <Align> (alinear) para restablecer el contador de líneas. Para enviar la salida a impresión oprima <Go>. Las opciones son PRINTER o FILE. Si desea imprimir la hoja de cálculo, elija PRINTER. Después de impreso un archivo, oprima Q de QUIT para salir del menú de impresión.
 - Para terminar la sesión, escriba /QUIT o /Q.

PAQUETE DE CÁMPUTO MINITAB

Se empleará Minitab para demostrar qué fácil es generar números aleatorios mediante un paquete estadístico de software.

El problema. La tabla P2.7 para el problema 7 en el cap. 2, presenta una población de 200 tamaños de familias y en este problema se le solicita seleccionar una muestra de 30 tamaños y probar la hipótesis de que el tamaño promedio de la familia no ha cambiado de 2.9 personas en los últimos diez años.

Solución con Minitab. Los comandos de Minitab para generar números aleatorios son

```

MTB > RANDOM 30 C1;
SUBC > INTEGER 1 200.
MTB > PRINT C1
C1
    55    75    70    53    33     9   141    40   146   136
   132   134   160   143   173   163   148   102   179   135
    22    91   146   128   148   182    62    81    60   197
MTB > STOP

```

- El comando RANDOM se usa para generar los 30 números aleatorios y colocarlos en C1. Nótese que se coloca un punto y coma (;) al final de la línea para obtener un indicador de subcomando.
- El subcomando INTEGER indica que los números aleatorios deberán ser enteros entre 1 y 200. Nótese que se incluye un punto (.) al final de la línea para indicar el fin de los subcomandos.
- El comando PRINT se utiliza para exhibir el contenido de C1.

BIBLIOGRAFÍA

- ALRECK, P. L. y SETTLE, R. B. *The Survey Research Handbook*. Homewood, Ill.: Richard D. Irwin, Inc., 1985.
- DANIELLS, L. M. *Business Forecasting for the 1980s-and Beyond*. Boston: Baker Library. Harvard Business School, 1980.
- DAVIDSON, T. A. y AYERS, J. L. "Selecting and Using External Data Sources." En *Handbook of Forecasting*, S. Makridakis y S. C. Wheelwright, eds. Nueva York: John Wiley & Sons, 1982.
- KINNEAR, T. C. y TAYLOR, J. R. *Marketing Research: An Applied Approach*. Nueva York: McGraw-Hill, 1983.
- VICHAS, R. P. *Complete Handbook of Profitable Marketing Research Techniques*. Englewood Cliffs, N. J.: Prentice Hall, 1982.
- ZIKMUND, W. G. *Exploring Marketing Research*. Nueva York: The Dryden Press, 1986.
- ZIKMUND, W. G. *Business Research Methods*, 3ra. ed. Nueva York: The Dryden Press, 1991.

Exploración de los patrones de datos y selección de la técnica de pronóstico

ACTIVIDAD

La generación de un pronóstico preciso y útil implica dos consideraciones básicas. La primera consiste en reunir datos que sean aplicables para la tarea de pronóstico y que contengan información que pueda producir pronósticos precisos. Esta importante tarea se analiza en el cap. 3. El segundo factor clave es seleccionar una técnica de pronóstico que utilice al máximo la información contenida en los datos y los patrones que éstos presenten.

El primer paso en la formulación de pronósticos, expuesto en el cap. 1 fue la recolección de datos. Después de esto, el pronosticador debe revisar los intentos anteriores para pronosticar la variable de interés. Una búsqueda de bibliografía y los comentarios con colegas de la compañía pueden ayudar a evaluar los éxitos y/o fracasos anteriores de enfoques alternativos. Después de realizar la investigación inicial, el pronosticador se encuentra en una mejor posición para la reducción de datos (paso 2).

Con buenos datos a la mano, un pronosticador puede empezar la importante tarea de explorar los patrones de datos. Este paso comprende la observación de los datos, la comprensión de lo que los datos sugieren y el uso de varios métodos gráficos para obtener una mejor visión en el proceso que generó los datos. Con frecuencia es útil la elaboración de una gráfica de la serie de tiempo de datos, si éstos se recopilaban a lo largo del tiempo. La fig. 4.2 muestra un ejemplo de este tipo de gráfica. El examen de los datos conduce al paso 3 del proceso de pronóstico (selección de una técnica de pronóstico).

Se pueden emplear dos métodos básicos de pronóstico. Las técnicas de pronóstico *cualitativas* se basan en el juicio humano y en la intuición, más que en la manipulación de datos históricos anteriores. Las técnicas cualitativas comunes incluyen al método Delphi, curvas de crecimiento, escritura de escenarios, investigación de mercado y grupos de enfoque. Con frecuencia, estas técnicas son importantes en el esquema general de pronóstico y se abordan con cierto detalle en el cap. 11.

Las técnicas de pronóstico *cuantitativas* se utilizan cuando existen suficientes datos históricos disponibles y cuando se juzga que estos datos son representativos de un futuro desconocido. Esta apreciación es un paso importante en el proceso de pronóstico, ya que

todas las técnicas cuantitativas se apoyan en la suposición de que el pasado puede extenderse hacia el futuro de manera significativa para proporcionar pronósticos precisos. Las técnicas cuantitativas se clasifican frecuentemente en dos categorías: estadísticas y determinísticas.

Las técnicas estadísticas se enfocan completamente en patrones, cambios en los patrones y perturbaciones causadas por influencias aleatorias. Este libro expone muchas de estas técnicas como los promedios móviles y la atenuación exponencial (cap. 5), descomposición de series de tiempo y proyecciones de tendencia (cap. 8) y la metodología Box–Jenkins (cap. 10).

Las técnicas estadísticas de pronóstico emplean básicamente dos enfoques. Uno está basado en la suposición de que los datos se pueden descomponer en componentes como tendencia, ciclo, estacionalidad e irregularidad. Se realiza una predicción mediante la combinación de las proyecciones de cada uno de estos componentes individuales.

El segundo enfoque se asocia con las metodologías de modelos econométricos de series de tiempo y de Box–Jenkins. Los fundamentos teóricos se basan principalmente en conceptos estadísticos y no se supone que los datos estén representados por componentes separados.

Las *técnicas determinísticas (causales)* comprenden la identificación y determinación de relaciones entre la variable por pronosticar y otras variables de influencia. En estas técnicas se incluyen los modelos de regresión y regresión múltiple, (caps. 6 y 7), indicadores básicos (cap. 8), modelos econométricos (cap. 9) y encuestas anticipadas y modelos de entrada–salida (que no se abordan en este libro).

Las dos secciones siguientes se refieren brevemente a los elementos clave de los patrones de datos que son importantes para la comprensión de las técnicas de pronóstico que se presentan en capítulos posteriores. Se presenta en primer término una exposición de los componentes con frecuencia reconocidos explícitamente en las series de tiempo, seguido de una introducción de los elementos básicos del análisis de autocorrelación que son la base de muchos procedimientos sofisticados de pronóstico. La siguiente sección se dedica a una cuestión más importante: una vez identificados los patrones generales en los datos históricos, ¿qué técnica de pronóstico se debe elegir? El capítulo concluye con secciones sobre la medición del error en los pronósticos y la determinación de la suficiencia de una técnica de pronóstico.

COMPONENTES DE SERIES DE TIEMPO

Como se expuso en el cap. 3, con frecuencia se realizan observaciones de datos a través del tiempo. Cualquier variable que conste de datos reunidos, registrados u observados sobre incrementos sucesivos de tiempo se denomina *serie de tiempo*.

Una *serie de tiempo* consta de datos que se reúnen, registran u observan sobre incrementos sucesivos de tiempo.

En el análisis de series de tiempo de datos, una tentación inmediata consiste en intentar explicar o contabilizar el comportamiento de las series. Para evitar esfuerzos inútiles, lo que se necesita es un enfoque sistemático para analizar las series. La descomposición clásica es un método que se basa en la suposición de que se pueden descomponer en componentes como tendencia, ciclo, estacionalidad e irregularidad. Una predicción se hace mediante la combinación de las proyecciones de cada componente individual. La técnica de descomposición clásica se aborda en el cap. 8. Otro enfoque, denominado *análisis de autocorrelación*, también se puede emplear para identificar componentes y patrones de datos. Esta técnica se expone en secciones posteriores de este capítulo.

Muchas variables macroeconómicas, como el Producto Nacional Bruto (PNB), el empleo y la producción industrial están dominadas por una fuerte tendencia. La *tendencia* de una serie de tiempo es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio. Las fuerzas básicas que ayudan a explicar la tendencia de una serie son el crecimiento de la población, la inflación de precios, el cambio tecnológico y los incrementos en productividad.

La *tendencia* de una serie de tiempo es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio.

El *componente cíclico* es la fluctuación en forma de onda alrededor de la tendencia, afectada por lo regular por las condiciones económicas generales. Los patrones cíclicos tienden a repetirse en los datos aproximadamente cada dos, tres o más años. Es común que las fluctuaciones cíclicas estén influidas por cambios de expansión y contracción económica, a los que comúnmente se hace referencia como el *ciclo de los negocios*.

El *componente cíclico* es la fluctuación en forma de onda alrededor de la tendencia.

El *componente estacional* se refiere a un patrón de cambio que se repite a sí mismo año tras año. En el caso de las series mensuales, el componente estacional mide la variabilidad de las series cada enero, febrero, etc. En las series trimestrales hay cuatro elementos estacionales, uno para cada trimestre. La variación estacional puede reflejar condiciones de clima, días festivos o la longitud de los meses calendario.

El *componente estacional* es un patrón de cambio que se repite a sí mismo año tras año.

El *componente aleatorio* mide la variabilidad de las series de tiempo después de que se retiran los otros componentes. Contabiliza la variabilidad aleatoria en una serie de tiempo

ocasionada por factores imprevistos y no recurrentes. La mayoría de los componentes irregulares se conforman de variabilidad aleatoria. Sin embargo, ciertos sucesos a veces impredecibles como huelgas, cambios de clima (sequías, inundaciones o terremotos), elecciones, conflictos armados o la aprobación de asuntos legislativos, pueden causar irregularidades en una variable.

El *componente aleatorio* mide la variabilidad de las series de tiempo después de retirar los otros componentes.

EXPLORACIÓN DE PATRONES DE DATOS MEDIANTE ANÁLISIS DE AUTOCORRELACIÓN

Cuando se mide una variable a través del tiempo, con frecuencia está correlacionada consigo misma cuando se desfasa uno o más periodos. Esta correlación se mide mediante el coeficiente de *autocorrelación*.

Autocorrelación es la correlación existente entre una variable desfasada uno o más periodos y la misma variable.

Los patrones de datos que incluyen componentes como tendencia, estacionalidad e irregularidad se pueden estudiar usando el enfoque del análisis de autocorrelación. Los coeficientes de autocorrelación para diferentes desfases de tiempo de una variable se emplean para identificar patrones en las series de tiempo de datos.

El concepto de autocorrelación se ilustra mediante los datos que se presentan en la tabla 4.1. Nótese que las variables Y_{t-1} y Y_{t-2} son en realidad valores de Y que se han desfasado uno y dos periodos, respectivamente. Los valores para el mes de marzo que se muestran en el renglón del periodo 3, son las ventas de marzo $y_t = 125$; las ventas de febrero, $y_{t-1} = 130$; y las ventas de enero, $y_{t-2} = 123$.

Comúnmente se usa la ecuación 4.1 para calcular el coeficiente de autocorrelación de primer orden (r_1) o la correlación entre Y_t y Y_{t-1} .

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (4.1)$$

en donde

r_1 = coeficiente de autocorrelación de primer orden

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

Y_t = observación en el periodo t

Y_{t-1} = observación en un periodo anterior o en el periodo $t - 1$.

La ecuación 4.2 es la fórmula para calcular el coeficiente de autocorrelación de orden k (r_k), entre observaciones separadas por k periodos: Y_t y Y_{t-k} .

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (4.2)$$

en donde

r_k = coeficiente de autocorrelación para un desfase de k periodos

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

Y_t = observación en el periodo de tiempo t

Y_{t-k} = observación en k periodos anteriores o en el periodo $t - k$.

Ejemplo 4.1

Harry Vernon reunió datos sobre el número de videocaseteras (VCR's) que vendió el año pasado para Vernon's Music Store. Los datos se presentan en la tabla 4.1. La tabla 4.2 muestra los cálculos que conducen al uso de la ecuación 4.1.

TABLA 4.1 DATOS AUTOCORRELACIONADOS

PERIODO, t	MES	DATOS ORIGINALES, Y_t	Y DESFASADA UN PERIODO, Y_{t-1}	Y DESFASADA DOS PERIODOS, Y_{t-2}
1	Enero	123		
2	Febrero	130		
3	Marzo	125		
4	Abril	138		
5	Mayo	145		
6	Junio	142		
7	Julio	141		
8	Agosto	146		
9	Septiembre	147		
10	Octubre	157		
11	Noviembre	150		
12	Diciembre	160		

TABLA 4.2 CÁLCULO DEL COEFICIENTE DE AUTOCORRELACIÓN DE PRIMER ORDEN PARA LOS DATOS DE LA TABLA 4.1

PERIODO 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
Sumas	1,704		0	-18	1,474	843

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

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

El coeficiente de autocorrelación de primer orden (r_1), o la correlación entre Y_t y Y_{t-1} , se calcula utilizando, en la ecuación 4.1, las sumas de la tabla 4.2.

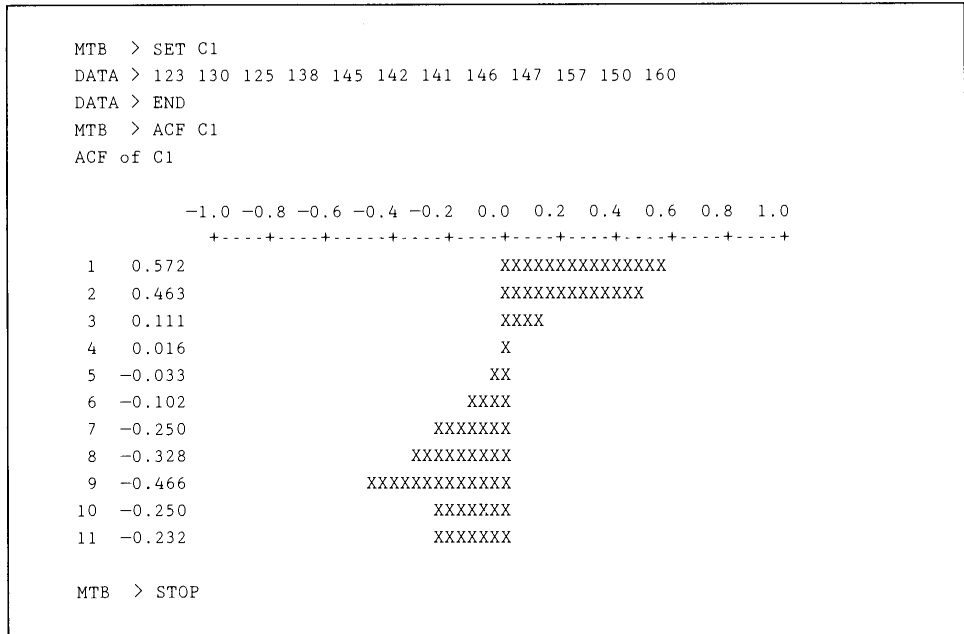
$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

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

Parece que existe alguna autocorrelación en esta serie de tiempo desfasada un periodo. La correlación entre Y_t y Y_{t-1} , o la autocorrelación para un desfase de 1 es .572. Esto significa que las ventas sucesivas de videocaseteras están de algún modo correlacionadas una con otra. Esta información puede proporcionar a Harry una valiosa visión interior de su serie de tiempo, puede ayudarlo a prepararse a utilizar un método de predicción avanzado y puede prevenirle del uso del análisis de regresión con estos datos. Esta última situación se estudiará en el capítulo 9.

El coeficiente de autocorrelación de segundo orden (r_2), o el coeficiente de correlación entre Y_t y Y_{t-2} para los datos de Harry, se calcula con la ecuación 4.2

$$r_2 = \frac{\sum_{t=1}^{n-2} (Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$



- El comando SET se usa para almacenar los datos en C1.
- El comando END se emplea para indicar el fin de la captura de datos.
- El comando ACF se utiliza para calcular coeficientes de autocorrelación para los datos en C1.

Nótese que el correlograma de Minitab es diferente al de la fig. 4.1. La escala de autocorrelación se muestra en la parte superior de la gráfica en lugar de en la parte inferior. El número de desfases y sus coeficientes de autocorrelación, ambos aparecen a la izquierda. Con esta gráfica, se pueden estudiar los patrones de datos, incluyendo tendencia, estacionariedad e irregularidad, mediante el enfoque del análisis de autocorrelación. Los coeficientes de autocorrelación para diferentes periodos desfasados de una variable se pueden utilizar en una serie de tiempo de datos para identificar lo siguiente:

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

Si una serie es aleatoria, la correlación entre Y_t y Y_{t-1} es cercana a cero y los valores sucesivos de la serie de tiempo no guardan relación entre sí.

Si una serie tiene una tendencia, Y_t y Y_{t-1} están altamente correlacionados y es típico que los coeficientes de autocorrelación sean diferentes de cero de manera significativa para varios de los primeros periodos de desfasamiento y caigan gradualmente hacia cero al

incrementarse el número de periodos. El coeficiente de autocorrelación para el periodo de desfase 1 es por lo regular muy grande (cercano a 1). El coeficiente de autocorrelación para el periodo de desfase 2 será también grande. Sin embargo, no lo será tanto como para el periodo de desfase 1, ya que se emplea un término menos para calcular su numerador.

Si una serie tiene un patrón estacional, se presentará un coeficiente de autocorrelación significativo en el periodo de desfase correspondiente: cuatro en los datos trimestrales o doce en los datos mensuales.

La tabla 4.3 muestra una serie de tiempo de 30 números aleatorios de tres dígitos, seleccionados de una tabla de números aleatorios. La fig. 4.2 muestra una gráfica de la serie de tiempo de datos. En teoría, las autocorrelaciones calculadas para estos datos deberían ser iguales a cero. Desde luego, los 30 valores de la tabla 4.3 sólo son un conjunto de un gran número de posibles muestras de tamaño 30. Cada muestra producirá autocorrelaciones diferentes. La mayoría de estas muestras producirá coeficientes de autocorrelación cercanos a cero. No obstante, es posible que sólo por azar, una muestra produzca un coeficiente de autocorrelación significativamente diferente de cero.

¿Qué hace un analista para determinar si un coeficiente de autocorrelación es significativamente diferente de cero, para los datos de la tabla 4.3? Teóricamente, se podría desarrollar una distribución muestral de coeficientes de autocorrelación, tomando un número infinito de muestras de 30 números aleatorios.

Quenouille¹ y otros han demostrado que los coeficientes de autocorrelación de datos aleatorios tienen una distribución que se puede aproximar a una curva normal con una media de cero y una desviación estándar de $1/\sqrt{n}$. Sabiendo lo anterior, el analista puede comparar los coeficientes de autocorrelación de la muestra con esta distribución teórica de muestreo y determinar si los datos provienen de una población cuya media sea cero en periodos de desfase en particular.

TABLA 4.3 SERIE DE TIEMPO CON 30 NÚMEROS ALEATORIOS SELECCIONADOS

PERIODO		PERIODO		PERIODO	
t	Y_t	t	Y_t	t	Y_t
1	646	11	707	21	173
2	477	12	709	22	145
3	560	13	39	23	674
4	688	14	164	24	533
5	892	15	30	25	67
6	386	16	708	26	296
7	747	17	379	27	838
8	533	18	458	28	242
9	127	19	590	29	717
10	54	20	766	30	196

¹ M.H. Quenouille, "The Joint Distribution of Serial Correlation Coefficients", *Annals of Mathematical Statistics*, Vol. 20 (1949), pp 561-571.

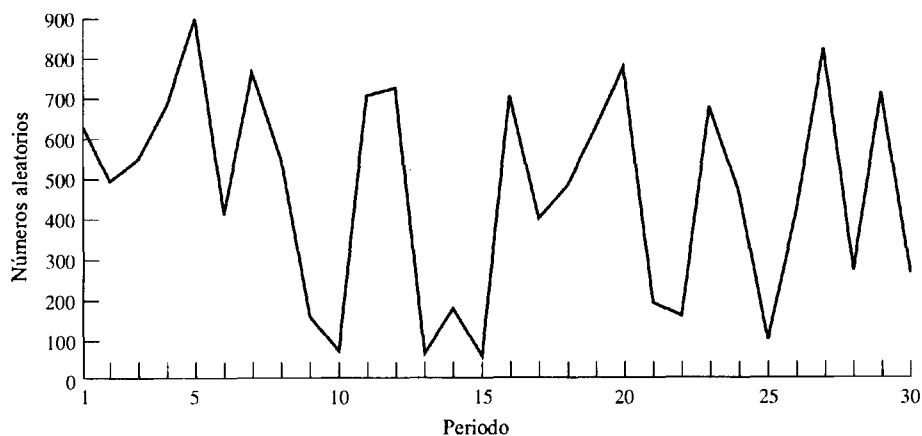


Figura 4.2 Gráfica de una serie aleatoria.

En realidad, los coeficientes de autocorrelación se pueden probar en forma simultánea para todos los periodos de desfase. Si la serie es en efecto aleatoria, la mayoría de los coeficientes de autocorrelación debe ubicarse dentro de un nivel especificado por 0, más o menos un cierto número de errores estándar. A un nivel específico de confianza, se puede considerar aleatoria una muestra si los coeficientes de autocorrelación calculados se encuentran todos dentro del intervalo producido por la ecuación 4.3.

$$0 \pm Z \frac{1}{\sqrt{n}} \quad (4.3)$$

en donde

Z = valor normal estándar para un nivel de confianza dado

n = número de observaciones en la serie de datos

Este procedimiento se ilustra en el ejemplo 4.2.

Ejemplo 4.2

Se desarrolla una prueba de hipótesis para determinar si la serie presentada en la tabla 4.3 es aleatoria. Las hipótesis nula y alterna para probar si los coeficientes de autocorrelación para un periodo de desfase en particular son diferentes de cero de manera significativa son

$$H_0: \rho_k = 0$$

$$H_1: \rho_k \neq 0$$

Ya que para esta serie aleatoria $n = 30$, el error estándar (desviación estándar de la distribución muestral de los coeficientes de autocorrelación), es $1/\sqrt{30} = .18$. Si se prueba la hipótesis nula

al nivel de significancia de .05, el valor correcto normal estándar z es 1.96 y el valor crítico es $1.96(.18) = .353$. La regla de decisión es:

Si un coeficiente de autocorrelación es menor que $-.353$ o mayor que $.353$, rechazar la hipótesis nula, en caso contrario, no rechazarla (rechazar H_0 si $r_k < -.353$ o $r_k > .353$).

Los coeficientes de autocorrelación de los datos de la tabla 4.3 aparecen graficados en el correlograma de la fig. 4.3. Las dos líneas de puntos paralelas al eje vertical son los límites de confianza de 95% ($-.353$ y $.353$). Se verifican 15 periodos de desfase y todos los coeficientes de autocorrelación se ubican dentro de estos límites. El analista concluye que esta serie es aleatoria, es decir, que se generó mediante un proceso aleatorio.

También se puede emplear el paquete de cómputo Minitab para resolver el ejemplo 4.2. Los comandos de Minitab son:

```
MTB > RANDOM 30 C1;
SUBC> INTEGER 1 999.
MTB > PRINT C1

C1
  646  477  560  688  892  386  747  533  127  54  707
  709   39  164   30  708  379  458  590  766  173  145
  674  533   67  296  838  242  717  196

MTB > ACF C1
ACF of C1

      -1.0  -0.8  -0.6  -0.4  -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +-----+-----+-----+-----+-----+-----+
1 -0.027                                     XX
2 -0.217                                XXXXXX
3 -0.018                                    X
4 -0.284                                XXXXXXXX
5 -0.116                                XXXX
6 -0.156                                XXXXX
7  0.083                                 XXX
8  0.016                                 X
9 -0.070                                 XXX
10 -0.311                               XXXXXXXX
11  0.117                               XXXX
12  0.101                               XXXX
13  0.032                               XX
14 -0.312                               XXXXXXXX
15  0.219                               XXXXXX

MTB > STOP
```

La serie se puede considerar aleatoria al nivel de confianza del 95%, ya que todos los coeficientes de autocorrelación calculados se encuentran entre el intervalo producido por la ecuación 4.3 ($-.353 > r_k < .353$).

Los comandos de Minitab para diferenciar una serie de datos de ejemplo son:

```
MTB > SET C1
DATA> 221 224 230 237 235 240 247 248 250 255 262 271
DATA> END
MTB > DIFFERENCES 1 FOR C1, STORE IN C2
MTB > PRINT C1-C2
```

ROW	C1	C2
1	221	*
2	224	3
3	230	6
4	237	7
5	235	-2
6	240	5
7	247	7
8	248	1
9	250	2
10	255	5
11	262	7
12	271	9

```
MTB > STOP
```

- El comando SET se usa para almacenar los datos en C1.
- El comando DIFFERENCES calcula la primera diferencia de los datos en C1 y almacena estas diferencias en C2.
- El comando PRINT se emplea para mostrar lo que está almacenado en las columnas C1 y C2.

Ejemplo 4.3

A Maggie Trymane, una analista de Sears, se le asigna la tarea de pronosticar las ventas de 1992. Ella reúne los datos de 1955 a 1991, que se muestran en la tabla 4.4. Estos datos se grafican en la serie de tiempo de la fig. 4.4. Primero, Maggie calcula un intervalo de confianza del 95% para los coeficientes de autocorrelación:

$$0 \pm 1.96 \frac{1}{\sqrt{37}}$$

$$0 \pm .322$$

A continuación, Maggie ejecuta los datos en Minitab (véanse las instrucciones en la sección “Paquete de Cómputo Minitab” al final de este capítulo), y produce un correlograma similar al que se muestra en la fig. 4.5. Al examinarlo, nota que las autocorrelaciones para los primeros siete periodos de desfaseamiento son significativamente diferentes de cero (.92, .83, .73, .64, .55, .46 y .37) y que esos valores gradualmente caen a cero. Maggie decide que los datos tienen una tendencia.

TABLA 4.4 VENTAS ANUALES
DE SEARS ROEBUCK & CO.,
1955-1991

AÑO	VENTAS DE SEARS
	Y_t
1955	3,307
1956	3,556
1957	3,601
1958	3,721
1959	4,036
1960	4,134
1961	4,268
1962	4,578
1963	5,093
1964	5,716
1965	6,357
1966	6,769
1967	7,296
1968	8,178
1969	8,844
1970	9,251
1971	10,006
1972	10,991
1973	12,306
1974	13,101
1975	13,639
1976	14,950
1977	17,224
1978	17,946
1979	17,514
1980	25,195
1981	27,357
1982	30,020
1983	35,883
1984	38,828
1985	40,715
1986	44,282
1987	48,440
1988	50,251
1989	53,794
1990	55,972
1991	57,242

Fuente: *Industry Surveys*, 1993.

Maggie sabe que la serie debe diferenciarse para quitar la tendencia y crear una serie estacionaria. Hace primero la diferenciación de los datos mediante Minitab (véase “Paquete de Cómputo Minitab”) y desarrolla un correlograma similar al de la fig. 4.6. En éste aprecia que el coeficiente de autocorrelación para el periodo de desfaseamiento 3, .65, es significativamente diferente de cero y que los dos primeros coeficientes de autocorrelación y el cuarto están cerca

TABLA 4.5 VENTAS POR TRIMESTRE DE OUTBOARD MARINE, 1984–1993
(EJEMPLO 4.4)

AÑO FISCAL (al término)	DICIEMBRE 31	MARZO 31	JUNIO 30	SEPTIEMBRE 30
1984	147.6	251.8	273.1	249.1
1985	139.3	221.2	260.2	259.5
1986	140.5	245.5	298.8	287.0
1987	168.8	322.6	393.5	404.3
1988	259.7	401.1	464.6	479.7
1989	264.4	402.6	411.3	385.9
1990	232.7	309.2	310.7	293.0
1991	205.1	234.4	285.4	258.7
1992	193.2	263.7	292.5	315.2
1993	178.3	274.5	295.4	311.8

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1988, 1992), p. 1768.

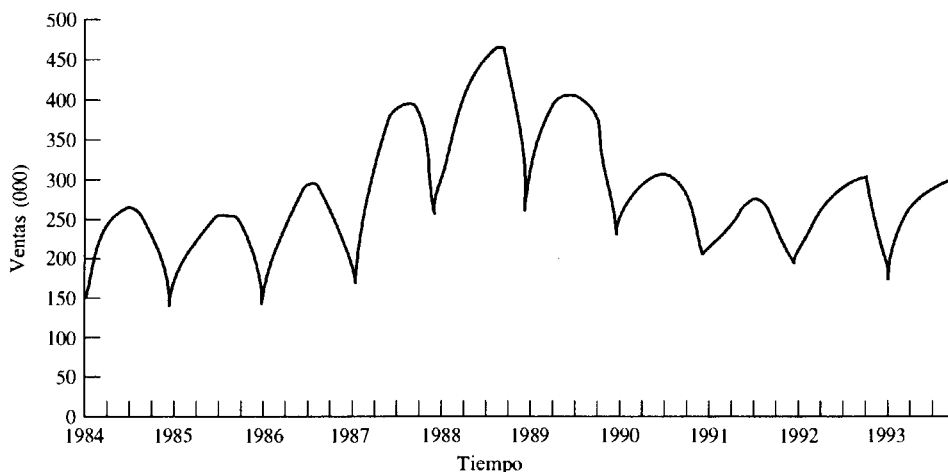


Figura 4.7 Gráfica de una serie estacional para los datos de Outboard Marine del ejemplo 4.4.

A continuación, Perkin calcula los coeficientes de autocorrelación que se muestran en la fig. 4.8. Nota que en el periodo de desfase 4, los coeficientes de autocorrelación son significativamente diferentes de cero ($.73 > .31$) y concluye que las ventas de Outboard Marine son estacionales en una base trimestral.

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

Un factor principal que influye en la selección de una técnica de pronóstico consiste en la identificación y comprensión de patrones históricos en los datos. Si se pueden reconocer patrones de tendencia, cíclicos o estacionales, entonces se pueden seleccionar las técnicas con la capacidad de utilizar eficazmente estos patrones.

Técnicas de pronóstico para datos estacionarios

Como se definió anteriormente, una *serie estacionaria* es aquella cuyo valor promedio no varía a través del tiempo. Estas situaciones se presentan cuando los patrones de demanda que influyen sobre la serie son relativamente estables. En su forma más simple, el pronóstico de series estacionarias comprende el uso de la historia disponible de las series para estimar su valor promedio, el cual se convierte después en el pronóstico de valores futuros. Las técnicas más sofisticadas comprenden la actualización de la estimación, al haber disponible nueva información. Estas técnicas son útiles cuando no son confiables las estimaciones iniciales o cuando la estabilidad del promedio es cuestionable. Además, las técnicas de actualización proporcionan cierto grado de tranquilidad con respecto a los cambios en la estructura subyacente de la serie.

Las técnicas de pronóstico estacionarias se emplean siempre que:

- *Las fuerzas que generan una serie se han estabilizado y el medio en el que existe la serie permanece relativamente sin cambios.* Ejemplos de ello son las fallas por semana en una línea de ensamble que tiene una tasa de producción uniforme., las ventas unitarias de un producto o servicio en la etapa de maduración de su ciclo de vida y el número de ventas resultantes de un nivel constante de esfuerzo.
- *Se requiere un modelo muy sencillo debido a la falta de datos o para facilitar su explicación o implementación.* Un ejemplo sería cuando un negocio u organización es nuevo y hay disponible muy poca información histórica.
- *Se puede lograr la estabilidad haciendo correcciones sencillas a factores como crecimiento de la población e inflación.* Ejemplos de esto son modificar el ingreso por el ingreso per cápita, o las ventas en dólares por montos en dólares constantes.
- *La serie se puede transformar en una serie estable.* Como ejemplo, está la transformación de series mediante logaritmos, raíces cuadradas o diferencias.
- *La serie es un conjunto de errores de pronóstico, de una técnica de pronóstico que se considera adecuada.* Véase el ejemplo 4.6.

Varias técnicas que se podrían considerar al pronosticar en series estacionarias son los métodos no formales, los métodos de promedio simple y los métodos de promedios móviles, atenuación exponencial y de Box–Jenkins.

Técnicas de pronóstico para datos con una tendencia

Anteriormente definimos a una serie *con tendencia* como una serie de tiempo que contiene un componente de largo plazo que representa el crecimiento o declinación de la serie a través de un periodo amplio. En otras palabras, se dice que una serie de tiempo tiene una tendencia si su valor promedio varía a través del tiempo, de modo que se espera que aumente o disminuya durante el periodo para el que se desea el pronóstico. Es común que las series económicas contengan una tendencia.

Las técnicas de pronóstico para series con tendencia se utilizan siempre que:

- *Una productividad creciente y la nueva tecnología conducen a cambios en el estilo de vida.* Como ejemplos se pueden citar la demanda de componentes electrónicos, que se incrementó con el advenimiento de la computadora; y el uso del ferrocarril que disminuyó con el advenimiento del avión.
- *El incremento en la población provoca un incremento en la demanda de bienes y servicios.* Ejemplos de esto son las utilidades por venta de bienes de consumo, la demanda en el consumo de energía eléctrica y el uso de materias primas.
- *El poder de compra del dólar afecta las variables económicas por causa de la inflación.* Los ejemplos son salarios, costos de producción y precios.
- *Aumenta la aceptación en el mercado.* Un ejemplo sería el periodo de crecimiento en el ciclo de vida de un nuevo producto.

Las técnicas a considerar al pronosticar series con tendencia son promedio móvil lineal, atenuación exponencial lineal de Brown, atenuación exponencial lineal de Holt, atenuación exponencial cuadrática de Brown, regresión simple, modelo de Gompertz, curvas de crecimiento y modelos exponenciales.

Técnicas de pronóstico para datos con estacionalidad

Anteriormente definimos una serie *estacional* como una serie de tiempo con un patrón de cambio que se repite a sí mismo año tras año. Por lo regular, el desarrollo de una técnica de pronóstico estacional comprende la selección de un método multiplicativo o uno de adición y estimar después índices estacionales a partir de la historia de la serie. Estos índices se usan posteriormente para incorporar la estacionalidad al pronóstico o para eliminar tales efectos de los valores observados. A este último proceso se le conoce como ajuste estacional de datos y se expone en el cap. 8.

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

- *El clima influye en la variable de interés.* Como ejemplos están el consumo de energía eléctrica, las actividades de verano e invierno (deportes como el patinaje), el guardarropa y las estaciones de desarrollo agrícola.
- *El año calendario influye en la variable de interés.* Ejemplos de ello son las ventas al menudeo influidas por días festivos, fines de semana de tres días y los calendarios escolares.

Las técnicas a considerar al pronosticar series estacionales son descomposición clásica, Census II, atenuación exponencial de Winter, regresión múltiple de series de tiempo y métodos de Box–Jenkins.

Técnicas de pronóstico para series cíclicas

El efecto *cíclico* se definió anteriormente como la fluctuación en forma de onda alrededor de la tendencia. Los patrones cíclicos tienden a repetirse en los datos cada dos, tres o más años. Es difícil establecer un modelo para estos patrones cíclicos, ya que no son estables. Las fluctuaciones en forma de onda hacia arriba y hacia abajo alrededor de la tendencia rara vez se repiten en intervalos fijos de tiempo y también varía la magnitud de las fluctuaciones. Para analizar datos cíclicos se pueden ampliar los métodos de descomposición (cap. 8). Sin embargo, debido al comportamiento irregular de los ciclos, el análisis de componentes cíclicos de una serie requiere a menudo de encontrar indicadores económicos principales o coincidentes.

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

- *El ciclo del negocio influye sobre la variable de interés.* Como ejemplos están los factores económicos de mercado y de competencia.
- *Se presentan cambios en el gusto popular.* Ejemplos de ello son la moda, la música y la alimentación.
- *Se presentan cambios en la población.* Podemos citar como ejemplos las guerras, escasez, epidemias y desastres naturales.
- *Se presentan cambios en el ciclo de vida del producto.* Ejemplos de ello son la introducción, crecimiento, maduración, saturación y declinación del mercado.

Las técnicas a considerar al pronosticar series cíclicas son la descomposición clásica, los indicadores económicos, los modelos econométricos, la regresión múltiple y los métodos de Box-Jenkins.

Otros factores por considerar al elegir una técnica de pronóstico

El horizonte en el tiempo para un pronóstico tiene una relación directa con la selección de una técnica de pronóstico. Para los pronósticos de corto y mediano plazos, se pueden aplicar diversas técnicas cuantitativas. Sin embargo, al aumentar el horizonte del pronóstico, algunas de estas técnicas se hacen menos aplicables. Por ejemplo, los promedios móviles, la atenuación exponencial y los modelos de Box-Jenkins no son muy buenos pronósticos de cambios económicos radicales, mientras que los modelos econométricos son más útiles para este fin. Los modelos de regresión, son apropiados para los periodos corto, mediano y largo. Las proyecciones de medias, promedios móviles, descomposición clásica y tendencia son técnicas cuantitativas apropiadas para horizontes de corto y mediano plazos. Las técnicas más complejas de Box-Jenkins y los modelos econométricos resultan también apropiados para pronósticos de corto y mediano plazos. Para horizontes mayores en el tiempo, se usan con frecuencia los métodos cualitativos (véase cap. 11).

En general, la aplicabilidad de las técnicas de pronóstico es algo que el pronosticador realiza con base en su experiencia. Es común que los administradores requieran de pronósticos en un tiempo relativamente corto. En esta situación, tienen ventaja los métodos de atenuación exponencial, proyección de tendencia, modelos de regresión y la descomposición clásica.

En todo caso, un pronóstico se presentará a los ejecutivos de la administración para su aprobación y uso en el proceso de toma de decisiones. De ahí que la facilidad de comprensión e interpretación de los resultados sea una consideración importante. Las técnicas de modelos de regresión, proyección de tendencias, descomposición clásica y atenuación exponencial, califican alto para este criterio.

Los costos de computadora ya no son una parte significativa en la elección de una técnica. Las computadoras de escritorio (microcomputadoras) y los paquetes de programas de pronóstico se han convertido en algo común en muchas organizaciones. Debido a estos desarrollos es probable que, en el futuro, otros criterios ignoren las consideraciones de costo de computadora.

Como parte final de la selección, el pronosticador debe evaluar cada técnica en términos de confiabilidad y aplicabilidad para el problema que le ocupa, su valor en términos de efectividad comparado con el de otras técnicas apropiadas, su nivel de precisión, su costo y aceptación por parte de la administración. La tabla 4.6 presenta un resumen de las técnicas de pronóstico que deben ser usadas con determinados patrones de datos.

TABLA 4.6 SELECCIÓN DE UNA TÉCNICA DE PRONÓSTICO

MÉTODO	PATRÓN DE DATOS	HORIZONTE EN EL TIEMPO	TIPO DE MODELO	REQUERIMIENTOS MÍNIMOS DE DATOS	
				NO ESTACIONALES	ESTACIONALES
No formales	ST, T, S	S	TS	1	
Promedios simples	ST	S	TS	30	
Promedios móviles	ST	S	TS	4-20	
Atenuación exponencial	ST	S	TS	2	
Atenuación exponencial lineal	T	S	TS	3	
Atenuación exponencial cuadrática	T	S	TS	4	
Atenuación exponencial estacional	S	S	TS		2*L
Filtración adaptiva	S	S	TS		5*L
Regresión simple	T	I	C	10	
Regresión múltiple	C, S	I	C	10*V	
Descomposición clásica	S	S	TS		5*L
Modelos de tendencia exponencial	T	I, L	TS	10	
Ajuste de curva-S	T	I, L	TS	10	
Modelos de Gompertz	T	I, L	TS	10	
Curvas de crecimiento	T	I, L	TS	10	
Census II	S	S	TS		6*L
Box-Jenkins	ST, T, C, S	S	TS	24	3*L
Indicadores principales	C	S	C	24	
Modelos econométricos	C	S	C	30	
Regresión múltiple de series de tiempo	T, S	I, L	C		6*L

Patrón de datos: ST, estacionarios; T, con tendencia; S, estacionales; C, cíclicos.

Horizonte en el tiempo: S, corto plazo (menos de 3 meses); I, mediano plazo; L, largo plazo.

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

Estacionales: L, longitud de la estacionalidad.

MEDICIÓN DEL ERROR EN EL PRONÓSTICO

Ya que las técnicas cuantitativas de pronóstico implican, por lo regular, series de tiempo de datos, se desarrolló una notación matemática para hacer referencia a cada periodo específico. Se empleará la letra y para denotar una variable de serie de tiempo, a menos que exista más de una variable. El periodo asociado con una observación se muestra como subíndice. Así, Y_t se refiere al valor de la serie de tiempo en el periodo t . Los datos trimestrales de Outboard Marine Corporation que se presentan en el ejemplo 4.4. se denotarían como $y_1 = 147.6$, $y_2 = 261.8$, $y_3 = 273.1$..., $y_{40} = 311.8$.

También se desarrolló una notación matemática para distinguir el valor real de una serie de tiempo y el valor de pronóstico. Se empleará el símbolo $\hat{}$ (acento circunflejo) sobre un valor, para indicar que se trata de un pronóstico. El valor de pronóstico para Y_t es $\hat{Y}_t$. Con frecuencia, se juzga la precisión de una técnica de pronóstico mediante la comparación de la serie original $Y_1, Y_2, \dots$ con la serie de pronóstico $\hat{Y}_1, \hat{Y}_2, \dots$.

La notación básica de pronóstico se resume como:

Notación Básica de Pronóstico

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

$\hat{Y}_t$ = valor del pronóstico para Y_t

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

Se han ideado diversos métodos para resumir los errores generados por una técnica particular de pronóstico. La mayoría de estas mediciones implican promediar alguna función de la diferencia entre el valor real y su valor de pronóstico. A menudo se denominan *residuales* a estas diferencias entre valores observados y los valores de pronóstico.

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

Para calcular el error o residual de cada periodo de pronóstico, se utiliza la ecuación 4.4.

$$e_t = Y_t - \hat{Y}_t \quad (4.4)$$

en donde

e_t = error del 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 pronóstico consiste en obtener la suma de los errores absolutos. La Desviación Absoluta de la Media (DAM) mide la precisión de un pronóstico mediante el promedio de la magnitud de los errores de pronóstico (valores absolutos de cada error). La DAM resulta de gran utilidad cuando el analista desea medir el error de pronóstico en las mismas unidades de la serie original. La ecuación 4.5 muestra cómo se calcula la DAM.

$$\text{DAM} = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n} \quad (4.5)$$

Otro método para evaluar una técnica de pronóstico es el Error Medio Cuadrado (EMC). Cada error o residual se eleva al cuadrado; luego, estos valores se suman y se divide entre el número de observaciones. Este enfoque penaliza los errores mayores de pronóstico ya que eleva cada uno al cuadrado. Esto es importante pues en ocasiones pudiera ser preferible una técnica que produzca errores moderados a otra que por lo regular tenga errores pequeños, pero que ocasionalmente arroje algunos en extremo grandes. La ecuación 4.6 muestra el cálculo del EMC.

$$\text{EMC} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n} \quad (4.6)$$

En ocasiones, resulta más útil calcular los errores de pronóstico en términos de porcentaje y no en cantidades. El Porcentaje de Error Medio Absoluto (PEMA) se calcula encontrando el error absoluto en cada periodo, dividiendo éste entre el valor real observado para ese periodo y después promediando estos errores absolutos de porcentaje. Este enfoque es útil cuando el tamaño o magnitud de la variable de pronóstico es importante en la evaluación de la precisión del pronóstico. El PEMA proporciona una indicación de qué tan grandes son los errores de pronóstico comparados con los valores reales de la serie. También se puede utilizar el PEMA para comparar la precisión de la misma u otra técnica sobre dos series completamente diferentes. La ecuación 4.7 muestra el cálculo del PEMA.

$$\text{PEMA} = \frac{\sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}}{n} \quad (4.7)$$

A veces resulta necesario determinar si un método de pronóstico está sesgado (pronóstico consistentemente alto o bajo). En estos casos, se emplea el Procentaje Medio de Error (PME), que se calcula encontrando el error en cada periodo, dividiendo esto entre

el valor real de ese periodo y promediando después estos porcentajes de error. Si un enfoque de pronóstico no está sesgado, la ecuación 4.8 producirá un porcentaje cercano a cero. Si el resultado es un porcentaje negativo grande, el método de pronóstico está sobrestimando de manera consistente. Si el resultado es un porcentaje positivo grande, el método de pronóstico está subestimando de manera consistente.

$$\text{PME} = \frac{\sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}}{n} \quad (4.8)$$

Una parte de la decisión para utilizar una técnica de pronóstico en particular es la determinación de si la técnica producirá errores de predicción que se juzguen como suficientemente pequeños. Es en efecto realista esperar que una técnica produzca errores de pronóstico relativamente bajos sobre una base consistente.

Las cuatro mediciones de precisión de un pronóstico que acabamos de describir se utilizan de la siguiente manera:

- La comparación de la precisión de dos técnicas diferentes.
- La medición de la utilidad o confiabilidad de una técnica.
- La búsqueda de una técnica óptima.

El ejemplo 4.5 ilustrará cómo calcular cada una de estas mediciones de error.

Ejemplo 4.5 La tabla 4.7 presenta los datos del número de clientes diarios que requieren trabajos de reparación Y_t , un pronóstico de estos datos $\hat{Y}_t$, en la Gary's Chevron Station. La técnica de pronóstico

TABLA 4.7 CÁLCULOS PARA LOS MÉTODOS DE EVALUACIÓN DE PRONÓSTICOS DEL EJEMPLO 4.5

PERIODO t	DATOS, Y_t CLIENTES	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	7.4	-7.4
3	60	54	6	6	36	10.0	10.0
4	55	60	-5	5	25	9.1	-9.1
5	62	55	7	7	49	11.3	11.3
6	62	62	0	0	0	0.0	0.0
7	65	62	3	3	9	4.6	4.6
8	63	65	-2	2	4	3.2	-3.2
9	70	63	7	7	49	10.0	10.0
Sumas			12	34	188	55.6	16.2

utiliza el número de clientes atendidos en el periodo anterior, como el pronóstico del periodo actual. Esta sencilla técnica se expondrá en el cap. 5. Los cálculos para evaluar este modelo mediante DAM, EMC, PEMA y PME se demuestran en la tabla 4.7.

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

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

$$\text{PEMA} = \frac{\sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}}{n} = \frac{55.6\%}{8} = 6.95\%$$

$$\text{PME} = \frac{\sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}}{n} = \frac{16.2\%}{8} = 2.03\%$$

La DAM indica que cada pronóstico está desviado en un promedio de 4.3 clientes. El EMC de 23.5 y el PEMA de 6.95% se compararán con el EMC y el PEMA de cualquier otro método empleado para pronosticar estos datos. Por último, un bajo PME de 2.03% indica que la técnica no está desviada; ya que su valor es cercano a cero, la técnica no sobrestima ni subestima en forma consistente el número diario de clientes atendidos.

DETERMINACIÓN DE LO ADECUADO DE UNA TÉCNICA DE PRONÓSTICO

Antes de proceder a evaluar una técnica elegida, se debe evaluar qué tan adecuada es la elección. El pronosticador deberá responder a las siguientes preguntas:

- ¿Son indicativos de una serie aleatoria los coeficientes de autocorrelación de los residuales? Esta pregunta se puede responder examinando los coeficientes de autocorrelación en la gráfica de residuales, como la que se muestra en el ejemplo 4.6.
- ¿Presentan los residuales una distribución aproximadamente normal? Esta pregunta se puede responder analizando un histograma de los residuales.
- ¿Tienen proporciones significativas de t todas las estimaciones de parámetros? La distribución t se expuso en el cap. 2 y la explicación de las proporciones de t se demostrarán en el cap. 6.
- ¿Es sencilla la técnica en su uso y fácil de entender por quienes toman las decisiones?

El gerente de control de calidad de una fábrica de lingotes de aluminio, requiere de un pronóstico semanal de los defectos de producción, para los altos directivos de la compañía.

Un banquero desea ver proyectada la utilidad mensual de un pequeño fabricante de bicicletas que solicita un préstamo mayor para triplicar su capacidad.

Una agencia del gobierno federal necesita proyecciones anuales de kilómetros promedio por litro en los automóviles americanos, para los próximos 10 años, a fin de elaborar recomendaciones de regulación.

Una compañía de ahorro y préstamo requiere pronosticar los préstamos fraudulentos durante los próximos dos años, a efecto de evitar la quiebra.

Una compañía que fabrica microprocesadores necesita de un pronóstico industrial sobre el número de computadoras personales vendidas durante los próximos cinco años, con el fin de planear su presupuesto de investigación y desarrollo.

GLOSARIO

Autocorrelación. Es la correlación que existe entre una variable desfasada uno o más periodos y la misma variable.

Componente aleatorio. Este componente mide la variabilidad de la serie de tiempo después de quitar otros componentes.

Componente cíclico. Es la fluctuación en forma de onda alrededor de la tendencia.

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

Correlograma. Es una herramienta gráfica para mostrar las autocorrelaciones de varios periodos desfasados de una serie de tiempo.

Residual. Es la diferencia entre el valor real observado y su valor de pronóstico.

Serie estacionaria. Es aquella cuyo valor estacionario no cambia a través del tiempo.

Serie de tiempo. Consiste en datos reunidos, registrados u observados en incrementos sucesivos de tiempo.

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

FÓRMULAS CLAVE

De primer orden

Coefficiente de autocorrelación

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (4.1)$$

De orden k -ésimo

$$\text{Coeficiente de autocorrelación} \quad r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (4.2)$$

Intervalos de confianza del
coeficiente de autocorrelación

$$0 \pm Z \frac{1}{\sqrt{n}} \quad (4.3)$$

Error de pronóstico o residual

$$e_t = Y_t - \hat{Y}_t \quad (4.4)$$

Desviación absoluta de la media

$$\text{DAM} = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n} \quad (4.5)$$

Error medio cuadrado

$$\text{EMC} = \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n} \quad (4.6)$$

Porcentaje de error medio absoluto

$$\text{PEMA} = \frac{\sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}}{n} \quad (4.7)$$

Porcentaje medio de error

$$\text{PME} = \frac{\sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}}{n} \quad (4.8)$$

PROBLEMAS

1. Explique las diferencias entre las técnicas de pronóstico cualitativas y cuantitativas.
2. ¿Qué es una serie de tiempo?
3. Describa cada uno de los cuatro componentes de una serie de tiempo.
4. ¿Qué es una autocorrelación?
5. ¿Qué mide un coeficiente de autocorrelación?
6. Describa cómo se emplean los correlogramas para analizar autocorrelaciones para varios desfases de una serie de tiempo.
7. Cada una de las siguientes declaraciones describe una serie ya sea estacionaria o no estacionaria. Indique cuál corresponde a cada una.
 - a. Una serie que tiene una tendencia.
 - b. Una serie que no tiene una tendencia pero es muy cíclica.
 - c. Una serie cuya media y varianza permanecen constantes a través del tiempo.
 - d. Una serie cuyo valor promedio cambia a través del tiempo.
 - e. Una serie que no presenta crecimiento o disminución.

8. A continuación se presentan descripciones para diversos tipos de series: aleatoria, estacionaria, con tendencia o estacional. Identifique cada tipo de serie.
 - a. Una serie que tiene propiedades estadísticas básicas como la media y la varianza que permanecen constantes a través del tiempo.
 - b. Los valores sucesivos de una serie de tiempo no se relacionan entre sí.
 - c. Existe una elevada relación entre cada valor sucesivo de una serie.
 - d. Aparece un coeficiente de autocorrelación significativo en el periodo de desfase 4.
 - e. Una serie que no presenta crecimiento o disminución.
 - f. Los coeficientes de autocorrelación son diferentes de cero en forma significativa durante los primeros periodos de desfase y después caen a cero al incrementarse el número de periodos.
9. Enliste algunas de las técnicas de pronóstico a considerar al pronosticar una serie estacionaria. Dé ejemplos de situaciones en las que se puedan aplicar estas técnicas.
10. Enliste algunas de las técnicas de pronóstico a considerar al pronosticar una serie con tendencia. Dé ejemplos de situaciones en las que se puedan aplicar estas técnicas.
11. Enliste algunas de las técnicas de pronóstico a considerar al pronosticar una serie estacional. Dé ejemplos de situaciones en las que se puedan aplicar estas técnicas.
12. Enliste algunas de las técnicas de pronóstico a considerar al pronosticar una serie cíclica. Dé ejemplos de situaciones en las que se puedan aplicar estas técnicas.
13. En la tabla siguiente se presenta el número de matrimonios en los Estados Unidos. Calcule las primeras diferencias para estos datos.

AÑO	MILES DE MATRIMONIOS
1985	2,425
1986	2,400
1987	2,421
1988	2,389
1989	2,404

Fuente: *Statistical Abstract of the United States*,
1988, 1989, 1991.

14. Calcule un intervalo de confianza de 95% para los coeficientes de autocorrelación de una serie que contiene 80 términos.
15. Calcule un intervalo de confianza de 90% para los coeficientes de autocorrelación de una serie que contiene 60 términos.
16. ¿Cuál método de evaluación de la técnica de pronóstico debería emplearse en cada una de las situaciones siguientes?

- a. El analista desea determinar si el método de pronóstico tiene algún sesgo.
 - b. El analista piensa que el tamaño o magnitud de la variable por pronosticar es importante en la evaluación de la precisión del pronóstico.
 - c. El analista requiere penalizar los errores de pronóstico mayores.
17. ¿Cuáles de las siguientes declaraciones son verdaderas, considerando los métodos empleados para evaluar pronósticos?
- a. El PEMA toma en cuenta la magnitud de los valores que se pronostican.
 - b. El EMC penaliza los errores mayores.
 - c. El PME se usa para determinar si un modelo está pronosticando muy alto o muy bajo de manera sistemática.
 - d. La ventaja del método DAM consiste en que relaciona el tamaño del error con la observación real.
18. A Allie White, oficial en jefe de préstamos del Dominion Bank, le gustaría analizar la cartera de préstamos del banco para los años de 1984 a 1989. Los datos son los siguientes:

PRESTAMOS DEL DOMINION BANK POR TRIMESTRE, 1984–1989
(MILLONES DE DOLARES)

CALENDARIO	MAR. 31	JUN. 30	SEPT. 30	DIC. 31
1984	2313	2495	2609	2792
1985	2860	3099	3202	3161
1986	3399	3471	3545	3851
1987	4458	4850	5093	5318
1988	5756	6013	6158	6289
1989	6369	6568	6646	6861

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990), p. 2018.

- a. Calcule las autocorrelaciones para los periodos de desfase 1 y 2. Haga pruebas para determinar si estos coeficientes de autocorrelación son significativamente diferentes de cero, al nivel de significancia de .05.
 - b. Use un programa de cómputo para determinar si los datos son estacionarios.
19. Esta pregunta se refiere al problema 18. Obtenga la primera diferenciación para los datos de préstamos trimestrales del Dominion Bank.
- a. Calcule el coeficiente de autocorrelación para el periodo de desfase 1, utilizando los datos diferenciados.
 - b. Use un programa de cómputo para determinar si estos datos diferenciados son estacionarios.
20. Analice los coeficientes de autocorrelación de las series que se muestran en las figs. 4.11 a 4.14. Describa brevemente cada serie.

[illegible]

Figura 4.11 Coeficientes de autocorrelación del problema 20.

[illegible]

Figura 4.12 Coeficientes de autocorrelación del problema 20.

TIME LAG	AUTOCORRELATION
12	-0.22
11	-0.16
10	-0.10
9	-0.13
8	-0.09
7	0.00
6	0.10
5	0.24
4	0.43
3	0.59
2	0.71
1	0.74

Figura 4.13 Coeficientes de autocorrelación del problema 20.

- b. Evalúe mediante DAM
- c. Evalúe mediante EMC
- d. Evalúe mediante PEMA
- e. Evalúe mediante PME
- f. Redacte un memorando resumiendo sus observaciones.

CASO DE ESTUDIO 4.1

MURPHY BROTHERS FURNITURE

En 1958, los hermanos Murphy establecieron una mueblería en el centro de Dallas. A través de los años alcanzaron un éxito considerable y extendieron su cobertura hacia el oeste y el medio oeste. Para 1994, su cadena de mueblerías se encuentra bien establecida en 36 estados.

Julie Murphy, hija de uno de los fundadores, terminó recientemente sus estudios de administración y se unió a la empresa. Su padre y su tío se habían especializado en muchas áreas, pero no en el área de las habilidades cuantitativas. En particular, ambos pensaban que no podían pronosticar con precisión las ventas futuras de Murphy Brothers mediante técnicas modernas computarizadas. Por este motivo acudieron a Julie para que los ayudara como parte de su nuevo trabajo.

Julie había estudiado pronósticos en su programa de administración y ordenó un paquete de programas, junto con Minitab, para utilizarlos en la computadora de su oficina. Consideró primero utilizar como su variable las ventas en dólares de Murphy, pero se encontró con que faltaban varios años del historial. Le preguntó a su padre al respecto y éste le respondió que, en su momento, “no pensó que fueran tan importantes”.

Julie pensó que era probable que las ventas de Murphy estuvieran relacionadas con las cifras de ventas nacionales y optó por buscar una variable apropiada en alguna de las diversas publicaciones

TABLA 4.8 VENTAS MENSUALES DE TODAS LAS TIENDAS AL MENUDEO, 1983–1992
(MILES DE MILLONES DE DÓLARES)

	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992
Ene.	81.3	93.1	98.8	105.6	106.4	113.6	122.5	132.6	130.9	140.2
Feb.	78.9	93.7	95.6	99.7	105.8	115.0	118.9	127.3	128.6	141.5
Mar.	93.8	103.3	110.2	114.2	120.4	131.6	141.3	148.3	149.3	153.0
Abr.	93.8	103.9	113.1	115.7	125.4	130.9	139.8	145.0	148.5	157.3
May.	97.8	111.8	120.3	125.4	129.1	136.0	150.3	154.1	159.8	163.8
Jun.	100.6	112.3	115.0	120.4	129.0	137.5	149.0	153.5	153.9	162.7
Jul.	99.4	106.9	115.5	120.7	129.3	134.1	144.6	148.9	154.6	164.2
Ago.	100.1	111.2	121.1	124.1	131.5	138.7	153.0	157.4	159.9	164.3
Sep.	97.9	104.0	113.8	124.4	124.5	131.9	144.1	145.6	146.7	158.8
Oct.	100.7	109.6	115.8	123.8	128.3	133.8	142.3	151.5	152.1	166.8
Nov.	103.9	113.5	118.1	121.4	126.9	140.2	148.8	156.1	155.6	164.7
Dic.	125.8	132.3	138.6	152.1	157.2	171.0	176.5	179.7	181.0	200.2

Fuente: *Survey of Current Business*, varios años.

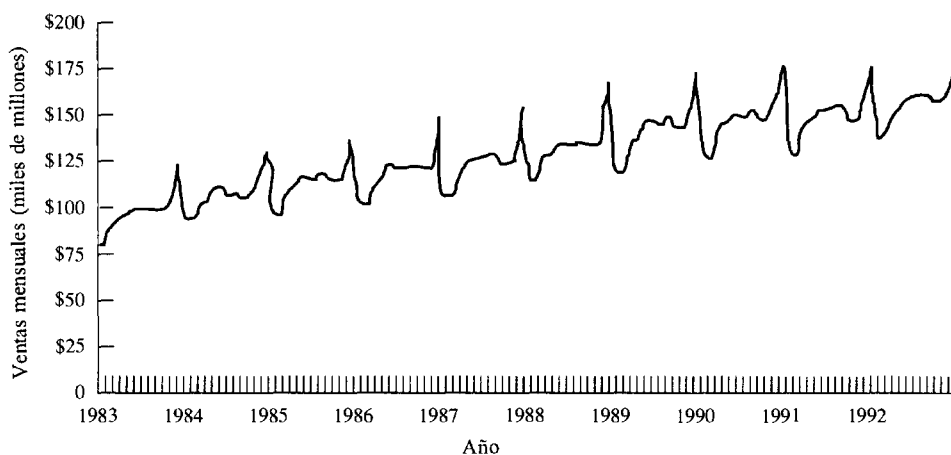


Figura 4.15 Estimación de ventas mensuales de todas las tiendas al menudeo en EUA, 1983–1992.

TIME LAG		AUTOCORRELATION
24	. I . *	0.53
23	. I . *	0.34
22	. I . *	0.28
21	. I . *	0.34
20	. I . *	0.38
19	. I . *	0.42
18	. I . *	0.42
17	. I . *	0.45
16	. I . *	0.45
15	. I . *	0.44
14	. I . *	0.42
13	. I . *	0.50
12	. I . *	0.73
11	. I . *	0.52
10	. I . *	0.47
9	. I . *	0.54
8	. I . *	0.58
7	. I . *	0.62
6	. I . *	0.63
5	. I . *	0.66
4	. I . *	0.66
3	. I . *	0.66
2	. I . *	0.65
1	. I . *	0.73

I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I.I
-1 0 +1

Figura 4.16 Autocorrelaciones de los datos de ventas de las tiendas al menudeo en EUA, para el caso de estudio 4.1.

federales. Después de buscar en una copia reciente de *Survey of Current Business*, encontró el historial de las ventas mensuales de todas las tiendas al menudeo en EUA y decidió usar esta variable como sustituto de su variable de interés, ventas en dólares de Murphy Brothers. Razonó que si podía establecer pronósticos precisos para las ventas nacionales, podría relacionar estos pronósticos con las propias ventas de Murphy y obtener el pronóstico deseado.

La tabla 4.8 presenta los datos reunidos por Julie y la fig. 4.15 muestra una gráfica de datos obtenida por su programa de cómputo. Julie comenzó su análisis empleando la computadora para desarrollar una gráfica de los coeficientes de autocorrelación.

Después de examinar el correlograma de la fig. 4.16, le pareció obvio a Julie que sus datos tuvieran una tendencia. Los primeros coeficientes de autocorrelación son muy grandes y gradualmente caen hacia cero al transcurrir el tiempo. Para hacer estacionaria la serie con el fin de considerar varios métodos de pronóstico, Julie decidió hacer una primera diferenciación de sus datos para ver si podía quitar la tendencia. El correlograma para la primera diferenciación de datos se muestra en la fig. 4.17.

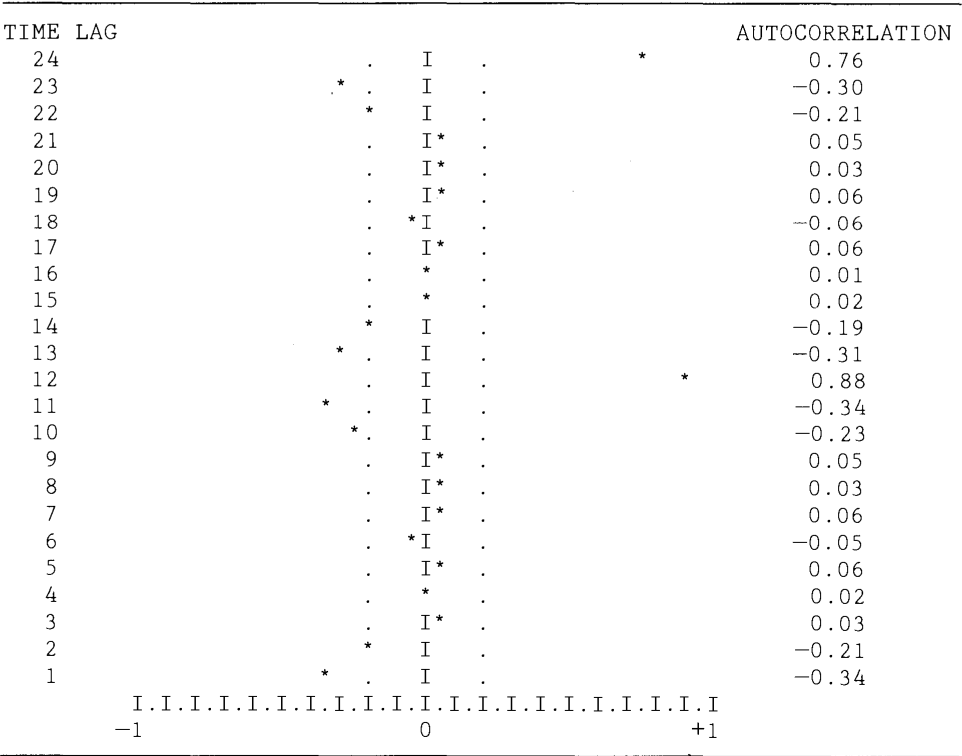


Figura 4.17 Autocorrelaciones de la primera diferenciación de los datos de ventas de las tiendas al menudeo en los EUA, para el caso de estudio 4.1.

PREGUNTAS

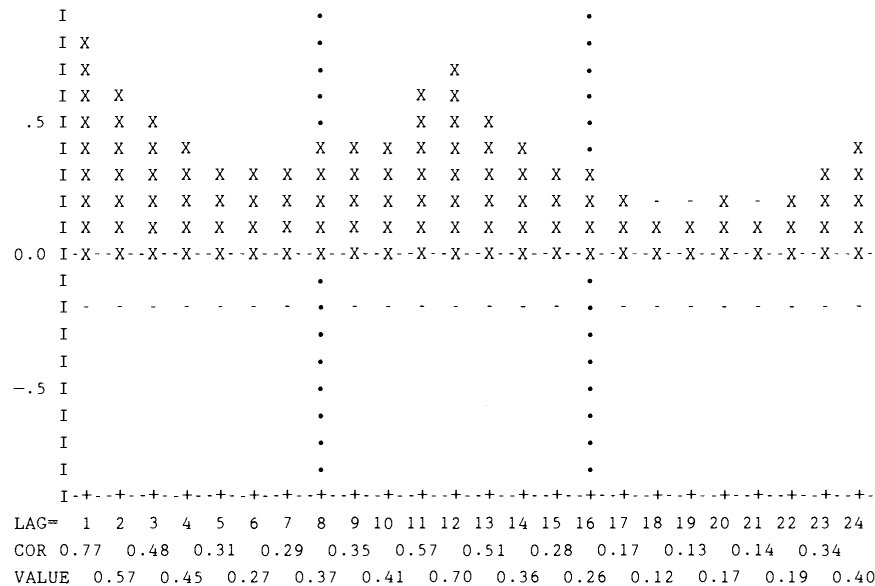
1. ¿A qué conclusión debe llegar Julie con respecto a las ventas al menudeo?
2. ¿Ha logrado Julie algún progreso para encontrar una técnica de pronóstico?
3. ¿Qué técnicas de pronóstico debería intentar Julie?
4. ¿Cómo sabrá Julie qué técnica funciona mejor?

CASO DE ESTUDIO 4.2

MR. TUX

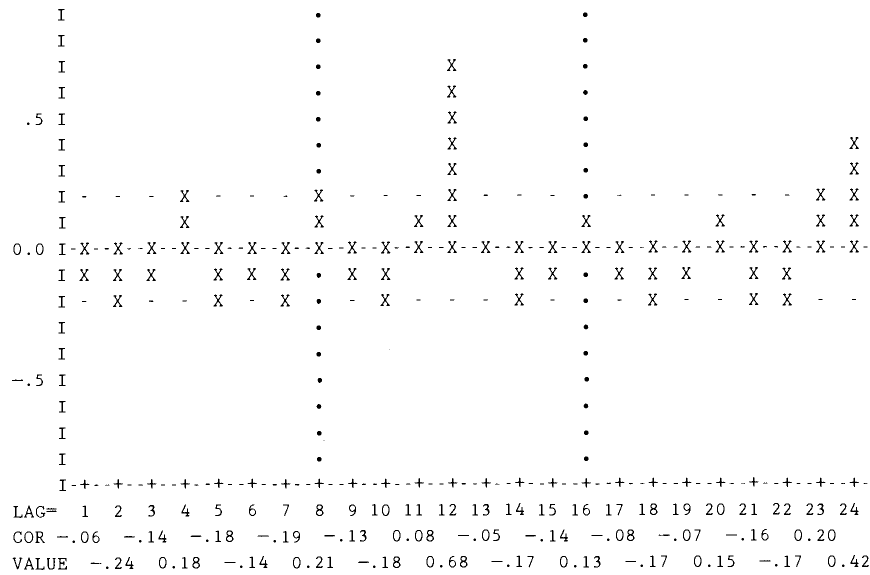
John Mosby, propietario de varias tiendas de renta Mr. Tux, comienza a pronosticar la variable más importante de su negocio, ventas mensuales en dólares (véanse los casos de estudio Mr. Tux al final de los caps. 1 y 2). Una de sus empleadas Virginia Perot, reunió los datos de ventas que se presentan en el caso de estudio del cap. 2.

John tiene un amigo con el que conversa sobre sus esfuerzos de pronóstico y éste le ofrece utilizar su programa de cómputo Sybil/Runner. John le echa un vistazo al programa y decide usarlo para pronosticar sus datos de ventas mensuales. Utiliza los 96 meses de datos reunidos y hace que el programa desarrolle el siguiente correlograma:



El programa reporta que "... hay algún tipo de patrón en los datos, es decir, los datos no presentan una distribución aleatoria, ya que hay 18 valores que se ubican fuera de los límites de control". Debido a que todos los coeficientes de autocorrelación son positivos y se desvanecen. John concluye que sus datos tienen una tendencia.

Entonces, solicita al programa la primera diferenciación de los datos y ejecuta un correlograma:



Esta vez, recibe el mismo mensaje, excepto que sólo cuatro coeficientes se ubican fuera de los límites de control. El programa establece que “...los datos son estacionarios en la primera diferenciación”. John advierte que se presentan coeficientes de autocorrelación significativos en los periodos de desfase 12 y 24 y concluye que sus datos mensuales son estacionales.

PERIODO DE DESFASE	COEFICIENTE
12	.68
24	.42

El programa termina calculando el porcentaje de la varianza en los datos originales explicada por los factores en el análisis:

FACTOR	% EXPLICADO
Datos	100
Tendencia	6
Estacionalidad	45
Aleatorio	49

TAREA

1. Resuma en un párrafo, dirigido a un directivo, no a un pronosticador, los resultados del análisis de John.
2. Describa los efectos de la tendencia y la estacionalidad que parecen estar presentes en las ventas de Mr. Tux.
3. ¿Cómo explicaría la línea “aleatorio = 49%” que aparece al final de la impresión de salida del programa Sybil/Runner?
4. Suponga que el periodo de desfase 36 es significativo, ¿concluiría que la estacionalidad tiene una tendencia? De ser así, explique qué significa una tendencia en la estacionalidad.

CASO DE ESTUDIO 4.3**CONSUMER CREDIT COUNSELING**

La operación de Consumer Credit Counseling (CCC) se describió en los caps. 1 (caso de estudio 1.2) y 3 (caso de estudio 3.4).

A Marv Harnishfeger, Director Ejecutivo, le preocupa el tamaño y calendario del equipo de personal para el resto de 1993. Marv explica el problema a Dorothy Mercer, recién electa presidenta del comité ejecutivo. Después de meditar el problema, Dorothy concluye que CCC necesita analizar el número de nuevos clientes adquiridos cada mes. Dorothy, quien trabajó en una empresa de servicio local, y está familiarizada con diversas técnicas de exploración de datos, estuvo de acuerdo en analizar el problema. Para ello, solicitó a Marv los datos de nuevos clientes atendidos por mes. Marv le proporcionó los datos mensuales del número de nuevos clientes atendidos por CCC en el periodo de enero de 1985 a marzo de 1993:

1985:	182	136	99	77	75	63	87	73	83	82	74	75
1986:	102	121	128	128	112	122	104	108	97	141	97	87
1987:	145	103	113	150	100	131	96	92	88	118	102	98
1988:	101	153	138	107	100	115	78	106	94	93	103	104
1989:	150	102	151	100	100	98	97	120	98	135	141	67
1990:	127	146	175	110	153	117	121	121	131	147	121	110
1991:	171	185	172	168	142	152	151	141	128	151	121	126
1992:	166	138	175	108	112	147	168	149	145	149	169	138
1993:	152	151	199									

TAREA

1. Explique cómo utilizó Dorothy el análisis de autocorrelación para explorar el patrón de datos para el número de clientes atendidos por CCC.
2. ¿Cuál fue la conclusión de Dorothy después de terminar el análisis?
3. ¿Qué tipo de técnica de pronóstico recomendó Dorothy para este conjunto de datos?

VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Se entra al paquete de exploración de datos seleccionando la opción 2 del menú principal. El sistema le preguntará el nombre del archivo de datos, con el mensaje:

```
What is the name of the data file? CREDIT
```

El menú le permite elegir entre seis diferentes técnicas de exploración. En la pantalla aparece:

```
Data file name: B:CREDIT
Variable:      NUMBER OF CLIENTS
```

```
***Exploratory Package***
```

```
1. Time Plot
```


2. 4235 HT Robust Smoothing
3. Box Plot (trend/cycle variation)
4. Aggregate Box Plot (seasonal variation)
5. Spread vs Level Plot
6. Autocorrelation function
7. Return to Main Menu

Which?

AUTOCORRELATION FUNCTION

99 Observations in the Working Series

Mean of the Working Series = 122.15

Standard Deviation of the Working Series = 30.043

Lag	Value	T-Value	-1.0	0.0	+1.0
1	0.51	5.05	:	[:****]*****	:
2	0.45	3.63	:	[:****]*****	:
3	0.37	2.65	:	[:****]***	:
4	0.36	2.41	:	[:****]***	:
5	0.31	1.99	:	[:****]	:
6	0.27	1.67	:	[:****]	:
7	0.28	1.68	:	[:****]	:
8	0.28	1.62	:	[:****]	:
9	0.22	1.22	:	[:****]	:
10	0.24	1.35	:	[:****]	:
11	0.31	1.70	:	[:****]	:
12	0.48	2.56	:	[:****]***	:
13	0.33	1.63	:	[:****]	:
14	0.35	1.68	:	[:****]	:
15	0.19	0.91	:	[:****]	:
16	0.22	1.04	:	[:****]	:
17	0.32	1.50	:	[:****]	:
18	0.12	0.54	:	[:****]	:
19	0.19	0.85	:	[:****]	:
20	0.23	1.01	:	[:****]	:
21	0.28	1.26	:	[:****]	:
22	0.18	0.77	:	[:****]	:
23	0.29	1.25	:	[:****]	:
24	0.44	1.89	:	[:****]	:
25	0.25	1.05	:	[:****]	:
26	0.22	0.91	:	[:****]	:
27	0.14	0.56	:	[:****]	:
28	0.16	0.66	:	[:****]	:
29	0.24	0.98	:	[:****]	:
30	-0.07	-0.26	:	[**:]	:
31	0.09	0.37	:	[**:]	:
32	-0.00	-0.01	:	[:]	:
33	0.13	0.50	:	[**:]	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 33 Degrees of Freedom = 261.7

Figura 4.18 Número de clientes—Análisis exploratorio de los datos del caso de estudio 4.3.

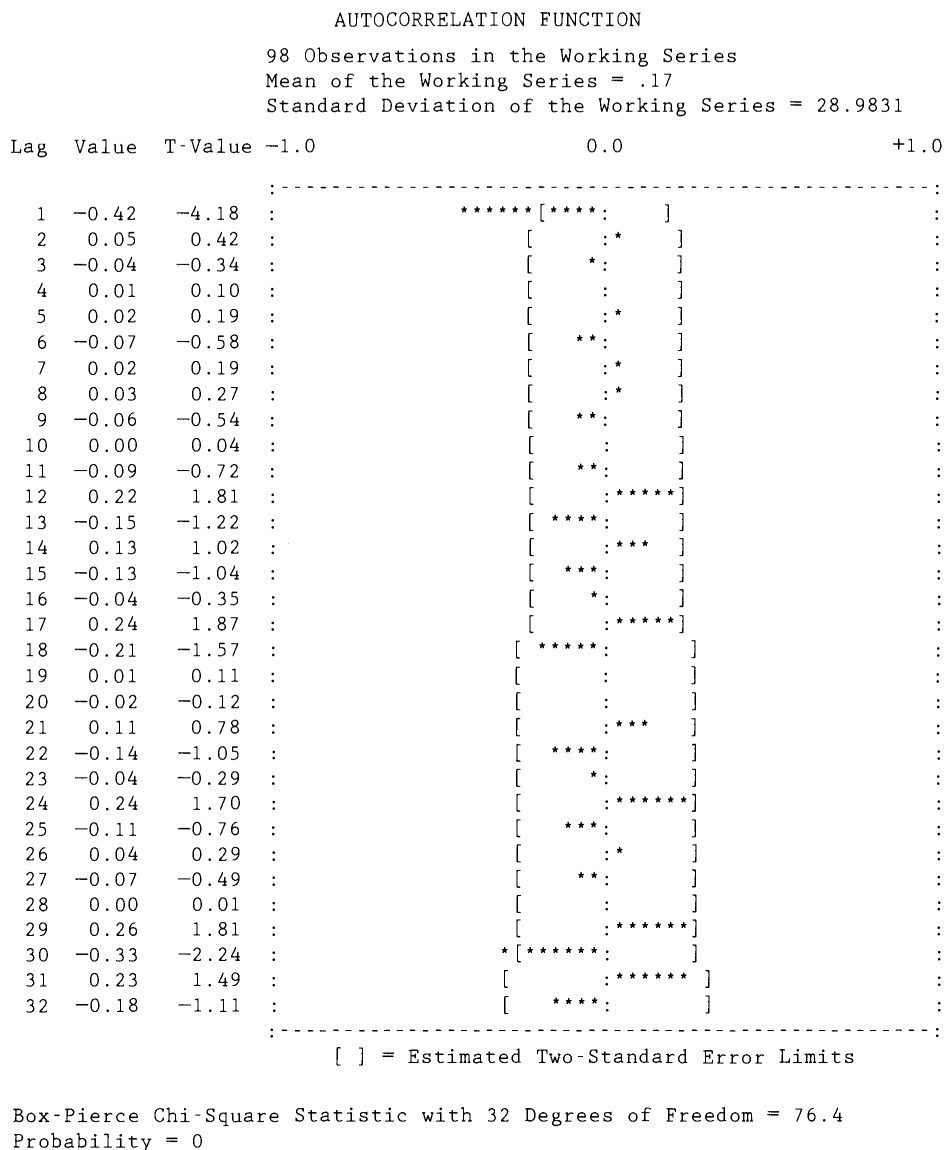


Figura 4.19 Número de clientes—Análisis exploratorio, primera diferenciación, para el caso de estudio 4.3.

El primer paso en cualquier análisis exploratorio debe ser una gráfica del periodo (Time Plot). Por lo regular, es mucho lo que se puede aprender al observar una gráfica de los datos sin procesar. Para nuestro ejemplo, elegimos la función de autocorrelación, opción 6. En la pantalla aparece lo siguiente:

```
*** Autocorrelation Function ***

Transformation to original data series:      7
(1=Reciprocal, 2=1/Sqrt, 3=1/4th root,
4=Log, 5=4th root, 6=Sqrt, 7=None)

Degree of regular differencing:             0

Degree of seasonal differencing:            0

Length of seasonality:                     12

Number of lags to print:                    A

Press ENTER to accept:
```

Para ejecutar un correlograma para nuestros datos, podemos utilizar las opciones por omisión ya provistas. Deseamos utilizar la serie de datos originales sin diferenciación y, ya que los datos son mensuales, la estacionalidad es 12. El número de periodos de desfase a imprimir tomará, por omisión, el valor de automático (número de puntos de datos entre 3). Al presionar <Enter>, se imprime la información de la fig. 4.18.

Los coeficientes de autocorrelación muestran que los datos no son estacionarios (tienen tendencia). La diferenciación se emplea para eliminar la tendencia. Los datos se vuelven a procesar usando la diferenciación regular de 1. La fig. 4.19 muestra el correlograma de los datos de la primera diferenciación.

PRONÓSTICO MEDIANTE HOJAS DE CÁLCULO

El problema

Harry Vernon desea emplear Lotus 1-2-3 (versión 2.0) para calcular los coeficientes de autocorrelación de primer orden para los datos que se presentan en la tabla 4.1. Dichos cálculos se muestran en la tabla 4.2.

La solución mediante hoja de cálculo

1. Cargue el archivo que guardó en el cap. 3. Escriba /FILE RETRIEVE o /FR. Esto equivale a seleccionar FILE en el menú principal y RETRIEVE en el submenú. En la pantalla aparecerá el mensaje:

```
Name of file to retrieve: A:/*WK1
```

En la tercera línea aparece el nombre de la hoja de cálculo que guardó en el cap. 3, VCR1.WK1. Use las teclas de flechas para resaltar este archivo y oprima <Enter>. En la pantalla aparece la hoja de cálculo VCR1.WK1.

- Coloque el apuntador de celda en A4, teclee cinco espacios y luego Y_t , oprima <→> y teclee cuatro espacios y después Y_{t-1} . Oprima <Enter>.
- Para crear la columna Y_{t-1} , oprima /COPY o /C. Coloque el apuntador de celda en B5. Lotus le solicita el rango de celdas a copiar, con el mensaje:

Enter range to copy FROM: B5..B5

Presione la tecla de punto para anclar el resaltado comenzando en la celda B5. Presione la tecla de flecha hacia abajo <↓> para extender el área resaltada hasta B15. Oprima <Enter>. Coloque ahora el apuntador de celda en C6 y presione <Enter> para copiar la columna en la pantalla. En la fig. 4.20 se muestra el resultado.

	A	B	C	D	E	F	G	H
1	VERNON'S MUSIC STORE							
2	NUMBER OF VCRS SOLD							
3								
4	Month	Yt	Yt-1					
5	January	123		-19		361		
6	February	130	123	-12	-19	144	228	
7	March	125	130	-17	-12	289	204	
8	April	138	125	-4	-17	16	68	
9	May	145	138	3	-4	9	-12	
10	June	142	145	0	3	0	0	
11	July	141	142	-1	0	1	0	
12	August	146	141	4	-1	16	-4	
13	September	147	146	5	4	25	20	
14	October	157	147	15	5	225	75	
15	November	150	157	8	15	64	120	
16	December	160	150	18	8	324	144	
17		1704		0	-18	1474	843	
18								
19	Mean =	142.00						
20								
21	r =	0.571913						
22								
23								
24								

Figura 4.20 Coeficiente de autocorrelación de primer orden para el número de videocasetas vendidas por Vernon's Music Store.

4. Para crear la columna $(Y_t - \bar{Y})$, calcule la media de la columna Y_t . Coloque el apuntador de celda en B19, escriba MEAN =. Presione <→>. Ahora está ubicado en C19. Lotus 1-2-3 tiene una función integrada que le permite calcular la media aritmética de los valores comprendidos en un rango. Escriba @AVG(B5..B16) y oprima <Enter>. En C19 aparece una media de 142. Ahora colóquese en D5 y escriba la fórmula +B5-142. El signo más (+) le indica al programa que se va a introducir una fórmula. Oprima <Enter> y en D5 aparece -19. Ahora presione /COPY para copiar la fórmula al resto de la columna. Oprima la tecla de flecha hacia abajo <↓> y la tecla de punto para indicar el comienzo del resaltado en la celda D6. Oprima <↓> para ampliar el resaltado hasta D16. Oprima <Enter>. El resultado se muestra en la fig. 4.20. Nótese que si mueve el indicador de celda a D16 y observa la línea superior, ésta mostrará +B16-142.
5. Para crear la columna $(Y_{t-1} - \bar{Y})$, coloque el indicador de celda en E6, escriba la fórmula +C6-142 y cópiela al resto de la columna.
6. Para crear la columna $(Y_t - \bar{Y})^2$, coloque el indicador de celda en F5, escriba la fórmula +D5*D5 y cópiela al resto de la columna.
7. Para crear la última columna, coloque el indicador de celda en G6, escriba la fórmula +D6*E6 y cópiela al resto de la columna.
8. Para sumar las columnas, coloque el apuntador de celda en B17 y escriba @SUM(B5..B16). Oprima <Enter> y en B17 aparecerá la suma 1704. Nótese que en la línea uno en la parte superior de la pantalla aparece la función de suma @SUM(B5..B16). Copie ahora la suma de B17 en C17 hasta G17.
9. Por último, calcule el coeficiente de autocorrelación colocando el apuntador de celda en B21 y escribiendo $r =$. Después de presionar <→>, escriba la fórmula +G17/F17. La respuesta aparecerá en C21.

PAQUETE DE CÓMPUTO MINITAB

El problema. En el ejemplo 4.3, Maggie Trymane, una analista de Sears, desea pronosticar las ventas para 1992. Ella requiere determinar el patrón para los datos de ventas de los años 1955 a 1991.

La solución con Minitab

```
MTB > READ 'SEARS.DAT' INTO C1
      31 ROWS READ
```

```
C1
  3307      3556      3601      3721      .      .      .
```

```
MTB > ACF C1
```

```
ACF of C1
      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +---+---+---+---+---+---+---+---+---+---+
  1  0.865                                     XXXXXXXXXXXXXXXXXXXX
  2  0.725                                     XXXXXXXXXXXXXXXXXXXX
  3  0.591                                     XXXXXXXXXXXXXXXXXXXX
  4  0.477                                     XXXXXXXXXXXXXXXX
  5  0.371                                     XXXXXXXXXXXX
  6  0.271                                     XXXXXXXX
  7  0.220                                     XXXXXXX
  8  0.158                                     XXXXX
  9  0.094                                     XXX
 10  0.041                                     XX
 11 -0.005                                     X
 12 -0.053                                     XX
 13 -0.100                                     XXXX
 14 -0.140                                     XXXX
 15 -0.174                                     XXXXX
```

```
MTB > DIFFERENCES 1 C1, STORE C2
```

```
MTB > ACF C2
```

```
ACF of C2
      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +---+---+---+---+---+---+---+---+---+---+
  1  0.275                                     XXXXXXXX
  2  0.248                                     XXXXXXX
  3  0.529                                     XXXXXXXXXXXXXXXX
  4  0.177                                     XXXXX
  5  0.055                                     XX
  6  0.069                                     XXX
  7  0.080                                     XXX
  8  0.024                                     XX
  9  0.007                                     X
 10 -0.020                                     X
```

```
11 -0.027          XX
12 -0.025          XX
13 -0.072         XXX
14 -0.087         XXX
15 -0.076         XXX
MTB > STOP
```

- El comando READ se usa para cargar en C1 el archivo de datos denominado SEARS.DAT.
- El comando ACF se emplea para calcular los coeficientes de autocorrelación de los datos de ventas de Sears.
- El comando DIFFERENCES se utiliza para calcular la primera diferenciación de los datos en C1 y almacenar estas diferencias en C2.
- El comando ACF se usa para calcular los coeficientes de autocorrelación para la primera diferenciación de los datos de ventas de Sears.

PAQUETE DE CÓMPUTO SAS

Problema. En el ejemplo 4.4, Perkin Kendell, un analista de Outboard Marine, desea usar la autocorrelación para determinar si sus datos son estacionales.

Solución con SAS

```
$ SAS/CC=PRINT
OPTIONS PAGESIZE=60 LINESIZE=80;
TITLE "SAS AUTOCORRELATION EXAMPLE";
DATA OUTBOARD;
INPUT SALES;
CARDS;
147.6
251.8
273.1
249.1
139.3
221.2
260.2
259.5
140.5
245.5
```

```
298.8
287.0
168.8
322.6
393.5
404.3
259.7
401.1
464.6
479.7
264.4
402.6
411.3
385.9
232.7
309.2
310.7
293.0
205.1
234.4
285.4
258.7
193.2
263.7
292.5
315.2
178.3
274.5
295.4
311.8

PROC ARIMA;
  IDENTIFY VAR=OUTBOARD;
ENDSAS;
```

- Las dos primeras líneas proporcionan instrucciones al sistema sobre comandos e impresión.
- La declaración TITLE da nombre al programa SAS.
- La declaración DATA proporciona un nombre al archivo de datos.
- La declaración INPUT da nombre a la columna de datos SALES (ventas).
- La declaración CARDS indica que las siguientes 40 líneas son imágenes de tarjetas que representan las ventas trimestrales de Outboard Marine de 1984 a 1992.
- El comando PROC ARIMA determina las autocorrelaciones significativas entre los periodos de desfase dados.
- El subcomando IDENTIFY indica que la variable para analizar es OUTBOARD.
- La declaración ENDSAS hace que termine la ejecución de SAS.

La salida de cómputo de esta ejecución de SAS se presenta en la fig. 4.21.


```

SAS AUTOCORRELATION EXAMPLE
      ARIMA Procedure

Name of variable = Y.
Mean of working series = 284.6475
Standard deviation      = 82.46956
Number of observations = 40

Autocorrelations
Lag  Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0    6801.228    1.00000  :      : : : : : : : : : : : : : : : : : : : : :
1    2866.123    0.42141  :      :      : : : : : : : : : : : : : : : : :
2    1320.944    0.19422  :      :      :      : : : : : : : : : : : : : : :
3    2121.483    0.31193  :      :      :      :      : : : : : : : : : : : : :
4    4986.262    0.73314  :      :      :      :      :      : : : : : : : : :
5    1206.109    0.17734  :      :      :      :      :      :      : : : : : : : :
6    -811.990    -0.11939  :      :      :      :      :      :      :      : : : : : : :
7    -356.159    -0.05237  :      :      :      :      :      :      :      :      : : : :
8    2022.852    0.29742  :      :      :      :      :      :      :      :      :      : :
9    -1311.359   -0.19281  :      :      :      :      :      :      :      :      :      :      :
10   -3004.888   -0.44182  :      :      :      :      :      :      :      :      :      :      :
      " ." marks two standard errors

Inverse Autocorrelations
Lag  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1    -0.08825  :      :      :      :      :      :      :      :      :      : :
2    -0.10834  :      :      :      :      :      :      :      :      :      : :
3     0.01038  :      :      :      :      :      :      :      :      :      : :
4    -0.50030  :      :      :      :      :      :      :      :      :      : :
5    -0.06139  :      :      :      :      :      :      :      :      :      : :
6     0.16114  :      :      :      :      :      :      :      :      :      : :
7     0.06282  :      :      :      :      :      :      :      :      :      : :
8     0.01640  :      :      :      :      :      :      :      :      :      : :
9     0.09375  :      :      :      :      :      :      :      :      :      : :
10   -0.04084  :      :      :      :      :      :      :      :      :      : :

Partial Autocorrelations
Lag  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1    0.42141  :      :      :      :      :      :      :      :      :      : :
2    0.02022  :      :      :      :      :      :      :      :      :      : :
3    0.27152  :      :      :      :      :      :      :      :      :      : :
4    0.66908  :      :      :      :      :      :      :      :      :      : :
5    -0.61920  :      :      :      :      :      :      :      :      :      : :
6    -0.38588  :      :      :      :      :      :      :      :      :      : :
7    -0.21644  :      :      :      :      :      :      :      :      :      : :
8    -0.10040  :      :      :      :      :      :      :      :      :      : :
9    -0.14334  :      :      :      :      :      :      :      :      :      : :
10   0.08322  :      :      :      :      :      :      :      :      :      : :

ARIMA Procedure
Autocorrelation Check for White Noise
Autocorrelations
To      Chi
Lag     Square  DF   Prob
6       41.03   6    0.000   0.421   0.194   0.312   0.733   0.177   -0.119

```

Figura 4.21 Análisis de autocorrelación para Outboard Marine.

Promedios móviles y métodos de atenuación

ACTIVIDAD

Tanto en organizaciones de compañías de negocios como de gobierno se ha puesto un gran énfasis en el mejoramiento de la toma de decisiones. Debido a que existimos en un medio influido por el tiempo, se ha hecho una meta común el asignar tiempo disponible junto con recursos competitivos que de alguna manera sea óptima. Esta meta se puede lograr, en parte, a través de pronósticos precisos. Hace veinte años los directivos tomaban su decisiones principalmente con base en su propio sentir e intuición sobre la economía y la industria. Ahora están complementando este “sentir” hacia la economía y la industria con técnicas de pronóstico ya sean éstas sencillas o complejas. El incremento en el uso de microcomputadoras continúa fortaleciendo esta tendencia.

Este capítulo describirá tres tipos de técnicas de pronóstico: no formales, de promedios y de atenuación. Los métodos *no formales* se emplean para desarrollar modelos sencillos que suponen que los periodos recientes son los mejores pronosticadores del futuro. Los métodos de *promedio* se desarrollan con base en un promedio de observaciones ponderadas. Los métodos de *atenuación* se basan en promedios de valores anteriores de una serie en una forma decreciente (exponencial).

La fig. 5.1 muestra un esquema del procedimiento de pronóstico para los métodos que se abordan en este capítulo. Visualícese en una escala de tiempo. Usted se encuentra en el punto t de la fig. 5.1 y puede mirar hacia atrás sobre observaciones pasadas de la variable de interés (Y_t) o hacia adelante en el futuro. Una vez seleccionada la técnica de pronóstico, ésta se ajusta a los datos conocidos, y se obtienen los valores de pronóstico ($\hat{Y}_t$). Ya que se tienen disponibles los valores de pronóstico, se pueden comparar con las observaciones conocidas y se puede calcular el error del pronóstico (e).

Una estrategia típica para la evaluación de métodos de pronóstico que se exponen en este capítulo, comprende los siguientes pasos:

1. Se elige un método de pronóstico con base en la intuición del analista con respecto al patrón de datos.
2. Se divide el conjunto de datos en dos secciones —una parte de inicialización y una parte de prueba.

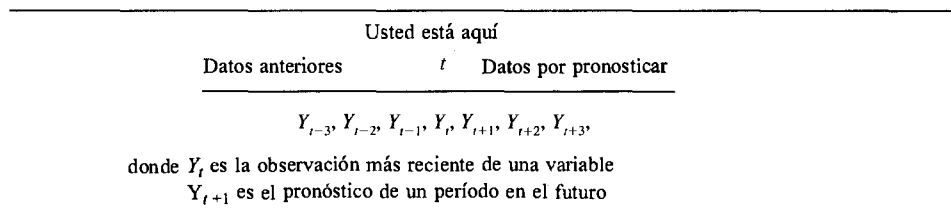


Figura 5.1 Esquema de pronóstico.

3. Se emplea la técnica de pronóstico elegida para desarrollar valores ajustados para la parte de inicialización de los datos.
4. Se utiliza la técnica para pronosticar la parte de prueba de los datos y se determina y evalúa el error de pronóstico (consulte el cap. 4 para un repaso de la medición de la precisión del pronóstico).
5. Se toma una decisión. La decisión pudiera ser la de emplear la técnica en su forma actual, modificarla, o desarrollar un pronóstico mediante otra técnica y comparar los resultados.

MODELOS NO FORMALES

Las técnicas no formales más sencillas suponen que los periodos recientes son los mejores pronosticadores del futuro. El modelo más sencillo es

$$\hat{Y}_{t+1} = Y_t \quad (5.1)$$

en donde Y_{t+1} es el pronóstico realizado en el periodo t para el periodo $t + 1$.

Ejemplo 5.1

La fig. 5.2 presenta las ventas de sierras por trimestre de la Acme Tool Company. Se emplea la técnica no formal para pronosticar que las ventas del próximo trimestre sean las mismas que las del trimestre anterior. La tabla 5.1 muestra datos de 1988 a 1994. Si se usan los datos de 1988 a 1993 como la parte de inicialización y 1994 como la parte de prueba, el pronóstico para el primer trimestre de 1994 es

$$\begin{aligned} \hat{Y}_{t+1} &= Y_t \\ y_{25} &= 650 \end{aligned}$$

El error de pronóstico se determina usando la ecuación 4.4. El error para el periodo 25 es

$$e_{25} = y_{25} - \hat{y}_{25} = 850 - 650 = 200$$

El pronóstico para el periodo 26 es 850 con un error de -250. La fig. 5.2 muestra que estos datos tienen una tendencia, así que se toma la decisión de modificar el modelo.

Una revisión de los datos del ejemplo 5.1 conduce al analista a concluir que se incrementan a través del tiempo. Cuando esto sucede, se dice que son *no estacionarios* o

TABLA 5.1 VENTAS DE SIERRAS DE LA
ACME TOOL COMPANY, 1988-1994

AÑO	TRIMESTRE	t	VENTAS
1988	1	1	500
	2	2	350
	3	3	250
	4	4	400
1989	1	5	450
	2	6	350
	3	7	200
	4	8	300
1990	1	9	350
	2	10	200
	3	11	150
	4	12	400
1991	1	13	550
	2	14	350
	3	15	250
	4	16	550
1992	1	17	550
	2	18	400
	3	19	350
	4	20	600
1993	1	21	750
	2	22	500
	3	23	400
	4	24	650
1994	1	25	850
	2	26	600
	3	27	450
	4	28	700

que tienen una *tendencia*. Si se usa la ecuación 5.1, las proyecciones son coherentemente bajas. La técnica puede adaptarse para tomar en cuenta esta tendencia agregando la diferencia entre éste y el último periodo. El modelo es

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \quad (5.2)$$

La ecuación 5.2 toma en cuenta la cantidad de cambio ocurrida entre trimestres. El pronóstico para el primer trimestre de 1994 es

$$\hat{y}_{24+1} = y_{24} + (y_{24} - y_{24-1})$$

$$\hat{y}_{25} = y_{24} + (y_{24} - y_{23})$$

$$\hat{y}_{25} = 650 + (650 - 400)$$

$$\hat{y}_{25} = 650 + 250$$

$$\hat{y}_{25} = 900$$

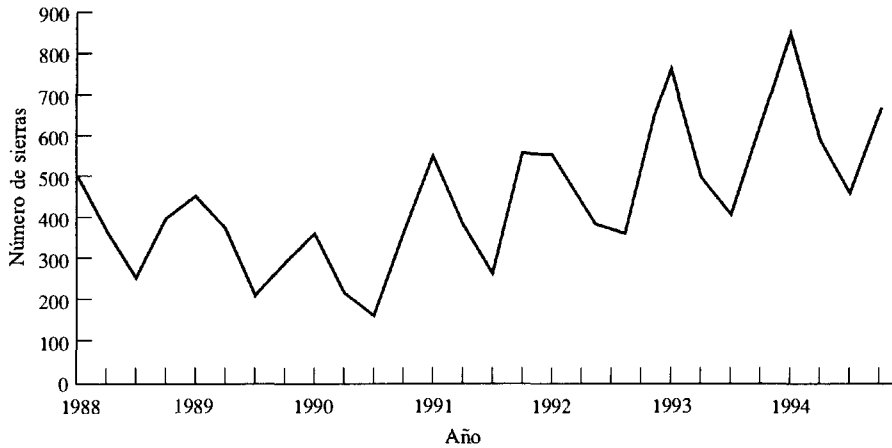


Figura 5.2 Ventas de sierras de la Acme Tool Co., 1988–1994.

Con este modelo, el error de pronóstico es

$$\begin{aligned}
 e_{25} &= y_{25} - \hat{y}_{25} \\
 e_{25} &= 850 - 900 \\
 e_{25} &= -50
 \end{aligned}$$

Para algunos fines, la proporción de cambio pudiera ser más apropiada que la cantidad absoluta de cambio, entonces el modelo sería

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} \quad (5.3)$$

El pronóstico para el primer trimestre de 1994, usando este modelo, es

$$\begin{aligned}
 \hat{y}_{24+1} &= y_{24} \cdot \frac{y_{24}}{y_{24-1}} \\
 \hat{y}_{25} &= y_{24} \cdot \frac{y_{24}}{y_{23}} \\
 \hat{y}_{25} &= 650 \cdot \frac{650}{400} \\
 \hat{y}_{25} &= 1,056
 \end{aligned}$$

Probablemente sería mejor si el pronosticador usara, al preparar su pronóstico, un *promedio* de los cambios absolutos anteriores o porcentajes de cambio.

La inspección visual de los datos de la tabla 5.1 indica que parece existir una variación estacional. Las ventas en el primer trimestre son por lo regular mayores que aquellas en cualquier otro trimestre. Si el patrón estacional es fuerte, entonces el modelo apropiado pudiera ser

$$\hat{Y}_{t+1} = Y_{t-3} \quad (5.4)$$

La ecuación 5.4 establece que en el siguiente trimestre, la variable tomará el mismo valor que tenía en el trimestre correspondiente del año anterior. El pronóstico para el primer trimestre de 1994 es

$$\begin{aligned} \hat{y}_{24+1} &= y_{24-3} \\ \hat{y}_{25} &= y_{21} \\ \hat{y}_{25} &= 750 \end{aligned}$$

La debilidad principal de este enfoque consiste en que ignora todo lo ocurrido desde el año pasado incluyendo cualquier patrón de tendencia. Existen varias formas de introducir información más reciente. Por ejemplo, el analista puede combinar enfoques y tomar en cuenta tanto las variaciones estacionales como las de tendencia. Un posible modelo es

$$\hat{Y}_{t+1} = Y_{t-3} + \frac{(Y_t - Y_{t-1}) + \cdots + (Y_{t-3} - Y_{t-4})}{4} \quad (5.5)$$

en donde el término y_{t-3} pronostica los patrones estacionales y el término que queda promedia la cantidad de cambio para los últimos cuatro trimestres (tendencia). El pronóstico para el primer trimestre de 1994, usando este modelo, sería

$$\begin{aligned} \hat{y}_{24+1} &= y_{24-3} + \frac{(y_{24} - y_{24-1}) + \cdots + (y_{24-3} - y_{24-4})}{4} \\ \hat{y}_{25} &= y_{21} + \frac{(y_{24} - y_{23}) + (y_{23} - y_{22}) + (y_{22} - y_{21}) + (y_{21} - y_{20})}{4} \\ \hat{y}_{25} &= 750 + \frac{(650 - 400) + (400 - 500) + (500 - 750) + (750 - 600)}{4} \\ \hat{y}_{25} &= 750 + 12.5 \\ \hat{y}_{25} &= 762.5 \end{aligned}$$

Resulta aparente que el número y complejidad de modelos no formales posibles, sólo está limitado por el ingenio del analista, aunque estas técnicas deben utilizarse de manera juiciosa.

MÉTODOS DE PROMEDIO

Con frecuencia, la administración enfrenta la situación donde se requiere actualizar los pronósticos diaria, semanal o mensualmente para inventarios que contienen cientos de miles de elementos. A menudo no es posible desarrollar técnicas sofisticadas de pronóstico para cada elemento. En vez de ello, para llevar a cabo esta tarea, se necesita alguna herramienta de pronóstico rápido, barato, muy sencillo y de corto plazo.

Es probable que un administrador que enfrente tal situación, emplee una técnica de promedio o de atenuación. Este tipo de técnicas utilizan una forma de promedio ponderado de observaciones anteriores para atenuar fluctuaciones de corto plazo. La suposición fundamental de estas técnicas es que las fluctuaciones en los valores anteriores representan puntos de partida aleatorios de alguna curva atenuada. Una vez que se identifica esta curva, se puede proyectar hacia el futuro para producir un pronóstico.

Promedios simples

Los datos históricos se pueden atenuar en muchas formas. El objetivo consiste en usar datos anteriores para desarrollar un modelo de pronóstico para periodos futuros. En esta sección, se considera el método de promedios simples. Al igual que en los métodos no formales, se toma una decisión para utilizar los primeros t puntos de datos como la parte de inicialización y el resto como la parte de prueba. A continuación, se usa la ecuación 5.6 para promediar (calcular la media de) la parte de inicialización de los datos y para pronosticar el siguiente periodo.

$$\hat{Y}_{t+1} = \sum_{i=1}^n \frac{Y_i}{n} \quad (5.6)$$

Por último, se determinan los errores de pronóstico y se toma una decisión con respecto a lo adecuado de la técnica de pronóstico.

Un *promedio simple* se obtiene encontrando la media de todos los valores pertinentes y usando después esta media para pronosticar el siguiente periodo.

Ejemplo 5.2

El método de promedios simples se demuestra usando las ventas de sierras por trimestre de Acme Tool Company que se presentan en la tabla 5.1. El pronóstico para el primer trimestre de 1994 es

$$\begin{aligned} \hat{y}_{24+1} &= \sum_{i=1}^n \frac{Y_i}{n} \\ \hat{y}_{25} &= \frac{9,800}{24} \\ \hat{y}_{25} &= 408.33 \end{aligned}$$

El error de pronóstico es

$$e_{25} = y_{25} - \hat{y}_{25}$$

$$e_{25} = 850 - 408.33$$

$$e_{25} = 441.67$$

El pronóstico para el segundo periodo de 1994 incluye un punto más de datos (850) que se incorpora a la parte de inicialización de la historia anterior. El pronóstico es

$$\hat{y}_{25+1} = \frac{10,650}{25}$$

$$\hat{y}_{26} = 426$$

El método simple no parece apropiado para estos datos. Este método debería emplearse cuando los datos son estacionarios; sin tendencia, estacionalidad u otros patrones sistemáticos.

Promedios móviles

El método de promedio simple usa, para pronosticar, la media de todos los datos. Pero, ¿qué sucede si el analista está más interesado en las observaciones recientes? Se puede especificar como conjunto un número de puntos de datos y calcular la media para las observaciones más recientes. Para describir este enfoque, se emplea el término *promedio móvil*. Al estar disponible cada nueva observación, se puede calcular una nueva media eliminando el valor más antiguo e incluyendo el más reciente. Entonces, se usa este promedio móvil para pronosticar el siguiente periodo. La ecuación 5.7 establece el modelo simple de promedio móvil.

$$M_t = \hat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \cdots + Y_{t-n+1})}{n} \quad (5.7)$$

en donde

M_t = promedio móvil en el periodo t

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

Y_t = valor real en el periodo t

n = número de términos en el promedio móvil

El promedio móvil para el periodo t es la media aritmética de las n observaciones más recientes.

Nótese que se asignan ponderaciones iguales a cada observación. Al hacerse disponible, cada nuevo punto de datos se incluye en el promedio y se descarta el más antiguo. La proporción de respuesta a los cambios en el patrón subyacente de los datos depende del número de periodos, n , que se incluyen en el promedio móvil.

Nótese que la técnica de promedio móvil se refiere sólo a los últimos n periodos de datos conocidos; el número de puntos de datos en cada promedio no cambia al correr del

tiempo. El modelo de promedio móvil funciona mejor con datos estacionarios. No maneja muy bien la tendencia o la estacionalidad, aunque lo hace mejor que el método de promedio simple.

En un promedio móvil, el analista debe escoger el número de periodos n . Un promedio móvil de orden 1 tomaría la última observación y_t y la usaría para pronosticar el siguiente periodo. Esto es simplemente el enfoque de pronóstico no formal de la ecuación 5.1. Para datos trimestrales, los datos en un promedio móvil de 4 trimestres arrojan un promedio de los 4 trimestres; y para datos mensuales, un promedio móvil de 12 meses elimina o promedia los efectos estacionales. Es frecuente que se utilicen los promedios móviles con datos trimestrales o mensuales para auxiliar en el examen de los componentes en una serie de tiempo, como se muestra en el cap. 8. Entre mayor sea el orden del promedio móvil, mayor será el efecto de atenuación. Empleado como un pronóstico, un promedio móvil grande presta poca atención a las fluctuaciones en la serie de datos.

Un *promedio móvil* se obtiene encontrando la media de un conjunto específico de valores y empleándolo después para pronosticar el siguiente periodo.

Ejemplo 5.3

La tabla 5.2 demuestra la técnica de pronóstico de promedio móvil con datos de Acme Tool Company. El pronóstico de promedio móvil para el primer trimestre de 1993 es

$$M_{20} = \hat{y}_{20+1} = \frac{y_{20} + y_{20-1} + y_{20-2} + y_{20-4+1}}{4}$$

$$\hat{y}_{21} = \frac{y_{20} + y_{19} + y_{18} + y_{17}}{4}$$

$$\hat{y}_{21} = \frac{550 + 400 + 350 + 600}{4} = \frac{1,900}{4} = 475$$

Cuando se conoce al valor del primer trimestre de 1993, se puede calcular el error de pronóstico:

$$e_{21} = y_{21} - \hat{y}_{21} = 750 - 475 = 275$$

El pronóstico para el primer trimestre de 1995 es

$$M_{28} = \hat{y}_{28+1} = \frac{y_{28} + y_{28-1} + y_{28-2} + y_{28-4+1}}{4}$$

$$\hat{y}_{29} = \frac{y_{28} + y_{27} + y_{26} + y_{25}}{4}$$

$$\hat{y}_{29} = \frac{700 + 450 + 600 + 850}{4} = \frac{2,600}{4} = 650$$

El analista debe emplear su juicio para determinar el número de semanas, meses o trimestres en los que basará el promedio móvil. Entre menor sea el número, se dará más

TABLA 5.2 PROMEDIO MÓVIL TRIMESTRAL DE ACME TOOL COMPANY

AÑO	TRIMESTRE	t	VENTAS Y_t	TOTAL MÓVIL	PRONÓSTICO DE PROMEDIO MÓVIL, $\hat{Y}_{t+1}$	e_t
1992	1	17	550			
	2	18	400			
	3	19	350			
	4	20	600	1,900		
1993	1	21	750	2,100	475	+275
	2	22	500	2,200	525	-25
	3	23	400	2,250	550	-150
	4	24	650	2,300	562.5	+87.5
1994	1	25	850	2,400	575	+275
	2	26	600	2,500	600	0
	3	27	450	2,550	625	-175
	4	28	700	2,600	637.5	+62.5
1995	1	29			650	

peso a los periodos recientes. Inversamente, entre más grande sea el número, se dará menor peso a los periodos más recientes. Es más deseable un número pequeño cuando hay cambios súbitos en el nivel de la serie. Un número pequeño asigna mayor peso a la historia reciente, lo que permite que el pronóstico se ajuste con más rapidez al nivel actual. Cuando existan fluctuaciones amplias no frecuentes en la serie, es deseable un número mayor.

Nótese que una suposición del promedio móvil es que los datos son estacionarios. El ejemplo 5.4 muestra lo que sucede cuando se emplea la técnica del promedio móvil en datos con tendencia. El examen de la columna de error de la tabla 5.3 muestra que todos los valores son positivos, lo que significa que el pronóstico no se ajustó a la tendencia. En la siguiente sección se presenta la técnica de promedio móvil doble, la cual está diseñada para manejar datos con tendencia.

Promedio móvil doble

Una forma de pronosticar series de tiempo que tienen una tendencia lineal, consiste en utilizar la técnica del promedio móvil doble. El método hace lo que su nombre implica: se calcula un conjunto de promedios móviles y después se calcula un segundo conjunto como promedio móvil del primero.

La fig. 5.3 muestra las rentas por semana de Movie Video Store. Además, se muestran en la gráfica el promedio móvil de tres semanas y el promedio móvil doble para estos datos. Nótese cómo el promedio móvil se ubica por debajo de los valores reales en periodos comparables. Esto ilustra lo que sucede cuando se usa la técnica de promedio móvil en

TABLA 5.3 RENTAS SEMANALES DE MOVIE VIDEO STORE PARA EL EJEMPLO 5.4

t	UNIDADES RENTADAS POR SEMANA	TOTAL MÓVIL TRES SEMANAS	PRONÓSTICO DE PROMEDIO MÓVIL $\hat{Y}_{t+1}$	e_t
1	654	—	—	
2	658	—	—	
3	665	1977	—	
4	672	1995	659	13
5	673	2010	665	8
6	671	2016	670	1
7	693	2037	672	21
8	694	2058	679	15
9	701	2088	686	15
10	703	2098	696	7
11	702	2106	699	3
12	710	2115	702	8
13	712	2124	705	7
14	711	2133	708	3
15	728	2151	711	17
16	—	—	717	

EMC = 133.

datos con tendencia. Nótese también que el promedio móvil doble se ubica por debajo del primer conjunto en proporción similar a la que se ubica el primer conjunto por debajo de los valores reales. Esta diferencia entre ambos conjuntos se emplea para pronosticar los valores reales.

La técnica de promedio móvil doble se resume mediante las ecuaciones 5.8 a 5.11.

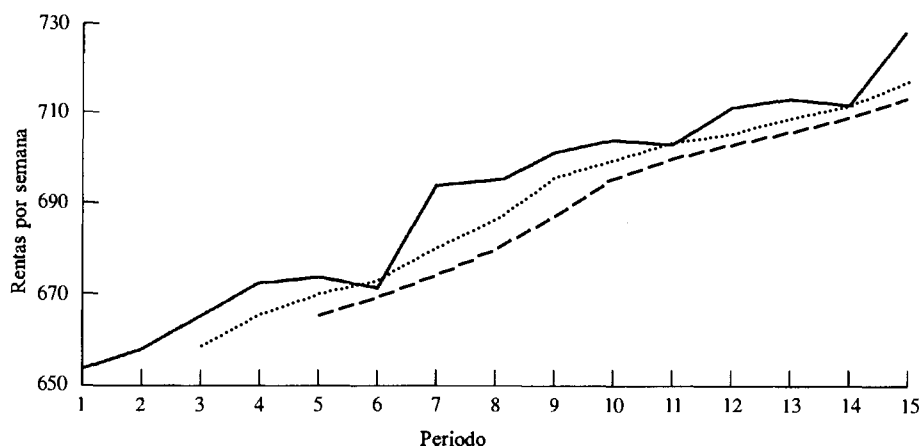


Figura 5.3 Promedios móviles dobles: Movie Video Store.

Primero, se utiliza la ecuación 5.7 para calcular el promedio móvil.

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \cdots + Y_{t-n+1}}{n}$$

Dado que $M_t = \hat{y}_{t+1}$. Entonces se usa la ecuación 5.8 para calcular el segundo promedio móvil.

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \cdots + M_{t-n+1}}{n} \quad (5.8)$$

La ecuación 5.9 se usa para calcular la diferencia entre ambos promedios móviles.

$$a_t = 2M_t - M'_t \quad (5.9)$$

La ecuación 5.10 es un factor de ajuste adicional, similar a la medición de una pendiente que cambia a través de la serie.

$$b_t = \frac{2}{n-1} (M_t - M'_t) \quad (5.10)$$

Por último, se utiliza la ecuación 5.11 para realizar el pronóstico de periodos m en el futuro.

$$\hat{Y}_{t+p} = a_t + b_t p \quad (5.11)$$

en donde

n = número de periodos en el promedio móvil

Y_t = valor real de la serie en el periodo t

p = número de periodos a pronosticar hacia el futuro

Ejemplo 5.4

La empresa Movie Video Store funciona con varios distribuidores de renta de videocintas en Denver, Colorado. La compañía está creciendo y requiere ampliar su inventario para dar cabida a la creciente demanda de servicios. El presidente de la compañía asigna a Jill Ottenbreit para que pronostique las rentas del próximo mes. Los datos de renta disponibles de las últimas 15 semanas se presentan en la tabla 5.4. Como primer paso, Jill intenta desarrollar un pronóstico utilizando un promedio móvil de tres semanas. El EMC (error medio cuadrado) para este modelo es 133. Puesto que los datos obviamente tienen tendencia, Jill encuentra que sus pronósticos están consistentemente subestimados de las rentas reales. Por esta razón, decide intentar con el promedio móvil doble. Los resultados se presentan en la tabla 5.4. A continuación se presentan los cálculos para comprender el pronóstico para la semana 16. Se usa la ecuación 5.7 para calcular el promedio móvil de tres semanas (columna 3).

$$\hat{y}_{15+1} = \frac{y_{15} + y_{15-1} + \cdots + y_{15-3+1}}{3}$$

$$\hat{y}_{16} = \frac{728 + 711 + 712}{3} = 717$$

TABLA 5.4 PRONÓSTICO DE PROMEDIO MÓVIL DOBLE PARA
MOVIE VIDEO STORE: EJEMPLO 5.4

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PERIODO	VENTAS POR SEMANA	PROMEDIO MÓVIL TRES SEMANAS	PROMEDIO MÓVIL DOBLE	VALOR DE a	VALOR DE b	PRONÓSTICO $a + bp$ ($p = 1$)	e_t
t	Y_t	M_t	M'_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	

EMC = 63.7.

Dado $M_t = 717$. Entonces, se emplea la ecuación 5.8 para calcular el segundo promedio móvil de tres semanas (columna 4).

$$M'_{15} = \frac{M_{15} + M_{15-1} + \cdots + M_{15-3+1}}{3}$$

$$= \frac{717 + 711 + 708}{3} = 712$$

Se usa la ecuación 5.9 para calcular la diferencia entre ambos promedios móviles (columna 5).

$$a_{15} = 2M_{15} - M'_{15} = 2(717) - 712 = 722$$

La ecuación 5.10 ajusta la pendiente (columna 6).

$$b_{15} = \frac{2}{3-1} (M_{15} - M'_{15}) = \frac{2}{2} (717 - 712) = 5$$

Se utiliza la ecuación 5.11 para efectuar el pronóstico de un periodo a futuro (columna 7).

$$\hat{y}_{15+1} = a_{15} + b_{15}(1) = 722 + 5 = 727$$

El promedio de cuatro semanas a futuro es

$$\hat{y}_{15+4} = a_{15} + b_{15}(4) = 722 + 5(4) = 742$$

Nótese que el EMC se redujo de 133 a 63.7.

Parece razonable que las observaciones más recientes son apropiadas para contener la información más importante. En la siguiente sección se introduce una técnica que da mayor énfasis a las observaciones más recientes

MÉTODOS DE ATENUACIÓN EXPONENCIAL

La atenuación exponencial es un método utilizado para revisar constantemente una estimación a la luz de experiencias más recientes. El método está basado en el promedio (atenuación) de valores anteriores de una serie, haciendo esto de forma decreciente (exponencial). Las observaciones se ponderan, asignando mayor peso a las más recientes. Las ponderaciones empleadas se designan como α para la observación más reciente, $\alpha(1 - \alpha)$ para la siguiente más reciente, $\alpha(1 - \alpha)^2$ para la siguiente y así sucesivamente.

En una forma atenuada, se podría pensar en un nuevo pronóstico (para el periodo $t + 1$), como el promedio ponderado de la nueva observación (en el periodo t) y el promedio anterior (para el periodo t). Se asigna la ponderación α al nuevo valor observado y $1 - \alpha$ al pronóstico anterior, suponiendo que $0 < \alpha < 1$. De esta forma,

$$\text{nuevo pronóstico} = \alpha \times (\text{nueva observación}) + (1 - \alpha) \times (\text{pronóstico anterior})$$

De manera más formal, la ecuación de atenuación exponencial es

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (5.12)$$

en donde

$\hat{Y}_{t+1}$ = nuevo valor atenuado o valor de pronóstico para el siguiente periodo

α = constante de atenuación ($0 < \alpha < 1$)

Y_t = nueva observación o valor real de la serie en el periodo t

$\hat{Y}_t$ = valor atenuado anterior o experiencia promedio de la serie atenuada al periodo $t - 1$

En un esfuerzo por interpretar mejor a alfa, la ecuación 5.12 se escribe como

$$\begin{aligned} \hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha) \hat{Y}_t \\ &= \alpha Y_t + \hat{Y}_t - \alpha \hat{Y}_t \\ &= \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) \end{aligned}$$

La atenuación exponencial es simplemente el pronóstico anterior ($\hat{Y}_t$) más α veces el error ($Y_t - \hat{Y}_t$), en el pronóstico anterior.

La *atenuación exponencial* es un procedimiento para revisar constantemente un pronóstico a la luz de la experiencia más reciente.

La constante de atenuación α sirve como el factor para ponderar. El valor real de α determina el grado hasta el cual la observación más reciente puede influir en el valor del pronóstico. Cuando α es cercana a 1, el nuevo pronóstico incluirá un ajuste sustancial de cualquier error ocurrido en el pronóstico anterior. Inversamente, cuando α está cercana a 0, el pronóstico es similar al anterior. Piense en $\hat{Y}_t$ como un promedio ponderado de todas las observaciones anteriores con ponderaciones que disminuyen de manera exponencial al ir tomando los datos hacia atrás (en el tiempo). Como se muestra en la tabla 5.5, la velocidad con que los valores anteriores pierden su importancia, depende de α .

La clave del análisis es el valor de α . Si se desea que los pronósticos sean estables y se atenúen las variaciones aleatorias, se requiere de un valor de α pequeño. Si se desea una respuesta rápida al cambio real en el patrón de observaciones, resulta más apropiado un mayor valor de α . Un método para estimar α consiste en un procedimiento iterativo que minimiza el error medio cuadrado (EMC) calculado en la ecuación 4.6. Se calculan los pronósticos para α igual a .1, .2, . . . , .9, y para cada uno se calcula la suma de los errores cuadrados del pronóstico. Para generar pronósticos futuros, se elige el valor de α que produce el error más pequeño.

TABLA 5.5. COMPARACIÓN DE CONSTANTES DE ATENUACIÓN

PERIODO	$\alpha = .1$		$\alpha = .6$	
	CÁLCULOS	PONDERACIÓN	CÁLCULOS	PONDERACIÓN
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
Los restantes		.590		.011
		Total 1.000		1.000

Ejemplo 5.5

La técnica de atenuación exponencial se muestra en tabla 5.6 para los datos de Acme Tool Company de 1988 a 1993, utilizando constantes de atenuación de .1 y .6. Los datos de 1994 se emplearán como la parte de prueba para ayudar a evaluar qué modelo se desempeña mejor. La serie atenuada en forma exponencial se calcula asignando inicialmente $\hat{y}_1$ igual a 500. Si hay datos anteriores disponibles, se podrían usar para desarrollar una serie atenuada hasta 1988

TABLA 5.6 VALORES ATENUADOS EXPONENCIALMENTE PARA EL EJEMPLO 5.5, VENTAS DE ACME TOOL COMPANY

PERIODO	VALOR REAL, Y_t	VALOR ATENUADO, $\hat{Y}_t(\alpha = .1)$	ERROR DE PRONÓSTICO, e_t	VALOR ATENUADO, $\hat{Y}_t(\alpha = .6)$	ERROR DE PRONÓSTICO, e_t
1	500	500	—	—	—
2	350	500	-150	500	-150
3	250	485 (1)	-235 (2)	410	-160
4	400	462 (3)	-62	314	+86
5	450	455	-5	366	+84
6	350	455	-105	416	-66
7	200	444	-244	376	-176
8	300	420	-120	270	+30
9	350	408	-58	288	+62
10	200	402	-202	325	-125
11	150	382	-232	250	-100
12	400	359	+41	190	+210
13	550	363	+187	316	+234
14	350	382	-32	456	-106
15	250	378	-128	392	-142
16	550	366	+184	307	+243
17	550	384	+166	453	+97
18	400	401	-1	511	-121
19	350	401	-51	444	-96
20	600	395	+205	388	+212
21	750	416	+334	515	+235
22	500	449	+51	656	-156
23	400	454	-54	562	-12
24	650	449	+201	465	+185
25		469 (4)		576 (4)	

Los números entre paréntesis se refieren a las explicaciones dadas en el texto

y utilizar esta experiencia como el valor inicial de la serie atenuada. A continuación se muestran los cálculos que conducen al pronóstico del periodo 3.¹

1.

$$\begin{aligned}\hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha)\hat{y}_t \\ \hat{y}_{2+1} &= \alpha Y_2 + (1 - \alpha)\hat{y}_2 \\ \hat{y}_3 &= .1(350) + (1 - .1)500 = 485\end{aligned}$$

2. El error para este pronóstico es

$$e_3 = 250 - 485 = -235$$

3. El pronóstico para el periodo 4 es

$$\hat{y}_4 = .1(250) + .9(485) = 462$$

¹ Los números de guía 1a 4 se muestran entre paréntesis en la tabla 5.6.

4. Los pronósticos de ventas para el primer trimestre de 1994, utilizando constantes de atenuación de .1 y .6 son 469 y 576, respectivamente.

Nótese lo estable que son los valores atenuados para la constante de atenuación de .1. Por otra parte, con base en la minimización del error medio cuadrado (EMC), resulta mejor la constante de atenuación de .6, que también es mejor si se compara la media del porcentaje de error absoluto (PEMA). Al comparar el pronóstico para cada constante de atenuación con las ventas reales del primer trimestre de 1994, también parece ser mejor la constante de atenuación de .6.

$\alpha = .1$	EMC = 25,317	PEMA = 40.61%	PME = -20.81%
$\alpha = .6$	EMC = 23,216	PEMA = 38.14%	PME = -10.33%
$\alpha = .34$	EMC = 21,421	PEMA = 35.41%	PME = -11.50%

Cuando se emplea un programa de computadora que elige en forma automática la constante de atenuación, α es igual a .34. El EMC se reduce a 21 421 y el PEMA equivale a 35.41%. El porcentaje medio de error (PME), a veces denominado *sesgo*, es de -11.50% para este modelo, lo que significa que los pronósticos basados en este modelo son consistentemente muy grandes. Se necesita investigar la posibilidad de una tendencia o variación estacional en los datos.

Otro factor que afecta el valor de pronósticos subsecuentes es la elección del valor inicial de $\hat{Y}_t$. La tabla 5.6 muestra que se usó Y_t como valor inicial de $\hat{Y}_t$, en el ejemplo 5.5.

$$\begin{aligned}\hat{Y}_t &= Y_t \\ \hat{y}_1 &= 500\end{aligned}$$

Esta elección tiende a dar mucho peso a Y_t en periodos posteriores. Por fortuna, la influencia del pronóstico inicial disminuye en gran medida al incrementarse t .

Otro enfoque para inicializar Y_t consiste en promediar las primeras n observaciones. La columna de atenuación comenzaría entonces con

$$\hat{Y}_t = \bar{Y} = \frac{\sum_{t=1}^n Y_t}{n}$$

Si se elige n igual a 4, entonces el valor inicial para el ejemplo 5.5 es

$$\hat{Y}_t = \bar{Y} = \frac{\sum_{t=1}^4 Y_t}{n} = \frac{500 + 350 + 250 + 400}{4} = 375$$

Rastreo

Debido a que la atenuación exponencial supone la continuidad en el futuro de algún patrón histórico, resulta útil desarrollar una medida que se pueda utilizar para determinar cuando

se modifica el patrón básico. La medida más común es la *señal de rastreo*. Esta señal comprende el cálculo de alguna medición del error a través del tiempo y establece límites, de modo que cuando el error rebase dichos límites, se alerte al pronosticador.

Una *señal de rastreo* comprende el cálculo de alguna medición del error a través del tiempo y establece límites, de modo que cuando el error rebase dichos límites, se alerte al pronosticador.

Por ejemplo, se puede emplear una señal de rastreo para determinar cuándo debe cambiarse el tamaño de alfa utilizado en una técnica de atenuación exponencial. Ya que por lo regular se pronostican un gran número de elementos, una práctica común consiste en continuar con el mismo tamaño de alfa durante varios periodos, antes de determinar si es necesaria una revisión. Por desgracia, la simplicidad de utilizar un modelo establecido de atenuación exponencial es un fuerte motivador para no efectuar un cambio. Aunque en algún punto es necesario actualizar alfa. Cuando un modelo produce pronósticos que contienen un amplio margen de error, un cambio resulta apropiado.

Un sistema de rastreo proporciona un método para monitorear la necesidad de cambio. Dicho sistema contiene un nivel de variaciones permisibles entre el pronóstico y los valores reales. Mientras el pronóstico se ubique dentro del nivel, no es necesario modificar a alfa. Sin embargo, si el pronóstico se sale del nivel, el sistema señala la necesidad de actualizar a alfa.

Por ejemplo, si las cosas van bien, la técnica de pronóstico debería subestimar y sobrestimar con la misma frecuencia. Se puede desarrollar una señal de rastreo con base en este razonamiento.

Sea U el número de subestimaciones en los últimos n pronósticos. En otras palabras, U es el número de errores de entre los últimos n que son positivos. Si el proceso está bajo control, el valor esperado de U sería $n/2$; pero se contempla una variabilidad de la muestra, por lo que no serían raros los valores cercanos a $n/2$. Por otro lado, los valores que no fueran cercanos a $n/2$ indicaría que la técnica está produciendo pronósticos con sesgo.

Ejemplo 5.5

Suponga que Acme Tool Company decidió usar la técnica de atenuación exponencial con alfa igual a .1, como se muestra en el ejemplo 5.4. Si el proceso está bajo control y el analista decide monitorear los 10 últimos valores de error, U tiene un valor esperado de 5. De hecho, los valores de 2, 3, 4, 6, 7 u 8 no serían alarmantes. Sin embargo, valores de 0, 1, 9 o 10 serían de preocupar, ya que la probabilidad de obtener dichos valores por casualidad sería de .024 (distribución binomial).

Si $2 \leq U \leq 8$, entonces el procedimiento está bajo control

Si $U < 2$ o $U > 8$, entonces el procedimiento está fuera de control

Suponga que, de los 10 pronósticos siguientes mediante esta técnica, sólo uno tiene un error positivo. Debido a que la probabilidad de obtener sólo un error positivo entre 10 es muy baja, .011, se considera que el proceso está fuera de control (sobrestimando) y se debería modificar el valor de alfa.

Otra forma de rastrear una técnica de diagnóstico consiste en determinar un nivel que contuviera todos los errores de pronóstico. Esto se puede lograr utilizando el EMC que se estableció al determinar el tamaño óptimo de alfa. Si la técnica de pronóstico es razonablemente precisa, el error de pronóstico debería estar en una distribución normal con una media de cero. Bajo esta condición, existe una probabilidad del 95% de que una observación real se ubique dentro de aproximadamente dos desviaciones estándar del pronóstico. En general, es suficiente un intervalo de confianza del 95% para los sistemas de rastreo. El ejemplo 5.6 ilustra este enfoque.

**Ejemplo
5.6**

En el ejemplo de Acme Tool Company, se determinó el valor óptimo de alfa como $\alpha = .34$ con un $EMC = 21,421$. La desviación estándar del error de pronóstico es $\sqrt{EMC}$ ó $\sqrt{21,421} = 146.4$. El error de pronóstico debería presentar una distribución normal con una media de cero. Bajo esta condición hay un 95% de posibilidad de que la observación real se ubique dentro de aproximadamente dos desviaciones estándar del pronóstico o

$$\pm 1.96 \sqrt{EMC} = \pm 1.96 \sqrt{21,421} = \pm 1.96(146.4) = \pm 286.9$$

De ahí que para el ejemplo, la variación permisible sea de 286.9. Si en cualquier pronóstico posterior el error es mayor a 286.9, existen motivos para pensar que se debe calcular un nuevo valor óptimo para alfa.

La exposición anterior sobre señales de rastreo se aplica también a los métodos aún por abordar en el resto del capítulo.

Nótese que en la técnica de atenuación exponencial simple existe la suposición de que los datos son estacionarios. El ejemplo 5.7 muestra lo que sucede cuando se emplea la técnica de atenuación exponencial en datos con tendencia. Siempre que exista una tendencia significativa, la atenuación exponencial se ubicará por debajo de los valores reales de pronóstico de la serie de tiempo, a través del tiempo. El examen de la columna de error en la tabla 5.7 muestra que todos los valores son positivos, lo que significa que el pronóstico no contempló la tendencia. A continuación se presenta la técnica de atenuación exponencial doble, la cual se diseñó para manejar datos con tendencia.

Atenuación exponencial doble

La técnica de atenuación exponencial doble, también conocida como *Método de Brown*, se usa para pronosticar series de tiempo que tienen una tendencia lineal. Los conceptos básicos son similares a los de los promedios móviles dobles. La técnica de atenuación exponencial doble se resume mediante el uso de las ecuaciones 5.13 a 5.17.

TABLA 5.7 RENTAS POR SEMANA DE MOVIE VIDEO STORE PARA EL EJEMPLO 5.7

t	UNIDADES RENTADAS POR SEMANA		
	Y_t	Y_{t+1}	e_t
1	654	654	—
2	658	654	4.0
3	665	655.6	9.4
4	672	659.4	12.6
5	673	664.4	8.6
6	671	667.8	3.2
7	693	669.1	23.9
8	694	678.7	15.3
9	701	684.8	16.2
10	703	691.3	11.7
11	702	696.0	6.0
12	710	698.4	11.6
13	712	703.0	9.0
14	711	706.6	4.4
15	728	708.4	19.6
16	—	716.2	

EMC = 147.3.

Debido a que los valores de las series no son pronósticos en sí mismos, las ecuaciones de actualización son más comprensibles si se adopta la siguiente notación.

A_t = valor atenuado exponencialmente de Y_t en el periodo t

A'_t = valor doblemente atenuado exponencialmente de Y_t en el periodo t

El valor simple atenuado exponencialmente se calcula ahora mediante la ecuación 5.13.

$$A_t = \alpha Y_t + (1 - \alpha)A_{t-1} \quad (5.13)$$

La ecuación 5.14 se usa para calcular el valor doblemente atenuado exponencialmente.

$$A'_t = \alpha A_t + (1 - \alpha)A'_{t-1} \quad (5.14)$$

La ecuación 5.15 se emplea para calcular la diferencia entre los valores atenuados exponencialmente.

$$a_t = 2A_t - A'_t \quad (5.15)$$

La ecuación 5.16 es un factor adicional de ajuste, similar a la medición de una pendiente que puede cambiar durante la serie.

$$b_t = \frac{\alpha}{1 - \alpha} (A_t - A'_t) \quad (5.16)$$

Por último, la ecuación 5.17 se usa para formular el pronóstico de p periodos en el futuro.

$$\hat{Y}_{t+p} = a_t + b_t p. \quad (5.17)$$

Ejemplo 5.7

Este ejemplo se refiere al ejemplo 5.4 en el que la empresa Movie Video Store pretendía pronosticar sus rentas del próximo mes. Los datos de renta de las últimas 15 semanas se presentan en la tabla 5.4. Jill intentó desarrollar un pronóstico mediante un promedio móvil de tres semanas y un promedio móvil doble. Decide emplear una atenuación exponencial de 0.4. Ahora tiene los resultados, mostrados en la tabla 5.7. Ya que es obvio que los datos tienen una tendencia, Jill se encuentra con que su pronóstico está siempre subestimando las rentas reales. Por esta razón, decide intentar la atenuación exponencial doble. El programa de computadora que utiliza selecciona automáticamente la mejor constante de atenuación minimizando el EMC. Los resultados se presentan en la tabla 5.8 para un valor de alfa de .4. Para poder comprender el pronóstico de la semana 16, se presentan a continuación los cálculos correspondientes.

Se emplea la ecuación 5.13 para calcular el valor simple atenuado exponencialmente (columna 3).

$$\begin{aligned} A_{15} &= \alpha y_{15} + (1 - \alpha)A_{15-1} \\ &= .4(728) + (1 - .4)(708.4) = 716.2 \end{aligned}$$

TABLA 5.8 PRONÓSTICO DE ATENUACIÓN EXPONENCIAL DOBLE PARA MOVIE VIDEO STORE, TÉCNICA DE BROWN, PARA EL EJEMPLO 5.7

(1) PERIODO t	(2) VENTAS POR SEMANA Y_t	(3) A_t	(4) A'_t	(5) VALOR DE a	(6) VALOR DE b	(7) PRONÓSTICO $a + bp$ ($p = 1$)	(8) e_t
1	654	654	654	654	0	654	—
2	658	655.6	654.6	656.6	.7	654	4.0
3	665	659.4	656.5	662.3	1.9	657.3	7.7
4	672	664.4	659.7	669.1	3.1	664.2	7.8
5	673	667.8	662.9	672.7	3.3	672.2	0.8
6	671	669.1	665.4	672.8	2.5	676.0	-5.0
7	693	678.7	670.7	686.7	5.3	675.3	17.7
8	694	684.8	676.3	693.3	5.7	692.0	2.0
9	701	691.3	682.3	700.3	6.0	699.0	2.0
10	703	696.0	687.8	704.2	5.5	706.3	-3.3
11	702	698.4	692.0	704.8	4.3	709.7	-7.7
12	710	703.0	696.4	709.6	4.4	709.1	0.9
13	712	706.6	700.5	712.7	4.1	714.0	-2.0
14	711	708.4	703.7	713.1	3.1	716.8	-5.8
15	728	716.2	708.7	723.7	5.0	716.2	11.8
16	—	—	—	—	—	728.7	—

EMC = 39.82.

La ecuación 5.14 se utiliza para calcular el valor doblemente atenuado exponencialmente (columna 4).

$$\begin{aligned} A'_{15} &= A_{15} + (1 - \alpha)A'_{15-1} \\ &= .4(716.2) + (1 - .4)(703.7) = 708.7 \end{aligned}$$

La ecuación 5.15 se utiliza para calcular las diferencias entre los valores atenuados exponencialmente (columna 5).

$$\begin{aligned} a_{15} &= 2A_{15} - A'_{15} \\ &= 2(716.2) - 708.7 = 723.7 \end{aligned}$$

La ecuación 5.16 es un valor adicional de ajuste, similar a la medición de una pendiente que cambia a través de la serie (columna 6).

$$\begin{aligned} b_{15} &= \frac{\alpha}{1 - \alpha} (A_{15} - A'_{15}) \\ &= \frac{.4}{.6} (716.2 - 708.7) = 5.0 \end{aligned}$$

Finalmente se emplea la ecuación 5.17 para hacer el pronóstico de un periodo en el futuro (columna 7).

$$\hat{y}_{15+1} = a_{15} + b_{15}(1) = 723.7 + 5.0(1) = 728.7$$

El pronóstico de cuatro semanas a futuro es

$$\hat{y}_{15+4} = a_{15} + b_{15}(4) = 723.7 + 5(4) = 743.7$$

Jill piensa que encontró una buena técnica de pronóstico. El EMC se redujo a 39.82. Como verificación final, decide calcular los coeficientes de autocorrelación de los residuales y probar para determinar si alguno de ellos es significativamente diferente de cero. Las hipótesis nula y alterna para probar si el coeficiente de autocorrelación para un periodo de desfaseamiento en particular es diferente de cero son

$$H_0: \rho_k = 0$$

$$H_1: \rho_k \neq 0$$

Como $n = 14$, el error estándar (desviación estándar de la distribución muestral de los coeficientes de autocorrelación), es $1/\sqrt{14} = .267$. Si se prueba la hipótesis nula al nivel de significancia de .05, el valor correcto estándar normal de z es 1.96 y el valor crítico es $1.96(.267) = .524$. La regla de decisión es

Si un coeficiente de autocorrelación es menor que $-.524$ o mayor que $.524$, rechazar la hipótesis nula; en caso contrario, no rechazarla (rechazar H_0 si $r_k < -.524$ o $r_k > .524$).

Los comandos de Minitab para ejecutar las autocorrelaciones para los residuales de la tabla 5.8 son

```

MTB > SET C1
DATA > 4.0 7.7 7.8 .8 -5 17.7 2 2 -3.3 -7.7 .9 -2 -5.8 11.8
DATA > END
MTB > ACF C1

ACF of C1

      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +-----+-----+-----+-----+-----+-----+
1 -0.062                                XXX
2 -0.059                               XX
3  0.080                                XXX
4 -0.139                               XXXX
5  0.051                                XX
6 -0.117                               XXXX
7  -0      XXXXXXXX
8  0.212                                XXXXXX
9 -0.160                                XXXXX
10 -0.126                               XXXX
11  0.003                                X
12  0.058                                XX

MTB > STOP

```

Se verificaron 12 periodos de desfaseamiento y los coeficientes de autocorrelación se ubican dentro de los límites apropiados. Jill concluye que esta serie de residuales es aleatoria.

Al igual que en la atenuación exponencial simple, la elección de alfa es un problema. Alfa se selecciona minimizando el EMC entre los valores reales y los pronósticos de estos valores utilizando el modelo. Algunos programas de computadora determinan este valor en forma automática. En caso contrario, éste se convierte en un proceso de prueba y error.

Un problema al emplear la atenuación exponencial doble es la determinación de los valores iniciales de la serie atenuada y el ajuste de la tendencia. En el ejemplo 5.7 se usó el valor real de la serie para A_t y A'_t . El problema consiste en que este procedimiento supone que no hay tendencia presente y dará por resultado un error. Tenderá a subestimar una tendencia positiva.

Se pueden calcular los valores de la pendiente de la tendencia y la intersección mediante el procedimiento de mínimos cuadrados que se expone en el cap. 6. Las estimaciones generadas para el periodo $t = 0$ son $a_0 = 650.3$ y $b_0 = 4.9$. Los valores iniciales se establecen mediante las ecuaciones 5.18 y 5.19

$$A_0 = a_0 - \frac{1 - \alpha}{\alpha} b_0 \quad (5.18)$$

$$A'_0 = a_0 - 2 \frac{1 - \alpha}{\alpha} b_0 \quad (5.19)$$

Ejemplo 5.8 Jill calcula los valores apropiados para el ejemplo 5.7 usando las ecuaciones 5.18 y 5.19.

$$A_0 = a_0 - \frac{1 - \alpha}{\alpha} b_0 = 650.3 - \frac{.6}{.4} 4.9 = 642.9$$

$$A'_0 = a_0 - 2 \frac{1 - \alpha}{\alpha} b_0 = 650.3 - 2 \frac{.6}{.4} 4.9 = 635.6$$

Jill comienza la atenuación empleando estos valores. La columna 3 es

$$\begin{aligned} A_1 &= \alpha y_1 + (1 - \alpha)A_{1-1} \\ &= .4(654) + (1 - .4)(642.9) = 647.3 \end{aligned}$$

La columna 4 es

$$\begin{aligned} A'_1 &= \alpha A_1 + (1 - \alpha)A'_{1-1} \\ &= .4(647.3) + (1 - .4)(635.6) = 640.3 \end{aligned}$$

La columna 5 es

$$\begin{aligned} a_1 &= 2A_1 - A'_1 \\ &= 2(647.3) - 640.3 = 654.3 \end{aligned}$$

La columna 6 es

$$\begin{aligned} b_1 &= \frac{\alpha}{1 - \alpha} (A_1 - A'_1) \\ &= \frac{.4}{.6} (647.3 - 640.3) = 4.7 \end{aligned}$$

La columna 7 es

$$y_{1+1} = a_1 + b_1(1) = 654.3 + 4.7(1) = 659$$

Atenuación exponencial ajustada a la tendencia: Método de Holt

Otra técnica que se usa con frecuencia para manejar una tendencia lineal se denomina *Método de dos parámetros de Holt*. La técnica de Holt atenúa en forma directa la tendencia

y la pendiente empleando diferentes constantes de atenuación para cada una de ellas. En el enfoque de Brown, sólo se usaba una constante de atenuación y los valores estimados de la tendencia serán muy sensibles a variaciones aleatorias. La técnica de Holt proporciona mayor flexibilidad al seleccionar las proporciones a las que se rastrearán la tendencia y la pendiente.

Las tres ecuaciones que se utilizan en esta técnica son:

1. La serie exponencialmente atenuada:

$$A_t = \alpha Y_t + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (5.20)$$

2. La estimación de la tendencia:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (5.21)$$

3. El pronóstico de p periodos en el futuro:

$$\hat{Y}_{t+p} = A_t + pT_t \quad (5.22)$$

en donde

A_t = nuevo valor atenuado

α = constante de atenuación de los datos ($0 \leq \alpha \leq 1$)

Y_t = nueva observación o valor real de la serie, en el periodo t

β = constante de atenuación de la estimación de la tendencia
($0 \leq \beta \leq 1$)

T_t = estimación de la tendencia

p = periodos a pronosticar en el futuro

$\hat{Y}_{t+p}$ = pronóstico de p periodos en el futuro

La primera ecuación, 5.20, es muy similar a la ecuación original de atenuación exponencial simple, 5.12, con excepción de que se incorpora un nuevo término (T_t) para la tendencia. La estimación de la tendencia se calcula tomando la diferencia entre los valores sucesivos de atenuación exponencial ($A_t - A_{t-1}$). Ya que los valores sucesivos se atenuaron con fines de aleatoriedad, su diferencia constituye una estimación de la tendencia en los datos.

Para atenuar la estimación de la tendencia se utiliza una segunda constante de atenuación, β . La ecuación 5.21 muestra que la estimación de la tendencia ($A_t - A_{t-1}$) se multiplica por β y se suma después a la anterior estimación de la tendencia (T_t), multiplicada por $(1 - \beta)$. La ecuación 5.21 es similar a la 5.20, excepto que la atenuación se realiza para la tendencia en vez de para los datos reales. El resultado de la ecuación 5.21 es una tendencia atenuada que excluye cualquier aleatoriedad.

La ecuación 5.22 muestra el pronóstico de p periodos a futuro. Se multiplica la estimación de la tendencia (T_t) por el número de periodos en el pronóstico (p) y el producto se suma después al nivel actual de los datos A_t , para eliminar aleatoriedad.

Ejemplo 5.9

La tabla 5.9 ilustra la atenuación exponencial ajustada a la tendencia, para los datos de Acme Tool Company. Para comenzar, se requieren dos valores estimados iniciales, el valor inicial estimado y el valor inicial de la tendencia. Por lo regular, el valor inicial atenuado se estima promediando unas cuantas observaciones anteriores de la serie. El valor inicial de la tendencia se estima mediante el uso de la pendiente de la ecuación de la tendencia obtenida de datos anteriores. Si no hay disponibles datos anteriores, se usa cero como estimación inicial.

El valor de alfa es similar al del modelo de atenuación (ecuación 5.12) y atenúa los datos para eliminar aleatoriedad. La constante de atenuación es como α , excepto que atenúa la tendencia en los datos. Ambas constantes de atenuación eliminan aleatoriedad ponderando valores anteriores.

En la tabla 5.9 se demuestra la técnica para $\alpha = .3$ y $\beta = .1$. A continuación se presentan los cálculos que conducen al pronóstico del periodo 3.

1. Actualización de la serie exponencialmente atenuada:

$$\begin{aligned} A_t &= \alpha Y_t + (1 - \alpha)(A_{t-1} + T_{t-1}) \\ A_2 &= .3y_2 + (1 - .3)(A_{2-1} + T_{2-1}) \\ &= .3(350) + .7(500 + 0) = 455 \end{aligned}$$

2. Actualización de la estimación de la tendencia

$$\begin{aligned} T_t &= \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \\ T_2 &= .1(A_2 - A_{2-1}) + (1 - .1)T_{2-1} \\ &= .1(455 - 500) + .9(0) = -4.5 \end{aligned}$$

3. Pronóstico de un periodo a futuro:

$$\begin{aligned} \hat{Y}_{t+p} &= A_t + pT_t \\ \hat{y}_{2+1} &= A_2 + pT_2 \\ \hat{y}_3 &= 455 + (-4.5) = 450.5 \end{aligned}$$

4. Determinación del error del pronóstico

$$\begin{aligned} e_t &= Y_t - \hat{Y}_t \\ e_3 &= y_3 - \hat{y}_3 = 250 - 450.5 = -200.5 \end{aligned}$$

El pronóstico para el periodo 25 se calcula como:

1. Actualización de la serie exponencialmente:

$$\begin{aligned} A_{24} &= .3y_{24} + (1 - .3)(A_{24-1} + T_{24-1}) \\ A_2 &= .3(650) + .7(517.6 + 9.8) = 564.2 \end{aligned}$$

2. Actualización de la estimación de la tendencia:

$$\begin{aligned} T_{24} &= .1(A_{24} - A_{24-1}) + (1 - .1)T_{24-1} \\ &= .1(564.2 - 517.6) + .9(9.8) = 13.5 \end{aligned}$$

3. Pronóstico de un periodo a futuro

$$\begin{aligned} \hat{y}_{24+1} &= A_{24} + pT_{24} \\ \hat{y}_{25} &= 564.2 + (13.5) = 577.7 \end{aligned}$$

TABLA 5.9 PRONÓSTICO DE ATENUACIÓN EXPONENCIAL DE LAS VENTAS DE ACME TOOL COMPANY, TÉCNICA DE HOLT, PARA EL EJEMPLO 5.9

(1) t	(2) Y_t	(3) A_t	(4) T_t	(5) $\hat{Y}_t$	(6) e_t
1	500	500.0	0	500	0
2	350	455.0	-4.5	500	-150
3	250	390.4	-10.5	450.5	-200.5
4	400	385.9	-9.9	379.8	20.2
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
9	350	307.4	-12.3	289.1	60.9
10	200	266.6	-15.2	295.1	-95.0
11	150	221.0	-18.2	251.4	-101.4
12	400	262.0	-12.3	202.8	197.2
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
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
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
25				577.7	

EMC = 21 488.3.

Con base en la minimización del error medio cuadrado (EMC), este modelo no resulta mejor que el de atenuación exponencial simple que utilizó una constante de atenuación de .34. Si se compara los porcentajes de error medio absoluto (PEMA), el método de Holt es mejor. Al comparar los promedios de ventas reales para el primer trimestre de 1994, el

modelo de Holt parece desempeñarse mejor. También el porcentaje medio de error (PME) resulta menor en el método de Holt.

$\alpha = .34$	EMC = 21,421	PEMA = 35.41%	PME = -11.50%
$\left. \begin{matrix} \alpha = .30 \\ \beta = .10 \end{matrix} \right\}$	EMC = 22,380	PEMA = 33.48%	PME = -6.65%
$\left. \begin{matrix} \alpha = .30 \\ \beta = .20 \end{matrix} \right\}$	EMC = 22,349	PEMA = 33.74%	PME = -6.22%

Al utilizar un programa de computadora que elige de manera automática las constantes de atenuación, α es igual a .3 y β es igual a .2. El EMC se reduce ligeramente a 22 349 y el PEMA se mantiene casi igual en 33.74%. El porcentaje medio de error (PME) es -6.22%, que significa que las predicciones basadas en este modelo son aún consistentemente grandes. Se necesita investigar la posibilidad de que haya en los datos una variación estacional.

Atenuación exponencial ajustada a la tendencia y a la variación estacional: Modelo de Winter

Al examinar los datos de Acme Tool Company en la tabla 5.1, se observa que las ventas son consistentemente altas durante el primer y cuarto trimestres y que son menores durante el tercero. Aparentemente existe un patrón estacional. El modelo de atenuación exponencial lineal y estacional de tres parámetros de Winter, que es una extensión del modelo de Holt, pudiera reducir el error de pronóstico. Se utiliza una ecuación adicional para determinar la estacionalidad. Esta estimación está dada por un índice estacional y se calcula con la ecuación 5.25, la cual muestra que la estimación del índice estacional (Y_t/A_t) se multiplica por γ ; se suma después a la estimación estacional anterior (S_{t-L}), multiplicada por $(1 - \gamma)$. La razón Y_t se divide entre A_t para expresar el valor en forma de índice en vez de hacerlo en términos absolutos, de modo que pueda promediarse con el índice estacional atenuado al periodo $t - L$.

Las cuatro ecuaciones que emplea el modelo de Winter son:

1. La serie exponencial atenuada:

$$A_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (5.23)$$

2. La estimación de la tendencia:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (5.24)$$

3. La estimación de la estacionalidad:

$$S_t = \gamma \frac{Y_t}{A_t} + (1 - \gamma)S_{t-L} \quad (5.25)$$

4. El pronóstico de p periodos en el futuro:

$$\hat{Y}_{t+p} = (A_t - pT_t)S_{t-L+p} \quad (5.26)$$

en donde

A_t = nuevo valor atenuado

α = constante de atenuación ($0 < \alpha < 1$)

Y_t = nueva observación o valor real de la serie en el periodo t

β = constante de atenuación de la estimación de la tendencia
($0 < \beta < 1$)

T_t = estimación de la tendencia

γ = constante de atenuación de la estimación de la estacionalidad
($0 < \gamma < 1$)

S_t = estimación de la estacionalidad

p = periodos a estimar a futuro

L = longitud de la estacionalidad

Y_{t+p} = pronóstico de p periodos a futuro

La ecuación 5.23 actualiza la serie atenuada. Una pequeña diferencia en esta ecuación la distingue de la correspondiente (5.20) en el modelo de Holt. En la ecuación 5.23, Y_t se divide S_{t-L} , lo cual ajusta Y_t a la estacionalidad, eliminando así los efectos estacionales que pudieran existir en el dato original Y_t .

Después de atenuar la estimación de la estacionalidad y la estimación de la tendencia en las ecuaciones 5.24 y 5.25, se obtuvo un pronóstico con la ecuación 5.26. Ésta es casi la misma que la fórmula correspondiente, ecuación 5.22, que se emplea para obtener el pronóstico en el modelo de Holt. La diferencia estriba en que esta estimación para un periodo futuro $t + p$, se multiplica por S_{t-L+p} . Este índice estacional es el último disponible, de ahí que se utilice para ajustar el pronóstico a la estacionalidad.

Ejemplo 5.10

La tabla 5.10 ilustra la atenuación exponencial ajustada a la tendencia y a la estacionalidad, para los datos de Acme Tool Company. Para comenzar los cálculos, se requieren las estimaciones iniciales de valor atenuado, tendencia y cada una de las cuatro estimaciones de estacionalidad. El valor inicial atenuado puede estimarse promediando unos cuantos valores anteriores de la serie. El valor inicial de la tendencia se puede estimar utilizando la pendiente de la ecuación de tendencia obtenida de datos anteriores. Las estimaciones de estacionalidad se pueden calcular para datos anteriores mediante la descomposición de la serie de tiempo (cap. 8).

El valor de α es similar al del modelo de atenuación exponencial simple (ecuación 5.12), y los datos para eliminar aleatoriedad. La constante de atenuación β es similar a alfa, salvo que atenúa la tendencia en los datos. La constante de atenuación γ es parecida a alfa y a beta, con excepción de que atenúa la estacionalidad de los datos.

En la tabla 5.10 se demuestra la técnica de Winter para $\alpha = .4$, $\beta = .1$ y $\gamma = .3$. Ya que no hay datos anteriores disponibles, se usa 500 como el valor inicial, 0 como el valor de estimación inicial de tendencia y 1.0 como la estimación inicial de estacionalidad. A continuación se presentan los cálculos que conducen a los valores atenuados actualizados para el periodo 2.

1. La serie exponencialmente atenuada:

$$A_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(A_{t-1} + T_{t-1})$$

$$A_2 = .4 \frac{y_2}{1.0} + (1 - .4)(A_{2-1} + T_{2-1})$$

$$= .4 \frac{350}{1.0} + (1 - .4)(500 + 0) = 140 + 300 = 440$$

2. La estimación de la tendencia:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1}$$

$$T_2 = .1 (A_2 - A_{2-1}) + (1 - .1)T_{2-1}$$

$$= .1(440 - 500) + (1 - .1)0 = -6 + 0 = -6$$

3. La estimación de la estacionalidad:

$$S_t = \gamma \frac{Y_t}{A_t} + (1 - \gamma)S_{t-L}$$

$$S_2 = .3 \frac{y_2}{A_2} + (1 - .3)1.0$$

$$= .3 \frac{350}{440} + (1 - .3)1.0 = .2386 + .7 = .9386$$

Los cálculos que conducen al pronóstico del periodo 6 se muestran a continuación.

1. La serie exponencialmente atenuada:

$$A_5 = .4 \frac{y_5}{S_{5-4}} + (1 - .4)(A_{5-1} + T_{5-1})$$

$$= .4 \frac{450}{1.0} + .6[368.4 + (-11.2)] = 394.189$$

2. La estimación de la tendencia:

$$T_5 = .1(A_5 - A_{5-1}) + (1 - .1)T_{5-1}$$

$$= .1(394.189 - 368.2) + .9(-11.2) = -7.5$$

3. La estimación de la estacionalidad

$$S_5 = .3 \frac{y_5}{A_5} + (1 - .3)S_{5-4}$$

$$= .3 \frac{450}{394.2} + .7(1.0) = .3425 + .7 = 1.0425$$

4. El pronóstico de un periodo a futuro

$$\hat{Y}_{t+p} = (A_t + pT_t)S_{t-L+p}$$

$$\hat{y}_{5+1} = (A_5 + pT_5)S_{5-4+1}$$

$$\hat{y}_6 = [394.2 + 1(-7.5)](.9386) = 362.9$$

La técnica de Winter resulta mejor que los modelos anteriores si se considera el minimizar el EMC. Al comparar los pronósticos de ventas reales para el primer trimestre de 1994, también la técnica de Winter parece desempeñarse mejor. El uso de un programa de computadora para seleccionar de manera automática las mejores constantes de atenuación, reduciría aún más el EMC.

TABLA 5.10 PRONÓSTICO DE ATENUACIÓN EXPONENCIAL DE LAS VENTAS DE ACME TOOL COMPANY, TÉCNICA DE WINTER, PARA EL EJEMPLO 5.10

(1) t	(2) Y_t	(3) A_t	(4) T_t	(5) S_t	(6) $\hat{Y}_{t+p}$	(7) e_t
1	500	500.0	0.0	1.00	500.0	0.0
2	350	440.0	-6.0	0.94	500.0	-150.0
3	250	360.4	-13.4	0.91	500.0	-250.0
4	400	368.2	-11.2	1.03	500.0	-100.0
5	450	394.2	-7.5	1.04	357.0	93.0
6	350	381.2	-8.1	0.93	362.9	-12.9
7	200	311.9	-14.2	0.83	338.8	-138.8
8	300	295.6	-14.4	1.02	305.5	-5.5
9	350	303.0	-12.2	1.08	293.2	56.8
10	200	260.3	-15.3	0.88	271.2	-71.2
11	150	219.5	-17.8	0.78	202.9	-52.9
12	400	277.5	-10.2	1.15	206.2	193.8
13	550	364.7	-0.5	1.21	287.6	262.4
14	350	377.0	0.8	0.90	321.7	28.3
15	250	354.1	-1.6	0.76	296.5	-46.5
16	550	403.1	3.5	1.21	404.9	145.1
17	550	426.4	5.5	1.23	490.3	59.7
18	400	437.5	6.0	0.90	387.3	12.7
19	350	450.1	6.7	0.77	337.6	12.4
20	600	471.9	8.2	1.23	554.1	45.9
21	750	531.8	13.4	1.28	591.0	159.0
22	500	548.8	13.7	0.90	491.7	8.3
23	400	546.4	12.1	0.76	430.9	-30.9
24	650	546.4	10.9	1.22	687.3	-37.3
25					751.9	
26					546.2	
27					449.6	
28					718.8	

EMC = 12 431.5.

El método de Winter proporciona una forma fácil de explicar la estacionalidad en un modelo, cuando los datos tienen un patrón estacional. Un método alternativo consiste en desestacionalizar primero o ajustar estacionalmente los datos. La *desestacionalización* es un proceso que quita a los datos en bruto el efecto de estacionalidad; este proceso se mostrará en el cap.8. Después se aplica el modelo de pronóstico a los datos sin estacionalidad y posteriormente los resultados se estacionalizan de nuevo para proporcionar pronósticos precisos.

Es frecuente que los pronosticadores encuentren este método más preciso que el uso del enfoque de Winter sobre los datos originales. Este enfoque es también más flexible que el método de Winter, pues permite utilizar la atenuación simple cuando los datos son estacionales, pero no hay presente una tendencia.

La atenuación exponencial es una técnica popular para los pronósticos de corto plazo. Sus mejores ventajas son un bajo costo y simplicidad. Es frecuente que los métodos de atenuación exponencial sean el único enfoque cuando se requieren pronósticos de inventarios que contienen miles de elementos.

Los pronósticos de promedios móviles simples y atenuación exponencial se basan en promedios ponderados de mediciones anteriores. La explicación consiste en que los valores anteriores contienen información de lo que ocurrirá en el futuro. Debido a que los valores anteriores incluyen fluctuaciones aleatorias, así como información relativa al patrón subyacente de una variable, se hace un intento de atenuar estos valores. Este enfoque supone que las fluctuaciones extremas representan aleatoriedad en una serie de observaciones de tiempo.

Los promedios móviles implican el cálculo de la media de cierto número de valores de una variable. Este promedio se convierte ahora en el pronóstico del siguiente periodo; después, el proceso se repite hasta formular un pronóstico del periodo futuro deseado. Este método asigna una ponderación igual a cada valor anterior. Sin embargo, se puede formular un argumento convincente para dar mayor importancia a los valores más recientes. Se emplea la atenuación exponencial debido a que se asigna una ponderación decreciente a las observaciones más antiguas. Ambos enfoques comprenden una regla o conjunto de reglas para determinar las ponderaciones.

APLICACIÓN EN LA ADMINISTRACIÓN

Los pronósticos son uno de los insumos más importantes que desarrollan los administradores para auxiliarse en el proceso de toma de decisiones. Virtualmente, cada decisión operativa importante depende en algún grado de un pronóstico. La acumulación de demanda se basa en un pronóstico de la demanda esperada; el departamento de producción tiene que programar los requerimientos de personal y los pedidos de materia prima para el siguiente o los dos siguientes meses; el departamento de finanzas debe determinar los requerimientos de contrataciones y suspensiones temporales. La lista de aplicaciones de los pronósticos resulta muy numerosa.

Los ejecutivos están en realidad conscientes de la importancia de los pronósticos. De hecho, se dedica una gran cantidad de tiempo al estudio de las tendencias en asuntos económicos y políticos y cómo pueden afectar los sucesos la demanda de productos y/o

servicios. Un aspecto de interés es la importancia que asignan los ejecutivos a los pronósticos cuantitativos, comparados con sus propias opiniones. Este aspecto es en especial sensible cuando participan sucesos que tienen un impacto significativo en la demanda. Un problema de los métodos de pronóstico cuantitativos es que dependen de datos históricos. Quizá por esta razón sean menos efectivos para predecir los cambios que a menudo resultan en una mayor o menor demanda. La computadora ha auxiliado a los administradores en la automatización de la sensibilidad a cambios agudos en la demanda, a través de sistemas adaptativos y modelos causales.

Los métodos no formales son útiles por su relativa simplicidad. Estos métodos simples tienden a ser menos costosos, más fáciles de implementar y más sencillos de comprender. Con frecuencia, el costo y la complejidad de los modelos más sofisticados tienen mayor peso que cualquier ganancia en precisión. Por estos motivos, los negocios pequeños encuentran prácticos los modelos no formales. Las empresas que no tienen computadoras y/o personal capaz de manejar modelos estadísticos, prefieren también los modelos no formales. A menudo, los administradores de empresas se enfrentan con la necesidad de preparar pronósticos a corto plazo para varios elementos diferentes. Un ejemplo clásico es el del gerente que debe programar la producción con base en algún pronóstico de demanda, para cientos de productos distintos, en una línea de producción. Por último, las empresas nuevas que no tienen grandes bases de datos históricos, encuentran útiles estos enfoques.

El método de promedio móvil realiza bien su trabajo de ajuste a los cambios en los patrones. Es económico de actualizar y no requiere almacenar muchos datos. Debido a que estas ventajas se ven de algún modo superadas por altos costos iniciales, los métodos de promedio móvil se usan con más frecuencia cuando son necesarios pronósticos repetidos.

La atenuación exponencial es una técnica popular cuya fuerza reside en una buena precisión a corto plazo combinada con una rápida actualización de bajo costo. La técnica se usa ampliamente cuando se requieren pronósticos regulares mensuales o semanales para un gran número de elementos. La atenuación exponencial se emplea cuando se requieren pronósticos de sistemas de inventario que contienen miles de elementos.

GLOSARIO

Atenuación exponencial Es un proceso utilizado para revisar constantemente un pronóstico a la luz de la experiencia más reciente.

Promedio móvil Se obtiene encontrando la media de un conjunto específico de valores y usándola después para pronosticar el siguiente periodo.

Promedio simple Se obtiene encontrando la media de todos los valores aplicables y empleando después esa media para pronosticar el siguiente periodo.

Señal de rastreo Comprende el cálculo de alguna medición de error a través del tiempo y el establecimiento de límites, de modo que cuando el error acumulativo rebasa dicho límite, se alerte al pronosticador.

FÓRMULAS CLAVE

Modelo no formal $\hat{Y}_{t+1} = Y_t$ (5.1)

Modelo no formal de tendencia $\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1})$ (5.2)

Modelo no formal de proporción de cambio $\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}}$ (5.3)

Modelo no formal estacional $\hat{Y}_{t+1} = Y_{t-3}$ (5.4)

Modelo no formal de tendencia y estacional

$$\hat{Y}_{t+1} = Y_{t-3} + \frac{(Y_t - Y_{t-1}) + \cdots + (Y_{t-3} - Y_{t-4})}{4} \quad (5.5)$$

Promedio simple $\hat{Y}_{t+1} = \frac{\sum_{t=1}^n Y_t}{n}$ (5.6)

Promedio móvil $\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2} + \cdots + Y_{t-n+1}}{n}$ (5.7)

Promedio móvil doble $M'_t = \frac{M_t + M_{t-1} + M_{t-2} + \cdots + M_{t-n+1}}{n}$ (5.8)

$$a_t = 2M_t - M'_t \quad (5.9)$$

$$b_t = \frac{2}{n-1} (M_t - M'_t) \quad (5.10)$$

$$\hat{Y}_{t+p} = a_t + b_t p \quad (5.11)$$

Atenuación exponencial $\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$ (5.12)

Atenuación exponencial doble $A_t = \alpha Y_t + (1 - \alpha) A_{t-1}$ (5.13)

$$A'_t = \alpha A_t + (1 - \alpha) A'_{t-1} \quad (5.14)$$

$$a_t = 2A_t - A'_t \quad (5.15)$$

$$b_t = \frac{\alpha}{1 - \alpha} (A_t - A'_t) \quad (5.16)$$

$$\hat{Y}_{t+p} = a_t + b_t p \quad (5.17)$$

Valores iniciales de la atenuación exponencial doble

$$A_0 = a_0 - \frac{1 - \alpha}{\alpha} b_0 \quad (5.18)$$

$$A'_0 = a_0 - 2 \frac{1 - \alpha}{\alpha} b_0 \quad (5.19)$$

$$\text{Modelo de Holt } A_t = \alpha Y_t + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (5.20)$$

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (5.21)$$

$$\hat{Y}_{t+p} = A_t + pT_t \quad (5.22)$$

$$\text{Modelo de Winter } A_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (5.23)$$

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (5.24)$$

$$S_t = \gamma \frac{Y_t}{A_t} + (1 - \gamma)S_{t-L} \quad (5.25)$$

$$\hat{Y}_{t+p} = (A_t + pT_t)S_{t-L+p} \quad (5.26)$$

PROBLEMAS

1. ¿Qué técnica de pronóstico revisa constantemente una estimación a la luz de la experiencia más reciente?
2. ¿Qué técnica de pronóstico usa el valor del periodo actual como el pronóstico del siguiente periodo?
3. ¿Qué técnica asigna igual ponderación a cada observación?
4. ¿Qué técnica se debe emplear si los datos tienen tendencia?
5. ¿Qué técnica se debe emplear si los datos son estacionales?
6. La Apex Mutual Fund invierte principalmente en acciones de tecnología. Los valores de posesión neta del fondo al final de cada mes de 1994, son los siguientes:

MES	PRECIO DEL FONDO
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

- a. Encuentre el valor del pronóstico para la Mutual Fund, para cada mes, empleando un modelo no formal. El valor para diciembre de 1993 fue 19.00.
 - b. Evalúe el método de pronóstico mediante DAM
 - c. Evalúe el método de pronóstico mediante EMC
 - d. Evalúe el método de pronóstico mediante PEMA
 - e. Evalúe el método de pronóstico mediante PME
 - f. Pronostique el precio de la Mutual Fund para enero de 1995.
 - g. Escriba un memorando resumiendo sus apreciaciones.
7. Con referencia al problema 6, utilice un promedio móvil de tres meses para pronosticar el precio de la Mutual Fund para enero de 1995. ¿Es este pronóstico mejor que el formulado con el método no formal? Explique.
8. Dada la serie siguiente:

PERIODO	Y_t	$\hat{Y}_t$	e_t
1	200	100	—
2	210		
3	215		
4	216		
5	219		
6	220		
7	225		
8	226		

- a. ¿Cuál es el pronóstico para el periodo 8, utilizando un promedio móvil de cinco meses?
 - b. Si se emplea una constante de atenuación de .4, ¿cuál sería el pronóstico de atenuación exponencial para el periodo 4?
 - c. En el apartado b, ¿cuál sería 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 1994 son las siguientes:

MES	DEMANDA
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

- a. Encuentre el valor de pronóstico del rendimiento de los bonos de obligacion general para cada mes, comenzando por abril, mediante un promedio móvil de tres meses.
 - b. Encuentre el valor del pronóstico del rendimiento de los bonos de obligacion general para cada mes, comenzando por junio, mediante un promedio móvil de cinco meses.
 - c. Evalúe estos métodos de pronóstico mediante DAM
 - d. Evalúe estos métodos de pronóstico mediante EMC
 - e. Evalúe estos métodos de pronóstico mediante PEMA
 - f. Evalúe estos métodos de pronóstico mediante PME
 - g. Pronostique el rendimiento para enero de 1995 usando la mejor técnica.
 - h. Redacte un memorando resumiendo sus apreciaciones.
10. Esta pregunta se refiere al problema 9. Utilice la atenuación exponencial con una constante de atenuación de .2 y un valor inicial de 9.29, para pronosticar el rendimiento de enero de 1995. ¿Es este pronóstico mejor que el realizado mediante el modelo de promedio móvil? Explique.
11. La Hughes Supply Company utiliza un método de administración de inventario para determinar las demandas mensuales de varios productos. Se registraron los valores de demanda de los últimos 12 meses y están disponibles para pronósticos posteriores. A continuación se presentan los valores de demanda de un dispositivo eléctrico, para los 12 meses de 1994:

MES	DEMANDA
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: registros de Hughes Supply Company.

Use la atenuación exponencial con una constante de atenuación de .5 y un valor inicial de 205, para pronosticar la demanda en enero de 1995.

12. La General American Investors Co., una compañía regulada de administración de inversiones, invierte principalmente en acciones de calidad media y alta. Jim Campbell está estudiando el valor de posesión por acción para esta compañía y le gustaría pronosticar esta variable para 1993.

Los datos son los siguientes:

	TRIMESTRE			
	1	2	3	4
1985	16.98	18.47	17.63	20.65
1986	21.95	23.85	20.44	19.29
1987	22.75	23.94	24.84	16.70
1988	18.04	19.19	18.97	17.03
1989	18.23	19.80	22.89	21.41
1990	21.50	25.05	20.33	20.60
1991	25.33	26.06	28.89	30.60
1992	27.44	26.69	28.71	28.56

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993), p. 2096.

Calcule el valor de posesión por acción utilizando los siguientes métodos de pronóstico: no formal, promedio móvil y atenuación exponencial. Al comparar estas técnicas considere que el valor real de posesión por acción para el primer trimestre de 1993 fue de 25.87. Redacte un informe para Jim indicando el método que debe utilizar y por qué.

13. Southdown, Inc., el tercer productor de cemento a nivel nacional, está impulsando un programa de quema de combustible de desperdicio. El costo total para Southdown se estima en 37 millones de dólares. Por esta razón, es en extremo importante para la compañía tener un pronóstico preciso de utilidades para el segundo trimestre de 1993. Los datos son:

	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			

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993), p. 900.

- Utilice la atenuación exponencial con una constante de atenuación de .4 y un valor inicial de 77.4 para pronosticar los ingresos por acción para el segundo trimestre de 1993.
- Utilice ahora una constante de atenuación de .6 y un valor inicial de 77.4 para pronosticar los ingresos por acción para el segundo trimestre de 1993.
- Estime la constante de atenuación que dará el mejor pronóstico.
- ¿Cómo es el valor de su pronóstico comparado con el de 142 de *Value Line*?

14. La Triton Energy Corporation realiza exploraciones y produce petróleo y gas. La presidenta de la compañía, Gail Freeman, desea que un analista pronostique las ventas por acción para 1993. Éste será un pronóstico importante, ya que los planes de restructuración de Triton se encuentran en un bache. Los datos son:

VENTAS POR ACCIÓN DE TRITON, 1974–1992

AÑO	VENTAS POR ACCIÓN	AÑO	VENTAS POR ACCIÓN
1974	.93	1984	1.93
1975	1.35	1985	5.17
1976	1.48	1986	7.72
1977	2.36	1987	5.33
1978	2.45	1988	8.12
1979	2.52	1989	10.65
1980	2.81	1990	12.06
1981	3.82	1991	11.63
1982	5.54	1992	6.58
1983	7.16		

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993), p. 1848.

Utilice tanto el promedio móvil doble de Brown como la técnica de Holt, para pronosticar 1993. Compare los resultados mediante las mediciones apropiadas. ¿Cómo resulta su pronóstico comparado con el de 8.30 de *Value Line*?

15. La Consolidated Edison Company vende electricidad (82% de sus ingresos), gas (13%) y vapor (5%) en la ciudad de Nueva York y en el condado de Westchester. A Bart Thomas, pronosticador de la compañía, se le asigna la tarea de pronosticar las utilidades por trimestre para 1993. Los datos reunidos son:

INGRESOS POR TRIMESTRE DE CONSOLIDATED EDISON
(MILLONES DE DÓLARES), 1985–1992

AÑO	MARZO 31	JUNIO 30	SEPTIEMBRE 30	DICIEMBRE 31
1985	1441	1209	1526	1321
1986	1414	1187	1411	1185
1987	1284	1125	1493	1192
1988	1327	1102	1469	1213
1989	1387	1218	1575	1371
1990	1494	1263	1613	1369
1991	1479	1330	1720	1344
1992	1456	1280	1717	1480

Fuente: *The Value Line Investment Survey* (Nueva York: Value Line, 1990, 1993), p. 168.

Determine la mejor técnica y formule el pronóstico para 1993.

CASO DE ESTUDIO 5.1**THE SOLAR ALTERNATIVE COMPANY²**

La Solar Alternative Company está por iniciar su tercer año en funcionamiento. La compañía fue fundada por Bob y Mary Johnson, ambos profesores de ciencias en la secundaria local. Los Johnson fundaron esta empresa para obtener ingresos complementarios a su sueldo académico. Con base en sus investigaciones sobre sistemas de energía solar, pudieron armar un sistema solar para calentar agua en el hogar. El sistema consta de un tanque de almacenamiento de 378.5 litros, de fibra de vidrio, dos paneles solares de .91 por (1.83 metros), controles electrónicos, tubería de PVC y diversas piezas para instalación.

El periodo de vida del sistema es de diez años. Aunque desde el punto de vista financiero esta situación no representa una oportunidad atractiva de inversión, existe suficiente interés en la novedad del concepto para proporcionar un nivel moderado de ventas. Después de costos y gastos, los Johnson obtienen alrededor de \$75 dólares libres de los \$2000 que cuesta un sistema instalado. El costo de materiales y equipo es del 75% del costo del sistema instalado. Un punto a favor del bajo margen de ganancia es que el producto no es lo suficientemente rentable como para tener competencia por parte de los contratistas en calefacción. Los Johnson tienen el negocio en su casa, con una oficina en el sótano y emplean su cochera en donde cabe sólo un automóvil, exclusivamente para almacenar componentes y materiales del sistema. Como resultado, los gastos adicionales son mínimos. Los Johnson disfrutan de un modesto ingreso complementario por el funcionamiento de la empresa, la cual también representa ciertas ventajas de carácter fiscal.

Bob y Mary están satisfechos con el crecimiento del negocio. Aunque las ventas varían de un mes a otro, en general el segundo año fue mucho mejor que el primero. Muchos clientes en el segundo año fueron vecinos de personas que compraron el sistema durante el primer año. Aparentemente, después de ver funcionar con éxito el sistema durante un año, otros se animaron a probar el concepto solar. Las ventas se suceden a lo largo del año. La mayor demanda se presenta al final del verano y principios del otoño, cuando por lo regular la gente hace planes para preparar sus hogares para la temporada de invierno, cuando emplearán calefacción.

Con el crecimiento anticipado del negocio, los Johnson piensan que necesitan un pronóstico de ventas para manejar con efectividad el próximo año. Por lo regular, toma de 60 a 90 días la recepción de los tanques de almacenamiento, después de hacer el pedido. Los paneles solares están disponibles de inmediato la mayor parte del año. Sin embargo, al final del verano y durante el otoño el tiempo requerido para terminar un proyecto puede llevar de 90 a 100 días. Aunque existe una competencia limitada, la pérdida de ventas puede convertirse en una posibilidad real si se le pide al cliente potencial que espere varios meses para instalarle. Tal vez sea más importante la necesidad de hacer una proyección de ventas para aprovechar los descuentos en compras por volumen. Al combinar estos factores con el alto costo de los componentes del sistema y el limitado espacio de almacenamiento disponible en la

MES	1993	1994	MES	1993	1994
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

² Este caso es una contribución de William P. Darrow de la Towson State University, Towson, Maryland

cochera, hacen necesario desarrollar un pronóstico confiable. En la siguiente tabla se presenta la historia de ventas de la compañía en sus dos primeros años:

PREGUNTAS

1. Identifique el modelo que deben emplear Bob y Mary como base en la planeación de su empresa en 1995 y exponga por qué se eligió dicho modelo.
2. Pronostique las ventas para 1995.

CASO DE ESTUDIO 5.2

MR. TUX

John Mosby, propietario de varias tiendas de renta de ropa de etiqueta Mr. Tux, comienza a pronosticar la variable más importante de su negocio, ventas mensuales en dólares. (Véanse los casos de Mr. Tux al final de los cap. 1 y 2). Una de sus empleadas, Virginia Perot, reunió los datos de ventas que se presentaron en el caso del cap. 2..

John usó el programa Sybil/Runner en el caso de estudio 4.2, para analizar los patrones de los datos. Ahora desea pronosticar por medio de las técnicas de promedio móvil y de atenuación exponencial.

Comienza pidiendo al programa que pronostique mediante atenuación exponencial simple. El programa le pide que indique la constante de atenuación (α) a utilizar, o que indique si desea que calcule el valor óptimo. John hace esto último y el programa encuentra que el valor óptimo de α es .400.

El programa calcula también varios resúmenes de mediciones de error. Estos valores resumen los errores en la predicción de valores reales de los datos históricos, mediante el método de atenuación exponencial simple en donde $\alpha = .400$. John decide registrar tres mediciones de error que son:

$$\begin{aligned} &= DAM = 281.66 \\ &= PME = -5.0\% \\ &= PEMA = 24.0\% \end{aligned}$$

La desviación absoluta media (DAM) es el error promedio en que se incurre al pronosticar valores anteriores. El porcentaje medio de error (PME) indica un ligero sesgo. El PEMA indica que el error promedio es de 24% de los valores reales, una cifra que John encuentra más bien elevada.

A continuación, John pide al programa que aplique a sus datos la atenuación exponencial de Holt. Este programa utiliza el método de atenuación exponencial, pero puede también contemplar una tendencia en los datos. De nuevo, John hace que el programa calcule las constantes óptimas a utilizar en la actualización de la serie y la tendencia. Dichos valores resultan ser $\alpha = .167$ y $\beta = .001$. Los tres resúmenes de medición de error, para el método de Holt, son

$$\begin{aligned} &= DAM = 245.84 \\ &= PME = -6.1\% \\ &= PEMA = 18.8\% \end{aligned}$$

El siguiente método que prueba John es la atenuación exponencial de Winter. Este método puede tomar en cuenta tanto la tendencia como los factores estacionales, y da por resultado las siguientes constantes de atenuación: $\alpha = .333$, $\beta = .667$, $\gamma = .333$. Las mediciones de error son

$$\begin{aligned} &= DAM = 334.52 \\ &= PME = -4.6\% \\ &= PEMA = 22.6\% \end{aligned}$$

Por último, John da instrucciones al programa para utilizar un promedio móvil de tres meses sobre sus datos. Las mediciones de error para este modelo son

$$DAM = 288.58$$

$$PME = 0.6 \%$$

$$PEMA = 8.9 \%$$

Al comenzar el estudio de los resultados de su análisis, John se desilusiona con los resultados obtenidos. Esperaba que alguno de los métodos le diera pronósticos precisos de los periodos anteriores: entonces podría usar ese método para pronosticar los niveles de ventas de los meses por venir del próximo año. Pero los errores promedio (DAM) y los porcentajes de error (PEMA) de estos métodos lo llevan a buscar otra forma de pronóstico.

TAREA

1. Resuma el nivel de error para el mejor método encontrado por John mediante el programa Sybil/Runner.
2. ¿Cuál es el valor ideal del PME al utilizar cualquier método de atenuación exponencial?
3. ¿Cuál es el significado de un signo negativo en el PME?
4. ¿Está usted de acuerdo con la decepción de John respecto a sus resultados?

CASO DE ESTUDIO 5.3

CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se describió en los caps.1 (caso de estudio 1.2) y 3 (caso de estudio 3.4).

El director ejecutivo Marv Harnishfeger, llegó a la conclusión de que la variable más importante que requería pronosticar CCC era el número de nuevos clientes por atender en el resto de 1993. Marv proporcionó a Dorothy Mercer los datos mensuales de nuevos clientes atendidos por CCC en el periodo de enero a marzo de 1993. Véase caso de estudio 4.3. En ese caso de estudio, Dorothy empleó el análisis de autocorrelación para explorar el patrón de datos. Con base en los resultados de esta investigación:

TAREA

1. Desarrolle un modelo no formal para pronosticar el número de clientes a atender por CCC en el resto de 1993.
2. Desarrolle un modelo de promedio móvil para pronosticar el número de clientes a atender por CCC en el resto de 1993.
3. Desarrolle un modelo de atenuación exponencial para pronosticar el número de clientes a atender por CCC en el resto de 1993.
4. Evalúe estos modelos mediante las técnicas de medición de error de pronóstico que se presentaron en el cap. 4.

5. Elija el mejor modelo y formule el pronóstico para el resto de 1993.
6. Determine por qué es adecuado el método de pronóstico que eligió.

CASO DE ESTUDIO 5.4

PROYECCIÓN DE INGRESO QUINQUENAL PARA DOWNTOWN RADIOLOGY

Downtown Radiology está desarrollando el centro médico de imágenes más completo y tecnológicamente avanzado que cualquier otro ubicado en el área. El equipo que se planea para el centro será igual o superior a las instalaciones de imágenes de todos los centros médicos de la región. El centro tendrá un digitalizador para tomografía computarizada CT serie 9800 y un equipo nuclear de imágenes por resonancia magnética (MRI). También incluirá ultrasonido, medicina nuclear, angiografía por sustracción digital (DSA), equipo de mamografía y equipo convencional de radiología y fluoroscopia. Se ofrecerán membresías mediante alguna forma de oferta pública, y Downtown desea una evaluación independiente del mercado, para lo cual solicitó a Professional Marketing Associates, Inc., una evaluación del mercado así como una proyección de ingresos para cinco años.

ENUNCIADO DEL PROBLEMA

El propósito de este estudio consiste en pronosticar los próximos cinco años para el centro médico de imágenes propuesto.

OBJETIVOS

Los objetivos del estudio son

- Identificar áreas de mercado para cada tipo de servicio a ofrecer por las nuevas instalaciones.
- Reunir y analizar los datos existentes sobre ingresos en el área de mercado para cada tipo de servicio a ofrecer por las nuevas instalaciones.
- Identificar las tendencias en la industria de atención a la salud, que afecten positiva o negativamente los ingresos por los servicios a proporcionar por las nuevas instalaciones.
- Identificar, en la planeación del nuevo riesgo, factores en empresas, comercialización o instalaciones, que impactarían positiva o negativamente las proyecciones de ingresos.
- Analizar los servicios anteriores de Downtown Radiology como una base de datos para el modelo de pronósticos a desarrollar.
- Utilizar el modelo apropiado de pronóstico cuantitativo para llegar a las proyecciones de cinco años de ingresos para el centro propuesto.

METODOLOGIA

Procedimientos

Con el propósito de formular la proyección de cinco años de ingresos, se implementaron los siguientes pasos: se realizó un análisis del número de servicios anteriores. Se desarrolló el método apropiado de pronóstico y se utilizó para determinar un punto de partida para la proyección de cada servicio.

1. Se determinó el área de mercado para cada tipo de servicio y se obtuvieron pronósticos de población para 1986 y 1990.
2. Se estudiaron patrones de referencias de doctores, para determinar el número de doctores que enviaban pacientes a Downtown Radiology, así como el número de pacientes que envió cada doctor.
3. Se obtuvieron cifras del National Center for Health Statistics. Estas estadísticas se compararon con números reales obtenidos de la comisión del hospital.
4. La participación en el mercado de Downtown Radiology se calculó con base en tomografías computarizadas reales en el área de mercado (la participación en el mercado para otros servicios se determinó con base en la participación de Downtown Radiology, comparada con las cifras proporcionadas por el National Center for Health Statistics).

Suposiciones

Para desarrollar el pronóstico cuantitativo, fue necesario formular ciertas suposiciones. Se hicieron las siguientes:

- El nuevo centro de imágenes estará funcionando con todo el equipo, excepto del de MRI, para el 1 de enero de 1985.
- El equipo nuclear de imágenes por resonancia magnética (MRI) estará funcionando en abril de 1985.
- La oferta de sociedad limitada tendrá un mercado exitoso por lo menos para 50 médicos en el área de servicio.
- Los doctores quienes tienen interés económico en el nuevo centro de imágenes, incrementarán el número de pacientes que envían al centro.
- No habrá otros equipos MRI en el área de servicio antes de 1987.
- El nuevo centro de imágenes ofrecerá servicios a precios menores que la competencia.
- Se llevará a cabo un esfuerzo efectivo de comercialización, con una concentración especial en empresas con gran número de empleados, grupos de aseguradoras y sindicatos.
- Durante los primeros seis meses de funcionamiento, el equipo MRI sustituirá aproximadamente el 60% de los estudios de tomografía de la cabeza, que actualmente se hacen con el equipo CT, y absorberá el 70% en los 12 meses siguientes.
- El público en general seguirá presionando a la industria de atención a la salud, para mantener los precios bajos.
- El costo de imágenes en la industria de atención a la salud se elevó 13.2% anualmente entre 1971 y 1981. La Health Care Financing Administration estima que la tasa promedio de incremento anual se reducirá aproximadamente a 11 o 12% entre 1981 y 1990 (*Industry Surveys*, abril de 1984).
- Las compañías aseguradoras reembolsarán a los pacientes entre el 80 y el 100% del costo de imágenes de resonancia magnética (*Imaging News*, febrero de 1984).

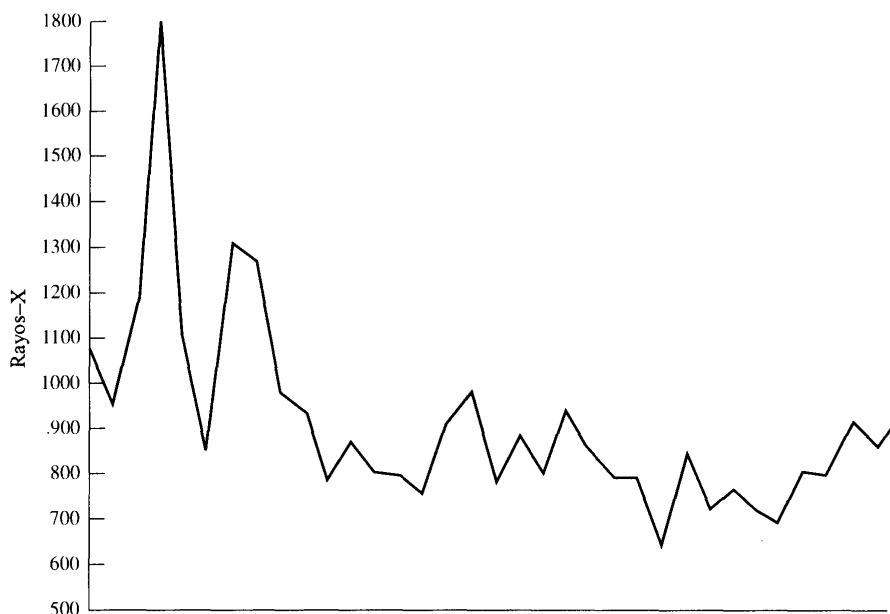


Figura 5.4 Rayos-X por mes, julio de 1981 a mayo de 1994.

Modelos

Se desarrolló un pronóstico para cada servicio con base en la experiencia anterior, las cifras de la industria y suposiciones razonables. Debido a que los modelos se desarrollaron basados en las suposiciones antes mencionadas, si éstas no son válidas, los modelos no serán precisos.

ANÁLISIS DE DATOS ANTERIORES

Rayos-X

Se analizó el número de servicios de rayos-X realizados entre julio de 1981 y mayo de 1984. Los datos comprenden rayos-X de diagnóstico, rayos-X gastrointestinales, imágenes del pecho, inyecciones y procedimientos especiales. Los datos se presentan en la fig. 5.4. El examen de estos datos indica que no están presentes patrones de tendencia, cíclicos o estacionales. Por esta razón, se eligió como método apropiado de pronóstico, la atenuación exponencial. Se examinaron varios factores de atenuación y se encontró que el mejor modelo lo proporcionaba una constante de atenuación de .3. Este valor significa que las observaciones históricas más recientes recibieron una ponderación de .3 y que al resto de la serie histórica se asignó una ponderación de .7. Este modelo se presenta en la tabla 5.11. El pronóstico para junio de 1984 es de 855 servicios de rayos-X.

Ultrasonido

Se analizó el número de servicios de ultrasonido efectuados entre julio de 1981 y mayo de 1984. La fig. 5.5 muestra el patrón de datos. Tampoco aquí se aprecia la presencia de un patrón de tendencia, cíclico o estacional. De nuevo se determinó que el mejor modelo era el de atenuación exponencial con un factor

TABLA 5.11 ATENUACIÓN EXPONENCIAL

DATOS POR ATENUAR: RAYOS-X			FACTOR DE ATENUACIÓN = .3		
PERIODO	RAYOS-X	VALOR ATENUADO	PERIODO	RAYOS-X	VALOR ATENUADO
1	1,072.00	1,072.00	19	882.00	863.94
2	953.00	1,036.00	20	794.00	842.06
3	1,157.00	1,072.51	21	946.00	873.87
4	1,800.00	1,290.76	22	848.00	866.11
5	1,085.00	1,229.03	23	785.00	841.78
6	837.00	1,111.42	24	779.00	822.94
7	1,301.00	1,168.29	25	636.00	766.85
8	1,262.00	1,196.41	26	846.00	790.60
9	977.00	1,130.58	27	712.00	767.02
10	937.00	1,072.51	28	767.00	767.02
11	780.00	984.76	29	719.00	752.61
12	873.00	951.23	30	689.00	733.53
13	808.00	908.26	31	804.00	754.67
14	796.00	874.58	32	796.00	767.07
15	752.00	837.81	33	910.00	809.95
16	901.00	856.77	34	849.00	821.66
17	978.00	893.14	35	932.00	854.76
18	770.00	856.20			

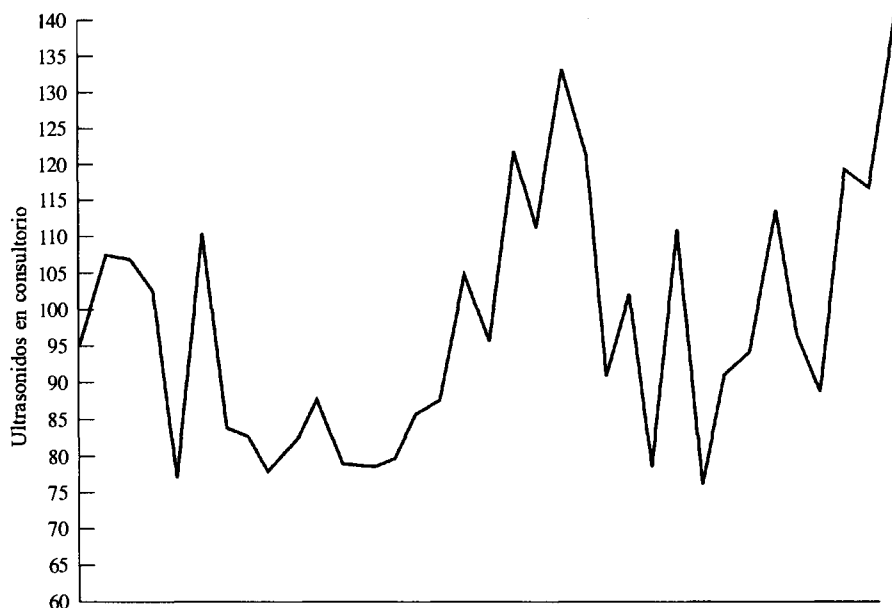


Figura 5.5. Ultrasonidos en consultorio por mes, julio de 1981 a mayo de 1984.

TABLA 5.12 ATENUACIÓN EXPONENCIAL

DATOS POR ATENUAR: ULTRASONIDOS EN CONSULTORIO			FACTOR DE ATENUACIÓN = .3		
PERIODO	VALOR ULTRASONIDOS ATENUADO		PERIODO	VALOR ULTRASONIDOS ATENUADO	
1	94.00	94.00	19	123.00	101.50
2	108.00	98.20	20	111.00	104.35
3	107.00	100.84	21	134.00	113.24
4	103.00	101.49	22	122.00	115.87
5	77.00	94.14	23	91.00	108.41
6	111.00	99.20	24	103.00	106.79
7	84.00	94.64	25	79.00	98.45
8	83.00	91.15	26	111.00	102.22
9	78.00	87.20	27	77.00	94.65
10	82.00	85.64	28	92.00	93.86
11	88.00	86.35	29	95.00	94.20
12	79.00	84.14	30	114.00	100.14
13	79.00	82.60	31	97.00	99.20
14	80.00	81.82	32	89.00	96.14
15	86.00	83.07	33	120.00	103.30
16	88.00	84.55	34	117.00	107.41
17	105.00	90.69	35	142.00	117.79
18	96.00	92.28			

de atenuación de .3. La tabla 5.12 muestra que el pronóstico para junio de 1984 es de 118 servicios de ultrasonido.

Se analizó el número de servicios de ultrasonido realizados por las dos unidades móviles propiedad de Downtown Radiology, en el periodo de julio de 1981 a mayo de 1984. En la fig. 5.6 se

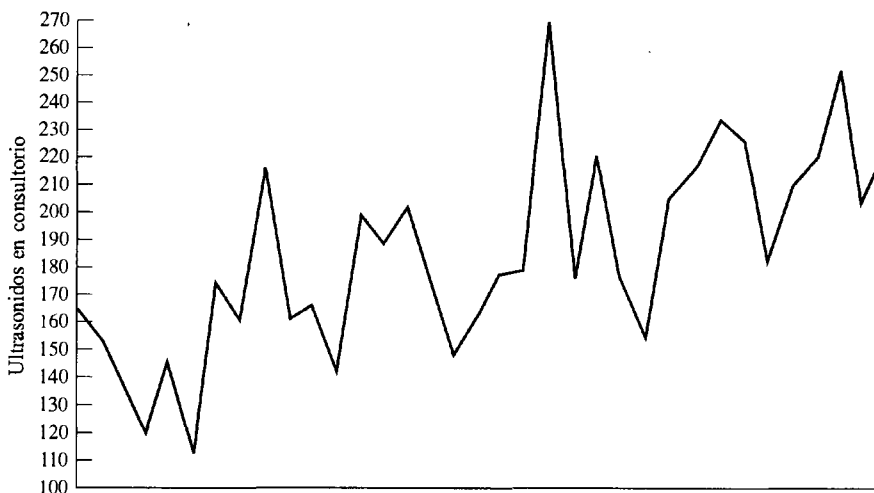


Figura 5.6 Ultrasonidos fuera de consultorio por mes, julio de 1981 a mayo de 1984.

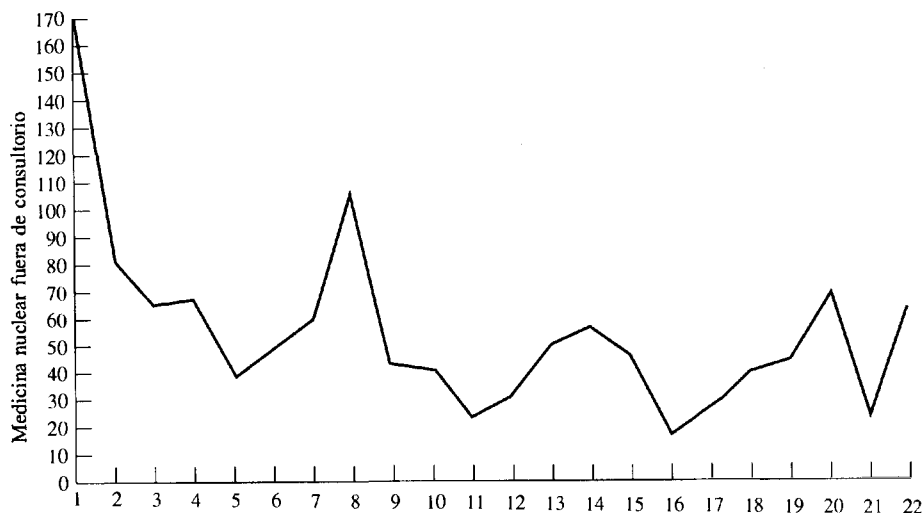


Figura 5.7 Medicina nuclear fuera de consultorio por mes, agosto de 1982 a mayo de 1984.

muestra el patrón de datos. En apariencia, existe una tendencia creciente que se puede moderar mediante el método de atenuación exponencial lineal de dos parámetros, de Holt. Se emplean las constantes de atenuación $\alpha = .37$ y $\beta = .18$; el pronóstico para junio de 1984 es de 226.

Servicios de medicina nuclear

Se analizó el número de servicios de medicina nuclear prestados por las unidades móviles propiedad de Downtown Radiology, entre agosto de 1982 y mayo de 1984. La fig. 5.7 muestra el patrón de datos. Los datos no son estacionales y no presentan patrón de tendencia o cíclico. Por este motivo, se escogió

TABLA 5.13 ATENUACIÓN EXPONENCIAL

DATOS POR ATENUAR: MEDICINA NUCLEAR			FACTOR DE ATENUACIÓN = .3		
PERIODO	SERVICIO FUERA DE CONSULTORIO	VALOR ATENUADO	PERIODO	SERVICIO FUERA DE CONSULTORIO	VALOR ATENUADO
14	167.00	167.00	25	29.00	41.95
15	78.00	140.30	26	49.00	44.06
16	64.00	117.41	27	56.00	47.64
17	65.00	101.69	28	44.00	46.55
18	37.00	82.28	29	14.00	36.79
19	48.00	72.00	30	24.00	32.95
20	58.00	67.80	31	36.00	33.87
21	104.00	78.66	32	41.00	36.01
22	41.00	67.36	33	65.00	44.70
23	39.00	58.85	34	19.00	36.99
24	21.00	47.50	35	61.00	44.19

la atenuación exponencial como el método apropiado de pronóstico. Se encontró que un factor de .3 daba el mejor modelo. Este modelo se presenta en la tabla 5.13. El pronóstico de servicios de medicina nuclear para junio de 1984 es de 44.

Tomografías computarizadas

Se analizó también el número de tomografías practicadas en el periodo de julio de 1981 a mayo de 1984. Los datos se presentan en la fig. 5.8. No se encontró estacionalidad y no parece haber tendencia en el número de tomografías. Sin embargo, parece haber un patrón cíclico. Sería importante conocer cuántas

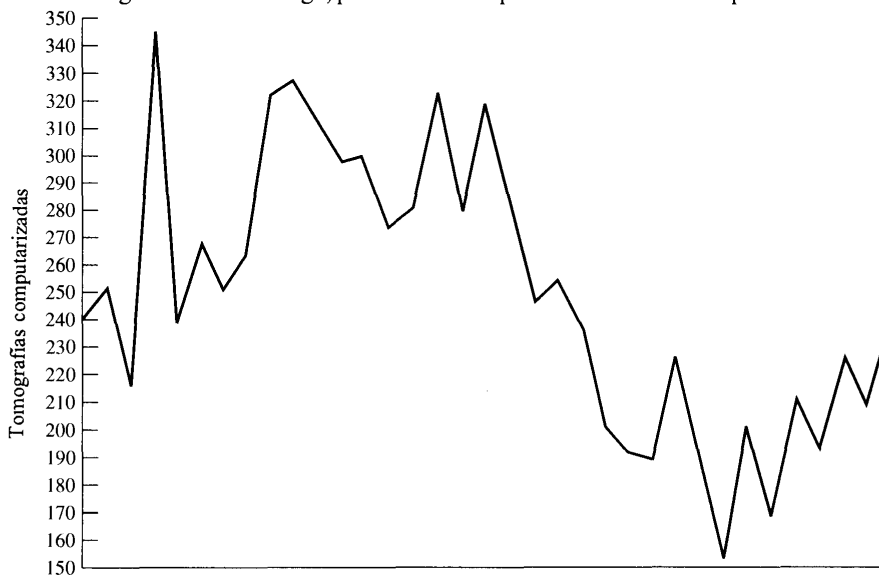


Figura 5.8 Tomografías computarizadas por mes, julio de 1981 a mayo de 1984.

tomografías se realizaron el mes pasado para el pronóstico de lo que sucederá este mes. Se examinó un modelo autorregresivo (caps. 9 y 10) y se comparó con un modelo de atenuación exponencial con una constante de atenuación de .5. Una constante de atenuación más grande asigna mayor ponderación en el pronóstico a las observaciones más recientes. Se determinó que era el mejor modelo y la tabla 5.14 muestra que la proyección del número de tomografías para junio de 1984 es de 223.

ANÁLISIS DEL ÁREA DE MERCADO

Se determinaron las áreas de mercado para los servicios que presta actualmente Downtown Radiology, examinando los registros de pacientes y los patrones de referencia de los doctores. Las áreas de mercado para los servicios que no presta Downtown Radiology en la actualidad, se determinaron investigando a la competencia y analizando las áreas geográficas que atienden.

Área de mercado para tomografías computarizadas

El área de mercado para tomografías computarizadas del centro médico de imágenes propuesta, 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 la proyección

TABLA 5.14 ATENUACIÓN EXPONENCIAL

DATOS POR ATENUAR: TOMOGRAFÍAS COMPUTARIZADAS			FACTOR DE ATENUACIÓN = .3		
PERIODO	TOMOGRAFÍAS COMPUTARIZADAS	VALOR ATENUADO	PERIODO	TOMOGRAFÍAS COMPUTARIZADAS	VALOR ATENUADO
1	238.00	238.00	19	277.00	290.80
2	250.00	244.00	20	244.00	267.40
3	214.00	229.00	21	253.00	260.20
4	344.00	286.50	22	234.00	247.10
5	236.00	261.25	23	199.00	223.05
6	266.00	263.63	24	190.00	206.53
7	248.00	255.81	25	187.00	196.76
8	261.00	258.41	26	226.00	211.38
9	322.00	290.20	27	190.00	200.69
10	327.00	308.60	28	151.00	175.85
11	312.00	310.30	29	201.00	188.42
12	297.00	303.65	30	166.00	177.21
13	299.00	301.33	31	210.00	193.61
14	272.00	286.66	32	191.00	192.30
15	279.00	282.83	33	226.00	209.15
16	322.00	302.42	34	207.00	208.08
17	276.00	289.41	35	238.00	223.04
18	320.00	304.60			

de los porcentajes correspondientes, el área de mercado para tomografías computarizadas tendrá en 1985 una población de 630 655 personas y en 1990, 696 018 personas.

Estimaciones cuantitativas

Para proyectar los ingresos es necesario determinar ciertas estimaciones cuantitativas. La estimación más importante comprende el número de doctores que participarán en la sociedad limitada. La estimación utilizada en cálculos posteriores es que participarán por lo menos el 8% de la población de doctores del condado de Spokane.

La siguiente incertidumbre que debe cuantificarse comprende la determinación de cómo se afectará el patrón de referencias con la participación de 50 doctores en la sociedad limitada. Se supone que 30 de los doctores que actualmente envían pacientes a Downtown Radiology, se unirán a la sociedad limitada. De los 30 que se adhieren, se supone que 10 no incrementarán sus referencias y que los otros 20 las duplicarán. También se supone que 20 doctores que nunca han enviado pacientes a Downtown Radiology se unirán a la sociedad limitada y comenzarán a enviar por lo menos la mitad de su trabajo a Downtown Radiology.

La cuantificación de referencias adicionales de doctores se aclarará con algunas observaciones cualitativas. El número de 50 doctores que se unirán a la sociedad limitada es conservador. Hay una fuerte posibilidad de que se unan doctores de áreas fuera del condado de Spokane. Tradicionalmente, el patrón de referencias de los doctores cambia muy lentamente. Sin embargo, es probable que la súbita naturaleza competitiva del mercado tenga un impacto en las referencias médicas. Si la sociedad limitada se ofrece a doctores con especialidades con un alto potencial de referencias radiológicas, el número de

referencias se incrementará más de lo proyectado. Es muy amplia la variabilidad en el número de referencias de doctores por servicio. Pocos doctores abarcan un gran porcentaje de los servicios de Downtown Radiology, si se recluta un nuevo número reducido de doctores con un alto porcentaje de referencias, se lograría un gran efecto en el número total de servicios prestados por Downtown.

Por último, se debe estimar el efecto que tendrá el nuevo centro de imágenes, en la participación de Downtown Radiology en el mercado. El nuevo centro tendrá el mejor equipo y estará preparado para efectuar todo el espectro de servicios a menor costo. El número de nuevos doctores que enviarán pacientes aumentará con base en la comunicación verbal de los nuevos doctores que inviertan. Si se formulan convenios entre el nuevo centro y las compañías de seguros, las empresas con un gran número de empleados y/o los sindicatos, Downtown podría incrementar su participación en el mercado por lo menos en 4% en 1985, 2% en 1986 y 1% en 1987, reteniendo esta participación en 1988 y 1989. En el resto de este informe, se hará referencia a este incremento en la participación en el mercado como el *efecto de imagen total*.

Proyecciones de ingresos

Se realizaron proyecciones de ingresos para cada servicio. En este caso de estudio, sólo se presentan las correspondientes a tomografías computarizadas.

Proyecciones de tomografías computarizadas

Con base en el modelo de atenuación exponencial y en lo sucedido en los primeros cinco meses de 1984, el pronóstico de tomografías computarizadas para 1984 (enero de 1984 a enero de 1985) es de 2600.

El National Center for Health Statistics reporta un índice de 261 tomografías por cada 100 mil habitantes, por mes. Utilizando la población proyectada de 630 655 personas para el área de mercado de tomografías computarizadas, el mercado sería de 19 752 servicios en 1985. El número real de tomografías computarizadas que se efectuaron en el área de mercado en 1983, se estimó en 21600. Esta estimación se basó en el conocimiento real de los servicios prestados por: Downtown Radiology (2600), Sacred Heart (4970), Deaconess (3 850), Valley (2 300) y Kootenai (1 820) así como estimaciones para Radiation Therapy (2 400) y Northwest Imaging (4 000). Si las estimaciones son precisas, Downtown Radiology tenía, en 1983, una participación aproximada en el mercado de 10.5%. También se analizaron los valores reales para 1982, resultando que Downtown Radiology tenía alrededor del 15.5% de participación en el mercado de tomografías computarizadas, en ese año. De ahí que se pronostica que Downtown Radiology promedia alrededor del 13% del mercado.

Con base en el incremento de referencias de doctores pertenecientes a la sociedad limitada y en un análisis del número de referencias promedio para tomografías computarizadas, se proyecta, a partir de esta fuente, un incremento de 320 tomografías computarizadas para 1985. Si se emplean los valores reales de 1983, la proporción para el área de mercado de tomografías computarizadas es de 3 568 (21 600/6.054) por cada 100 mil habitantes. Si este patrón continúa, el número de tomografías computarizadas en el área de mercado se incrementará a 22 514 ($3\,568 \times 6.31$) en 1985. De ahí que se proyecta que la participación en el mercado de Downtown Radiology será del 13% (2 920/22 514) Cuando aumenta al 4% la participación en el mercado con base en la suma total de imágenes, la participación en el mercado de Downtown Radiology aumenta al 17% y el número proyectado de tomografías computarizadas alcanzará 3 827 ($22\,514 \times .17$).

No obstante, la investigación parece indicar que el equipo MRI sustituirá eventualmente un gran número de tomografías computarizadas de la cabeza (*Applied Radiology* mayo/junio de 1983 y *Diagnostic Imaging*, febrero de 1984). El National Center for Health Statistics indica que el 60% de todas las tomografías computarizadas fueron de la cabeza. Los registros de Downtown Radiology mostraron que en 1982 el 59% de las tomografías computarizadas fueron de la cabeza y 54% en 1983. Si el 60% de las tomografías de Downtown son de la cabeza y el equipo MRI sustituye aproximadamente el 60% de éstas, se requieren nuevas proyecciones de este tipo de servicio para 1985. Considerando que el equipo MRI sólo funcionará la mitad del año, se pronostica una reducción de 689 tomografías ($3\ 827/2 \times .60 \times .60$).

El número proyectado de tomografías computarizadas para 1985 es de 3 138. El costo promedio de un estudio de este tipo es de \$360 dólares y el ingreso proyectado para los mismos es de \$1 129 680. En la tabla 5.15 se muestran los ingresos proyectados por tomografías computarizadas para los próximos cinco años. Se estima que el costo del servicio se incremente aproximadamente en un 11% anual.

Sin considerar el efecto del equipo MRI, la proyección de tomografías computarizadas para 1986 es de 4 363 ($6.31 \times 1.02 \times 3\ 568 \times .19$). Sin embargo, si el 60% de las tomografías son de la cabeza y el MRI absorbe el 70% de éstas, el número proyectado de tomografías se reduciría a 2 531 [$4\ 463 - (4\ 463 \times .60 \times .70)$].

TABLA 5.15 PROYECCIÓN DE INGRESOS DE CINCO AÑOS POR TOMOGRAFÍAS COMPUTARIZADAS

AÑO	SERVICIOS	INGRESOS
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

Sin considerar el efecto del equipo MRI, la proyección de tomografías computarizadas para 1987 es de 4 683 ($6.31 \times 1.04 \times 3\ 568 \times .20$). Considerando el efecto del equipo MRI, el número proyectado de tomografías se reduciría a 2 716 [$4\ 683 - (4\ 683 \times .60 \times .70)$].

Sin considerar el efecto del equipo MRI, la proyección de tomografías computarizadas para 1988 es de 4 773 ($6.31 \times 1.06 \times 3\ 568 \times .20$). Considerando el efecto del equipo MRI, el número proyectado de tomografías se reduciría a 2 482 [$4\ 773 - (4\ 773 \times .60 \times .80)$].

Sin considerar el efecto del equipo MRI, la proyección de tomografías computarizadas para 1989 es de 4 863 ($6.31 \times 1.08 \times 3\ 568 \times .20$). Considerando el efecto del equipo MRI, el número proyectado de tomografías se reduciría a 2 529 [$4\ 863 - (4\ 863 \times .60 \times .80)$].

PREGUNTA

El contador de Downtown Radiology proyectó que los ingresos serían considerablemente más altos. Debido a que la participación de propiedad se ofrecerá mediante algún tipo de oferta pública, la administración de Downtown Radiology debe tomar una decisión con respecto a la precisión de las proyecciones de Professional Marketing Associates. Se le pide analizar el informe. ¿Cuáles serían sus recomendaciones?

VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Se accesa el paquete de pronósticos mediante la selección de la opción 3 del menú principal. Si está usted empleando discos (no en un sistema de disco duro), el programa le solicitará que inserte el disco principal del sistema, con el mensaje:

```
Insert MAIN SYSTEM DISK in default drive
Press <ENTER> when ready
```

En este momento deberá sustituir el Data Management Disk por el Main System Disk. Presione <Enter> y el programa solicitará el nombre del archivo de datos a utilizar:

```
What is the name of the data file? CREDIT
```

El menú le permite elegir entre 13 diferentes técnicas de pronóstico mediante la siguiente pantalla:

```
Data file name:  B: CREDIT
Variable:        NUMBER OF CLIENTS

      *** Forecasting Analysis ***

  1. Simple Moving Average (unweighted)
  2. Single Exponential Smoothing
  3. Double Exponential Smoothing
  4. Holt's Two-Parameter Smoothing
  5. Harrison's Harmonic Smoothing
  6. Brown's Quadratic Exponential Smoothing
  7. Winter's Seasonal Smoothing
  8. Robust Decomposition
  9. Census X-11 Decomposition
 10. Regression Trend Analysis
 11. Multiple Regression Analysis
 12. Generalized Adaptive Filtering
 13. Box-Jenkins Analysis
 14. Return to Main Menu

Which?
```

El programa de análisis exploratorio de datos mostró que los datos de nuestro cliente presentaban una tendencia. Una técnica que elegimos intentar es la atenuación exponencial doble, opción 3. En la pantalla aparece:

NUMBER OF CLIENTS-DOUBLE EXPONENTIAL SMOOTHING

Smoothing Constant = 0.35 Lead Time = 1

RESIDUAL AUTOCORRELATION FUNCTION

Mean of the Residuals = .72

Standard Deviation of the Residuals = 26.9765

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.12	-1.20	:	[**:]	:
2	-0.06	-0.55	:	[*:]	:
3	-0.08	-0.77	:	[**:]	:
4	-0.05	-0.52	:	[*:]	:
5	-0.03	-0.33	:	[*:]	:
6	-0.10	-0.93	:	[**:]	:
7	-0.01	-0.10	:	[*:]	:
8	-0.03	-0.29	:	[**:]	:
9	-0.13	-1.19	:	[**:]	:
10	-0.03	-0.33	:	[*:]	:
11	0.03	0.31	:	[:*]	:
12	0.22	2.10	:	[:****]	:
13	-0.04	-0.36	:	[*:]	:
14	0.04	0.35	:	[:*]	:
15	-0.15	-1.32	:	[:****]	:
16	-0.01	-0.11	:	[:*]	:
17	0.18	1.55	:	[:****]	:
18	-0.17	-1.49	:	[:****]	:
19	-0.05	-0.43	:	[:*]	:
20	-0.00	-0.03	:	[:]	:
21	0.05	0.42	:	[:*]	:
22	-0.12	-0.97	:	[:****]	:
23	0.03	0.25	:	[:*]	:
24	0.26	2.13	:	[:****]	:
25	-0.04	-0.35	:	[:*]	:
26	-0.04	-0.34	:	[:*]	:
27	-0.08	-0.62	:	[**:]	:
28	0.08	0.61	:	[:**]	:
29	0.19	1.46	:	[:****]	:
30	-0.28	-2.14	:	[:*****]	:
31	0.08	0.56	:	[:**]	:
32	-0.17	-1.25	:	[:****]	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 32 Degrees for Freedom = 43.2

Probability = .089

NUMBER OF CLIENTS-DOUBLE EXPONENTIAL SMOOTHING

Smoothing Constant = 0.35 Lead Time = 1

SUMMARY STATISTICS

Mean of the Original Data = 122.15

Standard Deviation of the Original Data = 30.043

Mean of the Residuals = .72

Standard Deviation of the Residuals = 26.977

	Mean Absolute % Error	Mean % Error	Mean Square Error
Optimization Period	19.4225	1.0430	609.0
Remainder of Series	17.8037	-2.9088	757.1
Entire Series	18.2002	-1.9410	720.8

Figura 5.9 Atenuación exponencial doble para el caso de estudio 5.3.

*** Double Exponential Smoothing ***	
Smooth Constant:	A
Starting values [SO(1),SO(2):	A
Lead time:	1
Number of forecasts:	1
Time plot of original data, forecasts, and error:	Y
Residual autocorrelation function:	Y
Table of original data, forecasts, and error:	N
Summary statistics:	Y
Final smoothed statistics:	N
Press ENTER to accept:	

Para ejecutar la técnica de atenuación exponencial doble, podemos emplear las opciones por omisión que ya se muestran. Cuando se asigna la opción de automático (A) a la constante de atenuación, el programa intentará encontrar la mejor constante de atenuación en el nivel especificado (el valor por omisión para automático es de .1 a .9 con incrementos de .1). Para la atenuación exponencial doble, se requieren dos valores iniciales. La línea Lead Time, se refiere al periodo que se desea pronosticar. El valor por omisión 1 proporciona un pronóstico para el periodo inmediato siguiente. El número de pronósticos (Number of forecast) se refiere a cuántos pronósticos se desea que produzca el programa. Para nuestro ejemplo, podríamos haber elegido 9, para producir pronósticos para el resto de 1993.

A continuación, se presentan diversas opciones de salida. La fig. 5.9 muestra la salida cuando se responde Y a la función de autocorrelación residual y a las estadísticas de resumen y N a las otras opciones.

Por último el programa preguntará si desea almacenar los pronósticos de los residuales:

Do you want to save the forecasts of residuals (y/n)?

PRONÓSTICO MEDIANTE HOJAS DE CÁLCULO

ForeCalc es un programa de pronóstico que trabaja con datos registrados en hojas de cálculo Lotus 1-2-3. Inserte ForeCalc y enlázelo con su copia de Lotus 1-2-3 como un programa externo agregado. Esto amplía la capacidad de Lotus 1-2-3 al proporcionarle un conjunto adicional de menús. Forecalc emplea varios modelos diferentes de atenuación exponencial para el pronóstico de sus datos.

El problema. En los ejemplos 5.5 al 5.10, se utilizaron modelos de atenuación exponencial sobre los datos de ventas trimestrales de Acme Tool Company, de 1988 a 1993. Se requiere el pronóstico para 1994.

La solución mediante hoja de cálculo. Coloque el disco de ForeCalc en la unidad A y escriba INSTALLH después del indicador de comandos A:. Siga las indicaciones de instalación. Una vez concluida ésta, entre a Lotus 1-2-3 (versión 2.3) y capture los datos de la tabla 5.6 en las celdas A1 a A24 de la hoja de cálculo. Escriba /ADD-IN ATTACH ó /AA, lo que equivale a seleccionar los comandos ADD-IN FILE del menú principal y ATTACH del submenú. En la pantalla aparecerá la solicitud del nombre del archivo a incorporar:

```
Enter add-in to attach: G:\123R23\*.ADN
```

Se presentará una lista de programas que se pueden incorporar a 1-2-3. Seleccione:

```
FORECALC.ADN
```

El administrador de incorporación de programas presenta una lista de teclas de función disponibles:

```
No-Key 7 8 9 10
```

Si no desea asignar ninguna tecla de función al programa incorporado, seleccione NO-KEY.

En la línea de menú aparece el submenú de incorporación de programas. Seleccione INVOKE y elija FORECALC. En la línea de menú aparece el menú de ForeCalc:

```
Data Forecast Options Quit
```

1. Seleccione DATA RANGE o DR. Presione <Enter>. La pantalla solicita el rango de datos a considerar:

```
Enter data range: A1
```

Como los datos se encuentran en las celdas A1 a A24, escriba A1. . A24. La pantalla solicita el número de periodos a manejar por año:

```
Enter number of periods per year: 12
```

Ya que los datos son trimestrales, teclee un 4. La pantalla solicita el periodo inicial:

ForeCalc Forecasts			WINTERS
PERIOD	LOWER (95%)	FORECAST	UPPER (95%)
1/1994	608.236	756.359	904.482
2/1994	358.109	531.543	704.977
3/1994	207.819	404.522	601.225
4/1994	471.887	690.502	909.117

Exponential smoothing parameter values.

Level parameter:	0.591565
Trend parameter:	0.029564
Seasonal parameter:	0.366746

Initial point: 1/1988	Adjusted R-square: 0.799
Final point: 4/1993	Durbin-Watson: 1.527
Number of points: 24	Standard forecast error: 66.237
Mean value: 408.333	RMS Error: 61.959
Standard deviation: 150.843	MAP Error: 0.115
	Standardized BIC: 75.573

Figura 5.10 Atenuación exponencial de invierno para ventas de Acme Tools Company.

Enter starting period: 1

Los datos comienzan en el primer trimestre, así que teclee un 1. La pantalla solicita el año inicial:

Enter starting year: 1900

Como los datos comienzan en 1988, teclee 1988. En la pantalla reaparece el menú de ForeCalc.

- Después de registrar las especificaciones de sus datos, genere el pronóstico de la siguiente manera:

Seleccione: FORECAST

Seleccione: AUTOMATIC

Digite: 4 [número de periodos a pronosticar]

Presione: <Enter>

El programa elige automáticamente el mejor modelo a utilizar y optimiza los parámetros del modelo para asegurar un pronóstico preciso.

Seleccione: VIEW

En la pantalla (fig.5.10), se exhiben los pronósticos y sus límites de confianza asociados. ForeCalc también proporciona cierto número de estadísticas resumidas que describen qué tan bien se ajusta el modelo a su serie de datos.

Seleccione: DIAGNOSTICS

Estas estadísticas se muestran también en la fig.5.10. El programa ForeCalc permite además obtener gráficas e impresiones para cada una de las opciones.

BIBLIOGRAFIA

- AAKER, D. A. y JACOBSON, R. "The Sophistication of 'Naive' Modeling". *International Journal of Forecasting* 3 (314) (1987): 449-452.
- DALRYMPLE, D. J. y KING, B. E. "Selecting Parameters for Short-Term Forecasting Techniques." *Decision Sciences* 12 (1981): 661-669.
- GARDNER, E. S., Jr. y DANNENBRING, D. G. "Forecasting with Exponential Smoothing: Some Guidelines for Model Selection." *Decision Sciences* 11 (1980): 370-383.
- HARRISON, P. J. "Exponential Smoothing and Short-Term Sales Forecasting." *Management Science* 13 (11) (1967): 821-842.
- HOLT, C. C. "Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages." Office of Naval Research, Memorandum No. 52 (1957).
- LEDOLTER, J. y ABRAHAM, B. "Some Comments of the Initialization of Exponential Smoothing." *Journal of Forecasting* 3 (1) (1984): 79-84.
- MAKRIDAKIS, S. y WHEELWRIGHT, S. C. *Interactive Forecasting*, Palo Alto, Calif.: Scientific Press, 1977, pp. 145-152.
- MCKENZIE, E., "An Analysis of General Exponential Smoothing". *Operations Research* 24 (1976): 131-140.
- THIEL, H. y WAGE, S. "Some Observations on Adaptive Filtering." *Management Science*. 10 (2) (1964): 198 - 224.
- WINTERS, P. R. "Forecasting Sales by Exponentially Weighted Moving Averages." *Management Science* 6 (1960): 324-342.

Análisis de regresión

En el cap. 2, se expuso la relación lineal entre dos variables numéricas (correlación). Una vez establecida una relación lineal, se puede emplear el conocimiento de la variable independiente para pronosticar la variable dependiente. El modelo gráfico más sencillo para relacionar una variable dependiente con una sola variable independiente es una línea recta. En este capítulo se exponen los modelos de regresión lineal simple (línea recta) y cómo ajustarlos a un conjunto de puntos de datos mediante el método de mínimos cuadrados. Debido a que se requieren muchos cálculos en el análisis de regresión, se hace énfasis en aplicaciones de cómputo para las técnicas expuestas.

Para ampliar el análisis de la relación entre dos variables, que se abordó en el cap. 2, considere el siguiente ejemplo.

Ejemplo 6.1 Suponga que el Sr. Bump observa el precio y volumen de venta de galones de leche de 10 semanas elegidas en forma aleatoria. Los datos obtenidos se presentan en la tabla 6.1. Primero, el Sr. Bump genera un diagrama de dispersión de los datos, que se muestra en la fig. 6.1. En el diagrama parece ser que existe una relación lineal negativa entre Y , el número

TABLA 6.1 DATOS PARA EL EJEMPLO 6.1: GALONES DE LECHE

SEMANA	NIVEL SEMANAL DE VENTAS, Y (MILES DE GALONES)	PRECIO DE VENTA, X
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

de galones de leche vendidos y X , el precio de cada galón. Parece que al elevarse el precio, el volumen disminuye.

Ahora, el Sr. Bump desea medir el grado de esta relación aparente. Los cálculos que necesita se presentan en la tabla 6.2.

El coeficiente de correlación para esta muestra de datos se calcula a partir de la ecuación 2.16:

$$\begin{aligned}
 r &= \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{n\sum X^2 - (\sum X)^2} \cdot \sqrt{n\sum Y^2 - (\sum Y)^2}} \\
 &= \frac{(10)(149.3) - (14.4)(112)}{\sqrt{(10)(21.56) - 14.4^2} \cdot \sqrt{(10)(1,488) - 112^2}} \\
 &= \frac{1493 - 1612.8}{\sqrt{215.6 - 207.36} \cdot \sqrt{14,880 - 12,544}} \\
 &= \frac{-119.8}{\sqrt{8.24} \cdot \sqrt{2,336}} \\
 &= \frac{-119.8}{(287)(48.33)} = \frac{-119.8}{138.7} \\
 &= -.86
 \end{aligned}$$

La muestra de 10 puntos de datos revela un coeficiente de correlación simple de $-.86$, una relación lineal ciertamente negativa entre Y y X . De ahí que el Sr. Bump pueda concluir en forma tentativa que al aumentar el precio de un galón de leche, disminuye el número de galones vendidos.

TABLA 6.2 CÁLCULOS REQUERIDOS PARA EL EJEMPLO 6.1

	Y	X	XY	X^2	Y^2
$n = 10$	10	1.30	13.0	1.69	100
	6	2.00	12.0	4.0	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	14.4	149.3	21.56	1,488

La siguiente pregunta que pudiera presentarse es, ¿qué tanto disminuye el volumen al elevarse el precio? Esta pregunta sugiere dibujar una línea recta a través de los puntos exhibidos en el diagrama de dispersión. Después de dibujar esta línea, la pendiente de ésta mostrará la disminución promedio en Y para cada dólar de incremento en X . (Vea la fig. 6.1).

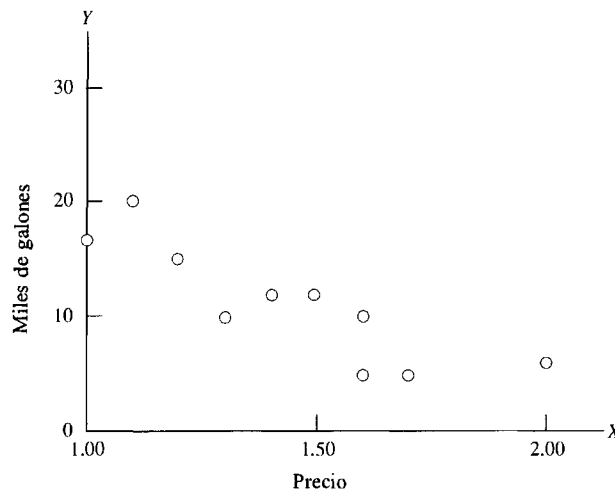


Figura 6.1 Diagrama de dispersión para el ejemplo 6.1.

LÍNEA DE REGRESIÓN

De hecho, el Sr. Bump hubiera podido dibujar una línea recta a través de los puntos de datos, intentando que la recta se ajustara lo más posible a los puntos. Sin embargo, un mejor procedimiento consiste en encontrar la ecuación de la línea que mejor se ajuste a los puntos. Esta línea recta es de la forma $\hat{Y} = b_0 + bX$. El primer término de esta ecuación, b_0 , se denomina *intersección de Y*, ya que es el valor que toma Y cuando X es igual a cero. Al segundo término, b , se le denomina *pendiente* de la recta, pues representa la cantidad de cambio en Y al incrementar X en una unidad (recuerde la exposición sobre la pendiente en el cap. 2). Entonces, lo que el Sr. Bump desea hacer es calcular los valores de b_0 y de b .

Pero antes es necesario definir qué significa que una línea se ajuste “mejor” a los puntos de datos reunidos. Existen muchas definiciones, pero la que se utiliza casi universalmente en el análisis de regresión es:

La línea que mejor se ajuste a un conjunto de puntos de datos X - Y , es aquella que minimiza la suma de las distancias al cuadrado de los puntos a la línea, medidas en dirección vertical o hacia Y . A esta línea se le conoce como *la línea de regresión* y su ecuación se denomina *ecuación de regresión*.

Bajo esta definición, el método que se emplea para encontrar los valores de b_0 y b se conoce como *método de mínimos cuadrados*. Para encontrar las ecuaciones para b y b_0 se requiere el uso del cálculo. La derivada de estas dos ecuaciones aparece en el Apéndice A, con los siguientes resultados:

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} \quad (6.1)$$

$$b_0 = \frac{\sum Y}{n} - \frac{b\sum X}{n} \quad (6.2)$$

Ejemplo 6.2

Los cálculos para los datos que se presentan en la tabla 6.2 son

$$b = \frac{(10)(149.3) - (14.4)(112)}{(10)(21.56) - (14.4)^2} = \frac{-119.8}{8.24} = -14.54$$

$$b_0 = \frac{112}{10} - (-14.54)\left(\frac{14.4}{10}\right) = 11.2 + 14.54(1.44) = 32.14$$

De esta manera, la ecuación de la recta que mejor se ajusta a los puntos de datos reunidos, bajo el método de mínimos cuadrados es

$$\begin{aligned} \hat{Y} &= b_0 + bX \\ &= 32.14 - 14.54X \end{aligned} \quad (6.3)$$

Esta ecuación se denomina *ecuación de regresión de la muestra*.

El Sr. Bump desearía ahora interpretar los valores en esta ecuación. La intersección de Y , b_0 , es el valor de Y cuando X es igual a cero. Una estricta interpretación sugeriría que el número de galones vendidos cuando $X = 0$ (esto es, si el precio de un galón de leche fuera cero), es de 32 140 galones. Esta interpretación no concuerda con el sentido común, ya que uno esperaría que se “vendiera” más leche si fuera gratis. El problema aquí ilustrado comprende la predicción de un valor para Y con base en un valor de X para el que no se ha reunido evidencia en la muestra. Es decir, ninguno de los puntos de la muestra tiene un valor de cero o cercano a cero para X . En este caso, como en muchos otros de análisis de regresión, no es posible hacer una interpretación útil de la intersección de Y . En términos más generales, no es prudente predecir valores de Y para cualquier X que esté más allá del nivel de las X reunidas en los datos de la muestra.

El valor de la pendiente, b , se pudiera interpretar como el cambio promedio en Y , que ocurre al incrementar X en 1. En este ejemplo, Y disminuye en un promedio de 14.54 (es decir, se venden 14 450 galones menos) cuando X aumenta en 1 (el costo de un galón de leche se eleva en \$1). Cada dólar de incremento en un galón de leche reduce la cantidad comprada en un promedio de 14 450 galones. O, para poner esta declaración en unidades más comprensibles, la evidencia de la muestra indica que cada incremento de 1 centavo en un galón de leche, reduce la cantidad comprada en un promedio de 145.4 galones.

Se puede ilustrar la relación X - Y dibujando, en el diagrama de dispersión, la línea recta que mejor se ajusta a los puntos de datos, como se muestra en la fig. 6.2. Nótese que las distancias verticales de los puntos hacia la línea se muestran como líneas punteadas. Si se elevaran al cuadrado estas distancias y se sumaran, esta suma sería la más baja posible para cualquier línea que pudiera dibujarse a través de estos puntos. Cualquier otra línea dibujada a través de estos puntos tendría una suma mayor de distancias al cuadrado. De modo que, de acuerdo con el procedimiento de mínimos cuadrados, esta línea representa el mejor ajuste para los 10 puntos de la muestra.

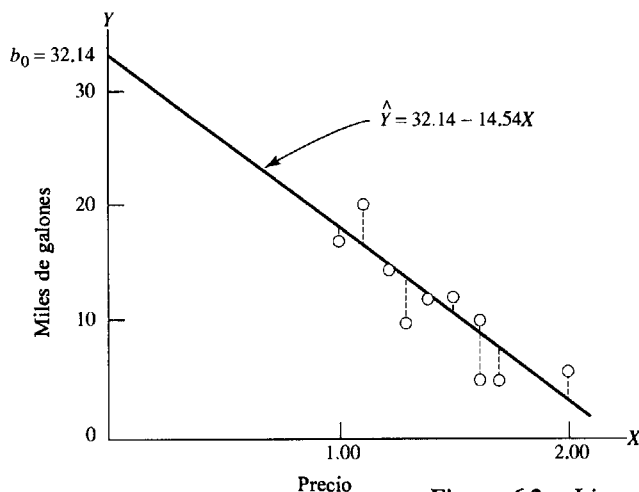


Figura 6.2 Línea de regresión para los datos del Sr. Bump.

En este punto, se debe destacar que a la línea de regresión calculada se le denomina correctamente línea de regresión de la *muestra*, debido a que es la línea que mejor se ajusta a los puntos de datos de la *muestra*, extraídos en forma aleatoria de la población de dichos puntos de datos. Como es común en los procesos de estadísticas, las estadísticas distinguen entre los valores de datos de la población y los valores de datos de una muestra. Para mantener claramente separados estos dos grupos, se emplea una notación diferente para describir las relaciones existentes entre X y Y , como se muestra en la tabla 6.3.

Los coeficientes de correlación de la población y de la muestra que aparecen en la tabla 6.3, se abordaron en el cap. 2. La ecuación de regresión de la muestra es la ecuación de la recta que mejor se ajusta a los puntos de datos observados de la muestra. En forma subyacente a la relación X - Y revelada por los valores de datos de la muestra, se encuentra la relación X - Y de toda la población, representada por el modelo de regresión de la población. Este modelo sugiere que cada valor de Y en la población es igual al valor de intersección β_0 , más una constante β veces su valor de X , más un término de error ϵ . Entonces, los términos de error son las diferencias entre los valores reales de la población y los valores representados por la ecuación de regresión del problema o $\epsilon = Y - (\beta_0 + \beta X)$. En una muestra, estas diferencias o errores se denominan *residuos*.

La mayoría de los paquetes de cómputo realizan correlación y análisis de regresión. Al final del capítulo, en la sección "Paquete de cómputo Minitab", se presentan las instrucciones de Minitab para los ejemplos 6.1 y 6.2.

TABLA 6.3 NOTACIÓN PARA LA POBLACIÓN Y LA MUESTRA

POBLACIÓN	MUESTRA
Coefficiente de correlación = ρ	Coefficiente de correlación = r
Modelo de regresión de la población:	Ecuación de regresión de la muestra:
$Y = \beta_0 + \beta X + \epsilon$	$\hat{Y} = b_0 + bX$

ERROR ESTÁNDAR DE LA ESTIMACIÓN

Después de calcular la ecuación de la línea de regresión de la muestra, el Sr. Bump podría interesarse en medir hasta qué grado los puntos de datos de la muestra se encuentran dispersos alrededor de la línea de regresión de la muestra. En particular, quisiera calcular la distancia típica de un punto de datos a la línea de regresión, medida en dirección de Y . Este concepto de medición de dispersión es similar a la noción de desviación estándar empleada para medir la dispersión de valores de datos alrededor de su media. En el análisis de regresión, la medición se denomina *error estándar de la estimación* y está representado por el símbolo $s_{y \cdot x}$. La ecuación correspondiente es

$$s_{y \cdot x} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} \quad (6.4)$$

El *error estándar de la estimación* mide la cantidad estándar en la cual los valores reales de Y difieren de los valores estimados ($\hat{Y}$).¹

Nótese la similitud entre la ecuación 6.4 y la fórmula de desviación estándar de la muestra en la estadística descriptiva simple. Véase cap. 2. En el numerador se encuentran las mediciones reales de datos (Y) y el valor “promedio” o “esperado” de la variable ($\hat{Y}$) como se determina mediante la línea de regresión de la muestra. Estas distancias se elevan al cuadrado, se suman y se dividen entre los grados de libertad, y se obtiene la raíz cuadrada. (Nótese que se pierden dos grados de libertad ya que se están estimando dos parámetros de la población mediante valores de datos de la muestra, a saber, b_0 y b .) Así, el error estándar mide la distancia típica o estándar desde un punto de datos a la línea de regresión de la muestra. Un análisis de regresión que tenga un error estándar pequeño implica puntos de datos muy cercanos a la línea de regresión. Por el contrario, un error estándar grande significa que los puntos de datos están muy dispersos alrededor de la línea.

Para fines de cálculo, la ecuación 6.4 se puede convertir en

$$s_{y \cdot x} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b \sum XY}{n - 2}} \quad (6.5)$$

Para el ejemplo del Sr. Bump, el error estándar es

¹ El error estándar de la estimación también se puede definir como la desviación estándar de las distribuciones supuestas de Y para cualquier X dada.

$$\begin{aligned}
 s_{y \cdot x} &= \sqrt{\frac{1,488 - (32.14)(112) + (14.54)(149.3)}{8}} \\
 &= \sqrt{\frac{59.14}{8}} = \sqrt{7.39} \\
 &= 2.72
 \end{aligned}$$

El error estándar es una medición de la distancia vertical estándar desde los puntos de datos de la muestra hacia la línea de regresión de la muestra. Se puede utilizar para comparar la dispersión de puntos de datos alrededor de la línea de regresión que se encuentran en esta situación, con la dispersión de puntos de datos en otras situaciones. Los analistas experimentados lo emplean para juzgar hasta qué grado se pudieran resumir los puntos de datos de la muestra mediante la línea de regresión de mejor ajuste. Además, se supone que la población de puntos de datos presenta una distribución normal alrededor de la línea de regresión de la población. Se puede usar $s_{y \cdot x}$ como una estimación de la desviación estándar de esa distribución normal. La importancia de realizar tal suposición de normalidad se abordará más adelante en este capítulo.

PREDICCIÓN DE Y

La siguiente tarea consiste en estimar el valor de Y para un valor dado de X . Para obtener una *predicción de punto*, se emplea la ecuación de regresión. Se sustituye el valor de X en la ecuación y se encuentra el valor de predicción para Y .

Ejemplo 6.3 Suponga que el Sr. Bump desea pronosticar la cantidad de leche vendida si el precio se estableciera en \$1.63 dólares. De la ecuación 6.3 tenemos:

$$\begin{aligned}
 \hat{Y} &= 32.14 - 14.54 X \\
 &= 32.14 - 14.54(1.63) \\
 &= 8.440 \quad \text{o} \quad 8\,440 \text{ galones}
 \end{aligned}$$

Obsérvese que este estimado es el valor de $\hat{Y}$; es decir, representa el punto sobre la línea de regresión en donde $x = 1.63$.

Desde luego, el Sr. Bump se da cuenta de que los puntos de datos que generaron la línea de regresión se encuentran dispersos alrededor de esa línea, según la medición por medio de $s_{y \cdot x}$. Para hacer una *predicción de intervalo* de Y cuando $X = 1.63$, se debe tomar en cuenta esta dispersión. Además, se debe considerar un segundo factor antes de hacer tal predicción.

Anteriormente se estableció que la línea de regresión calculada era una línea de regresión de la muestra, ya que se calculó a partir de una muestra aleatoria de diez puntos de datos y no de todos los puntos de datos de la población. Otras muestras aleatorias de diez puntos de datos podrían producir diferentes líneas de regresión de la muestra, de forma similar al caso donde varias muestras extraídas de una población tienen medias diferentes. De modo que para hacer una predicción de intervalo de Y se deben considerar tanto la dispersión de los puntos de datos de la muestra alrededor de la línea de regresión de la muestra, como la dispersión de varias líneas de regresión de la muestra alrededor de la verdadera línea de regresión de la población.

El *error estándar de pronóstico* mide la variabilidad de los valores de predicción de Y alrededor del verdadero valor de Y para un valor dado de X ; tomando en cuenta los dos factores mencionados. El error estándar de pronóstico es

$$s_f = s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2}} \quad (6.6)$$

Nótese que $s_{y \cdot x}$ es un término en este error estándar, ya que el primer término bajo el signo radical es 1. Los otros factores miden la variabilidad de las líneas de regresión de la muestra alrededor de la verdadera línea de regresión. Nótese también que en la ecuación 6.6 aparece el valor particular de X que se está utilizando. En otras palabras, el error de pronóstico *depende* del valor de X para el cual se desea un pronóstico. El error de pronóstico más pequeño posible es $X = \bar{X}$, pues así el tercer término bajo el radical equivale a cero $(X - \bar{X})^2 = 0$. Entre más alejado esté X de $\bar{X}$, mayor será el error de pronóstico.

Ejemplo 6.4

De manera gráfica, el intervalo de predicción del 95% del Sr. Bump para varios valores de X , se vería similar al de la fig. 6.3.

En forma más específica, el error estándar del pronóstico para $x = 1.63$ se calcula como:

$$\begin{aligned} s_f &= 2.72 \sqrt{1 + \frac{1}{10} + \frac{(1.63 - 1.44)^2}{.8240}} \\ &= 2.72(1.066) = 2.90 \end{aligned}$$

Nota: En este cálculo, $\sum(X - \bar{X})^2$ se determinó como se muestra en la tabla 6.4.

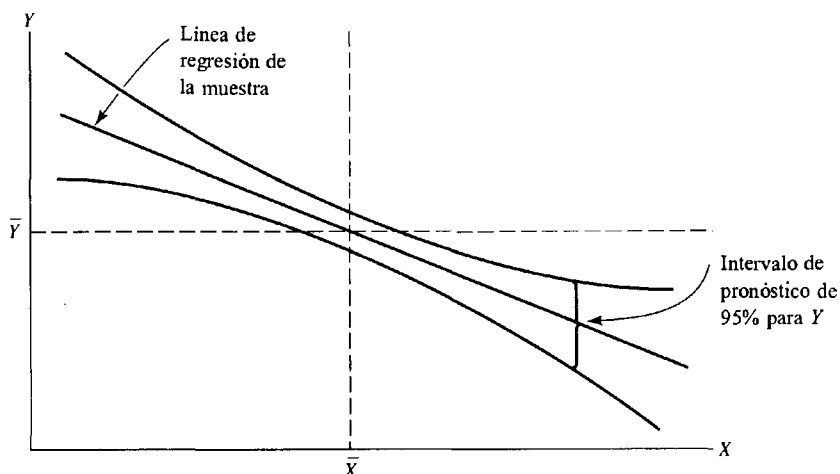


Figura 6.3 Intervalo de pronóstico para los datos del Sr. Bump.

TABLA 6.4 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
	$\Sigma(X - \bar{X})^2 = .8240$

Mediante la ecuación 6.7 se calcula un intervalo de predicción de 95% para Y cuando $X = 1.63$. Esta ecuación se puede emplear siempre que el tamaño de la muestra sea suficientemente grande ($n \geq 30$).

$$\hat{Y} \pm Zs_f \quad (6.7)$$

Al calcular un intervalo de predicción, se hacen suposiciones importantes. De hecho, todo el procedimiento de análisis de regresión se apoya sobre las siguientes suposiciones básicas:

1. *La población de Y valores presenta una distribución normal con respecto a la línea de regresión de la población.* Esta condición se muestra en la fig. 6.4. En la práctica, se obtienen resultados razonablemente precisos siempre que los valores de Y tengan una distribución aproximadamente 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.* Esto es, la varianza de la población no es más grande ni más pequeña al incrementarse el valor de los puntos de datos de X . La violación a esta suposición se conoce como *heterosedasticidad* y se aborda en detalle en el cap. 7; en el cap. 9 se presenta un ejemplo de esta condición y su cura.
3. *Los términos de error (ϵ) son independientes entre si.* Esta suposición implica una muestra aleatoria de puntos de datos X - Y . La violación a esta suposición se conoce como *correlación serial* y ocurre con frecuencia cuando los valores de datos X - Y se miden a través del tiempo. En el cap. 9 se trata la identificación y solución de este problema.
4. *Existe en la población una relación lineal entre X y Y .* Las líneas de regresión no lineal constituyen una ampliación al caso lineal expuesto en este capítulo. Existen técnicas para manejar relaciones no lineales entre X y Y ; algunas de ellas se comentan más adelante en este capítulo.

En realidad la predicción de intervalo dada por la ecuación 6.7 sólo es aproximadamente correcta para el ejemplo del Sr. Bump. Para un uso correcto del valor de Z de la

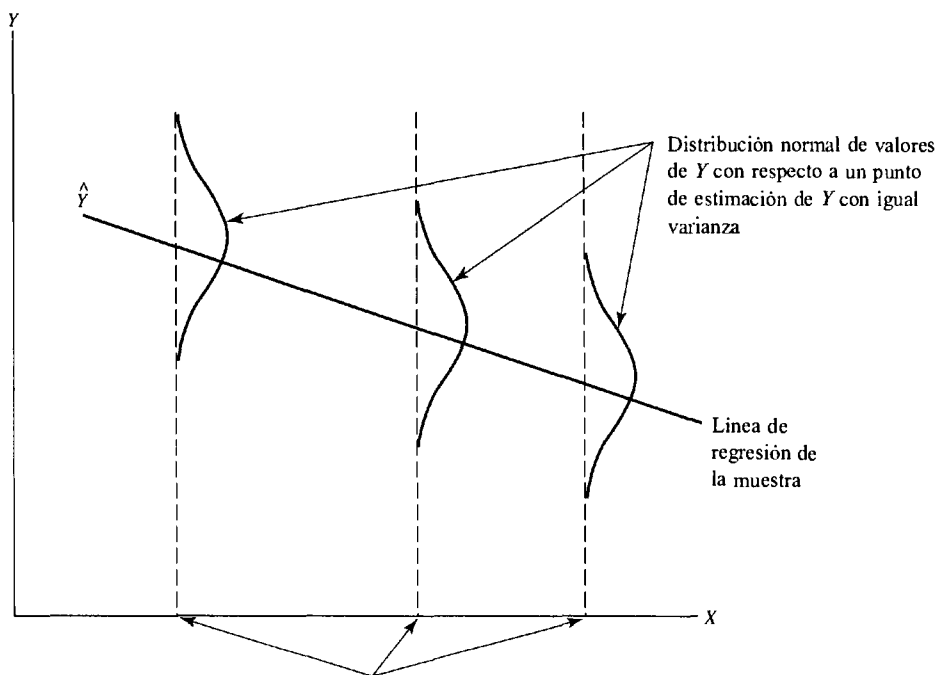


Figura 6.4 Suposición de normalidad.

tabla de la curva normal, no debe haber un tamaño de muestra reducido. Como regla de oro, se requiere de un tamaño de muestra de por lo menos 30, para utilizar la tabla normal. Para tamaños de muestra reducidos resulta apropiado calcular el intervalo de predicción mediante la distribución t . En este caso, el intervalo se convierte en

$$\hat{Y} \pm ts_f \quad (6.8)$$

(Para la tabla t , $df = n - 2$.)

Para los datos del Sr. Bump, un conjunto de 10 puntos de datos, resulta apropiada la fórmula de distribución t . Así, el intervalo es

$$\hat{Y} \pm ts_f$$

$$8.44 \pm 2.306(2.90)$$

$$8.44 \pm 6.69$$

$$1.75 \text{ a } 15.13 \quad \text{o bien} \quad 1\,750 \text{ galones a } 15\,130 \text{ galones}$$

La ecuación 6.8 predice un intervalo en cierta forma más amplio que el producido si se empleara la curva normal, lo que refleja la presencia de un tamaño de muestra reducido. Nótese la amplitud extrema del intervalo de confianza calculado. Es tan amplio que resulta virtualmente inútil en la predicción de Y . Aunque esta imprecisión se hace muy notoria en la amplitud del intervalo, no lo es en el punto de estimación calculado a partir de la ecuación

de regresión. Este grado de precisión implícito es el mayor beneficio de la estimación del intervalo.

Otro punto importante a recordar es que no resulta prudente predecir más allá del nivel de datos observados. Por lo tanto, se justifica que el Sr. Bump intente predecir Y cuando $x = 1.63$, ya que algunos de los valores originales de X se encuentran cerca de 1.63. Por otra parte, no haría bien en predecir Y cuando $x = 3.00$. No se han reunido valores de datos cercanos a este alto valor y, por esta razón, cualquier predicción que comprenda dicho valor de X sería altamente sospechosa. Para estimar la cantidad vendida cuando el precio unitario es de \$3 dólares, el Sr. Bump debe suponer que el modelo lineal aún es válido. Pudiera tener una buena razón para suponer esto, pero no posee evidencia directa que lo sustente.

COEFICIENTE DE DETERMINACIÓN

A continuación se expone una útil estadística denominada *coeficiente de determinación*.

Considere la fig. 6.5, que muestra la línea de regresión de la muestra del Sr. Bump junto con un punto de datos hipotético. Al considerar el punto de datos individual que se ubica bajo la línea de regresión, el Sr. Bump primero “esperaría” que este valor de Y fuera igual al promedio de todos los valores de Y , es decir $\bar{Y}$. Así, la distancia vertical entre Y y $\bar{Y}$ que se muestra en la fig. 6.5, se convierte en la diferencia total o desviación que debe explicarse.

Parte de esta distancia vertical es fácilmente explicable. Se esperaría que el valor de Y fuera igual a $\bar{Y}$, sólo si el valor de X de la observación fuera igual a $\bar{X}$. Sin embargo, éste no es el caso. El valor de X para esta observación es considerablemente más alto que $\bar{X}$. Por lo tanto, es comprensible por qué el valor de Y de esta observación pudiera haberse movido de $\bar{Y}$ a $\hat{Y}$: el valor de X de la observación es mayor que $\bar{X}$, y es sabido que X y Y

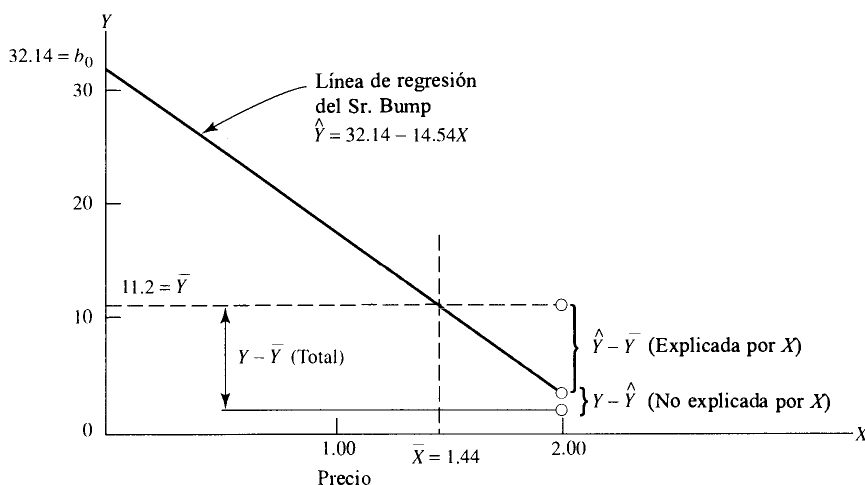


Figura 6.5 Variación explicada y no explicada para los datos del Sr. Bump.

tienen una fuerte correlación negativa ($r = -0.86$). De ahí que la distancia vertical $Y - \bar{Y}$ se "explique" mediante el movimiento de X , mientras que la distancia $Y - \hat{Y}$ no se explique por el movimiento de X .

Ejemplo 6.5

Para los datos del Sr. Bump, considere un punto de datos con un mayor valor de X ($x = 1.70$, $y = 5$). Para este punto de datos.

$$\begin{aligned}\hat{Y} &= 32.14 - 14.54(1.70) \\ &= 32.14 - 24.72 = 7.422\end{aligned}$$

Por lo tanto

$$\begin{aligned}\bar{Y} &= 11.2 \\ \hat{Y}_R &= 7.422 \\ y &= 5\end{aligned}$$

Estos valores se presentan en la fig. 6.6.

En este caso, la distancia de $\bar{Y}$ a $\hat{y}$ se explica mediante el movimiento de $\bar{X}$ (1.44) a su valor real (1.70), como se esperaba de acuerdo con el conocimiento de la relación entre Y y X ($r = -0.86$). En otras palabras, el movimiento de X de 1.44 a 1.70 explica el movimiento de y de 11.2 a 7.422, pero *no* explica el movimiento de 7.422 a 5.00.

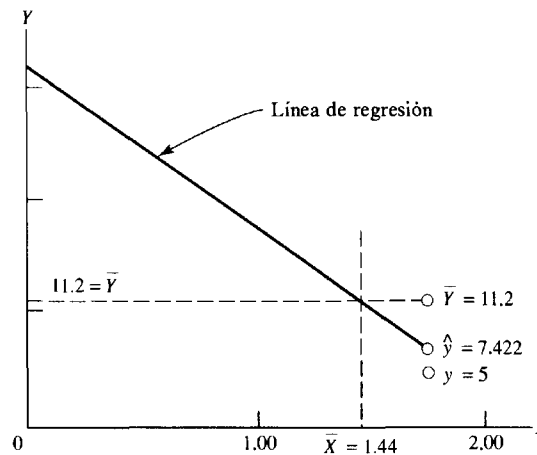


Figura 6.6 Pronóstico de Y para el ejemplo 6.5.

Para encontrar la porción de la desviación total que se explica mediante X para cualquier observación, se debe dividir la desviación explicada entre la desviación total, de la manera siguiente:

Porcentaje de desviación total explicada por

$$X = \frac{\hat{Y} - \bar{Y}}{Y - \bar{Y}}$$

Porcentaje de desviación total no explicada por

$$X = \frac{Y - \hat{Y}}{Y - \bar{Y}}$$

Por supuesto, se deben considerar todos los puntos de datos, no sólo uno. Además, el cálculo para todos los puntos de datos se realiza con diferencias al cuadrado. Este cálculo se hace de modo que el resultado final tenga una interpretación útil.

El símbolo del coeficiente de determinación de la muestra es r^2 y su valor se define como

$$r^2 = 1 - \frac{\text{Desviación no explicada}}{\text{Desviación total}} = 1 - \frac{\sum(Y - \hat{Y})^2}{\sum(Y - \bar{Y})^2} \quad (6.9)$$

El *coeficiente de determinación* mide el porcentaje de variabilidad en Y que puede explicarse a través del conocimiento de la variable independiente X .

La ecuación 6.9 es la forma usual de definir r^2 , y establece que el porcentaje de variabilidad en Y que se explica mediante la variabilidad en X , es 100%, menos el porcentaje no explicado por X . El valor de r^2 es un valor muy importante en cualquier análisis de regresión, ya que muestra el grado hasta el cual están relacionadas la variabilidad de X y Y .

En la práctica, r^2 se puede calcular mediante la siguiente fórmula equivalente en el cálculo:

$$r^2 = \frac{b_0 \sum Y + b \sum XY - n \bar{Y}^2}{\sum Y^2 - n \bar{Y}^2} \quad (6.10)$$

Ejemplo 6.6

Para los datos del Sr. Bump, r^2 se calcula con la ecuación 6.10.

$$\begin{aligned} r^2 &= \frac{(32.14)(112) - (14.54)(149.3) - (10)(11.2)^2}{1,488 - (10)(11.2)^2} \\ &= \frac{3,599.68 - 2,170.82 - 1,254}{1,488 - 1,254} = \frac{174.9}{234} = .747 \end{aligned}$$

Con base en este valor de r^2 , se formulan las siguientes declaraciones:

1. De la variabilidad en la cantidad de leche vendida, el 75% se puede explicar por la variabilidad en el precio de la leche.
2. De la variabilidad en la cantidad de leche vendida, el 25% *no se puede* explicar mediante la variabilidad en el precio de la leche. Esta porción en la variabilidad del volumen de leche se debe explicar mediante factores que aún no han sido identificados en este análisis de regresión (v. gr., clima, cantidad de publicidad, disponibilidad de productos sustitutos).

Nótese que r^2 es el cuadrado de r , el coeficiente de correlación (ignorando errores de redondeo). Esto es

$$\text{Coeficiente de determinación} = (\text{coeficiente de correlación})^2$$

$$r^2 = (r)^2$$

Por lo tanto

$$.75 = (-.86)^2$$

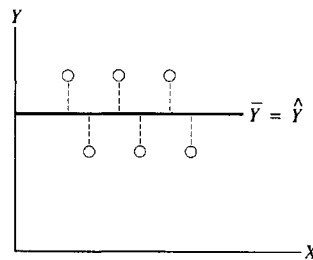
$$.75 = .75$$

Entonces, ¿por qué es necesario identificar ambos valores en un análisis de regresión? La respuesta es que cada uno tiene una ventaja sobre el otro.

La ventaja del coeficiente de correlación (r) es que revela relaciones tanto positivas como negativas. En el caso de los datos reunidos por el Sr. Bump, existe una relación negativa ($r = -.86$). En otros casos, se podría revelar una relación positiva mediante el valor de r . Como se verá en el siguiente capítulo, es importante identificar las relaciones existentes entre ciertos pares de variables, cuando se confrontan con un gran conjunto de variables y es necesario saber si existen relaciones positivas o negativas. Nótese que cuando se eleva al cuadrado el coeficiente de correlación, el valor es siempre positivo, perdiéndose la naturaleza de la relación.

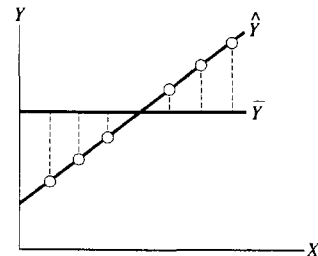
La ventaja del coeficiente de determinación (r^2) consiste en que tiene una interpretación muy útil. El valor de r^2 mide el porcentaje de variabilidad en Y que se explica por la variabilidad en X . Esta interpretación es muy útil, lo que hace de r^2 una de las estadísticas consultadas con mayor frecuencia en el análisis de regresión.

La fig. 6.7 ilustra dos casos extremos para r^2 , cuando es igual a cero y cuando es igual a 1. En el primer caso, *ninguna* variabilidad de las Y se explica mediante X : el diagrama de dispersión sugiere que no existe relación lineal entre X y Y . Cuando r cuadrada = 1, *toda* variabilidad de las Y se explica cuando se conoce X : todos los puntos de datos de la muestra se ubican sobre la línea de regresión.



$$\begin{aligned} r^2 &= 1 - \Sigma(Y - \hat{Y})^2 / \Sigma(Y - \bar{Y})^2 \\ &= 1 - 1 = 0 \end{aligned}$$

(a) Sin correlación lineal



$$\begin{aligned} r^2 &= 1 - \Sigma(Y - \hat{Y})^2 / \Sigma(Y - \bar{Y})^2 \\ &= 1 - 0 = 1 \end{aligned}$$

(b) Correlación lineal perfecta

Figura 6.7 Valores extremos de r^2 .

RESIDUOS

El análisis del Sr. Bump comenzó con sólo 10 cantidades de volumen de venta semanal (la variable Y). Si hubiera disponible más información, el Sr. Bump podría emplear la media $\bar{Y}$ de Y como pronosticador de cada volumen de venta semanal. Si lo hace, generará ciertos errores, uno por cada predicción que haga. Estos errores, denominados residuos, se expusieron anteriormente en este capítulo. Al estimar los valores de Y con $\bar{Y}$, el Sr. Bump minimiza los errores de estimación sobre un largo periodo, ya que emplea el volumen medio de ventas, 11 200 galones, como la mejor estimación del volumen de venta de cada semana. Los errores suman cero, ya que la media es el centro matemático de la distribución de Y . Las diez predicciones y los diez residuos se muestran en la tabla 6.5.

Además de los residuos, la tabla 6.5 muestra la suma de los residuos elevados al cuadrado. Este cálculo se efectúa con el propósito de poder medir la cantidad total de error; debido a que hay residuos tanto positivos como negativos, la suma de los residuos reales siempre es cero. Al valor de la sumatoria se le denomina *suma de residuos cuadrados* y representa la variabilidad total de Y alrededor de la media. Esta es la variabilidad total (233.60) que el Sr. Bump desea reducir, introduciendo el conocimiento de una variable relacionada, X . La fórmula para obtener la suma de los residuos cuadrados alrededor de la media de Y , o la suma total de cuadrados es

$$\Sigma(Y - \bar{Y})^2 \quad (6.11)$$

La suma de los residuos cuadrados se puede convertir en una varianza, dividiéndola entre el número de grados de libertad. Como se está estimando un parámetro de la población (μ) mediante una estadística de la muestra ($\bar{Y}$), se pierde un grado de libertad. Así, el grado

TABLA 6.5 RESIDUOS PARA LOS DATOS DEL SR. BUMP

Y REAL	PRONÓSTICO DE Y ($\bar{Y}$)	RESIDUO ($Y - \bar{Y}$)	$(Y - \bar{Y})^2$
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
		Sumas	233.60

de libertad apropiado es $n - 1$. De modo que la varianza total de Y alrededor de su media es

$$s_y^2 = \frac{\sum (Y - \bar{Y})^2}{n - 1} \quad (6.12)$$

Nota: La raíz cuadrada de esta varianza es la desviación estándar de Y .

Ejemplo 6.7

Para los datos del Sr. Bump,

$$s_y^2 = \frac{233.6}{9} = 25.96$$

Nótese el siguiente punto importante: esta varianza de los valores de Y se calculó usando $\bar{Y}$ como el mejor estimador de los valores semanales de Y . Esto es, no se ha utilizado información acerca de la variable relacionada X . Cuando el Sr. Bump incluye información con respecto a X , resultaría una menor varianza de Y puesto que X está correlacionada con Y y se podrían efectuar predicciones más precisas mediante la ecuación de regresión.

Suponga ahora que el Sr. Bump utiliza la ecuación de regresión calculada para predecir Y . Como él estableció una relación entre X y Y con base en los puntos de datos de la muestra ($r = -.86$), la línea de regresión debería generar predicciones más precisas para Y que las generadas cuando se utilizó $\bar{Y}$ para predecir los valores de Y . Si el Sr. Bump predice Y mediante esta ecuación de regresión, aún habrá residuos, aunque él espera que sean menores. En forma similar, tanto la suma de residuos cuadrados como la varianza de

TABLA 6.6 ANÁLISIS DE LOS RESIDUOS DE LOS DATOS DEL SR. BUMP

X	Y	PRONÓSTICO DE $Y(\hat{Y})$	RESIDUO	$(Y - \hat{Y})^2$
		USANDO $\hat{Y} = 32.14 - 14.54X$	$(Y - \hat{Y})$	
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
		Sumas	0.000	59.41

los valores de Y alrededor de la línea de regresión deberían ser menores. Se infiere que r^2 será más grande. La tabla 6.6 presenta este análisis de residuos.

En el cuadro 6.6 se puede apreciar que la suma de residuos cuadrados se reduce de 233.6 a 59.4; la reducción ocurre debido a que los residuos ahora comprenden la variabilidad de los valores de Y alrededor de la línea de regresión, en vez de alrededor de Y . En otras palabras, el conocimiento de la variable relacionada X reduce el error de la predicción, tal como el Sr. Bump pudiera esperar, a través del conocimiento del coeficiente de correlación de la muestra entre X y Y . La suma de residuos cuadrados es

$$\Sigma(Y - \hat{Y})^2 \quad (6.13)$$

La varianza de Y alrededor de los valores de predicción es

$$s_{y \cdot x}^2 = \frac{\Sigma(Y - \hat{Y})^2}{n - 2} \quad (6.14)$$

Ejemplo 6.8 Para los datos de el Sr. Bump

$$s_{y \cdot x}^2 = \frac{59.41}{8} = 7.43$$

Nótese que este resultado es el cuadrado del error estándar de la estimación, $s_{y \cdot x}$.

Del error original ¿qué porcentaje puede el Sr. Bump explicar a través del conocimiento de la variable relacionada X ? Este valor es el coeficiente de determinación r^2 .

$$\begin{aligned} r^2 &= 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2} \\ &= 1 - \frac{59.41}{233.6} = 1 - .254 = .746 \end{aligned}$$

La respuesta, por supuesto, es el mismo valor de r^2 que se obtuvo anteriormente, e indica que el conocimiento de X le permite a el Sr. Bump explicar el 74.6% de la variabilidad de Y alrededor de su media. También, el 25.4% de la variabilidad de Y no se explica mediante el conocimiento de X .

La relación entre la desviación total, la desviación explicada y la desviación no explicada es:

Desviación total = desviación explicada + desviación no explicada

$$\begin{aligned} \Sigma(Y - \bar{Y})^2 &= \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2 \\ 233.6 &= 174.2 + 59.4 \\ 233.6 &= 233.6 \end{aligned}$$

PRUEBA DE LA HIPÓTESIS

En el cap. 2, se advirtió que el coeficiente de correlación de la muestra r tiene su contraparte en la población, denominado ρ (rho). En forma similar, la población de todos los puntos $X - Y$ posee un coeficiente de determinación ρ^2 . Como siempre, debe tenerse presente que la información de la muestra produce estadísticas de muestra que permiten al pronosticador hacer inferencias sobre las relaciones existentes entre X y Y junto con todos los puntos de datos de la población. Una prueba estadística que se podría considerar es

$$H_0: \beta = 0$$

en donde β es la pendiente de la verdadera línea de regresión de la población. El no rechazar esta hipótesis nula significa que a diferencia del hecho de que la evidencia de la muestra produjo un valor diferente de cero para b , la evidencia no es lo suficientemente fuerte para rechazar la noción de que la línea de regresión para todos los puntos de datos de la población es plana (horizontal). Nótese que tal declaración es equivalente a establecer que $\rho = 0$. Una población de puntos de datos que tiene una línea de regresión plana tendrá también un coeficiente de correlación cero.

¿Cómo podría ser β cero cuando b es diferente de cero? Considere la fig. 6.8, en donde se muestra una población de puntos de datos a partir de la cual se seleccionó una muestra de cinco puntos (los puntos de la muestra se señalan con $\times$). Como sugiere este diagrama de dispersión, si se seleccionan suficientes puntos de datos de muestra, resultará obvio que la población de puntos de datos tiene una línea de regresión con una pendiente plana. Esto es, tanto β como ρ son cero. Sin embargo, en la muestra aleatoria de cinco puntos de datos, se seleccionaron puntos que se ubican muy cerca de una línea de regresión de tendencia ascendente. Sería erróneo concluir a partir de esta evidencia que X y Y están

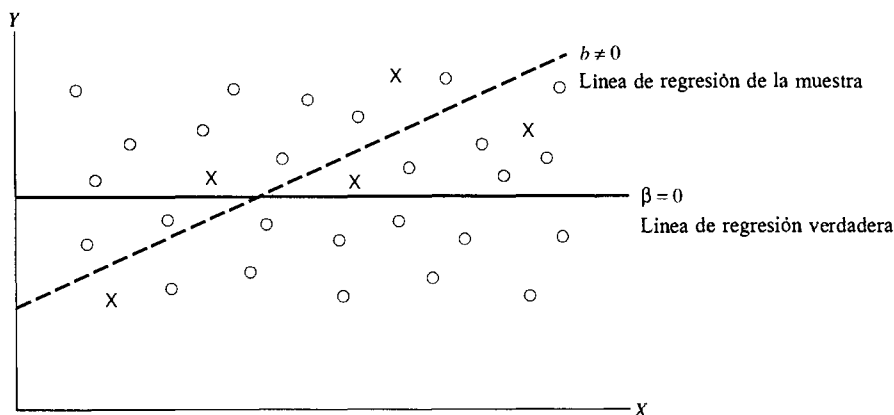


Figura 6.8 Puntos de datos de la población y de la muestra.

relacionadas en una forma lineal positiva. Sin embargo, si se prueba la hipótesis de que $\beta = 0$, es probable que el pronosticador no pueda rechazarla.

Ejemplo 6.9

Para los datos del Sr. Bump, la prueba aparece en la tabla 6.7.

De la tabla t para $n - 2$ grados de libertad

$$t_{.01, (df=8)} = 3.355$$

$$t_{.05, (df=8)} = 2.306$$

De ahí que, el Sr. Bump pueda rechazar la siguiente hipótesis nula: la línea de regresión de la población tiene una pendiente plana ($\beta = 0$).

TABLA 6.7 PRUEBA DE HIPÓTESIS PARA LA PENDIENTE

$$H_0: \beta = 0$$

$$s_b = \frac{s_{y \cdot x}}{\sqrt{\sum (X - \bar{X})^2}} \quad (6.15)$$

$$t = \frac{b - \beta}{s_b} \quad (6.16)$$

$$t = \frac{-14.54 - 0}{2.72/\sqrt{.824}} = \frac{-14.54}{3.00} = -4.8$$

Nota: Los valores utilizados se calcularon anteriormente en este capítulo. El término s_b es el error estándar del coeficiente de regresión.

Considere ahora dos situaciones especiales que ilustran la importancia de probar una hipótesis con respecto a la población así como considerar el valor de r^2 .

Situación

Suponga que $n = 1000$ y $r^2 = .10$. Se rechaza la siguiente hipótesis nula: $H_0: \beta = 0$. Sin embargo, la regresión no se considera útil.

Razón. El tamaño en extremo grande de la muestra produce una estimación muy precisa de ρ^2 , el coeficiente de determinación de la población. Se puede rechazar la noción de que este coeficiente sea igual a cero. Sin embargo, la relación entre X y Y , aunque el pronosticador confía en que existe, es muy débil ($\rho^2 = .10$). Por lo tanto, queda establecido, con un margen reducido de error, que existe una débil relación en la población entre X y Y , por lo que el pronosticador no puede utilizar el análisis de regresión.

Situación

Suponga que $n = 3$ y $r^2 = .95$. La siguiente hipótesis no se puede rechazar: $H_0: \beta = 0$.

Razón. Los tres puntos de datos casi se ubican sobre una línea recta. Sin embargo, es un tamaño de muestra en extremo reducido, lo que no constituye suficiente evidencia para hacer una declaración positiva acerca de todos los puntos de datos de la población. La hipótesis nula no se puede rechazar debido al elevado valor de r^2 . Se requiere reunir mayor evidencia de la muestra antes de poder efectuar declaraciones significativas con respecto a la población.

Algunos paquetes de cómputo no calculan la estadística t . En su lugar utilizan una prueba de proporción de varianzas o estadística F . Para probar la hipótesis de que un modelo en particular es bueno, se divide una estimación no desviada de la varianza explicada por el modelo, entre la estimación no desviada del error de varianza no explicado para el modelo. La ecuación 6.17 se emplea para calcular la estadística F .

$$F = \frac{\text{Varianza contabilizada}}{\text{Varianza total}} = \frac{(r_F^2 - r_R^2)/(K_F - K_R)}{(1 - r_F^2)/(n - K_F)} \quad (6.17)$$

en donde

$r_F^2 = r^2$ para el modelo completo

$r_R^2 = r^2$ para el modelo restringido

K_F = número de parámetros linealmente independientes a estimar en el modelo completo

K_R = número de parámetros linealmente independientes a estimar en el modelo restringido

n = tamaño de la muestra

$r_F^2 - r_R^2$ = porcentaje de varianza total explicada por la(s) variable(s) en el modelo completo pero no en el modelo restringido

$1 - r_F^2$ = porcentaje de la varianza total no explicada por el modelo completo

El modelo completo es el modelo que el analista está interesado en probar. El modelo restringido se forma mediante el uso de la media de la variable Y ($\bar{Y}$) como el pronosticador o variable independiente. r_F^2 es el porcentaje de la varianza explicada cuando se emplea X para predecir Y . r_R^2 es cero ya que el uso de la media de Y para predecir Y no explica la varianza de ninguna variable Y .

Para determinar el valor apropiado de F a partir de la tabla C.6, el analista necesita conocer dos parámetros: los grados de libertad tanto del numerador ($K_F - K_R$), como del denominador ($n - K_F$).

Ejemplo 6.10

La hipótesis nula para probar la ecuación del Sr. Bump, el modelo que usa el precio para predecir ventas, no es una buena hipótesis: el conocimiento del movimiento de la variable precio no explica un porcentaje significativo de la varianza de la variable ventas. Se usa la ecuación 6.17 para calcular la estadística F apropiada.

$$\begin{aligned} F &= \frac{(r_F^2 - r_R^2)/(K_F - K_R)}{(1 - r_F^2)/(n - K_F)} \\ &= \frac{(.746 - 0)/(2 - 1)}{(1 - .746)/(10 - 2)} \\ &= \frac{.746}{.03175} = 23.5 \end{aligned}$$

Debido a que la estadística F calculada (23.5) es mayor que el valor tabulado de F para 1 y 8 grados de libertad a los niveles de significancia de .05 (5.32) y .01 (11.26), se rechaza la hipótesis nula en ambos niveles de significancia. El Sr. Bump concluye que es un buen modelo y que explica un porcentaje significativo de la varianza de la variable dependiente.

SALIDA DE CÓMPUTO

El problema de análisis de regresión del Sr. Bump (datos de la tabla 6.1) se ejecuta en un programa de cómputo de regresión para producir la salida que se presenta en la tabla 6.8. El paquete utilizado produce una salida que es típica en comparación con las generadas por otros programas de análisis de regresión. Aunque no hay un formato estándar para presentar los resultados de un análisis de regresión, la salida que se muestra en la tabla 6.8 contiene los valores que exhiben la mayoría de los programas de análisis de regresión. Para explicar la terminología utilizada en la salida de cómputo, en la siguiente lista se presentan definiciones y cálculos relacionados con la tabla 6.8.

1. *Correlación X vs. Y* = $-.86$. El coeficiente de correlación de la muestra (r) indica la relación entre X y Y , o entre precio y ventas, respectivamente.
2. *Coeficiente de regresión* = -14.54 . Este valor es la proporción de cambio en Y (ventas) cuando X (precio) se incrementa en una unidad (b). Cuando el precio se incrementa en \$1 dólar, las ventas disminuyen en 14 540 unidades.
3. *Error estándar del coeficiente de regresión* = 3.0 . Este valor es la desviación estándar de la distribución de muestreo del valor del coeficiente de regresión (b).

$$s_b = \frac{s_{y \cdot x}}{\sqrt{\sum (X - \bar{X})^2}} = \frac{2.72}{\sqrt{.824}} = \frac{2.72}{.908} = 3.00$$

TABLA 6.8 SALIDA DE CÓMPUTO PARA EL PROBLEMA DE REGRESIÓN DEL SR. BUMP

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD. ERROR OF REG. COEF	COMPUTED T VALUE
2	1.44000	0.30258	-0.86 (1)	-14.54 (2)	3.00 (3)	-4.84 (4)
DEPENDENT						
1	11.20000	5.09466				
INTERCEPT		32.14 (5)		MULTIPLE CORRELATION	0.86 (7)	
STD. ERROR OF ESTIMATE		2.725 (6)		R SQUARED	0.746 (8)	
				CORRECTED R SQUARED	0.714 (9)	
ANALYSIS OF VARIANCE FOR THE REGRESSION (14)						
SOURCES OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE		
ATTRIBUTABLE TO REGRESSION	1	174.175	174.175	23.448		
DEVIATION FROM REGRESSION	8	59.425 (10)	7.428			
TOTAL	9	233.600 (11)				
CORRELATION MATRIX: 2 BY 2 (13)						
VAR.	1	2				
1	1.000	-0.863				
2	-0.863	1.000				
TABLE OF RESIDUALS						
CASE NO.	Y VALUE	Y ESTIMATE	RESIDUAL			
1	10.00000	13.23543	-3.23543 (12)			
2	6.00000	3.05824	2.94176			
3	5.00000	7.41989	-2.41989			
4	12.00000	10.32765	1.67235			
5	10.00000	8.87376	1.12624			
6	15.00000	14.68932	0.31068			
7	5.00000	8.87376	-3.87376			
8	12.00000	11.78154	0.21846			
9	17.00000	17.59706	-0.59706			
10	20.00000	16.14317	3.85683			

4. *Valor calculado* $t = -4.84$. El valor calculado t se usa para probar si el coeficiente de regresión de la población β es diferente de cero en forma significativa.

$$t = \frac{b - \beta}{s_b} = \frac{-14.54}{s_b}$$

$$s_b = \frac{s_{y \cdot x}}{\sqrt{\Sigma(X - \bar{X})^2}} = \frac{2.725}{\sqrt{.8240}} = 3.00$$

$$t = \frac{-14.54}{3.00} = -4.85 \text{ (error de redondeo)}$$

5. *Intersección* = 32.14. Este valor es el punto de intersección de y (b_0). De ahí que la ecuación completa de regresión sea

$$\hat{Y} = 32.14 - 14.54X$$

6. *Error estándar de la estimación* = 2.725. El error estándar de la estimación indica que los valores de Y se ubican en forma típica alrededor de casi 2.725 unidades de la línea de regresión.

$$s_{y \cdot x} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - 2}} = \sqrt{\frac{59.425}{10 - 2}} = \sqrt{7.428} = 2.725$$

7. *Correlación múltiple* = .86. El coeficiente de correlación múltiple es la raíz cuadrada positiva de r^2 e indica el grado de relación entre X y Y pero no el tipo de relación.

$$r^2 = .74$$

$$\sqrt{r^2} = \sqrt{.74}$$

$$r = .86$$

8. *r cuadrada* = .746. La regresión explica el 74.6% de la varianza en el volumen de ventas.

$$r^2 = 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2} = 1 - \frac{59.4}{233.6}$$

$$= 1 - .254 = .746$$

9. *r cuadrada corregida* = .714. El valor de r^2 se ajusta a los grados de libertad apropiados.

$$r_c^2 = 1 - \frac{\Sigma(Y - \hat{Y})^2 / (n - 2)}{\Sigma(Y - \bar{Y})^2 / (n - 1)} = 1 - \frac{59.425/8}{233.6/9}$$

$$= 1 - \frac{7.428}{25.956} = 1 - .286 = .714$$

10. *Suma de desviaciones cuadradas de la regresión* = 59.425. La suma de los residuos cuadrados es la diferencia entre la Y real y la predicción de Y ($\hat{Y}$), elevada al cuadrado y sumada (en ocasiones se le denomina como el *error de la suma de cuadrados*).

$$\Sigma(Y - \hat{Y})^2 = 59.425$$

11. *Suma total de cuadrados* = 233.6. Es el valor de la suma de las desviaciones cuadradas de la media.

$$\Sigma(Y - \bar{Y})^2 = 233.6$$

12. *Tabla de residuos*. Los residuos miden la diferencia entre lo que es realmente Y y lo que se predice que será cuando se usa la ecuación de regresión. Por ejemplo,

$$\text{Case 1: } Y - \hat{Y} = 10 - 13.23543 = -3.23543$$

$$\text{Case 2: } Y - \hat{Y} = 6 - 3.05824 = 2.94176$$

13. *Matriz de correlación*. La matriz proporciona la correlación entre todas las variables en el análisis. Como sólo comprende dos variables, sólo hay un coeficiente de correlación (-.863).
14. *Análisis de varianza*. El valor F (23.448) en esta tabla ANOVA (análisis de varianza), prueba la hipótesis nula de que la regresión no es significativa; es decir, la reducción en la suma de errores cuadrados de 233.6 a 59.425 se debe a la casualidad. Un valor alto de F permitirá rechazar esta hipótesis, sugiriendo una regresión *significativa*. El valor de F ($23.488 = 174.175/7.428$) se hace más grande en tanto se explique una mayor parte de la suma total de residuos cuadrados, mediante la regresión. En este caso el valor F tabulado ($df_n = 1$, $df_d = 8$, $\alpha = .01$) es 11.26. De ahí que se rechace la hipótesis de que la regresión no es significativa al nivel de significancia de 1%, ya que $23.488 > 11.26$. Los grados de libertad apropiados en cualquier análisis de regresión son $k - 1$, $n - k$ y $n - 1$ para los atribuibles a la regresión, la desviación de la regresión y el total, respectivamente, en donde k es el número de coeficientes de regresión calculados a partir de los datos de la muestra (todas las b).

Transformación de variables

Las técnicas en este capítulo se apoyan en la suposición de que existe una relación *lineal* entre la variable dependiente (Y) y la variable de predicción (X). Se supone también esta relación entre Y y pronósticos múltiples como se expone en el cap. 7.

En ocasiones, los datos de interés no muestran una relación lineal. Si la relación X - Y se acerca a ser lineal, la suposición de este capítulo sobre la relación de línea recta podría proporcionar suficiente precisión. Cuando éste no es el caso, se debe encontrar la forma de manejar una relación que no se puede aproximar a un modelo lineal.

Existen dos formas básicas para el manejo de relaciones no lineales. El primero consiste en ajustar los datos con una línea curva o modelo y emplear la relación establecida con fines de pronóstico. Este enfoque constituye una ampliación del modelo lineal descrito en este libro y es tema de más avanzados tratamientos del análisis de regresión.

El segundo enfoque consiste en convertir una o más variables a otra forma, de modo que la relación resultante con Y sea lineal. Cuatro de las transformaciones más comunes que se usan para generar nuevas variables de predicción son: el logaritmo de X , la raíz cuadrada de X , X cuadrada y el recíproco de X . Al graficar cada una de estas variables contra Y , la esperanza es que, aunque la relación X - Y en sí misma no es lineal, alguna de las transformaciones arroje una relación lineal. De ser así, se puede tratar a Y y a esta nueva variable mediante el modelo lineal expuesto en este capítulo, incluyendo el cálculo del coeficiente de regresión y la ecuación de regresión.

En el ejemplo siguiente, se emplea Minitab para graficar una relación X - Y simple que parece no ser lineal. Se instruye al programa para que calcule las cuatro transformaciones descritas anteriormente. Entonces se pueden graficar estas variables contra Y , para producir las gráficas que se muestran.

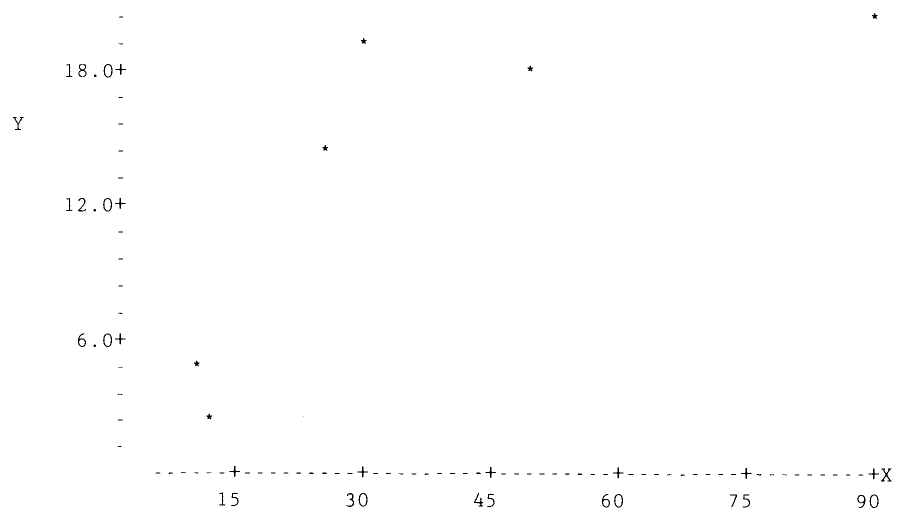
```
MTB > SET C1
DATA> 5 18 20 14 3 19
DATA> END
MTB > SET C2
DATA> 10 50 90 25 12 30
DATA> END
MTB > LOGE C2, PUT INTO C3
MTB > SQRT C2, PUT INTO C4
MTB > MULTIPLY C2 BY C2, PUT INTO C5
MTB > DIVIDE 1 BY C2, PUT INTO C6
MTB > NAME C1 'Y' C2 'X' C3 'LOGX' C4 'SQRTX' C5 'X2' C6 '1/X'
MTB > PRINT C1-C6
```

ROW	Y	X	LOGX	SQRTX	X2	1/X
1	5	10	2.30259	3.16228	100	0.100000
2	18	50	3.91202	7.07107	2500	0.020000
3	20	90	4.49981	9.48683	8100	0.011111
4	14	25	3.21888	5.00000	625	0.040000
5	3	12	2.48491	3.46410	144	0.083333
6	19	30	3.40120	5.47723	900	0.033333

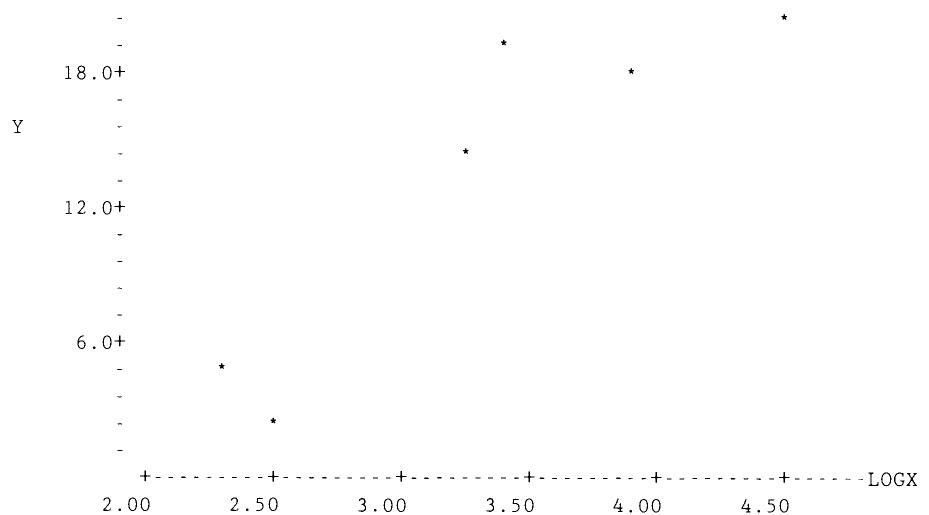
- El comando SET se usa para introducir Y y X en C1 y C2.
- El comando LOGE se emplea para calcular los logaritmos de los valores de X y almacenar los resultados en C3.

- El comando SQRT se utiliza para calcular las raíces cuadradas de los valores de X y almacenar los resultados en C4.
- El comando MULTIPLY se usa para elevar al cuadrado los valores de X y almacenar los resultados en C5.
- El comando DIVIDE se emplea para los recíprocos de X y almacenar los resultados en C6.

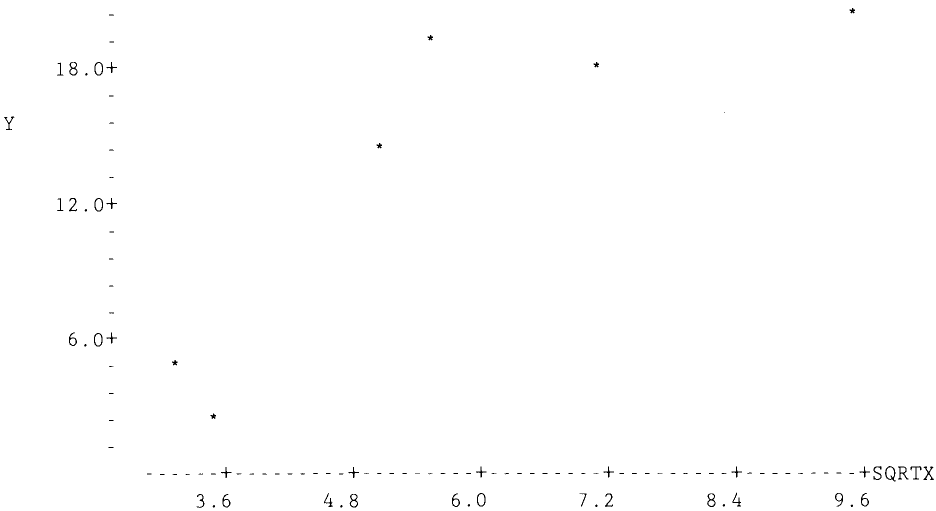
MTB > PLOT C1 VS C2



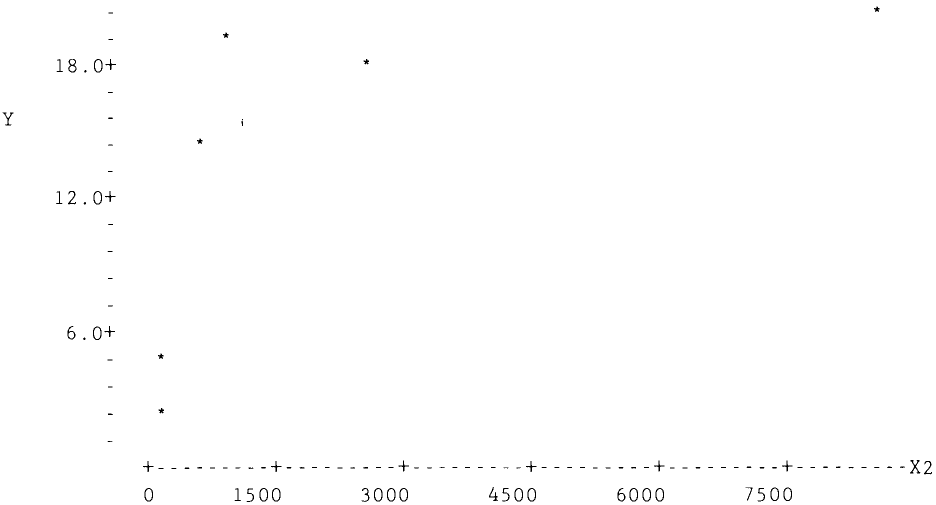
MTB > PLOT C1 VS C3



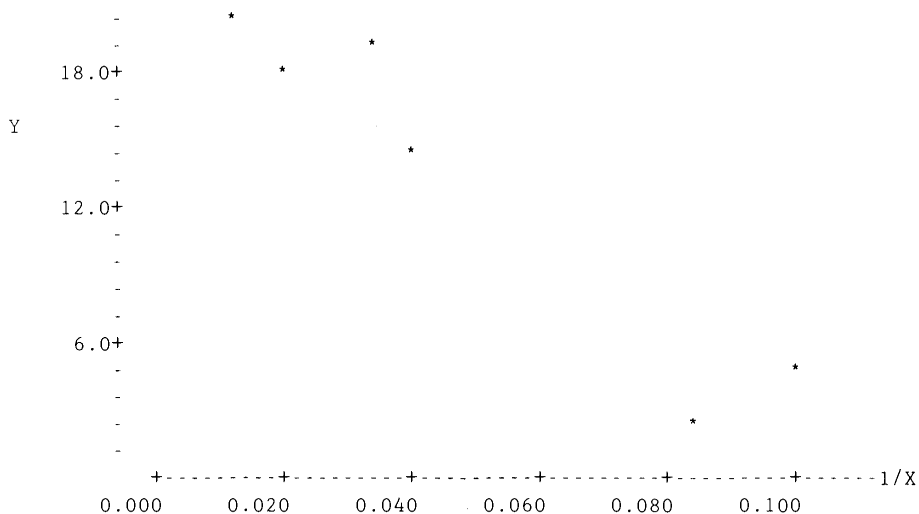
MTB > PLOT C1 VS C4



MTB > PLOT C1 VS C5



MTB > PLOT C1 VS C6



MTB > STOP

Nótese que la primera gráfica (X vs. Y) produce una relación no lineal. Es probable que el ajustar una línea recta a través de estos puntos de datos se produzca en el análisis y en los pronósticos subsecuentes, un grado de error no deseado.

En contraste, la segunda gráfica (Y vs. logaritmo de X) y la última gráfica (Y vs. recíproco de X) se muestran ambas prometedoras para el modelo lineal. Se podrían utilizar los métodos lineales de este capítulo y del cap. 7, para emplear estos dos pronósticos en un intento por encontrar una ecuación de predicción con un grado aceptable de precisión.

APLICACIÓN EN LA ADMINISTRACIÓN

El análisis de regresión es la herramienta estadística más ampliamente utilizada por la administración, cuando existe la necesidad de evaluar el impacto de una sola variable independiente sobre una variable dependiente. El análisis de regresión junto con el análisis de correlación, ayuda al pronosticador a caracterizar las relaciones entre variables. El pronosticador puede determinar tanto la importancia como la dirección de la relación entre variables.

La mayoría de los problemas que utilizan el análisis de regresión comprenden una versión más compleja denominada *análisis de regresión múltiple* (que se describirá en el siguiente capítulo), ya que la mayoría de las relaciones implican el estudio de más de una

variable independiente. No obstante, en ocasiones se utilizan el análisis de regresión simple y el análisis de correlación, algunos ejemplos son

- *Consumo de productos.* Un fabricante pudiera intentar predecir qué tanta cerveza bebe una persona por semana, a partir de variables como ingreso, edad educación o clase social.
- *Ventas.* Un detallista podría intentar predecir las ventas de un producto para una tienda versus otra, con base en diferenciales de precio, ingreso relativo a la comunidad de los alrededores, el trato amistoso del personal de la tienda, o el número y fuerza de los competidores en cada área de mercado.
- *Precios de la bolsa.* Un analista de bolsa de una empresa regional de corredores podría intentar predecir el precio de una nueva emisión de una empresa local, con base en la economía regional, el ingreso, la población o la imagen de la empresa.
- *Cartera vencida.* Un contador podría intentar predecir la cartera vencida que una compañía pudiera esperar encontrar en el siguiente trimestre fiscal, con base en el número de personas desempleadas, crédito excesivo, tasas de interés o ventas esperadas.
- *Requerimientos de personal.* Un director de personal de una gran compañía manufacturera podría intentar predecir los requerimientos de personal para el próximo año, a partir de la edad promedio de sus empleados, su escala de salario comparada con la de la comunidad de los alrededores, nuevos contratos de ventas esperados o la disponibilidad de trabajos competitivos.
- *Demanda de un centro comercial.* El gerente de un nuevo centro comercial podría tratar de anticiparse a la demanda, analizando el ingreso de la comunidad de los alrededores, el tamaño de la población o la proximidad y dimensiones de otros centros comerciales.

Una vez determinada la relación entre las variables independiente y dependiente, la administración puede, en algunos casos, tratar de controlar la variable dependiente a partir de este conocimiento. Por ejemplo, suponga que un gerente de comercialización 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 comercialización podría intentar controlar las ventas aumentando o disminuyendo la cantidad de publicidad que maximizará las utilidades. Siempre que el administrador tenga control sobre la variable independiente, hay posibilidad de un control parcial de la variable dependiente. De ahí que la ecuación de regresión y el coeficiente de determinación ayuden a la administración a determinar si vale la pena dicho control.

GLOSARIO

Línea de regresión. Una línea de regresión es aquella que mejor se ajusta a un conjunto de datos $X - Y$. Minimiza la suma de las distancias al cuadrado de los puntos a la línea, medidas en dirección vertical o de Y .

Error estándar de la estimación. El error estándar de la estimación mide la cantidad estándar en que los valores reales de Y difieren de los valores estimados ($\hat{Y}$).

FÓRMULAS CLAVE

Método de mínimos cuadrados: fórmula de la pendiente $b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$ (6.1)

Método de mínimos cuadrados: fórmula de intersección con Y $b_0 = \frac{\sum Y}{n} - \frac{b\sum X}{n}$ (6.2)

Ecuación de regresión $\hat{Y} = b_0 + bX$ (6.3)

Error estándar de la estimación: fórmula definitoria

$$s_{y \cdot x} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n - 2}} \quad (6.4)$$

Error estándar de la estimación: fórmula de cálculo

$$s_{y \cdot x} = \sqrt{\frac{\sum Y^2 - b_0 \sum Y - b \sum XY}{n - 2}} \quad (6.5)$$

Error estándar del pronóstico

$$s_f = s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum (X - \bar{X})^2}} \quad (6.6)$$

Intervalo de predicción: muestra grande $\hat{Y} \pm Zs_f$ (6.7)

Intervalo de predicción: muestra reducida $\hat{Y} \pm ts_f$ (6.8)

Coefficiente de determinación $r^2 = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}$ (6.9)

Coefficiente de determinación: fórmula de cálculo

$$r^2 = \frac{b_0 \sum Y + b \sum XY - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2} \quad (6.10)$$

$$\text{Suma total de cuadrados} \quad \Sigma(Y - \bar{Y})^2 \quad (6.11)$$

$$\text{Varianza total de } Y \text{ alrededor de su media} \quad s_y^2 = \frac{\Sigma(Y - \bar{Y})^2}{n - 1} \quad (6.12)$$

$$\text{Suma de residuos cuadrados} \quad \Sigma(Y - \hat{Y})^2 \quad (6.13)$$

Varianza total de Y alrededor de los valores predecidos

$$s_{y \cdot x}^2 = \frac{\Sigma(Y - \hat{Y})^2}{n - 2} \quad (6.14)$$

$$\text{Error estándar del coeficiente de regresión} \quad s_b = \frac{s_{y \cdot x}}{\sqrt{\Sigma(X - \bar{X})^2}} \quad (6.15)$$

$$\text{Estadística } t \text{ para el coeficiente de regresión} \quad t = \frac{b - \beta}{s_b} \quad (6.16)$$

$$\text{Prueba de proporción de varianza o estadística } F \quad F = \frac{(r_F^2 - r_R^2)/(K_F - K_R)}{(1 - r_F^2)/(n - K_F)} \quad (6.17)$$

PROBLEMAS

1. ¿Cuál de las siguientes situaciones es inconsistente?

- a. $\hat{Y} = 499 + .21X$ y $r = .75$
- b. $\hat{Y} = 100 + .9X$ y $r = -.70$
- c. $\hat{Y} = -20 + 1X$ y $r = .40$
- d. $\hat{Y} = -7 - 4X$ y $r = -.90$

2. Los ingresos por acción de AT&T (American Telephone & Telegraph) se estiman mediante el PNB (Producto Nacional Bruto). La ecuación de regresión es $\hat{Y} = 0.078 + 0.06X$, en donde el PNB se mide en miles de millones de dólares.

- a. Interprete la pendiente.
- b. Interprete la intersección con Y .

3. Considere los datos de la tabla siguiente, en donde X = gastos semanales de publicidad y Y = ventas semanales.

Y	X	Y	X
\$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

- a. ¿Existe una relación significativa entre los gastos de publicidad y las ventas?
- b. Establezca la ecuación de predicción.
- c. Pronostique las ventas para un gasto de publicidad de \$50 dólares.

- d. ¿Qué porcentaje de la varianza se puede explicar con la ecuación de predicción?
- e. Establezca la cantidad de varianza no explicada.
- f. Establezca la cantidad total de varianza.
4. En la tabla siguiente se muestran los tiempos requeridos de atención a clientes en las cajas de un supermercado y los valores de las compras. Responda con estos datos los apartados a, b, e y f del problema 3. Dé un punto e intervalo de estimación de 99% para Y si $X = 3.0$.

TIEMPO REQUERIDO DE ATENCIÓN (EN MINUTOS)	VALOR DE LAS COMPRAS (EN DÓLARES)	TIEMPO REQUERIDO DE ATENCIÓN (EN MINUTOS)	VALOR DE LAS COMPRAS (EN DÓLARES)
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

5. A Lori Franz, supervisora de mantenimiento de la Baltimore Transit Authority, le gustaría determinar si existe una relación entre el costo anual de mantenimiento de un autobús y su antigüedad. Si hay una relación, Lori piensa que puede pronosticar mejor el presupuesto anual de mantenimiento de autobuses. Por ello, reúne los siguientes datos:

AUTOBÚS	COSTO DE MANTENIMIENTO (\$) Y	ANTIGÜEDAD (EN AÑOS) X
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

- a. Grafique un diagrama de dispersión.
- b. ¿Qué clase de relación existe entre las dos variables?
- c. Calcule el coeficiente de correlación.
- d. Pruebe el nivel de correlación a un nivel de significancia de .05.
- e. ¿Debería usar Lori el análisis de regresión para pronosticar el presupuesto anual de mantenimiento de autobuses?
- f. Determine la ecuación de regresión de la muestra.
- g. Pronostique el costo de mantenimiento para un autobús que tiene 5 años de antigüedad.
6. Andrew Vazsonyi es el gerente de la cadena de supermercados Spendwise. A él le gustaría poder pronosticar las ventas de libros en rústica (libros por semana), con base en la cantidad de espacio disponible en los estantes (en pies). Andrew reúne datos para una muestra de 11 semanas:

SEMANA	NÚMERO DE LIBROS VENDIDOS Y	PIES DE ESPACIO EN ESTANTES X
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

- Grafique un diagrama de dispersión.
 - ¿Qué clase de relación existe entre las dos variables?
 - Calcule el coeficiente de correlación.
 - Pruebe el nivel de correlación a un nivel de significancia de .10.
 - ¿Debería usar el análisis de regresión para pronosticar las ventas de libros en rústica?
 - Determine la ecuación de regresión de la muestra.
 - Pronostique las ventas de libros en rústica para una semana durante la cual hay un espacio disponible de 4 pies.
7. En la siguiente tabla se presenta información de una empresa de órdenes por correo, para 12 ciudades:

CIUDAD	NÚMERO DE ÓRDENES RECIBIDAS POR CORREO, Y (EN MILES)	NÚMERO DE CATÁLOGOS DISTRIBUIDOS, X (EN MILES)	CIUDAD	NÚMERO DE ÓRDENES RECIBIDAS POR CORREO, Y (EN MILES)	NÚMERO DE CATÁLOGOS DISTRIBUIDOS, X (EN MILES)
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

- Determine si existe una relación lineal significativa entre estas dos variables (pruebe a un nivel de significancia de .05).
 - Determine la línea de regresión.
 - Determine el error estándar de la estimación.
 - ¿Qué porcentaje de la varianza de la variable órdenes por correo se explica mediante la variable catálogos distribuidos?
 - Pruebe para determinar si el coeficiente de regresión es diferente de cero en forma significativa. (Utilice un nivel de significancia de .01.)
 - Pronostique las órdenes por correo recibidas cuando se distribuyeron 10 000 catálogos.
 - Calcule la varianza explicada para la variable Y .
8. En una regresión de inversiones sobre la tasa de interés, se observaron los resultados que se muestran en la tabla siguiente, para un periodo de 10 años:

INVERSIÓN ANUAL (MILES DE DÓLARES)	TASA PROMEDIO DE INTERÉS (PORCENTAJE)	INVERSIÓN ANUAL (MILES DE DÓLARES)	TASA PROMEDIO DE INTERÉS (PORCENTAJE)
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

- ¿Es significativa la relación entre estas variables?
 - ¿Se puede desarrollar una ecuación de predicción efectiva?
 - ¿Podemos pronosticar la inversión anual si la tasa de interés dentro de 5 años es del 4%?
 - Calcule e interprete r^2 .
 - Exponga la correlación y causalidad de este ejemplo.
9. La ABC Investment Company está en el negocio de hacer licitaciones sobre inversiones ofrecidas por diversas compañías que desean financiamiento adicional. ABC tabuló su licitación sobre las últimas 25 emisiones a las que se aplicó licitación, en términos del porcentaje de licitación de valor nominal. Para estas emisiones, tabuló también la licitación como porcentaje de valor nominal, de su principal competidor. Ahora, ABC se pregunta si está empleando el mismo razonamiento que su competidor en la preparación de licitaciones. En otras palabras, ¿se podría usar la licitación de ABC para pronosticar la de su competidor? De no ser así, el competidor debe estar evaluando las emisiones en forma distinta. Los datos se presentan en la siguiente tabla:

EMISIÓN	LICITACIÓN DE ABC	LICITACIÓN DEL COMPETIDOR	EMISIÓN	LICITACIÓN DE ABC	LICITACIÓN DEL COMPETIDOR
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			

- ¿Hasta qué grado están ambas compañías empleando el mismo razonamiento en la preparación de sus licitaciones?
- Pronostique la licitación del competidor, si la de ABC es 101% del valor nominal. Dé ambos un punto e intervalo de predicción bajo la suposición de que los datos constituyen la totalidad de la *población* de datos de licitaciones.
- De acuerdo con el apartado b, ¿cuál es la posibilidad de ABC de ganar esta licitación en particular? (La licitación más baja gana.)
- Si los datos constituyen una *muestra* aleatoria de los datos de licitaciones, en vez de la población total, ¿qué cambios serían necesarios en los apartados b y c? Responda la pregunta del apartado c, suponiendo que los datos corresponden a una muestra aleatoria.

10. Evalúe las siguientes declaraciones:

- a. “Una r^2 alta representa una regresión significativa”.
- b. “En un problema de regresión, un tamaño de muestra muy grande siempre producirá resultados útiles”.

11. Ed Bogdanski, propietario de American Precast Company, le contrata a usted como analista de medio tiempo. Ed quedó en extremo complacido cuando usted determinó que existe una relación positiva entre el número de permisos de construcción y la cantidad de trabajo disponible en su compañía. Ahora, se pregunta si es posible emplear el conocimiento de las tasas de interés sobre primeras hipotecas, para predecir el número de permisos de construcción que se emitirán cada mes. Usted reúne una muestra aleatoria de nueve meses de datos:

MES	PERMISOS DE CONSTRUCCIÓN Y	TASA DE INTERÉS X
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 ecuación de regresión de la muestra.
 - c. Pruebe el coeficiente de regresión a un nivel de significancia de .05.
 - d. Al incrementarse la tasa de interés en 1%, ¿cuál es la disminución promedio en el número de permisos de construcción?
 - e. Calcule el coeficiente de determinación simple.
 - f. Escriba una oración comprensible para Ed, interpretando la r^2 calculada en el apartado e.
 - g. Redacte un memorando para Ed explicando los resultados de su análisis.
12. Considere la población de 140 observaciones que se presentan en la tabla siguiente. La Marshall Printing Company desea estimar la relación entre el número de copias producidas mediante una técnica de impresión offset (X) y su costo laboral directo asociado (Y). Seleccione una muestra aleatoria de 20 observaciones.
- a. Elabore un diagrama de dispersión.
 - b. Calcule el coeficiente de correlación de la muestra.
 - c. Determine la ecuación de la línea de regresión de la muestra.
 - d. Dibuje la línea de regresión sobre el diagrama de dispersión.
 - e. Calcule el error estándar de la estimación.
 - f. Calcule el coeficiente de determinación de la muestra e interprete su valor.
 - g. Pruebe la hipótesis de que la pendiente de la línea de regresión que pasa a través de los datos de la totalidad de la población, es cero.
 - h. Formule un punto e intervalo de estimación del 90% de confianza para el costo laboral directo, si el proyecto comprende 250 copias.

TABLA PARA EL PROBLEMA 12

OBS.	Y	X	OBS.	Y	X	OBS.	Y	X
(1)	1.0	10	(48)	2.2	180	(95)	2.0	330
(2)	0.9	10	(49)	2.4	180	(96)	2.4	340
(3)	0.8	10	(50)	1.6	180	(97)	2.2	340
(4)	1.3	20	(51)	1.8	190	(98)	2.0	340
(5)	0.9	20	(52)	4.1	190	(99)	2.5	350
(6)	0.6	30	(53)	2.0	190	(100)	2.8	350
(7)	1.1	30	(54)	1.5	200	(101)	2.3	350
(8)	1.0	30	(55)	2.1	200	(102)	2.7	350
(9)	1.4	40	(56)	2.5	200	(103)	2.8	360
(10)	1.4	40	(57)	1.7	220	(104)	3.1	360
(11)	1.2	40	(58)	2.0	220	(105)	2.5	370
(12)	1.7	50	(59)	2.3	220	(106)	2.9	370
(13)	0.9	50	(60)	1.8	220	(107)	2.6	370
(14)	1.2	50	(61)	1.3	230	(108)	3.0	380
(15)	1.3	50	(62)	1.6	230	(109)	3.2	380
(16)	0.7	60	(63)	2.8	230	(110)	2.9	390
(17)	1.0	60	(64)	2.2	230	(111)	2.6	390
(18)	1.3	70	(65)	2.6	230	(112)	2.5	390
(19)	1.5	70	(66)	1.4	240	(113)	2.7	400
(20)	2.0	70	(67)	1.6	240	(114)	3.1	400
(21)	0.8	80	(68)	1.7	240	(115)	2.4	400
(22)	0.6	80	(69)	1.5	250	(116)	3.0	400
(23)	1.8	80	(70)	2.2	250	(117)	3.4	420
(24)	1.0	90	(71)	2.5	250	(118)	3.5	420
(25)	2.0	100	(72)	2.4	260	(119)	3.1	420
(26)	0.5	100	(73)	2.0	260	(120)	2.9	420
(27)	1.5	100	(74)	2.7	260	(121)	2.8	430
(28)	1.3	110	(75)	2.0	270	(122)	3.3	430
(29)	1.7	110	(76)	2.2	270	(123)	2.5	440
(30)	1.2	110	(77)	2.4	270	(124)	2.8	440
(31)	0.8	110	(78)	1.8	280	(125)	2.4	450
(32)	1.0	120	(79)	2.8	290	(126)	2.6	450
(33)	1.8	120	(80)	2.2	290	(127)	3.0	450
(34)	2.1	120	(81)	2.4	290	(128)	3.4	460
(35)	1.5	130	(82)	2.1	290	(129)	3.0	460
(36)	1.9	130	(83)	1.9	290	(130)	3.3	470
(37)	1.7	140	(84)	2.4	300	(131)	3.4	470
(38)	1.2	150	(85)	2.5	300	(132)	3.1	470
(39)	1.4	150	(86)	2.9	300	(133)	3.6	480
(40)	2.1	150	(87)	2.0	300	(134)	3.0	480
(41)	0.9	160	(88)	1.9	310	(135)	2.9	480
(42)	1.1	160	(89)	2.5	310	(136)	3.2	480
(43)	1.7	160	(90)	2.6	310	(137)	2.6	490
(44)	2.0	160	(91)	3.2	320	(138)	3.8	490
(45)	1.6	170	(92)	2.8	320	(139)	3.3	490
(46)	1.9	170	(93)	2.4	320	(140)	2.9	500
(47)	1.7	170	(94)	2.5	320			

CASO DE ESTUDIO 6.1

TIGER TRANSPORT

La Tiger transport Company es una compañía de camiones que se dedica a hacer mudanzas locales y a través del país. Su preocupación actual es el precio que cobra por transportar cargas pequeñas a grandes distancias. La empresa está satisfecha con sus tarifas para camiones con carga completa; estas tarifas se basan en los costos variables de conductor, combustible y mantenimiento, más gastos generales y utilidad. Sin embargo, cuando un camión lleva menos de la mitad de su carga posible, existe cierto cuestionamiento sobre los bienes requeridos para llenar el camión. Para pronosticar las necesidades futuras de combustible y preparar presupuestos a largo plazo, la empresa quisiera determinar el costo de agregar carga a un camión parcialmente lleno.

La empresa piensa que el único costo adicional en que se incurre si se agrega carga adicional al camión, es el costo de combustible adicional, ya que entonces se reduciría el número de millas por galón del vehículo. Como uno de los factores empleados para determinar las tarifas por cargas pequeñas, la empresa desearía saber el costo del desembolso asociado al costo de combustible por carga adicional.

Usted es un recién graduado en administración que trabaja en el departamento de contabilidad de costos, y se le asigna el trabajo de investigar este asunto y aconsejar a la alta dirección sobre las consideraciones necesarias para una buena decisión sobre las tarifas. Usted comienza suponiendo que todos los camiones son iguales, de hecho son casi idénticos en dimensiones, capacidad total de carga y tamaño del motor. Supone también que todos los conductores obtendrán el mismo millaje en un viaje largo. El contador en jefe de la empresa está de acuerdo con que estas suposiciones son razonables.

Entonces, sólo le queda una variable que pudiera afectar los millas por galón de los camiones en largos recorridos: el peso de la carga. Usted encuentra que el departamento de contabilidad tiene registros de cada viaje realizado por los camiones de Tiger durante los últimos años. Estos registros incluyen el peso total de la carga, la distancia cubierta y el número de galones de combustible diesel consumidos. La proporción entre estas dos últimas cifras son las millas por galón del viaje.

Usted selecciona como población los viajes realizados en los últimos cuatro años, que ascienden a 5 428. Después, selecciona 40 números de una tabla de números aleatorios y, ya que los viajes están registrados uno tras otro, asigna el número 1 al primero, el número 2 al segundo y así sucesivamente. Entonces, sus 40 números aleatorios producen una selección aleatoria de 40 viajes por examinar. De estos viajes se registran el peso de la carga y las millas por galón, según se muestra en la tabla 6.9.

Como Tiger tiene una computadora con un paquete de análisis de regresión, usted emplea este programa para ejecutar los datos de la tabla 6.9. El reporte resultante aparece en la tabla 6.10.

Al estudiar el reporte de la tabla 6.10, usted decide que la muestra de datos produjo una ecuación útil. Esta conclusión se basa en una r^2 relativamente alta (76%), un alto valor de t (-10.9) y un alto valor de F (119). A partir del reporte, usted formula la ecuación de regresión

$$Y = 8.84833 - .06040X$$

en donde Y se mide en millas por galón y X se mide en miles de libras.

La pendiente de la ecuación de regresión (-0.6040) se interpreta de la siguiente manera: *cada 1000 libras adicionales de carga, reducen el millaje de un camión en un promedio de .06040 millas por galón.*

TABLA 6.9 DATOS DE TIGER TRANSPORT: PESO DE CARGA POR VIAJE (MILES DE LIBRAS Y MILLAS POR GALÓN)

PESO	MILLAS POR GALÓN	PESO	MILLAS POR GALÓN	PESO	MILLAS POR GALÓN	PESO	MILLAS POR GALÓN
60	5.3	58	4.9	63	5.0	63	5.0
55	5.0	60	5.1	65	4.9	62	4.9
80	4.0	74	4.5	72	4.6	77	4.6
72	4.2	80	4.3	81	4.0	76	4.5
75	4.5	53	5.9	64	5.3	51	5.7
63	5.1	61	5.5	78	4.4	74	4.2
48	7.2	80	3.5	62	4.9	78	4.3
79	3.9	68	4.1	83	3.8	50	6.1
82	3.8	76	4.5	79	4.1	79	4.3
72	4.4	75	4.4	61	4.8	55	4.7

Actualmente la empresa paga aproximadamente \$1.25 por galón de combustible diesel. De ahí, puede usted entonces calcular el costo de transportar por 100 millas, 1000 libras de carga adicional, de la manera siguiente:

Promedio de millas por galón de la tabla 6.10 = 4.7 (media de la variable 2)

Costo de 100 millas,

$$\frac{100(1.25)}{4.7} = \$26.60$$

TABLA 6.10 SALIDA DEL ANÁLISIS DE REGRESIÓN PARA TIGER TRANSPORT

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD. ERROR OF REG. COEF	COMPUTED T VALUE
1	68.59999	10.21763	-0.87055	-0.06040	0.00554	-10.90562
DEPENDENT 2	4.70499	0.70890				
INTERCEPT		8.84833				
MULT. CORR		0.87055		R SQUARE	0.75786	
STD. ERR.		0.35339		CORR-RSQUARE	0.75786	
ANALYSIS OF VARIANCE FOR THE REGRESSION						
SOURCES OF VARIATION		DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE	
ATTRIBUTABLE TO REGRESSION		1	14.85315	14.85315	118.93246	
DEVIATION FROM REGRESSION		38	4.74572	0.12489		
TOTAL		39	19.59886			

Costo del mismo viaje con 1000 libras adicionales,

$$\frac{100(1.25)}{(4.7 - .0604)} = \$26.94$$

Por lo que

El costo adicional de transportar 100 millas, 1000 libras adicionales = \$0.34.

Ahora, usted piensa que terminó parte de su asignación, es decir, la determinación de los costos de desembolso asociados con el agregar un peso de carga a un camión sin llenar. Por supuesto, usted se da cuenta de que existen otros factores que influyen en la decisión sobre la tarifa para cargas pequeñas.

TAREA

Prepare un memorando que resuma el análisis, para la alta dirección de la empresa. Incluya comentarios sobre el grado en que su trabajo mejora los pronósticos de necesidades de combustible e ingresos por camión.

CASO DE ESTUDIO 6.2

BUTCHER PRODUCTS, INC.

Gene Butcher es el propietario y presidente de Butcher Products, Inc. una pequeña empresa que fabrica ductos de fibra de vidrio para instalaciones de cableado eléctrico. Gene ha estado estudiando el número de unidades de ductos fabricadas por día durante los últimos dos años y medio y está preocupado por la amplia variabilidad de esta cifra. Para programar apropiadamente la producción, costo e ingresos, Gene necesita establecer una relación entre la producción y alguna otra variable.

Con base en su experiencia en la empresa, Gene no podía encontrar ninguna razón para la variabilidad de la producción, hasta que comenzó a pensar en las condiciones climáticas. Su razonamiento es que la temperatura externa pudiera tener algo que ver con la productividad de su fuerza de trabajo y la producción diaria alcanzada.

Gene selecciona en forma aleatoria varios días de sus registros y anota el número de unidades de ductos producidas para cada uno de estos días. A continuación acude a la oficina local de clima y, para cada uno de los días seleccionados, registra la temperatura alta del día. Con ello, está listo para ejecutar un estudio de correlación entre estas dos cifras, cuando se da cuenta de que sería probable que la producción estuviera relacionada con una desviación de la temperatura ideal, en vez de con la propia temperatura. Esto es, piensa que un día muy caluroso o muy frío podría tener un efecto negativo en la producción al compararlo con un día que tuviera la temperatura ideal. Decide entonces convertir sus lecturas de temperatura en desviaciones a partir de 65 grados Fahrenheit, temperatura que él entiende es la ideal para lograr una alta producción. Sus datos son los siguientes; Y representa el número de unidades producidas, mientras que X representa la diferencia absoluta (eliminando los signos negativos) entre la temperatura alta del día y 65 grados Fahrenheit:

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 realiza un análisis de regresión utilizando la computadora de la empresa y el paquete de cómputo Minitab. Después de acceder Minitab, teclea las siguientes instrucciones para el programa:

```
MTB > SET C1
DATA > 12 10 3 4 27 10 18 12 27 20 4 11
DATA > 15 25 8 35 5 12 2 7 17 8 12
DATA > END
MTB > SET C2
DATA > 485 512 625 585 318 405 379 497 316
DATA > 351 525 395 327 308 603 321 426 410
DATA > 515 498 357 429 401
DATA > END
MTB > NAME C1 'TEMP' C2 'UNITS'
MTB > REGRESS C2 ON 1 PREDICTOR C1
```

The regression equation is
UNITS = 552 - 8.91 TEMP

Predictor	Coef	Stdev	t-ratio	p
Constant	552.04	22.85	24.16	0.000
TEMP	-8.911	1.453	-6.13	0.000

s = 59.41 R-sq = 64.2% R-sq(adj) = 62.5%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	132758	132758	37.62	0.000
Error	21	74109	3529		
Total	22	206866			

Unusual Observations						
Obs.	TEMP	UNITS	Fit	Stdev.Fit	Residual	St.Resid
15	8.0	603.0	480.8	14.5	122.2	2.12R
16	35.0	321.0	240.2	34.0	80.8	1.66 X

MTB > STOP

Los resultados del análisis de regresión mostrados anteriormente, complacen a Gene. Los valores t son altos, lo que indica que ambos valores en la ecuación de regresión de la muestra (552 y -8.9) son significativos: para ambos, el nivel de significancia es cercano a cero, como lo indican los valores p para la prueba t . El valor de F es alto (37.62), así que se puede emplear la ecuación de regresión de la muestra en la población de todos los días de producción.

Al observar r cuadrada, Gene está en cierto modo decepcionado al encontrar que este valor, aunque es satisfactorio (64.2%), no es tan alto como hubiera esperado. Sin embargo, piensa que es lo suficientemente alto para comenzar a pensar en formas de aumentar los niveles de productividad.

PREGUNTAS

1. ¿Cuántas unidades pronosticaría para un día cuando la temperatura alta es de 89 grados Fahrenheit?
2. ¿Cuántas unidades pronosticaría para un día cuando la temperatura alta es de 41 grados Fahrenheit?
3. Con base en los resultados del análisis de regresión mostrados con anterioridad ¿qué acción aconsejaría que tomara Gene con el propósito de incrementar la producción diaria?
4. ¿Piensa usted que Gene desarrolló una herramienta efectiva de pronóstico?

CASO DE ESTUDIO 6.3

DEPARTAMENTO DE PERSONAL DE ACE

La Ace Manufacturing Company emplea varios miles de personas en la fabricación de teclados, gabinetes para el equipo y cables para la industria de las pequeñas computadoras. El presidente de Ace se ha preocupado recientemente con la tasa de ausentismo existente entre los empleados de la compañía y pidió al departamento de personal que atendiera esta situación. El departamento de personal se da cuenta de que un método efectivo para pronosticar el ausentismo fortalecería en gran medida su capacidad para elaborar planes adecuados.

El director de personal, Bill McGone, en un intento por dimensionar el problema, decide revisar algunos expedientes del personal. Decide seleccionar 15 expedientes en forma aleatoria y registrar el número de días de ausencia durante el último año fiscal, junto con la edad del empleado. Después de leer un artículo relacionado con personal en una revista reciente, él piensa que la edad puede tener un efecto significativo en el ausentismo. Si encuentra que la edad y los días de ausencia muestran una buena correlación en su pequeña muestra, pretende tomar una muestra de 200 a 300 expedientes y formular una buena ecuación de predicción.

A continuación se presentan los datos reunidos en la muestra inicial. El número de días de ausencia durante el último año fiscal está representado por Y , mientras que X es la edad del empleado.

Y	X	Y	X
3	25	9	56
4	36	12	60
7	41	8	51
4	27	5	33
3	35	6	37
3	31	2	31
5	35	2	29
7	41		

PREGUNTAS

1. ¿Qué tan bien se correlacionan los días de ausencia y la edad? ¿Se puede generalizar esta correlación a toda la fuerza de trabajo?
2. ¿Cuál es la ecuación de pronóstico para los días de ausencia, utilizando la edad como la variable de predicción?
3. ¿Qué porcentaje de la variabilidad en los días de ausencia se puede explicar a través del conocimiento de la edad?
4. ¿Qué tan bien ha resultado la ecuación de predicción? Al responder esta pregunta, utilice los procedimientos estadísticos apropiados para respaldar su respuesta.
5. Suponga que una persona recién contratada tiene 24 años de edad ¿cuántos días de ausencia le pronostica durante el año fiscal?
6. ¿Deberá Bill McGone proceder a tomar una muestra grande de empleados de la compañía con base en los resultados preliminares de la muestra?
7. ¿Se desarrolló un método efectivo de pronóstico?

CASO DE ESTUDIO 6.4

MR TUX

John Mosby ha oído que el análisis de regresión se utiliza a menudo para pronosticar series de tiempo de variables y, como tiene una computadora personal con un paquete de cómputo de regresión, decide intentarlo. Su variable dependiente serán sus volúmenes de ventas mensuales, los datos se presentaron al final del cap. 2.

Como primer intento, John decide usar el número de periodo como la variable de predicción, o X . El primer valor de ventas y de su serie, \$16 850, tendrá un valor x de 01, el segundo tendrá un valor de 02 y así sucesivamente. Su razonamiento es que la tendencia creciente que él sabe que existe en sus ventas se contabilizará mediante un valor siempre en aumento de X para explicar sus datos de ventas.

Después de realizar el análisis de regresión en su computadora, John registra los siguientes valores:

$$t = 8.877 \quad r\text{-cuadrada} = .49$$

$$F = 78.81 \quad \hat{Y} = 19,092 + 2,861.57X$$

El alto valor de t le indica a John que la línea pendiente de su regresión de la muestra (2 861.57) es significativa; es decir, rechaza la hipótesis nula de que la pendiente en la población es cero. El alto valor de F le permite rechazar la hipótesis nula de que la regresión no es significativa y concluye que el valor de r cuadrada de la población es significativo.

John está decepcionado con el valor de r cuadrada de la muestra (49%). Él esperaba un valor más alto de modo que pudiera usar su ecuación de regresión simple para pronosticar con precisión sus ventas. Se da cuenta de que este valor bajo puede deberse a la estacionalidad de sus ventas mensuales, un hecho que él conocía antes de comenzar sus esfuerzos de pronóstico. Diferencias estacionales considerables darían por resultado puntos de datos mensuales que no se agruparían alrededor de la línea de regresión, así como un valor de r cuadrada no satisfactorio.

Hay otro aspecto en los resultados de su regresión que le molesta. En la impresión aparece la declaración: Durbin-Watson = 1.00228. Como no entiende esta información, llama a su profesor de estadística de la universidad. Después de describirle los valores de la impresión de su regresión, el profesor le dice: “Tengo que impartir una clase en este momento, pero el valor bajo en la estadística Durbin-Watson significa que violaste una de las suposiciones del análisis de regresión”.

TAREA

1. Haga un comentario sobre la creencia de John de que sus ventas mensuales son altamente estacionales y el “bajo” valor de r cuadrada.
2. ¿Cuál es su opinión con respecto a lo adecuado del método de pronóstico de John?
3. ¿Cómo fue que John violó una de las suposiciones del análisis de regresión?

CASO DE ESTUDIO 6.5

CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC), se describió en los caps. 1 (caso de estudio 1.2) y 3 (caso de estudio 3.4).

El director ejecutivo, Marv Hamishfeger, llegó a la conclusión de que la variable más importante que CCC requería pronosticar era el número de clientes por atender en el resto de 1993. Marv proporcionó a Dorothy Mercer los datos mensuales del número de clientes atendidos por CCC para el periodo de enero de 1985 a marzo de 1993 (véase Caso de estudio 4.3). En el Caso de estudio 4.3, Dorothy utilizó el análisis de correlación para explorar el patrón de datos. En el Caso de estudio 5.3, utilizó los modelos de promedios móviles y atenuación exponencial para pronosticar el resto de 1993.

Dorothy se preguntó si el análisis de regresión se podría emplear para desarrollar un buen modelo de pronóstico. Le preguntó a Marv si se le ocurrían algunas variables para usar como pronóstico potencial. Marv pensó que el número de personas que reúnen cupones de descuento pudiera estar relacionado con el número de clientes atendidos.

Dorothy sólo pudo encontrar datos sobre personas que reúnen cupones de descuento, para el periodo de enero de 1989 a diciembre de 1992:

1989:	24,450	24,761	25,397	25,617	25,283	25,242
	25,163	25,184	25,417	25,411	25,565	26,543
1990:	26,784	27,044	27,567	28,080	28,142	28,412
	28,161	27,936	28,423	28,366	29,029	29,035
1991:	29,254	29,962	30,499	30,879	30,995	31,356
	30,863	31,288	31,492	31,577	31,912	32,050
1992:	32,383	32,625	33,499	34,076	34,191	33,788
	33,556	33,751	33,777	33,769	34,077	34,232

Marv también tenía conocimiento de un índice de actividad empresarial, calculado para el condado por el Economic Development Council. Este índice de actividad empresarial fue un indicador de los cambios relativos en las condiciones globales de los negocios de la región. Los datos de este índice son:

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									

TAREA

1. Determine si hay una relación significativa entre el número de nuevos clientes atendidos y el número de personas que reúnen cupones de descuento y/o el índice de actividad empresarial. No olvide las posibles transformaciones de datos.
2. Desarrolle una ecuación de regresión y utilícela para pronosticar el número de nuevos clientes atendidos para los tres primeros meses de 1993.
3. Compare los resultados de su pronóstico con las observaciones reales de los tres primeros meses de 1993.
4. ¿Sería el índice de actividad empresarial un buen pronóstico del número de nuevos clientes?
5. Suponga que desarrolló una buena ecuación de regresión. ¿Podría emplearla para pronosticar el resto de 1993? Explique su respuesta.

VERSIÓN ESTUDIANTEL DE FORECAST PLUS

Se accesa el paquete de pronósticos mediante la selección de la opción 3 del menú principal. Si está usted utilizando discos (no en un sistema de disco duro), el programa le solicitará que inserte el disco del sistema principal, por medio del mensaje:

```
Insert MAIN SYSTEM DISK in default drive
Press <ENTER> when ready
```

NUMBER OF CLIENTS-LEAST SQUARES REGRESSION TREND ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

90 Observations in the Residual Series

Mean of the Residual Series = 0

Standard Deviation of the Residual Series = 24.2935

Lag	Value	T-Value	-1.0	0.0	+1.0
1	0.22	2.23	:	[:****] *	:
2	0.12	1.13	:	[:***]	:
3	-0.00	-0.00	:	[:]	:
4	0.01	-0.06	:	[:]	:
5	-0.06	-0.52	:	[* :]	:
6	-0.10	-0.90	:	[** :]	:
7	-0.07	-0.65	:	[** :]	:
8	-0.05	-0.50	:	[* :]	:
9	-0.12	-1.12	:	[*** :]	:
10	-0.07	-0.61	:	[** :]	:
11	0.03	0.28	:	[* :]	:
12	0.27	2.47	:	[:*****] *	:
13	0.07	0.60	:	[* :]	:
14	0.10	0.83	:	[: **]	:
15	-0.11	-0.95	:	[*** :]	:
16	-0.08	-0.69	:	[* :]	:
17	0.06	0.47	:	[* :]	:
18	-0.23	-1.87	:	[*****]	:
19	-0.12	-1.00	:	[*** :]	:
20	-0.07	-0.53	:	[* :]	:
21	0.03	0.25	:	[: *]	:
22	-0.12	-0.96	:	[*** :]	:
23	0.05	0.36	:	[* :]	:
24	0.27	2.13	:	[:*****] *	:
25	0.02	0.16	:	[* :]	:
26	0.00	0.03	:	[:]	:
27	-0.08	-0.64	:	[** :]	:
28	-0.08	-0.60	:	[** :]	:
29	0.03	0.23	:	[* :]	:
30	-0.36	-2.70	:	** [*****]	:
31	-0.14	-0.99	:	[*****]	:
32	-0.25	-1.74	:	[*****]	:
33	-0.08	-0.52	:	[* :]	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 33 Degrees of Freedom = 60.9

Probability = .002

NUMBER OF CLIENTS-LEAST SQUARES REGRESSION TREND ANALYSIS

SUMMARY STATISTICS

Mean of the Original Data = 122.15

Standard Deviation of the Original Data = 30.043

Mean of the Residuals = 0

Standard Deviation of the Residuals = 24.294

Mean Absolute % Error	Mean % Error	Mean Square Error
16.7797	-4.1396	584.2

REGRESSION STATISTICS

R-Squared: 0.3461

F Statistic: 51.35 DF = 1, 97

	Coefficient	Estd Std Error	T-Value
Intercept	91.3834	4.9457	18.48
Slope	0.6154	0.0859	7.17

Figura 6.9 Análisis de regresión de tendencia para el Caso de estudio 6.5.

En este punto, usted necesita sustituir el Data Management Disk por el Main System Disk. Presione <Enter> y el programa le solicitará el nombre del archivo:

```
What is the name of the data file? CREDIT
```

El menú le permite elegir entre 13 diferentes técnicas de pronóstico. Una muestra de esta pantalla se presentó en la sección de Forecast Plus del cap. 5.

Para poder utilizar como variable de predicción el número de personas que reúnen cupones de descuento o el índice de actividad, y el número de clientes atendidos como la variable independiente, tendríamos que utilizar el programa de regresión múltiple (véase cap. 7). En vez de ello, probaremos el programa de análisis de regresión de tendencia, opción 10, el cual utiliza el periodo como variable de predicción. La pantalla muestra lo siguiente:

```
*** Regression Trend Analysis ***

Least Squares (1) or Robust (2):                1

Length of pre-analysis smoothing:                1

Power Transformation:                            8
[A=Automatic (least squares only), 1=reciprocal,
2=1/Sqrt, 3=1/(4th root), 4=Log, 5=1/4th root,
6=Sqrt, 7=S curve, 8=none]

Number of forecasts                             1

Time plot of original data, forecasts, and error: Y
Residual autocorrelation function:              Y
Table of original data, forecasts, and error:   N
Summary statistics:                             Y
```

Press ENTER to accept:

Para ejecutar este análisis de regresión de tendencia, podemos utilizar las opciones por omisión que ya están dadas. La fig. 6.9 muestra la salida correspondiente a la respuesta *Y* en la función de autocorrelación de residuos y en las estadísticas de resumen, y *N* en el resto de las opciones.

Por último, el programa le preguntará si desea almacenar los pronósticos de los residuos:

```
Do you want to save the forecasts of residuals (y/n)?
```



```

A1:
MEMU
X-Range Y-Range Output-Range Intercept Reset Go Quit
Specify independent variables (X range)
      A      B      C      D      E      F      G      H
1      Y      X
2
3      10      1.3
4      6      2      Constant      32.13592
5      5      1.7      Std Err of Y Est      2.725453
6      12      1.5      R Squared      0.745613
7      10      1.6      No. of Observations      10
8      15      1.2      Degrees of Freedom      8
9      5      1.6
10     12      1.4      X Coefficient(s) -14.5388
11     17      1      Std Err of Coef. 3.002445
12     20      1.1
13
14
15
16
17
18
19
20
03-Jan-94 06:02 PM
NUM

```

Figura 6.10 Análisis de regresión utilizando Lotus, para el ejemplo del Sr. Bump.

PRONÓSTICO MEDIANTE HOJAS DE CÁLCULO

El problema. En los ejemplos 6.1 al 6.8, se utilizó el análisis de regresión para determinar si se podría emplear el precio de venta para pronosticar las ventas semanales de galones de leche (tabla 6.1), para el ejemplo del Sr. Bump. Para este pronóstico, se puede usar Lotus 1-2-3 (versión 2.3).

La solución mediante hoja de cálculo

1. Se captura la variable *Y*, ventas semanales en las celdas A3 a la A12, y la variable *X*, precio de venta, en las celdas B3 a la B12 de la hoja de cálculo (véase fig. 6.10).
2. Escribir /DATA REGRESSION o /DR, lo que equivale a seleccionar DATA FILE del menú principal y REGRESSION en el submenú. El menú de regresión (véase fig. 6.10) muestra lo siguiente:

X-Range Y-Range Output-Range Intercept Reset Go Quit

Se selecciona *X-Range* y la pantalla solicita el rango de las *X*:

```
Enter independent variables, or X range: B3. . B12
```

Usted puede elegir entre escribir B3 .. B12 o definir la variable X resaltando dicho rango. Al seleccionar Y -Range, el programa solicita el rango de las Y :

```
Enter dependent variables, or Y range: A3..A12
```

Una vez definido el rango de Y , se selecciona Output-Range y en la pantalla se pide el rango donde se ubicarán los resultados:

```
Enter output range: D3..G11
```

Después de indicar este rango, seleccione GO para ejecutar la regresión. La salida se muestra en la fig. 6.10.

3. Para guardar la hoja de cálculo para uso futuro, escriba /FILE SAVE o /FS. Lotus le solicitará el nombre del archivo a guardar:

```
Enter name of file to save: A:\*.WK1
```

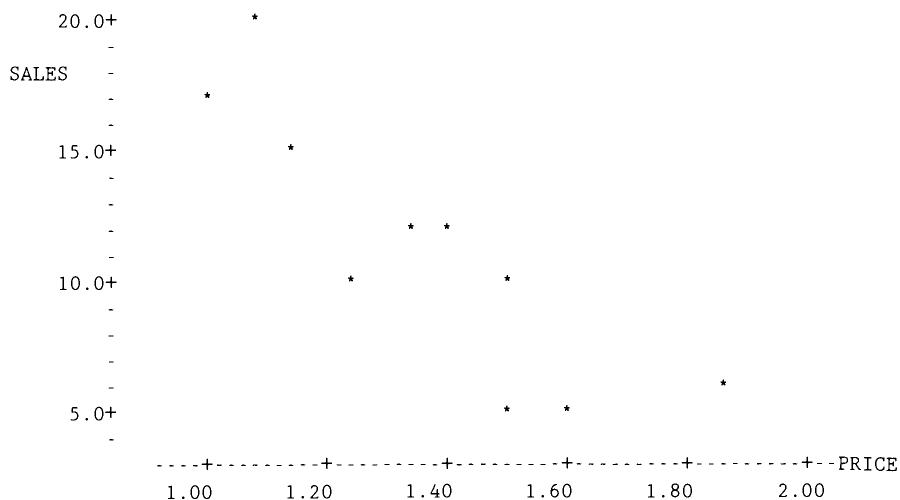
Asegúrese de que el directorio es el correcto y escriba BUMP. El programa borrará el asterisco, el punto y la extensión, pero no el directorio por omisión. Oprima <Enter> y Lotus guardará automáticamente su archivo con la extensión .WK1.

PAQUETE DE CÓMPUTO MINITAB

El problema. En los problemas 6.1 y 6.2, el Sr. Bump quería utilizar una ecuación de regresión para pronosticar ventas de leche.

La solución Minitab

```
MTB > SET C1
DATA > 10 6 5 12 10 15 5 12 17 20
MTB > END
MTB > SET C2
DATA > 1.30 2.00 1.70 1.50 1.60 1.20 1.60 1.40 1.00 1.10
MTB > END
MTB > NAME C1 'SALES' C2 'PRICE'
MTB > PLOT C1 VS C2
```



MTB > CORR C1-C2

Correlation of SALES and PRICE = -0.863

MTB > BRIEF 3

MTB > REGRESS C1 1 PREDICTOR C2;

SUBC > PREDICT 1.63.

The regression equation is

SALES = 32.1 - 14.5 PRICE

Predictor	Coef	Stdev	t-ratio	P
Constant	32.136	4.409	7.29	0.000
PRICE	-14.539	3.002	-4.84	0.000

s = 2.725 R-sq = 76.6% R-sq(adj) = 71.4%

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	1	174.18	174.18	23.45	0.000
Error	8	59.42	7.43		
Total	9	233.60			

Obs.	C2	C1	Fit	Stdev.Fit	Residual	St.Resid
1	1.30	10.000	13.235	0.959	-3.235	-1.27
2	2.00	6.000	3.058	1.889	2.942	1.50
3	1.70	5.000	7.420	1.163	-2.420	-0.98
4	1.50	12.000	10.328	0.880	1.672	0.65
5	1.60	10.000	8.874	0.987	1.126	0.44
6	1.20	15.000	14.689	1.123	0.311	0.13
7	1.60	5.000	8.874	0.987	-3.874	-1.52
8	1.40	12.000	11.782	0.870	0.218	0.08
9	1.00	17.000	17.597	1.577	-0.597	-0.27
10	1.10	20.000	16.143	1.336	3.857	1.62
	Fit	Stdev.Fit		95% C.I.		95% P.I.
	8.438	1.034		(6.054, 10.822)		(1.714, 15.161)

- Los comandos SET se usan para capturar los datos de ventas en C1 y los datos de precio en C2.
- El comando NAME se emplea para asignar nombres a las variables en cada columna.
- El comando PLOT se utiliza para desarrollar un diagrama de dispersión con C1 sobre el eje de las Y y C2 sobre el eje de las X .
- El comando CORR se usa para calcular el coeficiente de correlación entre C1 y C2.
- El comando BRIEF 3 se emplea para obtener una salida que incluya una tabla de residuos.
- El comando REGRESS se utiliza para desarrollar un análisis de regresión con las ventas como la variable dependiente (Y) y el precio como la variable independiente (X).
- El subcomando PREDICT se usa para calcular un intervalo de confianza del 95%, en este caso para un valor de x de 1.63.

PAQUETE DE CÓMPUTO SAS

Problema. En los ejemplos 6.1 y 6.2, el Sr. Bump desea emplear la correlación y la regresión para determinar si puede usar el precio de la leche para predecir las ventas semanales.

Solución SAS

```
$ SAS/CC=PRINT
OPTIONS PAGESIZE=60 LINESIZE=80;
TITLE "REGRESSION ANALYSIS FOR EXAMPLES 6.1 & 6.2";
DATA BUMP;
INPUT Y X;
CARDS;
10 1.30
6 2.00
5 1.70
12 1.50
10 1.60
15 1.20
5 1.60
12 1.40
17 1.00
20 1.10
PROC PRINT;
PROC CORR;
PROC REG;
MODEL Y=X;
ENDSAS;
```

REGRESSION ANALYSIS FOR EXAMPLES 6.1 & 6.2						
	OBS	Y	X			
	1	10	1.3			
	2	6	2.0			
	3	5	1.7			
	4	12	1.5			
	5	10	1.6			
	6	15	1.2			
	7	5	1.6			
	8	12	1.4			
	9	17	1.0			
	10	20	1.1			
Correlation Analysis						
Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
	10	11.20000	5.09466	112.00000	5.00000	20.00000
	10	1.44000	0.30258	14.40000	1.00000	2.00000
Pearson Correlation Coefficients / Prob > :R: under Ho: Rho=0 / N = 10						
		Y	X			
	Y	1.00000	-0.86349			
		0.0	0.0013			
	X	-0.86349	1.00000			
		0.0013	0.0			
Model: MODEL1						
Dependent Variable: Y						
Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F	
Model	1	174.17524	174.17524	23.448	0.0013	
Error	8	59.42476	7.42809			
C Total	9	233.60000				
Root MSE		2.72545	R-square	0.7456		
Dep Mean		11.20000	Adj R-sq	0.7138		
C.V.		24.33440				
Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > :T:	
INTERCEP	1	32.135922	4.40858773	7.289	0.0001	
X	1	-14.538835	3.00244533	-4.842	0.0013	

Figura 6.11 Salida de SAS para los ejemplos 6.1 y 6.2.

- La declaración TITLE da nombre al programa SAS.
- La declaración DATA asigna un nombre de archivo a los datos.
- La declaración INPUT nombra y especifica el orden de los dos campos de datos Y y X .
- La declaración CARDS indica que las 10 siguientes líneas son imágenes de tarjetas que representan el volumen semanal de ventas de galones de leche y el precio de venta.
- El comando PROC PRINT instruye a SAS para que enliste los datos.
- El comando PROC CORR indica que se desea una matriz de correlación.
- Los subcomandos PROC REG y MODEL indican que X y Y son las variables de la regresión, siendo Y la variable dependiente y X la variable independiente.

La salida de cómputo de esta ejecución de SAS se muestra en la fig. 6.11.

BIBLIOGRAFÍA

DRAPER, N. y SMITH, H., *Applied Regression Analysis*. Nueva York: John Wiley & Sons, 1966.

KLEINBAUM, D. y KUPPER, K., *Applied Regression Analysis and Other Multivariable Methods*. North Scituate, Mass.: Duxbury Press, 1978.

NETER, J., WASSERMAN, W. y KUTNER, M. *Applied Linear Regression Models*, Homewood, Ill.: Richard D. Irwin, Inc. 1983.

Regresión múltiple

ACTIVIDADES

En la regresión simple, se investiga la relación entre las variables independiente y dependiente. A menudo, la relación entre dos variables permite a una persona predecir con precisión la variable dependiente a partir del conocimiento de la variable independiente. Por desgracia, muchas de las situaciones de la vida real no son tan simples. Por lo regular, se necesita más de una variable independiente para pronosticar con precisión la variable dependiente. Cuando se emplea más de una variable independiente o de predicción, el problema se convierte en uno para el análisis de regresión múltiple. Los conceptos básicos siguen siendo los mismos, sólo que se utiliza más de una variable independiente para pronosticar la variable dependiente.

La regresión múltiple comprende el uso de más de una variable independiente para pronosticar una variable dependiente.

VARIABLES DE PREDICCIÓN

Como ejemplo, regresemos al problema en el que se pronostica el volumen de ventas de galones de leche a partir del conocimiento del precio por galón. El Sr. Bump se enfrenta con el problema de formular un pronóstico que no es del todo preciso. Él puede explicar casi el 75% de la varianza total del volumen de ventas de galones de leche vendidos mediante el uso de una variable independiente. De modo que el 25% ($1 - r^2$) de la variación total queda sin explicación. En otras palabras, a partir de la evidencia de la muestra, el Sr. Bump conoce el 75% de lo que debería saber para pronosticar a la perfección el volumen de ventas. Para hacer más preciso el trabajo de pronóstico, debe encontrar otra variable de predicción que le permita explicar más la varianza total. Si el Sr. Bump puede reducir la varianza no explicada, su pronóstico implicará menos error y más precisión.

Se debe buscar otra variable independiente que se relacione con el volumen de ventas de galones de leche. Sin embargo, esta nueva variable independiente o variable de predicción, no puede tener una relación muy alta con la variable independiente que ya se

está empleando (precio por galón). Si las dos variables independientes están altamente relacionadas entre sí, explicarán la misma variación y la incorporación de la segunda variable no mejorará el pronóstico. En áreas como la econometría y la estadística aplicada, existe una gran preocupación con respecto a este problema de intercorrelación entre las variables independientes que con frecuencia se denomina *colinealidad*. Una solución sencilla al problema de dos variables independientes altamente relacionadas, consiste simplemente en *no* utilizarlas juntas. El problema de colinealidad se abordará más adelante en este capítulo.

Para resumir, los atributos de una *buen variable de pronóstico* son: Una buena variable de pronóstico está relacionada con la variable dependiente. Una buena variable de pronóstico no está altamente relacionada con ninguna otra variable independiente.

MATRIZ DE CORRELACIÓN

El Sr. Bump decide que los gastos de publicidad podrían ayudarle a mejorar su pronóstico de volumen semanal de ventas. Investiga la relación entre los gastos de publicidad, el volumen de ventas y el precio por galón mediante el examen de una matriz de correlación. La *matriz de correlación* se crea mediante el cálculo de coeficientes de correlación simples para cada combinación de pares de variables.

En la tabla 7.1 se ilustra un ejemplo de matriz de correlación. El coeficiente de correlación que indica la relación existente entre las variables 1 y 2 se representa como r_{12} . Nótese que el primer subíndice (r_{12}) se refiere al renglón y el segundo subíndice (r_{12}) se refiere a la columna. Este enfoque estandarizado le permite a uno determinar, a simple vista, la relación entre cualquier par de variables. Por supuesto, la relación entre las variables 1 y 2 (r_{12}) es exactamente la misma que para las variables 2 y 1 (r_{21}). De ahí que sólo sea necesaria la mitad de la matriz. Además, la diagonal principal siempre contendrá unos, ya que siempre relaciona una variable consigo misma (r_{11} , r_{22} , r_{33}).

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 CÓMPUTO PARA LOS DATOS DEL SR.BUMP

VARIABLES	VARIABLES		
	VENTAS, 1	PRECIO, 2	PUBLICIDAD 3
Ventas, 1	1.00	-.86	.89
Precio, 2		1.00	-.65
Publicidad, 3			1.00

El Sr. Bump procesa sus datos en la computadora y obtiene la matriz de correlación que se muestra en la tabla 7.2. (La mayoría de los programas de regresión múltiple proporcionan matrices de correlación.) Una investigación de las relaciones entre los gastos de publicidad, el volumen de ventas y el precio por galón, indica que la nueva variable independiente podría contribuir a mejorar el pronóstico. La matriz de correlación muestra que los gastos de publicidad tienen una relación positiva alta ($r_{13} = .89$) con la variable dependiente, volumen de ventas; y una relación negativa moderada ($r_{23} = -.65$) con la variable independiente, precio por galón. Esta combinación de relaciones podría permitir que los gastos de publicidad explicaran parte de la varianza total del volumen de ventas que aún no se ha explicado por medio del precio por galón. Como se verá, cuando se emplean tanto el precio por galón como los gastos de publicidad para estimar el volumen de ventas, R^2 se incrementa al 93.2%.

El análisis de la matriz de correlación es un paso inicial importante en la solución de cualquier problema que comprenda varias variables independientes.

ECUACIÓN DE REGRESIÓN MÚLTIPLE

En la regresión simple, la variable dependiente se puede representar por Y y la variable independiente mediante X . En el análisis de regresión múltiple se utilizan X con subíndices para representar las variables independientes. La variable dependiente se continúa representando con Y , y las variables independientes se representan como $X_2, X_3, \dots, X_n$. En este sistema de notación, la nueva ecuación de regresión del Sr. Bump se escribe como

$$\hat{Y} = b_0 + b_2X_2 + b_3X_3 \quad (7.1)$$

en donde

Y = volumen de galones vendidos, estimado por la ecuación de regresión
 b_2, b_3 = coeficientes de regresión neta (el mejor conjunto ponderado entre las dos variables independientes, con el fin de lograr la máxima predicción)

X_2 = Precio por galón

X_3 = Gastos de publicidad (en cientos de dólares)

b_0 = Constante o intersección con Y

El análisis utiliza el método de mínimos cuadrados para obtener la ecuación de regresión lineal de mejor ajuste para tres variables. Mientras que en el problema con dos variables el método de mínimos cuadrados producía la línea recta de mejor ajuste, en este problema de tres variables, se emplea el método para obtener el plano de mejor ajuste (véase fig. 7.1). Los puntos se grafican en tres dimensiones a lo largo de los ejes Y , X_2 y X_3 . Los puntos se ubican por arriba y por abajo del plano, de tal forma que $\Sigma (Y - \hat{Y})^2$ es mínima. Los valores de b_2 y b_3 se derivan como el mejor conjunto de ponderaciones que minimizan la suma de las distancias al cuadrado entre los puntos de datos y el plano de regresión múltiple.

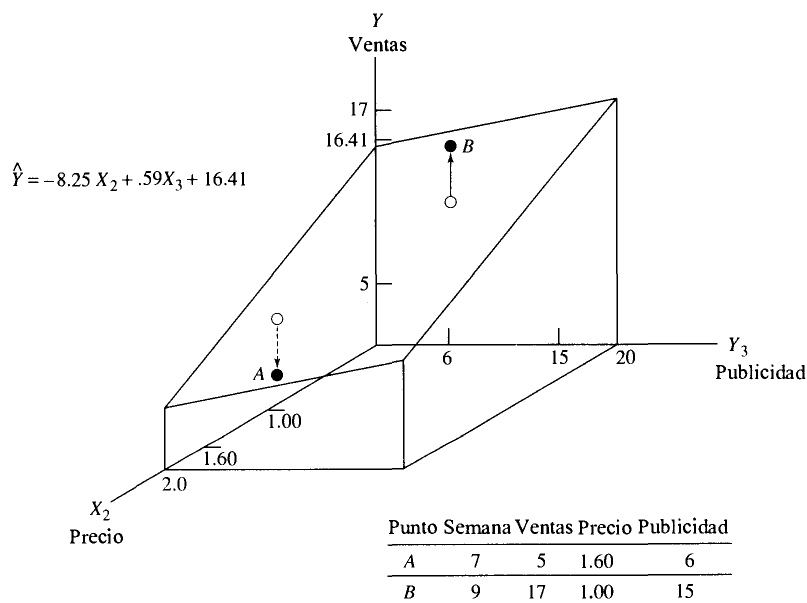


Figura 7.1 Plano de regresión múltiple para los datos del Sr. Bump.

En el cap. 6, el ejemplo comprendía dos variables y del proceso de minimización resultaban dos ecuaciones. Ahora, se deben resolver tres ecuaciones, para determinar los valores de b_2 , b_3 y b_0 .¹ Estas ecuaciones son

$$\begin{aligned}
 \Sigma Y &= nb_0 + b_2 \Sigma X_2 + b_3 \Sigma X_3 \\
 \Sigma X_2 Y &= b_0 \Sigma X_2 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 \\
 \Sigma X_3 Y &= b_0 \Sigma X_3 + b_2 \Sigma X_2 X_3 + b_3 \Sigma X_3^2
 \end{aligned} \tag{7.2}$$

¹ Como se muestra en el apéndice A, una función de la forma $F(b_2, b_3, b_0) = \Sigma (y - b_0 - b_2 x_2 - b_3 x_3)^2$ se minimiza a través del método de cálculo de tomar las derivadas parciales con respecto a b_0 , b_2 , b_3 e igualando estas derivadas a cero.

Ejemplo 7.1

Los cálculos de las sumas requeridas para el problema del Sr. Bump se muestran en la tabla 7.3. Sustituyendo en la ecuación 7.2 nos da

$$\begin{aligned} 112 &= 10b_0 + 14.4b_2 + 114b_3 \\ 149.3 &= 14.4b_0 + 21.56b_2 + 155.3b_3 \\ 1,480 &= 114b_0 + 115.3b_2 + 1,522b_3 \end{aligned}$$

Con frecuencia se denomina a estas ecuaciones como *ecuaciones normales* y se pueden resolver simultáneamente por métodos matriciales. Este enfoque lo exponen Younger, Neter y Wasserman, cuyos textos podrá consultar si le interesan las matemáticas del procedimiento.

En este ejemplo, se usó un programa de cómputo para obtener

$$\begin{aligned} b_2 &= -8.2476 \\ b_3 &= .5851 \\ b_0 &= 16.4064 \end{aligned}$$

Sustituyendo en la ecuación 7.1, la ecuación de regresión múltiple queda ahora como

$$\hat{Y} = 16.41 - 8.25 X_2 + .59 X_3$$

Esta ecuación resulta útil para pronosticar las ventas de la próxima semana. Si se planea un precio unitario de \$1.50 y gastos de publicidad de \$1000, el pronóstico es de 9,930 galones; esto es

$$\begin{aligned} \hat{Y} &= 16.41 - 8.25X_2 + .59X_3 \\ &= 16.41 - 8.25(1.5) + .59(10) \\ &= 9.93 \text{ (miles de galones)} \end{aligned}$$

TABLA 7.3 CÁLCULOS PARA EL ANÁLISIS DE REGRESIÓN MÚLTIPLE PARA LOS DATOS DEL SR. BUMP

SEMANA	VENTAS (EN MILES),	PRECIO POR GALÓN, X_2	PUBLICIDAD (CIENTOS DE DÓLARES)	X_2Y	X_3Y	X_2X_3	Y^2	X_2^2	X_3^2
	Y		X_3						
1	10	\$1.30	9	13	90	11.7	100	1.69	81
2	6	2.00	7	12	42	14.0	36	4.00	49
3	5	1.70	5	8.5	25	8.5	25	2.89	25
4	12	1.50	14	18	168	21.0	144	2.25	196
5	10	1.60	15	16	150	24.0	100	2.56	225
6	15	1.20	12	18	180	14.4	225	1.44	144
7	5	1.60	6	8	30	9.6	25	2.56	36
8	12	1.40	10	16.8	120	14.0	144	1.96	100
9	17	1.00	15	17	255	15.0	289	1.00	225
10	20	1.10	21	22	420	23.1	400	1.21	441
Totales	112	14.40	114	149.3	1,480	155.3	1,488	21.56	1,522
Medias	11.2	1.44	11.4						

El análisis para el ejemplo del cap. 6 para dos variables se puede realizar con una calculadora de mano. Sin embargo, con más de dos variables el análisis se complica de manera creciente, ya que el número de ecuaciones a resolver aumenta con cada nueva variable independiente. El mejor enfoque consiste en utilizar alguno de los diversos programas de cómputo disponibles para regresión múltiple.

COEFICIENTES DE REGRESIÓN

En el cap. 6 se expuso la interpretación de b_0 y b en la ecuación de regresión. Ahora considere la interpretación de b_0 , b_2 y b_3 en la ecuación de regresión múltiple. El valor de b_0 es de nuevo una intersección con Y . Sin embargo, ahora se interpreta como el valor $\hat{Y}$ cuando tanto X_2 como X_3 son iguales a cero. A los valores b_2 y b_3 se les denomina *coeficientes de regresión neta*. Cada uno mide cambios promedio en Y por unidad de cambio en la variable independiente relevante. No obstante, debido a que se está midiendo la influencia simultánea de todas las variables independientes sobre Y , el efecto neto de X_2 (o de cualquier otra X), se debe medir por separado de cualquier influencia de otras variables. De ahí que se diga que b_2 mide el promedio de cambio en Y por unidad de cambio en X_2 , manteniendo constantes las otras variables independientes.

El *coeficiente de regresión neta* mide el promedio de cambio en la variable dependiente por unidad de cambio en la variable independiente relevante, manteniendo constantes las demás variables independientes.

En el ejemplo actual, el valor de -8.25 de b_2 indica que cada incremento de 1 centavo en un galón de leche cuando los gastos de publicidad se mantienen *constantes*, reduce la cantidad de compras en un promedio de 8.25 galones. En forma semejante, el valor $.59$ de b_3 significa que si se aumentan en \$100 los gastos de publicidad, manteniendo constante el precio por galón, entonces el volumen de ventas se incrementará en un promedio de 590 galones.

Ejemplo 7.2

Para ilustrar estas influencias, consideremos un ejemplo en el que el precio es de \$ 1.00 por galón y se gastan \$1000 en publicidad. Así,

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25 X_2 + .59 X_3 \\ &= 16.41 - 8.25(1.00) + .59(10) \\ &= 16.41 - 8.25 + 5.9 = 14.06 \text{ o } 14,060 \text{ galones}\end{aligned}$$

Las ventas se pronostican en 14,060 galones de leche.

¿Cuál es el efecto en las ventas de un incremento de 1 centavo en el precio si los gastos de publicidad siguen siendo \$1,000?

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25(1.01) + .50(10) \\ &= 16.41 - 8.3325 + 5.9 = 13.9775\end{aligned}$$

Nótese que las ventas disminuyen en 82.5 galones ($14.06 - 13.9775 = .0825$).

¿Cuál es el efecto en las ventas de un incremento de \$100 en publicidad, si el precio se mantiene constante en \$1.00?

$$\begin{aligned}\hat{Y} &= 16.41 - 8.25(1.00) + .59(11) \\ &= 16.41 - 8.25 + 6.49 = 14.65\end{aligned}$$

Nótese que las ventas se incrementan en 590 galones ($14.65 - 14.06 = .59$).

INFERENCIA ESTADÍSTICA EN LA REGRESIÓN MÚLTIPLE

Cuando cada una de las mediciones que se emplean en el desarrollo de una ecuación de regresión múltiple representa una muestra de probabilidad de una población específica, se pueden hacer inferencias estadísticas sobre los parámetros de la población, con base en la siguiente ecuación de regresión múltiple:

$$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \quad (7.3)$$

Por ejemplo, se pueden usar datos de muestra para desarrollar una ecuación de regresión múltiple de muestra en la forma

$$\hat{Y} = b_0 + b_2 X_2 + b_3 X_3$$

en donde b_0 , b_2 y b_3 son estimaciones lineales, eficientes, no desviadas de los parámetros correspondientes de la población β_0 , β_2 y β_3 . Ya que en general se usan datos de muestra, el analista está por lo regular interesado en hacer inferencias sobre los parámetros de la población.

En el cap. 6 se expusieron ciertas suposiciones que son la base de este procedimiento de inferencia. El resumen es el siguiente:

1. Los valores de Y presentan una distribución normal con respecto al plano de regresión múltiple.
2. Los puntos de dispersión alrededor del plano de regresión permanecen constantes en cualquier parte del plano.
3. Los términos de error (ϵ) son independientes entre sí.
4. Existe una relación lineal entre cada X y Y en la población.

RESIDUOS

Como se mencionó en el cap. 6, un residuo es la diferencia entre lo que realmente se observa y lo que se pronostica mediante la ecuación de regresión ($Y - \hat{Y}$). Así, la ecuación de regresión múltiple del Sr. Bump para cada uno de los volúmenes semanales de venta anteriores es

$$Y = b_0 + b_2 X_2 + b_3 X_3 + e \quad (7.4)$$

en donde

Y = volumen real de ventas semanales en galones

b_2, b_3 = coeficientes de regresión neta

X_2 = precio por galón

X_3 = gastos de publicidad, en cientos de dólares

b_0 = constante

e = residuo [diferencia entre el valor real de Y y el valor de pronóstico de $Y(\hat{Y})$].

Si se sustituye $\hat{Y}$ por $b_0 + b_2X_2 + b_3X_3$ en la ecuación (7.4),

$$Y = \hat{Y} + e$$

ó

$$e = Y - \hat{Y}$$

Ejemplo 7.3

Si Mr. Bump sustituye los datos de la primera semana en la ecuación 7.1, se obtiene un pronóstico para las ventas de esa semana de 10,950 galones de leche.

$$\begin{aligned}\hat{Y} &= 16.4064 - 8.2476X_2 + .5851X_3 \\ &= 16.4064 - 8.2476(1.30) + .5851(9) = 10.9504\end{aligned}$$

Sin embargo, él ya sabe que las ventas de esa semana fueron de 10000 galones. Su pronóstico falló por 950 galones. En otras palabras, el residuo de este pronóstico es -950 galones.

$$\begin{aligned}e &= Y - \hat{Y} \\ &= 10.000 - 10.950 = -.950\end{aligned}$$

La tabla 7.4 ilustra el concepto de residuo para cada uno de los diez puntos de datos.

TABLA 7.4 RESIDUOS DEL MODELO COMPLETO PARA LOS DATOS DEL SR. BUMP

Y	X_2	X_3	PREDICCIÓN DE $Y(\hat{Y})$, USANDO $\hat{Y} = 16.4064$ $-8.2476X_2 + .5851X_3$	RESIDUO, $(Y - \hat{Y})$	$(Y - \hat{Y})^2$
10	1.30	9	10.9504	-.95041	.903279
6	2.00	7	4.0069	1.99311	3.972487
5	1.70	5	5.3110	-.31097	.096702
12	1.50	14	12.2264	-.22639	.051252
10	1.60	15	11.9867	-1.98673	3.947095
15	1.20	12	13.5305	1.46953	2.159518
5	1.60	6	6.7208	-1.72083	2.961256
12	1.40	10	10.7108	1.28925	1.662166
17	1.00	15	16.9353	.06471	.004187
20	1.10	21	19.6211	.37888	.143550
Sumas				.00000	15.901493

Los valores b_2 y b_3 se derivan como el mejor conjunto de ponderaciones que minimizan la suma de las distancias al cuadrado entre los puntos de datos y el plano de regresión múltiple. Al sumarizar los residuos $[\Sigma (Y - \hat{Y})]$ para los diez puntos de datos, la columna suma cero $\Sigma (Y - \hat{Y}) = 0$. Cuando los errores se elevan al cuadrado y se suman $[\Sigma (Y - \hat{Y})^2]$, la suma de los cuadrados es un valor mínimo, $\Sigma (Y - \hat{Y})^2 = 15.901$. No se pueden encontrar otros valores de b para estos puntos de datos, que produzcan una suma de residuos cuadrados menor de 15.901.

Cuando el Sr. Bump utilizó sólo el precio por galón para estimar la cantidad de galones vendidos, la suma de los residuos cuadrados $\Sigma (Y - \hat{Y})^2$ fue de 59.4. Al emplear el precio por galón y los gastos de publicidad como variables de predicción, la suma de los residuos cuadrados se reduce a 15.9, de modo que el Sr. Bump puede ahora explicar el 93.2% de la varianza del volumen de ventas, a través del conocimiento de las relaciones entre el precio por galón, los gastos de publicidad y el volumen de ventas.²

Desviación total = Desviación explicada – Desviación no explicada

$$\Sigma(Y - \bar{Y})^2 = \Sigma(\hat{Y} - \bar{Y})^2 + \Sigma(Y - \hat{Y})^2$$

$$233.6 = 217.7 + 15.9$$

$$R^2 = 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2} = 1 - \frac{15.9}{233.6} = 1 - .068 = .932$$

La información derivada hasta este punto se resume en la tabla 7.5.

TABLA 7.5 RESUMEN DEL ANÁLISIS DE LOS DATOS DEL SR. BUMP

VARIABLES EMPLEADAS PARA EXPLICAR LA VARIANZA DE Y	R^2	$\Sigma(Y - \hat{Y})^2$
Ninguna	.00	233.6
Precio	.75	59.4
Precio y gastos de publicidad	.93	15.9

ERROR ESTÁNDAR DE LA ESTIMACIÓN

Tal como en una regresión simple, el error estándar de la estimación es la desviación estándar de los residuos. Y mide la dispersión típica de los valores de Y alrededor del plano de regresión. El error estándar de la estimación se escribe como

$$s_{y \cdot x_2 \times x_3} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - k}} \quad (7.5)$$

² Las desviaciones se utilizan aquí para simplificar el ejemplo. Si las desviaciones se dividen entre los correspondientes grados de libertad, da como resultado las varianzas.

en donde

$S_{Y \cdot X_2 X_3}$ = error estándar de la estimación de la variable dependiente Y , a la cual se aplica una regresión contra las variables independientes X_2 y X_3 .

Y = volumen real de ventas semanales, en galones

$\hat{Y}$ = volumen estimado de galones vendidos a partir de la ecuación de regresión

n = número de observaciones

k = número de parámetros linealmente independientes a estimar en la ecuación de regresión múltiple (b 's).

El *error estándar de la estimación* mide la cantidad estándar en que los valores reales (Y) difieren de los valores estimados ($\hat{Y}$).

Ejemplo 7.4

El error estándar para los datos de el Sr. Bump es

$$S_{Y \cdot X_2 X_3} = \sqrt{\frac{15.901493}{10 - 3}} = \sqrt{2.27} = 1.51$$

La cantidad típica en que el valor real de volumen de leche vendido difiere de lo pronosticado mediante la ecuación de regresión múltiple es 1,510 galones.

SALIDA DE CÓMPUTO

En la tabla 7.6 se presenta la salida de computadora para el problema de el Sr. Bump. Una revisión de esta salida nos conduce a las siguientes observaciones (las explicaciones se relacionan por número con la tabla 7.6).

1. Los coeficientes de regresión son -8.2476 para el precio y $.5851$ para los gastos de publicidad. La ecuación de regresión es $\hat{Y} = 16.4064 - 8.2476X_2 + .5851X_3$.
2. La ecuación de regresión explica el 93.2% de la varianza de los volúmenes de ventas.
3. El error estándar de la estimación es 1.507 galones. Este valor es la cantidad estándar o típica en que el valor real difiere del valor de pronóstico.
4. En el cap. 6 se probó el coeficiente de correlación para determinar si era diferente de cero. En esta ecuación, ambos coeficientes de regresión son significativos, como lo indican los valores calculados para t de -3.76 y 4.38 (ya que, la tabla C.3 del Apéndice nos dice que para 7 grados de libertad, $t_{.01} = 3.499$).
5. La matriz de correlación se demostró en la tabla 7.2.
6. El cálculo de la desviación de la suma de cuadrados (15.901) se demostró en la tabla 7.4. A este término se le denomina suma de los residuos cuadrados $[\Sigma(Y - \hat{Y})^2]$.
7. El cálculo de la suma de cuadrados atribuible a la regresión (217.699), se expuso en la sección de residuos. A este término se le denomina suma explicada de cuadrados $[\Sigma(\bar{Y} - \hat{Y})^2]$.

TABLA 7.6 SALIDA DE CÓMPUTO PARA EL PROBLEMA DE REGRESIÓN MÚLTIPLE DEL SR. BUMP

VARIABLE	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD. ERROR OF REG. COEF.	COMPUTED T VALUE
2	1.44000	0.30258	-0.86349	-8.24760 (1)	2.19605	-3.75565 (4)
3	11.40000	4.97103	0.89150	0.58510 (1)	0.13367	4.37716 (4)
DEPENDENT						
1	11.20000	5.09466				
INTERCEPT		16.40639 (1)		MULTIPLE CORRELATION		0.96536
STD. ERROR OF ESTIMATE		1.50719 (3)		R SQUARED		0.93193 (2)
				CORRECTED R SQUARED		0.91248 (11)
ANALYSIS OF VARIANCE FOR THE REGRESSION						
SOURCES OF VARIATION		DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE	
ATTRIBUTABLE TO REGRESSION		2	217.699 (7)	108.849	47.917 (10)	
DEVIATION FROM REGRESSION		7	15.901 (6)	2.272		
TOTAL		9	233.600 (8)			
CORRELATION MATRIX: 3 BY 3						
VAR.	1	2	3 (5)			
1	1.000	-0.863	0.891			
2	-0.863	1.000	-0.654			
3	0.891	-0.654	1.000			
TABLE OF RESIDUALS (9)						
CASE NO.	Y VALUE	Y ESTIMATE	RESIDUAL			
1	10.00000	10.95041	-0.95041			
2	6.00000	4.00689	1.99311			
3	5.00000	5.31097	-0.31097			
4	12.00000	12.22639	-0.22639			
5	10.00000	11.98673	-1.98673			
6	15.00000	13.53047	1.46953			
7	5.00000	6.72083	-1.72083			
8	12.00000	10.71075	1.28925			
9	17.00000	16.93529	0.06471			
10	20.00000	19.62112	0.37888			

8. El cálculo de la suma total de cuadrados (233.6) se demostró en la tabla 6.5. A este término se le denomina suma total de cuadrados $[\Sigma(Y - \bar{Y})^2]$ y se puede calcular adicionando la suma de residuos cuadrados y la suma explicada de cuadrados.

9. La columna de residuos se demostró en la tabla 7.4.

10. El valor calculado $F(47.917)$ se usa para probar si la ecuación de regresión ($\hat{Y} = 16.4064 - 8.2476X_2 + .5851X_3$) explica un porcentaje significativo de la varianza de la variable Y . El cálculo del valor F es

$$F = \frac{(.93193 - 0)/(3 - 1)}{(1 - .93193)/(10 - 3)} = 47.917$$

considerando que el valor F para 2 y 7 grados de libertad en la tabla C.6 es de 9.55 al nivel de significancia de .01, se puede concluir que la ecuación de regresión múltiple explica un porcentaje significativo de la varianza de la variable Y .

11. El cálculo del valor corregido o ajustado de R^2 es

$$R_c^2 = 1 - \frac{\frac{\sum(Y - \hat{Y})^2}{n - k}}{\frac{\sum(Y - \bar{Y})^2}{n - 1}} = \frac{\frac{15.9}{10 - 3}}{\frac{233.6}{10 - 1}} = .91248$$

VARIABLES FICTICIAS

Abrimos este tema observando el ejemplo siguiente:

Ejemplo 7.5

Suponga que un analista desea investigar qué tan bien puede un examen de aptitud en particular, predecir el desempeño en el trabajo. Ocho mujeres y siete hombres hicieron la prueba, que mide la destreza manual en el manejo de objetos pequeños. Después, cada sujeto tomó un mes de capacitación intensiva como ensambladores de partes electrónicas, seguido de un mes en el puesto de ensamblador, durante el cual se evaluó su productividad mediante un índice con valores en un intervalo de 0 a 10 (0 significa improductivo),

Los datos se presentan en la tabla 7.7 y en la fig. 7.2 se muestra el diagrama de dispersión correspondiente. Cada trabajador que es mujer se representa con un 0 y cada hombre con un 1.

De inmediato resulta evidente que la relación entre este examen de aptitud y el desempeño en el trabajo sigue dos patrones distintos, uno que se aplica a las mujeres y otro a los hombres.

En ocasiones es necesario determinar cómo se relaciona una variable dependiente con una independiente cuando un factor *cualitativo* ejerce influencia en la situación. Esta relación se logra creando una *variable ficticia*. Existen muchas formas de identificar cuantitativamente las clases de una variable cualitativa. En este texto emplearemos 0 y 1.

Las *variables ficticias* o de indicador se usan para determinar la relación entre variables cualitativas independientes y una variable dependiente.

En la fig. 7.3 se muestra la técnica de la variable ficticia. Los puntos de datos para mujeres se muestran como 0 y el 1 representa a los hombres. Se trazan dos líneas paralelas en el diagrama de dispersión. La línea superior se ajusta a los datos para las mujeres y la inferior a los datos para los hombres.

TABLA 7.7 DATOS PARA EL EJEMPLO 7.5: VARIABLE FICTICIA DE ENSAMBLADORES ELECTRÓNICOS

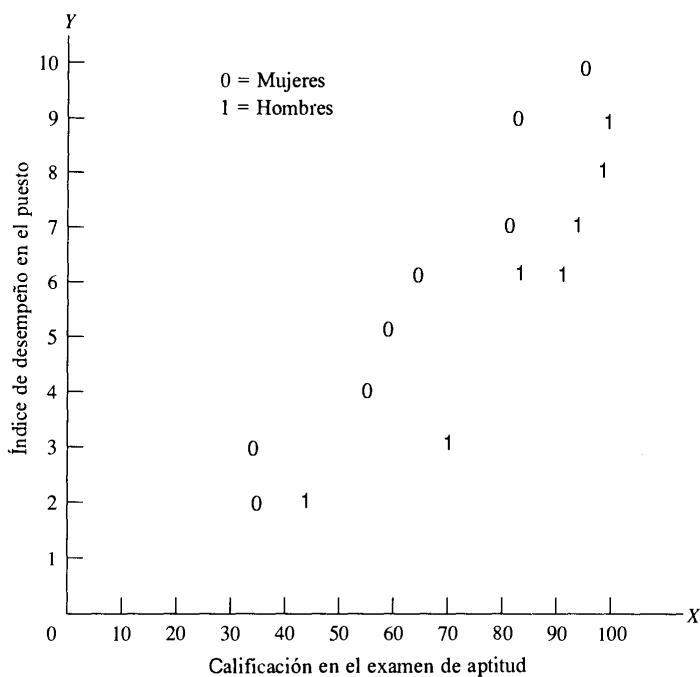
SUJETO	ÍNDICE DE DESEMPEÑO EN EL PUESTO, Y	CALIFICACIÓN EN EL EXAMEN DE APTITUD, X_2	SEXO X_3
1	5	60	0 (F)
2	4	55	0 (F)
3	3	35	0 (F)
4	10	96	0 (F)
5	2	35	0 (F)
6	7	81	0 (F)
7	6	65	0 (F)
8	9	85	0 (F)
9	9	99	1 (M)
10	2	43	1 (M)
11	8	98	1 (M)
12	6	91	1 (M)
13	7	95	1 (M)
14	3	70	1 (M)
15	6	85	1 (M)
	$\overline{87}$	$\overline{1,093}$	

$\bar{Y}_F$ = media del índice de desempeño en el puesto (mujeres) = 5.75

$\bar{Y}_M$ = media del índice de desempeño en el puesto (hombres) = 5.86

$\bar{X}_F$ = media de la calificación en el examen de aptitud (mujeres) = 64

$\bar{X}_M$ = media de la calificación en el examen de aptitud (hombres) = 83

**Figura 7.2** Diagrama de dispersión para los datos del ejemplo 7.5.

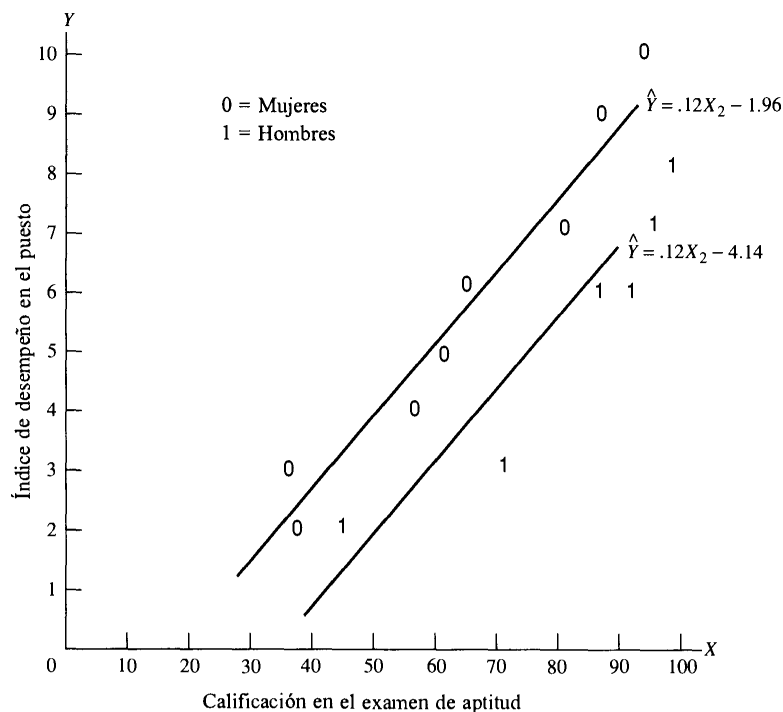


Figura 7.3 Líneas de regresión para las variables ficticias del ejemplo 7.5.

Cada una de estas líneas se obtiene mediante la solución de la ecuación de regresión múltiple 7.1.

$$\hat{Y} = b_0 + b_2X_2 + b_3X_3$$

en donde

$X_2 =$ calificación del examen

$X_3 = \begin{cases} 0 & \text{para mujeres} \\ 1 & \text{para hombres} \end{cases}$ variable ficticia

La ecuación simple es equivalente a las dos siguientes ecuaciones:

$$\hat{Y} = b_0 + b_2X_2 \quad \text{para mujeres}$$

$$\hat{Y} = b_0 + b_2X_2 + b_3 \quad \text{para hombres}$$

Nótese que b_3 representa el efecto de un hombre sobre el desempeño en el puesto y que b_2 representa el efecto de las diferencias del examen de aptitud (se supone que el valor de b_2 es el mismo para hombres y mujeres). El punto importante consiste en que una ecuación de regresión múltiple arrojará las dos líneas estimadas que se muestran en la fig. 7.3. La línea superior es la estimación para mujeres y la inferior para hombres. Se podría

visualizar X_3 como una variable “de interruptor”, que está “encendida” cuando se hace una observación para un hombre y “apagada” cuando se hace para una mujer.

Los comandos de Minitab para ejecutar los datos del problema 7.5 son:

```
MTB > SET C1
DATA > 5 4 3 10 2 7 6 9 9 2 8 6 7 3 6
MTB > END
MTB > SET C2
DATA > 60 55 35 96 35 81 65 85 99 43 98 91 95 70 85
MTB > END
MTB > SET C3
DATA > 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
MTB > END
MTB > NAME C1 'JOB' C2 'TEST' C3 'GENDER'
MTB > CORR C1-C3
```

	JOB	TEST
TEST	0.876	
GENDER	0.021	0.428

```
MTB > REGRESS C1 ON 2 PREDICTORS C2 C3

The regression equation is
JOB = -1.96 + 0.120 TEST -2.18 GENDER
```

Predictor	Coef	Stdev	t-ratio	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
Error	12	7.419	0.618		
Total	14	94.400			

SOURCE	DF	SEQ SS
TEST	1	72.486
GENDER	1	14.496

```
MTB > STOP
```

- Los comandos SET se usan para capturar los datos de desempeño en el puesto en C1, los datos de calificación del examen en C2 y la variable ficticia sexo en C3.
- El comando NAME se emplea para dar nombre a las variables en cada columna.
- El comando CORR se utiliza para calcular la matriz de correlación para las variables C1, C2 y C3.

- El comando REGRESS se usa para desarrollar un análisis de regresión, siendo el desempeño en el trabajo la variable dependiente (Y) la calificación en el examen (X_2) y el sexo (X_3) como las variables independientes.

Ejemplo 7.6

La ecuación de regresión múltiple estimada para los datos del ejemplo 7.5 se muestra en la salida de cómputo de Minitab. La ecuación es

$$\hat{Y} = -1.96 + .12X_2 - 2.18X_3$$

Para los dos valores (0 y 1) de X_3 , la ecuación resulta

$$\hat{Y} = -1.96 + .12X_2 - 2.18(0) = -1.96 + .12X_2 \quad \text{para mujeres}$$

y

$$\hat{Y} = -1.96 + .12X_2 - 2.18(1) = -4.14 + .12X_2 \quad \text{para hombres}$$

Estas ecuaciones se pueden interpretar de la siguiente manera. El valor del coeficiente de regresión $b_2 = .12$, que es la pendiente de cada una de las líneas, es el incremento promedio estimado en el índice de desempeño por cada unidad de incremento en la calificación del examen de aptitud. Este coeficiente se aplica tanto a hombres como a mujeres.

El otro coeficiente de regresión $b_3 = -2.18$ se aplica a los hombres. Si un hombre realizó el examen, el índice de desempeño en el puesto se reduce en 2.18 unidades, cuando se mantiene constante la calificación de aptitud.

Una revisión de las medias de las variables Y y X_2 , clasificadas por sexo, ayuda a comprender este resultado. La tabla 7.7 muestra que las medias del índice de desempeño para hombres (5.86) y para mujeres (5.75) fueron aproximadamente iguales. Sin embargo, en el examen de aptitud, los hombres califican considerablemente más alto (83) que las mujeres (64). De ahí que si dos aspirantes, un hombre y una mujer, hacen el examen de aptitud y ambos califican 70, el índice de desempeño en el puesto para la mujer será 2.18 puntos más alto que el del hombre, como se muestra a continuación:

$$\text{Mujer: } \hat{Y} = -1.96 + .12X_2 = -1.96 + .12(70) = 6.44$$

$$\text{Hombre: } \hat{Y} = -4.14 + .12X_2 = -4.14 + .12(70) = 4.26$$

Al observar la matriz de correlación en la salida de Minitab, se pueden apreciar algunos puntos interesantes. Existe una alta relación lineal entre el desempeño en el puesto y el examen de aptitud, $r_{12} = .88$. Mediante el conocimiento de esta relación, se puede explicar el 77% de la varianza de la variable índice de desempeño en el puesto ($.88^2 = .77$).

El coeficiente de correlación $r_{13} = .02$ indica que virtualmente no hay relación entre el sexo y el desempeño en el puesto. Esta conclusión resulta evidente también cuando se considera el hecho de que la media de los índices de desempeño para hombres y mujeres es casi igual (5.86 *versus* 5.75). A primera vista, se podría concluir que no es una información útil el conocer si el aspirante es hombre o mujer. No obstante, la relación moderada $r_{23} = .43$, entre el sexo y la calificación en el examen de aptitud, indica que el examen podría discriminar entre los sexos. Los hombres parecen calificar mejor que las mujeres (82 *versus* 64). Quizá exista un elemento de fuerza en el examen, que no existe en el puesto.

Cuando se emplean tanto el sexo como el resultado del examen para pronosticar el desempeño en el trabajo, se explica el 92% de la varianza. Este resultado sugiere que ambas

variables hacen una valiosa contribución en la predicción del desempeño. El examen de aptitud explica el 77% de la varianza y el sexo en conjunción con los resultados del examen agregan otro 15%. Si se prueban los valores calculados de t , 11.86 y -4.84 , se incluirán ambas variables en la ecuación final.

VALIDACIÓN DEL MODELO

Hasta ahora, no se ha verificado la validez de las suposiciones inherentes al modelo de regresión. Ahora es el momento de abordar aquellas situaciones en las que se violan las suposiciones e investigar las técnicas correctivas adecuadas.

Las suposiciones se enlistan en el recuadro de la sección “Inferencia estadística en la regresión múltiple”, presentado anteriormente en este capítulo. La primera suposición de normalidad se requiere para efectuar pruebas t y F , y se abordó en el cap. 6. En realidad, cuando el tamaño de la muestra es lo suficientemente grande (30 o más elementos), el teorema del límite central proporciona un razonamiento para usar estas pruebas estadísticas sin la suposición de normalidad.

La segunda suposición se refiere a la dispersión de los puntos de datos de la población alrededor del plano de la población. El analista supone que la dispersión de puntos de datos permanece constante en cualquier sitio a través del plano. Si en una situación dada no hay una varianza constante, existe *heteroscedasticidad*. Este concepto se expondrá en la siguiente sección.

La tercera suposición implica una muestra aleatoria de puntos de datos X - Y . Pero, por ejemplo, los datos económicos ordenados en una serie histórica no se pueden siempre considerar como aleatorios. Es decir, una observación de precio en un periodo dado, está por lo regular correlacionada con el valor del periodo anterior. Para describir esta situación se usa el término *correlación en serie*, la cual se abordará con amplitud en el cap. 9.

La cuarta suposición indica una relación lineal, ésta es una importante suposición de que el modelo está correctamente especificado. La graficación de los puntos de datos en un diagrama de dispersión es un primer paso invaluable para evitar una mala especificación. Con frecuencia aparece una curvatura pronunciada en la gráfica de los puntos de datos X - Y . En ocasiones se emplea la transformación de datos sobre estos datos curvilíneos de modo que los datos transformados resulten lineales. Esta técnica se demostró en el cap. 6.

Heteroscedasticidad

La heteroscedasticidad se presenta debido a cambios en las condiciones antecedentes no reconocidos en la especificación del modelo. Los cambios que alteran la estructura inherente a la economía y los cambios en leyes o políticas de gobierno ocasionan con frecuencia heteroscedasticidad. Un motivo común de esta condición es el cambio en la precisión de las mediciones con respecto a series de tiempo de datos.

Existe *heteroscedasticidad* cuando los errores o residuos no tienen una varianza constante a través de un nivel completo de valores.

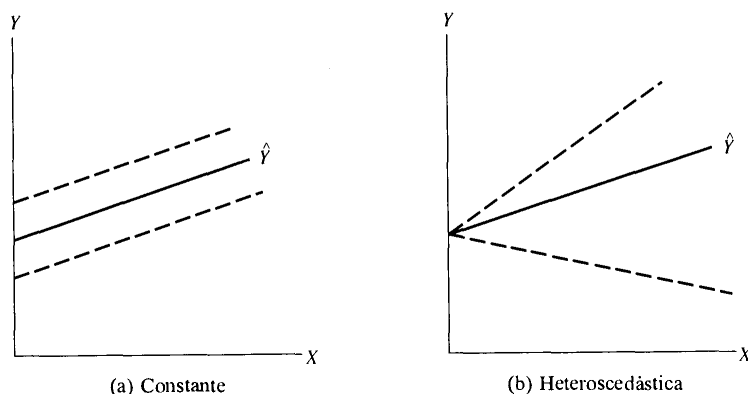


Figura 7.4 Varianza constante y heteroscedástica.

En la fig. 7.4 se ilustra el efecto de la heteroscedasticidad. La fig. 7.4(a) muestra una varianza constante. Dada X , las observaciones de Y se ubican dentro de una banda constante alrededor de la línea de regresión. De ahí que la varianza no depende del valor especificado para X . En la fig. 7.4(b), la varianza se incrementa al aumentar X . Si se emplea la ecuación de regresión para pronosticar Y , será menos probable que los intervalos de confianza contengan el valor verdadero de la variable Y , para valores grandes de la variable X . Esta condición se aborda con detalle en el cap. 9.

Colinealidad

Siempre que aparece más de una variable independiente en una ecuación de regresión, es posible que estas variables estén relacionadas entre sí. A esta situación se le conoce como *colinealidad*.

La *colinealidad* es la situación en la que las variables independientes de una ecuación de regresión múltiple están altamente intercorrelacionadas.

En un análisis, la colinealidad causa problemas en los siguientes aspectos:

1. Un coeficiente de regresión que tiene signo positivo en una ecuación de regresión de dos variables, pudiera cambiar a signo negativo en una ecuación de regresión múltiple que contenga otras variables con las que está altamente interrelacionado. (El cambio puede ser también a la inversa, de negativo a positivo.)
2. La estimación de los coeficientes de regresión fluctúa marcadamente entre una muestra y otra.

3. Con frecuencia se usa la regresión múltiple como una herramienta interpretativa para evaluar la importancia relativa de varias variables independientes. Cuando las variables de predicción están intercorrelacionadas, éstas explican la misma varianza en la estimación de la variable dependiente. Por esta razón, resulta en extremo difícil separar las influencias individuales de cada una de las variables independientes.

Estos problemas se exploran en el ejemplo siguiente:

Ejemplo 7.7

Se desarrolla una ecuación de regresión múltiple a partir de los datos que se presentan en la tabla 7.8. Se pronostica el gasto en alimentos a partir del conocimiento de su relación con las variables de ingreso y tamaño de la familia. Una rápida revisión de la matriz de correlación para estas variables, dada en la tabla 7.9, conduce a las siguientes conclusiones:

1. Tanto el ingreso como el tamaño de la familia están relacionados en forma positiva con el gasto en alimentos y tienen potencial como buenas variables de predicción ($r_{12} = .88$ y $r_{13} = .74$).
2. El ingreso y el tamaño de la familia están altamente interrelacionados ($r_{23} = .87$) y probablemente estén explicando la misma porción de la varianza del gasto en alimentos.

La ecuación de regresión múltiple para este ejemplo es

$$\hat{Y} = 3.52 + 2.28X_2 - .41X_3$$

Nótese que el coeficiente de regresión negativo ($-.41$) para la variable tamaño de la familia (X_3). Este coeficiente sugiere que un incremento de una persona en una familia, manteniendo constante el ingreso anual, *disminuye* en \$41 el gasto anual estimado en alimentos, sin importar si el ingreso familiar es de \$10 000 o \$20 000. Este resultado no es lógico, en especial si se observa que en la matriz de correlación el tamaño de la familia y el consumo de alimentos están relacionados *positivamente* ($r_{13} = .74$).

TABLA 7.8 DATOS PARA EL EJEMPLO 7.7

FAMILIA	GASTO ANUAL EN ALIMENTOS (CIENTOS DE DÓLARES)	INGRESO ANUAL (MILES DE DÓLARES)	TAMAÑO DE LA FAMILIA,
	Y	X_2	X_3
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

TABLA 7.9 SALIDA DE CÓMPUTO PARA EL EJEMPLO 7.7

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD ERROR OF REG. COEF.	COMPUTED T VALUE
2	6.9	2.38	.884	2.28	0.81261	2.80
3	3.0	1.56	.737	-0.41	1.23603	-0.33
DEPENDENT						
1	18.0	5.5				
INTERCEPT		3.52		MULTIPLE CORRELATION		.886
STD. ERROR OF ESTIMATE		2.89		R SQUARED		0.785
				CORRECTED R SQUARE		0.723
CORRELATION MATRIX						
	EXPENDITURES			INCOME	SIZE	
EXPENDITURES		1.00		.88	.74	
INCOME				1.00	.87	
SIZE					1.00	

Nótese también que cuando se usa sólo el ingreso para estimar el gasto en alimentos, se explica el 78.1% (.884²) de la varianza. Al agregar el tamaño de la familia, R^2 sólo aumenta a 78.5%. Resulta evidente que el tamaño de la familia está explicando la misma varianza que el ingreso.

Cuando existe una colinealidad extrema, no hay forma aceptable de realizar un análisis de regresión múltiple mediante el conjunto dado de variables independientes. Las dos soluciones que se sugieren son: (1) usar en la ecuación final sólo una de las variables altamente correlacionadas, o (2) crear y utilizar una nueva variable que sea una combinación de las dos variables altamente interrelacionadas.

SELECCIÓN DE LA MEJOR ECUACIÓN DE REGRESIÓN

¿Cómo se hace para desarrollar la mejor ecuación de regresión múltiple, para pronosticar una variable de interés? El *primer paso* comprende la selección de un conjunto completo de variables potenciales de predicción. Se incluye cualquier variable que pueda agregar precisión al pronóstico. En la selección de la ecuación final, por lo regular se enfrenta el dilema de proporcionar el pronóstico más preciso al menor costo. En otras palabras, al elegir variables de predicción a incluir en la ecuación final, el analista debe evaluarlas mediante dos criterios *opuestos*:

1. El analista desea que la ecuación incluya tantas variables de predicción como sea posible. Siempre que se incorpora una variable de predicción a una ecuación de regresión múltiple, R^2 permanece sin cambio o se incrementa. Mientras la muestra sea lo suficientemente grande (se recomienda un tamaño de muestra n de 10 para cada variable

empleada en la ecuación), cada nueva variable de predicción tiene el potencial de mejorar el pronóstico.

2. Como obtener y monitorear información sobre un gran número de X cuesta dinero, la ecuación debe contener el menor número posible de pronosticadores. Por lo regular, la ecuación más sencilla es la mejor.

La selección de la mejor ecuación de regresión implica por lo regular un acuerdo entre estos extremos y el juicio personal será una parte necesaria de cualquier solución.

Después de reunir una lista completa de pronosticadores potenciales, el *segundo paso* consiste en eliminar las variables independientes que no parezcan ser apropiadas. Una variable independiente: (1) pudiera no ser fundamental para el problema (debe haber cierta causalidad plausible entre las variables dependiente e independiente), (2) pudiera estar sujeta a errores de medición considerables, (3) pudiera duplicar otras variables independientes (colinealidad) o, (4) pudiera ser difícil medirla con precisión (no hay datos disponibles o son costosos).

El *tercer paso* consiste en recortar la lista de pronosticadores con el propósito de obtener la “mejor” selección de variables independientes. En la siguiente sección se abordarán las técnicas actualmente en uso. No se puede decir que alguno de los procedimientos de búsqueda dé por resultado el “mejor” conjunto de variables independientes. De hecho, es frecuente que no haya un único “mejor” conjunto. Para aumentar la confusión, las diversas técnicas no necesariamente conducen a la misma ecuación final de pronóstico. Todo el proceso de selección de variables es en extremo subjetivo. Por esta razón se debe analizar como primer paso la matriz de correlación, antes de utilizar cualquier procedimiento automático. La principal ventaja de los procedimientos automáticos de búsqueda es que el analista puede enfocar sus juicios sobre las áreas centrales del problema.

Para demostrar los diversos procedimientos de búsqueda, se presenta un ejemplo sencillo que tiene cinco variables potenciales independientes.

Ejemplo 7.8

Pam Weigand, gerente de personal de la Zurenko Pharmaceutical Company, está interesada en pronosticar si un aspirante en particular se convertirá en un buen vendedor. Pam decide emplear como variable dependiente (Y) las ventas del primer mes y elige las siguientes variables independientes para analizar:

X_2 = examen de aptitud en ventas

X_3 = edad, en años

X_4 = calificación en la prueba de ansiedad

X_5 = experiencia, en años

X_6 = PPG (punto promedio de grado) en secundaria

El gerente de personal reúne los datos que se muestran en la tabla 7.10 y asigna a su analista la tarea de obtener el “mejor” conjunto de variables independientes para pronosticar la habilidad en ventas.

TABLA 7.10 DATOS PARA EL EJEMPLO 7.8: ZURENKO PHARMACEUTICAL COMPANY

VENTAS DE UN MES (UNIDADES)	CALIFICACIÓN EN EL EXAMEN DE APTITUD	EDAD (EN AÑOS)	CALIFICACIÓN EN LA PRUEBA DE ANSIEDAD	EXPERIENCIA (EN AÑOS)	PPG EN SECUN- DARIA
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

El primer paso consiste en obtener, mediante un programa de cómputo, una matriz de correlación de todas las variables. La matriz proporcionará el conocimiento esencial con respecto a las relaciones básicas entre las variables.

La revisión de la matriz de correlación de la tabla 7.11 revela que el examen de aptitud en ventas, la edad, la experiencia y el PPG están relacionados en forma positiva con la habilidad en ventas y tienen potencial como buenas variables de pronóstico. La calificación de la prueba de ansiedad muestra una ligera correlación negativa con respecto a las ventas y es probable que no sea un pronosticador importante. Un mayor análisis indica que el PPG y la edad, así como la experiencia y la edad están interrelacionados. Esta presencia de interrelación es la que debe manejarse en el intento por encontrar el mejor conjunto posible de variables explicativas.

TABLA 7.11 MATRIZ DE CORRELACIÓN PARA EL EJEMPLO 7.8

	(1) VENTAS	(2) EXAMEN	(3) EDAD	(4) ANSIEDAD	(5) EXPERIENCIA	(6) PPG
(1) Ventas	1.000	.676	.798	-.296	.550	.622
(2) Examen		1.000	.228	-.222	.350	.318
(3) Edad			1.000	-.287	.540	.695
(4) Ansiedad				1.000	-.279	-.244
(5) Experiencia					1.000	.312
(6) PPG						1.000

A continuación se demuestran dos procedimientos: el de todas las regresiones posibles y el de regresión por pasos.

Todas las regresiones posibles

El procedimiento llama a investigar todas las ecuaciones de regresión posibles que comprendan las variables independientes potenciales. El analista comienza con una ecuación que no contiene variables independientes y analiza cada combinación posible con el propósito de encontrar el mejor conjunto de pronosticadores.

Con el enfoque de todas las regresiones posibles se pueden emplear diferentes criterios para comparar las diversas ecuaciones de regresión. Sólo expondremos aquí la técnica de R^2 , que comprende cuatro pasos.

El procedimiento requiere primero el ajuste de cualquier modelo posible de regresión que comprenda la variable dependiente y cualquier número de variables independientes. Cada variable independiente puede, ya sea estar o no en la ecuación (dos posibles resultados) y éste es un hecho real para cada variable independiente. De modo que en conjunto, existen 2^i ecuaciones (donde i equivale al número de variables independientes). Así que si hay ocho variables independientes por considerar ($i = 8$), entonces se deben examinar $2^i = 256$ ecuaciones.

El segundo paso del procedimiento consiste en dividir las ecuaciones en grupos, de acuerdo con el número de parámetros por estimar.

Ejemplo 7.9

En la tabla 7.12 se presentan los resultados de la ejecución de todas las regresiones posibles para el ejemplo de la Zurenko Pharmaceutical Company. Nótese que la tabla 7.12 está dividida en seis grupos de ecuaciones de regresión resultantes. Esta división coincide con el número de parámetros que contiene cada ecuación.

El tercer paso es el que está relacionado con la selección de la mejor variable (o variables) independiente(s) para cada grupo de parámetros. La ecuación que tenga la R^2 más alta es la mejor. En la tabla 7.13 se presenta la mejor ecuación de cada uno de los grupos mostrados en la tabla 7.12.

El cuarto paso comprende la realización de la decisión subjetiva “¿Cuál ecuación es mejor?” Por una parte el analista desea la R^2 más alta posible y por otro, desea la ecuación

TABLA 7.12 VALORES DE R^2 PARA TODAS LAS REGRESIONES POSIBLES DE ZURENKO PHARMACEUTICAL

VARIABLES INDEPENDIENTES UTILIZADAS	NÚMERO DE PARÁMETROS	GRADOS DE LIBERTAD	R^2
Ninguna	1	29	0
X_2	2	28	.457
X_3	2	28	.637
X_4	2	28	.088
X_5	2	28	.302
X_6	2	28	.387
X_2, X_3	3	27	.8948
X_2, X_4	3	27	.479
X_2, X_5	3	27	.569
X_2, X_6	3	27	.641
X_3, X_4	3	27	.642
X_3, X_5	3	27	.657
X_3, X_6	3	27	.646
X_4, X_5	3	27	.324
X_4, X_6	3	27	.409
X_5, X_6	3	27	.527
X_2, X_3, X_4	4	26	.8951
X_2, X_3, X_5	4	26	.8948
X_2, X_3, X_6	4	26	.8953
X_2, X_4, X_5	4	26	.575
X_2, X_4, X_6	4	26	.646
X_2, X_5, X_6	4	26	.701
X_3, X_4, X_5	4	26	.659
X_3, X_4, X_6	4	26	.650
X_3, X_5, X_6	4	26	.669
X_4, X_5, X_6	4	26	.531
X_2, X_3, X_4, X_5	5	25	.8951
X_2, X_3, X_4, X_6	5	25	.8955
X_2, X_3, X_5, X_6	5	25	.8953
X_2, X_4, X_5, X_6	5	25	.701
X_3, X_4, X_5, X_6	5	25	.671
X_2, X_3, X_4, X_5, X_6	6	24	.8955

TABLA 7.13 LAS MEJORES ECUACIONES DE REGRESIÓN PARA ZURENKO PHARMACEUTICAL

NÚMERO DE PARÁMETROS	VARIABLES INDEPENDIENTES	GRADOS DE LIBERTAD	R^2	F
1	Ninguna	29	0	49.09
2	X_3	28	.637	66.220
3	X_2, X_3	27	.8948	.124
4	X_2, X_3, X_6	26	.8953	.054
5	X_2, X_3, X_4, X_6	25	.8955	.000
6	X_2, X_3, X_4, X_5, X_6	24	.8955	

más sencilla posible. Se debe también destacar que un tamaño de muestra de $n = 10$ es deseable más no esencial. El enfoque de todas las regresiones posibles supone que n excede al número de parámetros.

**Ejemplo
7.10**

El analista intenta encontrar el punto en el que no vale la pena incorporar variables independientes adicionales en el problema de la Zurenko Pharmaceutical, debido a que conduce a un muy pequeño incremento de R^2 . Los resultados de la tabla 7.13 indican claramente que no es necesario agregar variables después de la del examen (X_2) y la de edad (X_3). Por lo tanto, la ecuación final es

$$\hat{Y} = b_0 + b_2X_2 + b_3X_3$$

y explica el 89.48% de la varianza de Y .

Draper y Smith³ resumen mejor el procedimiento de todas las regresiones posibles:

En general, no se justifica el análisis de todas las regresiones posibles. Aunque significa que el analista observó todas las posibilidades, también significa que examinó un gran número de ecuaciones de regresión, las cuales un pensamiento inteligente hubiera descartado. El volumen de tiempo de cómputo utilizado es un desperdicio y el esfuerzo físico de examinar todas las salidas de cómputo es enorme cuando se examinan más de unas cuantas variables. Es preferible algún tipo de procedimiento que haga esta tarea más corta.

Regresión por pasos

El procedimiento de regresión por pasos incorpora al modelo una variable independiente a la vez, paso a paso. Mediante este procedimiento se puede manejar en la computadora un gran número de variables en una ejecución.

En esencia, este enfoque calcula una secuencia de ecuaciones de regresión, incorporando o eliminando en cada paso una variable independiente. En otras palabras, el programa de cómputo registra las variables en pasos individuales de la mejor a la peor, siempre que cumplan con el criterio estadístico establecido.

Se registra primero la variable independiente que explica la mayor cantidad de varianza en la variable dependiente. La siguiente variable por registrar, explica la mayor cantidad de varianza en conjunción con la primera y así sucesivamente. En cada paso, se registra en la ecuación la variable que explica la mayor cantidad de varianza no explicada por la variable que ya está incluida en el modelo.

Se debe señalar que algunos programas de regresión por pasos permiten que una variable independiente se registre en la ecuación en las primeras etapas, para ser eliminada posteriormente debido a su relación con otras variables incorporadas al modelo en etapas subsecuentes. Como una verificación en este paso, se evalúa la estadística parcial F para cada variable en la ecuación de regresión en cualquier etapa y se compara con un punto crítico predeterminado, elegido de la distribución F correspondiente. Esta comparación proporciona una verificación de la contribución que hace cada variable, como si fuera la

³ N. R. Draper y H. Smith, *Applied Regression Analysis* (Nueva York: Wiley, 1966), p. 167.

variable más reciente registrada, sin importar cuándo se registró realmente en la ecuación. Toda variable que no contribuye se elimina del modelo.

El proceso por pasos funciona de la siguiente manera. El analista proporciona al paquete de cómputo de regresión por pasos la variable dependiente (Y) y diversas variables independientes (X) potencialmente importantes.

Paso 1. La computadora ejecuta todos los modelos posibles de regresión simple de la forma:

$$Y = \beta_0 + \beta_2 X_2 + \epsilon$$

La variable independiente que produce el mayor valor (absoluto) de t se emplea como el mejor pronosticador de Y y se identifica como X_2 . Nótese que en el paso 1 la computadora siempre elegirá la variable independiente que tenga la correlación (absoluta) más alta con la variable dependiente.

Paso 2. La computadora ajusta todos los modelos posibles de dos variables que contengan X_2 . Se ejecuta cada una de las otras variables independientes en combinación con X_2 , para llegar a la forma:

$$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Se retiene la nueva variable independiente que produce el mayor valor (absoluto) de t y se utiliza para predecir Y en conjunción con X_2 . Este nuevo pronóstico se identifica como X_3 . Nótese que aunque X_3 estará correlacionada con Y , no estará altamente correlacionada con X_2 .

Debe notarse que es probable que X_2 y X_3 estén correlacionadas en cierto grado y que la inclusión de X_3 modifica el valor de β_2 y su estadística t . Algunos de los mejores programas de cómputo verifican de nuevo β_2 para asegurar que sigue siendo lo suficientemente útil como para permanecer en el modelo. Si no es así, la computadora busca la mejor variable de remplazo.

Paso 3. La computadora ajusta todos los modelos posibles de tres variables que contengan X_2 y X_3 . Se ejecuta cada una de las variables independientes en combinación con X_2 y X_3 para llegar a la forma:

$$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

Se retiene la variable independiente que pronostica el mayor valor (absoluto) de t y se emplea para pronosticar Y en conjunción con X_2 y X_3 . Este nuevo pronosticador se identifica como X_4 . Nótese que, aunque X_4 está correlacionado con Y , no estará altamente correlacionado con X_2 o X_3 .

El proceso por pasos continúa hasta llegar al paso donde ninguna de las variables restantes tiene un coeficiente β que sea diferente de cero en forma significativa en algún nivel especificado de alfa. En este punto, la computadora no puede incorporar una nueva variable que contribuya al pronóstico de Y . Por lo tanto, el resultado del procedimiento por

pasos es un modelo que sólo contiene variables independientes con valores que son significativos en el nivel de significación especificado.

El procedimiento por pasos se ilustra en el ejemplo 7.11.

Ejemplo 7.11

Ahora, paso a paso, el problema de Zurenko está resuelto, tal y como funcionaría un programa de regresión por pasos.

Pam examina la matriz de correlación que se muestra en la tabla 7.11 y decide que al ejecutar el análisis de regresión por pasos, la variable edad se registrará primero en el modelo, ya que tiene la mayor correlación con ventas ($r_{1,3} = .798$) y explicará el 63.68% ($.798^2$) de la varianza de las ventas.

Ella advierte que la calificación en el examen de aptitud se registrará en segundo lugar en el modelo, ya que está fuertemente relacionada con las ventas ($r_{1,2} = .676$), pero no altamente relacionada con la variable edad ($r_{2,3} = .228$) contemplada en el modelo.

Pam también nota que las otras variables no califican como buenas variables de pronóstico. La calificación en la prueba de ansiedad no será un buen pronosticador debido a que no está bien relacionada con las ventas ($r_{1,4} = -.296$). Las variables de experiencia y PPG tienen potencial como buenas variables de pronóstico ($r_{1,5} = .550$) y ($r_{1,6} = .622$). Sin embargo, ambas tienen un problema de colinealidad múltiple con la variable edad ($r_{3,5} = .540$) y ($r_{3,6} = .695$).

Los comandos de Minitab para ejecutar un análisis de regresión por pasos para el ejemplo 7.11 se demuestran en la sección “Paquete de cómputo Minitab” al final del capítulo. La salida para esta ejecución de regresión por pasos se muestra en la tabla 7.14.

TABLA 7.14 SALIDA DE MINITAB DE LA REGRESIÓN POR PASOS PARA EL EJEMPLO 7.11 DE ZURENKO PHARMACEUTICAL

STEPWISE REGRESSION OF SALES ON 5 PREDICTORS, WITH N = 30

STEP	1	2
CONSTANT	-100.85	-86.79
AGE	6.97	5.93
T-RATIO	7.01	10.60
APTITUDE		0.200
T-RATIO		8.13
S	6.85	3.75
R-SQ	63.70	89.48

Aunque la salida del proceso por pasos de Minitab contiene estadísticas t , en realidad Minitab utiliza la estadística F para decidir si se debe incorporar una variable al modelo. En este punto debe mencionarse que existe una relación matemática entre las estadísticas t y F ($t^2 = F$). En cada paso, el programa Minitab calcula una estadística F para cada una de las variables que no está en el modelo. La variable con la mayor estadística F se incorpora al modelo, siempre que dicha estadística sea mayor que el valor límite de corte especificado; en este ejemplo, $F_{ENTER} = 4.2$ (se usa $F = 4$ a menos que se especifique otro valor). El

procedimiento por pasos de Minitab realiza una verificación para asegurar que aún sean válidas las variables incorporadas al modelo en pasos anteriores.

En el análisis por pasos se procede de la siguiente manera:

Paso 1. El modelo después del paso uno es:

$$\text{Ventas} = -100.85 + 6.97 (\text{edad})$$

Este modelo explica el 68.70% de la varianza de las ventas y tiene la mayor estadística F , $F = 49.14$. Como 49.14 es mayor que 4.2, la edad se incorpora al modelo.

Si el análisis se realiza utilizando la estadística t de la salida, las hipótesis nula y alternativa para determinar si el coeficiente de regresión de la edad es diferente de cero en forma significativa, son

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

Si la prueba se realiza a un nivel de significación de .05, la estadística t basada en 28 grados de libertad ($n - k = 30 - 2$) es 2.048. La regla de decisión es

Si la estadística t calculada es mayor que 2.048 o menor que -2.048, rechazar la hipótesis nula (rechazar H_0 si $t > 2.048$ o $t < -2.048$).

Como la estadística t calculada que se muestra en la salida de Minitab, 7.01 ($t^2 = F = 49.13$), es mayor que el valor crítico ($7.01 > 2.048$), se rechaza la hipótesis nula. El coeficiente de regresión de la variable edad es diferente de cero en forma significativa. El resultado significa que la edad es una buena variable y el procedimiento continúa con el paso 2.

Paso 2. El modelo después del paso dos es:

$$\text{Ventas} = -86.79 + 5.93 (\text{edad}) + 0.200 (\text{aptitud})$$

Este modelo explica el 89.48% de la varianza de las ventas. Las hipótesis nula y alternativa para determinar si el coeficiente de regresión de la calificación en el examen de aptitud es diferente de cero en forma significativa, son:

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0$$

Si la prueba se realiza en un nivel de significación de .05, la estadística crítica de t con base en 27 grados de libertad ($n - k = 30 - 3$) es 2.052. La regla de decisión es

Si la estadística t calculada es mayor que 2.052 o menor que -2.052, rechazar la hipótesis nula (rechazar H_0 si $t > 2.052$ o $t < -2.052$).

Ya que el valor calculado de t que se encuentra en la salida de Minitab es 8.13, se rechaza la hipótesis nula. El coeficiente de regresión de la calificación en el examen de aptitud es diferente de cero en forma significativa. El resultado significa que la calificación en el examen de aptitud es una buena variable cuando se utiliza en conjunción con la edad. Nótese que el índice t para el coeficiente de regresión de la variable edad, 10.6, es aún significativo. El procedimiento continúa con el paso 3.

Paso 3. La computadora ajusta todos los modelos posibles de tres variables que contengan X_2 y X_3 . Ninguna de las variables restantes es significativa al ejecutarlas en combinación con X_2 y X_3 , así que concluye el procedimiento por pasos.

En realidad, se desarrolla el mejor modelo de tres variables y se descarta, ya que ninguno de los índices t resulta significativo. El mejor modelo de tres variables es:

$$\text{Ventas} = -89.42 + 6.12 (\text{edad}) + 0.202 (\text{aptitud}) - .592 (\text{PPG})$$

Este modelo explica el 89.50% de la varianza de las ventas. Las hipótesis nula y alternativa para determinar si el coeficiente de regresión del PPG es diferente de cero en forma significativa, son:

$$H_0: \beta_6 = 0$$

$$H_1: \beta_6 \neq 0$$

Si se efectúa la prueba en un nivel de significación de .05, la estadística crítica de t basada en 26 grados de libertad ($n - k = 30 - 4$) es 2.056. La regla de decisión es:

Si la estadística t calculada es mayor que 2.056 o menor que -2.056, rechazar la hipótesis nula (rechazar H_0 si $t > 2.056$ o $t < -2.056$).

Si se ejecuta este paso, el valor calculado de t en la salida de Minitab es -36. La hipótesis nula no se rechaza. El coeficiente de regresión de la variable PPG no es diferente de cero en forma significativa. El resultado significa que el PPG no es una buena variable cuando se utiliza en conjunción con la edad y la calificación en el examen de aptitud.

Notas finales sobre la regresión por pasos

La técnica de regresión por pasos es en extremo fácil de usar. Por desgracia, también es en extremo fácil de usar equivocadamente. A menudo, los analistas que desarrollan un modelo de regresión generan un gran conjunto de variables independientes potenciales y dejan después que el procedimiento de regresión por pasos determine cuáles son significativas. El problema consiste en que cuando se analiza un gran conjunto de variables independientes se realizan muchas pruebas de t y esto es probable que dé por resultado un error de tipo I. Por esta razón, el modelo final pudiera contener una variable que no esté relacionada en forma lineal con la variable dependiente, pero que se incorporó al modelo por mera casualidad.

Un segundo problema comprende la selección inicial de variables independientes potenciales. Al elegir estas variables, con frecuencia se omiten términos de orden alto (curvilíneos, no lineales y de interacción), para mantener manejable el número de variables. En consecuencia, se podrían omitir inicialmente del modelo diversas variables importantes. Resulta obvio que la selección intuitiva del analista, de las variables independientes, es crítica para el desarrollo de un modelo de regresión exitoso.

USO DE LA REGRESIÓN PARA PRONOSTICAR DATOS ESTACIONALES

En el cap. 8 se presentará un modelo de series de tiempo *multiplicativo*, en el que la fluctuación estacional será proporcional al nivel de la tendencia, para cada observación. Esta sección presenta un modelo de series de tiempo *de agregación*, en el que una cantidad constante se agrega a la estimación de la tendencia de la serie histórica, correspondiente al incremento esperado en el valor de la variable dependiente debido a factores estacionales. En el modelo multiplicativo, la estimación de la tendencia se multiplica por un porcentaje fijo. En el modelo de agregación, se *agrega* una cantidad constante a la estimación de la tendencia. La ecuación 7.6 se emplea para hacer una regresión de datos trimestrales mediante este método.

$$\hat{Y}_t = b_0 + b_2S_2 + b_3S_3 + b_4S_4 \quad (7.6)$$

en donde

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

$S_2 = 1$ si t es el primer trimestre del año; 0 en caso contrario

$S_3 = 1$ si t es el segundo trimestre del año; 0 en caso contrario

$S_4 = 1$ si t es el tercer trimestre del año; 0 en caso contrario

b_0 = constante

b_2, b_3, b_4 = coeficientes de regresión

Las variables S_2, S_3 y S_4 son variables ficticias que representan el primero, segundo y tercer trimestres, respectivamente. Nótese que los cuatro niveles de la variable cualitativa se describen sólo con tres variables ficticias. Esto se debe a que la media del cuarto trimestre se contabilizará mediante la intersección b_0 . Si S_2, S_3 y S_4 son todos igual a cero, el cuarto trimestre se representa mediante b_0 .

Ejemplo 7.12 James Brown, pronosticador de la Washington Water Power Company, intenta pronosticar el consumo eléctrico de clientes residenciales para 1994. Como sabe que los datos son estacionales, decide emplear la ecuación 7.6 para desarrollar su modelo de regresión estacional. James reúne datos trimestrales de 1980 a 1993. En la tabla 7.15 se muestran los datos de consumo eléctrico, medidos en millones de kilowatt-horas.

TABLA 7.15 DATOS DE CONSUMO ELÉCTRICO DE LA WASHINGTON WATER POWER COMPANY, 1980–1993, PARA EL EJEMPLO 7.12

AÑO	TRIMESTRE	KILOWATTS (EN MILLONES)	AÑO	TRIMESTRE	KILOWATTS (EN MILLONES)
1980	1	1071	1987	1	933
	2	648		2	582
	3	480		3	490
	4	746		4	708
1981	1	965	1988	1	953
	2	661		2	604
	3	501		3	508
	4	768		4	708
1982	1	1065	1989	1	1036
	2	667		2	612
	3	486		3	503
	4	780		4	710
1983	1	926	1990	1	952
	2	618		2	628
	3	483		3	534
	4	757		4	733
1984	1	1047	1991	1	1085
	2	667		2	692
	3	495		3	568
	4	794		4	783
1985	1	1068	1992	1	928
	2	625		2	655
	3	499		3	590
	4	850		4	814
1986	1	975	1993	1	1018
	2	623		2	670
	3	496		3	566
	4	728			

Fuente: Washington Water Power Annual Report, varios años

James crea las variables ficticias S_2 , S_3 y S_4 , que representan el primero, segundo y tercer trimestres, respectivamente. Los datos para los cuatro trimestres de 1980 son

Y_t	S_2	S_3	S_4
1071	1	0	0
648	0	1	0
480	0	0	1
746	0	0	0

James registra todos los datos de la tabla 7.15 en un archivo y utiliza Minitab para realizar el análisis.

Los comandos de Minitab para ejecutar el análisis estacional son

```
MTB > READ 'USAGE.DAT' C1-C4
      55 ROWS READ
```

ROW	C1	C2	C3	C4
1	1071	1	0	0
2	648	0	1	0
3	480	0	0	1
4	746	0	0	0
.	.	.	.	.

```
MTB > NAME C1 'USAGE' C2 '1ST QT' C3 '2ND QT' C4 '3RD QT'
MTB > REGRESS C1 3 PREDICTORS C2-C4;
SUBC> PREDICT 0 0 0;
SUBC> PREDICT 1 0 0;
SUBC> PREDICT 0 1 0;
SUBC> PREDICT 0 0 1;
```

The regression equation is
 USAGE = 768 + 236 1ST QT - 125 2ND QT - 249 3RD QT

Predictor	Coef	Stdev	t-ratio	P
Constant	768.46	11.96	64.25	0.000
1ST QT	235.61	16.61	14.19	0.000
2ND QT	-125.46	16.61	-7.55	0.000
3RD QT	-248.68	16.61	-14.97	0.000

s = 43.12 R-sq = 95.0% R-sq(adj) = 94.7%

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	3	1794906	598302	321.75	0.000
Error	51	94835	1860		
Total	54	1889741			

SOURCE	DF	SEQ SS
1ST QT	1	1377949
2ND QT	1	112
3RD QT	1	416845

Fit	Stdev.Fit	95% C.I.	95% P.I.
768.46	11.96	(744.45, 792.48)	(678.60, 858.32)
1004.07	11.52	(980.93, 1027.21)	(914.44, 1093.70)
643.00	11.52	(619.86, 666.14)	(553.37, 732.63)
519.79	11.52	(496.64, 542.93)	(430.16, 609.42)

```
MTB > STOP
```

- El comando READ se usa para almacenar los datos almacenados en el archivo USAGE.DAT, en las columnas C1 a C4.
- El comando NAME se emplea para dar nombre a las variables.
- El comando REGRESS se utiliza para desarrollar una ecuación de regresión, con C1 como la variable dependiente y C2, C3 y C4 como las variables independientes.
- Los subcomandos PREDICT se emplean para pronosticar el consumo del cuarto trimestre de 1993 y los tres primeros trimestres de 1994.

El modelo de regresión estacional es:

$$\hat{Y} = b_0 + b_2S_2 + b_3S_3 + b_4S_4$$

$$\text{Consumo} = 768 + 236 (\text{1er trimestre}) - 125 (\text{2o. trimestre}) - 249 (\text{tercer trimestre})$$

James advierte que este modelo explica el 95% de la varianza del consumo eléctrico para clientes residenciales. Su pronóstico para el cuarto trimestre de 1993 es:

$$\text{Cuarto trimestre: Consumo} = 768 + 236(0) - 125(0) - 249(0) = 768$$

Nótese que la constante 768 es el pronóstico para el cuarto trimestre. Este valor es también el promedio o media del consumo eléctrico del cuarto trimestre. Los pronósticos de James para cada trimestre de 1994 son:

$$\text{Primer trimestre: Consumo} = 768 + 236(1) - 125(0) - 249(0) = 1004$$

$$\text{Segundo trimestre: Consumo} = 768 + 236(0) - 125(1) - 249(0) = 643$$

$$\text{Tercer trimestre: Consumo} = 768 + 236(0) - 125(0) - 249(1) = 519$$

$$\text{Cuarto trimestre: Consumo} = 768 + 236(0) - 125(0) - 249(0) = 768$$

PRONÓSTICOS ECONOMÉTRICOS

En este capítulo y en el cap. 6, el análisis de regresión utilizó el conocimiento de una o más variables independientes para pronosticar la variable de interés. El uso de una variable independiente permite al analista pronosticar el valor de una variable dependiente, siempre que exista entre ambas una relación lineal significativa. Cuando la variable dependiente se encuentra en una serie histórica, el análisis de regresión se conoce con el nombre de *modelo econométrico*.

Para que la ecuación de regresión sea útil en el pronóstico de un valor futuro de una variable, como por ejemplo ventas al menudeo, el analista debe obtener una estimación del

valor de la variable independiente para el mismo periodo en el futuro. El analista debe tener presente al desarrollar un modelo de regresión lineal, que para pronosticar algún valor futuro de la variable dependiente, debe conocer el valor o valores de la(s) variable(s) independiente(s), si es que la ecuación de regresión va a ser útil. Por ejemplo, si se usa el PNB para pronosticar ventas al menudeo, se debe estimar el valor del PNB. Si no se puede obtener el valor del PNB para el próximo año, no se puede utilizar para pronosticar las ventas al menudeo del próximo año.

La esencia de la construcción de un modelo econométrico mediante el análisis de regresión consiste en la identificación y especificación de factores de causa, a utilizar en una ecuación de regresión. Sin embargo, recordamos que en la exposición del cap. 6 se dijo que el uso del análisis estadístico no le permite al analista afirmar causa y efecto. En la construcción de un modelo econométrico, el analista está desarrollando un modelo a partir de la teoría de que las variables independientes ejercen una influencia en el comportamiento de la variable dependiente, en formas que pueden explicarse con base en el sentido común.

Hoy en día se utilizan modelos econométricos de gran escala para modelar empresas específicas dentro de una industria, industrias selectas dentro de la economía y la economía total. Los modelos econométricos comprenden un gran número de ecuaciones de regresión lineal múltiple simultáneas. Por ello, los modelos econométricos son sistemas de ecuaciones simultáneas que comprenden gran cantidad de variables independientes.

Ejemplo 7.13

Una economista de una compañía nacional de aparatos electrodomésticos debe desarrollar una ecuación de regresión que se usará para pronosticar las ventas mensuales de hornos de microondas. La economista decide utilizar como variables independientes el ingreso personal disponible, el precio típico de un horno de microondas y el número de casas por amueblar con un retardo de tres meses. Su razonamiento al incluir la última variable es que los hornos de microondas son uno de los últimos elementos que se incorporan al hogar, así que hay un retardo de tres meses entre amueblar una casa y la compra de un horno de microondas. La ecuación de regresión que desarrolla es

$$\hat{Y}_t = b_0 + b_2X_2 + b_3X_3 + b_4X_4$$

en donde

$\hat{Y}_1$ = pronóstico de ventas de hornos de microondas para el próximo mes

X_2 = estimación del ingreso personal disponible para el próximo mes

X_3 = amueblado de casas hace tres meses

X_4 = precio de un horno de microondas típico, este mes

Si la analista está pronosticando las ventas de hornos de microondas para enero de 1995, debe obtener la siguiente información: una estimación del ingreso personal disponible para enero, dato que por lo regular proporcionan las agencias de gobierno; el número de casas por amueblar desde octubre de 1994; y el precio de venta durante el presente mes (diciembre de 1994). La analista posee el potencial para desarrollar un buen modelo. Ella sólo tendrá que estimar una de las variables independientes, el ingreso personal disponible y tendrá datos reales tanto para el precio de un horno de microondas típico durante diciembre, como del número de casas por amueblar desde octubre.

SOBREAJUSTE

El sobreajuste comprende la selección de un modelo que se apegue a las excentricidades de los datos de la muestra bajo análisis. Al aplicar dicho modelo a otros conjuntos de datos de la misma población, no se desempeña de la misma manera, ni siquiera aproximadamente.

Es más probable que ocurra un sobreajuste cuando el tamaño de la muestra es reducido, en especial si se incluyen en el modelo un gran número de variables independientes. La experiencia indica que el tamaño n de la muestra deberá ser por lo menos de 10 para cada variable independiente. (Si hay cuatro variables independientes, se requiere de un tamaño de muestra de por lo menos 40 elementos.)

Una forma de prevención para evitar el sobreajuste consiste en desarrollar el modelo a partir de una porción de los datos y luego aplicar los resultados a otro segmento. Si el modelo toma en cuenta principalmente las excentricidades de la muestra en particular a partir de la cual fue desarrollado, aparecerá este resultado al comparar los errores estimados para los dos conjuntos de datos. Si una medición comparable del error medio cuadrado $[\Sigma (Y - \hat{Y})^2]$ es sustancialmente mayor en el conjunto que utilizó el modelo para estimar nuevas observaciones, se ha presentado un sobreajuste.

APLICACIÓN EN LA ADMINISTRACIÓN

Inicialmente, el análisis de regresión múltiple fue concebido por estadísticos y economistas matemáticos (econometristas), para auxiliar en el pronóstico de la actividad económica de diversos segmentos de la economía. Este uso continúa siendo una de sus aplicaciones principales. Muchos de los reportes y pronósticos sobre el futuro de la economía que aparecen en *The Wall Street Journal*, *Fortune*, *Business Week* y en otras publicaciones similares, están basados en modelos econométricos. El gobierno de EUA utiliza ampliamente el análisis de regresión para la predicción de utilidades futuras, gastos, niveles de ingreso, tasas de interés, tasas de natalidad, desempleo y requerimientos de beneficios del Seguro Social, así como de una multitud de otros sucesos. De hecho, cada departamento principal en el gobierno de EUA hace uso de las herramientas descritas en este capítulo.

En forma similar, las entidades empresariales han adoptado y modificado según necesidades, el análisis de regresión como ayuda en el pronóstico de sucesos futuros. En el medio actual, pocas empresas pueden sobrevivir sin un pronóstico relativamente preciso de las ventas del mañana, los gastos, requerimientos de capital y flujo de efectivo. Mientras que empresas pequeñas o menos complejas pudieran arreglárselas con pronósticos intuitivos, las grandes o más complejas empresas han adoptado el análisis de regresión, que les permite estudiar las relaciones entre múltiples variables independientes y cómo afectarán éstas el futuro.

Por desgracia, la gran notoriedad que recibe el análisis de regresión por su utilidad como herramienta de predicción del futuro, tiende a opacar un hecho igualmente importante: su habilidad para ayudar a evaluar y controlar el presente. Dado que la ecuación de regresión proporciona información tanto de *fuerza* como de *dirección*, la administración puede evaluar y modificar sus estrategias actuales.

Suponga, por ejemplo, que un fabricante de mermeladas desea saber sobre quién concentrar sus esfuerzos de comercialización en la presentación de un nuevo sabor. Se puede emplear el análisis de regresión para ayudar a determinar el perfil de los principales consumidores de mermeladas. Por ejemplo, una compañía podría intentar predecir el número de sabores que un ama de casa pudiera tener en un momento dado, con base en algunas variables independientes como las siguientes:

- Número de niños en casa
- Edad de los niños
- Sexo de los niños
- Propiedad *versus* renta de la casa
- Tiempo empleado en compras
- Ingreso

Una reflexión, incluso superficial, sobre el ejemplo de las mermeladas, conduce rápidamente al analista a percatarse de que el análisis de regresión tiene numerosas posibilidades de uso en estudios de segmentación de mercado, para determinar qué variables parecen impactar la participación en el mercado, la frecuencia de compras, la propiedad del producto, la lealtad a un producto y marca, así como muchas otras áreas.

Los científicos en agricultura emplean el análisis de regresión para explorar las relaciones entre el rendimiento de un producto (*v.gr.*, toneladas de maíz por hectárea) y el tipo y cantidad de fertilizante, la precipitación pluvial, la temperatura, los días de sol y las plagas. Las granjas modernas están equipadas con mini y microcomputadoras junto con paquetes de cómputo que ayudan en este proceso.

Los investigadores médicos utilizan el análisis de regresión para establecer lazos entre la presión sanguínea y variables independientes como la edad, clase social, peso, hábito de fumar y raza. Los doctores exploran el impacto de las comunicaciones, el número de consultas y la edad del paciente en relación con la satisfacción del paciente con el servicio.

Los directores de personal exploran las relaciones de los niveles salariales con la ubicación geográfica, las tasas de desempleo, el crecimiento de la industria, la afiliación sindical, el tipo de industria o los salarios competitivos. Los analistas financieros buscan causas de los precios altos de acciones, analizando el rendimiento de dividendos, los ingresos por acción, las aperturas de la bolsa, las expectativas de tasas de interés del consumidor, los niveles de ahorro y las tasas de inflación.

Los gerentes de publicidad intentan con frecuencia estudiar el impacto de los presupuestos en publicidad, la selección de los medios, la copia de mensajes, la frecuencia de la publicidad o la opinión de elección de las personas sobre el cambio en la actitud de consumo. En forma similar, los comercializadores pretenden determinar las ventas a partir de los gastos de publicidad, los niveles de precios, los gastos en mercados competitivos, el ingreso disponible del consumidor, así como una gran diversidad de otras variables.

Un ejemplo final ilustrará mejor la versatilidad del análisis de regresión. Los analistas en ubicación de bienes raíces encuentran que el análisis de regresión puede ser muy útil en el señalamiento de áreas geográficas en las que hay una alta o baja penetración de tipos

específicos de tiendas al menudeo. Por ejemplo, una cadena de ferreterías podría buscar una ciudad potencial para ubicar una nueva tienda, mediante el desarrollo de un modelo de regresión diseñado para pronosticar la venta de artículos en una determinada ciudad. Los investigadores podrían concentrar sus esfuerzos en aquellas ciudades en las que el modelo predijera ventas mayores a las realmente alcanzadas (como se puede determinar por muchos medios). La hipótesis es que no son potenciales las ventas de artículos de ferretería en estas ciudades.

En resumen, el análisis de regresión ha proporcionado a la administración una herramienta poderosa y versátil, para estudiar las relaciones entre una variable dependiente y diversas variables independientes. La meta consiste en comprender y controlar mejor los sucesos presentes así como predecir mejor los futuros.

GLOSARIO

Coefficiente neto de regresión. El coeficiente neto de regresión mide el cambio promedio en la variable dependiente por unidad de cambio en la variable independiente pertinente, manteniendo constantes el resto de las variables independientes.

Colinealidad. La colinealidad es la situación en la cual las variables independientes de una ecuación de regresión múltiple se encuentran altamente intercorrelacionadas.

Error estándar de la estimación. El error estándar de la estimación mide la cantidad estándar en la que los valores reales de Y difieren de los valores estimados ($\hat{Y}$).

Heteroscedasticidad. Existe heteroscedasticidad cuando los errores o residuos no tienen una varianza constante a través de un rango completo de valores.

Regresión múltiple. La regresión múltiple comprende el uso de más de una variable independiente, para pronosticar una variable dependiente.

Variables ficticias. Las variables ficticias se emplean para determinar las relaciones entre variables independientes cualitativas y una variable dependiente.

FÓRMULAS CLAVE

$$\text{Ecuación de pronóstico de regresión múltiple} \quad \hat{Y} = b_0 + b_2X_2 + b_3X_3 \quad (7.1)$$

Ecuaciones normales

$$\begin{aligned} \Sigma Y &= nb_0 + b_2\Sigma X_2 + b_3\Sigma X_3 \\ \Sigma X_2Y &= b_0\Sigma X_2 + b_2\Sigma X_2^2 + b_3\Sigma X_2X_3 \\ \Sigma X_3Y &= b_0\Sigma X_3 + b_2\Sigma X_2X_3 + b_3\Sigma X_3^2 \end{aligned} \quad (7.2)$$

Ecuación de regresión múltiple de la población

$$Y = \beta_0 + \beta_2X_2 + \beta_3X_3 + \epsilon \quad (7.3)$$

$$\text{Ecuación de regresión múltiple de la muestra} \quad Y = b_0 + b_2X_2 + b_3X_3 + e \quad (7.4)$$

$$\text{Error estándar de la estimación} \quad s_{y \cdot x_2x_3} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n - k}} \quad (7.5)$$

$$\text{Regresión estacional} \quad \hat{Y}_t = b_0 + b_2S_2 + b_3S_3 + b_4S_4 \quad (7.6)$$

PROBLEMAS

1. ¿Cuáles son las características de una buena variable de pronóstico?
2. ¿Cuáles son las suposiciones de la técnica de regresión múltiple?
3. ¿Qué mide el coeficiente neto de regresión en la regresión múltiple?
4. ¿Qué mide el error estándar de la estimación en la regresión múltiple?
5. Su ecuación de regresión múltiple es $\hat{Y} = 7.52 + 3X_2 - 12.2X_3$. Formule un punto de estimación, dados $x_2 = 20$ y $x_3 = 7$.
6. Explique cada uno de los siguientes conceptos:
 - a. Matriz de correlación
 - b. R^2
 - c. Colinealidad
 - d. Residuo
 - e. Variable ficticia
 - f. Regresión por pasos
7. La mayoría de las soluciones de cómputo para el análisis de regresión comienzan con una matriz de correlación. Esta matriz debe ser el primer paso cuando se analiza un problema que comprende más de una variable independiente. Responda las siguientes preguntas con respecto a la matriz de correlación que se muestra en la tabla siguiente:

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

- a. ¿Por qué todos los elementos de la diagonal principal son iguales a 1?
 - b. ¿Por qué aparece en blanco la mitad inferior de la matriz, bajo la diagonal principal?
 - c. Si la variable 1 es la variable dependiente, ¿cuáles variables independientes tienen el mayor grado de relación con ella?
 - d. ¿Qué clase de relación existe entre las variables 1 y 4?
 - e. ¿Muestra esta matriz de correlación alguna evidencia de colinealidad?
 - f. En su opinión, ¿qué variable o variables se incluirán en el mejor modelo de pronóstico? Explique el porqué.
 - g. Si se ejecuta el problema en un programa por pasos, ¿qué variable se registrará primero en el modelo?
8. A Jennifer Dahl, supervisora de la cadena Circle O, le gustaría pronosticar el tiempo que toma atender a un cliente en caja. Para ello, decide utilizar las siguientes variables independientes:

monto de las compras y número de artículos comprados. Jennifer reúne una muestra de datos para 18 clientes:

CLIENTE	TIEMPO DE ATENCIÓN (EN MINUTOS) Y	MONTO (EN DÓLARES) X_2	NÚMERO DE ARTÍCULOS X_3
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

- Determine la mejor ecuación de regresión.
 - Al comprar un artículo adicional, ¿cuál es el incremento promedio en el tiempo de atención?
 - Calcule el residuo para el cliente 18.
 - Calcule el error estándar de la estimación.
 - Interprete el apartado d en términos de las variables empleadas en este problema.
 - Calcule el punto de estimación del tiempo de atención si el cliente compra 14 artículos por un monto de \$70 dólares.
 - Calcule un intervalo de estimación del 95% para su pronóstico del apartado f.
 - ¿Cuáles serían las conclusiones de Jennifer?
9. Las ventas de cerveza de la Shapiro One-Stop Store se analizan utilizando como variables la temperatura y el número de personas (21 o más) que pasan por la calle. Se ejecuta una regresión con las siguientes variables:

Y = número de paquetes de seis cervezas

X_2 = temperatura alta

X_3 = contador de tráfico

Se selecciona una muestra aleatoria de 20 días. La salida parcial de cómputo arroja los siguientes resultados.

CORRELATION MATRIX: 3 BY 3				
VAR.	1	2	3	
1	1.000	0.827	0.822	
2	0.827	1.000	0.680	
3	0.822	0.680	1.000	
VARIABLE NO.	MEAN	STANDARD DEVIATION	REGRESSION COEFFICIENT	STD ERRO OF REG. COEF.
2	64.09999	17.46243	0.78207	0.22694
3	537.54980	195.59874	0.06795	0.02026
DEPENDENT				
1	59.95000	27.45083		
INTERCEPT	-26.70621			
ANALYSIS OF VARIANCE FOR THE REGRESSION				
SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION	2	11589.03500	5794.51560	36.11064
DEVIATION FROM REGRESSION	17	2727.91400	160.46553	
TOTAL	19	14316.94900		

- Analice la matriz de correlación.
 - Pruebe, a un nivel de significación de .01, la significación de los coeficientes de regresión.
 - Pronostique el volumen de cerveza vendida si la temperatura alta es de 60 grados Fahrenheit y el contador de tráfico es de 500 personas.
 - Calcule R^2 e interprete su significado en términos del problema.
 - Calcule el error estándar de la estimación.
 - Explique cómo afecta a las ventas de cerveza el incremento de un grado en la temperatura alta.
 - Indique sus conclusiones para este análisis con respecto a la precisión de la ecuación de pronóstico, así como de la contribución de las variables independientes.
10. El gerente de ventas de una gran distribuidora de partes automotrices, Hartman Auto Supplies, desea desarrollar un modelo para pronosticar desde el mes de mayo, las ventas anuales de una región. Si es posible pronosticar las ventas regionales, se podrá pronosticar el total de ventas de la compañía. Las dos variables independientes por investigar son el número de distribuidoras en la región que almacenan partes de la compañía y el número de automóviles registrados para cada región al 1o. de mayo. Los datos se presentan en el siguiente cuadro.

REGIÓN	VENTAS ANUALES (EN MILLONES), Y	NÚMERO DE DISTRIBUIDORAS X_2	NÚMERO DE AUTOMÓVILES REGISTRADOS (EN MILLONES) X_3
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

- Analice la matriz de correlación.
 - ¿Qué tanto error comprende el pronóstico de la región 1
 - Pronostique las ventas anuales para la región 12, dadas 2 500 distribuidoras y 20.2 millones de automóviles registrados.
 - Exponga la precisión del pronóstico efectuado en el apartado c.
 - Indique cómo se calculó el error estándar de la estimación.
 - Interprete los coeficientes netos de regresión. ¿Son válidos?
 - ¿Cómo podría mejorarse esta ecuación de regresión?
11. El gerente de ventas de Hartman Auto Supplies decide investigar una nueva variable, ingreso personal por región. Los datos correspondientes se presentan en la tabla siguiente:

REGIÓN	INGRESO PERSONAL (EN MILES DE MILLONES)	REGIÓN	INGRESO PERSONAL (EN MILES DE MILLONES)
1	98.5	7	67.6
2	31.1	8	19.7
3	34.8	9	67.9
4	32.7	10	61.4
5	68.8	11	85.6
6	94.7		

- ¿Contribuye el ingreso personal por región en el pronóstico de ventas?
 - Pronostique las ventas anuales de la región 12 para un ingreso personal de \$40 mil millones, utilizando las tres variables.
 - Exponga la precisión del pronóstico formulado en el apartado b.
 - ¿Qué variables incluiría en su modelo final de pronóstico? ¿Por qué?
12. La Nelson Corporation decide desarrollar una ecuación de regresión múltiple para pronosticar el comportamiento de las ventas. Se entrevista a una muestra aleatoria de 14 vendedores y se les

aplica un examen de aptitud. Además, se calcula un índice de esfuerzo empleado por cada vendedor, con base en la proporción de kilómetros recorridos en su automóvil con respecto al total de kilómetros proyectados para una cobertura adecuada de territorio. El análisis de regresión arroja los siguientes resultados:

$$\hat{Y} = 16.57 + .65X_2 + 20.6X_3$$

(.05) (1.69)

Las cantidades entre paréntesis indican el error estándar de los coeficientes netos de regresión. El error estándar de la estimación es 3.56. El error estándar de la variable ventas es $s_y = 16.57$. Las variables son

Y = comportamiento de las ventas, en miles

X_2 = calificación en el examen de aptitud

X_3 = índice de esfuerzo

- a. ¿Son bastante diferentes de cero los coeficientes de regresión, en el nivel de significación .01?
 - b. Interprete el coeficiente neto de regresión del índice de esfuerzo.
 - c. Pronostique el comportamiento de ventas para un vendedor que tiene una calificación de 75 en el examen de aptitud y un índice de esfuerzo de .5.
 - d. Calcule la suma de residuos cuadrados, $\Sigma (Y - \hat{Y})^2$.
 - e. Calcule la suma total de cuadrados, $\Sigma (Y - \bar{Y})^2$.
 - f. Calcule R^2 e interprete su significado en términos de este problema.
 - g. Calcule R_C^2 corregida.
13. El Sr. Palmer, propietario de un lote de venta de casas móviles, desea pronosticar sus ventas brutas. Selecciona en forma aleatoria datos de 40 meses sobre las cuatro variables siguientes:

X_2 = número de vendedores empleados

X_3 = temperatura promedio mensual

X_4 = número de líneas de casas móviles en inventario durante el mes

X_5 = gastos de publicidad

Los datos se presentan en la tabla P.13.

- a. Ejecute una matriz de correlación de un programa de regresión múltiple y analicela.

TABLA P. 13

VENTAS BRUTAS (EN MILES), Y	NÚMERO DE VENDEDORES, X_2	TEMPERATURAS PROMEDIO, X_3	NÚMEROS DE LÍNEAS X_4	PUBLICIDAD X_5
\$54.3	5	16.3	6	\$ 716
79.9	7	26.1	9	792
57.1	5	35.9	8	492
89.3	7	46.2	8	650
115.0	11	57.4	15	865
126.0	12	65.2	16	1,293
76.5	10	67.4	7	790
81.1	7	67.1	7	802
56.7	6	59.8	8	484
138.8	11	43.7	16	1,501
47.9	7	36.3	5	326
42.5	7	29.4	4	202
39.4	5	25.9	5	215
68.9	5	36.3	10	609
60.3	7	37.0	10	600
87.7	8	41.6	13	764
46.9	6	54.9	6	304
44.2	6	66.2	8	252
84.4	8	72.5	5	746
64.1	5	70.2	8	629
115.3	11	54.2	8	1,044
40.7	5	44.9	4	158
79.2	5	36.0	7	716
39.5	4	27.9	5	176
14.7	4	31.8	6	102
24.1	5	33.6	4	209
117.3	10	35.2	11	1,501
67.9	7	45.3	9	631
73.2	10	56.3	11	692
63.7	10	58.2	6	618
36.1	7	69.7	5	140
58.6	5	74.1	6	544
97.2	7	55.2	11	901
41.8	5	44.2	4	100
93.9	6	35.4	9	862
21.8	5	25.8	2	175
17.3	3	22.6	4	111
65.9	6	30.7	7	594
32.4	4	41.4	7	131
61.1	6	42.0	9	589

- b. Exponga la importancia de cada variable independiente.
- c. ¿Qué ecuación debe emplear el Sr. Palmer para pronosticar las ventas brutas, si desea incluir todas las variables independientes estadísticamente significativas? Utilice un nivel de significación de .01.
- d. ¿Qué ecuación de regresión aconsejaría usted al Sr. Palmer? ¿Por qué?
- e. ¿Existe un problema de colinealidad en la ecuación elegida en el apartado c? De ser así, ¿qué efectos pudiera tener?
- f. Pronostique las ventas brutas de acuerdo con lo siguiente:

Hay cuatro vendedores empleados.

La temperatura alta del mes es 40 grados Fahrenheit.

Hay en el inventario cinco líneas diferentes de casas móviles.

Los gastos en publicidad son de \$750. (Nota: Use la ecuación que seleccionó en el apartado d.)

- g. Exponga la precisión de este pronóstico.
14. La Srta. Haight, una corredora de bienes raíces, desea pronosticar la importancia de cuatro factores en la determinación de precios de terrenos. Para ello, acumula datos sobre precio, superficie, elevación y pendiente, así como índices de vista para 50 terrenos. Ejecuta los datos en un programa de correlación y obtiene la matriz correspondiente que se presenta en la tabla siguiente. La Srta. Haight ejecuta los datos en un programa de regresión por pasos.

VARIABLE	VARIABLE				
	PRECIO	SUPERFICIE	ELEVACIÓN	PENDIENTE	VISTA
Precio	1.00	.59	.66	.68	.88
Superficie		1.00	.04	.64	.41
Elevación			1.00	.13	.76
Pendiente				1.00	.63
Vista					1.00

- a. Determine qué variable se registrará primero en el modelo, cuál en segundo término, cuál en tercero y cuál al final.
 - b. ¿Qué variable(s) se incluirá(n) en la mejor ecuación de predicción?
15. National Presto es un fabricante de aparatos pequeños electrodomésticos y artículos para el hogar, incluyendo ollas de presión, calentadores, enlatadoras, freidoras, planchas, hornos rosticeros, freidoras profundas, ollas para palomitas de maíz, abrelatas, cafeteras, rebanadoras, batidoras y artículos portátiles varios. En la tabla P.15 se muestran las ventas trimestrales en millones de dólares. Las mejores ventas de Presto se presenta principalmente en la temporada navideña, de modo que existe un fuerte efecto estacional. Desarrolle un modelo de regresión múltiple empleando variables ficticias para pronosticar las ventas del tercer y cuarto trimestres de 1993. Escriba un informe resumiendo sus resultados.

TABLA P. 15

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		

Fuente: *The Value Line Investment Survey*
(Nueva York: Value Line, 1993), p.132.

CASO DE ESTUDIO 7.1

EL MERCADO DE VALORES*

A principios de 1982, Judy Johnson, vicepresidente de finanzas de una gran empresa privada de inversiones del noroeste, se enfrentó con un problema financiero. La compañía necesitaba dinero para pagar sus deudas de corto plazo por vencer y para continuar la construcción de una planta carbonífera.

La principal preocupación de Judy era estimar el mercado de valores de 10 o 30 años; la compañía necesitaba decidir entre utilizar un financiamiento de equidad o una deuda de largo plazo. Para tomar esta decisión, se requería de un pronóstico confiable de la tasa de interés que pagaría al momento de emitir las obligaciones.

Judy convocó a una reunión con su equipo financiero para discutir el problema del mercado de valores. Un miembro del grupo, Ron Peterson, recién graduado de la maestría en administración de empresas, opinó que se podría desarrollar un modelo de regresión múltiple para pronosticar las tasas de las obligaciones. Como la vicepresidenta no estaba familiarizada con la regresión múltiple, desvió la discusión en otra dirección. Después de una hora de intercambios improductivos, Judy pidió a Ron que le tuviera un informe sobre su escritorio el lunes siguiente.

Ron sabía que la clave para el desarrollo de un buen modelo de pronóstico estaba en la identificación de las variables independientes que se relacionaran con la tasa de interés pagada al momento de la emisión de las obligaciones. Después de discutir el problema con varias personas de la empresa, Ron decidió investigar las siguientes variables: un índice de las obligaciones de empresas de servicio (Moody's), un índice de ingresos sobre cargos fijos, las tasas de obligaciones de la tesorería, la madurez de las obligaciones (10 o 30 años) y la tasa líder de préstamos al momento de la emisión.

Ron reunió los datos que pensó podría relacionar con las tasas de interés de obligaciones para las emisiones de obligaciones de empresas de servicios durante 1980 y 1981. Al principio, no tenía la

*Los datos para este caso fueron aportados por una estudiante de Maestría en Administración de Empresas de la Eastern Washington University, Dorothy Mercer. El análisis lo efectuaron los estudiantes de MAE Tak Fu, Ron Hand, Dorothy Mercer, Mary Lou Redmond y Harold Wilson.

certeza de cómo manejar el índice de obligaciones de las empresas de servicio. Finalmente, decidió considerar sólo las empresas de servicios con un índice igual o ligeramente mayor que el de su compañía. Esta decisión le proporcionó una muestra de 93 emisiones para analizar. Aunque aún estaba preocupado con respecto a la validez de utilizar los índices de obligaciones como datos de intervalo. De modo que llamó a su exprofesor de estadística y aprendió que las variables ficticias resolverían el problema. Así, codificó los índices de obligaciones de la siguiente manera:

$X_2 = 1$ si el índice de las obligaciones es A; 0 en caso contrario

$X_3 = 1$ si el índice de las obligaciones es AA; 0 en caso contrario

Si el índice de las obligaciones es BAA, tanto X_2 como X_3 son cero.

El siguiente paso para Ron fue seleccionar un programa de regresión múltiple de la biblioteca de programas de la computadora e ingresar los datos. En la ecuación del modelo completo se incluyeron las siguientes variables:

Variable 1: Y = tasa de interés pagada por la obligación al momento de la emisión

Variable 2: $X_2 = 1$ si el índice de la obligación es A

Variable 3: $X_3 = 1$ si el índice de la obligación es AA

Variable 4: X_4 = índice de ingresos sobre cargos fijos

Variable 5: X_5 = Tasas de las obligaciones de la Tesorería de los EUA (para 10 y 30 años) al momento de la emisión

Variable 6: X_6 = madurez de las obligaciones (10 o 30 años)

Variable 7: X_7 = tasa líder de préstamos al momento de la emisión

Los datos reales se presentan en el Apéndice D.

Ron decidió analizar la matriz de correlación que se presenta en la tabla 7.16. No le sorprendió encontrar una alta relación positiva entre la tasa de interés pagada al momento de la emisión de las obligaciones y la tasa de interés de los bonos de la Tesorería, $r_{15} = .883$. También esperaba una relación positiva relativamente alta entre la variable dependiente y la tasa líder de préstamos ($r_{17} = .596$). No le sorprendió mucho el descubrir que estas dos variables también estaban relacionadas entre sí (colinealidad, $r_{57} = .713$). La relación negativa entre la variable dependiente y la duración de maduración de las obligaciones (10 o 30 años), $r_{16} = -.221$, también fue un resultado razonable para Ron.

TABLA 7.16 MATRIZ DE CORRELACIÓN PARA EL ESTUDIO DEL MERCADO DE VALORES

VARIABLE	VARIABLE						
	1	2	3	4	5	6	7
1	1.000	-.347	.173	.097	.883	-.221	.596
2	-.347	1.000	-.399	.037	-.256	.278	-.152
3	.173	-.399	1.000	.577	.291	.010	.342
4	.097	.037	.577	1.000	.253	.094	.255
5	.883	-.256	.291	.253	1.000	-.477	.713
6	-.221	.278	.010	.094	-.477	1.000	-.314
7	.596	-.152	.342	.255	.713	-.314	1.000

A continuación, Ron ejecutó el modelo completo conteniendo todas las variables de predicción. La revisión de los valores t , que se presentan en la tabla 7.17, indicó que el índice de ingresos sobre cargos fijos, variable 4 y la tasa líder de interés, variable 7, no hacían ninguna contribución al pronóstico de la variable dependiente.

TABLA 7.17 EJECUCIÓN DEL MODELO COMPLETO PARA EL ESTUDIO DEL MERCADO DE VALORES

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD ERROR OF REG. COEF.	COMPUTED T VALUE
2	0.47312	0.50198	-0.34654	-0.82853	0.13423	-6.17266
3	0.15054	0.35954	0.17257	-0.88937	0.22474	-3.95737
4	2.59053	0.64437	0.09701	-0.24165	0.11352	-2.12877
5	12.63572	1.45429	0.88303	1.25751	0.05964	21.08649
6	21.18279	9.98356	-0.22122	0.06284	0.00659	9.53710
7	16.78168	2.94729	0.59627	-0.00313	0.02720	-0.11502
DEPENDENT						
1	14.99006	1.69549				
INTERCEPT		-1.02621				
MULT. CORR		0.95441		R SQUARED		0.91090
STD. ERROR		0.52344		CORRECTED R SQUARED		0.90578
ANALYSIS OF VARIANCE FOR THE REGRESSION						
SOURCE OF VARIATION			DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION			6	240.90656	40.15109	146.54047
DEVIATION FROM REGRESSION			86	23.56342	0.27399	
TOTAL			92	264.46997		

Ron concluyó que la variable 4 no estaba relacionada con la variable dependiente ($r_{14} = .097$) y que la variable 7 era colineal con la variable 5. Decidió eliminar ambas. Los resultados de la ejecución de cómputo para este modelo se muestran en la tabla 7.18. Los valores calculados de t fueron significativos para cada una de las variables independientes.

En su informe para Judy, Ron incluyó los siguientes comentarios:

1. El modelo de regresión múltiple apropiado, $\hat{Y} = -1.276 - .929X_2 - 1.175X_3 + 1.233X_5 + .061X_6$, explica el 90.6% de la varianza de la variable dependiente.
2. El error estándar de la estimación, que es el error típico o promedio al utilizar el modelo de pronóstico es .53. Por ello, alrededor del 95% de los valores reales de la variable dependiente se ubicarán dentro de 1.06 unidades de un pronóstico dado.
3. Los coeficientes de regresión son válidos y confiables.

Ron quedó muy satisfecho con su esfuerzo y supone que Judy también lo estará.

PREGUNTA

¿Qué preguntas cree usted que formulará Judy a Ron?

TABLA 7.18 MODELO FINAL PARA EL ESTUDIO DEL MERCADO DE VALORES

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	STD ERROR OF REG. COEF.	COMPUTED T VALUE
2	0.47312	0.50198	-0.34654	-0.92932	0.12637	-7.35422
3	0.15054	0.35954	0.17257	-1.17506	0.17810	-6.59793
5	12.63572	1.45429	0.88303	1.23307	0.04599	26.80913
6	21.18279	9.98356	-0.22122	0.06147	0.00665	9.24634
DEPENDENT						
1	14.99006	1.69549				
INTERCEPT		-1.27627				
MULT. CORR		0.95195		R SQUARED		0.90621
STD. ERROR		0.53093		CORRECTED R SQUARED		0.90304
ANALYSIS OF VARIANCE FOR THE REGRESSION						
SOURCE OF VARIATION			DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARES	F VALUE
ATTRIBUTABLE TO REGRESSION			4	239.66434	59.91608	212.55716
DEVIATION FROM REGRESSION			88	24.80563	0.28188	
TOTAL			92	264.46997		

CASO DE ESTUDIO 7.2

VENTAS DE RESTAURANTE

Jim Price, quien estaba trabajando en su grado de maestría en administración de empresas, laboraba en un pequeño restaurante cerca de la Marquette University en Milwaukee, Wisconsin. Un día, la gerente del restaurante le pidió a Jim que acudiera a su oficina; le indicó que estaba muy interesada en pronosticar las ventas semanales y quería saber si Jim estaba dispuesto a colaborar. Como Jim acababa de tomar un curso de estadísticas avanzadas dijo que le encantaría el reto.

Jim pidió a la gerente que le proporcionara cualquier información histórica que tuviera disponible. Ella le indicó que las ventas del restaurante de la semana anterior se recopilaban cada lunes por la mañana. Jim comenzó su análisis obteniendo los datos de ventas semanales de la semana que terminó el domingo 4 de enero de 1981 hasta la semana que terminó el domingo 26 de diciembre de 1982; un total de 104 observaciones.

La media de ventas semanales para las 104 semanas resultó ser de \$4 862. La fig. 7.5 es una gráfica de las ventas semanales a través del tiempo. La gráfica indica que las ventas semanales fueron muy inconsistentes, con una fluctuación de \$1 870 a \$7 548, con muy poca tendencia. Como Jim concluyó recientemente un curso sobre análisis de regresión, decidió usar las ventas semanales como la variable dependiente y buscar la posibilidad de encontrar predictores variables independientes útiles.

Jim probó con tres predictores. El primero fue el periodo. El segundo predictor fue una variable ficticia que indicaba si la universidad estaba o no en sesión completa durante esa semana (0 = no en sesión completa, 1 = en sesión completa). Una revisión de los datos de ventas en la fig. 7.5 reveló que las ventas semanales disminuían siempre que la universidad no estaba en sesión completa, es decir, durante los recesos de navidad, primavera y verano. Jim no se sorprendió, pues el restaurante se ubica cerca del campus de la universidad y la mayoría de sus clientes son miembros de la comunidad de

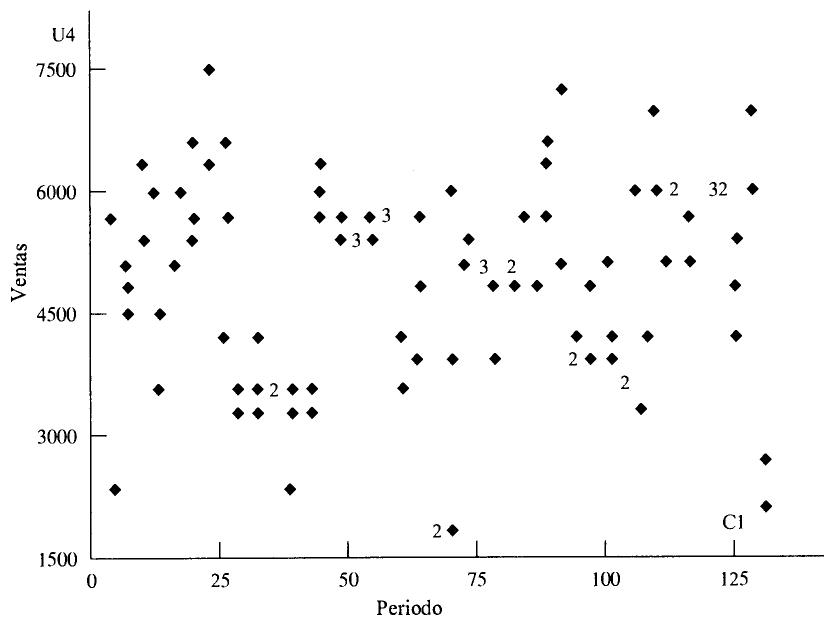


Figura 7.5 Ventas de restaurante, de enero de 1981 a diciembre de 1982.

Marquette. El tercer pronóstico que probó Jim fueron las ventas retrasadas una semana, ya que la revisión de la fig. 7.5 indicaba que las ventas de dos semanas adyacentes eran frecuentemente similares.*

Por medio de Minitab, Jim calculó las correlaciones simples entre los tres pronósticos potenciales y la variable dependiente, ventas semanales. Los resultados se presentan en la matriz de correlación de la tabla 7.19. Como Jim esperaba, casi no había tendencia en las ventas semanales como lo 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 está o no en sesión completa tiene un buen potencial como pronóstico de las ventas de la semana actual. Las ventas de la semana actual estuvieron relacionadas en forma moderada, $r = .58$, con las ventas retrasadas una semana. No obstante, Jim también notó que la variable ficticia estaba moderadamente relacionada, $r = .49$, con las ventas retrasadas una semana.

De nuevo, con la ayuda del paquete de cómputo Minitab, Jim experimentó con diversos modelos de regresión. Los resultados de éstos se presentan en la tabla 7.20. Como los datos manifestaron una

TABLA 7.19 MATRIZ DE CORRELACIÓN DE LAS VENTAS DE RESTAURANTE

	VENTAS ACTUALES	PERIODO	VARIABLE FICTICIA	VENTAS RETRASADAS UNA SEMANA
Ventas actuales	1.000	.049	.772	.580
Periodo		1.000	-.048	.120
Variable ficticia			1.000	.490
Ventas retrasadas				1.000

tendencia casi nula, el pronóstico “periodo” agregó muy poco valor productivo al modelo de regresión. Nótese que el modelo 4 tiene sólo una R^2 ligeramente más alta que el modelo 2, y que ambos modelos poseen una cantidad significativa de autocorrelación (la autocorrelación se abordará en forma más completa en el cap. 9). Además, los modelos 3 y 5 tienen la misma R^2 , mientras que el modelo 7 sólo tiene una R^2 ligeramente superior al modelo 6. Por otra parte, el pronóstico “ventas retrasadas” incorpora una buena cantidad de poder predictivo al modelo de regresión. El modelo 6 tiene una R^2 significativamente más alta que el modelo 2, sin tener una autocorrelación significativa.

TABLA 7.20 MODELOS DE REGRESIÓN DEL RESTAURANTE

PRONOSTICO(S) DEL MODELO	R2	ESTADÍSTICA DURBIN-WATSON	¿AUTOCORRELACIÓN SIGNIFICATIVA EN EL NIVEL .05?	COLINEALIDAD
1. Periodo	.0024	.81	Sí	Ninguna
2. Variable ficticia	.596	1.30	Sí	Ninguna
3. Ventas desfasadas	.336	1.89	No	Ninguna
4. Periodo y ficticia	.603	1.32	Sí	Muy poca
5. Periodo y ventas desfasadas	.336	1.89	No	Poca
6. Ficticia y ventas retrasadas	.649	1.74	No	Alguna
7. Periodo, ficticia y ventas retrasadas	.651	1.73	No	Moderada

Jim decidió seleccionar el modelo de regresión 6 para pronosticar las ventas semanales, por las siguientes razones:

1. El modelo 6 tiene la segunda R^2 mayor, sólo .002 por debajo del modelo 7.
2. Los parámetros del modelo 6 fueron cada uno diferentes de cero en forma significativa al nivel de .01.
3. El modelo 6 no exhibió una cantidad significativa de autocorrelación.
4. El modelo 6 es más simple que el modelo 7 y no tiene tanta colinealidad.

El modelo de regresión utilizado por Jim fue $\hat{Y} = 2\,614.3 + 1\,610.7$ (variable ficticia) + .2605 (ventas retrasadas). La $R^2 = .649$ significa que se puede explicar el 64.9% de la variación en las ventas semanales mediante el conocimiento de si la universidad estaba o no en sesión completa y de las ventas de la semana previa. La ecuación de regresión implica que el promedio de ventas semanales es de alrededor de \$1 611 más alto cuando Marquette se encuentra en sesión completa, manteniendo constantes las ventas de la semana anterior.

Jim quedó satisfecho con su esfuerzo pero se preguntó si otro tipo de modelo de pronóstico pudiera ser más efectivo. Por este motivo, decidió tomar un curso sobre pronósticos.

PREGUNTAS

1. ¿Fue correcto el uso de la variable ficticia por parte de Jim?
2. ¿Fue correcto que Jim empleara las ventas retrasadas una semana como variable de predicción?
3. ¿Está usted de acuerdo con las conclusiones de Jim?
4. ¿Sería otro tipo de modelo de pronóstico más efectivo para el pronóstico de las ventas semanales?

CASO DE ESTUDIO 7.3

MR. TUX

John Mosby ejecutó un análisis de regresión simple utilizando el periodo como variable de predicción y observó un decepcionante valor de r -cuadrada de .49 (véase el caso de Mr. Tux al final del cap. 6). Como sabe que sus ventas tienen un fuerte componente estacional, decide ejecutar un modelo de regresión múltiple estacional.

Para realizar su análisis, John emplea el paquete de cómputo Sibyl/Runner. Este programa puede incluir cualesquier variables de pronóstico, y crear de manera automática variables ficticias para el efecto estacional. Como los datos de Mr. Tux son mensuales, Sibyl/Runner codifica una variable estacional para cada enero como 1, otra para cada febrero como 1 y así sucesivamente. Sólo hay 11 variables estacionales, para los meses de enero a noviembre. Diciembre es el mes base y no tiene un ajuste estacional en la ecuación final de regresión.

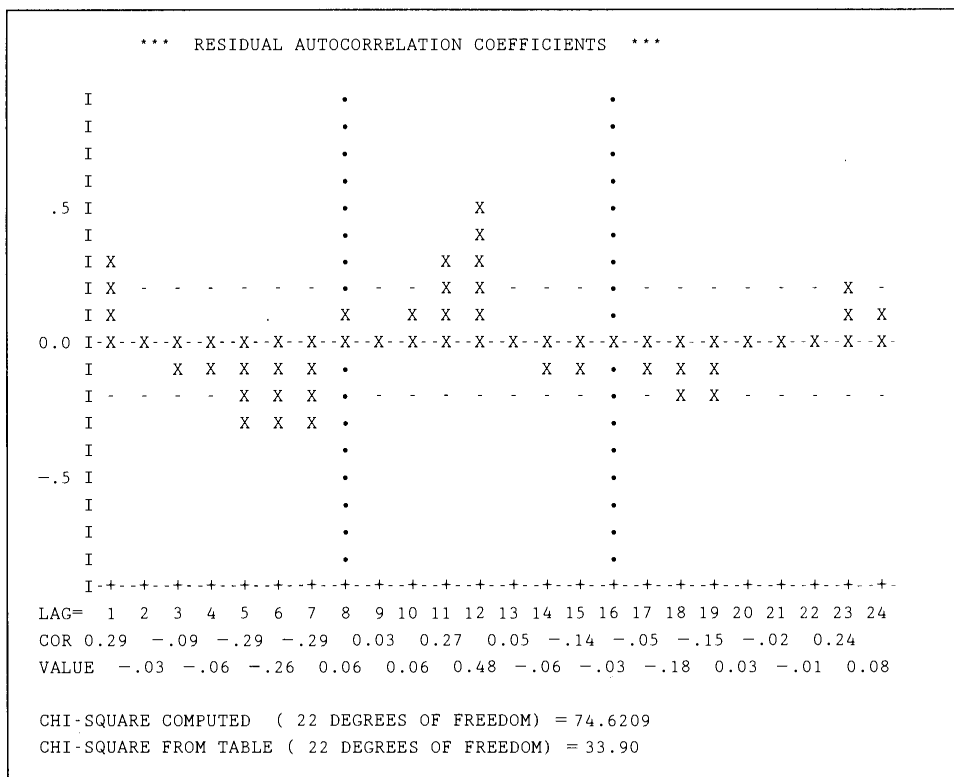
John ejecuta el programa utilizando las 12 variables de predicción: una para el número de mes, del 1 al 96, para representar el periodo y las otras once para representar el efecto mensual. A continuación se presentan las estadísticas clave generadas por su programa de cómputo:

R -cuadrada = .877	F = 49.385
Durbin-Watson = 1.41	DAM = 277.3
EMC = 124 795.4	PEMA = 72.71%
PME = 30.69%	

Las mediciones de error se efectuaron por medio del programa para las 96 observaciones. John calculó a mano el PEMA para los últimos 12 meses de datos y resultó ser de 21.25%. Él está muy decepcionado con las altas cifras de las mediciones de error, pues esperaba que resultaran pronósticos precisos al incluir los efectos estacionales.

Al revisar los valores de t de su regresión, John encontró que tanto para el término constante como para el coeficiente de regresión del periodo, los valores de t eran diferentes de cero en forma significativa (-2.3 y 19.5).

Al graficar los coeficientes de autocorrelación de los residuos del modelo (véase la gráfica siguiente), John se encontró con más problemas.



En especial le molestó la declaración al final de la impresión de Sibyl/Runner:

Este análisis indica que no se puede defender en forma estadística el método de pronóstico, ya que existe cierto patrón en los residuos que se puede pronosticar. Las autocorrelaciones que no parecen aleatorias están en los periodos 23, 12, 11, 7, 6, 5, 1.

PREGUNTAS

1. ¿Qué tan válido es el método de pronóstico de John?
2. Explique los índices t para las variables ficticias.
3. John está comprensiblemente molesto al fallar en el desarrollo de un buen método de pronóstico utilizando el periodo y los efectos estacionales. ¿Como le explicaría usted esta falla?
4. ¿Qué cree usted que debería intentar ahora John en su esfuerzo para pronosticar los volúmenes de ventas de Mr. Tux?

CASO DE ESTUDIO 7.4

CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC), se describió en el cap. 1 (caso de estudio 1.2) y en el cap. 3 (caso de estudio 3.4).

El director ejecutivo, Marv Harnishfeger, llegó a la conclusión de que la variable más importante que requería pronosticar CCC era el número de nuevos clientes por atender en el resto de 1993. Marv le proporcionó a Dorothy Mercer los datos mensuales de número de nuevos clientes atendidos en el periodo de enero de 1985 a marzo de 1993 (véase Caso de estudio 4.3). En el Caso de estudio 4.3, Dorothy usó el análisis de autocorrelación para explorar el patrón de datos. En el Caso de estudio 6.5, probó el número de personas que reúnen cupones de descuento y un índice de actividad empresarial para desarrollar un modelo de regresión para pronosticar el resto de 1993.

Dorothy no quedó satisfecha con los resultados de su modelo de regresión. Decidió intentar la regresión múltiple y pidió a Marv que pensara en otras variables que pudieran relacionarse con el número de clientes atendidos. Marv le indicó que podía intentar el número de quiebras registradas y el número de permisos de construcción emitidos.

Dorothy localizó datos para ambas variables para el periodo de enero de 1986 a diciembre de 1992:

QUIEBRAS

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

PERMISOS DE CONSTRUCCIÓN EMITIDOS

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

Dorothy desarrolló un modelo de regresión múltiple que utilizaba el número de personas que reúnen cupones de descuento, el índice de actividad empresarial, el número de quiebras registradas y el número de permisos de construcción emitidos. También creó un modelo con base en la suposición de que los datos fueran estacionales.

PREGUNTAS

1. Determine si existen relaciones significativas entre el número de clientes atendidos y cualquiera de estas variables.

2. Desarrolle una ecuación de regresión y úsela para pronosticar el número de nuevos clientes atendidos para el resto de 1993.
3. Compare los resultados de su pronóstico con las observaciones reales de 1993.
4. Suponga que desarrolló una buena ecuación de regresión. ¿Podría emplearla para pronosticar el resto de 1993?
5. Determine si el modelo de agregación utilizando variables ficticias para la estacionalidad fue un buen modelo.
6. Utilice el modelo estacional para pronosticar los tres primeros meses de 1993 y compare sus resultados.

VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Forecast Plus puede utilizar archivos de datos creados mediante otros programas como Lotus 1-2-3 y dBase. Sin embargo, pudiera ser necesario emplear un programa de conversión para modificar el formato del archivo de datos. Forecast Plus sólo puede leer archivos ASCII. El programa de conversión (FX2FR) se ejecuta a partir del disco principal, Data Management Disk.

Por ejemplo, creamos un archivo de datos (CLIENT), en ASCII, que contiene las variables clientes, cupones, permisos e índice. El primer registro de este archivo, que representa los datos de enero de 1989 es

```
15024450 61107
```

Para convertir este archivo de manera que pueda ejecutarse con Forecast Plus, escriba el nombre del programa de conversión:

```
A:\FX2FR
```

El sistema solicitará el nombre del archivo a convertir

```
What is the name of the data file? CLIENT
```

A continuación, se le pedirá el formato que tiene dicho archivo ("Enter data file format"):

- ```
1. StatPac file
2. Constant field width
3. Variable field width
```

```
Which? 3
```

Como nuestro archivo tiene campos de longitud variable, la elección es 3. Enseguida se solicita el número de variables

```
Enter the number of variables: 4
```

Después de registrar 4, el programa solicita la longitud de cada variable:

```
Field width for Variable # 1: 3
Field width for Variable # 2: 5
Field width for Variable # 3: 3
Field width for Variable # 4: 3
```

Se le pedirá el nombre del archivo convertido ("Enter Name of Output File: CLIENTFP"). El archivo de datos ya ha sido convertido y se puede ejecutar en Forecast Plus. Las variables fueron separadas por comas y ahora el primer registro es

```
150,24450,61,107
```

A continuación, se debe crear el archivo de etiquetas. En el menú de administración de datos (Data Management Menu), se selecciona la opción 1 (ENTER NEW DATA); el programa le solicitará el nombre del archivo de datos al que desea incorporarle un archivo de etiquetas ("What data file would you like to add to?"). Después de responder CLIENTFP, el programa indica lo siguiente:

```
The label file B:CLIENTFP does not exist

Do you want to create it (y/n)? y

 *** Labels File Creation ***

 Creating labels file B:CLIENTFP.LBL

Number of variables: 4

Type of labeling: YEAR

Starting label: 1989
Starting observation: 1
```

```

Number of observations between each label: 12
Increment of the labels: 1

*** Labels File Creation ***

Variable 1 label: CLIENTS
Variable 2 label: FOOD
Variable 3 label: PERMITS
Variable 4 Label: INDEX

```

Ahora, se llama al paquete de pronósticos seleccionando la opción 3 del menú principal. Se le solicita el nombre del archivo de datos

```

What is the name of the data file? CLIENTFP

```

El menú le permite elegir entre 13 diferentes técnicas de pronóstico. La pantalla correspondiente se muestra en la sección de Forecast Plus del cap. 5.

Para utilizar los cupones de descuento, los permisos y el índice de actividad como nuestras variables de predicción y los clientes atendidos como la variable dependiente, se utiliza el programa de regresión múltiple, opción 11. La pantalla muestra lo siguiente.

```

*** Multiple Regression **

Enter the variables to use in the multiple regression. Enter
the dependent variable, followed by the independent variables
(LIST will list the variables):

CLIENTS,FOOD,PERMITS,INDEX

Descriptive statistics: Y
Regression statistics: Y
Regression coefficients Y
Simple correlation matrix: Y
Partial correlation matrix: Y
Inverted correlation matrix: N
Number of records to use in residual analysis: 50

Time plot of original data, forecasts, and error: Y
Residual autocorrelation function: Y
Interactive prediction N

Press ENTER to accept:

```

Nótese que en el orden de registro de las variables fue primero la dependiente, seguida de las independientes, separadas por comas. Usted podría haber deseado usar números de variable en lugar de nombres. Para ejecutar este modelo de regresión múltiple, se pueden usar las opciones por omisión ya proporcionadas por el sistema. La fig. 7.6 muestra los resultados cuando las respuestas son Y, Y, Y, Y, Y, N, 0, N, Y y N.

Por último se le pregunta si desea almacenar los pronósticos sobre residuos

Do you want to save the forecasts or residuals (y/n)?

#### CLIENTS - MULTIPLE REGRESSION ANALYSIS

##### Variables in the equation - DESCRIPTIVE STATISTICS

| Var. | Variable label | Mean       | Standard Dev. |
|------|----------------|------------|---------------|
| DV   | CLIENTS        | 127.6471   | 40.8949       |
| IV1  | FOOD           | 27781.6275 | 7674.1176     |
| IV2  | PERMITS        | 196.9608   | 96.2239       |
| IV3  | INDEX          | 106.4118   | 27.7994       |

##### REGRESSION STATISTICS

Coefficient of multiple determination = 0.7377 (Corrected = 0.7267)  
 Coefficient of multiple correlation = 0.8589 (Corrected = 0.8525)  
 Standard error of multiple estimate = 21.6043 (Corrected = 22.0498)

F-Ratio = 44.0513  
 Degrees of freedom = 3 & 47  
 Probability of chance = 0.0000  
 Number of observations = 51

##### REGRESSION COEFFICIENTS

Constant = -7.9141

| Var. | Coeff.  | Beta    | F-ratio | Prob. | Std. Error |
|------|---------|---------|---------|-------|------------|
| IV1  | -0.0011 | -0.2088 | 0.229   | 0.634 | 0.0023     |
| IV2  | -0.0094 | -0.0221 | 0.037   | 0.849 | 0.0488     |
| IV3  | 1.5818  | 1.0752  | 7.271   | 0.010 | 0.5866     |

##### SIMPLE CORRELATION MATRIX

|     |   | DV    | IV1   | IV2   |
|-----|---|-------|-------|-------|
| IV1 | I | 0.828 |       |       |
| IV2 | I | 0.503 | 0.700 |       |
| IV3 | I | 0.857 | 0.979 | 0.625 |

**Figura 7.6** Regresión múltiple para el caso de estudio 7.4.

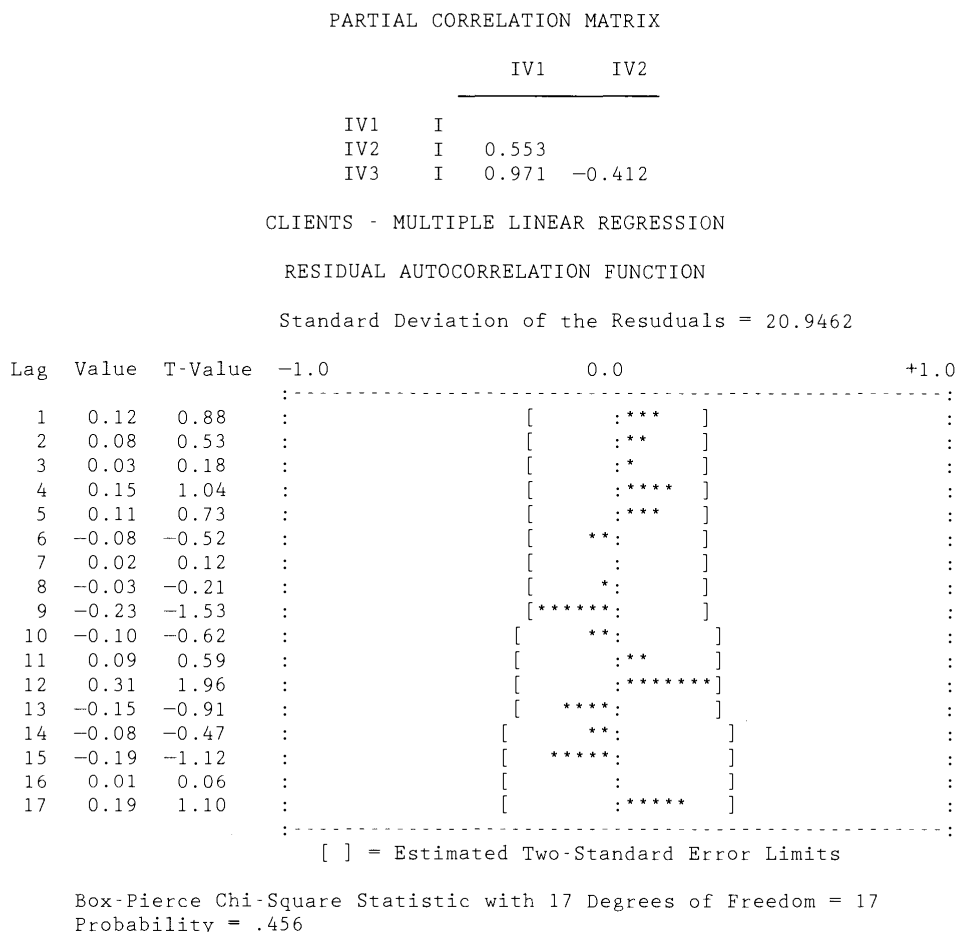


**Figura 7.6** (Continuación)

## PRONÓSTICOS MEDIANTE HOJAS DE CÁLCULO

**El problema.** En los ejemplos 7.1 al 7.4, se utilizó el análisis de regresión múltiple para determinar si se podían emplear el precio de venta y los gastos en publicidad para pronosticar las ventas semanales de galones de leche (tabla 7.3) para el ejemplo del Sr. Bump. Este pronóstico se puede llevar a cabo mediante Lotus 1-2-3 (versión 2.3).

### La solución mediante hoja de cálculo

1. Para acceder la hoja de cálculo BUMP.WK1, escriba /FILE RETRIEVE o /FR. Cuando aparezca el mensaje



```
Name of file to retrieve: A:\
```

En la tercera línea deberá ver el nombre de la hoja de cálculo que almacenó al final del cap. 6. Después de señalar BUMP.WK1, oprima <Enter>.

2. Para borrar los resultados de la regresión anterior, escriba /RANGE ERASE o /RE y aparecerá al mensaje

```
Enter range to erase:
```

Escriba el rango D3..G12.

3. Ingrese ahora la variable  $X$  en la columna C.
4. Escriba /DATA REGRESSION o /DR, que equivale a seleccionar Data File del menú principal y Regression del submenú. El menú de regresión muestra lo siguiente

```
X-Range Y-Range Output-Range Intercept Reset Go Quit
```

Se selecciona  $X$ -Range y en la pantalla se solicita el rango respectivo

```
Enter independent variables, or X range: B3..C12
```

Usted puede ya sea escribir B3..C12 o definir las variables  $X$  resaltándolas. Una vez definido el rango de  $X$ , se selecciona Output-Range para indicar el área en que se escribirán los resultados

```
Enter output range: E3..H11
```

Ya seleccionado el rango para los resultados, escriba GO para ejecutar la regresión. La salida se muestra en la fig. 7.7.

```

E3:
Fill Table Sort Query Distribution Matrix Regression Parse
Calculate linear regression

```

|           | A        | B   | C  | D | E | F | G | H |
|-----------|----------|-----|----|---|---|---|---|---|
| 1         | Y        | X2  | X3 |   |   |   |   |   |
| 2         |          |     |    |   |   |   |   |   |
| 3         | 10       | 1.3 | 9  |   |   |   |   |   |
| 4         | 6        | 2   | 7  |   |   |   |   |   |
| 5         | 5        | 1.7 | 5  |   |   |   |   |   |
| 6         | 12       | 1.5 | 14 |   |   |   |   |   |
| 7         | 10       | 1.6 | 15 |   |   |   |   |   |
| 8         | 15       | 1.2 | 12 |   |   |   |   |   |
| 9         | 5        | 1.6 | 6  |   |   |   |   |   |
| 10        | 12       | 1.4 | 10 |   |   |   |   |   |
| 11        | 17       | 1   | 15 |   |   |   |   |   |
| 12        | 20       | 1.1 | 21 |   |   |   |   |   |
| 13        |          |     |    |   |   |   |   |   |
| 14        |          |     |    |   |   |   |   |   |
| 15        |          |     |    |   |   |   |   |   |
| 16        |          |     |    |   |   |   |   |   |
| 17        |          |     |    |   |   |   |   |   |
| 18        |          |     |    |   |   |   |   |   |
| 19        |          |     |    |   |   |   |   |   |
| 20        |          |     |    |   |   |   |   |   |
| 04-Jan-94 | 10:17 AM |     |    |   |   |   |   |   |

```

Regression Output:
Constant 16.40636
Std Err of Y Est 1.507196
R Squared 0.931928
No. of Observations 10
Degrees of Freedom 7
X Coefficient(s) -8.24758 0.585100
Std Err of Coef. 2.196057 0.133671

```

NUM

**Figura 7.7** Regresión múltiple mediante Lotus para el ejemplo del Sr. Bump.

## PAQUETE DE CÓMPUTO MINITAB

**El problema.** En el ejemplo 7.11, Pam Weigand deseaba ejecutar una regresión por pasos para los datos de la Zurenko Pharmaceutical Company, con el propósito de pronosticar qué aspirante se convertiría en un buen vendedor.

### La solución con Minitab

```

MTB > READ 'ZURENKO.DAT' C1-C6
30 ROWS READ

```

| ROW | C1 | C2 | C3   | C4  | C5 | C6  |
|-----|----|----|------|-----|----|-----|
| 1   | 44 | 10 | 22.1 | 4.9 | 0  | 2.4 |
| 2   | 47 | 19 | 22.5 | 3.0 | 1  | 2.6 |
| 3   | 60 | 27 | 23.1 | 1.5 | 0  | 2.8 |
| 4   | 71 | 31 | 24.0 | 0.6 | 3  | 2.7 |
| .   | .  | .  | .    | .   | .  | .   |

```

MTB > NAME C1 'SALES' C2 'APTITUDE' C3 'AGE' C4 'ANXIETY'
MTB > NAME C5 'EXP' C6 'GPA'
MTB > STEP C1 USING PREDICTORS C2-C6

```

STEPWISE REGRESSION OF SALES ON 5 PREDICTORS, WITH N = 30

| STEP     | 1       | 2      |
|----------|---------|--------|
| CONSTANT | -100.85 | -86.79 |

|          |       |       |
|----------|-------|-------|
| AGE      | 6.97  | 5.93  |
| T-RATIO  | 7.01  | 10.60 |
| APTITUDE |       | 0.200 |
| T-RATIO  |       | 8.13  |
| S        | 6.85  | 3.75  |
| R-SQ     | 63.70 | 89.48 |

- El comando READ se usa para capturar los datos del archivo ZURENKO.DAT en las columnas C1 a C6.
- El comando NAME se emplea para dar nombre a las variables en cada columna.
- El comando STEP se utiliza para desarrollar una regresión por pasos usando C1 como la variable dependiente y C2 a C6 como las variables independientes.
- La salida para el paso 1 muestra que la ecuación es

$$\hat{Y} = -100.85 + 6.97 (\text{edad})$$

y explica el 63.7% de la varianza.

- La salida para el paso 2 muestra que la ecuación final es

$$\hat{Y} = -86.79 + 5.93 (\text{edad}) + .20 (\text{aptitud})$$

y explica el 89.48% de la varianza.

## PAQUETE DE CÓMPUTO SAS

**El problema.** En los problemas 10 y 11, el gerente de ventas de Hartman Auto Supplies desea desarrollar un modelo para pronosticar el total de las ventas anuales de una región. Decide desarrollar un modelo de regresión por pasos utilizando como variables independientes el número de distribuidoras, el número de automóviles registrados y el ingreso personal.

### La solución SAS

```
$ SAS/CC=PRINT
OPTIONS PAGESIZE=60 LINESIZE=80;
TITLE 'STEPWISE ANALYSIS FOR PROBLEMS 10 & 11';
DATA HARTMAN;
INPUT SALES OUTLETS AUTOS INCOME;
CARDS;
52.3 2011 24.6 98.5
26.0 2850 22.1 31.1
20.2 650 7.9 34.8
16.0 480 12.5 32.7
30.0 1694 9.0 68.8
```

```

46.2 2302 11.5 94.7
35.0 2214 20.5 67.6
 3.5 125 4.1 19.7
33.1 1840 8.9 67.9
25.2 1233 6.1 61.4
38.2 1699 9.5 85.6
PROC STEPWISE;
 MODEL SALES=OUTLETS AUTOS INCOME/
 SLENTY=.05;
ENDSAS;

```

- La declaración TITLE da nombre al programa SAS.
- La declaración DATA proporciona un nombre de archivo para los datos.
- La declaración INPUT da nombre y especifica el orden de los cuatro campos en las líneas de datos.
- La declaración CARDS indica que las siguientes 11 líneas son imágenes de tarjeta que representan a la variable  $Y$  y a las tres variables  $X$ .
- El comando PROC STEPWISE y el subcomando MODEL, indican que se van a registrar en una ecuación de regresión las variables distribuidoras, automóviles e ingreso (OUTLETS, AUTOS, INCOME), para predecir la variable independiente ventas (SALES).
- La declaración SLENTY = .05, especifica el nivel de significación para que una variable se registre en el modelo final de regresión.

En la fig. 7.8, se muestra la salida de cómputo para esta ejecución de SAS.

| STEPWISE ANALYSIS FOR PROBLEMS 10 & 11          |                         |                                          |                        |       |        |
|-------------------------------------------------|-------------------------|------------------------------------------|------------------------|-------|--------|
| Stepwise Procedure for Dependent Variable SALES |                         |                                          |                        |       |        |
| Step 1                                          | Variable INCOME Entered | R-square = 0.87682179 C(p) = 25.76209947 |                        |       |        |
|                                                 | DF                      | Sum of Squares                           | Mean Square            | F     | Prob>F |
| Regression                                      | 1                       | 1660.02452235                            | 1660.02452235          | 64.06 | 0.0001 |
| Error                                           | 9                       | 233.20456856                             | 25.91161873            |       |        |
| Total                                           | 10                      | 1893.22909091                            |                        |       |        |
| Variable                                        | Parameter Estimate      | Standard Error                           | Type II Sum of squares | F     | Prob>F |
| INTERCEP                                        | 1.03230539              | 3.88620209                               | 1.82835352             | 0.07  | 0.7965 |
| INCOME                                          | 0.47426771              | 0.05925344                               | 1660.02452235          | 64.06 | 0.0001 |

**Figura 7.8** Salida de cómputo de SAS para los problemas 10 y 11.

```

Bounds on condition number: 1, 1
.....

Step 2 Variable AUTOS Entered R-square = 0.96503519 C(p) = 4.29970153

 DF Sum of Squares Mean Square F Prob>F
Regression 2 1827.03269576 913.51634788 110.40 0.0001
Error 8 66.19639515 8.27454939
Total 10 1893.22909091

Variable Parameter Standard Type II
 Estimate Error Sum of Squares F Prob>F
INTERCEP -4.02689773 2.46798811 22.02923556 2.66 0.1414
AUTOS 0.62092180 0.13821024 167.00817341 20.18 0.0020
INCOME 0.43016878 0.03489322 1257.59286751 151.98 0.0001

Bounds on condition number: 1.085938, 4.343753
.....

All variables left in the model are significant at the 0.1500 level.
No other variable met the 0.0500 significance level for entry into the
model.

```

Summary of Stepwise Procedure for Dependent Variable SALES

| Step | Variable<br>Entered Removed | Number<br>In | Partial<br>R**2 | Model<br>R**2 | C(p)    | F       | Prob>F |
|------|-----------------------------|--------------|-----------------|---------------|---------|---------|--------|
| 1    | INCOME                      | 1            | 0.8768          | 0.8768        | 25.7621 | 65.0649 | 0.0001 |
| 2    | AUTO                        | 2            | 0.0882          | 0.9650        | 4.2997  | 20.1834 | 0.0020 |

**Figura 7.8** (Continuación)

## BIBLIOGRAFÍA

- BELSLEY, D. A. "Colinearity and Forecasting." *Journal of Forecasting* 3 (1984): 183–196.
- DRAPER, N. R. y SMITH, H. *Applied Regression Analysis*. Nueva York: John Wiley & Sons, 1966, p. 167.
- LARSEN, W. A. y McCLEARY, S. J. "The Use of Partial Residual Plots in Regression Analysis." *Technometrics* 14 (1972): 781–790.
- LAVENBACH, H. y CLEARY, J. P. *The Professional Forecaster*. Belmont, Calif.: Lifetime Learning Publications, 1982.
- NETER, J. y WASSERMAN, W. *Applied Linear Statistical Models*. Homewood, Ill.: Richard D. Irwin, 1985.
- YOUNGER, M. S. *A Handbook for Linear Regression*. North Scituate, Mass.: Duxbury Press, 1979.

## Análisis de series de tiempo

### ACTIVIDADES

Los análisis de regresión y de correlación se refieren a la relación lineal entre dos o más variables. Se emplea el conocimiento de la variable independiente  $X$  para predecir la variable dependiente  $Y$ . En el análisis de series de tiempo la variable independiente es el tiempo. La variable bajo estudio ( $Y$ ) toma diferentes valores a través del tiempo. De ahí que cualquier variable clasificada en orden cronológico sea una serie histórica. Los periodos pueden ser años, trimestres, meses, semanas y, en algunos casos, días u horas. Las series de tiempo se analizan para descubrir patrones anteriores de variabilidad, que puedan emplearse para predecir valores futuros y asistir a las operaciones de administración de empresas. Los análisis de series de tiempo no dan la respuesta a lo que nos reserva el futuro, pero resulta valioso en el proceso de pronóstico y ayuda a reducir errores en ellos.

Las publicaciones como *Statistical Abstract of the United States*, *Survey of Current Business*, *Monthly Labor Review*, *Federal Reserve Bulletin* y los reportes anuales de corporaciones, contienen series de tiempo de todo tipo. Los datos, que regularmente se reportan en forma mensual, trimestral o anual, y abarcan precios, producción, ventas, empleo, desempleo, horas trabajadas, combustible empleado, energía producida, ingreso, etc., llenan las páginas de éstas y otras publicaciones económicas y de negocios.

Es importante que los administradores comprendan el pasado y empleen los datos históricos y el buen juicio para elaborar planes inteligentes que cubran la demanda del futuro. Los pronósticos se formulan para auxiliar a la administración en la determinación de estrategias alternativas.

Los pronósticos a largo plazo son por lo regular para 5, 10 e incluso 20 años a futuro. Estos pronósticos son esenciales para permitir el tiempo suficiente para que los departamentos de metas, manufactura, finanzas, ventas y otros de una compañía, desarrollen planes para probables plantas nuevas, financiamiento, desarrollo de nuevos productos y nuevos métodos de ensamble. Los pronósticos a largo plazo son de particular interés para la alta dirección.

Las organizaciones empresariales en EUA deben pronosticar el nivel de ventas tanto a corto como a largo plazos. La competencia por los dólares del consumidor, la presión

para obtener un beneficio para los accionistas y el deseo de abarcar una cada vez mayor participación en el mercado, son algunas de las principales fuerzas motivadoras en los negocios. De ahí que sea necesaria una declaración, denominada *pronósticos*, de las expectativas de la administración, para tener disponibles las materias primas, las instalaciones productivas y el equipo de personal para cubrir la demanda proyectada.

Desde luego, existe la alternativa de no planear a futuro. Sin embargo, en un medio empresarial dinámico, la falta de esta planeación puede resultar desastrosa. Una compañía que hubiese ignorado la tendencia hacia la televisión a color y los circuitos de estado-sólido, habría perdido con rapidez toda participación en el mercado.

En el análisis de series de tiempo, las consideraciones subjetivas son en extremo importantes, ya que a la fecha no se ha encontrado un enfoque satisfactorio de probabilidad para dicho análisis. Aunque las consideraciones subjetivas serían necesarias en la formulación de pronósticos, incluso si hubiera disponible un enfoque de probabilidad adaptable al análisis de series de tiempo. Siempre que se examina el pasado en busca de pistas con respecto al futuro, sólo es relevante hasta el grado en que las condiciones causales que estuvieron en efecto anteriormente, se mantengan constantes. La gran multitud de factores causales en el trabajo tiende a cambiar constantemente, así que la conexión entre pasado, presente y futuro se debe reevaluar continuamente.

Aunque las técnicas del análisis de series de tiempo no eliminan las evaluaciones subjetivas, sí aportan una útil contribución al proporcionar un enfoque conceptual a los pronósticos. Los pronósticos se elaboran con la ayuda de un conjunto de procedimientos formales y los juicios se indican de manera explícita.

## DESCOMPOSICIÓN

Un enfoque al análisis de series de tiempo comprende un intento de identificar los factores que ejercen influencia sobre cada uno de los valores periódicos de una serie. Este procedimiento de identificación se denomina *descomposición*. Cada componente se identifica por separado de tal manera que la serie histórica pueda proyectarse al futuro y utilizarse en pronósticos tanto de corto como de largo plazos.

Los cuatro componentes que se encuentran en una serie histórica se presentaron en el capítulo 4. Ellos son: tendencia, variaciones cíclicas, variaciones estacionales y fluctuaciones irregulares.

1. *Tendencia*. La tendencia es el componente de largo plazo que constituye la base del crecimiento (o declinación) de una serie histórica. Las fuerzas básicas que producen o afectan la tendencia de una serie son: cambios en la población, inflación, cambio tecnológico e incremento en la productividad.
2. *Variación cíclica*. El componente cíclico es un conjunto de fluctuaciones en forma de onda o ciclos, de más de un año de duración producidos por cambios en las condiciones económicas. Representa la diferencia entre los valores esperados de una variable (tendencia) y los valores reales —la variación residual que fluctúa alrededor de la tendencia.

3. *Variación estacional.* Las fluctuaciones estacionales se encuentran típicamente en los datos clasificados por trimestre, mes o semana. La variación estacional se refiere a un patrón de cambio regularmente recurrente a través del tiempo. El movimiento se completa dentro de la duración de un año y se repite a sí mismo año tras año.
4. *Fluctuación irregular.* El comportamiento irregular está compuesto por fluctuaciones causadas por sucesos impredecibles o no periódicos, como un clima poco usual, huelgas, guerras, rumores de guerras, elecciones y cambios en las leyes.

Para estudiar los componentes de una serie histórica, el analista debe considerar sus relaciones matemáticas. El enfoque que se utiliza con mayor frecuencia consiste en tratar los datos originales de una serie histórica como un producto de sus componentes; es decir, una serie anual es un producto de tendencia y fluctuaciones cíclicas, que se expresa simbólicamente como  $T \times C$ , como se muestra en la ecuación 8.1. En esta composición multiplicativa,  $T$  se mide en las unidades de los datos reales y  $C$  es el valor de un índice.

$$Y = TC \quad (8.1)$$

en donde

$Y$  = valor real

$T$  = tendencia

$C$  = componente cíclico

En una serie medida en periodos menores a un año, como las series de tiempo mensuales y trimestrales, se considera que cada valor original es el producto multiplicativo de cuatro componentes, como se muestra en la ecuación 8.2.

$$Y = TSCI \quad (8.2)$$

Más adelante, en este capítulo, se abordará la descomposición de series de tiempo mensuales y trimestrales.

## ÍNDICE DE PRECIOS

Diversas series sobre producción, ventas y otras situaciones económicas contienen datos disponibles sólo en valores de dólares. Estos datos son afectados tanto por la cantidad física de bienes vendidos, como por su precio. La inflación y la amplia variación de precios a través del tiempo pueden causar problemas en el análisis. Por ejemplo, un volumen de dólares incrementado podría ocultar una disminución en las ventas provocada por precios inflados. Por ello, con frecuencia es necesario conocer qué tanto del cambio de los valores en dólares representa un cambio real en la cantidad física y qué tanto se debe al cambio en los precios por la inflación. En estos casos, es deseable expresar los valores en dólares en términos de “dólares constantes”.



El concepto de *poder de compra* es importante. El poder de compra actual de 1 dólar se define como

$$\text{Poder de compra actual de 1 dólar} = \frac{100}{\text{índice de precios actual}} \quad (8.3)$$

Así, si en noviembre de 1994 el índice de precios al consumidor (con 1982 como 100), alcanza 150, el poder de compra actual del dólar del consumidor en noviembre de 1994 es

$$\text{Poder de compra actual de 1 dólar} = \frac{100}{150} = 0.67 \text{ o } .67 \text{ de dólar}$$

El dólar de 1994 compraba sólo dos tercios de los bienes y servicios que pudieron adquirirse con el dólar del periodo base (1982).

Para expresar los valores en dólares en términos de *dólares constantes*, se utiliza la siguiente ecuación:

$$\text{valor deflactado del dólar} = (\text{valor del dólar}) \times (\text{poder de compra de 1 dólar}) \quad (8.4)$$

Suponga que las ventas de automóviles se elevaron de 300,000 dólares en 1993 a 350,000 dólares en 1994, mientras que el índice de precios de un automóvil nuevo (con 1982 como año base) se elevó de 135 a 155. Las ventas deflactadas para 1993 y 1994 serían

$$\text{Ventas deflactadas en 1993} = (300,000 \text{ dólares}) \left( \frac{100}{135} \right) = 222,222 \text{ dólares}$$

$$\text{Ventas deflactadas en 1994} = (350,000 \text{ dólares}) \left( \frac{100}{155} \right) = 225,806 \text{ dólares}$$

Nótese que las ventas reales en dólares tuvieron un incremento dimensionable de 350,000 dólares – 300,000 dólares = 50,000 dólares. Sin embargo, las ventas deflactadas sólo se incrementaron en 225,806 dólares – 222,222 dólares = 3,584 dólares.

La finalidad de deflactar los valores en dólares consiste en eliminar el efecto de los cambios de precio. Este ajuste se denomina *deflación de precios* o bien, se le refiere como la expresión de una serie en dólares constantes.

La *deflación de precios* es el proceso mediante el cual se expresan términos de una serie en dólares constantes.

El proceso de deflación es relativamente simple. En él se utiliza un índice calculado a partir de los precios de bienes cuyos valores se van a deflactar. Por ejemplo, las ventas de una tienda de zapatos deben deflactarse mediante un índice de precios de zapatos, no por un índice general de precios. Para deflactar valores en dólares que representan a más de un tipo de bien, el analista debe desarrollar un índice de precios mediante la combinación de los índices de precios adecuados, en la proporción correcta.

**Ejemplo 8.1**

El Sr. Burnham desea estudiar el crecimiento a largo plazo de Burnham Furniture Store. La tendencia de largo plazo de su negocio debe evaluarse mediante el volumen físico de ventas. Si no se puede efectuar esta evaluación, los cambios de precios reflejados en las ventas en dólares no seguirán un patrón congruente y sólo oscurecerán el patrón real de crecimiento. Si se emplean las ventas en dólares, las ventas reales en dólares deben dividirse entre un índice de precios adecuado para obtener las ventas que se miden en dólares constantes.

El índice de precios al consumidor (IPC) no es adaptable al caso de Burnham, ya que contiene elementos tales como rentas, alimentos y servicios personales que no vende la empresa, aunque algunos componentes de este índice pudieran ser apropiados. Burnham está consciente de que el 70% de sus ventas corresponden a muebles y el 30% a aparatos. De ahí, él puede multiplicar el componente de muebles al menudeo del IPC por .70 y el componente de aparatos por .30, para obtener un índice combinado de precios. La tabla 8.1 ilustra este enfoque, en donde los cálculos para 1986 son

$$111.6(.70) + 105.3(.30) = 109.7$$

Las ventas se deflactan para 1986 en términos del poder de compra de 1982.

$$\text{Ventas deflactadas en 1986} = (42.1) \left( \frac{100}{109.7} \right) = 38.4$$

La tabla 8.1 muestra que mientras las ventas reales aumentaron congruentemente de 1986 a 1993, el volumen físico se mantuvo más bien estable. Es evidente que el incremento en las ventas se debió a elevación en los precios que, a su vez, fue generada por la tendencia inflacionaria de la economía.

**TABLA 8.1 DATOS PARA EL EJEMPLO 8.1: VENTAS DE BURNHAM FURNITURE, 1986–1993**

| AÑO  | VENTAS DE<br>BURNHAM<br>(EN MILES) | ÍNDICE DE PRECIOS                        |                                           | ÍNDICE DE<br>PRECIOS*<br>(1982 = 100) | VENTAS<br>DEFLACTADAS**<br>(MILES DE<br>DÓLARES<br>DE 1982) |
|------|------------------------------------|------------------------------------------|-------------------------------------------|---------------------------------------|-------------------------------------------------------------|
|      |                                    | DE MUEBLES<br>AL MENUDEO<br>(1982 = 100) | DE APARATOS<br>AL MENUDEO<br>(1982 = 100) |                                       |                                                             |
| 1986 | 42.1                               | 111.6                                    | 105.3                                     | 109.7                                 | 38.4                                                        |
| 1987 | 47.2                               | 117.2                                    | 108.5                                     | 114.6                                 | 41.2                                                        |
| 1988 | 48.4                               | 124.2                                    | 109.8                                     | 119.9                                 | 40.4                                                        |
| 1989 | 50.6                               | 128.3                                    | 114.1                                     | 124.0                                 | 40.8                                                        |
| 1990 | 55.2                               | 136.1                                    | 117.6                                     | 130.6                                 | 42.3                                                        |
| 1991 | 57.9                               | 139.8                                    | 122.4                                     | 134.6                                 | 43.0                                                        |
| 1992 | 59.8                               | 145.7                                    | 128.3                                     | 140.5                                 | 42.6                                                        |
| 1993 | 60.7                               | 156.2                                    | 131.2                                     | 148.7                                 | 40.8                                                        |

\* Elaborado para muebles (70%) y aparatos (30%).

\*\* Ventas divididas entre el índice de precios por 100.

**TENDENCIA**

Las tendencias son movimientos de largo plazo en una serie histórica que se pueden describir mediante una línea recta o curva. Las fuerzas básicas que producen o afectan la

tendencia de una serie son: cambios en la población, cambios de precios, cambios tecnológicos, incrementos en la productividad y ciclos de vida de los productos.

El incremento en la población puede provocar que las ventas al menudeo de una comunidad se eleven cada año durante varios años. Más aún, las ventas en dólares corrientes pudieran haber sido impulsadas hacia arriba durante el mismo periodo debido a incrementos generales en los precios de bienes al menudeo –aun cuando el volumen físico de bienes vendidos no haya variado.

El cambio tecnológico pudiera causar que una serie histórica se mueva hacia arriba o hacia abajo. El desarrollo y mejoramiento del automóvil, junto con las mejoras a los caminos, ha incrementado el registro de automóviles. Sin embargo, el automóvil, producido en volúmenes crecientes, provocó también una tendencia a la baja en la producción de vagones tirados por caballos y calesas.

Los incrementos en la productividad –los cuales, a su vez, se pueden deber al cambio tecnológico– incorporan una pendiente hacia arriba a muchas series de tiempo. Cualquier medición total de salida, como las ventas de los fabricantes, está afectada por cambios en la productividad.

Antes de medir la tendencia de una serie histórica, se debe primero conocer el propósito de su medición. Este conocimiento guía al analista en la elección del método y dimensión de la serie histórica a utilizar en la medición. Existen dos propósitos fundamentales: proyectar la tendencia y eliminarla de los datos originales.

En el análisis de la tendencia la variable independiente es el tiempo. El analista debe graficar los datos tanto sobre escalas aritméticas como semilogarítmicas antes de elegir el método de medición. Al hacerlo, el analista obtiene una guía adicional para escoger la ecuación de tendencia, ya que la imagen general de los datos es aparente. Si la gráfica de la serie indica un movimiento en línea recta en la escala aritmética, el analista ajustará a los datos una línea recta de tendencia. Si parece haber una tendencia no lineal, se debe desarrollar la curva de tendencia apropiada. Si los datos se grafican en papel semilogarítmico y una línea recta parece ajustar, se seleccionará un modelo exponencial que indique una tasa constante de crecimiento.

Sin embargo, no siempre una gráfica de los datos y una declaración de la finalidad de la medición permiten al analista efectuar la elección final. En ocasiones es necesario calcular y graficar dos o más tendencias junto con los datos originales para ver cuál de ellas se ajusta mejor a la serie histórica.

Qué es lo que constituye un “mejor” ajuste es un asunto de juicio. La tendencia que mejor se ajusta es aquella que hace el trabajo que el analista tiene que hacer. Ningún método es superior para medir tendencias. A veces, una tendencia dibujada a mano sobre una serie histórica es suficiente para revelar una imagen de la forma general y la dirección de la serie. No obstante, para dibujar a mano una tendencia con propiedad, el analista debe ser capaz de reconocer los ciclos principales y las fluctuaciones estacionales a través de las cuales debe pasar la tendencia. En ocasiones resulta difícil lograr esta apreciación, a menos que el analista esté muy familiarizado con la serie en particular que esté analizando. De ahí que muchos analistas seleccionan un método objetivo que pueda establecerse como una ecuación, con el fin de evitar las decisiones subjetivas que requiere el método manual.

El método más ampliamente utilizado para describir tendencias de línea recta se denomina *método de mínimos cuadrados*. Este enfoque calcula la línea que mejor se ajusta a un grupo de puntos, en conformidad matemática con un criterio establecido.<sup>1</sup> La ecuación de tendencia es

$$\hat{Y} = b_0 + bX \quad (8.5)$$

en donde

$\hat{Y}$  = valor de predicción de tendencia de la variable  $Y$ , codificada en periodos ( $X$ ).

$b_0$  = valor de la tendencia cuando  $x = 0$

$b$  = incremento o disminución promedio en  $\hat{Y}$  (tendencia), para cada incremento de un periodo de  $X$

$X$  = el valor del periodo seleccionado

### Ejemplo 8.2

En la tabla 8.2 se muestran los datos anuales de registros de automóviles nuevos de pasajeros en EUA de 1960 a 1992 y se grafican en la figura 8.1. Se emplean los valores de 1960 a 1992 para desarrollar una ecuación de tendencia. Los registros son la variable dependiente  $Y$  y el periodo o variable independiente se codifica como  $x$  (1960 = 1, 1961 = 2, etc.).

La  $r^2$  es .205, lo que significa que la variable tiempo explica sólo una pequeña cantidad, 20.5% de la varianza de la variable de registros de automóviles nuevos de pasajeros. El valor estimado de la tendencia para registros de automóviles de pasajeros para 1960 es de 8,106.7, como se muestra en la columna de regresión de la tabla 8.2. Con base en la pendiente de la ecuación de tendencia, se espera que los registros se incrementen

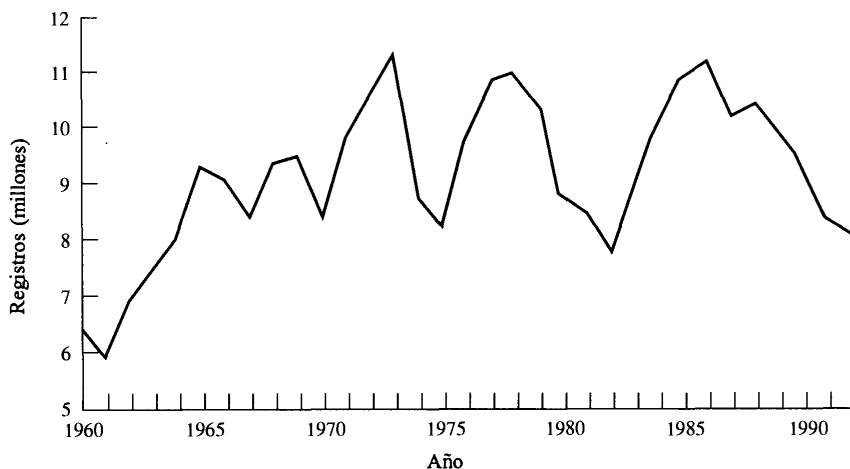


**Figura 8.1** Datos de registros del ejemplo 8.2.

<sup>1</sup> Como se expuso en el capítulo 6, el método de mínimos cuadrados minimiza la suma de las distancias al cuadrado, medidas en dirección vertical, desde los puntos de datos hacia la línea de la tendencia.

**TABLA 8.2** REGISTROS DE AUTOMÓVILES NUEVOS DE PASAJEROS EN ESTADOS UNIDOS, 1960–1992

| AÑO  | REGISTROS<br>(EN MILLONES) |     | $\hat{Y}$ | COMPONENTE<br>CÍCLICO |
|------|----------------------------|-----|-----------|-----------------------|
|      | $Y$                        | $X$ |           |                       |
| 1960 | 6.577                      | 1   | 8.1067    | 81.13                 |
| 1961 | 5.855                      | 2   | 8.1704    | 71.66                 |
| 1962 | 6.939                      | 3   | 8.2341    | 84.27                 |
| 1963 | 7.557                      | 4   | 8.2977    | 91.07                 |
| 1964 | 8.065                      | 5   | 8.3614    | 96.45                 |
| 1965 | 9.314                      | 6   | 8.4251    | 110.55                |
| 1966 | 9.009                      | 7   | 8.4888    | 106.13                |
| 1967 | 8.357                      | 8   | 8.5525    | 97.71                 |
| 1968 | 9.404                      | 9   | 8.6162    | 109.14                |
| 1969 | 9.447                      | 10  | 8.6798    | 108.84                |
| 1970 | 8.388                      | 11  | 8.7435    | 95.93                 |
| 1971 | 9.831                      | 12  | 8.8072    | 111.62                |
| 1972 | 10.409                     | 13  | 8.8709    | 117.34                |
| 1973 | 11.351                     | 14  | 8.9346    | 127.05                |
| 1974 | 8.701                      | 15  | 8.9982    | 96.70                 |
| 1975 | 8.168                      | 16  | 9.0619    | 90.14                 |
| 1976 | 9.752                      | 17  | 9.1256    | 106.86                |
| 1977 | 10.826                     | 18  | 9.1893    | 117.81                |
| 1978 | 10.946                     | 19  | 9.2530    | 118.30                |
| 1979 | 10.357                     | 20  | 9.3167    | 111.17                |
| 1980 | 8.761                      | 21  | 9.3803    | 93.40                 |
| 1981 | 8.444                      | 22  | 9.4440    | 89.41                 |
| 1982 | 7.754                      | 23  | 9.5077    | 81.55                 |
| 1983 | 8.924                      | 24  | 9.5714    | 93.24                 |
| 1984 | 10.118                     | 25  | 9.6351    | 105.01                |
| 1985 | 10.889                     | 26  | 9.6988    | 112.27                |
| 1986 | 11.140                     | 27  | 9.7624    | 114.11                |
| 1987 | 10.183                     | 28  | 9.8261    | 103.63                |
| 1988 | 10.398                     | 29  | 9.8898    | 105.14                |
| 1989 | 9.833                      | 30  | 9.9535    | 98.79                 |
| 1990 | 9.160                      | 31  | 10.0172   | 91.44                 |
| 1991 | 9.234                      | 32  | 10.0808   | 81.68                 |
| 1992 | 8.054                      | 33  | 10.1445   | 79.39                 |

Fuente: Datos del Departamento de Comercio, *Survey of Current Business* (varios años).

en un promedio de 63,680 cada año. La figura 8.2 muestra la ecuación de tendencia ajustada a los datos reales. La ecuación de tendencia proyecta que los registros en 1992 serán de 10,144,530.

$$\hat{Y} = 8.043 + .06368X$$

$$\hat{Y}(1992) = 8.043 + .06368(33) = 10.144$$

En realidad, los registros de automóviles nuevos de pasajeros en 1992 fueron 8,054,000. Es evidente que al pronosticar registros, debe considerarse algún otro factor además de la tendencia.

Los comandos de Minitab para resolver el ejemplo 8.1 son

```
MTB > READ 'REG92.DAT' C1
 33 ROWS READ
```

```
C1
 6.577 5.855 6.939 7.557 . . .
```

```
MTB > PRINT C1
```

```
C1
 6.577 5.855 6.939 7.557 8.065 9.314 9.009 8.357
 9.404 9.447 8.388 9.831 10.409 11.351 8.701 8.168
 9.752 10.826 10.946 10.357 8.761 8.444 7.754 8.924
 10.118 10.889 11.140 10.183 10.398 9.833 9.160 8.234
 8.054
```

```
MTB > SET C2
```

```
DATA> 1:33
```

```
DATA> END
```

```
MTB > NAME C1 'REG' C2 'TIME'
```

```
MTB > REGRESS C1 1 PREDICTOR C2, TRESRES C3, DHATS C4;
```

```
SUBC> PREDICT 34.
```

The regression equation is

REG = 8.04 + 0.0637 TIME

| Predictor | Coef    | Stdev   | t-ratio | P     |
|-----------|---------|---------|---------|-------|
| Constant  | 8.0430  | 0.4384  | 18.35   | 0.000 |
| TIME      | 0.06368 | 0.02250 | 2.83    | 0.008 |

s = 1.231      R-sq = 20.5%      R-sq(adj) = 18.0%

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F    | P     |
|------------|----|--------|--------|------|-------|
| Regression | 1  | 12.134 | 12.134 | 8.01 | 0.008 |
| Error      | 31 | 46.952 | 1.515  |      |       |
| Total      | 32 | 59.085 |        |      |       |

| Fit    | Stdev.Fit | 95% C.I.         | 95% P.I.         |
|--------|-----------|------------------|------------------|
| 10.208 | 0.438     | ( 9.314, 11.103) | ( 7.543, 12.873) |

```
MTB > STOP
```

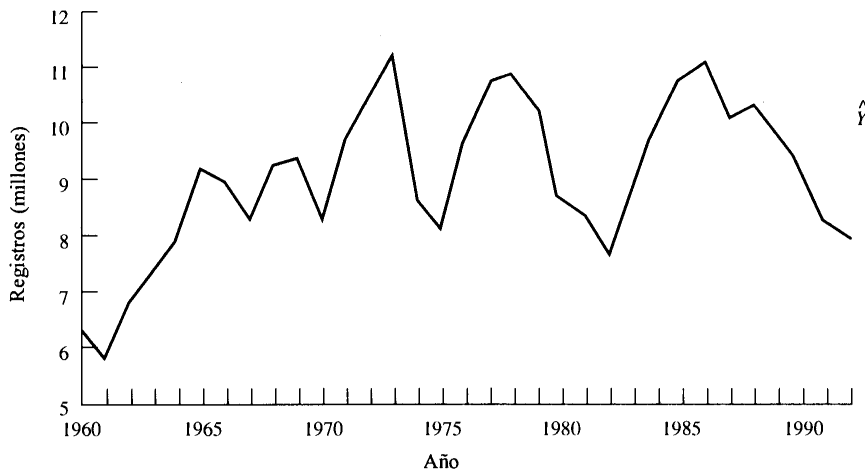


**Figura 8.2** Ecuación de tendencia para los datos del ejemplo 8.2.

- El comando READ se usa para leer un archivo que contiene los datos de registros que se presentan en la tabla 8.2.
- El comando SET se emplea para generar enteros del 1 al 33 y almacenarlos en C2.
- El comando NAME se utiliza para dar nombre a las variables en cada columna.
- El comando REGRESS se usa para desarrollar un análisis de regresión con los registros como variable dependiente ( $Y$ ) y el tiempo como variable independiente ( $X$ ).
- El comando DHATS calcula las estimaciones de  $Y$  y las almacena en C4.
- El subcomando PREDICT se usa para calcular un intervalo de confianza del 95%, en este caso para un valor de  $x$  de 34.

Aunque es probable que la tendencia de mínimos cuadrados se utilice con mayor frecuencia que cualquier otra para describir el crecimiento a largo plazo de una serie histórica, en ocasiones es necesario utilizar tendencias curvas para obtener una descripción lógica del cambio. Existen disponibles muchas ecuaciones para calcular tendencias de línea curva. Algunas de las ecuaciones de tendencia más útiles comprenden la curva exponencial y la curva de crecimiento de Gompertz. A continuación se presenta una breve descripción de estos métodos.

### **Tendencia no lineal**

La figura 8.3 muestra que el ciclo de vida de un típico producto nuevo se divide en tres estados principales: introducción, crecimiento, y por último madurez y saturación. El

tiempo, representado en el eje horizontal, puede variar de días a años, dependiendo de la naturaleza del mercado. Un modelo lineal no funcionaría con este tipo de datos. Los modelos lineales asumen que una variable se incrementa en una cantidad constante en cada periodo. La etapa de crecimiento de la mayoría de los productos nuevos parece incrementarse en una cantidad o proporción constante.

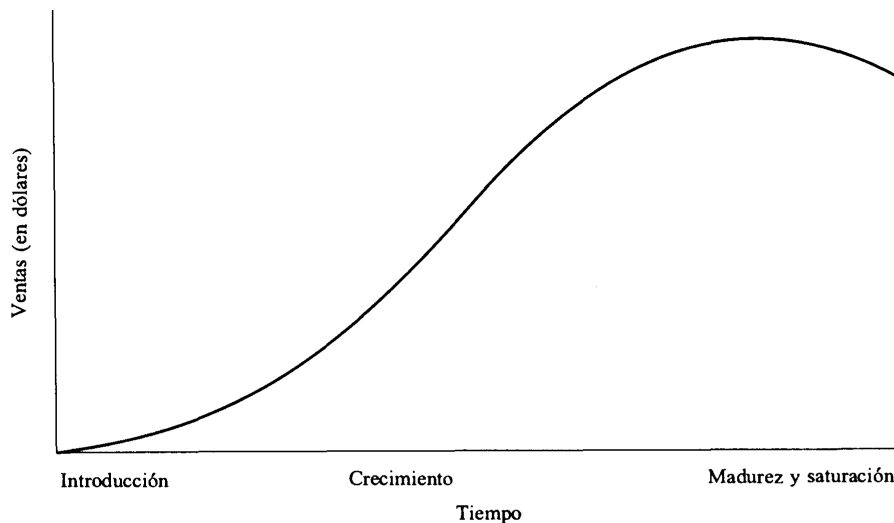


**Figura 8.3** Ciclo de vida de un producto nuevo típico.

Cuando una serie histórica comienza lentamente y después parece incrementarse en una proporción creciente tal (véase figura 8.4), que la diferencia porcentual entre una observación y otra es constante, se puede ajustar a una tendencia exponencial. En la figura 8.4, el número de vendedores en una mutualidad parece crecer a una tasa constante en vez de a una cantidad constante. Una tendencia lineal pudiera indicar un crecimiento promedio de 8 vendedores por año. Un ajuste exponencial podría indicar un crecimiento promedio de 32% anual para los datos de la figura 8.4. Si el modelo exponencial estimó 50 vendedores en la mutualidad para 1992, el incremento estimado para 1993 sería de 16 ( $50 \times .32$ ) en lugar de 8.

Si el analista no es cuidadoso, el modelo exponencial puede conducir a pronósticos muy imprecisos. Nótese en la figura 8.4 que el modelo exponencial siempre pronosticaría una cantidad creciente del número de vendedores en la mutualidad. ¿Qué pasa cuando tiene lugar la maduración y saturación del mercado y la tasa de crecimiento disminuye? El pronóstico será mayor al valor real de la variable. Otro problema se refiere al uso de mínimos cuadrados ordinarios para ajustar el modelo exponencial. Se ha argumentado que



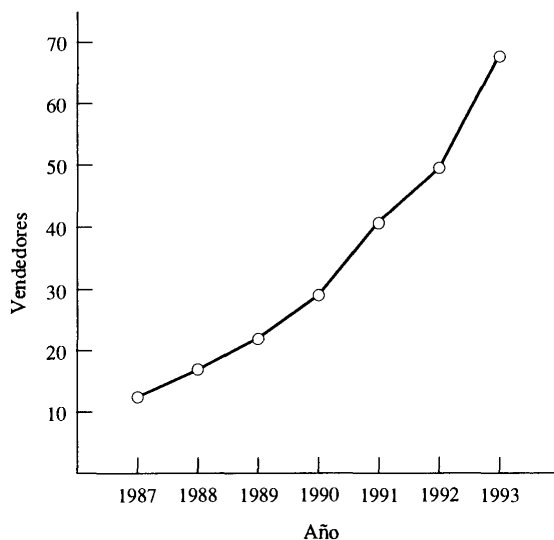


**Figura 8.4** Gráfica del fondo mutualista de vendedores.

este tipo de pronósticos estará desviado a menos que se incluya en la fórmula una corrección de varianza.

Las curvas de crecimiento de Gompertz o de tipo logístico representan la tendencia de muchas industrias y líneas de productos a crecer a una tasa de declinación cuando maduran. La curva de Gompertz pudiera ser apropiada si la gráfica de datos refleja una situación en la que las ventas comienzan bajas, luego el producto tiene su auge de ventas y, por último, las ventas se desvanecen al alcanzar la saturación. La figura 8.5 muestra la

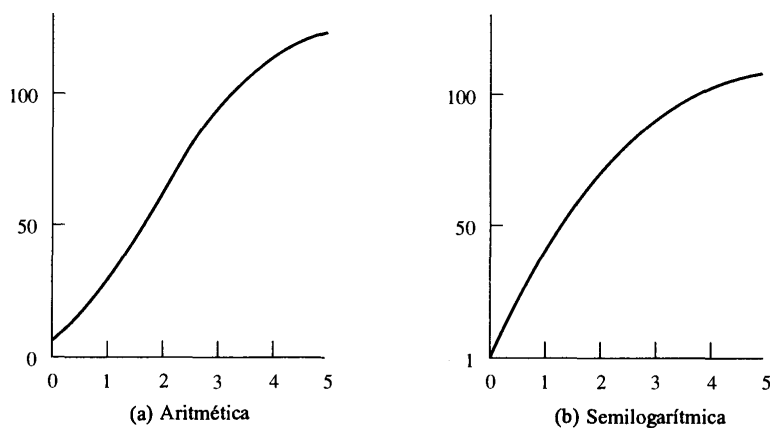


**Figura 8.5** Curvas de crecimiento de Gompertz.

forma general de las curvas de Gompertz en gráficas tanto aritmética como semilogarítmica de ventas. Los cálculos para la curva de Gompertz son en extremo complejos y no están comprendidos en el alcance de este texto. Para los procedimientos de cálculo, deberán consultarse libros especializados.<sup>2</sup>

En la decisión de qué tendencia utilizar, se debe conocer el propósito de calcular la tendencia. Por ejemplo, si el propósito consiste en obtener una estimación de gastos para un año futuro, se requiere un conocimiento de las fuerzas básicas que producen o afectan la tendencia. La correcta elección de una tendencia es un aspecto de juicio, de ahí que requiera de experiencia y sentido común por parte del analista. La línea o curva que mejor se ajusta a los puntos de datos pudiera no tener sentido al proyectarla como la tendencia del futuro.

## VARIACIÓN CÍCLICA

El análisis de la tendencia de la variable dependiente tiene un valor práctico directo para los pronósticos a largo plazo. Sin embargo, el análisis del componente cíclico es de un valor dudoso en el pronóstico. El componente cíclico es la fluctuación en forma de onda o ciclo de más de ocho meses de duración, debida a condiciones económicas cambiantes. Los economistas le han prestado mucha atención al análisis de ciclos en los negocios y a sus causas, pero no está comprendido en el alcance de este texto considerar las numerosas teorías que se refieren a este problema.

El componente cíclico de las series de tiempo se identifica eliminando o promediando los efectos de la tendencia. Ya que este componente constituye lo que queda después de dichos ajustes, se le refiere como el *método residual*.

Los pasos específicos comprendidos en el método residual dependen de si el análisis comienza con series de datos mensuales, trimestrales o anuales. Si los datos son mensuales o trimestrales, entonces los efectos tanto de la tendencia como de los componentes estacionales se deben eliminar. Si los datos son anuales, entonces sólo deben eliminarse los efectos del componente de la tendencia.

En forma simbólica, la descomposición de una serie histórica anual se representa como

$$C = \frac{Y}{T} = \frac{TC}{T} \quad (8.6)$$

En la determinación del efecto relativo del componente cíclico en cada valor anual, se acepta al valor de  $\hat{Y}$  como una indicación precisa de la tendencia y se trata a la discrepancia (residuo) como el componente cíclico.

<sup>2</sup> Para una descripción de los métodos matemáticos para ajustar curvas de crecimiento, véase F. E. Croxton, D. J. Cowden y B. W. Bolch, *Practical Business Statistics*, 14. ed. (Englewood Cliffs, N. J.: Prentice-Hall, 1969), pp. 332-338, o S. Makridakis y S. C. Wheelwright, *Interactive Forecasting*, Vol. 1 (Palo Alto, Calif.: Scientific Press, 1977), pp. 100-104.

**Ejemplo 8.3**

En la tabla 8.1 se presentan los índices cíclicos para el registro anual de nuevos automóviles de pasajeros en EUA de 1960 a 1992. Cada índice cíclico se calcula dividiendo el número real de registros para cada año ( $Y$ ) entre el número esperado de registros ( $\hat{Y}$ ). La ecuación de tendencia,  $\hat{Y} = 8.043 + .06368X$  sirve para calcular el valor esperado de la tendencia. Esta proporción se multiplica por 100 para convertir el componente cíclico a porcentaje o índice.

Por ejemplo, el valor cíclico para 1960 en la tabla 8.1 se calcula (ignorando errores de redondeo), como sigue. La estimación de la tendencia para 1960 es

$$\hat{Y} = 8.043 + .06368(1) = 8.10668$$

El valor cíclico para 1960 es

$$C = \frac{6.577}{8.10668} (100) = 81.13$$

El índice cíclico muestra la posición de cada valor de  $Y$  en relación con la línea de tendencia. Esta posición se refleja como porcentaje. Por ejemplo, en 1960, el valor de  $Y$  era 81.13% de la línea de tendencia. En 1973,  $Y$  era el 127.05% de la línea de tendencia. Como los valores cíclicos se muestran como porcentajes de la línea de tendencia, se puede decir que se ha eliminado la tendencia de la serie, dejando sólo el componente cíclico para evaluación. En 1960, los registros de automóviles nuevos de pasajeros fueron aproximadamente 19% menores de lo esperado con base en la estimación de la tendencia. En 1973, los registros de automóviles nuevos de pasajeros fueron aproximadamente 27% superiores a lo esperado con base en la estimación de la tendencia.

The MINITAB commands to solve Example 8.3 are

```
MTB > DIVIDE C1 BY C4, PUT INTO C5
MTB > LET C6 = C5*100
MTB > PRINT C6
```

C6

|         |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|
| 81.131  | 71.661  | 84.272  | 91.073  | 96.455  | 110.551 | 106.128 |
| 97.715  | 109.144 | 108.839 | 95.934  | 111.625 | 117.339 | 127.046 |
| 96.697  | 90.135  | 106.864 | 117.811 | 118.297 | 111.167 | 93.398  |
| 89.411  | 81.555  | 93.236  | 105.012 | 112.272 | 114.111 | 103.632 |
| 105.139 | 98.790  | 91.443  | 81.680  | 79.393  |         |         |

```
MTB > STOP
```

- Los índices cíclicos se calculan dividiendo los valores reales de  $Y$  almacenados en C1, entre los valores estimados de  $Y$  almacenados en C4. El resultado, almacenado en C5, se multiplica por 100 para crear un número índice, que se almacena en C6.
- El comando PRINT lista el contenido de C6 (índice cíclico).

La figura 8.6 muestra una gráfica cíclica que se desarrolló para ayudar a analizar el componente cíclico. Nótese que la ecuación de tendencia se representa como 100% o como la línea base. Resulta más fácil apreciar diversos patrones cíclicos en una gráfica de este

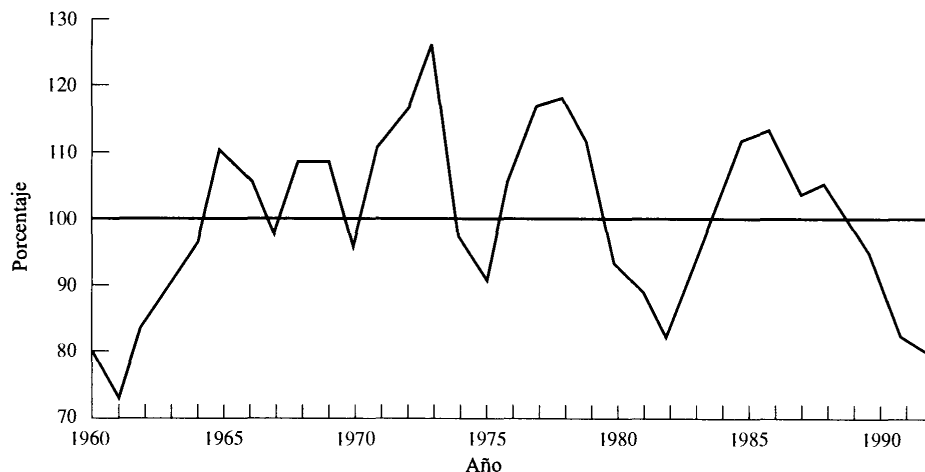


**Figura 8.6** Gráfica cíclica para los datos del ejemplo 8.3.

tipo; la cual sirve también para comparar el componente cíclico de la variable de interés con el patrón cíclico de otras variables y/o indicadores de los negocios.

Los índices relativos responden las siguientes preguntas para cualquier serie histórica:

1. ¿Presenta ciclos la serie?
2. De ser así, ¿qué tan extremo es el ciclo?
3. ¿Sigue la serie el estado general de la economía (ciclo de los negocios)?

#### **Ejemplo 8.4**

Se puede emplear la figura 8.6 para responder preguntas con respecto al patrón cíclico de registros de automóviles nuevos de pasajeros. Los registros fueron débiles (abajo de 100) de 1960 a 1974. Luego, el patrón cíclico osciló hacia arriba y hacia abajo alrededor de la tendencia hasta 1970 cuando se incrementó en forma consistente hasta alcanzar el punto más alto de 127 en 1973. A esto siguió una caída dramática a 90 en 1975. El patrón alcanzó un pico de nuevo en 1978 de 118, pero luego cayó a uno bajo de 81 en 1982. En 1986 llegó otro pico, 114, seguido de una declinación de seis años hasta llegar a un punto bajo de 79 en 1992.

Los indicadores cíclicos muestran que hay un efecto cíclico definido para el registro de automóviles nuevos: Los indicadores varían de 71.66 a 127.05. En contraste, se podría esperar que indicadores cíclicos para leche o jabón para ropa fueran muy leves o inexistentes, mientras que otros como los de abrigos de pieles y joyería fueran extremos. El patrón cíclico de registro de automóviles nuevos parece estar influido en cierto grado por las condiciones económicas. Los picos se presentan cuando la economía es fuerte y las caídas cuando es débil. Sin embargo, parece haber otros factores que afectan el patrón cíclico de la variable.

Una forma de investigar los patrones cíclicos es a través del estudio de indicadores empresariales. Un *indicador empresarial* es una serie histórica relacionada con los negocios que se usa para ayudar a evaluar el estado general de la economía, en particular con referencia al ciclo de los negocios. Mucha gente de negocios y economistas siguen sistemáticamente el movimiento de tales series estadísticas para obtener su información en

la forma de una imagen que se desdobra, que esté actualizada, que sea comprensible, relativamente objetiva y capaz de ser leída y entendida en el menor tiempo posible.

Los *indicadores empresariales* son series de tiempo relacionadas con los negocios que ayudan a evaluar el estado general de la economía.

La lista más importante de indicadores estadísticos se originó durante la aguda recesión en los negocios de 1937–1938. El secretario del tesoro, Henry Morgenthau, solicitó al Departamento Nacional de Investigación Económica (NBER: National Bureau of Economic Research) que concibiera un sistema que señalara cuándo la recesión se acercaría a su final. Bajo la dirección de Wesley Mitchell y Arthur F. Burns, los economistas del NBER seleccionaron 21 series que de acuerdo con su desempeño en el pasado prometían ser indicadores relativamente confiables de la recuperación económica. Debido a que el departamento ha revisado la lista en diversas ocasiones, en la tabla 8.3 aparece una lista reciente que consta de 22 indicadores.

**TABLA 8.3 INDICADORES CÍCLICOS EMPRESARIALES**

| INDICADORES CONDUCENTES                                                                                  | INDICADORES COINCIDENTES                                         | INDICADORES RETRASADOS                                                                     |
|----------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| Horas semanales promedio en manufacturas                                                                 | Empleados en nóminas no agrícolas, en miles                      | Duración promedio del desempleo, en semanas                                                |
| Promedio semanal de solicitudes iniciales de seguro de desempleo, miles                                  | Ingreso personal menos pagos de transferencia, miles de millones | Proporción entre inventarios de manufactura y comercio y ventas en 1987                    |
| Pedidos nuevos de fabricantes, bienes de consumo y materiales, miles de millones                         | Índice de producción industrial, 1987 = 100                      | Promedio de tasa líder cargada por los bancos, porcentaje                                  |
| Cambio en el costo laboral por unidad de producción, manufactura, porcentaje                             | Venta de manufacturas y comercio, en millones                    | Créditos comerciales e industriales pendientes, millones                                   |
| Desempeño de distribuidores, índice de difusión más lenta de entregas, porcentaje                        |                                                                  | Proporción del crédito a plazos al consumidor con respecto al ingreso personal, porcentaje |
| Contratos y pedidos para planta y equipo, miles de millones                                              |                                                                  | Cambio en el Índice de Precios al Consumidor para servicios, porcentaje                    |
| Índice de unidades de casas privadas nuevas autorizadas por permisos de construcción locales, 1967 = 100 |                                                                  |                                                                                            |
| Cambio en las órdenes no cubiertas por fabricantes, bienes duraderos, miles de millones                  |                                                                  |                                                                                            |
| Cambio en precios de materiales sensibles, porcentaje                                                    |                                                                  |                                                                                            |
| Índice de precios de acciones, 500 acciones comunes, 1941–1943 = 100                                     |                                                                  |                                                                                            |
| Oferta de dinero M2, miles de millones                                                                   |                                                                  |                                                                                            |
| Índice de expectativas del consumidor, U. de Michigan, 1961:1 = 100                                      |                                                                  |                                                                                            |

El NBER ha encontrado que ciertas series de tiempo pueden ser útiles como indicadores directos de expansiones y contracciones cíclicas en la actividad empresarial. Después de estudios continuos de ciclos en los negocios, se ha aprendido que no todas las áreas de la economía se expanden en forma simultánea durante las expansiones, ni todas se contraen en forma concurrente durante los periodos de contracción. Un estudio de series de tiempo económicas individuales y su relación con este movimiento cíclico general indica que la periodicidad de la mayoría de las series se apega sólo en una pequeña medida –y en ciertas series en ninguna– a la periodicidad del ciclo de los negocios. Sin embargo, existen excepciones a esta aseveración genérica.

El departamento identificó 22 indicadores estadísticos –once de ellos clasificados como conducentes, 4 como coincidentes y 7 como retrasados– que han demostrado su utilidad como indicadores de condiciones de los negocios. Mensualmente se publican cifras actualizadas de estas series en *Business Conditional Digest*, del Departamento de Comercio de los Estados Unidos y en *Survey of Current Business*; y semanalmente en *Statistical Indicator Report*, de Statistical Indicator Associates.

Para guía de pronosticadores y otros usuarios, cada serie en la lista del NBER ha sido evaluada y calificada por el departamento en términos de los siguientes elementos: (1) grado hasta el cual la serie mide o representa una actividad que tiene un papel clave en el proceso cíclico, (2) precisión estadística y dependencia de las series, (3) grado de apego de las series a los ciclos históricos de los negocios, (4) congruencia con la que la serie se ha movido en cierta fase específica (conducente, coincidente o retrasada), con el movimiento cíclico en la economía, (5) grado de atenuación (v. gr., ausencia relativa de fluctuaciones mayores irregulares) y, (6) oportunidad de publicación.<sup>3</sup> Las series seleccionadas para la lista corta (tabla 8.3) tienen todas altas calificaciones y comprenden poca duplicidad de cobertura.

Como se mencionó, se identificaron tres grupos de indicadores: los que proporcionan advertencias adelantadas de cambios probables en la actividad económica, los indicadores conducentes; aquellos que reflejan el desempeño actual de la economía, los indicadores coincidentes; y aquellos que confirman los cambios previamente señalados, los indicadores retrasados.

1. *Indicadores conducentes.* En la práctica, se estudian los componentes de las series conducentes para ayudar a anticipar los momentos cruciales en la economía. La *Survey of Current Business* publica esta lista cada mes, junto con los valores reales de cada serie para varios meses anteriores y el año más reciente. También, se calcula, para cada mes y año, un índice compuesto de indicadores conducentes; con frecuencia se reporta el valor mensual más reciente en la prensa popular para indicar la dirección general de la economía futura. En julio de 1993 este índice compuesto fue de 151.5 (1982 = 100), que subió con respecto a junio de 1992 en el que fue de 148.8. Ejemplos de indicadores conducentes son los permisos de construcción de casas nuevas y el mercado de valores.

<sup>3</sup> G. H. Moore y J. Shiskin, *Indicators of Business Expansions and Contractions* (New York: National Bureau of Economic Research, 1957), pp. 8-28.

2. *Indicadores coincidentes*. Los cuatro indicadores coincidentes de la tabla 8.3 proporcionan una medida del desempeño actual de la economía de EUA. Cada mes se calcula un índice de las cuatro series, el cual se ubicó en 123.3 en junio de 1992 y subió a 126.9 en julio de 1993. Ejemplos de estos indicadores son el ingreso personal y las ventas.
3. *Indicadores retrasados*. Estos indicadores tienden a ubicarse en forma retrasada detrás del estado general de la economía, tanto en las alzas como en las bajas. También para esta lista se calcula un índice compuesto, que fue de 105.4 en junio de 1992 y bajó a 104.2 en julio de 1993. Como ejemplos se pueden citar la tasa líder de interés y los préstamos comerciales pendientes.

Una de las mayores dificultades de un indicador consiste en determinar cuándo alcanza un punto cíclico crítico. El componente de fluctuación de las series económica y de negocios contiene movimientos irregulares de corto plazo, adicionales a los cíclicos. De ahí que las gráficas de datos semanales, mensuales o trimestrales tengan la apariencia desigual de una sierra, como se muestra en la figura 8.6. En consecuencia, resulta difícil identificar los puntos cíclicos críticos cerca del momento en el que ocurren. Recuerde que los puntos críticos existen sólo como consecuencia de una declinación seguida de un aumento, o viceversa, en el componente cíclico. Por ello, pueden pasar varios meses antes de identificar, con cierta seguridad, un auténtico giro cíclico hacia arriba o hacia abajo en una serie conducente.

Otro problema que surge en el uso del enfoque del indicador cíclico se debe a que no presenta uniformidad en el lapso mediante el cual un indicador conducente dado, antecede a cambios cíclicos en la economía. En vez de ello, se puede apreciar una variabilidad considerable en el periodo conducente de ciclo a ciclo. En consecuencia los indicadores conducentes pudieran señalar que puede esperarse una recesión o recuperación en algún momento futuro, pero proporciona poca ayuda para establecer el momento del cambio. Para establecer esta conexión, con frecuencia resultan útiles los análisis de los indicadores coincidentes y retrasados. Por ejemplo, si los indicadores conducentes señalan un giro cíclico hacia abajo, los analistas comenzarán a buscar con detenimiento signos de debilidad en los indicadores coincidentes. De ahí que en la práctica, el uso efectivo de los indicadores cíclicos requiera de una continua evaluación de las series de los tres grupos.

Otra dificultad más consiste en que, en ocasiones, los indicadores emiten “señales falsas”, es decir, señalan un punto crítico que no se materializa. Algunas veces, estas señales se deben a factores que se pueden identificar por adelantado, como una gran huelga, de modo que se pueden descartar dichas señales. En otros casos, ha sido imposible distinguir entre las señales falsas y las precisas hasta después de consumado el hecho.

En resumen, los indicadores empresariales han demostrado ser de gran utilidad en el pronóstico de puntos críticos, en especial en donde la mayoría de los otros métodos tiende a fallar. Los indicadores tienen sus limitaciones, sin embargo, se han desempeñado tan bien como cualquier otro método, a través del tiempo.

En su artículo titulado “Signos tempranos de advertencia para la economía”, Geoffrey H. Moore y Julius Shiskin dicen lo siguiente con respecto a la utilidad de los indicadores estadísticos:<sup>4</sup>

<sup>4</sup> G. H. Moore y J. Shiskin, “Early Warning Signals for the Economy”, en *Statistics: A guide to Business and Economics* (San Francisco: Holden-Day, 1976) p. 81.

A partir de los datos registrados, parece ser claro que los indicadores de ciclo en los negocios resultan útiles al juzgar el tono de los negocios actuales y los prospectos de corto plazo. Aunque debido a sus limitaciones, los indicadores deben usarse junto con otros datos y con plena conciencia de la confianza y expectativas de los negocios y el consumidor, las políticas gubernamentales y los sucesos internacionales. Debemos también anticipar que con frecuencia los indicadores no serán fáciles de interpretar, que las interpretaciones algunas veces varían entre los analistas y que las señales que ofrecen no serán correctamente interpretadas.

Los indicadores proporcionan una imagen sensible y reveladora del flujo y reflujo de las mareas económicas, que un avezado analista de la escena económica, política e internacional, puede utilizar para mejorar sus posibilidades de formular un pronóstico válido de las tendencias económicas de corto plazo. Si el analista está consciente de sus limitaciones y está atento al mundo que lo rodea, encontrará en los indicadores útiles señales que le servirán de guía para evaluar la economía y sus necesidades.

Como un ejemplo de análisis y predicción, se presenta el patrón cíclico de registro de automóviles nuevos.

### **Ejemplo 8.5**

Para explicar el patrón cíclico de la variable: registro de automóviles nuevos, que se muestra en la figura 8.6, se deben investigar diversos factores. Si la variable de registro de automóviles de pasajeros está relacionada de manera lógica con la economía, en alguna medida los indicadores coincidentes que proporcionan la medición del desempeño actual de la actividad económica, debieran proporcionar ayuda en el pronóstico. Muchas series de indicadores las pronostican expertos y es frecuente que estos pronósticos estén disponibles en fuentes publicadas. Si existen estos pronósticos y tienen un buen registro de confiabilidad, pudieran proporcionar una base para el pronóstico de registro de automóviles nuevos.

La figura 8.6 muestra que la variable de registro de automóviles está relacionada con la economía. Cuando en 1961, 1970, 1975, 1982, 1991 y 1992, descendieron los registros de automóviles, la economía estuvo deprimida. Cuando los registros alcanzaron picos en 1965, 1968, 1973, 1978 y 1986, la economía era saludable.

Una variable que podría explicar por lo menos parte de esta fluctuación cíclica es el índice de precios de automóviles nuevos, que genera el gobierno federal y se publica en *Survey of Current Business*. Presumiblemente, cuando el precio es alto, los registros debieran ser bajos y cuando los precios son bajos, los registros debieran aumentar. El índice de precios de automóviles nuevos se ajustó a la tendencia para eliminar los efectos de la inflación y se graficaron los ciclos resultantes en la figura 8.7. En ella se muestran las gráficas cíclicas tanto para registros de automóviles nuevos como para el índice de precios de automóviles nuevos (periodo base 1982–1984).

La figura 8.7 revela una relación entre el registro de automóviles nuevos y el índice de precios de automóviles nuevos. Sin embargo, no parece haber una conexión entre ambos índices cíclicos. Debe haber otro factor, distinto a los precios de los automóviles, que esté provocando que los registros de automóviles exhiban el efecto cíclico que se muestra en las figuras 8.6 y 8.7. Un analista interesado en explicar la naturaleza cíclica de los registros de automóviles nuevos, podría buscar variables económicas adicionales y graficarlas contra los registros, esperando encontrar una o más variables con una fuerte relación.



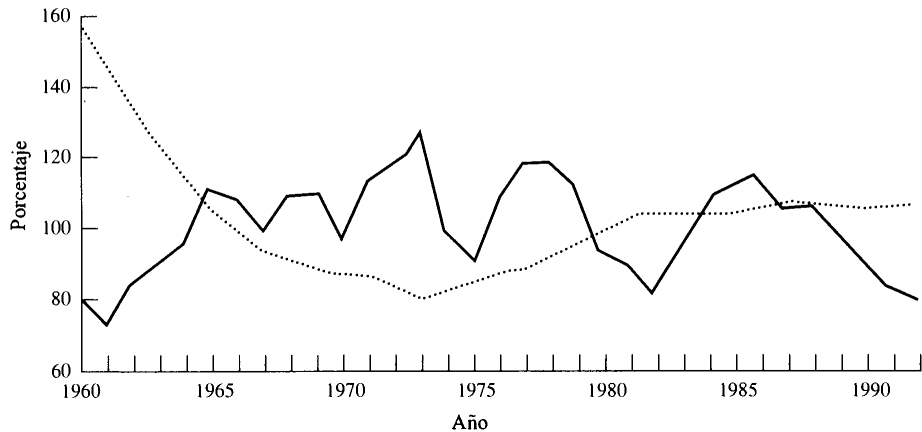


**Figura 8.7** Gráfica cíclica de registros de automóviles e índice de precios de automóviles nuevos.

Hasta ahora, la decisión muestra que se pueden separar y estudiar en forma individual los factores que generan una variación en una serie histórica. El proceso de separar la serie histórica se denomina *análisis*; el proceso de reunirla nuevamente es la *síntesis*.

Para comprender la variable de registro de automóviles nuevos, se analizan los componentes individuales. La descomposición del año 1992 se lleva a cabo mediante la determinación de los componentes de tendencia y cíclico (véase tabla 8.2).

$$T = \text{tendencia (1992)} = 10.14453$$

$$C = \text{cíclico (1992)} = 79.39$$

Para unir nuevamente la serie, el analista emplea el modelo multiplicativo original y sustituye los componentes correspondientes.

$$Y = TC$$

$$Y(1992) = 10.14453(.7939) = 8.054$$

También se puede utilizar el modelo multiplicativo para estimar los registros de automóviles nuevos para 1993. Todo lo que necesita el analista son estimaciones precisas de los componentes de tendencia y cíclico irregular.

La estimación de la tendencia se determina mediante la ecuación apropiada de tendencia, en este caso  $\hat{Y} = 8.043 + .06368X$ . Puesto que  $X = 1$  representa al 1o. de julio de 1960, el valor codificado de  $X$  para 1993 es 34 y la estimación de la tendencia es

$$\hat{Y} = 8.043 + .06368(34) = 10.208$$

A continuación, se estima el componente cíclico para 1993. Éste es un proceso subjetivo en el que el analista utiliza la información disponible proveniente de economistas destacados,

boletines de asociaciones comerciales, proyecciones internas de ventas y cualquier otra información que considere el curso futuro de los negocios de la empresa. Asimismo, el analista estudia el patrón anterior de índices cíclicos de los datos históricos, que aparecen en las figuras 8.6 y 8.7. El índice cíclico para el registro de automóviles nuevos ha disminuido de manera congruente desde 1988. ¿Continuará este patrón, o se nivelarán los registros, o inclusive aumentarán nuevamente? ¿Cómo se comportará la economía en 1993? ¿Qué pronostican los expertos para el ciclo de los negocios? ¿Cuáles son las políticas propuestas para los precios de automóviles nuevos en 1993? Una vez contestadas estas preguntas, el analista puede elegir una estimación del ciclo para 1993.

El aspecto más importante del proceso de estimación del ciclo consiste en que el pronosticador formule su predicción en la dirección correcta. Si el pronosticador predice con éxito si el patrón cíclico de una variable en particular continuará aumentando o disminuyendo, se nivelará, o tendrá un giro hacia arriba o hacia abajo, el pronóstico será razonablemente preciso. Si el analista indica la dirección correcta, no es tan importante el valor real empleado para la estimación.

En el último trimestre de 1992, el ingreso personal disponible real se incrementó en 3.8% y la tasa de desempleo declinó a 7.3%. También fueron favorables los factores específicos de la industria de vehículos automotores, para la compra de automóviles por el consumidor. Los programas de incentivos sobre ventas se mantuvieron en el último trimestre, tan atractivos como en el anterior; las tasas de interés se mantuvieron bajas y los precios de automóviles nuevos se incrementaron en una tasa anual de sólo 0.9%. El componente cíclico ( $C$ ) para 1992 fue de 79.39. Se estima el componente cíclico para 1993 para nivelarlo y regresar hacia arriba. Se estima un componente cíclico de .83 y se utiliza en la ecuación 8.1, para pronosticar 1993

$$\begin{aligned}\hat{Y}(1993) &= T \times C \\ &= 10.208(.83) = 8.473\end{aligned}$$

Resulta en extremo difícil estimar  $C$  para más de un año o dos a futuro. De ahí que cualquier pronóstico de largo plazo para dicha variable sea, en el mejor de los casos, riesgoso. Si la ecuación de tendencia se ajusta bien a los datos anteriores, entonces es razonable una proyección a futuro de cinco a diez años. Mientras que se espere que el crecimiento o declinación de la tendencia permanezca congruente con el comportamiento anterior, se puede anticipar un pronóstico preciso.

## VARIACIÓN ESTACIONAL

El análisis de tendencia tiene implicaciones en la planeación de largo plazo. El análisis del componente estacional de una serie histórica tiene implicaciones más inmediatas de corto plazo y es de gran importancia para los niveles medio e inferior de la administración. Por ejemplo, los planes de comercialización deben tener en cuenta los patrones estacionales esperados en las compras del consumidor. La identificación del comportamiento estacional en una serie histórica difiere del análisis de tendencia en por lo menos dos formas.

1. Mientras que la tendencia se determina en forma directa a partir de los datos disponibles, el componente estacional se determina eliminando los otros componentes, de modo que al final sólo quede el estacional.

2. Mientras que la tendencia se determina mediante una ecuación o líneas de mejor ajuste, se debe calcular un valor estacional por separado para cada mes (o trimestre) del año, por lo regular en la forma de un número índice. Al igual que en el análisis de tendencia, se han desarrollado diversos métodos para medir la variación estacional; la mayoría de los cálculos del índice estacional que ahora se emplean, son variaciones del *método de la proporción al promedio móvil*.

El componente estacional en una serie histórica se mide en la forma de un *número índice*. Su cálculo, que representa el grado de influencia estacional para un segmento del año en particular, comprende la comparación de los valores medidos o esperados para ese segmento (mes, trimestre, etc.), con el promedio general de todos los segmentos del año. De este modo, un índice estacional de 100 para un mes en particular, indica que el valor esperado para ese mes es de 1/12 del total del periodo anual centrado en ese mes. Un índice estacional de 125 para otro mes, indica que el valor esperado para ese mes es 25% mayor que 1/12 del total anual. Un índice mensual de 80 indica que el nivel de actividad esperado para ese mes es 20% menor que 1/12 del nivel de actividad total del año. Así, un número índice mensual indica las altas y bajas esperadas en los niveles de actividad mensual o trimestral, eliminando los efectos ocasionados por los componentes de tendencias cíclico e irregular.

Los *números índice* son porcentajes que reflejan cambios a través del tiempo.

En el siguiente ejemplo, se ilustra el método de la proporción al promedio móvil, para medir la variación estacional.

### **Ejemplo 8.6**

En el análisis de registros de automóviles nuevos, los datos mensuales para los años de 1986 a 1992 se utilizan como base para el análisis del componente estacional, en lugar de los datos mensuales para el periodo completo de 1960 a 1992. Una de las razones para esta reducción de datos es la de simplificar los cálculos en esta exposición. Otra y más importante razón es que si hubiese ocurrido cualquier cambio significativo en el patrón estacional de la compra de automóviles nuevos en los años recientes, el uso del periodo completo de 1960 a 1992 del análisis estacional pudiera resultar en proyecciones más pobres de los patrones estacionales esperados para 1993, que utilizando sólo datos de 1986 a 1992.

La tabla 8.4 presenta los registros mensuales de automóviles nuevos, en miles, para el periodo de enero de 1986 a diciembre de 1987, para ilustrar el principio de los cálculos de la proporción al promedio móvil. El primer paso para los datos mensuales es calcular un promedio móvil de 12 meses (para datos trimestrales, se calculará un promedio móvil de cuatro trimestres). Ya que en este promedio móvil se incluyen todos los meses del año, se eliminan los efectos ocasionados por el componente estacional, dejando sólo los efectos ocasionados por la tendencia de largo plazo, y los componentes cíclico e irregular en los promedios móviles.

A continuación se presentan los pasos para calcular índices estacionales mediante el método de la proporción al promedio móvil, y se muestran en la tabla 8.4.

**TABLA 8.4** REGISTROS DE AUTOMÓVILES NUEVOS DE PASAJEROS: DATOS MENSUALES, EN MILES, 1986-1987

| PERIODO    | REGISTROS | TOTAL<br>MÓVIL DE<br>12 MESES                                      | TOTAL<br>MÓVIL DE<br>2 AÑOS | PROMEDIO<br>MÓVIL<br>CENTRADO<br>12 MESES | ÍNDICE<br>ESTACIONAL |
|------------|-----------|--------------------------------------------------------------------|-----------------------------|-------------------------------------------|----------------------|
| 1986       |           |                                                                    |                             |                                           |                      |
| Enero      | 913       | 10,965<br>10,852<br>10,701<br>10,682<br>10,671<br>10,583<br>10,534 | 21,817}3                    | 909.04}4                                  | 102.75}5             |
| Febrero    | 822       |                                                                    |                             |                                           |                      |
| Marzo      | 848       |                                                                    |                             |                                           |                      |
| Abril      | 906       |                                                                    |                             |                                           |                      |
| Mayo       | 918       |                                                                    |                             |                                           |                      |
| Junio      | 1012      |                                                                    |                             |                                           |                      |
| Julio      | 934       |                                                                    |                             |                                           |                      |
| Agosto     | 894       |                                                                    |                             |                                           |                      |
| Septiembre | 1149      |                                                                    |                             |                                           |                      |
| Octubre    | 948       |                                                                    |                             |                                           |                      |
| Noviembre  | 719       |                                                                    |                             |                                           |                      |
| Diciembre  | 902       |                                                                    |                             |                                           |                      |
| 1987       |           |                                                                    |                             |                                           |                      |
| Enero      | 800       | 10,499                                                             | 21,033                      | 876.4                                     | 91.29                |
| Febrero    | 671       | 10,508                                                             | 21,007                      | 875.2                                     | 76.66                |
| Marzo      | 829       | 10,314                                                             | 20,822                      | 867.6                                     | 95.55                |
| Abril      | 895       | 10,185                                                             | 20,499                      | 854.1                                     | 104.79               |
| Mayo       | 830       | 10,184                                                             | 20,369                      | 848.7                                     | 97.80                |
| Junio      | 963       | 10,122                                                             | 20,306                      | 846.1                                     | 113.82               |
| Julio      | 899       |                                                                    |                             |                                           |                      |
| Agosto     | 903       |                                                                    |                             |                                           |                      |
| Septiembre | 955       |                                                                    |                             |                                           |                      |
| Octubre    | 819       |                                                                    |                             |                                           |                      |
| Noviembre  | 718       |                                                                    |                             |                                           |                      |
| Diciembre  | 840       |                                                                    |                             |                                           |                      |

**Paso 1** Calcule el total móvil de 12 meses y coloque el total para enero a diciembre de 1986 frente al 1o. de julio de 1986.

**Paso 2** De modo que los promedios subsiguientes se ubiquen en el centro de cada mes, calcule un total móvil de dos años.

**Paso 3** Como el total de dos años contiene los datos de 24 meses (enero de 1986 una vez, febrero a diciembre de 1986 dos veces, y enero de 1987 una vez), este total se centra frente al 15 de julio de 1986.

$$10,965 + 10,852 = 21,817$$

**Paso 4** Divida el total móvil de dos años entre 24, con el fin de obtener el promedio móvil centrado de 12 meses.

$$\frac{21,817}{24} = 909.04$$

**Paso 5** La proporción al promedio móvil se calcula dividiendo el valor real de cada mes entre el promedio móvil centrado de 12 meses, y multiplicando por 100 de modo que la proporción sea un número índice en la forma de un porcentaje.

$$\frac{934}{909.04} (100) = 102.75$$

Se repite este procedimiento para cada mes.

A continuación, se determina la proporción de porcentaje promedio para cada mes. La tabla 8.5 muestra el uso de este *método de media modificado*. Este enfoque elimina los valores más alto y más bajo y promedia los índices restantes.

**TABLA 8.5 CÁLCULO DE LA MEDIA MODIFICADA PARA ENERO, PARA EL EJEMPLO 8.5**

|       |        |                                    |
|-------|--------|------------------------------------|
| 80.00 |        | Eliminar el valor más alto (91.29) |
| 83.63 | 83.63  | y el más bajo (80.00) y promediar  |
| 86.85 | 86.85  | el resto:                          |
| 89.78 | 89.78  |                                    |
| 90.67 | 90.67  | $350.93 \div 4 = 87.73$            |
| 91.29 | 350.93 |                                    |

Por último, se calculan los índices estacionales multiplicando cada una de las proporciones promedio por un valor de modo que la suma de los 12 índices mensuales sea igual exactamente a 1 200. Debido a que el multiplicador debe ser mayor que 1 si el total de los promedios antes del ajuste es menor que 1,200, y menor que 1 si el total es mayor que 1,200, se define al multiplicador como

$$\text{Multiplicador} = \frac{1,200}{\text{Total real}} \quad (8.7)$$

**TABLA 8.6** RESUMEN DE LOS ÍNDICES ESTACIONALES MENSUALES PARA EL  
EJEMPLO 8.6

| MES        | 1986   | 1987   | 1988   | 1989   | 1990   | 1991   | 1992   | MEDIA<br>MENSUAL<br>MODIFICADA | MEDIA DEL<br>ÍNDICE<br>ESTACIONAL<br>AJUSTADO<br>*1.002311 |
|------------|--------|--------|--------|--------|--------|--------|--------|--------------------------------|------------------------------------------------------------|
| Enero      |        | 91.29  | 89.78  | 86.85  | 80.00  | 83.63  | 90.67  | 87.73                          | 87.9                                                       |
| Febrero    |        | 76.66  | 94.04  | 85.86  | 85.63  | 82.98  | 86.76  | 85.31                          | 85.5                                                       |
| Marzo      |        | 95.55  | 106.80 | 99.07  | 102.04 | 95.15  | 98.34  | 98.75                          | 99.0                                                       |
| Abril      |        | 104.79 | 99.16  | 100.20 | 100.03 | 96.95  | 102.96 | 100.59                         | 100.8                                                      |
| Mayo       |        | 97.80  | 102.59 | 105.73 | 109.71 | 107.98 | 103.24 | 104.89                         | 105.1                                                      |
| Junio      |        | 113.82 | 113.42 | 115.21 | 114.08 | 115.99 | 124.92 | 114.77                         | 115.0                                                      |
| Julio      | 102.75 | 106.72 | 102.02 | 102.29 | 103.75 | 110.81 |        | 103.88                         | 104.1                                                      |
| Agosto     | 99.55  | 106.60 | 104.75 | 109.46 | 103.99 | 98.95  |        | 103.72                         | 104.0                                                      |
| Septiembre | 128.96 | 111.48 | 109.86 | 119.69 | 111.47 | 108.06 |        | 113.12                         | 113.4                                                      |
| Octubre    | 106.55 | 95.39  | 95.06  | 100.96 | 104.97 | 101.38 |        | 100.67                         | 100.9                                                      |
| Noviembre  | 81.19  | 83.58  | 90.02  | 84.98  | 97.10  | 89.55  |        | 87.04                          | 87.2                                                       |
| Diciembre  | 102.51 | 97.45  | 105.73 | 89.01  | 94.86  | 92.23  |        | 96.76                          | 97.0                                                       |
|            |        |        |        |        |        |        |        | 1197.23                        | 1200.0                                                     |

La tabla 8.6 muestra los cálculos para los datos de registro de automóviles nuevos. El multiplicador es

$$\frac{1200}{1197.23} = 1.002311$$

La última columna de la tabla 8.6 lista los valores de los índices estacionales calculados para cada mes, determinados mediante el ajuste (multiplicando por 1.002311). Una revisión de estos valores muestra que el pico estacional en el número de registro de automóviles nuevos ocurre en los primeros meses del verano y la baja estacional ocurre en los meses de enero y febrero. El clima juega un papel importante en la explicación del patrón estacional de esta variable. El registro de automóviles nuevos tiende a subir en los meses de buen clima como junio y julio y a bajar en los meses de clima malo como enero y febrero.

Los resultados del análisis de variación estacional se pueden emplear para (1) programar la producción, (2) desestacionalizar las ventas, (3) evaluar las ventas actuales, la producción y los embarques y/o, (4) pronosticar ventas mensuales.

### **Tendencia estacional**

Una vez identificado el componente estacional, el siguiente paso consiste en calcular una ecuación de tendencia estacional (ecuación de tendencia mensual o trimestral). La tendencia se definió anteriormente en este capítulo como el crecimiento o declinación de largo plazo

de una serie histórica. Por esta razón, la tendencia de una serie mensual o trimestral debe ser congruente con su crecimiento o declinación de largo plazo. Se puede calcular la tendencia de datos estacionales utilizando alguno de los enfoques siguientes:

1. Si no hay datos disponibles de largo plazo, la tendencia estacional debe calcularse a partir de los datos estacionales.
2. Si la tendencia de los datos estacionales parece similar al crecimiento de largo plazo descrito mediante la ecuación de tendencia de largo plazo, se puede calcular la tendencia estacional a partir de los datos estacionales.
3. Si hay disponibles datos mensuales o trimestrales para la serie anual completa, se puede calcular la ecuación de tendencia estacional utilizando todos estos valores.
4. Si ya se calculó la ecuación de tendencia de largo plazo, se puede convertir matemáticamente a una ecuación de tendencia estacional con los datos estacionales disponibles.

### Ejemplo 8.7

La figura 8.2 muestra la ecuación de tendencia de largo plazo para el registro de automóviles nuevos de pasajeros de 1960 a 1992. El enfoque más simple para desarrollar una ecuación de tendencia estacional consiste en utilizar el procedimiento de mínimos cuadrados sobre los datos mensuales de 1986 a 1992. Ésta es una solución razonable si la ecuación de tendencia mensual resultante refleja con precisión el patrón de crecimiento de largo plazo. De acuerdo con los datos de la figura 8.2, la pendiente de una ecuación de tendencia mensual desarrollada a partir de los datos mensuales de 1986 a 1992, sería negativa. Debido a que la dirección de la tendencia de largo plazo ya fue establecida como positiva, dicha ecuación tendrá que convertirse para su uso con los datos estacionales. Este proceso de conversión se demuestra en el Apéndice A.

Los datos mensuales de esta serie, medidos en miles, se presentan en la primera columna del cuadro 8.7 y se emplean como la variable  $Y$ . Si se usaran los datos reales para desarrollar

**TABLA 8.7** CÁLCULOS PARA LOS COMPONENTES DE CORTO PLAZO

| PERIODO    | DATOS<br>Y | REGRESIÓN<br>T | AJUSTE EST.<br>TCI | CI     | C      | I      |
|------------|------------|----------------|--------------------|--------|--------|--------|
| 1986       |            |                |                    |        |        |        |
| Enero      | 913.00     | 811.09         | 1038.25            | 128.01 |        |        |
| Febrero    | 822.00     | 811.53         | 961.34             | 118.46 |        |        |
| Marzo      | 848.00     | 811.97         | 856.75             | 105.51 | 114.00 | 92.55  |
| Abril      | 906.00     | 812.42         | 898.64             | 110.61 | 110.04 | 100.53 |
| Mayo       | 918.00     | 812.86         | 873.22             | 107.43 | 108.39 | 99.11  |
| Junio      | 1012.00    | 813.30         | 879.69             | 108.16 | 108.41 | 99.77  |
| Julio      | 934.00     | 813.74         | 897.08             | 110.24 | 111.17 | 99.16  |
| Agosto     | 894.00     | 814.18         | 859.95             | 105.62 | 112.74 | 93.69  |
| Septiembre | 1149.00    | 814.63         | 1013.35            | 124.39 | 111.32 | 111.75 |
| Octubre    | 948.00     | 815.07         | 939.49             | 115.27 | 112.06 | 102.86 |
| Noviembre  | 719.00     | 815.51         | 824.20             | 101.07 | 113.23 | 89.26  |
| Diciembre  | 902.00     | 815.95         | 930.01             | 113.98 | 107.56 | 105.96 |

(continúa)

TABLA 8.7 (continuación)

| PERIODO    | DATOS<br>Y | REGRESIÓN<br>T | AJUSTE EST.<br>TCI | CI     | C      | I      |
|------------|------------|----------------|--------------------|--------|--------|--------|
| 1987       |            |                |                    |        |        |        |
| Enero      | 800.00     | 816.40         | 909.75             | 111.44 | 105.01 | 106.12 |
| Febrero    | 671.00     | 816.84         | 784.74             | 96.07  | 106.61 | 90.20  |
| Marzo      | 829.00     | 817.28         | 837.55             | 102.48 | 103.01 | 99.49  |
| Abril      | 895.00     | 817.72         | 887.73             | 108.56 | 101.17 | 107.30 |
| Mayo       | 830.00     | 818.16         | 789.51             | 96.50  | 103.04 | 93.65  |
| Junio      | 963.00     | 818.60         | 837.10             | 102.26 | 103.75 | 98.57  |
| Julio      | 899.00     | 819.05         | 863.46             | 105.42 | 102.58 | 102.77 |
| Agosto     | 903.00     | 819.49         | 868.61             | 105.99 | 103.07 | 102.84 |
| Septiembre | 955.00     | 819.93         | 842.26             | 102.72 | 102.67 | 100.05 |
| Octubre    | 819.00     | 820.37         | 811.65             | 98.94  | 102.68 | 96.36  |
| Noviembre  | 718.00     | 820.82         | 823.05             | 100.27 | 102.90 | 97.45  |
| Diciembre  | 840.00     | 821.26         | 866.09             | 105.46 | 105.40 | 100.05 |
| 1988       |            |                |                    |        |        |        |
| Enero      | 774.00     | 821.70         | 880.18             | 107.12 | 108.19 | 99.01  |
| Febrero    | 810.00     | 822.14         | 947.30             | 115.22 | 108.67 | 106.03 |
| Marzo      | 919.00     | 822.58         | 928.48             | 112.87 | 107.98 | 104.53 |
| Abril      | 852.00     | 823.03         | 845.08             | 102.68 | 107.26 | 95.73  |
| Mayo       | 883.00     | 823.47         | 839.92             | 102.00 | 104.79 | 97.34  |
| Junio      | 981.00     | 823.91         | 852.75             | 103.50 | 103.23 | 100.26 |
| Julio      | 883.00     | 824.35         | 848.09             | 102.88 | 102.72 | 100.16 |
| Agosto     | 901.00     | 824.79         | 866.68             | 105.08 | 101.69 | 103.33 |
| Septiembre | 937.00     | 825.23         | 826.38             | 100.14 | 102.19 | 97.99  |
| Octubre    | 807.00     | 825.68         | 799.76             | 96.86  | 103.97 | 93.16  |
| Noviembre  | 764.00     | 826.12         | 875.78             | 106.01 | 103.11 | 102.81 |
| Diciembre  | 896.00     | 826.56         | 923.83             | 111.77 | 103.50 | 107.99 |
| 1989       |            |                |                    |        |        |        |
| Enero      | 733.00     | 827.00         | 833.56             | 100.79 | 104.45 | 96.49  |
| Febrero    | 722.00     | 827.45         | 844.39             | 102.05 | 103.44 | 98.65  |
| Marzo      | 833.00     | 827.89         | 841.59             | 101.66 | 101.40 | 100.25 |
| Abril      | 843.00     | 828.33         | 836.15             | 100.94 | 101.16 | 99.78  |
| Mayo       | 885.00     | 828.77         | 841.83             | 101.58 | 99.97  | 101.61 |
| Junio      | 950.00     | 829.21         | 825.80             | 99.59  | 100.03 | 99.55  |
| Julio      | 830.00     | 829.66         | 797.19             | 96.09  | 100.15 | 95.94  |
| Agosto     | 880.00     | 830.10         | 846.48             | 101.97 | 98.91  | 103.09 |
| Septiembre | 956.00     | 830.54         | 843.14             | 101.52 | 97.42  | 104.21 |
| Octubre    | 800.00     | 830.98         | 792.82             | 95.41  | 95.40  | 100.00 |
| Noviembre  | 668.00     | 831.42         | 765.74             | 92.10  | 91.92  | 100.19 |
| Diciembre  | 694.00     | 831.86         | 715.55             | 86.02  | 90.07  | 95.50  |
| 1990       |            |                |                    |        |        |        |
| Enero      | 619.00     | 832.31         | 703.92             | 84.57  | 89.74  | 94.25  |
| Febrero    | 657.00     | 832.75         | 768.37             | 92.27  | 89.19  | 103.45 |
| Marzo      | 773.00     | 833.19         | 780.97             | 93.73  | 90.78  | 103.25 |
| Abril      | 751.00     | 833.63         | 744.90             | 89.36  | 91.74  | 97.40  |
| Mayo       | 824.00     | 834.08         | 783.80             | 93.97  | 91.21  | 103.03 |
| Junio      | 858.00     | 834.52         | 745.83             | 89.37  | 90.36  | 98.91  |

(continúa)



TABLA 8.7 (continuación)

| PERIODO    | DATOS<br>Y | REGRESIÓN<br>T | AJUSTE EST.<br>TCI | CI    | C     | I      |
|------------|------------|----------------|--------------------|-------|-------|--------|
| Julio      | 779.00     | 834.96         | 748.20             | 89.61 | 89.89 | 99.68  |
| Agosto     | 777.00     | 835.40         | 747.41             | 89.47 | 89.33 | 100.16 |
| Septiembre | 825.00     | 835.84         | 727.60             | 87.05 | 90.77 | 95.90  |
| Octubre    | 769.00     | 836.29         | 762.10             | 91.13 | 89.67 | 101.63 |
| Noviembre  | 705.00     | 836.73         | 808.15             | 96.58 | 88.04 | 109.70 |
| Diciembre  | 683.00     | 837.17         | 704.21             | 84.12 | 87.10 | 96.58  |
| 1991       |            |                |                    |       |       |        |
| Enero      | 599.00     | 837.61         | 681.18             | 81.32 | 84.99 | 95.68  |
| Febrero    | 590.00     | 838.05         | 690.01             | 82.33 | 81.64 | 100.85 |
| Marzo      | 669.00     | 838.49         | 675.90             | 80.61 | 81.68 | 98.69  |
| Abril      | 675.00     | 838.94         | 669.52             | 79.81 | 81.81 | 97.55  |
| Mayo       | 744.00     | 839.38         | 707.71             | 84.31 | 82.60 | 102.07 |
| Junio      | 792.00     | 839.82         | 688.46             | 81.98 | 81.93 | 100.06 |
| Julio      | 755.00     | 840.26         | 725.15             | 86.30 | 81.42 | 106.00 |
| Agosto     | 675.00     | 840.71         | 649.29             | 77.23 | 80.85 | 95.52  |
| Septiembre | 737.00     | 841.15         | 649.99             | 77.27 | 81.07 | 95.32  |
| Octubre    | 692.00     | 841.59         | 685.79             | 81.49 | 79.18 | 102.92 |
| Noviembre  | 610.00     | 842.03         | 699.25             | 83.04 | 80.43 | 103.24 |
| Diciembre  | 628.00     | 842.47         | 647.50             | 76.86 | 81.34 | 94.49  |
| 1992       |            |                |                    |       |       |        |
| Enero      | 619.00     | 842.91         | 703.92             | 83.51 | 80.97 | 103.14 |
| Febrero    | 590.00     | 843.36         | 690.01             | 81.82 | 80.65 | 101.45 |
| Marzo      | 665.00     | 843.80         | 671.86             | 79.62 | 80.86 | 98.47  |
| Abril      | 693.00     | 844.24         | 687.37             | 81.42 | 81.38 | 100.05 |
| Mayo       | 692.00     | 844.58         | 658.24             | 77.93 | 82.14 | 94.87  |
| Junio      | 837.00     | 845.13         | 727.57             | 86.09 | 80.11 | 107.46 |
| Julio      | 754.00     | 845.57         | 724.19             | 85.65 | 78.64 | 108.90 |
| Agosto     | 611.00     | 846.01         | 587.73             | 69.47 | 78.08 | 88.97  |
| Septiembre | 711.00     | 846.45         | 627.06             | 74.08 | 76.91 | 96.32  |
| Octubre    | 642.00     | 846.89         | 636.24             | 75.13 | 75.32 | 99.74  |
| Noviembre  | 593.00     | 847.34         | 679.76             | 80.22 |       |        |
| Diciembre  | 639.00     | 847.78         | 658.84             | 77.71 |       |        |
| 1993       |            |                |                    |       |       |        |
| Enero      |            | 848.22         |                    |       |       |        |
| Febrero    |            | 848.66         |                    |       |       |        |
| Marzo      |            | 849.10         |                    |       |       |        |
| Abril      |            | 849.54         |                    |       |       |        |
| Mayo       |            | 849.99         |                    |       |       |        |
| Junio      |            | 850.43         |                    |       |       |        |
| Julio      |            | 850.87         |                    |       |       |        |
| Agosto     |            | 851.31         |                    |       |       |        |
| Septiembre |            | 851.76         |                    |       |       |        |
| Octubre    |            | 852.20         |                    |       |       |        |
| Noviembre  |            | 852.64         |                    |       |       |        |
| Diciembre  |            | 853.08         |                    |       |       |        |

Pronóstico de tendencia =  $810.649 + .442 * \text{periodo X}$

la ecuación de tendencia, los meses representarían a la variable independiente ( $X$ ) con enero de 1986 como 1, febrero de 1986 como 2 y así sucesivamente. Entonces, se usaría la ecuación 8.5 para calcular la ecuación de tendencia estacional.

Para este ejemplo, la ecuación de tendencia de largo plazo se convirtió en una ecuación estacional (véase el Apéndice A):

$$\hat{Y} = 810.649 + .4422X$$

Utilizando la ecuación estacional, la tendencia estimada para diciembre de 1985 (cuando  $X = 0$ ) es 810 649. El incremento promedio por mes es de aproximadamente 442 registros.

En la segunda columna de la tabla 8.7 se muestran los valores de tendencia resultantes mediante esta ecuación. Nótese que la tendencia estimada para enero de 1986 es 811,091 (810,649 + 442). Nótese también que el incremento promedio de cada mes es de aproximadamente 442 registros.

## DATOS AJUSTADOS EN FORMA ESTACIONAL

En cada ejemplar del *Survey of Current Business* aparece un gran número de indicadores económicos. Muchos de estos indicadores contienen *datos ajustados en forma estacional* para que los usuarios del índice aprecien los patrones que son independientes de las variaciones estacionales. La eliminación de las variaciones estacionales ayuda a clarificar las fuerzas o debilidades básicas de los datos. Por ejemplo, los registros de automóviles nuevos pudieran incrementarse en un 10% de noviembre a diciembre, pero ¿es este incremento una indicación de que las ventas de automóviles están terminando un trimestre sobresaliente? Se puede responder esta pregunta con más precisión si se conocen las fluctuaciones en los registros de automóviles nuevos. La tabla 8.6 indica que un incremento es típico entre noviembre y diciembre. Quizás un ajuste del incremento estacional mostraría que todo el incremento se debe a factores estacionales.

Si se dividen los valores mensuales originales de una serie histórica entre sus índices estacionales correspondientes se dice que los datos resultantes están desestacionalizados o ajustados a la variación estacional. Como los valores resultantes aún incluyen movimientos de tendencia cíclicos e irregulares, el proceso de desestacionalización de datos puede representarse como

$$Y = TSCI \tag{8.2}$$

$$TCI = \frac{Y}{S} = \frac{TSCI}{S} \tag{8.8}$$

Resultan *datos ajustados en forma estacional* cuando los valores originales de una serie histórica, mensuales o trimestrales, se dividen entre sus índices estacionales correspondientes.

**Ejemplo 8.8**

La tabla 8.7 lista los datos ajustados en forma estacional para los registros de automóviles nuevos, así como diversos cálculos de corto plazo que se abordarán más adelante en este capítulo. Estos valores se calcularon dividiendo los valores reales mensuales de la tabla 8.7 entre los índices estacionales de la tabla 8.6 y multiplicando el resultado por 100. Debido a que se elimina el efecto del componente estacional, el número de registros para junio de cada año, un mes de altos registros, no es marcadamente más alto que el mes de febrero, que es un mes de bajos registros. De ahí que en los datos ajustados en forma estacional de la tabla 8.7 se haya eliminado el efecto del componente estacional, aunque aún se incluyan los efectos de los componentes de tendencia, cíclico e irregular. Los datos ajustados en forma estacional (*TCI*) se muestran en la tabla 8.7.

Un ejemplo de cálculo para enero de 1986 es

$$Y = 913$$

$$S = .879 \quad (\text{tabla 8.6})$$

$$Y = TSCI$$

$$\frac{Y}{S} = TCI$$

$$\frac{913}{.879} = 1038.68$$

Nótese que la salida de cómputo de la tabla 8.7 tiene un valor ligeramente distinto debido al error de redondeo.

## VARIACIONES DE CORTO PLAZO CÍCLICA E IRREGULAR

Ahora se pueden calcular las variaciones cíclica e irregular, para el análisis de corto plazo. Como los datos se ajustaron en forma estacional,

$$\frac{TSCI}{S} = TCI$$

se dividen estos valores desestacionalizados entre los valores de tendencia mensual correspondientes:

$$\frac{TCI}{T} = CI$$

Los valores *CI* se muestran en la columna 4 de la tabla 8.7. Por ejemplo, para enero de 1986,

$$Y = 913$$

$$S = .879 \quad (\text{tabla 8.6})$$

$$T = 811.09 \quad (\text{tabla 8.7})$$

$$TCI = \frac{TSCI}{S} = \frac{913}{.879} = 1038.68$$

$$CI = \frac{TCI}{T} = \frac{1038.68}{811.09} (100) = 128.06$$

Después de ajustar los datos para las influencias estacional y de tendencia, se analizan las variaciones cíclicas e irregulares. Se desarrolla un promedio móvil para atenuar las irregularidades. El proceso de cálculo es simple y se ilustra en el ejemplo 8.9.

### Ejemplo 8.9

La tabla 8.8 presenta los cálculos para un promedio móvil de cinco meses para la columna *CI* de la tabla 8.7. Nótese que pudo haberse empleado un promedio móvil de 7, 9 u 11 meses, para atenuar las variaciones aleatorias.

Los promedios móviles de cinco meses se incluyen en la columna 5 (con encabezado *C*) de la tabla 8.7 y, presumiblemente reflejan sólo el efecto del componente cíclico de corto plazo. Cuando se compara la columna cíclico (*C*) de la tabla 8.7 con la columna cíclico-irregular (*CI*), parece que el promedio móvil de cinco meses ha atenuado las irregularidades de la columna *CI*. El componente cíclico (columna *C*) se eleva y cae de manera atenuada y el ciclo es más fácil de identificar.

Como se muestra en la tabla 8.7, el índice cíclico comienza en marzo de 1986 con un valor alto (114), indicando buena temporada en la economía para ventas de automóviles. Esta buena temporada continúa hasta finales de 1989, en que una recesión económica parece deprimir las ventas de automóviles nuevos. Como se muestra en el resto de la columna *C*, esta mala temporada continúa hasta octubre de 1992, en que *C* alcanza un punto bajo de 75.32. Este índice cíclico de corto plazo se equipara al componente cíclico de largo plazo analizado en la figura 8.6.

**TABLA 8.8** PROCEDIMIENTO DE CÁLCULO DEL PROMEDIO MÓVIL DE 5 MESES PARA LOS DATOS DEL CUADRO 8.7.

| PERIODO<br>(1982) | <i>CI</i> | TOTAL MÓVIL<br>DE 5 MESES |  | <i>C</i> |
|-------------------|-----------|---------------------------|--|----------|
| Enero             | 128.01    |                           |  |          |
| Febrero           | 118.46    |                           |  |          |
| Marzo             | 105.51    | 570.02                    |  | 114.00   |
| Abril             | 110.61    | 550.17                    |  | 110.03   |
| Mayo              | 107.43    | 541.95                    |  | 108.39   |
| Junio             | 108.16    |                           |  |          |
| Julio             | 110.24    |                           |  |          |

Por último, la identificación del componente irregular se desarrolla en la ecuación 8.9 mediante la división de la columna  $CI$  de la tabla 8.7 entre la columna  $C$  y multiplicando el resultado por 100.

$$I = \frac{CI}{C} (100) \quad (8.9)$$

El componente irregular mide la variabilidad de las series de tiempo después de eliminar los otros componentes. La mayor parte del componente irregular está hecho de variabilidad aleatoria. Sin embargo, en ocasiones sucesos impredecibles causan irregularidades en una variable.

**Ejemplo  
8.10**

La última columna de la tabla 8.7 ( $I$ ) muestra los índices irregulares para los registros de automóviles nuevos de pasajeros. Estos valores se calcularon dividiendo los índices cíclico irregular de la columna 4 entre los índices cíclicos de la columna 5 y multiplicando el resultado por 100. El índice irregular para marzo de 1986 es

$$I = \frac{CI}{C} (100)$$

$$\frac{105.51}{114} (100) = 92.55$$

A veces, las variaciones irregulares en una serie histórica pueden explicarse. Una revisión de la columna irregular de la tabla 8.7 muestra que las irregularidades se presentaron en septiembre (111.75) y noviembre (89.26) de 1986. Estas irregularidades fueron provocadas por problemas de salarios y huelgas en la industria automotriz.

Una razón para descomponer una serie histórica consiste en aislar y examinar los componentes de la serie. Después de que el analista puede observar uno a la vez los componentes de tendencia, estacional, cíclico e irregular de una serie estacional, se pudieran ganar revelaciones en los patrones de los valores de los datos originales. Además de esta ganancia en conocimiento del proceso de descomposición, la identificación de componentes individuales hace mucho más fácil el pronóstico a futuro de la serie.

## PRONÓSTICO ESTACIONAL

Al pronosticar una serie histórica estacional, se invierte el proceso de descomposición. En vez de separar la serie en componentes para su examen, los componentes se recombinan para desarrollar los pronósticos de periodos futuros. Para desarrollar estos pronósticos, se emplea el modelo multiplicativo  $Y = TSCI$ .

**Ejemplo 8.11**

El pronóstico de corto plazo para enero de 1993 de los registros de automóviles nuevos de pasajeros se desarrolla utilizando las tablas 8.6 y 8.7.

1. *Tendencia.* Las estimaciones mensuales de la ecuación de tendencia se muestran en la tabla 8.7. El valor de la tendencia para enero de 1983 se calculó de la siguiente manera: como enero de 1986 está representado por  $X = 1$ , enero de 1993 lo está por  $X = 85$  y la proyección de la tendencia es 947.49

$$\begin{aligned} Y &= 810.649 + .4422X \\ &= 810.649 + .4422(85) \\ &= 848.23 \end{aligned}$$

2. *Estacional.* En la tabla 8.6 se encuentra que el índice estacional ajustado para el mes de enero es de 87.9.

3. *Cíclico.* El analista predice el índice cíclico utilizando toda la información reunida sobre el patrón cíclico. El analista debe responder las siguientes preguntas: ¿Para los últimos meses de 1992, el patrón cíclico que se muestra en la tabla 8.7 fue ascendente o descendente? ¿Se identificó algún indicador conducente? ¿Cuál es el pronóstico económico para 1993? Para llevar a término la demostración de este ejemplo, se utiliza un índice cíclico de 83.

4. *Irregular.* Puesto que las fluctuaciones irregulares son variaciones aleatorias, es común utilizar una estimación del 100% o 1.0. El pronóstico para enero de 1993 es

$$\begin{aligned} Y(\text{enero de 1993}) &= TSCI \\ &= (848.23) (.879) (.83) (1.0) \\ &= 618.84 \end{aligned}$$

El número real de registros de automóviles nuevos de pasajeros para enero de 1993 fue de 639 000. El pronóstico subestimó los registros en 20,160.

$$e_t = 639,000 - 618.840 = 20,160$$

En la práctica real, la importancia de los componentes individuales determina su uso en el pronóstico de corto plazo. Si una variable es estacional en extremo, el análisis de variación estacional proporciona al proceso de pronóstico un insumo importante si no es que total. Si se descubre un indicador conducente del cual se depende, el pronóstico pudiera basarse en él por completo. De ahí que si un componente domina el análisis, pudiera proporcionar, por sí solo, un pronóstico práctico y preciso.

## EL MÉTODO DE DESCOMPOSICIÓN CENSUS II

En la década de 1920, el National Bureau of Economic Research (Departamento Nacional de Investigación Económica), desarrolló un método que se utiliza comúnmente para el ajuste estacional y la descomposición de una serie histórica. Se considera a Julius Shiskin como el que contribuyó principalmente en el desarrollo de este enfoque. El método Census II se ha empleado ampliamente por el departamento de censos y otras agencias gubernamentales

y se utiliza cada vez más en el mundo de los negocios. Para una explicación detallada de este enfoque, deberá consultarse el Bureau of Census Technical Paper No. 15, "The X-11 Variant of the Census Method II Seasonal Adjustment Program".

La variación X-11 del enfoque Census II consta de cuatro fases. La primera fase pretende ajustar los datos de la serie histórica mensual a las variaciones por día comercial. Estos ajustes son necesarios debido a que el número de días laborables o comerciales varía de un año a otro, para un mes dado. En los sectores como el de ventas al menudeo, este factor puede afectar las ventas mensuales en forma significativa para un mes dado, de un año a otro.

La segunda fase del sistema X-11 comprende las estimaciones preliminares de factores estacionales y el ajuste de la serie a la estacionalidad. En esta fase se realiza una separación previa de la estacionalidad, de los factores tendencia/ciclo y se intenta después aislar la aleatoriedad.

En la tercera fase se afina el ajuste para proporcionar factores estacionales más precisos. Esta fase comprende también la estimación de los componentes de tendencia, cíclico e irregular.

En la última fase se generan estadísticas de resumen, que el analista puede emplear para determinar qué tan exitosamente aisló el método los factores estacionales, así como para desarrollar estimaciones futuras de tendencia/ciclo en los datos, para fines de pronóstico. Las pruebas en esta fase se basan en consideraciones intuitivas y no son estadísticas en el sentido matemático riguroso. Para una descripción más detallada del método Census II, se remite al lector a la obra *Forecasting: Methods and Applications* de Makridakis, Wheelwright y McGee.

## **APLICACIÓN A LA ADMINISTRACIÓN**

En años recientes, el análisis de series de tiempo ha ido a la vanguardia de las herramientas estadísticas para uso en pronósticos de sucesos futuros que están, en alguna forma, entrelazados con la economía. Los fabricantes están en extremo interesados en los ciclos de altibajos de la propia economía así como de las economías extranjeras de modo que puedan predecir mejor la demanda de sus productos, que a su vez impacta sus niveles de inventarios, requerimientos de personal, flujos de efectivo y casi todas las demás actividades de negocios dentro de la empresa.

La complejidad de estos problemas es enorme. Tómese, por ejemplo, el problema de predecir la demanda de petróleo y sus derivados. Al final de la década de 1960, el precio por barril de petróleo era muy bajo, y parecía haber una demanda mundial insaciable de gasolina y petróleo. Luego vino una conmoción en el precio del petróleo al principio y mediados de la década de 1970. ¿Cuál sería la demanda futura del petróleo? ¿cuáles los precios? Empresas como Exxon y General Motors estuvieron obviamente muy interesadas en estas preguntas. Si los precios del petróleo continuaban la escalada en forma logarítmica ¿disminuiría la demanda de automóviles grandes? ¿Cuál sería la demanda de electricidad? En términos generales, los analistas pronosticaron que la demanda de energía y entre ésta la del petróleo, no sería nada elástica; así, los precios continuarían sobrepasando a la inflación. Sin embargo, estos pronósticos no tomaron en cuenta un gran giro hacia abajo en el ciclo de los negocios a principios de

la década de 1980 y la mayor elasticidad de demanda de energía por el consumidor, que la pronosticada. Para 1980, el mundo comenzó a ver que el mercado se inundaba de petróleo con una caída radical en los precios. Parecía difícil creer que los consumidores se estuvieran beneficiando de hecho una vez más con la guerra de precios de la gasolina.

La demanda de petróleo se ve afectada no sólo por los sucesos cíclicos de largo plazo sino también por los sucesos estacionales y aleatorios, al igual que la mayoría de otros pronósticos de demanda para cualquier tipo de producto o servicio. Por ejemplo, considere las industrias de servicios y menudeo. Todos los pronosticadores de negocios y futuristas se encuentran prediciendo que en los próximos 25 años veremos un movimiento gradual del empleo de las manufacturas hacia los campos del menudeo y los servicios. De ahí que como el menudeo es un negocio en extremo estacional y cíclico, y como las proyecciones de inventarios son críticas para los detallistas, se usará más ampliamente el análisis de series históricas por parte de un número creciente de detallistas sofisticados. La supervivencia durante los periodos de gran cambio en la competencia de los que seremos testigos dependerá en ser más sofisticado que el propio competidor. Una indicación de que ya está ocurriendo un mayor interés en utilizar herramientas estadísticas más sofisticadas es el auge en el mercado de las mini y microcomputadoras en el campo del menudeo. También hay un auge en la demanda de paquetes estadísticos de cómputo.

Los fabricantes tendrán una continua necesidad de proyecciones estadísticas de sucesos futuros. ¿Cuál será la tasa de inflación? ¿Cómo afectará ésta los ajustes en el costo de vida que pudieran incluirse en el contrato laboral de una compañía? ¿Cómo afectarán estos ajustes a los precios y la demanda? ¿Cuál es el grupo de habilidades administrativas proyectado para 1990? ¿Cuál será el efecto de las estrategias de auge y caída del gasto y los impuestos?

Una de las principales preguntas que afecta a casi todos los segmentos de la economía es: ¿Cuál será la población futura de jóvenes? Los demógrafos observan de cerca la tasa actual de fertilidad y utilizan casi cualquier técnica de pronóstico de series de tiempo disponible para intentar proyectar esta variable. Errores de cálculo muy pequeños tendrían grandes impactos en todo, desde la fabricación de juguetes para bebés hasta la validez del financiamiento del sistema de Seguro Social. Es interesante notar que los demógrafos están observando ciclos de negocios de muy largo plazo (20 años o más por ciclo), en un intento por predecir qué es lo que hará la población de esta generación de mujeres en edad de procrear, con respecto a tener hijos. ¿Tendrán uno o dos hijos, como las familias de las décadas de 1960 y 1970, o volverán a tener dos o tres como las generaciones anteriores? Estas decisiones determinarán la composición de edades de la población en los próximos 50 a 75 años.

Los científicos políticos están interesados en el uso del análisis de series de tiempo para estudiar los patrones de cambio del gasto del gobierno en defensa y programas de bienestar social. Es obvio que estas tendencias tienen un gran impacto en el futuro de industrias completas.

Por último, en el campo legal ha aparecido un interesante microcosmos de aplicaciones del análisis de series históricas. Los abogados están haciendo un uso cada vez mayor de testigos expertos para testificar sobre el valor presente del ingreso futuro de una persona o empresa, el costo incurrido por la pérdida de un empleo debido a discriminación, o el efecto sobre el mercado de una huelga ilegal. Con frecuencia, estas cuestiones pueden responderse mejor a través del uso juicioso del análisis de series históricas.



## GLOSARIO

**Deflación de precios.** La deflación de precios es el proceso mediante el cual se expresan términos de una serie en dólares constantes.

**Indicadores empresariales.** Los indicadores empresariales son series de tiempo relacionadas con los negocios, que se usan para ayudar a evaluar el estado general de la economía.

**Números índice.** Los números índice son porcentajes que muestran cambios a través del tiempo.

## FÓRMULAS CLAVE

$$\text{Serie histórica: datos anuales} \quad Y = TC \quad (8.1)$$

$$\text{Serie histórica: datos mensuales o trimestrales} \quad Y = TSCI \quad (8.2)$$

$$\text{Poder de compra actual de \$1} \quad \frac{100}{\text{Índice actual de precios}} \quad (8.3)$$

$$\text{Valor deflactado del dólar} \quad (\text{Valor del dólar}) \times (\text{Poder de compra de \$1}) \quad (8.4)$$

$$\text{Ecuación de tendencia} \quad \hat{Y} = b_0 + bX \quad (8.5)$$

$$\text{Componente cíclico-irregular} \quad CI = \frac{Y}{T} = \frac{TCI}{T} \quad (8.6)$$

$$\text{Multiplicador} \quad \frac{1,200}{\text{Total real}} \quad (8.7)$$

$$\text{Datos ajustados en forma estacional} \quad TCI = \frac{Y}{S} = \frac{TSCI}{S} \quad (8.8)$$

$$\text{Componente irregular} \quad I = \frac{CI}{C} (100) \quad (8.9)$$

## PROBLEMAS

1. ¿Cuál es la finalidad de deflactar una serie histórica que está medida en dólares?
2. En el periodo base de junio de 1994, el precio de una cierta cantidad de bienes fue de \$1 289.73. En los meses más recientes, el precio índice para estos bienes fue 284.7. ¿Cuál sería el costo de dichos bienes si se compraran en el último mes?
3. Deflacte los siguientes volúmenes de ventas en dólares utilizando los índices de precios de las mercancías que se muestran. Estos índices son para todas las mercancías con 1982 = 100.

|      |         | VOLUMEN DE VENTAS<br>(EN DÓLARES) | ÍNDICE DE PRECIOS DE<br>MERCANCÍA (1982 = 100) |
|------|---------|-----------------------------------|------------------------------------------------|
| 1993 | Enero   | 358,235                           | 118.0                                          |
|      | Febrero | 297,485                           | 118.4                                          |
|      | Marzo   | 360,321                           | 118.7                                          |
|      | Abril   | 378,904                           | 119.2                                          |
|      | Mayo    | 394,472                           | 119.7                                          |
|      | Junio   | 312,589                           | 119.6                                          |
|      | Julio   | 401,345                           | 119.3                                          |

Fuente: *Survey of Current Business*.

4. Explique el concepto de descomposición de una serie histórica.
5. ¿Cuáles son los componentes que se analizan en la descomposición de una serie histórica anual?  
¿Y de una mensual o trimestral?
6. ¿Cuáles son las fuerzas básicas que afectan la tendencia secular de la mayoría de las variables?
7. ¿Qué clase de modelo de tendencia debería utilizarse en cada uno de los siguientes casos?
  - a. La variable se incrementa en una tasa constante.
  - b. La variable se incrementa en una tasa constante hasta que alcanza la saturación y se nivela.
  - c. La variable se incrementa en una cantidad constante.
8. ¿Cuál es la fuerza básica que afecta el componente cíclico en la mayoría de las variables?
9. Describa cada una de estas series, establecidas por el departamento de comercio de EUA: series conducentes, series coincidentes, series retrasadas.
10. ¿Cuál es la fuerza básica que afecta el componente estacional en la mayoría de las variables?
11. Las estimaciones del servicio *Value Line* de las ventas y el crecimiento de ingresos de compañías individuales se derivan mediante correlación con las ventas, ingresos y dividendos de los componentes correspondientes de las Cuentas Nacionales de Ingreso como gasto de capital. Un analista de *Value Line*, Jason Black, está revisando la tendencia en la variable gasto de capital entre 1977 y 1993. Los datos son los siguientes:

GASTO DE CAPITAL (MILES DE MILLONES \$), 1977–1992

| AÑO  | (MILES DE<br>MILLONES \$) | AÑO  | (MILES DE<br>MILLONES \$) |
|------|---------------------------|------|---------------------------|
| 1977 | 214                       | 1986 | 437                       |
| 1978 | 259                       | 1987 | 443                       |
| 1979 | 303                       | 1988 | 545                       |
| 1980 | 323                       | 1989 | 571                       |
| 1981 | 369                       | 1990 | 578                       |
| 1982 | 367                       | 1991 | 556                       |
| 1983 | 357                       | 1992 | 566                       |
| 1984 | 416                       | 1993 | 623                       |
| 1985 | 443                       | 1994 | 680*                      |

Fuente: *The Value Line Investment Survey* (New York: Value Line, 1988, 1990), p. 175.

\* Estimación de Value Line para 1994.

- a. Grafique los datos.
  - b. Determine el modelo de tendencia apropiado para los años 1977–1993.
  - c. Si el modelo apropiado es lineal, calcule el modelo de tendencia de línea recta para los años 1977–1993.
  - d. ¿Cuál es el incremento promedio en el gasto de capital por año, ocurrido desde 1987?
  - e. Estime el valor de la tendencia para el gasto de capital en 1994.
  - f. Compare su estimación de tendencia con la de *Value Line*.
  - g. ¿Cuál(es) factor(es) ejerce(n) influencia sobre el gasto de capital?
12. Esta pregunta se refiere al problema 11. Analice el componente cíclico para la variable gasto de capital.
- a. ¿Afecta la economía a este componente cíclico?
  - b. ¿Qué componente es más importante, el de tendencia o el cíclico?
  - c. ¿Cómo pronosticaría para 1994?
13. Una gran compañía está considerando recortar su publicidad en TV en favor de videos empresariales para repartir a sus clientes. Esta acción está en consideración después de que el presidente de la compañía leyó un artículo en una revista, referente a los videos como arma de ventas. Antes de tomar esta acción, el presidente quisiera investigar la historia de la publicidad televisiva en el país, en particular, el efecto cíclico.

A continuación se presentan los totales de dólares gastados, en millones, en la publicidad televisiva en EUA. Calcule los índices cíclicos para esta serie histórica anual.

| AÑO  | $Y$    | $\hat{Y}$ |
|------|--------|-----------|
| 1980 | 11,424 | 11,432.7  |
| 1981 | 12,811 | 13,222.1  |
| 1982 | 14,566 | 15,011.5  |
| 1983 | 16,542 | 16,800.8  |
| 1984 | 19,670 | 18,590.2  |
| 1985 | 20,770 | 20,379.6  |
| 1986 | 22,585 | 22,169.0  |
| 1987 | 23,904 | 23,958.3  |
| 1988 | 25,686 | 25,747.7  |
| 1989 | 26,891 | 27,537.1  |

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

Analice los picos y bajas del componente cíclico de la publicidad televisiva en dólares en EUA. Pronostique esta variable para 1990.

14. Suponga que los siguientes índices estacionales específicos para el mes de marzo se basan en el método de proporción de promedio móvil:

102.2 105.9 114.3 122.4 109.8 98.9

¿Cuál es el índice estacional para marzo utilizando el método modificado de la media?

15. El valor esperado de la tendencia para octubre es de \$850. Suponiendo un índice estacional de 112 para el mes de octubre, ¿cuál sería el pronóstico para octubre?

16. Para el mes de diciembre, están dados los siguientes índices estacionales específicos:

75.4 86.8 96.9 72.6 80.0 85.4

- Si el ajuste es de 1.05 y se emplea el método modificado de la media y si la tendencia esperada para diciembre es de \$900, ¿cuál es el pronóstico para diciembre?
17. Un hotel de veraneo cerca de Portland, Maine, ha estado registrando por varios años sus ventas mensuales, aunque nunca ha analizado estos datos. El hotel calcula los índices estacionales para sus ventas mensuales. ¿Cuáles de las siguientes declaraciones sobre el índice son correctas?
- La suma de los 12 números índices mensuales debe ser 1,200.
  - Un índice de 85 en mayo indica que las ventas son menores en un 15% del promedio de las ventas mensuales.
  - Un índice de 130 en enero indica que las ventas están 30% por arriba del promedio mensual de ventas.
  - El índice de cualquier mes debe estar entre cero y 200.
  - El índice promedio para cada uno de los 12 meses debería ser 100.
18. Al preparar un informe para June Bancrock, gerente de la Kula Department Store, usted incluye las siguientes estadísticas de las ventas del último año. Al observarlas, la Sra. Bancrock le dice "Este informe confirma lo que le he estado diciendo; el negocio está mejorando cada vez más." ¿Es precisa esta afirmación? ¿Por qué?

| MES        | VENTAS<br>(EN DÓLARES) | ÍNDICE<br>ESTACIONAL<br>AJUSTADO |
|------------|------------------------|----------------------------------|
| Enero      | 125                    | 51                               |
| Febrero    | 113                    | 50                               |
| Marzo      | 189                    | 87                               |
| Abril      | 201                    | 93                               |
| Mayo       | 206                    | 95                               |
| Junio      | 241                    | 99                               |
| Julio      | 230                    | 96                               |
| Agosto     | 245                    | 89                               |
| Septiembre | 271                    | 103                              |
| Octubre    | 291                    | 120                              |
| Noviembre  | 320                    | 131                              |
| Diciembre  | 419                    | 189                              |

Fuente: Registros de Kula Department Store.

19. A continuación se presentan los niveles de ventas trimestrales de Goodyear Tire. ¿Parece haber un efecto estacional significativo en sus niveles de ventas? Analice esta serie histórica para obtener cuatro índices estacionales y determine el alcance del componente estacional en las ventas de Goodyear.

| 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      | 2,850* | 3,100* | 3,100* | 3,050* |

Fuente: *The Value Line Investment Survey* (New York: Value Line, 1988, 1989, 1993, 1994), p. 130.

\* Estimación de Value Line para 1994.

- a. ¿Usaría usted para el pronóstico el componente de tendencia, el estacional, o ambos?
  - b. Pronostique el último trimestre de 1994.
  - c. Compare sus pronósticos con los de *Value Line*.
20. A continuación se presentan los valores de datos que representan las ventas mensuales de todas las tiendas al menudeo en EUA, en miles de millones de dólares. Analice esta serie incluyendo comentarios de los cuatro componentes de la misma. Pronostique 1993 y compare sus resultados con los valores reales que aparecen en la tabla.

| AÑO        | 1984  | 1985  | 1986  | 1987  | 1988  | 1989  | 1990  | 1991  | 1992  | 1993  |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Enero      | 93.1  | 98.8  | 105.6 | 106.4 | 113.6 | 122.5 | 132.6 | 130.9 | 142.1 | 148.5 |
| Febrero    | 93.7  | 95.6  | 99.7  | 105.8 | 115.0 | 118.9 | 127.3 | 128.6 | 143.1 | 145.3 |
| Marzo      | 104.3 | 110.2 | 114.2 | 120.4 | 131.6 | 141.3 | 148.3 | 149.3 | 154.7 | 164.6 |
| Abril      | 103.9 | 113.1 | 115.7 | 125.4 | 130.9 | 139.8 | 145.0 | 148.5 | 159.1 | 170.2 |
| Mayo       | 111.8 | 120.3 | 125.4 | 129.1 | 136.0 | 150.3 | 154.1 | 159.8 | 165.8 | 176.0 |
| Junio      | 112.3 | 115.0 | 120.4 | 129.0 | 137.5 | 149.0 | 153.5 | 153.9 | 164.6 | 175.6 |
| Julio      | 106.9 | 115.5 | 120.7 | 129.3 | 134.1 | 144.6 | 148.9 | 154.6 | 166.0 | 176.7 |
| Agosto     | 111.2 | 121.1 | 124.1 | 131.5 | 138.7 | 153.0 | 157.4 | 159.9 | 166.3 |       |
| Septiembre | 104.0 | 113.8 | 124.4 | 124.5 | 131.9 | 144.1 | 145.6 | 146.7 | 160.6 |       |
| Octubre    | 109.6 | 115.8 | 123.8 | 128.3 | 133.8 | 142.3 | 151.5 | 152.1 | 168.7 |       |
| Noviembre  | 113.5 | 118.1 | 121.4 | 126.9 | 140.2 | 148.8 | 156.1 | 155.6 | 167.2 |       |
| Diciembre  | 132.3 | 138.6 | 152.1 | 157.2 | 171.0 | 176.5 | 179.7 | 182.0 | 204.1 |       |

Fuente: *Survey of Current Business*, 1989, 1993.

21. Los índices ajustados de manera estacional que se presentan en el siguiente cuadro, reflejan el cambio en el volumen del negocio del Mt. Spokane Resort Hotel, que proporciona servicios al turismo familiar en el verano y el entusiasmo por el esquí en nieve durante los meses de invierno. No se esperan variaciones cíclicas agudas durante 1994.

| MES     | ÍNDICE<br>ESTACIONAL<br>AJUSTADO | MES        | ÍNDICE<br>ESTACIONAL<br>AJUSTADO |
|---------|----------------------------------|------------|----------------------------------|
| Enero   | 120                              | Julio      | 153                              |
| Febrero | 137                              | Agosto     | 151                              |
| Marzo   | 100                              | Septiembre | 95                               |
| Abril   | 33                               | Octubre    | 60                               |
| Mayo    | 47                               | Noviembre  | 82                               |
| Junio   | 125                              | Diciembre  | 97                               |

Fuente: Registros de Mt. Spokane Resort Hotel.

- a. Si en enero de 1994 hubieron 600 turistas en el hotel, ¿cuál es la estimación razonable para febrero?
  - b. La ecuación de tendencia mensual es  $Y = 140 + 5X$ , en donde  $X = 0$  representa el 15 de enero de 1988. ¿Cuál es el pronóstico para cada mes de 1994?
  - c. ¿Cuál es el número promedio de nuevos turistas por mes?
22. Seleccione un indicador conducente de la lista corta del NBER.
- a. Explique por qué este indicador conduce el ciclo de los negocios en los puntos críticos.
  - b. Exponga la importancia de este indicador como barómetro de la actividad empresarial en los últimos años.
23. ¿Cuál es la posición actual del ciclo de los negocios? ¿Se está expandiendo o contrayendo? ¿Cuándo ocurrirá el siguiente punto crítico?

## CASO DE ESTUDIO 8.1

### THE SMALL ENGINE DOCTOR

“The Small Engine Doctor” (el doctor de motores pequeños), es el nombre de un negocio desarrollado por Thomas Brown. Tom Brown era cartero del Servicio Postal de EUA. Desde su infancia siempre se interesó

en las reparaciones, desarmando viejos aparatos domésticos con el propósito de comprender “qué los hacía sonar”. Al crecer y convertirse en un típico habitante de los suburbios, adquirió equipo diverso de jardinería y para podar pasto. Cuando encontró que una escuela de la comunidad local ofrecía un curso sobre reparación de motores pequeños, aprovechó la oportunidad. Tom comenzó la reparación de motores pequeños desmantelando su propio equipo, dándole mantenimiento y ensamblándolo de nuevo. Poco después de terminar el curso, comenzó a reparar, para sus amigos y vecinos, segadoras, rotocultivadoras, máquinas para retirar nieve y otros equipos de jardinería. En el proceso, adquirió varios manuales de equipo y herramientas especiales.

No pasó mucho tiempo antes de que Tom decidiera convertir su afición en un negocio de tiempo parcial. Colocó un anuncio en una circular de una tienda local bajo el nombre de “The Small Engine Doctor”. Durante los dos últimos años, el negocio creció lo suficiente para proporcionar un agradable complemento a su salario regular. Aunque el crecimiento fue bienvenido, antes de entrar en su tercer año de funcionamiento, había cierto número de dudas. El negocio se encontraba en la casa de Tom. El sótano se dividió en una estancia, un taller y una oficina. La política de Tom es la de tener en existencia sólo un número limitado de refacciones, ordenando las que se requirieran según fuera necesario. Ésta parecía ser la única forma de poder manejar la gran variedad de refacciones que podían requerirse para reparar motores de una docena o más de fabricantes de equipo de jardinería.

Las refacciones probaron ser el problema más grave del negocio. Tom comenzó comprándolas en las distribuidoras de equipo. Esta práctica tenía varias desventajas. Primero, tenía que pagar una comisión por las refacciones y, segundo, la mayoría de las veces el distribuidor tenía que ordenar a su vez una o más refacciones para una determinada reparación. Las partes solicitadas al fabricante tenían tiempos de entrega de entre 30 y 120 días. Como resultado, Tom modificó su política y comenzó a ordenar las refacciones directamente a la fábrica. Se encontró con que los cargos por manejo y envío se comían sus utilidades, aún cuando el precio era de sólo el 60% del precio al menudeo. Sin embargo, los dos problemas mayores que generaban las refacciones eran ventas perdidas y espacio de almacenamiento. Tom atrajo clientes por la calidad de su servicio y sus precios razonables, que posiblemente se debían a sus bajos costos. Por desgracia, muchos clientes potenciales preferían ir a los distribuidores de equipo en vez de esperar varios meses para la reparación. El problema de mayor presión era el espacio de almacenamiento. Mientras que una pieza del equipo estaba esperando a que llegaran las refacciones, tenía que almacenarlo en sus instalaciones. No tardó mucho para que tanto su taller como su cochera para un automóvil se llenaran con equipo en espera de refacciones. De hecho, en el segundo año de funcionamiento, Tom suspendió la publicidad como táctica para limitar clientes debido a su falta de espacio.

Para su tercer año de funcionamiento, Tom Brown consideró ampliar las existencias de inventario. Esta práctica reduciría sus costos de compra haciendo posible obtener descuentos por volumen y términos de embarque más favorables. También esperaba que le diera un mejor tiempo de respuesta a sus clientes, mejorando tanto el flujo de efectivo como las ventas. Los riesgos de esta estrategia son el costo de tener un inventario no controlado y la obsolescencia de las refacciones.

Antes de comprometerse a sí mismo en el mayor inventario de refacciones, Tom desea tener un pronóstico confiable de la actividad de los negocios en el próximo año. Confía lo suficiente en su conocimiento de la mezcla del producto para agregar un pronóstico de órdenes de reparación de clientes como una base para hacer pedidos selectivos de refacciones. El pronóstico es complicado debido a los patrones de demanda estacional y a una tendencia al incremento en las ventas.

Tom planea desarrollar un pronóstico de ventas para su tercer año de funcionamiento. A continuación se presenta una historia de ventas para los dos primeros años.

| MES     | 1992 | 1993 | MES        | 1992 | 1993 |
|---------|------|------|------------|------|------|
| 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   |

## PREGUNTAS

1. Grafique los datos en un horizonte de dos años, de 1992 a 1993. Conecte los puntos de datos para hacer un diagrama de dispersión.
2. Desarrolle una ecuación de tendencia lineal mediante 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 de tendencia lineal correspondientes. Grafique una tendencia lineal ajustada de manera estacional.
4. Calcule un pronóstico de tendencia ajustada en forma exponencial, utilizando los valores ajustados constantes de ( $\alpha = .1$ ,  $\beta = .1$ ) ( $\alpha = .25$ ,  $\beta = .25$ ) y ( $\alpha = .5$ ,  $\beta = .5$ ). Grafique cada uno de los pronósticos. Extrapole los pronósticos exponenciales hasta el final del tercer año.
5. Calcule los valores de la DAM para los dos modelos que en forma visual parezcan dar los pronósticos más precisos.
6. Identifique cuál modelo usaría como base para su planeación del negocio en 1994 y exponga por qué seleccionó ese modelo entre los otros.

## CASO DE ESTUDIO 8.2

### MR. TUX

John Mosby se ha estado anticipando a la descomposición de su serie histórica mensual de ventas en dólares. Él sabe que su serie tiene un fuerte efecto estacional y le gustaría medirlo por dos razones. La primera, su banco está renuente de permitirle realizar pagos mensuales variables sobre su préstamo. John ha explicado que, debido a la estacionalidad de sus ventas y flujo mensual de efectivo, le gustaría hacer pagos adicionales en algunos meses y, reducir incluso a cero, los pagos en los otros meses. El banco desea ver alguna verificación de la afirmación de John de que sus ventas tienen un fuerte efecto estacional.

La segunda razón es que John desea poder pronosticar sus ventas mensuales. Requiere de estos pronósticos para fines de planeación, en especial debido al crecimiento de su negocio. Tanto los banqueros como quienes arriesgan capital desean ciertos pronósticos sólidos sobre los cuales basar sus decisiones de inversión. John sabe que su negocio está mejorando y que los prospectos futuros lucen brillantes, pero los inversionistas requieren documentación al respecto.





**TABLA 8.10** CÁLCULO DE LOS COMPONENTES DE CORTO PLAZO PARA EL CASO DE ESTUDIO 8.2

| PERIODO    | DATOS<br>Y | REGRESIÓN<br>T | AJUSTE EST.<br>TCI | CI     | C      | I      |
|------------|------------|----------------|--------------------|--------|--------|--------|
| 1986       |            |                |                    |        |        |        |
| Enero      | 16850.00   | 21937.77       | 53651.10           | 244.38 |        |        |
| Febrero    | 12753.00   | 24815.35       | 26348.65           | 106.18 |        |        |
| Marzo      | 26901.00   | 27676.93       | 31209.57           | 112.76 | 161.59 | 69.78  |
| Abril      | 61494.00   | 30538.51       | 34342.66           | 112.46 | 139.23 | 80.77  |
| Mayo       | 147862.00  | 33400.09       | 77548.08           | 232.18 | 143.46 | 161.84 |
| Junio      | 57990.00   | 36261.67       | 48075.48           | 132.58 | 140.37 | 94.45  |
| Julio      | 51318.00   | 39123.25       | 49807.67           | 127.31 | 128.87 | 98.79  |
| Agosto     | 53599.00   | 41984.83       | 40858.39           | 97.32  | 103.63 | 93.91  |
| Septiembre | 23038.00   | 44846.41       | 24653.47           | 54.97  | 90.38  | 60.82  |
| Octubre    | 41396.00   | 47707.99       | 50556.44           | 105.97 | 76.03  | 139.38 |
| Noviembre  | 19330.00   | 50569.57       | 33542.06           | 66.33  | 73.98  | 89.65  |
| Diciembre  | 22707.00   | 53431.15       | 29693.33           | 55.57  | 84.52  | 65.75  |
| 1987       |            |                |                    |        |        |        |
| Enero      | 15395.00   | 56292.73       | 49018.32           | 87.08  | 72.90  | 119.44 |
| Febrero    | 30826.00   | 59154.31       | 63688.82           | 107.67 | 77.40  | 139.10 |
| Marzo      | 25589.00   | 62015.89       | 29687.53           | 47.87  | 96.89  | 49.41  |
| Abril      | 103184.00  | 64877.47       | 57625.36           | 88.82  | 95.58  | 92.93  |
| Mayo       | 197608.00  | 67739.05       | 103637.99          | 153.00 | 84.59  | 180.86 |
| Junio      | 68600.00   | 70600.63       | 56871.49           | 80.55  | 93.27  | 86.37  |
| Julio      | 39909.00   | 73462.21       | 38734.45           | 52.73  | 91.39  | 57.69  |
| Agosto     | 91368.00   | 76323.79       | 69649.62           | 91.26  | 78.56  | 116.16 |
| Septiembre | 58781.00   | 79185.37       | 62902.84           | 79.44  | 76.12  | 104.36 |
| Octubre    | 59679.00   | 82046.95       | 72885.26           | 88.83  | 81.58  | 108.89 |
| Noviembre  | 33443.00   | 84908.53       | 58031.41           | 68.35  | 82.84  | 82.50  |
| Diciembre  | 53719.00   | 87770.11       | 70246.88           | 80.04  | 83.15  | 96.25  |
| 1988       |            |                |                    |        |        |        |
| Enero      | 27773.00   | 90631.69       | 88430.38           | 97.57  | 77.71  | 125.56 |
| Febrero    | 36653.00   | 93493.27       | 75727.84           | 81.00  | 88.53  | 91.50  |
| Marzo      | 51157.00   | 96354.85       | 59350.70           | 61.60  | 93.71  | 65.73  |
| Abril      | 217509.00  | 99215.43       | 121472.64          | 122.43 | 91.59  | 133.67 |
| Mayo       | 206229.00  | 102078.01      | 108159.39          | 105.96 | 93.92  | 112.82 |
| Junio      | 110081.00  | 104939.59      | 91260.51           | 86.96  | 99.35  | 87.53  |
| Julio      | 102893.00  | 107801.17      | 99864.77           | 92.54  | 94.62  | 97.91  |
| Agosto     | 128857.00  | 110662.75      | 98227.40           | 88.76  | 96.73  | 91.76  |
| Septiembre | 104776.00  | 113524.33      | 112123.11          | 98.77  | 97.88  | 100.91 |
| Octubre    | 111036.00  | 116385.91      | 135606.95          | 116.51 | 97.05  | 120.06 |
| Noviembre  | 63701.00   | 119247.48      | 110536.11          | 92.69  | 95.31  | 97.26  |
| Diciembre  | 82657.00   | 122109.07      | 108088.32          | 88.52  | 91.18  | 97.08  |
| 1989       |            |                |                    |        |        |        |
| Enero      | 31416.00   | 124970.65      | 100029.84          | 80.04  | 83.08  | 96.34  |
| Febrero    | 48341.00   | 127832.23      | 99876.11           | 78.13  | 84.84  | 92.09  |
| Marzo      | 85651.00   | 130693.80      | 99369.53           | 76.03  | 89.40  | 85.05  |
| Abril      | 242673.00  | 133555.39      | 135526.03          | 101.48 | 92.96  | 109.16 |
| Mayo       | 289554.00  | 136416.97      | 151860.23          | 111.32 | 99.27  | 112.14 |
| Junio      | 164373.00  | 139278.55      | 136270.23          | 97.84  | 102.58 | 95.38  |
| Julio      | 160608.00  | 142140.13      | 155881.17          | 109.67 | 102.89 | 106.59 |
| Agosto     | 176096.00  | 145001.70      | 134237.58          | 92.58  | 99.24  | 93.28  |
| Septiembre | 142363.00  | 147863.28      | 152345.78          | 103.03 | 105.33 | 97.81  |
| Octubre    | 114907.00  | 150724.86      | 140334.56          | 93.11  | 104.64 | 88.98  |

(continúa)

TABLA 8.10 (continuación)

| PERIODO    | DATOS<br>Y | REGRESIÓN<br>T | AJUSTE EST.<br>TCI | CI     | C      | I      |
|------------|------------|----------------|--------------------|--------|--------|--------|
| Noviembre  | 113552.00  | 153586.44      | 197039.23          | 128.29 | 106.75 | 120.18 |
| Diciembre  | 127042.00  | 156448.02      | 166129.38          | 106.19 | 106.62 | 99.59  |
| 1990       |            |                |                    |        |        |        |
| Enero      | 51604.00   | 159309.59      | 164309.28          | 103.14 | 117.38 | 87.87  |
| Febrero    | 80366.00   | 162171.17      | 166042.16          | 102.39 | 109.27 | 93.70  |
| Marzo      | 208938.00  | 165032.75      | 242403.16          | 146.88 | 103.53 | 141.88 |
| Abril      | 263830.00  | 167894.34      | 147341.61          | 87.76  | 103.87 | 84.49  |
| Mayo       | 252216.00  | 170755.92      | 132277.84          | 77.47  | 99.79  | 77.63  |
| Junio      | 219566.00  | 173617.50      | 182026.91          | 104.84 | 88.59  | 118.34 |
| Julio      | 149082.00  | 176479.08      | 144694.39          | 81.99  | 92.06  | 89.06  |
| Agosto     | 213888.00  | 179340.66      | 163046.34          | 90.91  | 94.21  | 96.50  |
| Septiembre | 178947.00  | 182202.23      | 191495.14          | 105.10 | 94.84  | 110.82 |
| Octubre    | 133650.00  | 185063.81      | 163225.16          | 88.20  | 100.94 | 87.38  |
| Noviembre  | 116946.00  | 187925.39      | 202928.61          | 107.98 | 102.11 | 105.75 |
| Diciembre  | 164154.00  | 190786.97      | 214659.73          | 112.51 | 98.41  | 114.33 |
| 1991       |            |                |                    |        |        |        |
| Enero      | 58843.00   | 193648.56      | 187358.56          | 96.75  | 106.94 | 90.48  |
| Febrero    | 82386.00   | 196510.14      | 170215.63          | 86.52  | 104.91 | 82.57  |
| Marzo      | 224803.00  | 199371.72      | 260809.20          | 130.82 | 99.19  | 131.88 |
| Abril      | 354301.00  | 202233.30      | 197867.11          | 97.84  | 104.85 | 93.31  |
| Mayo       | 328263.00  | 205094.88      | 172161.64          | 83.94  | 107.28 | 78.24  |
| Junio      | 313647.00  | 207956.45      | 260022.92          | 125.04 | 105.18 | 118.88 |
| Julio      | 214561.00  | 210818.03      | 208246.30          | 98.78  | 103.75 | 95.21  |
| Agosto     | 337192.00  | 213679.61      | 257040.70          | 120.29 | 103.06 | 116.72 |
| Septiembre | 183482.00  | 216541.19      | 196348.14          | 90.67  | 99.87  | 90.79  |
| Octubre    | 144618.00  | 219402.78      | 176620.25          | 80.50  | 101.55 | 79.27  |
| Noviembre  | 139750.00  | 222264.36      | 242498.88          | 109.10 | 97.34  | 112.09 |
| Diciembre  | 184546.00  | 225125.94      | 241325.81          | 107.20 | 106.58 | 100.58 |
| 1992       |            |                |                    |        |        |        |
| Enero      | 71043.00   | 227987.52      | 226203.86          | 99.22  | 115.35 | 86.01  |
| Febrero    | 152930.00  | 230849.09      | 315964.78          | 136.87 | 112.87 | 121.26 |
| Marzo      | 250559.00  | 233710.67      | 290690.47          | 124.38 | 108.72 | 114.40 |
| Abril      | 409567.00  | 236572.25      | 228731.61          | 96.69  | 107.55 | 89.90  |
| Mayo       | 394747.00  | 239433.83      | 207030.00          | 86.47  | 98.41  | 87.86  |
| Junio      | 272874.00  | 242295.41      | 226220.88          | 93.37  | 96.62  | 96.64  |
| Julio      | 230303.00  | 245157.00      | 223525.00          | 91.18  | 93.95  | 97.05  |
| Agosto     | 375402.00  | 248018.58      | 286168.09          | 115.38 | 93.36  | 123.59 |
| Septiembre | 195409.00  | 250880.16      | 290111.50          | 83.35  | 99.26  | 83.97  |
| Octubre    | 173518.00  | 253741.73      | 211915.48          | 83.52  | 107.10 | 77.98  |
| Noviembre  | 181702.00  | 256603.31      | 315295.41          | 122.87 |        |        |
| Diciembre  | 258713.00  | 259464.89      | 338311.97          | 130.39 |        |        |
| 1993       |            |                |                    |        |        |        |
| Enero      |            | 262326.47      |                    |        |        |        |
| Febrero    |            | 265188.06      |                    |        |        |        |
| Marzo      |            | 268049.63      |                    |        |        |        |
| Abril      |            | 270911.22      |                    |        |        |        |
| Mayo       |            | 273772.81      |                    |        |        |        |
| Junio      |            | 276634.38      |                    |        |        |        |
| Julio      |            | 279495.97      |                    |        |        |        |
| Agosto     |            | 282357.53      |                    |        |        |        |

(continúa)

TABLA 8.10 (continuación)

| PERIODO                                               | DATOS<br>Y | REGRESIÓN<br>T | AJUSTE EST.<br>TCI | CI | C | I |
|-------------------------------------------------------|------------|----------------|--------------------|----|---|---|
| Septiembre                                            |            | 285219.13      |                    |    |   |   |
| Octubre                                               |            | 288080.72      |                    |    |   |   |
| Noviembre                                             |            | 290942.28      |                    |    |   |   |
| Diciembre                                             |            | 293803.88      |                    |    |   |   |
| PRONÓSTICO FINAL = 19092.19 + 2861.579834 * X PERIODO |            |                |                    |    |   |   |

negocios en 1993, y como el último valor de *C* en octubre de 1992 ascendió (107.10), decide utilizar los siguientes valores de *C* para sus pronósticos:

|      |           |     |
|------|-----------|-----|
| 1992 | Noviembre | 102 |
|      | Diciembre | 103 |
| 1993 | Enero     | 104 |
|      | Febrero   | 104 |
|      | Marzo     | 105 |
|      | Abril     | 105 |
|      | Mayo      | 106 |
|      | Junio     | 107 |

Con respecto al valor irregular (*I*) para estos meses, John no prevé ningún suceso inusual, excepto para marzo de 1993. En ese mes, él planea hacer una promoción con tarifas reducidas en una de sus tiendas que está terminando de remodelar. Como esta promoción estará respaldada con publicidad en radio y televisión, espera que las ventas en esa tienda sean 50% más altas de lo normal. Para sus ventas mensuales globales, piensa que este efecto representará alrededor de un 15% más en las ventas totales.

Con las cifras estimadas por John y su salida de cómputo, él hace los siguientes pronósticos de ventas de Mr. Tux para los primeros seis meses de 1993:

| PRONÓSTICO DE VENTAS | T                    | S       | C      | I      |
|----------------------|----------------------|---------|--------|--------|
| Enero                | 85,665 = 262,326.47  | (.314)  | (1.04) | (1.0)  |
| Febrero              | 133,485 = 265,188.06 | (.484)  | (1.04) | (1.0)  |
| Marzo                | 279,003 = 268,049.63 | (.862)  | (1.05) | (1.15) |
| Abril                | 509,462 = 270,911.22 | (1.791) | (1.05) | (1.0)  |
| Mayo                 | 553,410 = 273,772.81 | (1.907) | (1.06) | (1.0)  |
| Junio                | 356,975 = 276,634.38 | (1.206) | (1.07) | (1.0)  |

John queda perturbado después de estudiar los pronósticos para 1993, al observar un amplio margen en las proyecciones de ventas mensuales —de \$85,665 a 553,410. Aunque sabe que sus

volumenes mensuales tienen una variabilidad considerable, le preocupa lo amplio de las fluctuaciones. John ha estado pensando en extender su ubicación actual en Spokane al área de Seattle. Recientemente descubrió que en Seattle hay diversos eventos “de etiqueta”, que son diferentes de su mercado actual en Spokane. Como los eventos se llevan a cabo en el otoño, cuando el mercado de Spokane es menor (véanse los índices estacionales para octubre y noviembre), percibe la ventaja de nivelar su negocio ingresando al mercado de Seattle.

Sin embargo, su verdadera preocupación se enfoca en sus peores meses, enero y febrero. Recientemente ha estado considerando comprar una máquina de confección de camisas de etiqueta que vio en una exhibición, pensando en que pudiera concentrarse en esa actividad durante los meses de invierno. Si recibe una reacción positiva de compradores potenciales de camisas en dicho periodo, podría intentarlo. Como se puede ver, los índices estacionales de su salida de cómputo han enfocado su atención en los extremos de la oscilación de sus niveles mensuales de ventas.

## PREGUNTAS

1. Suponga que el banco de John pide dos frases que justifiquen la solicitud de John para hacer pagos extra a su préstamo en algunos meses y ningún pago en otros. Escriba esas dos frases.
2. Suponga que el próximo año John duplica su negocio en Seattle con respecto a Spokane. Determine los índices estacionales en Seattle que serían ideales para equilibrar los ingresos mensuales de Mr. Tux.
3. Sin contar a Seattle, ¿qué volumen tendría John que vender de su máquina de confección de camisas para que los meses de enero y febrero estuvieran en el “promedio”?

## CASO DE ESTUDIO 8.3

### CONSUMER CREDIT COUNSELING

El funcionamiento de Consumer Credit Counseling (CCC) se describió en los capítulos 1 (Caso de estudio 1.2) y 3 (Caso de estudio 3.4).

El director ejecutivo Marv Harnishfeger, llegó a la conclusión de que la variable más importante que CCC requería pronosticar era el número de nuevos clientes por atender en el resto de 1993. Marv proporcionó a Dorothy Mercer los datos del número de clientes atendidos por CCC en el periodo de enero de 1985 a marzo de 1993 (véase caso de estudio 4.3).

Dorothy le da a usted estos datos y le pide que lleve a cabo un análisis de descomposición de la serie histórica. Ella hace énfasis en que desea comprender por completo los componentes de tendencia, cíclico y estacional. Dorothy desea conocer la importancia de cada componente. También desea saber si aparece alguna irregularidad en los datos. Su instrucción final es que pronostique el resto de 1993.

## TAREA

Escriba un informe que proporcione a Dorothy la información solicitada. Al programar el resto de 1993, suponga que el componente cíclico mantendrá la misma dirección que indican los tres primeros meses.

## VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Para pronosticar el número de clientes, decidimos utilizar el método de descomposición Census X-11. Para utilizar este programa, necesitamos ajustar los datos para tomar en cuenta los ciclos de días comerciales, es decir, los días que son comerciales (laborables) contra aquellos que no lo son. Primero, se selecciona el menú de administración de datos (opción 1). A continuación, se elige en el programa el ajuste para días comerciales (opción 6). Este programa utiliza por omisión 5 días comerciales por semana y 9 días festivos, como se describe en la sección de la tabla de parámetros.

Después del ajuste, se accesa el programa de pronóstico mediante la elección de la opción 3 del menú principal. Si emplean discos (no un sistema de disco duro), el sistema le solicitará que inserte el disco del sistema principal

```
Insert MAIN SYSTEM DISK in default drive
Press <ENTER> when ready
```

En este punto, usted necesita reemplazar el Data Management Disk por el Main System Disk. Presione <Enter> y el programa pedirá el nombre del archivo a utilizar

```
What is the name of the data file? CREDIT
```

El menú le permite elegir entre 13 diferentes técnicas de pronóstico. La pantalla correspondiente se mostró en la sección de Forecast Plus del cap. 5.

Seleccionamos descomposición Census X-11, opción 9.

```
*** Census X-11 Decomposition ***

Quarterly or monthly labeling [Q/M]: M
Starting year: 1985
Starting month/quarter: 1

Table of original data series: N
Final seasonal-irregular ratios (by percent): N
Final seasonal factors (by percent): N
Final seasonally adjusted series: N
Final trend cycle: N
Final irregular factors (by percent): N
Months for cyclical dominance moving average: N
Side by side plot of original data and adjusted data: Y
Side by side plot of irregular, seasonal, and trend: Y
Box plot of irregular variation: Y
Box plot of seasonal variation: Y
Component by season plot: Y
```

```
Press ENTER to accept:
```

Para ejecutar esta descomposición X-11, podríamos usar las opciones por omisión que ya se incluyen. La figura 8.8. muestra la salida cuando se responde *Y* a los factores estacionales finales y a la graficación de la variación estacional, y *N* al resto de las opciones.

Finalmente, el programa preguntará si desea almacenar los pronósticos de residuos

Do you want to save the forecasts of residuals (y/n)?

## PRONÓSTICO MEDIANTE HOJAS DE CÁLCULO

El Spreadsheet Forecaster es un paquete de herramientas que contiene 23 modelos prácticos de pronóstico empresarial, listos para su uso. Estos modelos son patrones que funcionan con Lotus 1-2-3 (versión 2.01 o posterior). Cada modelo se suministra con datos de muestra, una gráfica preestablecida e instrucciones paso a paso.

| NUMBER OF CLIENTS - X11 DECOMPOSITION |       |       |       |      |      |       |      |      |      |       |       |      |
|---------------------------------------|-------|-------|-------|------|------|-------|------|------|------|-------|-------|------|
| FINAL SEASONAL FACTORS                |       |       |       |      |      |       |      |      |      |       |       |      |
| YEAR                                  | JAN   | FEB   | MAR   | APR  | MAY  | JUN   | JUL  | AUG  | SEP  | OCT   | NOV   | DEC  |
| 1985                                  | 117.1 | 130.1 | 109.6 | 99.2 | 96.3 | 103.9 | 82.8 | 89.2 | 88.5 | 103.1 | 98.4  | 82.3 |
| 1986                                  | 116.7 | 129.6 | 111.0 | 99.1 | 95.7 | 102.7 | 83.2 | 89.8 | 88.4 | 102.9 | 99.4  | 83.2 |
| 1987                                  | 114.3 | 128.1 | 114.2 | 98.7 | 95.9 | 101.1 | 83.6 | 89.6 | 88.4 | 103.4 | 100.1 | 84.1 |
| 1988                                  | 113.6 | 126.1 | 116.4 | 97.8 | 95.5 | 99.4  | 85.1 | 89.4 | 89.2 | 103.7 | 100.5 | 85.4 |
| 1989                                  | 111.0 | 125.6 | 119.2 | 95.4 | 94.8 | 97.9  | 87.5 | 89.7 | 90.8 | 102.5 | 100.8 | 86.5 |
| 1990                                  | 110.8 | 124.1 | 121.0 | 92.9 | 93.3 | 96.2  | 91.0 | 90.4 | 92.6 | 101.3 | 101.5 | 86.9 |
| 1991                                  | 108.8 | 123.0 | 123.2 | 90.3 | 92.8 | 95.1  | 94.0 | 90.9 | 94.2 | 100.4 | 101.8 | 86.3 |
| 1992                                  | 108.9 | 121.3 | 123.3 | 89.4 | 93.1 | 95.1  | 95.5 | 90.7 | 95.1 | 100.6 | 101.3 | 85.7 |
| 1993                                  | 108.5 | 121.8 | 122.6 |      |      |       |      |      |      |       |       |      |

| PROJECTED ONE YEAR AHEAD SEASONAL FACTORS |       |       |       |      |      |      |      |      |      |       |       |      |
|-------------------------------------------|-------|-------|-------|------|------|------|------|------|------|-------|-------|------|
| YEAR                                      | JAN   | FEB   | MAR   | APR  | MAY  | JUN  | JUL  | AUG  | SEP  | OCT   | NOV   | DEC  |
| 1993                                      |       |       |       | 89.0 | 93.2 | 95.1 | 96.3 | 90.6 | 95.5 | 100.6 | 101.0 | 85.4 |
| 1994                                      | 108.3 | 122.1 | 122.3 |      |      |      |      |      |      |       |       |      |

**Figura 8.8** Census X-11 para el caso de estudio 8.3.

## NUMBER OF CLIENTS - X11 DECOMPOSITION

## BOX PLOT OF SEASONAL VARIATION

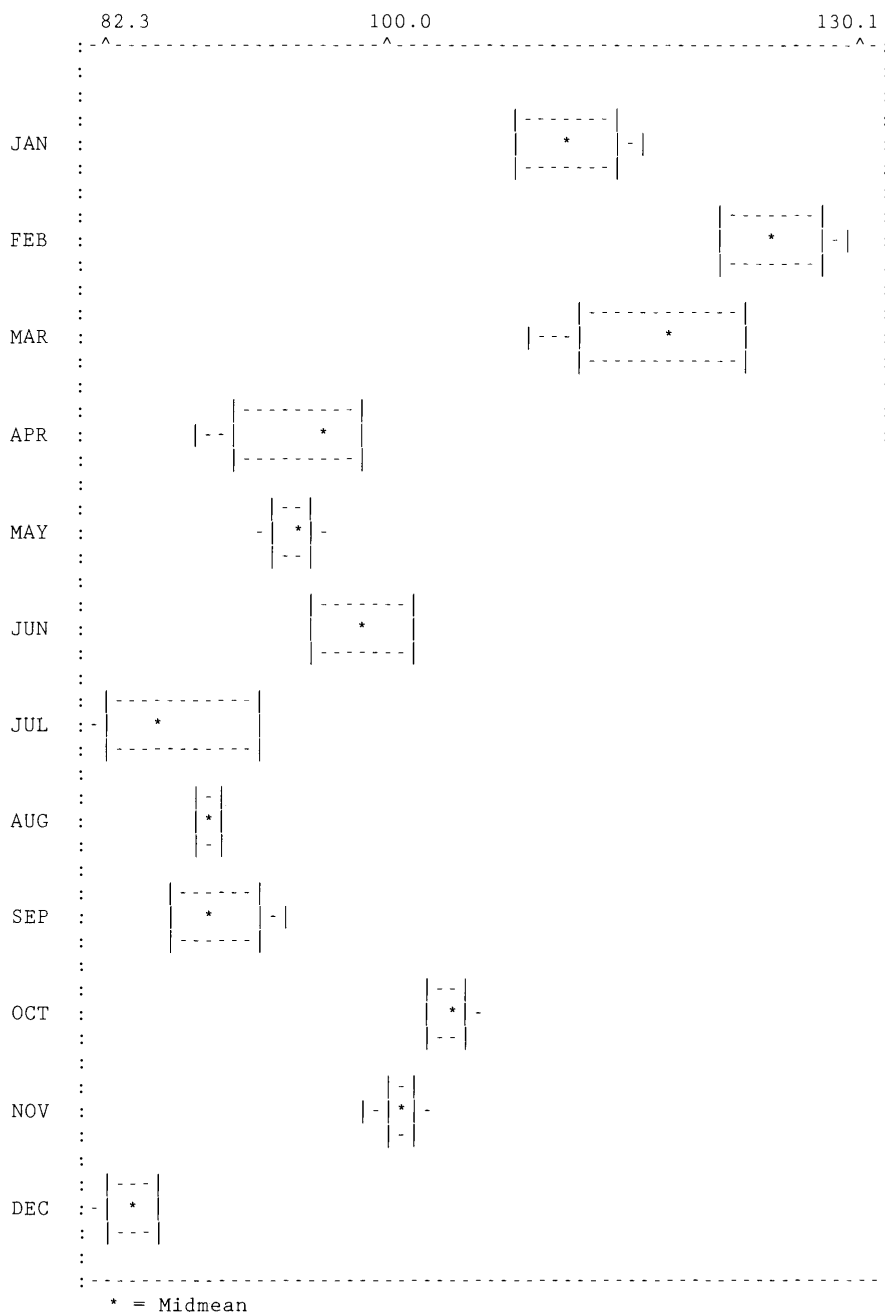


Figura 8.8 (continuación)



**El problema.** En el capítulo 4, ejemplo 4.4, se analizaron los datos de ventas trimestrales (tabla 4.5) de Outboard Marine. Desarrollar un modelo estacional multiplicativo para estos datos.

**La solución mediante hoja de cálculo.** Coloque el disco de Spreadsheet Forecaster en la unidad A. Entre a Lotus 1-2-3.

1. Para cargar la hoja de cálculo de *Spreadsheet Forecaster* que desarrolla modelos estacionales multiplicativos, escriba /FILE RETRIEVE o /FR. En la pantalla aparecerá el siguiente mensaje solicitando el nombre del archivo

Name of file to retrieve: A:\

En la tercera línea verá el nombre de la hoja de cálculo que desea emplear. Presione <Enter> después de señalar MULTIQTR.WK1. La figura 8.9 presenta los datos de muestra en una hoja de muestra.

2. Coloque el indicador de celda en la celda B3 y escriba OUTBOARD MARINE SALES. Coloque el indicador en la celda B5 y escriba DOLLARS.
3. Ubique el indicador en la celda B9 y escriba 1984. Nótese que el resto de los años cambian de manera automática.

|                    |                                                    |                                      |        |             |             |        |               |             |            |             |           |      |
|--------------------|----------------------------------------------------|--------------------------------------|--------|-------------|-------------|--------|---------------|-------------|------------|-------------|-----------|------|
| B2:                | [W4]                                               | "MULTIPLICATIVE SEASONAL ADJUSTMENT" |        |             |             |        |               |             |            |             |           | MENU |
| Worksheet          | Range                                              | Copy                                 | Move   | File        | Print       | Graph  | Data          | System      | Add-In     | Quit        |           |      |
| Global             | Insert                                             | Delete                               | Column | Erase       | Titles      | Window | Status        | Page        | Learn      |             |           |      |
|                    | A                                                  | B                                    | C      | D           | E           | F      | G             | H           | I          | J           | K         |      |
| 1                  | MULTIPLICATIVE SEASONAL ADJUSTMENT, QUARTERLY DATA |                                      |        |             |             |        |               |             |            |             |           |      |
| 2                  | TITLE1: MULTIPLICATIVE SEASONAL ADJUSTMENT         |                                      |        |             |             |        |               |             |            |             |           |      |
| 3                  | TITLE2: GAS GRILL SALES                            |                                      |        |             |             |        |               |             |            |             |           |      |
| 4                  | X-AXIS: QUARTER                                    |                                      |        |             |             |        |               |             |            |             |           |      |
| 5                  | Y-AXIS: UNITS                                      |                                      |        |             |             |        |               |             |            |             |           |      |
| 6                  |                                                    |                                      |        |             |             |        |               |             |            |             |           |      |
| 7                  |                                                    |                                      |        |             |             |        |               |             |            |             |           |      |
| 8                  | Year                                               | Qtr                                  | Per    | Actual data | Moving Avg. | Ratio  | Sum of ratios | # of ratios | Avg. ratio | Seas. index | Adj. data |      |
| 9                  | 1978                                               | 1                                    | 1      | 201         | 0.0         | 0.00   | 1.62          | 2           | 0.808      | 0.807       | 249.2     |      |
| 10                 |                                                    | 2                                    | 2      | 253         | 0.0         | 0.00   | 1.83          | 2           | 0.913      | 0.912       | 277.5     |      |
| 11                 |                                                    | 3                                    | 3      | 312         | 250.8       | 1.24   | 4.16          | 3           | 1.385      | 1.383       | 225.7     |      |
| 12                 |                                                    | 4                                    | 4      | 237         | 252.8       | 0.94   | 1.80          | 2           | 0.901      | 0.899       | 263.6     |      |
| 13                 | 1979                                               | 1                                    | 5      | 209         | 254.5       | 0.82   |               | Sum         | 4.008      | 4.000       | 259.1     |      |
| 14                 |                                                    | 2                                    | 6      | 260         | 275.0       | 0.95   |               |             |            |             | 285.2     |      |
| 15                 |                                                    | 3                                    | 7      | 394         | 276.3       | 1.43   |               |             |            |             | 285.0     |      |
| 16                 |                                                    | 4                                    | 8      | 242         | 280.0       | 0.86   |               |             |            |             | 269.1     |      |
| 17                 | 1980                                               | 1                                    | 9      | 224         | 281.8       | 0.80   |               |             |            |             | 277.7     |      |
| 18                 |                                                    | 2                                    | 10     | 267         | 303.0       | 0.88   |               |             |            |             | 292.9     |      |
| 19                 |                                                    | 3                                    | 11     | 479         | 322.5       | 1.49   |               |             |            |             | 346.4     |      |
| 20                 |                                                    | 4                                    | 12     | 320         | 0.0         | 0.00   |               |             |            |             | 355.8     |      |
| 04-Jan-94 03:17 PM |                                                    |                                      |        |             |             |        |               |             |            |             |           |      |
| NUM CAPS           |                                                    |                                      |        |             |             |        |               |             |            |             |           |      |

**Figura 8.9** Ajuste estacional multiplicativo de Spreadsheet Forecaster.

## 1-2-3 Spreadsheet Computing Series

MULTIPLICATIVE SEASONAL ADJUSTMENT, QUARTERLY DATA

TITLE1: MULTIPLICATIVE SEASONAL ADJUSTMENT

TITLE2: OUTBOARD MARINE SALES

X-AXIS: QUARTER

Y-AXIS: DOLLARS

|      |     |     | Actual | Moving |       | Sum of | # of   | Avg.  | Seas. | Adj.  |
|------|-----|-----|--------|--------|-------|--------|--------|-------|-------|-------|
| Year | Qtr | Per | data   | Avg.   | Ratio | ratios | ratios | ratio | index | data  |
| 1984 | 1   | 1   | 147.6  | 0.0    | 0.00  | 6.17   | 9      | 0.686 | 0.686 | 215.1 |
|      | 2   | 2   | 251.8  | 0.0    | 0.00  | 9.24   | 9      | 1.027 | 1.027 | 245.2 |
|      | 3   | 3   | 273.1  | 230.4  | 1.19  | 11.56  | 10     | 1.156 | 1.156 | 236.2 |
|      | 4   | 4   | 249.1  | 228.3  | 1.09  | 10.17  | 9      | 1.130 | 1.131 | 220.3 |
| 1985 | 1   | 5   | 139.3  | 220.7  | 0.63  |        | Sum    | 3.999 | 4.000 | 203.0 |
|      | 2   | 6   | 221.2  | 217.5  | 1.02  |        |        |       |       | 215.4 |
|      | 3   | 7   | 260.2  | 220.1  | 1.18  |        |        |       |       | 225.0 |
|      | 4   | 8   | 259.5  | 220.4  | 1.18  |        |        |       |       | 229.5 |
| 1986 | 1   | 9   | 140.5  | 226.4  | 0.62  |        |        |       |       | 204.8 |
|      | 2   | 10  | 245.5  | 236.1  | 1.04  |        |        |       |       | 239.1 |
|      | 3   | 11  | 298.8  | 243.0  | 1.23  |        |        |       |       | 258.4 |
|      | 4   | 12  | 287    | 250.0  | 1.15  |        |        |       |       | 253.8 |
| 1987 | 1   | 13  | 168.8  | 269.3  | 0.63  |        |        |       |       | 246.0 |
|      | 2   | 14  | 322.6  | 293.0  | 1.10  |        |        |       |       | 314.1 |
|      | 3   | 15  | 393.5  | 322.3  | 1.22  |        |        |       |       | 340.3 |
|      | 4   | 16  | 404.3  | 345.0  | 1.17  |        |        |       |       | 357.6 |
| 1988 | 1   | 17  | 259.7  | 364.7  | 0.71  |        |        |       |       | 378.5 |
|      | 2   | 18  | 401.1  | 382.4  | 1.05  |        |        |       |       | 390.6 |
|      | 3   | 19  | 464.6  | 401.3  | 1.16  |        |        |       |       | 401.8 |
|      | 4   | 20  | 479.7  | 402.5  | 1.19  |        |        |       |       | 424.3 |
| 1989 | 1   | 21  | 264.4  | 402.8  | 0.66  |        |        |       |       | 385.3 |
|      | 2   | 22  | 402.6  | 389.5  | 1.03  |        |        |       |       | 392.0 |
|      | 3   | 23  | 411.3  | 366.1  | 1.12  |        |        |       |       | 355.7 |
|      | 4   | 24  | 385.9  | 358.1  | 1.08  |        |        |       |       | 341.3 |
| 1990 | 1   | 25  | 232.7  | 334.8  | 0.70  |        |        |       |       | 339.1 |
|      | 2   | 26  | 309.2  | 309.6  | 1.00  |        |        |       |       | 301.1 |
|      | 3   | 27  | 310.7  | 286.4  | 1.08  |        |        |       |       | 268.7 |
|      | 4   | 28  | 293    | 279.5  | 1.05  |        |        |       |       | 259.1 |
| 1991 | 1   | 29  | 205.1  | 260.8  | 0.79  |        |        |       |       | 298.9 |
|      | 2   | 30  | 234.4  | 254.5  | 0.92  |        |        |       |       | 228.3 |
|      | 3   | 31  | 285.4  | 245.9  | 1.16  |        |        |       |       | 246.8 |
|      | 4   | 32  | 258.7  | 242.9  | 1.06  |        |        |       |       | 228.8 |
| 1992 | 1   | 33  | 193.2  | 250.3  | 0.77  |        |        |       |       | 281.6 |
|      | 2   | 34  | 263.7  | 252.0  | 1.05  |        |        |       |       | 256.8 |
|      | 3   | 35  | 292.5  | 266.2  | 1.10  |        |        |       |       | 253.0 |
|      | 4   | 36  | 315.2  | 262.4  | 1.20  |        |        |       |       | 278.8 |
| 1993 | 1   | 37  | 178.3  | 265.1  | 0.67  |        |        |       |       | 259.9 |
|      | 2   | 38  | 274.5  | 265.9  | 1.03  |        |        |       |       | 267.3 |
|      | 3   | 39  | 295.4  | 265.0  | 1.11  |        |        |       |       | 255.5 |
|      | 4   | 40  | 311.8  | 0.0    | 0.00  |        |        |       |       | 275.8 |

Figura 8.10 Salida de Lotus 1-2-3 para los datos de Outboard Marine con Spreadsheet Forecaster.

4. Posicione el indicador de celdas en la celda B9 y registre el primer valor de datos, 147.6. Oprima la flecha hacia abajo <↓> y continúe registrando los datos hasta ingresar el valor 311.8 en la celda D48. En la columna J se muestra el índice estacional final para cada trimestre. Los datos ajustados en forma estacional aparecen en la columna K.
5. Para almacenar la hoja de cálculo para su uso futuro, escriba /FILE SAVE o /FS, y el programa le solicitará el nombre del archivo

```
Enter name of file to save: A:*.WK1
```

Asegúrese de que el directorio actual es el correcto. Escriba OUTBOARD y presione <Enter>.

6. Para imprimir la hoja de cálculo, escriba /PRINT PRINTER RANGE o /PPR y aparecerá el mensaje que le solicita el rango a imprimir:

```
Enter print range:
```

El rango correspondiente es A1..K48. oprima <Enter> y después GO. Los resultados se muestran en la figura 8.10.

## PAQUETE DE CÓMPUTO MINITAB

**El problema.** El problema 20 requiere que el analista calcule los cuatro índices estacionales para las ventas de Goodyear Tire.

### La solución mediante Minitab.

```
MTB > READ 'PROB20.DAT' C1
C1
 2292 2450 2363 2477 . . .
MTB > LET C2 = C1
MTB > LET C3 = C1
MTB > LET C4 = C1
MTB > DELETE 1 C2
MTB > DELETE 1:2 C3
MTB > DELETE 1:3 C4
MTB > LET C5 = C1+C2+C3+C4
 I I I
*** UNEQUAL COLUMN LENGTHS DURING OPERATION
RESULT HAS SHORTER LENGTH
```

El primer paso consiste en calcular un total móvil de cuatro trimestres. Los datos se copian en las columnas 2, 3 y 4. Enseguida, se usa el comando DELETE para eliminar la primera observación en la columna 2, las primeras dos observaciones en la columna 3 y las primeras tres observaciones en la columna 4. Luego, se emplea el comando SET para agregar las columnas 1 a 4 y almacenar el resultado del total móvil de cuatro trimestres en C5. En este punto, Minitab proporciona una advertencia indicando que las columnas son desiguales en longitud. Esta advertencia se ignora a lo largo de este proceso. A continuación, se utiliza el mismo tipo de idea para calcular un total móvil de dos años y, eventualmente, los índices estacionales.

```
MTB > LET C6 = C5
MTB > DELETE 1 C6
MTB > LET C7 = C5+C6
MTB > PRINT C1-C7
```

| ROW | C1   | C2   | C3   | C4   | C5    | C6    | C7    |
|-----|------|------|------|------|-------|-------|-------|
| 1   | 2292 | 2450 | 2363 | 2477 | 9582  | 9353  | 18935 |
| 2   | 2450 | 2363 | 2477 | 2063 | 9353  | 9261  | 18614 |
| 3   | 2363 | 2477 | 2063 | 2358 | 9261  | 9214  | 18475 |
| 4   | 2477 | 2063 | 2358 | 2316 | 9214  | 9103  | 18317 |
| 5   | 2063 | 2358 | 2316 | 2366 | 9103  | 9308  | 18411 |
| 6   | 2358 | 2316 | 2366 | 2268 | 9308  | 9483  | 18791 |
| 7   | 2316 | 2366 | 2268 | 2533 | 9483  | 9646  | 19129 |
| 8   | 2366 | 2268 | 2533 | 2479 | 9646  | 9905  | 19551 |
| 9   | 2268 | 2533 | 2479 | 2625 | 9905  | 10253 | 20158 |
| 10  | 2533 | 2479 | 2625 | 2616 | 10253 | 10513 | 20766 |
| 11  | 2479 | 2625 | 2616 | 2793 | 10513 | 10690 | 21203 |
| 12  | 2625 | 2616 | 2793 | 2656 | 10690 | 10811 | 21501 |
| 13  | 2616 | 2793 | 2656 | 2746 | 10811 | 10838 | 21649 |
| 14  | 2793 | 2656 | 2746 | 2643 | 10838 | 10856 | 21694 |
| 15  | 2656 | 2746 | 2643 | 2811 | 10856 | 10879 | 21735 |
| 16  | 2746 | 2643 | 2811 | 2679 | 10879 | 10869 | 21748 |
| 17  | 2643 | 2811 | 2679 | 2736 | 10869 | 10918 | 21787 |
| 18  | 2811 | 2679 | 2736 | 2692 | 10918 | 10978 | 21896 |
| 19  | 2679 | 2736 | 2692 | 2871 | 10978 | 11199 | 22177 |
| 20  | 2736 | 2692 | 2871 | 2900 | 11199 | 11274 | 22473 |
| 21  | 2692 | 2871 | 2900 | 2811 | 11274 | 11079 | 22353 |
| 22  | 2871 | 2900 | 2811 | 2497 | 11079 | 11000 | 22079 |
| 23  | 2900 | 2811 | 2497 | 2792 | 11000 | 10938 | 21938 |
| 24  | 2811 | 2497 | 2792 | 2838 | 10938 | 10907 | 21845 |
| 25  | 2497 | 2792 | 2838 | 2780 | 10907 | 11188 | 22095 |
| 26  | 2792 | 2838 | 2780 | 2778 | 11188 | 11462 | 22650 |
| 27  | 2838 | 2780 | 2778 | 3066 | 11462 |       |       |
| 28  | 2780 | 2778 | 3066 |      |       |       |       |
| 29  | 2778 | 3066 |      |      |       |       |       |
| 30  | 3066 |      |      |      |       |       |       |

```
MTB > LET C8 = C7/8
MTB > LET C9 = (C3/C8)*100
MTB > PRINT C9
```

El promedio móvil centrado de cuatro trimestres (C8) se calcula dividiendo el total móvil de dos años (C7) almacenado en la columna 7, entre 8. Cada índice estacional trimestral se calcula dividiendo el valor real de cada trimestre (C3) entre el promedio móvil centrado de cuatro trimestres para ese trimestre y multiplicando por 100, de modo que el resultado sea un número índice. Los siguientes comandos se emplean para calcular las medianas de cada trimestre.

```
MTB > SET C10
DATA> 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
DATA> END
MTB > DESCRIBE C9;
SUBC> BY C10.
```

|    | C10 | N | MEDIAN |
|----|-----|---|--------|
| C9 | 1   | 6 | 95.98  |
|    | 2   | 6 | 103.19 |
|    | 3   | 7 | 99.84  |
|    | 4   | 7 | 101.13 |

```
MTB > STOP
```

El comando SET se usa para identificar cada trimestre. Los comandos DESCRIBE y BY se utilizan para calcular varias estadísticas de C9 para cada valor de C10. Sólo es relevante la parte de la salida que se muestra aquí. El último paso, que no se muestra, consiste en ajustar estos valores de modo que sumen 400.

## BIBLIOGRAFÍA

- BURMAN, J. P. "Seasonal Adjustment -A Survey." *TIMS Studies in Management Sciences* 12 (1979): 45-57.
- CROXTON, F. E., COWDEN, D. J. y BLOCH, B. W. *Practical Business Statistics*. Englewood Cliffs, N. J.: Prentice-Hall, 1969.
- MAKRIDAKIS, S. y WHEELWRIGHT, S. C. *Interactive Forecasting*. Palo Alto, Calif.: Scientific Press, 1977.
- MOORE, G. H. y SHISKIN, J. "Indicators of Business Expansions and Contractions." New York: National Bureau of Economic Research, 1957, pp. 8-28.
- MOORE, G. H. y SHISKIN, J. "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, p. 81.

## Regresión de datos de series de tiempo

### ACTIVIDAD

Muchas aplicaciones en la economía y en los negocios comprenden series de tiempo. Se puede llevar a cabo la regresión de datos mensuales, trimestrales o anuales mediante las técnicas descritas en los capítulos anteriores. Sin embargo, las mediciones de series de tiempo no pueden considerarse muestras probabilísticas; en vez de ello, están sujetas a tendencias, ciclos, variaciones estacionales e irregulares y a fluctuaciones aleatorias. De ahí que con frecuencia surjan problemas de interpretación.

Los modelos básicos de regresión expuestos en los caps. 6 y 7 suponen que los residuos ( $Y - \hat{Y}$ ) son variables aleatorias no correlacionadas o bien, variables aleatorias independientes normales. Por ello se supone que los términos de error sucedan de manera aleatoria. Esta suposición no sólo es válida para la mayoría de las series debido a que los términos de error tienden a incrementarse siempre que se relaciona a través del tiempo a una variable tomada de una industria creciente. Los términos de error se hacen mayores debido a que la variable o variables que comprende tienden a crecer en una tasa constante en vez de hacerlo en una cantidad constante. Esta condición se denomina *heteroscedasticidad* y se abordó en el cap. 7.

### EL PROBLEMA DE LA HETEROSCEDASTICIDAD EN LA REGRESIÓN DE DATOS DE SERIES DE TIEMPO

Para abordar este tema emplearemos un ejemplo.

#### Ejemplo 9.1

Suponga que un analista está comprometido con la planeación a futuro de Reynolds Metals Co., un productor de aluminio, y desea establecer una base cuantitativa para proyectar ventas futuras. Como la compañía vende en forma regional, una medida del gasto personal disponible por región tendrá una relación cercana. La tabla 9.1 muestra las ventas y el ingreso para el periodo 1973 – 1993. El analista correlaciona estas variables e intenta utilizar el ingreso para pronosticar las ventas de Reynolds para 1994.

Primero, hace una gráfica de los datos en una escala aritmética, como se muestra en la fig. 9.1. La relación parece ser lineal.

Al ejecutar los datos en un programa de cómputo, los resultados se muestran en la tabla 9.2. El ajuste es bueno; el 90% de la varianza de la variable ventas se explica mediante la ecuación lineal  $Y = -792 + 4.22517X$ . Sin embargo, una revisión minuciosa de la fig. 9.1 indica que los residuos satisfacen la suposición de varianza constante. Los residuos son

**TABLA 9.1 DATOS PARA EL EJEMPLO 9.1: VENTAS DE REYNOLDS METALS**

| AÑO  | VENTAS DE<br>REYNOLDS METALS<br>(MILES DE DÓLARES)<br>Y | INGRESO PERSONAL<br>DISPONIBLE<br>(MILLONES DE DÓLARES)<br>X | VENTAS DE<br>REYNOLDS METALS<br>RETRASADAS UN AÑO |
|------|---------------------------------------------------------|--------------------------------------------------------------|---------------------------------------------------|
| 1973 | 295                                                     | 273.4                                                        | —                                                 |
| 1974 | 400                                                     | 291.3                                                        | 295                                               |
| 1975 | 390                                                     | 306.9                                                        | 400                                               |
| 1976 | 425                                                     | 317.1                                                        | 390                                               |
| 1977 | 547                                                     | 336.1                                                        | 425                                               |
| 1978 | 555                                                     | 349.4                                                        | 547                                               |
| 1979 | 620                                                     | 362.9                                                        | 555                                               |
| 1980 | 720                                                     | 383.9                                                        | 620                                               |
| 1981 | 880                                                     | 402.8                                                        | 720                                               |
| 1982 | 1,050                                                   | 437.0                                                        | 880                                               |
| 1983 | 1,290                                                   | 472.2                                                        | 1,050                                             |
| 1984 | 1,528                                                   | 510.4                                                        | 1,290                                             |
| 1985 | 1,586                                                   | 544.5                                                        | 1,528                                             |
| 1986 | 1,960                                                   | 588.1                                                        | 1,586                                             |
| 1987 | 2,118                                                   | 630.4                                                        | 1,960                                             |
| 1988 | 2,116                                                   | 685.9                                                        | 2,118                                             |
| 1989 | 2,477                                                   | 742.8                                                        | 2,116                                             |
| 1990 | 3,199                                                   | 801.3                                                        | 2,477                                             |
| 1991 | 3,702                                                   | 903.1                                                        | 3,199                                             |
| 1992 | 3,316                                                   | 983.6                                                        | 3,702                                             |
| 1993 | 2,702                                                   | 1,076.7                                                      | 3,316                                             |

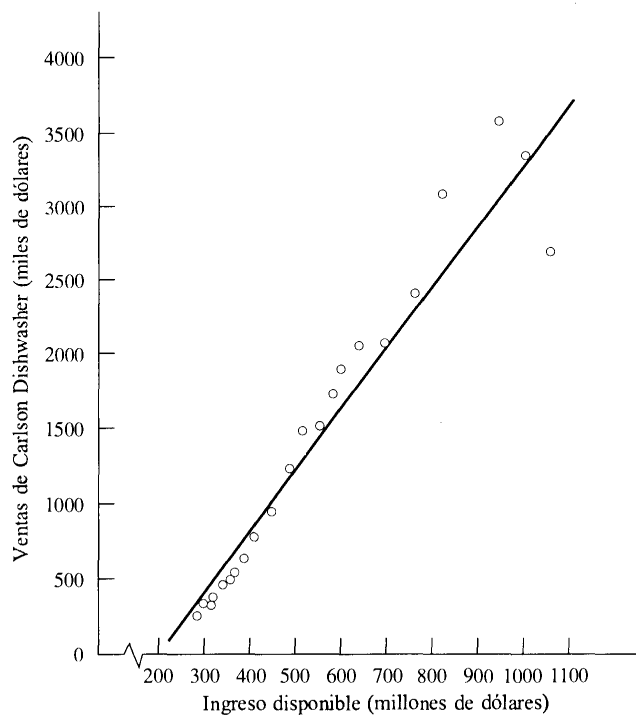


**Figura 9.1** Ajuste lineal de las ventas de Reynolds Metals.

TABLA 9.2 SALIDA DE CÓMPUTO PARA LAS VENTAS DE REYNOLDS METALS: AJUSTE LINEAL

| VARIABLE NO.                            | MEAN       | STANDARD DEVIATION | CORRELATION X VS Y | REGRESSION COEFFICIENT | STD. ERROR OF REG. COEF. | COMPUTED T VALUE |
|-----------------------------------------|------------|--------------------|--------------------|------------------------|--------------------------|------------------|
| 2                                       | 542.85     | 241.16             | 0.95               | 4.22517                | 0.31627                  | 13.45            |
| DEPENDENT                               |            |                    |                    |                        |                          |                  |
| 1                                       | 1517.90450 | 1078.69840         |                    |                        |                          |                  |
| INTERCEPT                               |            | -792               |                    | MULTIPLE CORRELATION   |                          | 0.951            |
| STD. ERROR OF ESTIMATE                  |            | 341.1              |                    | R SQUARE               |                          | 0.905            |
| DURBIN-WATSON STATISTIC                 |            | 0.87               |                    | CORRECTED R SQUARED    |                          | 0.900            |
| ANALYSIS OF VARIANCE FOR THE REGRESSION |            |                    |                    |                        |                          |                  |
| SOURCE OF VARIATION                     |            | DEGREES OF FREEDOM | SUM OF SQUARES     | MEAN SQUARES           | F VALUE                  |                  |
| ATTRIBUTABLE TO REGRESSION              |            | 1                  | 21061184           | 21061184               | 181.018                  |                  |
| DEVIATION FROM REGRESSION               |            | 19                 | 2210624            | 116348                 |                          |                  |
| TOTAL                                   |            | 20                 | 23271808           |                        |                          |                  |
| TABLE OF RESIDUALS                      |            |                    |                    |                        |                          |                  |
| CASE NO.                                | Y VALUE    | Y ESTIMATE         | RESIDUAL           |                        |                          |                  |
| 1                                       | 295        | 371.36             | -76.36             |                        |                          |                  |
| 2                                       | 400        | 447.53             | -47.53             |                        |                          |                  |
| 3                                       | 390        | 513.91             | -123.91            |                        |                          |                  |
| 4                                       | 425        | 557.32             | -132.32            |                        |                          |                  |
| 5                                       | 547        | 638.16             | -91.16             |                        |                          |                  |
| 6                                       | 555        | 694.76             | -139.76            |                        |                          |                  |
| 7                                       | 620        | 752.20             | -132.20            |                        |                          |                  |
| 8                                       | 720        | 841.56             | -121.56            |                        |                          |                  |
| 9                                       | 880        | 921.98             | -41.98             |                        |                          |                  |
| 10                                      | 1050       | 1067.51            | -17.51             |                        |                          |                  |
| 11                                      | 1290       | 1217.29            | 72.71              |                        |                          |                  |
| 12                                      | 1528       | 1379.84            | 148.16             |                        |                          |                  |
| 13                                      | 1586       | 1524.94            | 61.06              |                        |                          |                  |
| 14                                      | 1960       | 1710.47            | 249.53             |                        |                          |                  |
| 15                                      | 2118       | 1890.46            | 227.54             |                        |                          |                  |
| 16                                      | 2116       | 2126.62            | -10.62             |                        |                          |                  |
| 17                                      | 2477       | 2368.74            | 108.26             |                        |                          |                  |
| 18                                      | 3199       | 2617.67            | 581.33             |                        |                          |                  |
| 19                                      | 3702       | 3050.84            | 651.16             |                        |                          |                  |
| 20                                      | 3316       | 3393.38            | -77.38             |                        |                          |                  |
| 21                                      | 2702       | 3789.54            | -1087.54           |                        |                          |                  |



mucho mayores al crecer las ventas a través de los años. Este resultado se presenta debido a que las variables se incrementan en una tasa constante.

Una solución al problema de heteroscedasticidad consiste en expresar en dólares constantes los valores tanto de la variable dependiente como de la independiente. La expresión en dólares constantes de una serie de tiempo (véase exposición sobre deflación de precios en el cap. 8), con frecuencia supera el problema del incremento en los errores de pronóstico a través del tiempo. Otra solución es realizar una transformación logarítmica (véase la exposición en el cap. 6).

## **EL PROBLEMA DE CORRELACIÓN SERIAL DURANTE LA REGRESIÓN DE DATOS DE SERIES DE TIEMPO**

Otro problema con las de series de tiempo de datos es su movimiento en ciclos más bien que en una forma puramente aleatoria. Es frecuente que este fenómeno provoque resultados de varios errores positivos o negativos en secuencia. Este punto se ilustra revisando la fig. 9.1. Los primeros diez residuos (1973–1982) son negativos, lo que significa que las diez primeras estimaciones o predicciones basadas en la ecuación de regresión son todas muy altas. Los siguientes cinco residuos (1983–1987) son positivos, lo que representa que todas estas predicciones son muy bajas. El valor de cada año depende o está relacionado con el valor del año anterior, en vez de que sea independiente de él. Este hecho no debe sorprender. Considere, por ejemplo, una serie de precios. Si las observaciones sucesivas de un año al siguiente fueran en realidad independientes entre sí, viviríamos en una economía caótica. En un mundo así, los precios se determinarían como números extraídos de una tabla de números aleatorios. El conocimiento del precio en un año no ejercería influencia en el precio del año siguiente.

Los datos económicos ordenados en una serie de tiempo pueden considerarse como una muestra aleatoria. Una observación en un periodo dado sobre precios, inventarios, producción, acciones y otras variables económicas, está por lo regular correlacionada con (dependiente de) el valor de la misma variable en el periodo anterior. Para describir esta situación se emplea el término *correlación serial*. Los residuos ( $Y - \hat{Y}$ ) no son independientes entre una observación y la siguiente. De modo que el conocimiento del error en un año ayuda al analista a anticipar el error en el año siguiente. Los residuos están autocorrelacionados.

Existe *correlación serial* cuando las observaciones sucesivas a través del tiempo se encuentran relacionadas entre sí.

La forma más común de correlación serial es de primer orden, en el que cada término de error es una función del término de error del periodo anterior. Esta situación se ilustra en la ecuación 9.1.

$$Y_t = \beta_0 + \beta X_t + \epsilon_t \quad (9.1)$$

en donde  $\epsilon_t$  depende de su propio valor en el periodo anterior o

$$\epsilon_t = \rho \epsilon_{t-1} + v_t \quad (9.2)$$

en donde

$\epsilon_t$  = término de error de la ecuación

$\rho$  = parámetro (coeficiente de correlación serial de primer orden) que mide la correlación entre los términos de error

$v_t$  = término de error independiente con distribución normal

Para este tipo de correlación serial, todo lo que se necesita es que el nivel de un término de error ( $\epsilon_{t-1}$ ) afecte directamente el nivel del siguiente término de error ( $\epsilon_t$ ). La magnitud del coeficiente de correlación de primer orden ( $\rho$ ), indica la fuerza de la correlación serial en la ecuación 9.2. Si  $\rho$  es cero, entonces no existe correlación serial y los términos de error son independientes ( $\epsilon_t = v_t$ ).

Considere los datos que se presentan en la fig. 9.2, que grafica una serie de tiempo ficticia. La línea recta es una línea de regresión con respecto al tiempo. La correlación serial es obvia de inmediato. De 1983 a 1986 los términos de error ( $Y - Y$ ) son todos positivos (todas las estimaciones de la regresión son muy bajas), mientras que de 1987 a 1994 son negativos (todas las estimaciones de la regresión son muy altas). Se anticipa un valor negativo para 1995. Resulta fácil predecir la dirección de cada término de error sucesivo.

En la economía y en los negocios, una de las principales causas de residuos positivos correlacionados en forma serial es una especificación de error tal, semejante a omitir una variable o a una forma funcional incorrecta. Cuando se relacionan en forma positiva los efectos de secuencia en el tiempo de una variable "faltante", los residuos tienden a ser correlacionados serialmente en forma positiva en la ecuación de regresión, debido a que incluyen los efectos de la variable faltante. De ahí que no haya sido explicada una parte importante de la variación de la variable dependiente.

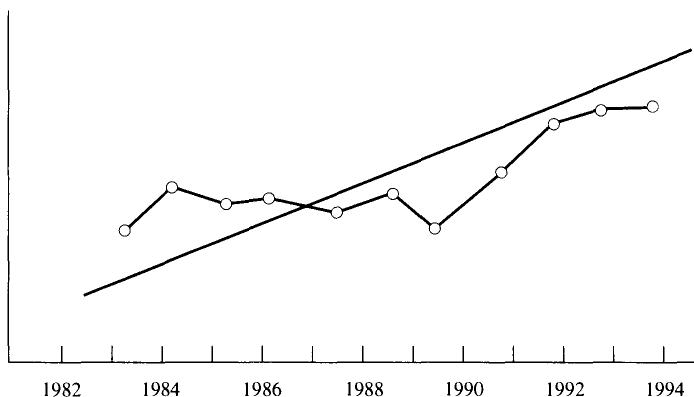


**Figura 9.2** Correlación serial positiva.

Por ejemplo, considere la variable de registros de automóviles nuevos que se expuso en el cap. 8. Si se emplean como variables de predicción el ingreso personal y el precio, el modelo pudiera explicar una gran parte de la varianza de los registros de automóviles nuevos. Por desgracia, podría estar presente una correlación serial. La especificación del modelo dejó fuera variables potencialmente clave como la población en edad de conducir, que tiene un impacto definitivo en la venta de automóviles y, en consecuencia, en su registro.

Si los residuos en una ecuación de regresión están autocorrelacionados en forma positiva, el uso del método de mínimos cuadrados presenta varios problemas.

1. El error estándar de la estimación subestima seriamente la variabilidad de los términos de error.
2. Los intervalos de confianza y pruebas que emplean las distribuciones  $t$  y  $F$  ya no son estrictamente aplicables.
3. El error estándar del coeficiente de regresión subestima la variabilidad del coeficiente de regresión estimado.

La fig. 9.3 ilustra la presencia de una correlación serial positiva en un modelo con una sola variable independiente. El residuo asociado con la primera observación, indicado en la gráfica por  $\hat{Y} - Y$ , resulta ser positivo. Este resultado conduce a series de cinco términos positivos de error y a una serie de cuatro residuos negativos. Nótese que la pendiente de regresión estimada es menor que la pendiente real. La línea de regresión, ajustada mediante mínimos cuadrados, se ajusta más a los puntos de datos que la línea de regresión verdadera. Este resultado nos lleva a una  $r^2$  que es artificialmente alta. Aún más, el error estándar de la estimación será menor que el error estándar real. Si para realizar pruebas estadísticas se

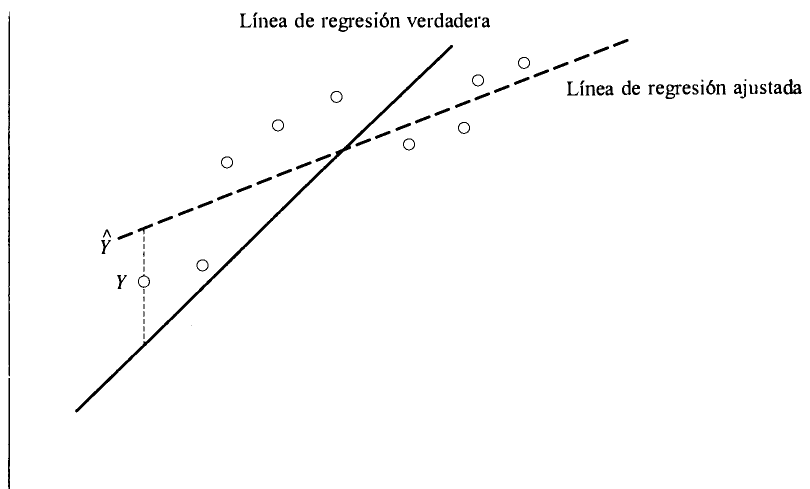


**Figura 9.3** Correlación serial positiva y procedimiento de mínimos cuadrados.

emplea el error estándar de mínimos cuadrados, se sobreestimaré el éxito del procedimiento de regresión.

Cuando existe una correlación serial positiva y el primer residuo es positivo (véase fig. 9.3), la línea de regresión estimada o ajustada tendrá una pendiente subestimada (muy pequeña) y una intersección sobrestimada (muy grande). Si, por el contrario, el primer residuo es negativo, la línea de regresión estimada o ajustada tendrá una pendiente sobrestimada y una intersección subestimada. En ambos casos la línea de regresión de mínimos cuadrados se ajusta más a los puntos de datos observados que la línea de regresión real. El problema no consiste en una desviación sino en una gran varianza en la estimación. No es una falla el procedimiento de mínimos cuadrados. Cualquier otro procedimiento de estimación (como el ajuste visual), se ajustaría en la misma forma al único patrón de datos oscilante. En una sección posterior se abordarán técnicas que podrían mejorar el procedimiento de estimación.

## PRUEBA DE DURBIN-WATSON PARA CORRELACIÓN SERIAL

Un enfoque que se usa con frecuencia para determinar si existe correlación serial es la *prueba de Durbin – Watson*. La prueba comprende la determinación de si el parámetro de autocorrelación  $\rho$  que se muestra en la ecuación 9.2, es igual a cero.

$$\epsilon_t = \rho\epsilon_{t-1} + v_t$$

Las hipótesis a considerar son

$$H_0: \rho = 0$$

$$H_1: \rho > 0$$

La hipótesis alternativa es  $\rho > 0$ , ya que los residuos en las aplicaciones de series de tiempo tienden a mostrar una correlación positiva.

El primer paso en el cálculo consiste en ajustar a los datos la línea de regresión de mínimos cuadrados. A continuación se calculan los residuos y la estadística de Durbin–Watson

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (9.3)$$

en donde

$e_t$  = error o diferencia entre punto y línea  
 $e_{t-1}$  = error o diferencia entre punto y línea  
 para el periodo anterior

$$\sum_{t=2}^n (e_t - e_{t-1})^2 = \text{diferencia entre el residuo actual y el anterior, elevados al cuadrado y sumados para todas las observaciones.}$$

$$\sum_{t=1}^n e_t^2 = \text{cada uno de los residuos elevados al cuadrado y sumados después.}$$

Aunque no hay disponible un procedimiento de prueba exacto, Durbin y Watson han proporcionado las fronteras inferior ( $I$ ) y superior ( $S$ ), de manera que se pueda efectuar una prueba de correlación serial después de calcular el valor  $DW$ . Las reglas de decisión son

1. Cuando la estadística de Durbin–Watson es mayor que la frontera superior ( $S$ ), el coeficiente de autocorrelación es igual a cero (no existe autocorrelación positiva).
2. Cuando la estadística de Durbin–Watson es menor que la frontera inferior ( $I$ ), el coeficiente de autocorrelación es mayor que cero (existe autocorrelación positiva).
3. Cuando la estadística de Durbin–Watson se ubica entre las fronteras inferior y superior, la prueba no ofrece una conclusión (no sabemos si existe autocorrelación positiva).<sup>1</sup>

La *prueba de Durbin–Watson* se emplea para determinar si existe una autocorrelación positiva.

Si  $DW > S$ , concluir  $H_0$ .

Si  $DW < I$ , concluir  $H_1$ .

Si  $DW$  está entre las fronteras superior e inferior

( $I \leq DW \leq S$ ), la prueba no ofrece conclusión.

Las fronteras críticas de  $S$  e  $I$  están dadas en la tabla C.7 del Apéndice C. Para encontrar los valores apropiados de  $S$  e  $I$ , el analista necesita conocer el tamaño de la muestra, el nivel de significación y el número de variables independientes. En la tabla de Durbin–Watson del Apéndice C, el tamaño de la muestra aparece en la columna de la izquierda y el número de variables independientes está determinado a partir de la parte superior de cada columna. Si, por ejemplo se emplearan tres variables independientes, habría que buscar en la columna  $p - 1 = 3$ .

Cuando no hay conclusión en los resultados, se necesita de un mayor número de observaciones. Pudiera no ser posible cubrir este requerimiento con los datos de la serie de tiempo, de modo que Durbin y Watson ofrecen una prueba aproximada que requiere de un mínimo de 40 observaciones.

<sup>1</sup> J. Durbin y G.S. Watson, “Testing for serial Correlation in Least-Squares Regression II”, *Biometrika* 38 (1951): 159-178.

**Ejemplo 9.2**

En la tabla 9.3 aparecen los cálculos que muestran la implementación de la ecuación 9.3. Los datos corresponden a Reynolds Metals (ejemplo 9.1, tabla 9.1). La columna de residuos se obtuvo de la salida de cómputo de la tabla 9.2. Los cálculos para 1974, de las otras tres columnas son los siguientes:

$$\begin{aligned}e_t - e_{t-1} &= -47.53 - (-76.36) = 28.83 \\(e_t - e_{t-1})^2 &= 28.83^2 = 831.17 \\e_t^2 &= -47.53^2 = 2,259.1\end{aligned}$$

La estadística de Durbin – Watson se calcula como

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{1,926,035.14}{2,210,641.78} = .87$$

Esta respuesta coincide con la salida de cómputo de la tabla 9.2. Utilizando el nivel de significancia de .01 para una muestra de 21 y una variable independiente, se obtiene

$$I = .97$$

$$S = 1.16$$

Como  $DW = .87$ , se ubica por debajo de  $I = .97$ , se rechaza la hipótesis nula y se concluye que los residuos están autocorrelacionados en forma positiva ( $\rho > 0$ ).

**TABLA 9.3** CÁLCULOS DE DURBIN– WATSON PARA EL EJEMPLO 9.2:  
VENTAS DE REYNOLDS METALS

| AÑO     | VENTAS,<br>$Y$ | INGRESO,<br>$X$ | RESIDUOS,<br>$e_t$ | $e_t - e_{t-1}$ | $(e_t - e_{t-1})^2$ | $e_t^2$      |
|---------|----------------|-----------------|--------------------|-----------------|---------------------|--------------|
| 1973    | 295            | 273.4           | -76.36             | —               | —                   | 5,830.85     |
| 1974    | 400            | 291.3           | -47.53             | 28.83           | 831.17              | 2,259.10     |
| 1975    | 390            | 306.9           | -123.91            | -76.38          | 5,833.90            | 15,353.69    |
| 1976    | 425            | 317.1           | -132.32            | -8.41           | 70.73               | 17,508.58    |
| 1977    | 547            | 336.1           | -91.16             | 41.16           | 1,694.15            | 8,310.15     |
| 1978    | 555            | 349.4           | -139.76            | -48.60          | 2,361.96            | 19,532.86    |
| 1979    | 620            | 362.9           | -132.20            | 7.56            | 57.15               | 17,476.84    |
| 1980    | 720            | 383.9           | -121.56            | 10.64           | 113.21              | 14,776.83    |
| 1981    | 880            | 402.8           | -41.98             | 79.58           | 6,332.98            | 1,762.32     |
| 1982    | 1,050          | 437.0           | -17.51             | 24.47           | 598.78              | 306.60       |
| 1983    | 1,290          | 472.2           | 72.71              | 90.22           | 8,139.65            | 5,286.74     |
| 1984    | 1,528          | 510.4           | 148.16             | 75.45           | 5,692.70            | 21,951.39    |
| 1985    | 1,586          | 544.5           | 61.06              | -87.10          | 7,586.41            | 3,728.32     |
| 1986    | 1,960          | 588.1           | 249.53             | 188.47          | 35,520.94           | 62,265.22    |
| 1987    | 2,118          | 630.4           | 227.54             | -21.99          | 483.56              | 51,774.45    |
| 1988    | 2,116          | 685.9           | -10.62             | -238.16         | 56,720.19           | 112.78       |
| 1989    | 2,477          | 742.8           | 108.26             | 118.88          | 14,132.45           | 11,720.23    |
| 1990    | 3,199          | 801.3           | 581.33             | 473.07          | 223,795.22          | 337,944.57   |
| 1991    | 3,702          | 903.1           | 651.16             | 69.83           | 4,876.23            | 424,009.35   |
| 1992    | 3,316          | 983.6           | -77.38             | -728.54         | 530,770.53          | 5,987.66     |
| 1993    | 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 |

Nótese en la tabla 9.3 que los términos adyacentes de error  $e_t$  y  $e_{t-1}$  tienden a ser del mismo signo y magnitud. Cuando existe una correlación serial positiva, la diferencia  $(e - e_{t-1})$  tiende a ser muy pequeña, lo que lleva a tener un numerador pequeño en la estadística  $DW$ .

## **SOLUCIÓN A PROBLEMAS DE CORRELACIÓN SERIAL**

Una vez que se ha descubierto la correlación serial en la regresión de una serie de tiempo de datos, es necesario eliminarla antes de poder evaluar la efectividad de la ecuación de regresión. El método apropiado para eliminar la correlación serial depende, en primer lugar, de qué fue lo que la causó. La correlación serial puede surgir debido a un error de especificación, tal como la omisión de una variable, o puede aparecer porque los términos independientes de error se encuentran intercorrelacionados en una ecuación correctamente especificada.

La solución al problema de la correlación serial comienza con una evaluación de la especificación del modelo. ¿Es correcta la forma funcional? ¿Se omitió alguna variable importante? ¿Existen errores de especificación que pudieran tener cierto patrón a través del tiempo, que pudieran haber introducido una correlación serial en los residuos? Como en los negocios, una de las causas principales de los residuos autocorrelacionados en forma positiva es la omisión de una o más variables clave, el mejor enfoque para resolver el problema consiste en encontrarlas. A esta técnica se le conoce en ocasiones como mejoramiento de la especificación del modelo. Por desgracia, no siempre es posible utilizar este método. La variable faltante, aunque conocida para el analista, pudiera no ser cuantificable. Por ejemplo, se podría esperar que la inversión en negocios, en periodos futuros, estuviera relacionada con la actitud de los inversionistas potenciales. Sin embargo, resulta difícil cuantificar la variable “actitud”. No obstante, siempre que sea posible, se debe especificar el modelo de acuerdo con una visión del sentido teórico. El problema de la autocorrelación serial no se resolverá mediante la aplicación de ninguna técnica correctiva a un modelo sin sentido teórico.

Sólo se puede considerar la posibilidad de un ajuste, después de haber revisado minuciosamente la especificación de la ecuación. Se abordarán varias técnicas para eliminar la correlación serial.

Un enfoque para la eliminación de la correlación serial, denominado regresión de cambios porcentuales, genera nuevas variables mediante el uso de cambios porcentuales (incremento o disminución), entre un periodo y otro. Para demostrar esta generación de nuevas variables, se emplean los datos de Reynolds Metals de la tabla 9.1. La cantidad en que cambió  $Y$  entre 1973 y 1974 fue de 105. Como porcentaje de los valores previos de  $Y$  (295), este cambio representa un incremento porcentual de  $105/295 = .356$  o 35.6%. Este valor es la nueva variable independiente para el periodo 2 (1974). Ya que no se emplean los valores reales de las variables, la correlación serial será eliminada. Por desgracia, el cambio porcentual de la variable independiente pudiera no funcionar muy bien para explicar el cambio porcentual en la variable dependiente.

Un segundo enfoque para eliminar la correlación serial, la creación de un modelo autorregresivo, genera una nueva variable de pronóstico, utilizando la variable  $Y$  retrasada

uno o más periodos. En el ejemplo de Reynolds Metals la variable de pronóstico para 1974 es las ventas de 1973 (295).

Otro enfoque para la eliminación de la correlación serial, denominado primera diferenciación, genera nuevas variables que utilizan las diferencias reales (incrementos o disminuciones) entre periodo y periodo. De nuevo se usan los datos de Reynolds Metals de la tabla 9.1. La cantidad en que cambió  $Y$  de 1973 a 1974 fue de 105. Este valor es la nueva variable independiente para el periodo 2 (1974). El enfoque de primera diferenciación supone que la relación entre los términos de error medidos mediante el coeficiente de autocorrelación es 1.

También, el enfoque iterativo para eliminar la correlación serial, genera nuevas variables que utilizan las diferencias reales (incremento o disminución) entre un periodo y otro. El enfoque iterativo estima el coeficiente de autocorrelación o relación entre los términos de error.

A continuación se presenta un ejemplo de cada uno de los enfoques para la eliminación de la correlación serial.

### **Error de especificación en el modelo (omisión de una variable)**

El ejemplo 9.3 muestra cómo puede resolverse el problema de una variable faltante.

**Ejemplo 9.3** La Novak Corporation desea desarrollar un modelo de pronóstico para la proyección de ventas futuras. Como la corporación tiene sucursales regionales, se elige como posible variable de predicción al ingreso personal disponible en una base regional. La tabla 9.4 muestra las ventas de Novak para 1977 – 1993. El ingreso personal disponible se presentó en la tabla 9.1.

**TABLA 9.4** DATOS PARA EL EJEMPLO 9.3:  
VENTAS DE NOVAK CORPORATION

| AÑO  | VENTAS<br>(MILLONES) | AÑO  | VENTAS<br>(MILLONES) |
|------|----------------------|------|----------------------|
| 1977 | 8.0                  | 1986 | 16.4                 |
| 1978 | 8.2                  | 1987 | 17.8                 |
| 1979 | 8.5                  | 1988 | 18.6                 |
| 1980 | 9.2                  | 1989 | 20.0                 |
| 1981 | 10.2                 | 1990 | 21.9                 |
| 1982 | 11.4                 | 1991 | 24.9                 |
| 1983 | 12.8                 | 1992 | 27.3                 |
| 1984 | 13.6                 | 1993 | 29.1                 |
| 1985 | 14.6                 |      |                      |



Los comandos de Minitab para resolver este ejemplo son:

```
MTB > READ 'NOVAK.DAT' C1-C2
 17 ROWS READ
```

| ROW | C1  | C2    |
|-----|-----|-------|
| 1   | 8.0 | 336.1 |
| 2   | 8.2 | 349.4 |
| 3   | 8.5 | 362.9 |
| 4   | 9.2 | 383.9 |
| .   | .   | .     |

```
MTB > NAME C1 'SALES' C2 'INCOME'
MTB > REGRESS C1 1 PREDICTOR C2, RES IN C3, DHATS IN C4;
SUBC> RESIDUALS C5;
SUBC> DW.
```

The regression equation is  
 $\text{SALES} = -1.50 + 0.0292 \text{ INCOME}$

| Predictor | Coef      | Stdev     | t-ratio | P     |
|-----------|-----------|-----------|---------|-------|
| Constant  | -1.5046   | 0.3290    | -4.57   | 0.000 |
| INCOME    | 0.0291916 | 0.0005129 | 56.92   | 0.000 |

$s = 0.4767$        $R\text{-sq} = 99.5\%$        $R\text{-sq(adj)} = 99.5\%$

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F       | P     |
|------------|----|--------|--------|---------|-------|
| Regression | 1  | 736.15 | 736.15 | 3239.89 | 0.000 |
| Error      | 15 | 3.41   | 0.23   |         |       |
| Total      | 16 | 739.56 |        |         |       |

Unusual Observations

| Obs. | INCOME | SALES  | Fit    | Stdev.Fit | Residual | St.Resid |
|------|--------|--------|--------|-----------|----------|----------|
| 17   | 1077   | 29.100 | 29.926 | 0.270     | -0.826   | -2.10R   |

R denotes an obs. with a large st. resid.

Durbin-Watson statistic = 0.72

```
NTB > NAME C4 'ESTIMATE' C5 'RESIDUAL'
MTB > PRINT C1 C2 C4 C5
```

| ROW | SALES | INCOME | ESTIMATE | RESIDUAL |
|-----|-------|--------|----------|----------|
| 1   | 8.0   | 336.1  | 8.3067   | -0.30671 |
| 2   | 8.2   | 349.4  | 8.6949   | -0.49495 |
| 3   | 8.5   | 362.9  | 9.0890   | -0.58904 |
| 4   | 9.2   | 383.9  | 9.7020   | -0.50206 |
| 5   | 10.2  | 402.8  | 10.2538  | -0.05378 |
| 6   | 11.4  | 437.0  | 11.2521  | 0.14786  |

|    |      |        |         |          |
|----|------|--------|---------|----------|
| 7  | 12.8 | 472.2  | 12.2797 | 0.52032  |
| 8  | 13.6 | 510.4  | 13.3948 | 0.20520  |
| 9  | 14.6 | 544.5  | 14.3902 | 0.20977  |
| 10 | 16.4 | 588.1  | 15.6630 | 0.73701  |
| 11 | 17.8 | 630.4  | 16.8978 | 0.90222  |
| 12 | 18.6 | 685.9  | 18.5179 | 0.08209  |
| 13 | 20.0 | 742.8  | 20.1789 | -0.17892 |
| 14 | 21.9 | 801.3  | 21.8866 | 0.01337  |
| 15 | 24.9 | 903.1  | 24.8583 | 0.04166  |
| 16 | 27.3 | 983.6  | 27.2083 | 0.09175  |
| 17 | 29.1 | 1076.7 | 29.9260 | -0.82599 |

MTB > STOP

- El comando READ se usa para almacenar los datos de ventas de Novak en C1 y el ingreso personal disponible en C2.
- El comando TRESIDS se emplea para calcular residuos estandarizados y almacenarlos en C3.
- El comando DHATS se utiliza para calcular las estimaciones de  $Y$  y almacenarlas en C4.
- El subcomando RESIDUAL se usa para calcular la diferencia entre la observación real (C1) y la estimación de  $Y$  (C4). El resultado se almacena en C5.
- El subcomando DW se emplea para calcular la estadística Durbin-Watson.

Utilizando un nivel de significación de .01 para una muestra de 17 y una variable independiente, se obtiene

$$I = .87$$

$$S = 1.10$$

La estadística de Durbin-Watson de .72 indica que existe una correlación serial (.72 < .87). Es evidente que en el modelo falta una o más variables. Una parte importante de la variación de las ventas no ha sido explicada. Este resultado es verdadero incluso aunque la salida de Minitab indique que el ingreso personal disponible está explicando el 99.5% de la varianza de la variable ventas.

La mejor solución a este dilema es encontrar la variable o variables que faltan. Se agrega al modelo la tasa de desempleo y los resultados de la ejecución se muestran en la tabla 9.5. Los comandos de MINITAB para este ejemplo se detallan en la sección del paquete de cómputo Minitab al final del capítulo.

Con un nivel de significación de .01 para una muestra de 17 y dos variables independientes, se obtiene

$$I = .77$$

$$S = 1.25$$

TABLA 9.5 SALIDA DE CÓMPUTO DE MINITAB PARA EL EJEMPLO 9.3

The regression equation is

$$\text{SALES} = -0.014 + 0.0297 \text{ INCOME} - 0.350 \text{ UNEMP}$$

| Predictor | Coef      | Stdev     | t-ratio | P     |
|-----------|-----------|-----------|---------|-------|
| Constant  | -0.0140   | 0.2498    | -0.06   | 0.956 |
| INCOME    | 0.0297492 | 0.0002480 | 119.96  | 0.000 |
| UNEMP     | -0.34987  | 0.04656   | -7.51   | 0.000 |

s = 0.2199      R-sq = 99.9%      R-sq(adj) = 99.9%

#### Analysis of Variance

| SOURCE     | DF | SS     | MS     | F       | P     |
|------------|----|--------|--------|---------|-------|
| Regression | 2  | 738.88 | 369.44 | 7637.92 | 0.000 |
| Error      | 14 | 0.68   | 0.05   |         |       |
| Total      | 16 | 739.56 |        |         |       |

| SOURCE | DF | SEQ SS |
|--------|----|--------|
| INCOME | 1  | 736.15 |
| UNEMP  | 1  | 2.73   |

#### Unusual Observations

| Obs. | INCOME | SALES   | Fit     | Stdev.Fit | Residual | St.Resid |
|------|--------|---------|---------|-----------|----------|----------|
| 17   | 1077   | 29.1000 | 29.0430 | 0.1713    | 0.0570   | 0.41 X   |

X denotes an obs. whose X value gives it large influence.

Durbin-Watson statistic = 1.98

La estadística de Durbin-Watson de 1.98 que aparece en la tabla 9.5 indica que se ha eliminado el problema de correlación serial ( $1.98 > 1.25$ ).

Se puede utilizar la regresión  $\hat{Y} = -0.14 + .03X_2 - .35X_3$  para estimar las ventas de Novak Corporation, con el conocimiento de que los términos de error son independientes. Las estimaciones de expertos para la región, del ingreso personal disponible (\$1 185 millones) y de la tasa de desempleo (7.8%), se emplean para pronosticar las ventas de Novak para 1991 en \$32.8 millones.

$$\hat{Y} = -.014 + .03(1,185.0) - .35(7.8) = 32.8$$

## REGRESIÓN DE CAMBIOS PORCENTUALES

El ejemplo 9.4 muestra como usar el método de regresión de cambios porcentuales, para corregir la correlación serial.

### Ejemplo 9.4

Fred Gardner está comprometido en el pronóstico, en miles de dólares, de las ventas de Sears Roebuck para la región occidental. Como su variable independiente, eligió el ingreso personal disponible. La tabla 9.6 muestra las ventas de Sears y el ingreso personal disponible para el periodo 1973 – 1993.

Los comandos de Minitab para la ejecución de estos datos son

```
MTB > READ 'SEARS.DAT' C1-C2
 21 ROWS READ
```

| ROW | C1   | C2    |
|-----|------|-------|
| 1   | 3307 | 273.4 |
| 2   | 3556 | 291.3 |
| 3   | 3601 | 306.9 |
| 4   | 3721 | 317.1 |
| .   | .    | .     |

```
MTB > REGRESS C1 1 PREDICTOR C2:
SUBC> DW.
```

The regression equation is  
 $C1 = -524 + 14.0 C2$

| Predictor | Coef    | Stdev  | t-ratio | P     |
|-----------|---------|--------|---------|-------|
| Constant  | -524.3  | 188.4  | -2.78   | 0.012 |
| C2        | 14.0496 | 0.3185 | 44.11   | 0.000 |

s = 343.5      R-sq = 99.0%      R-sq(adj) = 99.0%

Analysis of Variance

| SOURCE     | DF | SS        | MS        | F       | P     |
|------------|----|-----------|-----------|---------|-------|
| Regression | 1  | 229603712 | 229603712 | 1945.86 | 0.000 |
| Error      | 19 | 2241929   | 117996    |         |       |
| Total      | 20 | 231845648 |           |         |       |

Durbin-Watson statistic = 0.63

The MINITAB output indicates that Fred can explain 99% of the Sears sales variance for the western region through the knowledge of its relationship with disposable income for the western region. However, the Durbin-Watson statistic is .63. At the .01 significance level for a sample of 20 and one independent variable the test statistics are

L = .95  
 U = 1.15

The Durbin-Watson statistic indicates positive serial correlation (.63 < .95).

In an attempt to eliminate the relationship among residuals, a regression line is fitted to the percentage changes from year to year rather than to the actual data. MINITAB is used to perform these computations. First, the differences are computed for the two variables and stored in columns C3 and C4.

```
MTB > DIFFERENCES 1 FOR C1, STORE IN C3
MTB > DIFFERENCES 1 FOR C2, STORE IN C4
```

Next, each of the variables are lagged and stored in C5 and C6.

```
MTB > LAG 1 DATA IN C1, PUT IN C5
MTB > LAG 1 DATA IN C2, PUT IN C6
```

The differences for Sears sales which are stored in C3 are divided by Sears sales lagged one period to compute the percentage change from the previous year. This result is stored in C7. The differences for disposable income which are stored in C4 are divided by disposable income lagged one period to compute the percentage change from the previous year. This result is stored in C8.

```
MTB > DIVIDE C3 BY C5, STORE IN C7
MTB > DIVIDE C4 BY C6, STORE IN C8
```

Next, the results of each of the columns is printed.

```
MTB > PRINT C1-C8
```

| ROW | C1    | C2     | C3   | C4      | C5    | C6    | C7       | C8       |
|-----|-------|--------|------|---------|-------|-------|----------|----------|
| 1   | 3307  | 273.4  | *    | *       | *     | *     | *        | *        |
| 2   | 3556  | 291.3  | 249  | 17.900  | 3307  | 273.4 | 0.075295 | 0.065472 |
| 3   | 3601  | 306.9  | 45   | 15.600  | 3556  | 291.3 | 0.012655 | 0.053553 |
| 4   | 3721  | 317.1  | 120  | 10.200  | 3601  | 306.9 | 0.033324 | 0.033236 |
| 5   | 4036  | 336.1  | 315  | 19.000  | 3721  | 317.1 | 0.084655 | 0.059918 |
| 6   | 4134  | 349.4  | 98   | 13.300  | 4036  | 336.1 | 0.024281 | 0.039572 |
| 7   | 4268  | 362.9  | 134  | 13.500  | 4134  | 349.4 | 0.032414 | 0.038638 |
| 8   | 4578  | 383.9  | 310  | 21.000  | 4268  | 362.9 | 0.072634 | 0.057867 |
| 9   | 5093  | 402.8  | 515  | 18.900  | 4578  | 383.9 | 0.112495 | 0.049232 |
| 10  | 5716  | 437.0  | 623  | 34.200  | 5093  | 402.8 | 0.122325 | 0.084906 |
| 11  | 6357  | 472.2  | 641  | 35.200  | 5716  | 437.0 | 0.112141 | 0.080549 |
| 12  | 6769  | 510.4  | 412  | 38.200  | 6357  | 472.2 | 0.064810 | 0.080898 |
| 13  | 7296  | 544.5  | 527  | 34.100  | 6769  | 510.4 | 0.077855 | 0.066810 |
| 14  | 8178  | 588.1  | 882  | 43.600  | 7296  | 544.5 | 0.120888 | 0.080073 |
| 15  | 8844  | 630.4  | 666  | 42.300  | 8178  | 588.1 | 0.081438 | 0.071927 |
| 16  | 9251  | 685.9  | 407  | 55.500  | 8844  | 630.4 | 0.046020 | 0.088039 |
| 17  | 10006 | 742.8  | 755  | 56.900  | 9251  | 685.9 | 0.081613 | 0.082957 |
| 18  | 11200 | 801.3  | 1194 | 58.500  | 10006 | 742.8 | 0.119328 | 0.078756 |
| 19  | 12500 | 903.1  | 1300 | 101.800 | 11200 | 801.3 | 0.116071 | 0.127044 |
| 20  | 13101 | 983.6  | 601  | 80.500  | 12500 | 903.1 | 0.048080 | 0.089137 |
| 21  | 13640 | 1076.7 | 539  | 93.100  | 13101 | 983.6 | 0.041142 | 0.094652 |

Finally, the percentage changes for disposable income (C8) are used to predict the percentage change in Sears sales (C7). Since percentage changes are being used instead of actual values, the regression is run without using a constant or Y-intercept.

```
MTB > NAME C7 'SALES' C8 'INCOME'
MTB > REGRESS C7 1 PREDICTOR C8;
SUBC> NONCONSTANT;
SUBC> DW.
```

The regression equation is  
SEARS = 1.01 INCOME

20 cases used 1 cases contain missing values

| Predictor  | Coef   | Stdev   | t-ratio | P     |
|------------|--------|---------|---------|-------|
| Noconstant | 1      |         |         |       |
| INCOME     | 1.0297 | 0.09616 | 10.53   | 0.000 |

s = 0.03201

Analysis of Variance

| SOURCE     | DF | SS      | MS      | F      | P     |
|------------|----|---------|---------|--------|-------|
| Regression | 1  | 0.11375 | 0.11375 | 110.98 | 0.000 |
| Error      | 19 | 0.01947 | 0.00102 |        |       |
| Total      | 20 | 0.13322 |         |        |       |

Durbin-Watson statistic = 1.27

En los resultados, los residuos están distribuidos en forma más aleatoria, como lo indica la estadística Durbin – Watson de 1.27 ( $1.27 > 1.15$ ). Los diversos errores estándar calculados para estos cambios porcentuales ( $S_{y,x}$ ,  $S_b$ ) son más válidos que los calculados a partir de los valores originales. Desde luego, este resultado no necesariamente significa que el pronóstico en sí sea más preciso.

En este punto, Fred puede acudir por ayuda a su conocimiento de la regresión múltiple. Si los cambios en las ventas de Sears están relacionados en forma simultánea con otras variables que afectan las ventas, la precisión de la estimación debería mejorar. Por ejemplo, la introducción en el modelo del cambio en la tasa de desempleo (véase tabla 9.6). La tabla 9.7 muestra los resultados de la salida de cómputo. La estadística de Durbin–Watson mejoró a 1.82. Por lo tanto, aparentemente no hay problemas de correlación serial.

Fred enfrenta ahora el dilema de incorporar o no más variables al modelo. Cabe señalar que siempre que se incorpora una nueva variable al modelo, con fines de pronóstico, se deben usar estimaciones de expertos para esa variable. Es posible que la precisión del pronóstico se reduzca mediante este procedimiento.

Si el analista decide utilizar el modelo demostrado en la tabla 9.7, habría necesidad de conseguir estimaciones expertas, tanto para el ingreso personal disponible como para la tasa de desempleo. En la tabla 9.8 y a continuación se muestra el procedimiento de pronóstico.

1. Estimación del ingreso personal disponible para 1994.
2. Estimación de la tasa de desempleo para 1994.
3.  $(1\ 185.0 - 1\ 076.7) = 108.3$ , y

$$\frac{108.3}{1,076.7} = 10.1\%$$

**TABLA 9.6** DATOS PARA EL EJEMPLO 9.4: VENTAS DE SEARS, INGRESO DISPONIBLE Y TASA DE DESEMPLEO EN EUA, 1973 – 1993

| AÑO  | VENTAS DE SEARS,<br>$Y$ | INGRESO DISPONIBLE,<br>$X_2$ | TASA DE DESEMPLEO (%)<br>$X_3$ | PORCENTAJE DE CAMBIO RESPECTO AL AÑO ANTERIOR |              |              |
|------|-------------------------|------------------------------|--------------------------------|-----------------------------------------------|--------------|--------------|
|      |                         |                              |                                | $\Delta Y$                                    | $\Delta X_2$ | $\Delta X_3$ |
| 1973 | 3,307                   | 273.4                        | 4.4                            | —                                             | —            | —            |
| 1974 | 3,556                   | 291.3                        | 4.1                            | 7.5                                           | 6.5          | -6.8         |
| 1975 | 3,601                   | 306.9                        | 4.3                            | 1.3                                           | 5.4          | 4.9          |
| 1976 | 3,721                   | 317.1                        | 6.8                            | 3.3                                           | 3.3          | 58.1         |
| 1977 | 4,036                   | 336.1                        | 5.5                            | 8.5                                           | 6.0          | -19.1        |
| 1978 | 4,134                   | 349.4                        | 5.5                            | 2.4                                           | 4.0          | .0           |
| 1979 | 4,268                   | 362.9                        | 6.7                            | 3.2                                           | 5.7          | 21.8         |
| 1980 | 4,578                   | 383.9                        | 5.5                            | 7.3                                           | 5.8          | -17.9        |
| 1981 | 5,093                   | 402.8                        | 5.7                            | 11.2                                          | 6.3          | 3.6          |
| 1982 | 5,716                   | 437.0                        | 5.2                            | 12.2                                          | 8.5          | -8.8         |
| 1983 | 6,357                   | 472.2                        | 4.5                            | 11.2                                          | 8.1          | -13.5        |
| 1984 | 6,769                   | 510.4                        | 3.8                            | 6.5                                           | 8.1          | -15.6        |
| 1985 | 7,296                   | 544.5                        | 3.8                            | 7.8                                           | 6.7          | .0           |
| 1986 | 8,178                   | 588.1                        | 3.6                            | 12.1                                          | 8.0          | -5.3         |
| 1987 | 8,844                   | 630.4                        | 3.5                            | 8.1                                           | 7.2          | -2.8         |
| 1988 | 9,251                   | 685.9                        | 4.9                            | 4.6                                           | 8.8          | 40.0         |
| 1989 | 10,006                  | 742.8                        | 5.9                            | 8.2                                           | 8.3          | 20.4         |
| 1990 | 11,200                  | 801.3                        | 5.6                            | 11.9                                          | 7.9          | -5.1         |
| 1991 | 12,500                  | 903.1                        | 4.9                            | 11.6                                          | 12.7         | -12.5        |
| 1992 | 13,101                  | 983.6                        | 5.6                            | 4.6                                           | 8.9          | 14.3         |
| 1993 | 13,640                  | 1,076.7                      | 8.5                            | 4.0                                           | 9.5          | 51.8         |

4.  $(7.8 - 8.5) = .7$ , y

$$\frac{-.7}{8.5} = -8.2\%$$

5.  $\text{Sears} = 1.05 (\text{ingreso}) - 0.0694 (\text{desempleo})$   
 $= 1.05(10.1) - 0.0694(-8.2)$   
 $= 11.174\%$  cambio estimado o porcentaje de incremento en las ventas de Sears

6.  $\hat{Y}(1994) = Y(1993) \times \Delta Y(1994) + Y(1993)$   
 $= 13\,640 \times .11174 + 13\,640$   
 $= 1\,524 + 13\,640$   
 $= 15\,164$ , pronóstico de ventas de Sears en la región occidental

La estimación de las ventas de Sears en la región occidental para 1994 es de \$15,164. Los diversos errores estándar calculados para este modelo de regresión múltiple son válidos y se pueden utilizar para desarrollar intervalos de confianza. Si se desea mayor precisión, se pueden introducir al modelo otras variables importantes (población, número de tiendas, gasto de consumo e índice de precios).

**TABLA 9.7** SALIDA DE CÓMPUTO PARA EL CAMBIO PORCENTUAL CON RESPECTO AL AÑO ANTERIOR DE LAS VENTAS DE SEARS, INGRESO DISPONIBLE Y TASA DE DESEMPLEO

The regression equation is

$$\text{SEARS} = 1.05 \text{ INCOME} - 0.0694 \text{ UNEMP}$$

20 cases used 1 cases contain missing values

| Predictor  | Coef     | Stdev   | t-ratio | p     |
|------------|----------|---------|---------|-------|
| Noconstant |          |         |         |       |
| INCOME     | 1.05230  | 0.08739 | 12.04   | 0.000 |
| UNEMP      | -0.06942 | 0.02874 | -2.42   | 0.027 |

s = 0.02859

Analysis of Variance

| SOURCE     | DF | SS       | MS       | F     | P     |
|------------|----|----------|----------|-------|-------|
| Regression | 2  | 0.118511 | 0.059255 | 72.52 | 0.000 |
| Error      | 18 | 0.014708 | 0.000817 |       |       |
| Total      | 20 | 0.133219 |          |       |       |

| SOURCE | DF | SEQ SS   |
|--------|----|----------|
| INCOME | 1  | 0.113745 |
| UNEMP  | 1  | 0.004766 |

Durbin-Watson statistic = 1.82

**TABLA 9.8** PROCEDIMIENTO DE PRONÓSTICO PARA LAS VENTAS DE SEARS

| AÑO  | $Y$                 | $X_2$                | $X_3$            | $X_4$<br>$\Delta Y$ | $X_5$<br>$\Delta X_2$ | $X_6$<br>$\Delta X_3$ |
|------|---------------------|----------------------|------------------|---------------------|-----------------------|-----------------------|
| 1993 | 13,640              | 1,076.7              | 8.5              | 4.0                 | 9.5                   | 51.8                  |
| 1994 | 15,164 <sup>6</sup> | 1,185.0 <sup>1</sup> | 7.8 <sup>2</sup> | 11.174 <sup>5</sup> | 10.1 <sup>3</sup>     | -8.2 <sup>4</sup>     |

*Nota:* Los exponentes se refieren a los pasos en el texto.

## MODELOS AUTORREGRESIVOS

Una forma de resolver el problema de correlación serial consiste en aprovechar la correlación entre observaciones adyacentes. A este método se le conoce como *modelo autorregresivo*. Se retrasa la variable dependiente uno o más periodos y se utiliza como una variable independiente. Por ejemplo, en la ventas de Reynolds Metals,  $Y_t$  se retrasa un periodo y se utiliza junto con el ingreso personal como una variable independiente. Este modelo se escribe



$$\hat{Y}_t = b_0 + b_2X_2 + b_3X_3 \quad (9.4)$$

en donde  $X_3 = Y_{t-1}$

Un *modelo autorregresivo* expresa un pronóstico como una función de valores previos de esa serie de tiempo.

### Ejemplo 9.5

En la tabla 9.1 se presentan los datos para el modelo de la ecuación 9.4. Nótese que se pierde un año de datos, ya que no se conocen las ventas de Reynolds para 1972. El tamaño de la muestra es de 20 en lugar de 21.

Los resultados de la ejecución de este modelo de regresión múltiple se presentan en la tabla 9.9. La estadística de Durbin-Watson mejora a 1.92. Sin embargo, la estadística de Durbin-Watson no debería emplearse cuando una variable dependiente retrasada aparece como variable independiente, ya que tiende a presentar una desviación hacia arriba. En su lugar debe utilizarse la estadística  $h$  de Durbin, para efectuar pruebas de residuos correlacionados en forma serial.

La estimación de ventas de Reynolds Metals para 1994 es \$2 607 000.

$$\hat{Y}_{1994} = .666 - .37X_2 + 1.127X_3$$

en donde  $X_3 = Y_{1993}$ . De modo que,

$$\hat{Y}_{1994} = .666 - .37(1,185) + 1.127(2,702) = 2,607$$

Una revisión de la matriz de correlación en la tabla 9.9 nos lleva a la conclusión de que pudiera haber un problema de colinealidad. La alta interrelación entre las variables independientes,  $r_{23} = .986$ , indica que es probable que estén explicando la misma varianza en la variable dependiente. El problema de colinealidad se hace obvio cuando se analiza el coeficiente de regresión de la variable ingreso personal disponible,  $b_2 = -.37$ . Como la variable de ingreso muestra una relación positiva alta con las ventas,  $r_{12} = .968$ , el coeficiente de regresión no es significativo. El modelo se podría mejorar si la única variable independiente fuera las ventas de Reynolds Metals retrasada un periodo. Este modelo comprendería un modelo autorregresivo de orden 1. Un modelo alternativo que se explicará en la exposición de las técnicas de Box-Jenkins en el cap. 10.

### Mínimos cuadrados generalizados

Una forma de resolver problemas de autocorrelación consiste en desarrollar un modelo que reconozca la relación entre los residuos de una manera apropiada. Originalmente cada punto de datos estaba representado por la ecuación 9.1., es decir

$$Y_t = \beta_0 + \beta X_t + \epsilon_t$$

<sup>2</sup> Johnston, *Econometric Methods*, 2a. Ed. (Nueva York: McGraw-Hill, 1972).

<sup>3</sup> La estimación del ingreso personal disponible para esta región en 1994 es de \$1,185.

**TABLA 9.9** SALIDA DE CÓMPUTO PARA LAS VENTAS DE REYNOLDS METALS:  
MODELO AUTORREGRESIVO

| VARIABLE NO.                            | MEAN    | STANDARD DEVIATION | CORRELATION X VS Y | REGRESSION COEFFICIENT | STD. ERROR OF REG. COEF. | COMPUTED T VALUE |
|-----------------------------------------|---------|--------------------|--------------------|------------------------|--------------------------|------------------|
| 2                                       | 2.709   | 0.179              | 0.96792            | -0.370                 | 0.40752                  | -0.91            |
| 3                                       | 3.041   | 0.348              | 0.98769            | 1.127                  | 0.20928                  | 5.38             |
| DEPENDENT                               |         |                    |                    |                        |                          |                  |
| 1                                       | 3.089   | 0.331              |                    |                        |                          |                  |
| INTERCEPT                               |         | 0.666              |                    | MULTIPLE CORRELATION   |                          | 0.98826          |
| STD. ERROR OF ESTIMATE                  |         | 0.05346            |                    | R SQUARE               |                          | 0.97665          |
| DURBIN-WATSON STATISTIC                 |         | 1.91897            |                    | CORRECTED R SQUARED    |                          | 0.97390          |
| ANALYSIS OF VARIANCE FOR THE REGRESSION |         |                    |                    |                        |                          |                  |
| SOURCES OF VARIATION                    |         |                    | DEGREES OF FREEDOM | SUM OF SQUARES         | MEAN SQUARES             | F VALUE          |
| ATTRIBUTABLE TO REGRESSION              |         |                    | 2                  | 2.032                  | 1.016                    | 355.544          |
| DEVIATION FROM REGRESSION               |         |                    | 17                 | 0.049                  | 0.003                    |                  |
| TOTAL                                   |         |                    | 19                 | 2.081                  |                          |                  |
| CORRELATION MATRIX: 3 BY 3              |         |                    |                    |                        |                          |                  |
| VAR.                                    | 1       | 2                  | 3                  |                        |                          |                  |
| 1                                       | 1.000   | 0.968              | 0.988              |                        |                          |                  |
| 2                                       | 0.968   | 1.000              | 0.986              |                        |                          |                  |
| 3                                       | 0.988   | 0.986              | 1.000              |                        |                          |                  |
| TABLE OF RESIDUALS                      |         |                    |                    |                        |                          |                  |
| CASE NO.                                | Y VALUE | Y ESTIMATE         | RESIDUAL           |                        |                          |                  |
| 1                                       | 2.60206 | 2.53630            | 0.06576            |                        |                          |                  |
| 2                                       | 2.59106 | 2.67722            | -0.08616           |                        |                          |                  |
| 3                                       | 2.62839 | 2.65926            | -0.03087           |                        |                          |                  |
| 4                                       | 2.73799 | 2.69196            | 0.04603            |                        |                          |                  |
| 5                                       | 2.74429 | 2.80920            | -0.06491           |                        |                          |                  |
| 6                                       | 2.79239 | 2.81021            | -0.01782           |                        |                          |                  |
| 7                                       | 2.85733 | 2.85536            | 0.00197            |                        |                          |                  |
| 8                                       | 2.94448 | 2.92081            | 0.02368            |                        |                          |                  |
| 9                                       | 3.02119 | 3.00590            | 0.01529            |                        |                          |                  |
| 10                                      | 3.11059 | 3.07987            | 0.03072            |                        |                          |                  |
| 11                                      | 3.18412 | 3.16809            | 0.01603            |                        |                          |                  |
| 12                                      | 3.20030 | 3.24054            | -0.04024           |                        |                          |                  |
| 13                                      | 3.29226 | 3.24639            | 0.04587            |                        |                          |                  |
| 14                                      | 3.32592 | 3.33883            | -0.01290           |                        |                          |                  |
| 15                                      | 3.32551 | 3.36319            | -0.03768           |                        |                          |                  |
| 16                                      | 3.39393 | 3.34992            | 0.04400            |                        |                          |                  |
| 17                                      | 3.50501 | 3.41481            | 0.09020            |                        |                          |                  |
| 18                                      | 3.56844 | 3.52075            | 0.04769            |                        |                          |                  |
| 19                                      | 3.52022 | 3.57848            | -0.05826           |                        |                          |                  |
| 20                                      | 3.43168 | 3.51006            | -0.07837           |                        |                          |                  |

Se necesita crear un nuevo modelo que reconozca que el término de error ( $\epsilon_t$ ) está constituido por una fracción del término de error previo más cierto efecto aleatorio. En la ecuación 9.2, el término de error se escribió como

$$\epsilon_t = \rho\epsilon_{t-1} + v_t$$

o

$$v_t = \epsilon_t - \rho\epsilon_{t-1}$$

en donde

$\rho$  = correlación entre residuos

$v_t$  = error aleatorio

$v_t = \epsilon_t$  cuando  $\rho = 0$

El nuevo modelo se convierte en

$$Y_t = \beta_0 + \beta X_t + \rho\epsilon_{t-1} + v_t \quad (9.5)$$

Si los residuos no están relacionados y  $\rho$  es igual a cero, el resultado es la ecuación 9.1, esto es

$$\begin{aligned} Y_t &= \beta_0 + \beta X_t + \rho\epsilon_{t-1} + v_t \\ &= \beta_0 + \beta X_t + 0\epsilon_{t-1} + v_t \\ &= \beta_0 + \beta X_t + v_t \\ &= \beta_0 + \beta X_t + \epsilon_t \text{ en donde } v_t = \epsilon_t \end{aligned}$$

Si los términos de error están relacionados y  $\rho$  es mayor que cero, es deseable reconocer la estructura relacionada de los términos de error.

Si se conoce  $\rho$ , entonces se puede utilizar una diferenciación generalizada para ajustar el procedimiento de mínimos cuadrados de modo que los términos de error sean independientes. Para describir este procedimiento, se emplea el hecho de que el modelo lineal, ecuación 9.1, se mantiene para todos los periodos. En particular

$$Y_{t-1} = \beta_0 + \beta X_{t-1} + \epsilon_{t-1} \quad (9.6)$$

Ahora, se puede transformar la ecuación multiplicando la ecuación 9.6 por  $\rho$  y restándola de la ecuación 9.1.

$$Y_t = \beta_0 + \beta X_t + \epsilon_t \quad [\text{Ecuación 9.1}]$$

$$\rho(Y_{t-1}) = \rho\beta_0 + \rho\beta X_{t-1} + \rho\epsilon_{t-1} \quad [\rho \times \text{Ecuación 9.6}]$$

$$Y_t - \rho(Y_{t-1}) = (\beta_0 - \rho\beta_0) + (\beta X_t - \rho\beta X_{t-1}) + (\epsilon_t - \rho\epsilon_{t-1}) \quad [\text{resta}]$$

$$Y'_t = \beta_0(1 - \rho) + \beta X'_t + v_t \quad (9.7)$$

La ecuación transformada tiene un proceso de error que se distribuye independientemente con una media igual a cero y una varianza constante. De ahí que la regresión

ordinaria de mínimos cuadrados, aplicada a la ecuación 9.7, arrojará estimaciones válidas de parámetros de regresión. Recuerde, las variables fueron transformadas a la forma

$$\begin{aligned}Y'_t &= Y_t - \rho Y_{t-1} \\X'_t &= X_t - \rho X_{t-1}\end{aligned}$$

La ecuación 9.7 se puede reescribir como

$$Y'_t = \beta'_0 + \beta X'_t + v_t \quad (9.8)$$

A la ecuación 9.8 se le llama versión generalizada de mínimos cuadrados de la ecuación 9.7, en donde

1. El término de error no está correlacionado en forma serial.
2. El coeficiente de regresión  $\beta$  es el mismo que el coeficiente de regresión en la ecuación original correlacionada en forma serial, ecuación 9.1.

### Primera diferenciación

Un procedimiento de estimación que se utiliza comúnmente, conocido como *primera diferenciación*, supone que  $\rho$  es igual a 1. Si  $\rho = 1$ , el modelo transformado de la ecuación 9.7 se convierte en

$$\begin{aligned}Y'_t &= \beta_0(1 - \rho) + \beta X'_t + \epsilon_t \\&= \beta_0(1 - 1) + \beta X'_t + \epsilon_t = \beta X'_t + \epsilon_t\end{aligned} \quad (9.9)$$

Por lo tanto, se deriva que el coeficiente de regresión  $\beta$  se pueda estimar mediante el método de mínimos cuadrados para efectuar la regresión a través del origen con las variables transformadas

$$\begin{aligned}Y'_t &= Y_t - \rho Y_{t-1} \\X'_t &= X_t - \rho X_{t-1}\end{aligned}$$

Nótese que estas variables transformadas son primeras diferencias. Este enfoque resulta efectivo para una diversidad de aplicaciones.

#### Ejemplo 9.6

La tabla 9.10 contiene las variables transformadas  $Y'_t$  y  $X'_t$ , basadas en la transformación de primeras diferencias para las ventas de Sears en la región occidental. El cuadro también contiene los cálculos para estimar la regresión lineal a través del origen. Nótese que el coeficiente estimado de regresión,  $b = .0117$ , es similar al obtenido con mínimos cuadrados ordinarios aplicados a las variables originales,  $b = .014$ . Sin embargo, el error estándar del coeficiente de regresión,  $s_b = .00114$ , es considerablemente mayor que  $s_b = .00032$ . Probablemente es más preciso el nuevo error estándar del coeficiente de regresión  $s_b = .00114$ . El error estándar original,  $s_b = .00034$ , para los mínimos cuadrados ordinarios aplicados a las variables originales es probable que subestime el verdadero error estándar debido a la correlación serial.

**TABLA 9.10** DATOS PARA EL EJEMPLO 9.6: CÁLCULOS DE REGRESIÓN LINEAL A TRAVÉS DEL ORIGEN DE LA PRIMERA DIFERENCIACIÓN PARA LAS VENTAS DE SEARS EN LA REGIÓN OCCIDENTAL

| AÑO     | VENTAS,<br>Y<br>(000) | INGRESO<br>DISPONIBLE,<br>X | $Y'_t$ | $X'_t$ | $X'_t Y'_t$ | $(X'_t)^2$ | $(Y'_t - bX'_t)^2$ |
|---------|-----------------------|-----------------------------|--------|--------|-------------|------------|--------------------|
| 1973    | 3.307                 | 273.4                       |        |        |             |            |                    |
| 1974    | 3.556                 | 291.3                       | .249   | 17.9   | 4.5         | 320.4      | .0016              |
| 1975    | 3.601                 | 306.9                       | .045   | 15.6   | .7          | 243.4      | .0189              |
| 1976    | 3.721                 | 317.1                       | .120   | 10.2   | 1.2         | 104.0      | .0000              |
| 1977    | 4.036                 | 336.1                       | .315   | 19.0   | 6.0         | 361.0      | .0086              |
| 1978    | 4.134                 | 349.4                       | .098   | 13.3   | 1.3         | 176.9      | .0033              |
| 1979    | 4.268                 | 362.9                       | .134   | 13.5   | 1.8         | 182.3      | .0006              |
| 1980    | 4.578                 | 383.9                       | .310   | 21.0   | 6.5         | 441.0      | .0041              |
| 1981    | 5.093                 | 402.8                       | .515   | 18.9   | 9.7         | 357.2      | .0864              |
| 1982    | 5.716                 | 437.0                       | .623   | 34.2   | 21.3        | 1,169.6    | .0497              |
| 1983    | 6.357                 | 472.2                       | .641   | 35.2   | 22.6        | 1,239.0    | .0525              |
| 1984    | 6.769                 | 510.4                       | .412   | 38.2   | 15.7        | 1,459.2    | .0012              |
| 1985    | 7.296                 | 544.5                       | .527   | 34.1   | 18.0        | 1,162.8    | .0164              |
| 1986    | 8.178                 | 588.1                       | .882   | 43.6   | 38.5        | 1,901.0    | .1383              |
| 1987    | 8.844                 | 630.4                       | .666   | 42.3   | 28.2        | 1,789.3    | .0293              |
| 1988    | 9.251                 | 685.9                       | .407   | 55.5   | 22.6        | 3,080.3    | .0587              |
| 1989    | 10.006                | 742.8                       | .755   | 56.9   | 43.0        | 3,237.6    | .0080              |
| 1990    | 11.200                | 801.3                       | 1.194  | 58.5   | 69.8        | 3,422.3    | .2596              |
| 1991    | 12.500                | 903.1                       | 1.300  | 101.8  | 132.3       | 10,363.2   | .0119              |
| 1992    | 13.101                | 983.6                       | .601   | 80.5   | 48.4        | 6,480.3    | .1162              |
| 1993    | 13.640                | 1,076.7                     | .539   | 93.1   | 50.2        | 8,667.6    | .3028              |
| Totales |                       |                             |        |        | 542.3       | 46,158.4   | 1.1681             |

$$b = \frac{\sum X'_t Y'_t}{\sum (X'_t)^2} = \frac{542.3}{46,158.4} = .0117$$

$$s_b^2 = \frac{\sum (Y'_t - bX'_t)^2 / (n - 1)}{\sum (X'_t)^2} = \frac{1.1681/19}{46,158.4} = \frac{.0614789}{46,158.4} = .0000013$$

$$s_b = .00114$$

Los comandos Minitab para resolver este problema son

```
MTB > READ 'SEARS.DAT' C1 C2
 21 ROWS READ
```

| ROW | C1   | C2    |
|-----|------|-------|
| 1   | 3307 | 273.4 |
| 2   | 3556 | 291.3 |
| 3   | 3601 | 306.9 |
| 4   | 3721 | 317.1 |

```
MTB > DIFFERENCES 1 FOR C1, STORE IN C3
MTB > DIFFERENCES 1 FOR C2, STORE IN C4
MTB > NAME C3 'SEARS' C4 'INCOME'
MTB > REGRESS C3 1 PREDICTOR C4;
SUBC> NOCONSTANT.
```

The regression equation is  
SEARS = 0.0117 INCOME

20 cases used 1 cases contain missing values

| Predictor   | Coef     | Stdev    | t-ratio | p     |
|-------------|----------|----------|---------|-------|
| Nonconstant |          |          |         |       |
| INCOME      | 0.011747 | 0.001154 | 10.18   | 0.000 |

s = 0.2479

Analysis of Variance

| SOURCE     | DF | SS     | MS     | F      | p     |
|------------|----|--------|--------|--------|-------|
| Regression | 1  | 6.3696 | 6.3696 | 103.62 | 0.000 |
| Error      | 19 | 1.1679 | 0.0615 |        |       |
| Total      | 20 | 7.5375 |        |        |       |

### Enfoque iterativo

El enfoque iterativo también comprende la transformación de las variables originales que se mostraron en la ecuación 9.8.

$$Y'_t = Y_t - \rho Y_{t-1}$$

$$X'_t = X_t - \rho X_{t-1}$$

Este proceso se basa en el modelo transformado, ecuación 9.7, la cual se puede resolver mediante métodos ordinarios de mínimos cuadrados. Por desgracia, el modelo transformado

$$Y'_t = \beta_0(1 - \rho) + \beta X'_t + \epsilon_t$$

no se puede utilizar directamente ya que se desconoce el parámetro de autocorrelación  $\rho$ , necesario para obtener las variables transformadas de la ecuación 9.8. Sin embargo, se han desarrollado técnicas para estimar  $\rho$ . En el libro *Econometric Models and Economic Forecasts* de Pindyck y Rubinfeld,<sup>4</sup> se exponen diversos enfoques. Estos enfoques utilizan por lo general la noción de que  $\rho$  es un coeficiente de correlación asociado con errores de

<sup>4</sup> R. S. Pindyck y D. L. Rubinfeld, *Econometric Models and Economic Forecasts* (Nueva York: McGraw-Hill, 1976, 1976), pp. 108-113.

periodos adyacentes. El enfoque iterativo no siempre funciona, debido a que se tiende a subestimar el parámetro de autocorrelación  $\rho$ .

Cochrane y Orcutt desarrollaron uno de los procedimientos más populares para estimar  $\rho$  en la presencia de la correlación serial de primer orden. Este procedimiento iterativo produce estimaciones sucesivas de  $\rho$  hasta que una iteración posterior resulte en un cambio mínimo. Este enfoque emplea la noción de que  $\rho$  es un coeficiente de correlación asociado con errores de periodos adyacentes.

La estimación inicial de  $\rho$  se deriva de los residuos utilizando la ecuación 9.10.

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2} \quad (9.10)$$

El valor  $\rho$  se sustituye por  $\hat{\rho}$  en la ecuación 9.7, formando

$$Y'_t = \beta_0(1 - \hat{\rho}) + \beta X'_t + v_t \quad (9.11)$$

La ecuación estimada transformada arroja valores de parámetros para la intersección original ( $\beta_1$ ) y para todos los coeficientes de regresión ( $\beta_2, \dots, \beta_k$ ). Estas estimaciones revisadas de parámetros se sustituyen en la ecuación original y se obtienen nuevos residuos. Los nuevos residuos estándar se calculan mediante la ecuación 9.12.

$$\hat{\epsilon}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_2 - \dots - \hat{\beta}_k X_k \quad (9.12)$$

Mediante la ejecución de la regresión

$$\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + v_t$$

se pueden utilizar estos residuos de segunda ronda para obtener una nueva estimación de  $\rho$ . La segunda estimación de  $\rho$  se compara con la primera y, si estos valores son razonablemente cercanos, se emplea la segunda estimación, si no, se hace otra iteración.

### Ejemplo 9.7

La salida de cómputo de Minitab muestra los resultados del uso de la variable independiente ingreso personal, para predecir las ventas de Sears con el enfoque ordinario de mínimos cuadrados. Si los datos de Sears se ejecutan en miles de dólares, la ecuación de predicción es

$$\hat{Y} = -.524 + .014X$$

en donde

$$S_b = .0003185$$

$$t = 44.1$$

$$r^2 = .99$$

$$DW = .63$$

Si se emplean los mínimos cuadrados generalizados (utilizando el método Cochrane – Orcutt), el resultado es

$$\hat{Y} = 2.34 + .010933X$$

en donde

$$S_b = .00137$$

$$t = 7.96$$

$$r^2 = .78$$

$$\rho = .936$$

Los resultados para este ejemplo se muestran en el ejemplo de TSP al final del capítulo. Nótese que la estimación final de  $\rho$  es .936, lo que significa que  $Y$  se ejecutó en realidad como  $Y'_t = Y_t - .936Y_{t-1}$  y  $X$  como  $X'_t = X_t - .936X_{t-1}$ . Cuando se compara la ecuación generalizada de mínimos cuadrados con la ecuación ordinaria de mínimos cuadrados, se observa que el valor de  $t$  ha disminuido. Este resultado tiene sentido, ya que una de las consecuencias de la correlación serial es que los mínimos cuadrados ordinarios subestiman el error estándar del coeficiente de regresión. De hecho, una razón para ajustar la correlación serial es evitar cometer errores de inferencia debidos a valores muy altos de  $t$ . Por último, la correlación serial no provoca ninguna desviación en la estimación del coeficiente de regresión; sin embargo, se incrementa la cantidad en la que cualquier estimación dada difiere de la verdadera  $\beta$ . La solución de mínimos cuadrados generalizados (.010933) proporciona aproximadamente el mismo resultado que con el método ordinario de mínimos cuadrados (.014).

Al aplicar el análisis de regresión a las series de tiempo de datos, con frecuencia los residuos se correlacionan. Los problemas surgen debido a que el análisis de regresión supone que los residuos son independientes. Esta situación se describe con el término *correlación serial*. La  $r^2$  de un modelo que contiene correlación serial es artificialmente alta; además, el error estándar subestima seriamente la variabilidad de los residuos y los coeficientes de regresión se vuelven ineficientes.

Una causa principal de autocorrelación en los residuos es la omisión de una o más variables clave. Por lo regular, esta omisión significa que una parte importante de la variación de la variable dependiente no ha sido explicada. Cuando los efectos de secuencia en el tiempo de una variable faltante se relacionan en forma positiva, los residuos se autocorrelacionan, pues incluyen los efectos de la variable faltante. La mejor solución a este problema consiste en buscar la variable faltante para incluirla en el modelo.

## APLICACIÓN A LA ADMINISTRACIÓN

Para este capítulo también son apropiadas las aplicaciones descritas en el cap. 8. Sin embargo, las técnicas descritas en este capítulo permiten al analista detectar y corregir el problema de correlación serial. De esta forma, se fortalece la capacidad de la administración para manejar series de tiempo de datos. El resultado neto es que la administración puede



tratar con una mayor variedad de datos dependientes del tiempo y sentir la confianza de que las predicciones son sensatas. Las áreas en las que estas técnicas tienen una utilidad particular son:

|                                            |                                               |
|--------------------------------------------|-----------------------------------------------|
| Pronóstico de ventas                       | Proyección de precios de acciones y bonos     |
| Proyección de costos de materias primas    | Proyección de penetración de nuevos productos |
| Estimaciones de requerimientos de personal | Estudios de la relación publicidad-ventas     |
| Control de inventarios                     |                                               |

Ninguna de las aplicaciones anteriores es nueva ni representa algo que no se hubiera podido enfrentar antes. No obstante, las herramientas empleadas en las proyecciones son más complejas y proporcionan respuestas más válidas y confiables. De manera específica, se deben encontrar soluciones al problema de la correlación serial, que por lo regular acompaña la regresión de la variable dependiente de una serie de tiempo.

## GLOSARIO

**Correlación serial** Existe correlación serial cuando las observaciones sucesivas a través del tiempo se relacionan entre sí.

**Modelo autorregresivo** Un modelo autorregresivo expresa un pronóstico como una función de los valores previos de esa serie de tiempo.

## FÓRMULAS CLAVE

**Modelo lineal ordinario de mínimos cuadrados**  $Y_t = \beta_0 + \beta X_t + \epsilon_t$  (9.1)

**Correlación serial de primer orden**  $\epsilon_t = \rho\epsilon_{t-1} + v_t$  (9.2)

**Estadística de Durbin – Watson**  $DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$  (9.3)

**Modelo autorregresivo**  $\hat{Y}_t = b_0 + b_2X_t + b_3Y_{t-1}$  (9.4)

**Correlación serial de primer orden**  $Y_t = \beta_0 + \beta X_t + \rho\epsilon_{t-1} + v_t$  (9.5)

**Modelo lineal ordinario de mínimos cuadrados: periodo previo**

$$Y_{t-1} = \beta_0 + \beta X_{t-1} + \epsilon_{t-1} \quad (9.6)$$

**Modelo lineal ordinario transformado de mínimos cuadrados**

$$Y'_t = \beta_0(1 - \rho) + \beta X'_t + v_t \quad (9.7)$$

$$\text{Mínimos cuadrados generalizados } Y'_t = \beta'_0 + \beta X'_t + v_t \quad (9.8)$$

$$\text{Modelo de primera diferenciación } = \beta X'_t + \epsilon_t \quad (9.9)$$

$$\text{Estimación inicial de } \rho \text{ de Cochrane–Orcutt } \hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2} \quad (9.10)$$

$$\text{Modelo lineal de Cochrane–Orcutt } Y'_t = \beta_0(1 - \hat{\rho}) + \beta X'_t + v_t \quad (9.11)$$

Residuos estimados de Cochrane–Orcutt

$$\hat{\epsilon}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_2 - \dots - \hat{\beta}_k X_k \quad (9.12)$$

$$\text{Coeficientes estandarizados } \beta = b\left(\frac{s}{s_y}\right) \quad (9.13)$$

## PROBLEMAS

1. ¿Por qué es un problema la correlación serial al analizar series de tiempo de datos?
2. ¿Cuál es la causa principal de la correlación serial?
3. ¿Qué suposición fundamental se no se satisface con más frecuencia al analizar variables de series de tiempo?
4. ¿Qué estadística se emplea para detectar la correlación serial?
5. Usted prueba una serie de 32 observaciones con dos variables independientes al nivel de significación de .01, y la estadística calculada de Durbin–Watson es igual a 1.0. ¿Cuál es su conclusión?
6. Usted prueba una serie de 61 observaciones con una variable independiente al nivel de significación de .05, y la estadística calculada de Durbin–Watson es igual a 1.6. ¿Cuál es su conclusión?
7. ¿Cómo se elimina el problema de correlación serial?
8. ¿Cómo funciona un modelo autorregresivo?
9. ¿Cómo trabajan las técnicas de primera diferenciación e iterativa en la resolución de diferencias de correlación serial?
10. Tamson Russell, una economista que trabaja para el gobierno, intenta determinar la función de demanda de combustible para automóviles de pasajeros en EUA. Tamson desarrolló un modelo que empleaba el precio real por galón de gasolina regular, para predecir el consumo por año de combustible para motores. Con este modelo, ella sólo pudo explicar el 83.5% de la varianza. Tamson decidió agregar al modelo una variable que representara la población de EUA. Determine si existen problemas de correlación serial. Los datos se muestran en la tabla P.10.

TABLA P.10

| AÑO  | COMBUSTIBLE PARA MOTOR<br>CONSUMIDO POR AUTOMÓVILES<br>(MILES DE MILLONES DE GALONES) | PRECIO DE<br>LA GASOLINA | POBLACIÓN DE EUA<br>(MILLONES) |
|------|---------------------------------------------------------------------------------------|--------------------------|--------------------------------|
|      | $Y$                                                                                   | $X_2$                    | $X_3$                          |
| 1973 | 78.8                                                                                  | .39                      | 211.9                          |
| 1974 | 75.1                                                                                  | .53                      | 213.9                          |
| 1975 | 76.4                                                                                  | .57                      | 216.0                          |
| 1976 | 79.7                                                                                  | .59                      | 218.0                          |
| 1977 | 80.4                                                                                  | .62                      | 220.2                          |
| 1978 | 81.7                                                                                  | .63                      | 222.6                          |
| 1979 | 77.1                                                                                  | .86                      | 225.1                          |
| 1980 | 71.9                                                                                  | 1.19                     | 227.7                          |
| 1981 | 71.0                                                                                  | 1.33                     | 230.1                          |
| 1982 | 70.1                                                                                  | 1.26                     | 232.5                          |
| 1983 | 69.9                                                                                  | 1.22                     | 234.8                          |
| 1984 | 68.7                                                                                  | 1.21                     | 236.3                          |
| 1985 | 69.3                                                                                  | 1.16                     | 238.5                          |
| 1986 | 71.4                                                                                  | .92                      | 240.7                          |
| 1987 | 70.6                                                                                  | .95                      | 242.8                          |
| 1988 | 71.7                                                                                  | .95                      | 245.1                          |

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

11. Se solicita a Decision Science Associates que realice un estudio de factibilidad para un hotel de destino propuesto a ubicarse a 2.4 kilómetros de la presa Grand Coulee. Mark Craze no está satisfecho con el modelo de regresión que utilizó el precio de un galón de gasolina regular para predecir el número de personas que acudirían al centro de visitantes de la presa. Después de graficar los datos en un digrama de dispersión, Mark decide utilizar una variable ficticia para representar celebraciones importantes en el área general. Mark usa un 1 para representar una celebración y un 0 cuando no la hay. Nótese que el 1 en los datos de 1974 representa la Feria Mundial Expo 74, celebrada en Spokane, Washington, el 1 en 1983 representa la celebración del quincuagésimo aniversario de la presa Grand Coulee y, el 1 en 1986 representa la Feria Mundial celebrada en Vancouver, Canadá. Mark también decide utilizar el periodo como variable de predicción.

Escriba un reporte para que Mark presente a su jefe. Indique si hay problemas de correlación serial. Señale también qué información adicional sería importante para decidir si se recomienda la construcción de ese hotel.

12. Jim Jackson, un analista de precios de la Washington Water Power Company está preparando una caso de tarifas y necesita pronosticar el ingreso por consumo eléctrico residencial para 1993. Jim decide investigar tres variables de predicción potenciales: uso residencial por kilowatt – hora (kWh), cargo residencial por kWh (centavos/kWh) y, el número de clientes de electricidad residenciales. Reúne datos de 1968 a 1992. Los datos se muestran en la tabla P.12.

TABLA P.11

| AÑO  | NÚMERO DE<br>VISITANTES<br>$Y$ | PERIODO<br>$X_2$ | PRECIO DE<br>LA GASOLINA<br>$X_3$ | CELEBRACION<br>$X_4$ |
|------|--------------------------------|------------------|-----------------------------------|----------------------|
| 1973 | 268,528                        | 1                | .39                               | 0                    |
| 1974 | 468,136                        | 2                | .53                               | 1                    |
| 1975 | 390,129                        | 3                | .57                               | 0                    |
| 1976 | 300,140                        | 4                | .59                               | 0                    |
| 1977 | 271,140                        | 5                | .62                               | 0                    |
| 1978 | 282,752                        | 6                | .63                               | 0                    |
| 1979 | 244,006                        | 7                | .86                               | 0                    |
| 1980 | 161,524                        | 8                | 1.19                              | 0                    |
| 1981 | 277,134                        | 9                | 1.31                              | 0                    |
| 1982 | 382,343                        | 10               | 1.22                              | 0                    |
| 1983 | 617,737                        | 11               | 1.16                              | 1                    |
| 1984 | 453,881                        | 12               | 1.13                              | 0                    |
| 1985 | 471,417                        | 13               | .86                               | 0                    |
| 1986 | 654,147                        | 14               | .90                               | 1                    |

Fuente: Grand Coulee Dam Visitors Center y *Statistical Abstract of the United States, 1988*.

TABLA P.12

| AÑO  | INGRESOS<br>(MILLONES)<br>$Y$ | USO POR<br>KwH<br>$X_2$ | CARGO<br>(CENTAVOS/KwH)<br>$X_3$ | NÚMERO DE<br>CLIENTES<br>$X_4$ |
|------|-------------------------------|-------------------------|----------------------------------|--------------------------------|
| 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                        |

Fuente: "Financial and Operating Supplement", *Washington Water Power Annual Report*, varios años.

Jim testificó ante la comisión de tarifas de Idaho y se le preguntó si había problemas de correlación serial. él no supo la respuesta y le pide a usted que envíe una respuesta por escrito a la comisión.

13. Paul Raymond, presidente de Washington Water Power, está preocupado por la posibilidad de un intento de tomar el control y por el hecho de que el número de accionistas comunes ha disminuido desde 1983. Le da a usted instrucciones de estudiar el número de accionistas comunes desde 1968 y pronosticarlo para 1993. Usted decide investigar tres variables de predicción potenciales: ingresos por acción (común), dividendos por acción (común) y proporción de pago. Usted reúne los datos de 1968 a 1992 que se presentan en la tabla P.13.

TABLA P.13

| AÑO  | ACCIONISTAS<br>COMUNES<br>$Y$ | INGRESOS<br>POR ACCIÓN<br>$X_2$ | DIVIDENDOS<br>POR ACCIÓN<br>$X_3$ | PROPORCIÓN<br>DE PAGO<br>$X_4$ |
|------|-------------------------------|---------------------------------|-----------------------------------|--------------------------------|
| 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                             |

Fuente: "Financial and Operating Supplement", *Washington Water Power Annual Report*, varios años.

- a. Ejecute estos datos en la computadora y encuentre el mejor modelo de predicción.
  - b. ¿Hay en este modelo un problema de correlación serial?
  - c. Si hay un problema de correlación serial, escriba un memorando para Paul que exponga las diferentes soluciones al problema de autocorrelación e incluya su recomendación final.
14. Thompson Airlines determinó que el 5% del número total de pasajeros locales que usan líneas aéreas en EUA vuelan en aviones de Thompson. Se le asigna a usted la tarea de pronosticar el número de pasajeros que volará en Thompson Airlines en 1995. Los datos se presentan en la tabla P.14.

TABLA P.14

| AÑO  | NÚMERO DE PASAJEROS (MILES) | AÑO  | NÚMERO DE PASAJEROS (MILES) | AÑO  | NÚMERO DE PASAJEROS (MILES) | AÑO  | NÚMERO DE PASAJEROS (MILES) |
|------|-----------------------------|------|-----------------------------|------|-----------------------------|------|-----------------------------|
| 1962 | 22.8                        | 1968 | 39.5                        | 1974 | 61.9                        | 1990 | 117.2                       |
| 1963 | 26.1                        | 1969 | 45.4                        | 1975 | 69.9                        | 1991 | 124.9                       |
| 1964 | 29.4                        | 1970 | 46.3                        | 1976 | 79.9                        | 1992 | 136.6                       |
| 1965 | 34.5                        | 1971 | 45.8                        | 1977 | 96.3                        | 1993 | 144.8                       |
| 1966 | 37.6                        | 1972 | 48.0                        | 1978 | 109.0                       | 1994 | 147.9                       |
| 1967 | 40.3                        | 1973 | 54.6                        | 1979 | 116.0                       |      |                             |

- a. Desarrolle un modelo de regresión de serie histórica, utilizando el periodo como variable independiente y el número de pasajeros como variable dependiente.
  - b. ¿Están los términos de error de este modelo dispersos de manera uniforme?
  - c. Transforme la variable número de pasajeros para que los términos de error queden dispersos de manera uniforme.
  - d. Ejecute un programa de cómputo para el modelo transformado en el apartado c.
  - e. ¿Son los términos de error independientes en la ejecución del modelo del apartado d?
  - f. Si los términos de error son dependientes, ¿qué problemas se enfrentan al emplear este modelo?
  - g. Pronostique el número de pasajeros de Thompson Airlines para 1995.
15. La Thomas Furniture Company llega a la conclusión de que se puede mejorar la programación de la producción desarrollando un método preciso de predicción de las ventas trimestrales. El analista de la compañía, el Sr. Estes, decide investigar la relación entre el número de permisos de construcción de casas habitación y las ventas de muebles en el área de Springfield. Estes piensa que los permisos se emiten entre 1 y 2 trimestres antes de que se efectúen las ventas. Además, él pregunta si las estaciones afectan las ventas de muebles. Estes decide considerar otra variable independiente:

$$X_3 = \begin{cases} 0 & \text{si las ventas corresponden al primero o segundo trimestre} \\ 1 & \text{si las ventas corresponden al tercer o cuarto trimestre} \end{cases}$$

Los datos se presentan en la tabla P.15.

TABLA P.15

| AÑO  | TRIMESTRE | VENTAS<br>(EN MILES),<br>$Y$ | PERMISOS<br>$X_2$ | AÑO  | TRIMESTRE | VENTAS<br>(EN MILES),<br>$Y$ | PERMISOS<br>$X_2$ |
|------|-----------|------------------------------|-------------------|------|-----------|------------------------------|-------------------|
| 1989 | 3         |                              | 19                | 1992 | 1         | 120                          | 72                |
|      | 4         |                              | 3                 |      | 2         | 150                          | 31                |
| 1990 | 1         | 120                          | 35                |      | 3         | 660                          | 19                |
|      | 2         | 80                           | 11                |      | 4         | 270                          | 14                |
|      | 3         | 400                          | 11                | 1993 | 1         | 200                          | 75                |
|      | 4         | 200                          | 16                |      | 2         | 280                          | 41                |
| 1991 | 1         | 75                           | 32                |      | 3         | 800                          | 17                |
|      | 2         | 120                          | 10                |      | 4         | 320                          | 10                |
|      | 3         | 270                          | 12                |      |           |                              |                   |
|      | 4         | 155                          | 21                |      |           |                              |                   |

- Desarrolle un modelo de regresión que utilice los permisos de construcción como la variable de predicción.
  - Pruebe si el modelo tiene autocorrelación.
  - Desarrolle un modelo de regresión que use tanto los permisos como el efecto estacional.
  - ¿Es el modelo de regresión múltiple mejor que el modelo de regresión simple? (Pruebe con un nivel de significación de .05.)
  - ¿Tiene su mejor modelo un problema de autocorrelación? De ser así, ¿cómo se puede corregir?
  - Pronostique las ventas de Thomas Furniture Company para 1994 por trimestres.
16. Los datos de la tabla P.16 muestran las ventas trimestrales ajustadas en forma estacional de Dickson Corporation y de la industria total para los últimos 20 trimestres.

TABLA P.16

| AÑO  | TRIMESTRE | VENTAS DE<br>DICKSON<br>(EN MILES),<br>$Y$ | VENTAS DE<br>LA INDUSTRIA<br>(EN MILLONES)<br>$X_2$ | AÑO  | TRIMESTRE | VENTAS DE<br>DICKSON<br>(EN MILES),<br>$Y$ | VENTAS DE<br>LA INDUSTRIA<br>(EN MILLONES)<br>$X_2$ |
|------|-----------|--------------------------------------------|-----------------------------------------------------|------|-----------|--------------------------------------------|-----------------------------------------------------|
| 1989 | 1         | 83.8                                       | 31.8                                                | 1991 | 3         | 98.2                                       | 37.1                                                |
|      | 2         | 85.6                                       | 32.5                                                |      | 4         | 97.2                                       | 36.6                                                |
|      | 3         | 87.8                                       | 33.2                                                | 1992 | 1         | 100.1                                      | 37.6                                                |
|      | 4         | 86.1                                       | 32.4                                                |      | 2         | 102.6                                      | 38.3                                                |
| 1990 | 1         | 89.6                                       | 33.8                                                |      | 3         | 105.4                                      | 39.3                                                |
|      | 2         | 91.0                                       | 34.3                                                |      | 4         | 107.9                                      | 40.2                                                |
|      | 3         | 93.9                                       | 35.3                                                | 1993 | 1         | 110.1                                      | 41.1                                                |
|      | 4         | 94.6                                       | 35.7                                                |      | 2         | 111.1                                      | 41.4                                                |
| 1991 | 1         | 96.4                                       | 36.4                                                |      | 3         | 110.1                                      | 41.1                                                |
|      | 2         | 96.0                                       | 36.3                                                |      | 4         | 111.1                                      | 41.4                                                |

- a. Ajuste el modelo de regresión lineal, obtenga los residuos y grafíquelos contra el tiempo. ¿Qué es lo que encuentra?
  - b. Calcule la estadística de Durbin–Watson y determine si existe autocorrelación.
  - c. Estime el coeficiente de regresión  $b_2$  mediante el enfoque de primera diferenciación.
  - d. Estime el error estándar del coeficiente de regresión mediante el enfoque de primera diferenciación.
  - e. ¿Son más precisas sus estimaciones utilizando el enfoque de primera diferenciación?
17. Se hace un estudio para tratar de relacionar los ahorros personales con el ingreso personal (en miles de millones de dólares) para el periodo 1935–1954. Los datos se presentan en la tabla P.17.

TABLA P.17

| AÑO  | AHORRO<br>PERSONAL,<br>Y | INGRESO<br>PERSONAL,<br>X | AÑO  | AHORRO<br>PERSONAL,<br>Y | INGRESO<br>PERSONAL,<br>X | AÑO  | AHORRO<br>PERSONAL,<br>Y | INGRESO<br>PERSONAL,<br>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                       |      |                          |                           |

- a. Evalúe el modelo de regresión simple en el que se use el ingreso personal para predecir el ahorro personal. En concreto, 1) haga una prueba de significación ( $\alpha = .01$ ) del coeficiente de regresión; 2) pruebe la contribución de la variable ingreso personal a la predicción del ahorro personal utilizando la prueba  $F$  ( $\alpha = .01$ ); (3) pruebe si hay autocorrelación. ¿Cómo puede mejorarse el modelo?
  - b. Desarrolle una variable ficticia  $X_3$  para los años de guerra. Haga  $X_3 = 0$  para los tiempos de paz y  $X_3 = 1$  para los tiempos de guerra. Se consideran años de guerra 1941 a 1945. Ejecute un programa de cómputo para este modelo de regresión múltiple y evalúe los resultados. Específicamente, 1) determine si los años de guerra hacen una contribución importante a la predicción del ahorro personal ( $\alpha = .01$ ); 2) **pruebe si hay autocorrelación.**  
**¿Es el modelo de regresión múltiple mejor que el modelo de regresión simple?**
18. Use el método de Cochran–Orcutt para corregir la correlación serial en los datos de Reynolds Metals del ejemplo 9.1.
19. Utilice el método de Cochran–Orcutt para corregir la correlación serial en los datos de Novak Corporation del ejemplo 9.3.

## CASO DE ESTUDIO 9.1

**LA COMPAÑÍA DE SU ELECCIÓN**

Como se ha mencionado varias veces en este capítulo, las diversas variables en cada compañía, medidas cada año, trimestre y mes, son mediciones vitales para la salud de la compañía. Para cada una de ellas, puede haber muchas otras variables que se encuentren altamente correlacionadas y que pudieran proporcionar valiosas revelaciones y poder de pronóstico.



La finalidad de este caso de estudio es la de estimular la identificación de una importante variable de serie histórica para la compañía de su elección y analizar después los patrones en los datos mediante el análisis de correlación. Además, puede usted emplear programas de cómputo de análisis de regresión para ver si se puede encontrar una buena ecuación de pronóstico.

## PREGUNTAS

1. Identifique una compañía u organización de su interés, que sea una compañía local o nacional que tenga registros publicados, incluyendo la medición de variables de series de tiempo.
2. Identifique una variable clave de la compañía de su elección y registre sus valores para varios años, trimestres o meses.
3. Ya sea a mano o por computadora, calcule varios coeficientes de autocorrelación y grafíquelos en un correlograma.
4. Con base en el patrón del correlograma, describa los patrones de su serie histórica.
5. Calcule una  $Y$  retrasada y la primera diferenciación de sus datos, calcule los coeficientes de autocorrelación, grafíquelos en un correlograma y describa los patrones resultantes.
6. Identifique diversas variables de pronóstico potenciales que usted piense que podrían estar correlacionadas con la variable dependiente. En el proceso puede utilizar los registros de la compañía junto con otras fuentes de datos.
7. Ejecute los datos en un programa de cómputo de análisis de regresión y obtenga una matriz de correlación. Ensamble diversas combinaciones de variables de predicción que piense puedan tener éxito y ejecute un análisis de regresión para cada una. Vea si puede encontrar una buena ecuación de pronóstico para su variable dependiente.
8. Verifique su modelo para asegurarse de que no hay problemas de correlación serial.

## CASO DE ESTUDIO 9.2

### **ÍNDICE DE ACTIVIDAD EMPRESARIAL DEL CONDADO DE SPOKANE**

Antes de 1973 el condado de Spokane, Washington, no tenía una medición actualizada de la actividad empresarial en general. Sin embargo, lo que sucede en esta área como un todo, afecta cada negocio local, agencias de gobierno e individuos. Los planes y políticas elaborados por una unidad económica estarían incompletos sin algún conocimiento confiable de desempeño reciente de la economía de la que la unidad forma parte. Un índice de actividad empresarial de Spokane serviría como un insumo vital en la formulación de estrategias y decisiones en organizaciones tanto públicas como privadas.

Un índice de actividad empresarial es un indicador de cambio relativo en las condiciones económicas generales dentro de una región determinada. A nivel nacional, el nivel producto nacional bruto (que emite el Department of Commerce) y el índice de producción industrial (que emite la Federal Reserve Board) se consideran en general excelentes indicadores. Cada una de estas series se basa en miles de piezas de información (cuya recolección, edición y cálculo son empresas costosas y que tardan tiempo). Para un área local como el condado de Spokane, Washington, es más deseable una versión simplificada, capaz de proporcionar información razonablemente precisa y actual a un costo moderado.

Es común utilizar la regresión múltiple para construir un índice de actividad empresarial. Existen tres preguntas esenciales que debe enfrentar la generación de dicho índice.

- ¿Cuáles son los componentes del índice?
- ¿Representan estos índices de manera adecuada los cambios en las condiciones generales en los negocios?
- ¿Qué valor se debería asignar a cada uno de los componentes elegidos?

Las respuestas a estas preguntas se pueden obtener a través del análisis de regresión.

El Dr. Shik Chun Young, profesor de economía en la Eastern Washington University, está intentando desarrollar un índice de actividad económica para el condado de Spokane. Young selecciona el ingreso personal como la variable independiente. En el condado, el nivel de ingreso personal se considera como el mejor indicador disponible de las condiciones de las empresas locales. El ingreso personal mide el ingreso total recibido por los hogares antes del pago de impuestos personales. Como las actividades productivas son típicamente remuneradas por medios monetarios, el ingreso personal pudiera, de hecho, verse como una representación razonable del desempeño general de la economía. ¿Por qué entonces es necesario construir otro índice si el ingreso personal puede servir como un buen indicador de la actividad empresarial? Por desgracia, los datos de ingreso personal a nivel de condado los estima el Departamento de Comercio de EUA sobre una base anual y se liberan 16 meses más tarde. En consecuencia, estos datos son de poca utilidad para la planeación a corto plazo. La tarea de Young consiste en establecer un índice de actividad empresarial actualizado.

Las variables independientes se extraen de aquellos datos locales que están fácilmente disponibles en una base mensual. Actualmente, hay disponibles alrededor de 50 series de dichos datos mensuales y que van desde empleo, actividades bancarias y transacciones en bienes raíces hasta consumo de energía eléctrica. Si se incluyeran todas las series en el análisis de regresión, el esfuerzo tendría baja productividad ya que sólo unas cuantas de estas series serían significativas estadísticamente. De ahí que se requiera cierto conocimiento de la relación entre el ingreso personal y los datos disponibles, con el fin de determinar qué series incluir en la ecuación de regresión. A partir del conocimiento de Young de la economía de Spokane, se seleccionaron las siguientes 10 series:

$X_2$ , empleo total

$X_3$ , empleo en manufacturas

$X_4$ , empleo en la construcción

$X_5$ , empleo en comercio al mayoreo y menudeo

$X_6$ , empleo en servicios

$X_7$ , débitos bancarios

$X_8$ , depósitos bancarios

$X_9$ , permisos de construcción emitidos

$X_{10}$ , hipotecas en bienes raíces

$X_{11}$ , consumo total de electricidad

El primer paso en el análisis es la ejecución del modelo,

$$\hat{Y} = b_0 + b_2X_2 + b_3X_3 + \dots + b_{11}X_{11}$$

en donde

$Y$  = ingreso personal

$b_0$  = intersección con  $Y$

$b_2, b_3, \dots, b_{11}$  = coeficientes de las variables independientes respectivas

La  $R^2$  total corregida es .96, lo que significa que las 10 variables en conjunto explican el 96% de la varianza en la variable independiente, ingreso personal. Sin embargo, otras estadísticas de la regresión indican problemas. Primero, de estas 10 variables independientes sólo 3 tienen un valor  $t$  calculado que es significativo a un nivel de .05, éstas son: empleo total, empleo en servicios y débitos bancarios. Segundo, la matriz de correlación muestra un alto grado de interdependencia entre diversas variables independientes—multicolinealidad. Por ejemplo, el empleo total y los débitos bancarios tienen un coeficiente de correlación de .88; el consumo total de electricidad y los depósitos bancarios, .76; y los permisos de construcción y las hipotecas sobre bienes raíces, .68. Tercero, existe autocorrelación, como lo indica la estadística de Durbin – Watson de .91. Este fenómeno es más bien común en series de tiempo de datos en donde cada observación no es independientes de otras observaciones en la misma serie.

Como la aleatoriedad de las observaciones es uno de los conceptos básicos en la inferencia estadística, Young elige enfrentar primero el problema de autocorrelación. Adopta el método de primera diferenciación para minimizar la interdependencia entre las observaciones de cada serie histórica. Ahora, las 10 variables independientes se miden por la diferencia entre periodos en vez de por el valor absoluto de cada periodo. De modo que para poder distinguir los conjuntos de datos, se emplea una nueva designación de las variables independientes utilizadas.

$\Delta X_2$ , cambio en el empleo total

$\Delta X_3$ , cambio en el empleo en manufacturas

$\Delta X_4$ , cambio en el empleo en la construcción

$\Delta X_5$ , cambio en el empleo en comercio al mayoreo y menudeo

$\Delta X_6$ , cambio en el empleo en servicios

$\Delta X_7$ , cambio en los débitos bancarios

$\Delta X_8$ , cambio en los depósitos bancarios

$\Delta X_9$ , cambio en los permisos de construcción emitidos

$\Delta X_{10}$ , cambio en las hipotecas en bienes raíces

$\Delta X_{11}$ , cambio en el consumo total de electricidad

La ecuación de regresión se convierte en

$$\Delta \hat{Y} = b_0 + b_2 \Delta X_2 + b_3 \Delta X_3 + \dots + b_{11} \Delta X_{11}$$

en donde

$\Delta Y$  = cambio en el ingreso personal

$b_0$  = intersección con  $Y$

$b_2, b_3, \dots, b_{11}$  = coeficientes de regresión de las variables independientes respectivas

Con esta ecuación, se ejecuta una regresión con base en los datos de la primera diferenciación y produce una estadística de Durbin–Watson de 1.71. Ésta indica que no se mantiene una correlación seria.

El siguiente paso consiste en determinar cuál de las 10 variables son pronosticadores significativos de la variable dependiente. Con el fin de seleccionar la mejor ecuación, se hacen regresiones de una variedad de posibles combinaciones de las 10 variables contra  $Y$ . Los criterios empleados en la selección son los siguientes:

- Una  $R^2$  corregida satisfactoriamente alta
- Bajos coeficientes de correlación entre las variables independientes
- Que cada uno de los coeficientes de regresión de las variables independientes sea significativo a un nivel de .05.

Después de un cuidadoso escrutinio de los resultados de las regresiones, Young encuentra que la ecuación que contiene  $\Delta X_5$ ,  $\Delta X_6$  y  $\Delta X_{11}$  como variables independientes, reúne mejor los criterios mencionados.

Sin embargo, Young razona que (además de los usos industriales y comerciales) el consumo total de electricidad incluye el consumo residencial, que no debería tener una relación significativa con la actividad económica en el corto plazo. Para probar esta hipótesis, Young subdivide el consumo de electricidad total en cuatro variables:  $\Delta X_{12}$ , cambio en el uso de electricidad residencial;  $\Delta X_{13}$ , cambio en el uso de electricidad comercial;  $\Delta X_{14}$  cambio en el uso de electricidad industrial; y  $\Delta X_{15}$ , cambio en los usos comercial e industrial. Cada una de estas variables, combinadas con  $\Delta X_5$  y  $\Delta X_6$ , se emplean para producir cuatro nuevas ecuaciones de regresión (véase tabla 9.11).

El análisis estadístico indica que la ecuación D es la mejor. Al compararla con las ecuaciones previas que contienen  $\Delta X_5$ ,  $\Delta X_6$  y  $\Delta X_{11}$  como variables independientes, la ecuación A es la única que muestra deterioro en la significación estadística. El resultado confirma la noción de Young de que los usos de electricidad comercial e industrial son mejores pronosticadores que el ingreso personal y el uso total de electricidad, que incluye al consumo residencial.

Por lo tanto, se selecciona la ecuación D como la ecuación final de regresión, y los resultados son

$$\Delta Y = -1.86 + 17.10 \Delta X_5 + 23.01 \Delta X_6 + .007 \Delta X_{15}$$

(4.07)                      (56.1)                      (.002)

$$N = 15 \quad R_c^2 = .835$$

$$DW = 1.769 \quad F = 26.26$$

Las cifras entre paréntesis bajo los coeficientes de regresión son los errores estándar de la estimación de los coeficientes, todos ellos significativos a un nivel de .05. Los valores  $t$  de los coeficientes son 4.10, 4.20 y 2.97 para  $\Delta X_5$ ,  $\Delta X_6$  y  $\Delta X_{15}$ , respectivamente. La  $R_c^2$  indica que cerca del 84% de la varianza

**TABLA 9.11** ECUACIONES DE REGRESIÓN Y VARIABLES DE YOUNG

| ECUACIÓN | VARIABLES INDEPENDIENTES                | VARIABLE DEPENDIENTE |
|----------|-----------------------------------------|----------------------|
| A        | $\Delta X_5, \Delta X_6, \Delta X_{12}$ | $\Delta Y$           |
| B        | $\Delta X_5, \Delta X_6, \Delta X_{13}$ | $\Delta Y$           |
| C        | $\Delta X_5, \Delta X_6, \Delta X_{14}$ | $\Delta Y$           |
| D        | $\Delta X_5, \Delta X_6, \Delta X_{15}$ | $\Delta Y$           |

TABLA 9.12 MATRIZ DE COEFICIENTES DE CORRELACIÓN

|                 | $\Delta X_5$ | $\Delta X_6$ | $\Delta X_{15}$ |
|-----------------|--------------|--------------|-----------------|
| $\Delta X_5$    | 1.000        | .452         | .113            |
| $\Delta X_6$    | .452         | 1.000        | .122            |
| $\Delta X_{15}$ | .113         | .122         | 1.000           |

del cambio en el ingreso personal se explica mediante las tres variables independientes. La estadística *DW* muestra que no hay problema de autocorrelación. Además, la matriz de coeficientes de correlación de la tabla 9.12 demuestra un bajo nivel de interdependencia entre las tres variables independientes.

Para fines de construcción de índices, las variables independientes en la ecuación final de regresión se convierten en componentes del índice. La ponderación, o peso específico de los componentes puede determinarse a partir de los coeficientes de regresión. (Recuerde que los coeficientes de regresión representan el cambio promedio en la variable dependiente para un incremento de una unidad en la variable independiente.) Sin embargo, como las variables en la ecuación de regresión no tienen la misma unidad de medida (por ejemplo  $\Delta Y$  se mide en miles de dólares y  $\Delta X_{15}$  en miles de kilowatts – hora), se deben transformar a términos relativos los coeficientes de regresión. Esta transformación se logra mediante el cálculo de sus coeficientes  $\beta$ .

$$\beta = b \left( \frac{s}{s_y} \right) \quad (9.13)$$

en donde

$b$  = coeficiente de regresión de la variable independiente

$s$  = desviación estándar de la variable independiente

$s_y$  = desviación estándar de la variable dependiente

Es común encontrar disponibles los valores de todas estas estadísticas en la salida de cómputo de la regresión. Por lo tanto, los coeficientes estandarizados de las tres variables independientes son

$$\beta_5 = .4959 \quad \frac{.4959}{1.2811} = .3871$$

$$\beta_6 = .4833 \quad \frac{.4833}{1.2811} = .3772$$

$$\beta_{15} = \frac{.3019}{1.2811} \quad \frac{.3019}{1.2811} = .2357$$

Por último, debido a que la suma de las ponderaciones en un índice debe ser igual a 100%, se normalizan los coeficientes estandarizados.

| COMPONENTE      | PESO ESPECÍFICO |
|-----------------|-----------------|
| $\Delta X_5$    | .3871           |
| $\Delta X_6$    | .3772           |
| $\Delta X_{15}$ | .2357           |
| Total           | 1.0000          |

Una vez determinados los componentes y sus respectivos pesos específicos, se puede obtener el índice mediante los siguientes pasos:

**Paso 1.** Calcule el porcentaje de cambio de cada componente desde el periodo base.

**Paso 2.** Multiplique el porcentaje de cambio por el peso específico correspondiente.

**Paso 3.** Sume los porcentajes ponderados de cambio obtenidos en el paso 2.

En la fig. 9.4 se compara el índice elaborado para la actividad económica del condado de Spokane con el PNB de EUA, en dólares constantes (1967 = 100).

## PREGUNTAS

1. Young eligió resolver primero el problema de autocorrelación. ¿Hubiera sido mejor eliminar primero la multicolinealidad y después abordar la autocorrelación?
2. ¿Cómo afecta al análisis el tamaño reducido de la muestra?
3. ¿Debería haberse hecho a través del origen la regresión efectuada sobre la primera diferenciación?
4. ¿Hay algún uso potencial para datos retrasados?

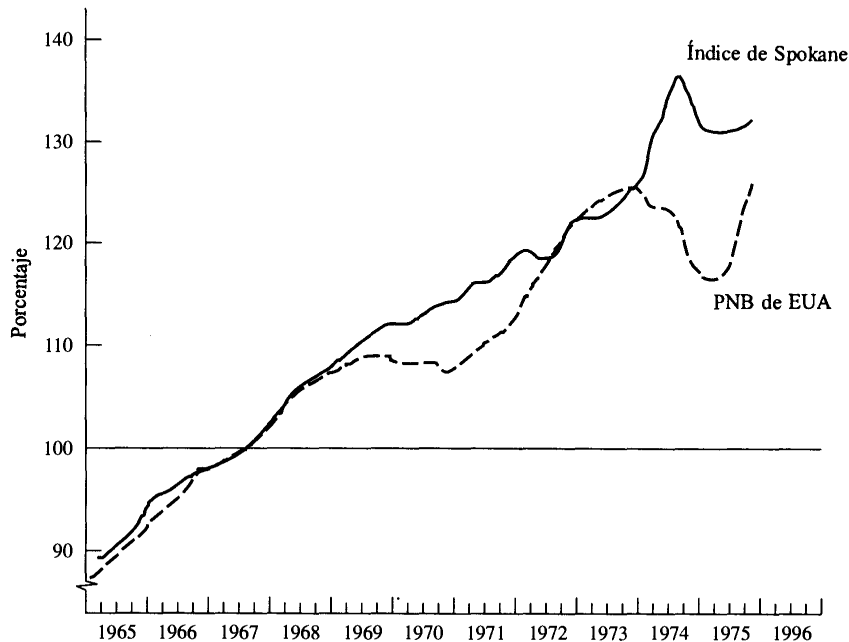


**Figura 9.4** Índice de actividad empresarial del condado de Spokane y PNB de EUA, en dólares constantes (1967 = 100).

5. ¿A qué conclusiones se puede llegar al comparar el índice de actividad empresarial del condado de Spokane con el PNB?

## CASO DE ESTUDIO 9.3

**MR. TUX**

El propietario de Mr. Tux, John Mosby, decidió intentar un modelo autorregresivo en el pronóstico de sus ventas mensuales (los datos aparecieron en el Caso de estudio Mr. Tux al final del cap. 2). Como él sabe que sus datos tienen un fuerte componente estacional (esto es, varían considerablemente de mes a mes), desea reflejar este efecto en su modelo de pronóstico.

Para ello, decide intentar una regresión múltiple con el programa de cómputo Sybil/Runner, utilizando dos variables: un número de periodo para modelar el tiempo y los valores originales de  $Y$  retrasados 12 meses para modelar el componente estacional. Al utilizar  $Y$  retrasada 12 meses no puede utilizar los primeros 12 meses de sus datos de ventas, pero como empezó con 96 periodos, esto le deja un tamaño de muestra de 84.

Ejecuta los datos con el programa utilizando dos variables de predicción y obtiene los siguientes valores clave en la ejecución de cómputo:

| VARIABLE | MEAN     | STANDARD DEVIATION |
|----------|----------|--------------------|
| 2 TREND  | 42.500   | 24.393             |
| 3 LAGGED | 1085.354 | 888.505            |
| 1 SALES  | 1407.093 | 997.094            |

CORRELATION COEFFICIENTS

| VAR | 2    | 3    | 1    |
|-----|------|------|------|
| 2   | 1.00 |      |      |
| 3   | 0.73 | 1.00 |      |
| 1   | 0.70 | 0.95 | 1.00 |

REGRESSION NUMBER 1 DEPENDENT VARIABLE IS SALES

\*\*\*\*\*

| VARIABLE | COEF    | STD.ERROR | T-RATIO | PCT VARIATION EXPLAINED |
|----------|---------|-----------|---------|-------------------------|
| CONSTANT | 235.194 | 68.202    | 3.4     |                         |
| 2 TREND  | 0.609   | 2.033     | 0.3     | 1.04%                   |
| 3 LAGGED | 1.056   | 0.056     | 18.9    | 89.55%                  |

R-SQUARED = 0.906 R = 0.952

F-TEST = 390.024 STD ERROR OF EST. = 309.572

DEGREES OF FREEDOM OF NUMER. = 2 OF DENUMER. = 81

DURBIN-WATSON TEST = 1.82





de los modelos de regresión desarrollados para pronosticar el número de clientes atendidos, se vea afectada por esta condición. Como a ella le gustó el informe que usted presentó utilizando un modelo de análisis de descomposición de series de tiempo, le asigna ahora la tarea de verificar esta situación.

Usted recuerda vagamente, en algún punto de su experiencia educativa, haber desarrollado un modelo autorregresivo y pregunta a Dorothy si le gustaría probar esa posibilidad. Ella decide que sería una buena forma de ocupar su tiempo.

## TAREA

Redacte un memorando para Dorothy que le proporcione la información que solicitó respecto al problema de correlación serial. Desarrolle un modelo autorregresivo y pronostique el resto de 1993. Redacte un segundo memorando para Dorothy presentando los resultados de este análisis.

## VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Hemos decidido emplear un modelo autorregresivo para pronosticar el número de clientes. Usaremos un archivo de datos denominado INDEX que contiene el número de clientes atendidos y el índice de actividad de enero de 1985 a marzo de 1993. Los datos de los primeros cinco meses son

| CLIENTES | INDICE |
|----------|--------|
| 182      | 110    |
| 136      | 108    |
| 99       | 105    |
| 77       | 104    |
| 75       | 104    |

Primero, se selecciona el menú de administración de datos (opción 1) y se elige a continuación la transformación de un archivo de datos (opción 5). este programa nos permite crear nuevas variables o calcular nuevos valores para las variables ya existentes. La pantalla muestra lo siguiente

```

-- VARIABLE LABELS --

1. CLIENTS 2. INDEX

-- KEY WORD LIST --

WRITE IF-THEN SELECT RECODE COMPUTE NEW LAG DIFFERENCE NORMALIZE DUMMY
SUM

NEW CLAG
COMPUTE CLAG = CLIENTS
LAG (1) CLIENTS INDEX

```

El comando NEW se utiliza para crear una nueva variable llamada CLAG. El comando COMPUTE se emplea para asignar a la nueva variable CLAG los mismos valores de la variable CLIENTS. El comando LAG(1) se utiliza para retrasar un periodo las variables CLIENTS e INDEX. Al presionar Escape (ESC), la pantalla muestra

Enter Selection:

- ```
==> 1. Execute the commands
      2. Re-enter the editing mode
      3. Quit & return to Data Management Menu
```

Which?

Después de seleccionar la opción 1 para ejecutar los comandos, en la pantalla aparece

SYNTAX CHECK COMPLETED

Input file name = B:INDEX

Output file name = NOT SPECIFIED

Enter the name for the new output file (or FILES)
-OR- Press <ENTER> to write over the existing file:

Se asigna al nuevo archivo de salida el nombre CRELAG, el cual contiene

CLIENTES	INDICE	CLAG
136	108	182
99	105	136
77	104	99
75	104	77

Se requiere crear un nuevo archivo de etiquetas (CRELAG.LBL). Recuerde que ahora los datos comienzan en febrero.

Ahora, se llama al paquete de pronóstico seleccionando la opción 3 del menú principal. El paquete solicita el nombre del archivo

```
What is the name of the data file? CRELAG
```

El menú le permite elegir entre 13 diferentes técnicas de pronóstico. La pantalla correspondiente se presentó en la sección de Forecast Plus del cap. 5.

Para utilizar como nuestras variables de pronóstico el índice de actividad y los clientes atendidos, retrasadas un periodo, y los clientes atendidos como nuestra variable dependiente, usamos el programa de regresión múltiple, opción 13. La pantalla muestra lo siguiente

```
*** Multiple Regression ***
```

```
Enter the variables to use in the multiple regression. Enter
the dependent variable, followed by the independent variables
(LIST will list the variables):
```

```
CLIENTS, INDEX, CLAG
```

```
Descriptive statistics:      Y
Regression statistics:      Y
Regression coefficients:    Y
Simple correlation matrix:   Y
Partial correlation matrix:  Y
Inverted correlation matrix: N
Number of records to use in residual analysis: 50

Time plot of original data, forecasts, and error: Y
Residual autocorrelation function: Y
Interactive prediction       N
```

```
Press ENTER to accept:
```

Nótese que primero se colocó la variable dependiente, seguida de las variables independientes separadas por comas. Es probable que usted quiera emplear números de variable en lugar de nombres. Para ejecutar este modelo de regresión múltiple, podemos usar las opciones por omisión que se proporcionan. Sin embargo, si queremos calcular una estadística de Durbin-Watson, nuestra elección del número de registros a utilizar en el análisis de residuos necesitará ser de 99. La fig. 9.5 muestra la salida de cómputo cuando las respuestas son Y, Y, Y, Y, Y, N, O, N, Y y N.

Por último, el programa preguntará si desea almacenar los pronósticos o los residuos.

```
Do you want to save the forecasts or residuals (y/n)?
```

CLIENTS - MULTIPLE REGRESSION ANALYSIS					
Variables in the equation - DESCRIPTIVE STATISTICS					
Var.	Variable label	Mean	Standard Dev.		
DV	CLIENTS	121.3673	29.1613		
IV1	INDEX	109.1020	7.7213		
IV2	CLAG	121.5408	29.5733		
REGRESSION STATISTICS					
Coefficient of multiple determination = 0.3853 (Corrected = 0.3789)					
Coefficient of multiple correlation = 0.6207 (Corrected = 0.6156)					
Standard error of multiple estimate = 23.1022 (Corrected = 23.2222)					
F-Ratio	= 29.7763				
Degrees of freedom	= 2 & 95				
Probability of chance	= 0.0000				
Number of observations	= 98				
REGRESSION COEFFICIENTS					
Constant = -113.9440					
Var.	Coeff.	Beta	F-ratio	Prob.	Std. Error
IV1	2.0492	0.5426	18.868	0.000	0.4718
IV2	0.0966	0.0980	0.615	0.435	0.1232
SIMPLE CORRELATION MATRIX					
		DV	IV1		
IV1	I	0.618			
IV2	I	0.513	0.765		
PARTIAL CORRELATION MATRIX					
		IV1			
IV1	I				
IV2	I	0.765			
CLIENTS - MULTIPLE LINEAR REGRESSION					
RESIDUAL AUTOCORRELATION FUNCTION					
Standard Deviation of the Residuals = 22.8628					
Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.03	-0.29	:	[*:]	:
2	0.13	1.26	:	[:***]	:
3	0.03	0.26	:	[:*]	:
4	0.04	0.35	:	[:*]	:

Figura 9.5 Modelo autorregresivo para el Caso de estudio 9.4.

5	0.02	0.19	:	[	:	]	:
6	-0.03	-0.29	:	[	*	:	:
7	-0.00	-0.02	:	[	:	]	:
8	-0.01	-0.06	:	[	:	]	:
9	-0.07	-0.70	:	[	**	:	:
10	0.00	0.01	:	[	:	]	:
11	-0.02	-0.17	:	[	:	]	:
12	0.19	1.83	:	[	:	*****	:
13	-0.07	-0.64	:	[	**	:	:
14	0.08	0.75	:	[	:	**	:
15	-0.11	-1.06	:	[	***	:	:
16	-0.06	-0.56	:	[	**	:	:
17	0.15	1.40	:	[	:	*****	:
18	-0.20	-1.80	:	[	*****	:	:
19	-0.06	-0.52	:	[	*	:	:
20	-0.06	-0.51	:	[	:	0.51	:
21	0.05	0.43	:	[	:	*	:
22	-0.14	-1.22	:	[	*****	:	:
23	-0.03	-0.27	:	[	*	:	:
24	0.21	1.78	:	[	:	*****	:
25	-0.09	-0.70	:	[	**	:	:
26	0.04	0.32	:	[	:	*	:
27	-0.02	-0.18	:	[	*	:	:
28	0.00	0.01	:	[	:	]	:
29	0.15	1.25	:	[	:	*****	:
30	-0.33	-2.61	:	**	[	*****	:
31	0.07	0.49	:	[	:	**	:
32	-0.21	-1.55	:	[	*****	:	:

[] = Estimated Two-Standard Error Limits							

Box-Pierce Chi-Square Statistic with 32 Degrees of Freedom = 40.7
 Probability = .14

Figura 9.5 (continuación)

PRONÓSTICOS MEDIANTE HOJA DE CÁLCULO

El problema. En el cap. 4, ejemplo 4.4, se analizaron las ventas trimestrales (véase tabla 4.5) de Marine Outboard. Desarrolle un modelo estacional autorregresivo para estos datos.

La solución mediante hoja de cálculo

1. Para acceder el archivo OUTBOARD.WK1 escriba /FILE RETRIEVE o /FR. Lotus solicitará el nombre del archivo con el mensaje

Name of file to retrieve: A:\

En la tercera línea, usted verá el nombre de la hoja de cálculo que desea acceder. Después de señalar OUTBOARD.WK1, presione <Enter>. En la fig. 8.10 se presenta una visualización de esta hoja de cálculo.

2. Para borrar los títulos, escriba /WORKSHEET DELETE ROW o /WDR y el programa le pedirá el rango de celdas a borrar

Enter range of rows to delete:

Escriba el rango A1...A6. Borre en forma similar las columnas E a la K.

3. Escriba a partir de la columna A los títulos que desee utilizar.
4. Utilice el comando /MOVE para desplazar el número de periodo a la columna E. Para borrar la columna C, escriba /WORKSHEET DELETE COLUMN o /WDC.
5. Para crear una columna con las ventas retrasadas un periodo, presione /COPY o /C. Coloque el apuntador de celdas en C3. El programa le solicita el rango de celdas a copiar:

Enter range to copy FROM: C3..C3

Presione la tecla de punto (.) para anclar lo resaltado comenzando en la celda C5. Oprima la tecla de flecha hacia abajo <↓> para extender lo resaltado hasta C38. Oprima <Enter>. Coloque ahora el indicador de celdas en D7 y presione <Enter> para copiar la columna. La fig. 9.6 muestra el resultado.

G3:	READY										
	A	B	C	D	E	F	G	H	I	J	K
1	Year	Qt	Sales	T	Lagged						
2											
3	1984	1	147.6	1							
4		2	251.8	2							
5		3	273.1	3							
6		4	249.1	4							
7	1985	1	139.3	5	147.6						
8		2	221.2	6	251.8						
9		3	260.2	7	273.1						
10		4	259.5	8	249.1						
11	1986	1	140.5	9	139.3						
12		2	245.5	10	221.2						
13		3	298.8	11	260.2						
14		4	287	12	259.5						
15	1987	1	168.8	13	140.5						
16		2	322.6	14	245.5						
17		3	393.5	15	298.8						
18		4	404.3	16	287						
19	1988	1	259.7	17	168.8						
20		2	401.1	18	322.6						
04-Jan-94 04:00 PM NUM											

Figura 9.6 Hoja de cálculo de Lotus para un modelo autorregresivo.

1-2-3 Spreadsheet Computing Series

Year	Qt	Sales	T	Lagged	
1984	1	147.6	1		Regression Output:
	2	251.8	2		
	3	273.1	3		
	4	249.1	4		
1985	1	139.3	5	147.6	Constant 86.656
	2	221.2	6	251.8	Std Err of Y Est 52.914
	3	260.2	7	273.1	R Squared 0.6296
	4	259.5	8	249.1	No. of Observations 36
1986	1	140.5	9	139.3	Degree of Freedom 33
	2	245.5	10	221.2	X Coefficient (-1.118 0.7990)
	3	298.8	11	260.2	
	4	287	12	259.5	
1987	1	168.8	13	140.5	
	2	322.6	14	245.5	Std Err of Coe 0.8740 0.1069
	3	393.5	15	298.8	
	4	404.3	16	287	
1988	1	259.7	17	168.8	
	2	401.1	18	322.6	
	3	464.6	19	393.5	
	4	479.7	20	404.3	
1989	1	264.4	21	259.7	
	2	402.6	22	401.1	
	3	411.3	23	464.6	
	4	385.9	24	479.7	
1990	1	232.7	25	264.4	
	2	309.2	26	402.6	
	3	310.7	27	411.3	
	4	293	28	385.9	
1991	1	205.1	29	232.7	
	2	234.4	30	309.2	
	3	285.4	31	310.7	
	4	258.7	32	293	
1992	1	193.2	33	205.1	
	2	263.7	34	234.4	
	3	292.5	35	285.4	
	4	315.2	36	258.7	
1993	1	178.3	37	193.2	
	2	274.5	38	263.7	
	3	295.4	39	292.5	
	4	311.8	40	315.2	

Figura 9.7 Modelo autorregresivo para Outboard Marine.

6. Escriba /DATA REGRESSION o /DR, que equivale a seleccionar DATA FILE en el menú principal y REGRESSION en el submenú. El submenú de regresión muestra lo siguiente:

X-Range	Y-Range	Output-Range	Intercept	Reset	Go	Quit

Se selecciona X – Range y el programa solicita las variables independientes o rango de X a utilizar

Enter independent variables, or X range:

Usted puede, ya sea escribir D7 . . E42 o bien definir las variables X resaltándolas. Se selecciona Y – Range y el programa solicita la variable dependientes o rango de Y a utilizar

Enter dependent variable, or Y-range:

Indique el rango C7...C42 para definir el rango de la variable dependiente Y. Después se selecciona Output – Range y en la pantalla se pide el rango en el que se colocarán los resultados

Enter output range:

Una vez seleccionado el rango correspondiente, escriba GO para ejecutar la regresión. Los resultados aparecen en la fig. 9.7.

PAQUETE DE CÓMPUTO MINITAB

Los comandos de Minitab para analizar los datos de Novak del ejemplo 9.3 son

```
MTB > READ 'NOVAK.DAT' C1-C3
      17 ROWS READ

      ROW      C1      C2      C3
      1      8.0      336.1    5.5
      2      8.2      349.4    5.5
      3      8.5      362.9    6.7
      4      9.2      383.9    5.5
      ...

MTB > NAME C1 'SALES' C2 ' INCOME' C3 'UNEMP'
MTB > REGRESS C1 2 PREDITORS C2 C3, RES IN C4, DHATS C5;
SUBC> RESIDUALS C6;
SUBC> DW.
```

The regression equation is
 SALES = - 0.014 + 0.0297 INCOME - 0.350 UNEMP

Predictor	Coef	Stdev	t-ratio	p
Constant	-0.0140	0.2498	-0.06	0.956
INCOME	0.0297492	0.0002480	119.96	0.000


```
UNEMP      -0.34987      0.04656      -7.51      0.000
```

```
s = 0.2199      R-sq = 99.9%      R-sq(adj) = 99.9%
```

Analysis of Variance

SOURCE	DF	SS	MS	F	P
Regression	2	738.88	369.44	7637.92	0.000
Error	14	0.68	0.05		
Total	16	739.56			

SOURCE	DF	SEQ SS
INCOME	1	736.15
UNEMP	1	2.73

Unusual Observations

Obs.	INCOME	SALES	Fit	Stdev.Fit	Residual	St Resid
17	1077	29.1000	29.0430	0.1713	0.0570	0.41 X

X denotes an obs. whose X value gives it large influence.

Durbin-Watson statistic = 1.98

```
MTB > NAME C5 'ESTIMATE' C6 'RESIDUAL'
```

```
MTB > PRINT C1 C2 C3 C5 C6
```

ROW	SALES	INCOME	UNEMP	ESTIMATE	RESIDUAL
1	8.0	336.1	5.5	8.0604	-0.060394
2	8.2	349.4	5.5	8.4561	-0.256058
3	8.5	362.9	6.7	8.4378	0.062171
4	9.2	383.9	5.5	9.4824	-0.282405
5	10.2	402.8	5.7	9.9747	0.225310
6	11.4	437.0	5.2	11.1670	0.232953
7	12.8	472.2	4.5	12.4591	0.340875
8	13.6	510.4	3.8	13.8405	-0.240452
9	14.6	544.5	3.8	14.8549	-0.254899
10	16.4	588.1	3.6	16.2219	0.178064
11	17.8	630.4	3.5	17.5153	0.284685
12	18.6	685.9	4.9	18.6766	-0.076576
13	20.0	742.8	5.9	20.0194	-0.019434
14	21.9	801.3	5.6	21.8647	0.035278
15	24.9	903.1	4.9	25.1381	-0.238096
16	27.3	983.6	5.6	27.2880	0.012003
17	29.1	1076.7	8.5	29.0430	0.056976

```
MTB > STOP
```

- El comando READ se usa para almacenar los datos de ventas de Novak en C1, el ingreso personal disponible en C2 y la tasa de desempleo en C3.

- El comando TRESIDS se emplea para calcular los residuos estandarizados y almacenarlos en C4.
- El comando DHATS se utiliza para calcular las estimaciones de Y y almacenarlas en C5.
- El subcomando RESIDUAL se usa para calcular la diferencia entre las observaciones reales (C1) y las estimaciones de Y (C5). El resultado se almacena en C6.
- El subcomando DW se emplea para calcular la estadística de Durbin–Watson.

PAQUETE DE CÓMPUTO SAS

Problema. En el ejemplo 4.4, Perkin Kendell, analista de Outboard Marine, usó la autocorrelación para determinar que sus datos eran estacionales. Ahora, él desea desarrollar un modelo autorregresivo utilizando el tiempo y las ventas retrasadas cuatro periodos para predecir las ventas trimestrales.

Solución mediante SAS

```
$ SAS/CC=PRINT
OPTIONS PAGESIZE=60 LINESIZE=80;
TITLE "SEASONAL AUTOREGRESSIVE MODEL";
DATA OUTBOARD;
  INPUT YT XT;
  LABEL YT=SALES
        YT4=LAGSALES
        XT=TIME;
  YT4=LAG4(YT);
CARDS;
147.6 1
251.8 2
273.1 3
249.1 4
139.3 5
221.2 6
260.2 7
259.5 8
140.5 9
245.5 10
298.8 11
287.0 12
168.8 13
```

```
322.6 14
393.5 15
404.3 16
259.7 17
401.1 18
464.6 19
479.7 20
264.4 21
402.6 22
411.3 23
385.9 24
232.7 25
309.2 26
310.7 27
293.0 28
205.1 29
234.4 30
285.4 31
258.7 32
193.2 33
263.7 34
292.5 35
315.2 36
178.3 37
274.5 38
295.4 39
311.8 40
PROC REG;
  MODEL YT=YT4 XT/ DW;
EBDSAS;
```

- El comando TITLE da nombre a la ejecución de SAS.
- La declaración DATA asigna nombre a los datos.
- La declaración INPUT da nombre y asigna el orden correcto a los dos campos de datos.
- La declaración LABEL proporciona etiquetas para cada variable especificada.
- La declaración $YT4=LAG4(YT)$ asigna a YT4 el valor de YT retrasada 4 periodos. Por ejemplo, el valor de YT en el caso 1 es de 147.6. En consecuencia, el valor de YT4 en el caso 5 es también 147.6.
- La declaración CARDS indica que las siguientes 40 líneas son imágenes de tarjetas que representan la variable YT (ventas) y la variable XT (tiempo).
- La declaración PROC REG y el subcomando MODEL indican que YT es la variable dependiente y que las variables de predicción son XT y YT4.
- El subcomando DW calcula la estadística de Durbin–Watson. La salida de cómputo para esta ejecución de SAS se muestra en la fig. 9.8.

SEASONAL AUTOREGRESSIVE MODEL					
Model: MODEL1					
Dependent Variable: YT		SALES			
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	157086.20093	78543.10046	28.052	0.0001
Error	33	92397.06657	2799.91111		
C Total	35	249483.26750			
Root MSE	52.91419		R-square	0.6296	
Dep Mean	290.67500		Adj R-sq	0.6072	
C.V.	18.20390				
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob>T
INTERCEP	1	86.656861	33.43635742	2.592	0.0141
YT4	1	0.799002	0.10692437	7.473	0.0001
XT	1	-1.118219	0.87406456	-1.279	0.0097
Variable	DF	Variable Label			
INTERCEP	1	Intercept			
YT4	1	LAGSALES			
XT	1	TIME			
Durbin-Watson D		0.508			
For Number of Obs.)		36			
1st Order Autocorrelation		0.725			

Figura 9.8 Modelo autorregresivo estacional para Outboard Marine.

EJEMPLO DE TSP

Un paquete estadístico que con frecuencia usan los economistas y que funciona tanto en macro como microcomputadoras es TSP. Este paquete permite al usuario ejecutar diversas técnicas de uso común en el área de econometría.

En esta sección se demostrará TSP para ilustrar el uso de paquetes de programas estadísticos en la resolución de un problema de correlación serial mediante el método de Cochrane – Orcutt.

EL PROBLEMA

En el ejemplo 9.7 se utilizó el ingreso personal disponible para predecir las ventas de Sears. Hubo un problema de correlación serial, de modo que se utilizó el método de Cochrane–Orcutt para obtener un modelo válido.

La solución mediante TSP

El archivo de comandos (consulte a su instructor sobre la información que requiera para ejecutar TSP en su sistema de cómputo en particular) para ejecutar los datos del ejemplo 9.7 es

```

PROGRAM
LINE *****
1 NAME SEARS 'SEARS'
1 FREQ A;
2 SMPL 1,21;
3 LOAD SALES;
3 3.307, 3.556, 3.601, 3.721, 4.036, 4.134, 4.268, 4.578, 5.093,
3 5.716, 6.357, 6.769, 7.296, 8.178, 8.844, 9.251, 10.006, 11.2,
3 12.5, 13.101, 13.640;
4 LOAD INCOME;
4 273.4, 291.3, 306.9, 317.1, 336.1, 349.4, 362.9, 383.9, 402.8,
4 437, 472.2, 510.4, 544.5, 588.1, 630.4, 685.9, 742.8, 801.3
4 903.1, 983.6, 1076.7;
5 AR1 (PRINT, METHOD=CORC) SALES C INCOME;
6 END;

EXECUTION
*****

```

Después de la ejecución, la salida de cómputo para los datos de Sears es:

```

EQUATION 1
*****
FIRST-ORDER SERIAL CORRELATION OF THE ERROR
COCHRANE-ORCUTT ITERATIVE TECHNIQUE

MEAN OF DEPENDENT VARIABLE =      7.29225
STANDARD DEVIATION =      3.37732
DEPENDENT VARIABLE: SALES

OPTIONS FOR ROUTINE
*****
FAIR   = TRUE           MAXIT   = 20           METHOD  = CORC
PRINT  = TRUE           RMAX    = 0.90000       RMTN    = -0.90000
RSTART = 0              RSTEP   = 0.10000       TOL     = 0.0050
TSCS   = FALSE          UNNORM  = FALSE        WEIGHT  =

INITIAL ESTIMATE
SSR =      2.2419
PARAMETER ESTIMATES:
-.5243      0.1405E-01
CONVERGENCE ACHIEVED AFTER 7 ITERATIONS
ITERATION      RHO      STD ERROR OF EQN
*****
1             0.815355    0.248105
2             0.870710    0.244495
3             0.897917    0.242975
4             0.913437    0.242195

```

5	0.923445	0.241729
6	0.930504	0.241421
7	0.935808	0.241201
FINAL VALUE OF RHO = 0.935808		
STANDARD ERROR OF RHO = 0.788240E-01		
T-STATISTIC FOR RHO = 11.8721		
STATISTICS BASED ON RHO-TRANSFORMED VARIABLES		

SUM OF SQUARED RESIDUALS = 1.04720		
STANDARD ERROR OF THE REGRESSION = 0.241201		
MEAN OF DEPENDENT VARIABLE = 0.951593		
STANDARD DEVIATION = 0.499369		
R-SQUARED = 0.778978		
ADJUSTED R-SQUARED = 0.766699		
F-STATISTIC(1, 18) = 63.4400		
DURBIN-WATSON STATISTIC(ADJ FOR 0 GAPS) = 1.0813		
NUMBER OF OBSERVATIONS = 20		
ESTIMATED STANDARD		
VARIABLE	COEFFICIENT	ERROR T-STATISTIC
C	2.340258	1.778342 1.315977
INCOME	0.1093294E-01	0.1372636E-02 7.964924

BIBLIOGRAFÍA

- COCHRANE, D. y ORCUTT, G. N. "Application of Least Squares to Relationships Containing Autocorrelated Error Terms." *Journal of the American Statistical Association* 44 (1949): 32 – 61.
- DURBIN, J. y WATSON, G. S. "Testing for Serial Correlation in Least Squares Regression II." *Biometrika* 38 (1951): 159 – 178.
- JOHNSTON, J. *Econometric Methods*, 2a. ed. Nueva York: McGraw – Hill, 1972.
- KING, J. L. y BESSLER, D. A. "A comparison of Multivariate Forecasting Procedures for Economic Time Series." *International Journal of Forecasting* 1 (1) (1985): 5 – 24.
- LESSER, C. E. V. "A survey on Econometrics". *Journal of the Royal Statistical Society, Series A*, 34 (1968): 530 – 566.
- MAKRIDAKIS, S., WHEELWRIGHT, S. C y MCGEE, V. E. *Forecasting: Methods and Applications*. Nueva York: John Wiley & Sons, 1983.
- PINDYCK, R. S. y RUBINFELD, D. L. *Econometric Models and Economic Forecasts*. Nueva York: McGraw–Hill, 1976, pp. 108 – 113.
- 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)

ACTIVIDAD

Los enfoques de pronóstico de atenuación exponencial (cap. 5), del análisis de correlación y de regresión (caps. 6–7) y del análisis de descomposición de series de tiempo (cap. 8), asumen que los valores de la serie que se pronostica son independientes estadísticamente o no relacionados entre sí. En los caps. 6 y 7 se emplearon ecuaciones de regresión lineal para pronosticar la variable dependiente mediante sustitución de valores de la(s) variable(s) independiente(s) en la ecuación de regresión y calculando un valor de predicción para Y .

Este capítulo trata con un tipo de modelos que produce pronósticos con base en una síntesis de los patrones históricos en los datos. Los modelos de promedio móvil autorregresivo integrado (ARIMA: Autoregressive integrated moving-average) son una clase especializada de técnicas de filtración que ignoran por completo a las variables independientes en la formulación de pronósticos. ARIMA es un dispositivo altamente refinado de ajuste de curvas que utiliza valores reales y anteriores de la variable dependiente, para producir pronósticos precisos de corto plazo. Un ejemplo de estos pronósticos es la predicción de precios del mercado de valores, creados por analistas corredores de bolsa y que se basan por completo en patrones anteriores del movimiento de los precios de acciones.

La metodología ARIMA es apropiada si las observaciones de una serie histórica son dependientes estadísticamente o si están relacionadas entre sí. En este cap. se exponen los métodos de Box–Jenkins.

TÉCNICA DE BOX-JENKINS

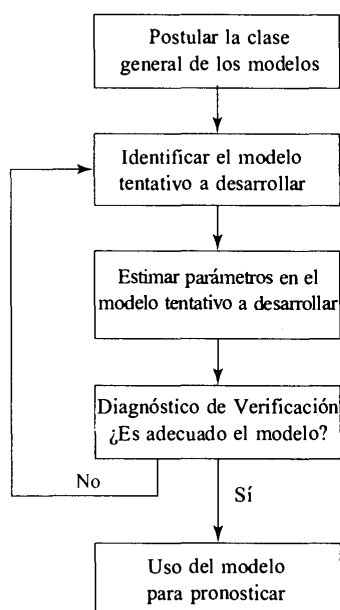
El *método Box–Jenkins* de pronóstico es diferente de la mayoría de los métodos. Esta técnica *no* asume ningún patrón particular en los datos históricos de la serie a pronosticar. Utilizan un enfoque iterativo de identificación de un modelo útil a partir de modelos de tipo general. El modelo elegido se verifica contra los datos históricos para ver si describe la serie con precisión. El modelo se ajusta bien si los residuos entre el modelo de pronóstico y los puntos de datos históricos son reducidos, distribuidos de manera aleatoria e independientes. Si el modelo especificado *no* es satisfactorio, se repite el proceso utilizando otro modelo diseñado

para mejorar el original. Este proceso se repite hasta encontrar un modelo satisfactorio. La fig. 10.1 ilustra el enfoque.

Los modelos ARIMA o modelos de promedio móvil autorregresivo integrado son un tipo general de los modelos de Box-Jenkins para series de tiempo estacionarias. Recuerde que una serie histórica estacionaria es aquella cuyo valor promedio no cambia a través del tiempo. Este grupo incluye a los modelos AR sólo con términos autorregresivos, los modelos MA sólo con términos de promedio móvil y los modelos ARIMA que comprenden tanto términos autorregresivos como de promedio móvil. La metodología de Box-Jenkins permite al analista seleccionar el modelo que mejor se ajuste a sus datos.

Se puede efectuar la selección del modelo apropiado comparando la distribución de los coeficientes de autocorrelación de la serie histórica que se está ajustando, con las distribuciones teóricas para los distintos modelos. En las figs. 10.2, 10.3 y 10.4 se muestran distribuciones teóricas de los coeficientes de autocorrelación para algunos de los modelos ARIMA más comunes.

Las técnicas de Box-Jenkins aplican métodos autorregresivos y de promedio móvil a los problemas de pronóstico de series de tiempo.



Fuente: G. P. Box y G. M. Jenkins, *Time Series Analysis Forecasting and Control* (San Francisco: Holden-Day, 1970), p. 19. Reimpreso con autorización.

Figura 10.1 Diagrama de flujo del método de Box-Jenkins.

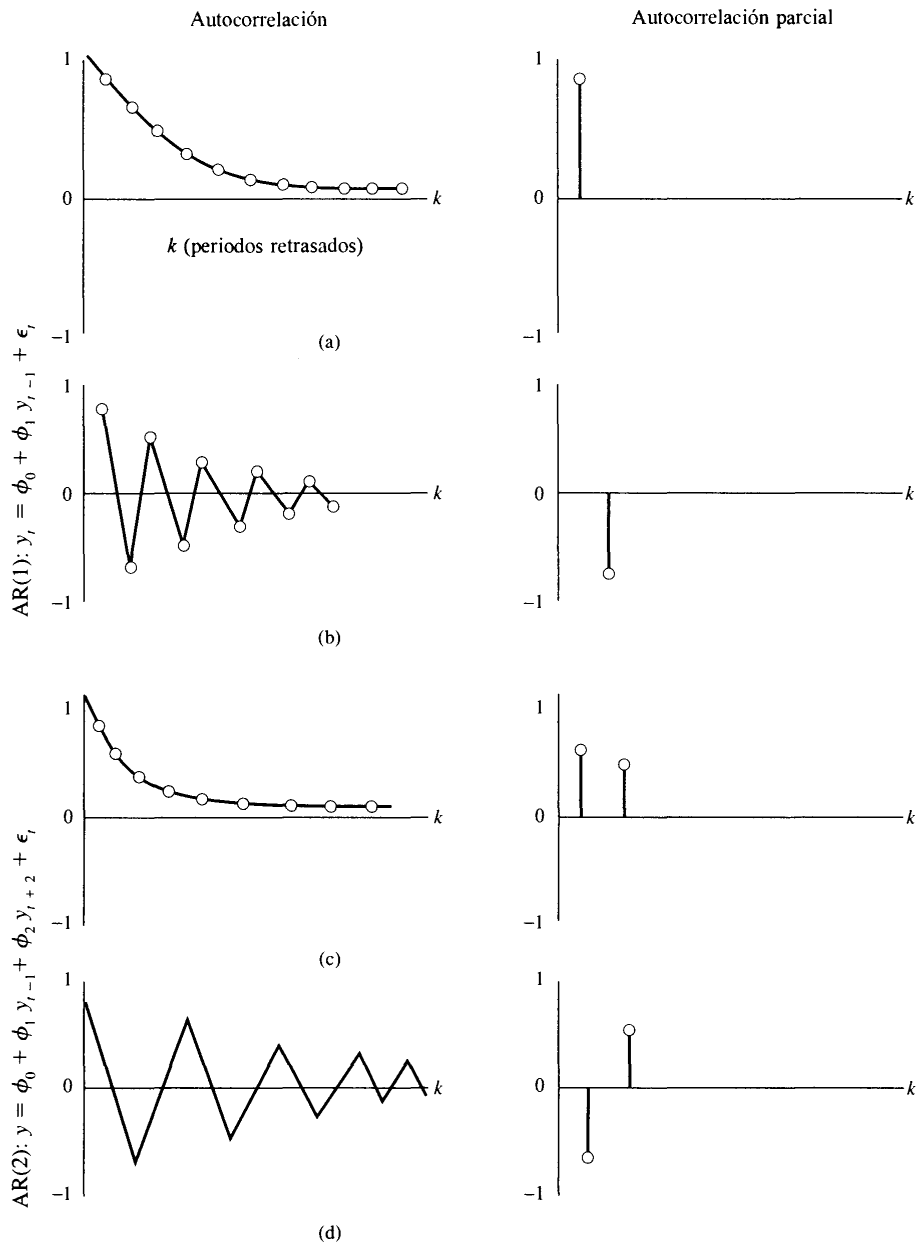


Figura 10.2 Coeficientes de autocorrelación y de autocorrelación parcial de los modelos AR(1) y AR(2).

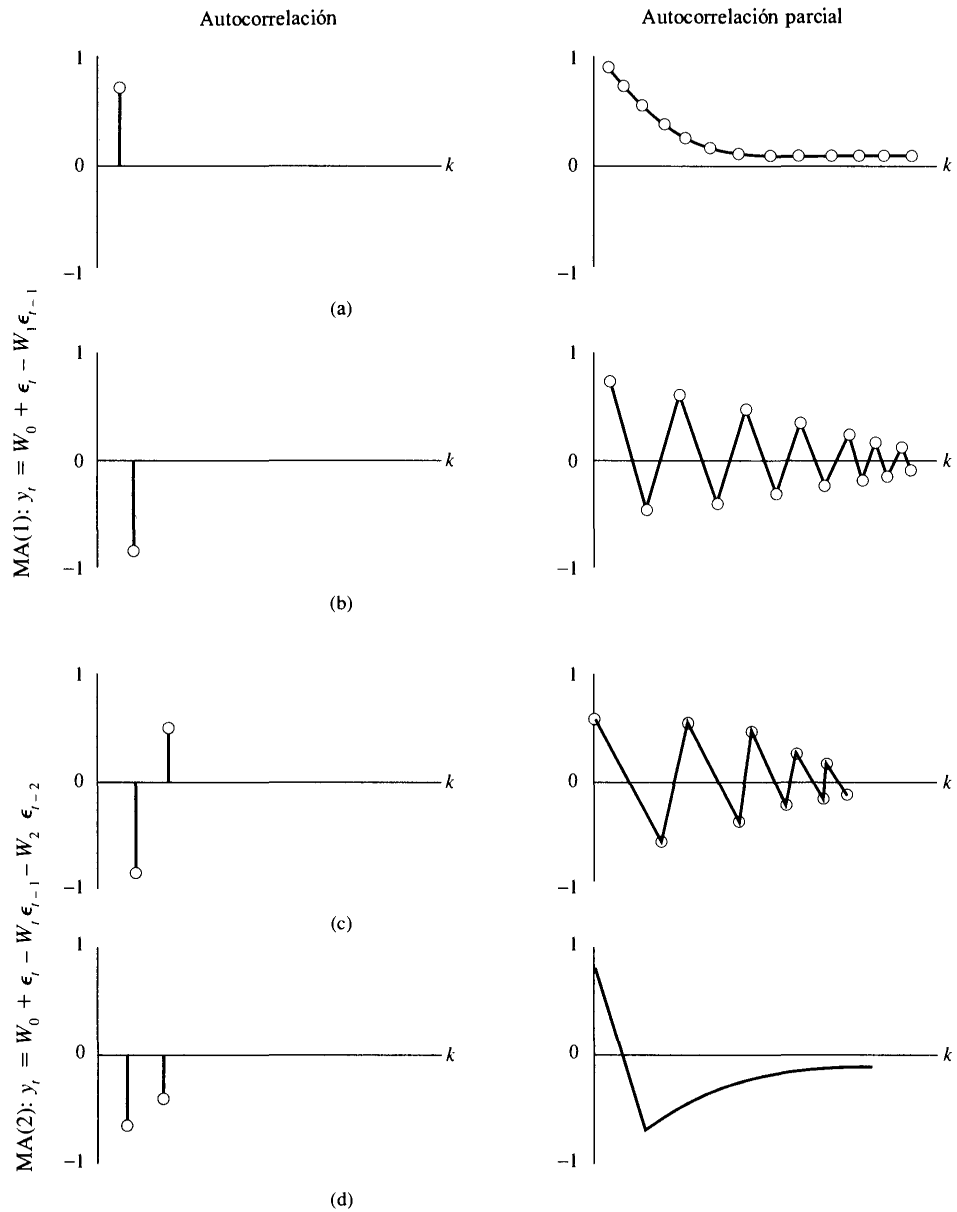


Figura 10.3 Coeficientes de autocorrelación y de autocorrelación parcial de los modelos MA(1) y MA(2).

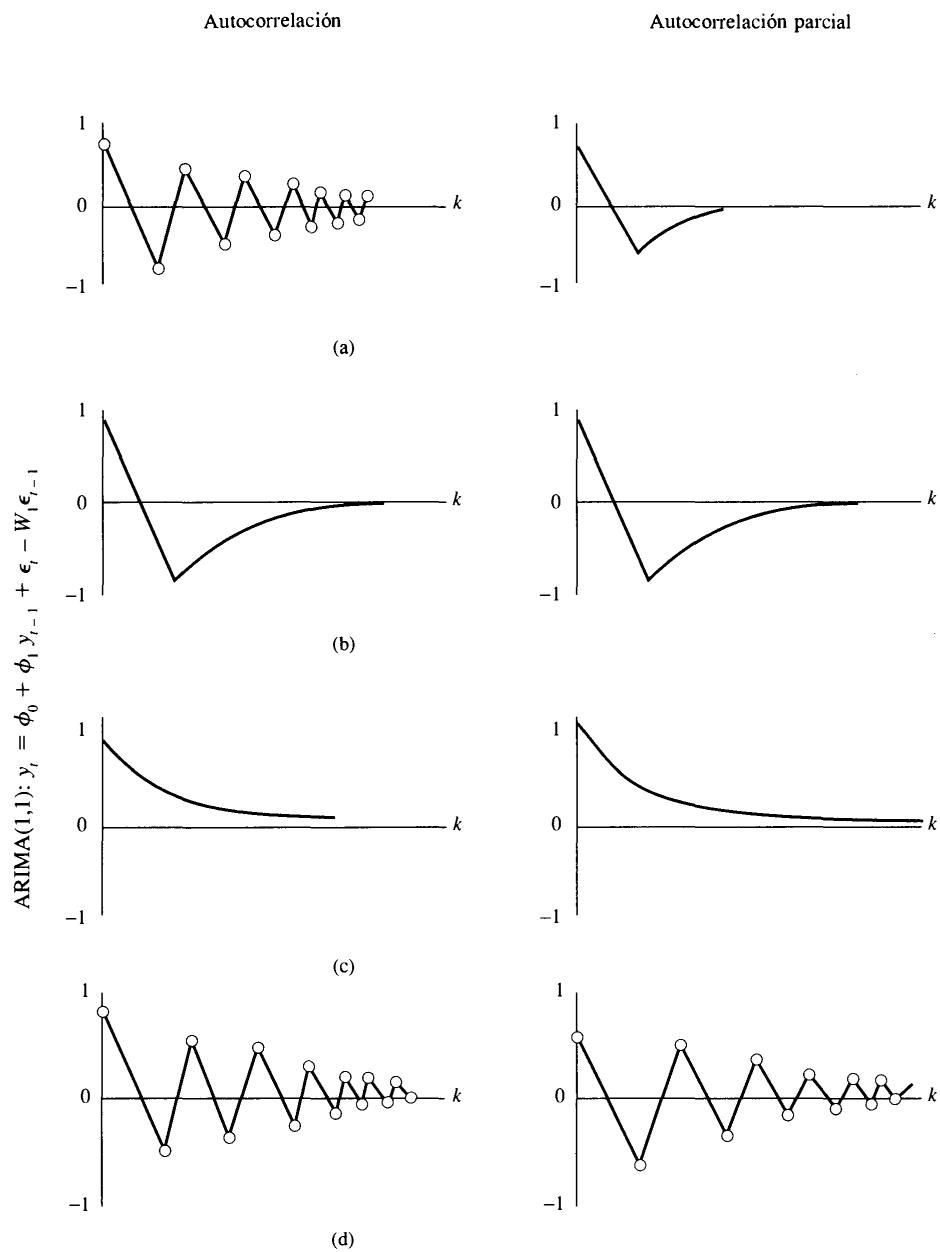


Figura 10.4 Coeficientes de autocorrelación y de autocorrelación parcial de un modelo mixto ARIMA(1,1).

Al seleccionar un modelo, recuerde que las distribuciones que se muestran son teóricas y que es muy improbable que las autocorrelaciones de datos reales sean idénticas exactamente a cualquiera de las distribuciones teóricas. Sin embargo, usted debe, mediante prueba y error, poder ubicar en forma adecuada la mayoría de las series de tiempo de datos, y al ganar experiencia, la tarea se hará más fácil.

AUTOCORRELACIONES PARCIALES

Al principio, el analista pudiera no darse cuenta del orden apropiado del proceso autorregresivo para ajustarlo a una serie histórica. A este mismo tipo de problema se enfrentó al decidir el número de variables independientes a incluir en un modelo de regresión múltiple. Las *autocorrelaciones parciales* se emplean para ayudar a identificar el grado de relación entre los valores reales de una variable y valores anteriores de la misma, mientras que se mantienen constantes los efectos de las otras variables (periodos retrasados).

La exposición de la derivación de autocorrelaciones parciales rebasa el alcance de este texto.¹ En su lugar, esta sección se concentrará en la forma totalmente mecánica en la cual puede aplicarse a cada grupo de modelos.

Modelo autorregresivo

Un modelo autorregresivo toma la forma

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (10.1)$$

en donde

Y_t = variable dependiente

$Y_{t-1}, Y_{t-2}, Y_{t-p}$ = variables independientes que son variables dependientes desfasadas un número específico de periodos

$\phi_0, \phi_1, \phi_2, \phi_p$ = coeficientes de regresión

ϵ_t = término de residuo que representa sucesos aleatorios no explicados por el modelo

Los modelos autorregresivos se presentaron en la ecuación 9.4. Sin embargo, la ecuación 10.1 difiere en varias formas importantes. En la ecuación 9.4 los coeficientes de regresión se estiman mediante el método lineal de mínimos cuadrados. En la ecuación 10.1 los coeficientes de regresión se encuentran por medio de un método de mínimos cuadrados no lineal.

Por lo regular el método de mínimos cuadrados no lineal utiliza una técnica de solución iterativa para calcular los parámetros en vez de usar un cálculo directo. Se emplean estimaciones preliminares como puntos iniciales; luego estas estimaciones se mejoran

¹ Para una exposición de autocorrelaciones parciales, consúltese a G. Box y G. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco: Holden-Day, 1976).

sistemáticamente hasta encontrar valores óptimos. Además, la varianza para la ecuación 10.1 se calcula de una forma distinta que toma en cuenta el hecho de que las variables independientes están relacionadas entre sí. Por último, La ecuación 10.1 pudiera o no contener un término constante. No se emplea el término constante cuando los valores de la variable dependiente (las Y) se expresan como derivaciones de su media ($Y' = Y - \bar{Y}$).

La fig. 10.2 muestra las ecuaciones de un modelo AR de orden 1, modelo AR(1) y de un modelo AR de orden 2, modelo AR(2). Se pueden agregar términos para representar un modelo AR(p), en donde p es el número de observaciones anteriores a incluir en el pronóstico del siguiente periodo. Las figs. 10.2(a) y (b) ilustran el comportamiento de las funciones teóricas de autocorrelación y autocorrelación parcial para un modelo AR(1). Nótese qué tan diferente se comportan las funciones de autocorrelación y de autocorrelación parcial. Los coeficientes de autocorrelación descienden a cero en forma gradual, mientras que los coeficientes de autocorrelación parcial caen a cero después del primer periodo de retraso. Las figs. 10.2(c) y (d) muestran un modelo AR(2). De nuevo, los coeficientes de autocorrelación descienden a cero, mientras que los coeficientes de autocorrelación parcial caen a cero, después del segundo periodo de retraso. En general, este tipo de patrón se mantendrá para cualquier modelo AR(p). Sin embargo, debe recordarse que las funciones de autocorrelación de la muestra diferirán de estas funciones teóricas debido a la variación de la muestra.

Ejemplo 10.1

En la tabla 10.1 se muestra un ejemplo del uso de un modelo de pronóstico AR(2). Suponga que se muestran las ventas de los últimos cinco periodos de una serie que contiene 100 observaciones. Se elige el modelo AR(2) y el programa de cómputo de Box-Jenkins calcula los coeficientes de regresión de $\phi_1 = .59$ y $\phi_2 = .40$. El pronóstico del periodo t sin utilizar una constante en la ecuación es

$$\begin{aligned}\hat{Y}_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \\ &= .59(250) + .40(240) = 243.5\end{aligned}$$

Por desgracia, no todas las series de datos pueden manejarse con modelos autorregresivos. Por esta razón, el enfoque de Box-Jenkins considera también el modelo de promedio móvil (MA).

TABLA 10.1 PRONÓSTICO CON UN MODELO AUTORREGRESIVO PARA EL EJEMPLO 10.1

PERIODO	VALORES	VALORES DE PRONÓSTICO	RESIDUO ϵ_t
$t - 5$	240	238.3	1.7
$t - 4$	230	240.1	-10.1
$t - 3$	225	231.7	-6.7
$t - 2$	240	224.8	+15.2
$t - 1$	250	231.6	+18.4
t		243.5	

Modelos de promedio móvil

Un modelo de promedio móvil toma la forma

$$Y_t = w_0 + \epsilon_t - w_1\epsilon_{t-1} - w_2\epsilon_{t-2} + \dots + w_q\epsilon_{t-q} \quad (10.2)$$

en donde

Y_t = variable dependiente

w_0, w_1, w_2, w_q = peso específico

ϵ_t = residuo o error

$\epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-q}$ = valores previos de residuos

La ecuación 10.2 es similar a la ecuación 10.1 excepto que la variable dependiente Y_t depende de los valores previos de los residuos, en vez de la misma variable. Los modelos de promedio móvil (MA) proporcionan pronósticos de Y_t con base en una combinación lineal de errores anteriores, mientras que los modelos autorregresivos (AR) expresan Y_t como una función lineal de cierto número de valores anteriores reales de Y_t . Es una costumbre presentar los pesos específicos con coeficientes negativos, aun cuando los pesos pueden ser positivos o negativos. La suma de $w_1 + w_2 + \dots + w_q$ no necesita ser igual a 1, y los valores de w_1 no se “mueven” con las nuevas observaciones, como sucede con los cálculos de promedio móvil en el cap. 5. Nótese que el nivel promedio μ de una serie MA(q) es igual a un término constante, w_0 , en el modelo, ya que $E(\epsilon_t) = 0$ para todos los valores de t . El nombre *promedio móvil* pudiera parecer inadecuado ya que de hecho el modelo es similar a la atenuación exponencial.

La fig. 10.3 muestra las ecuaciones de un modelo MA de orden 1, MA(1) y de un modelo MA(2). Se pueden incorporar términos para representar un modelo MA(q), en donde q es el número de términos de error anteriores a incluir en el pronóstico del siguiente periodo. Las figs. 10.3(a) y (b) ilustran también el comportamiento de los coeficientes de autocorrelación teóricos del modelo MA(1). Nótese que afortunado es que las funciones de autocorrelación y de autocorrelación parcial se comporten muy diferente en los modelos AR y MA. Los coeficientes de autocorrelación del modelo MA(1) caen a cero después del primer periodo de retraso, mientras que los coeficientes de autocorrelación parcial descenden a cero en forma gradual. Además, los coeficientes de autocorrelación del modelo MA(2) caerán a cero después del segundo periodo de retraso, mientras que los parciales descenderán gradualmente a cero. [Véase figs. 10.3(c) y (d).] Una vez más, se debe mencionar que las funciones de autocorrelación de la muestra diferirán de estas funciones teóricas debido a la variación de la muestra.

Ejemplo 10.2

Utilizando los términos de error que aparecen en la tabla 10.1, se presenta un ejemplo del uso de un modelo de pronóstico MA(2). Si el programa de Box-Jenkins calcula los valores de w como $w_1 = -.9$ y $w_2 = .35$, entonces

$$\begin{aligned} \hat{Y}_t &= w_0 - w_1\epsilon_{t-1} - w_2\epsilon_{t-2} \\ &= 237 - (-.9)(18.4) - (.35)(15.2) = 248.2 \end{aligned}$$

Modelos autorregresivos de promedio móvil

Además de los modelos AR y MA, ambos se pueden combinar en un tercer tipo de modelo general denominado ARIMA. La ecuaciones 10.1 y 10.2 se combinan para formar

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - w_1 \epsilon_{t-1} - w_2 \epsilon_{t-2} - \cdots - w_q \epsilon_{t-q} \quad (10.3)$$

Los modelos ARIMA (p, q) utilizan combinaciones de errores anteriores y valores anteriores y ofrecen un potencial para ajustar modelos que no pudieron ajustarse en forma adecuada mediante los modelos AR y MA por sí solos.

La fig. 10.4 muestra la ecuación de un modelo ARIMA(1, 1) y el comportamiento teórico de los coeficientes de autocorrelación y de autocorrelación parcial.

Una diferencia importante entre la metodología de Box-Jenkins y los métodos previos consiste en que Box-Jenkins no hace suposiciones acerca del número de términos o de los pesos específicos a asignar a los términos. El analista selecciona el modelo apropiado, incluyendo el número de términos; después el programa calcula los coeficientes, utilizando un método no lineal de mínimos cuadrados. Se pueden hacer pronósticos de periodos futuros y se pueden construir intervalos de confianza para estas estimaciones.

APLICACIÓN DE LA METODOLOGÍA

Como se muestra en la fig.10.1, el enfoque de Box-Jenkins comprende tres etapas separadas. Éstas son: identificación del modelo, estimación y prueba del modelo y aplicación del modelo.

Etapas 1: Identificación del modelo

1. El primer paso en la identificación del modelo consiste en determinar si la serie es estacionaria, es decir, si el valor de la media varía a través del tiempo.

Si la serie no es estacionaria, en general se puede convertir a una serie estacionaria mediante el método de diferenciación. El analista especifica el grado de diferenciación y el algoritmo de Box-Jenkins convierte los datos en una serie estacionaria y realiza los cálculos subsecuentes utilizando los datos convertidos.

2. Una vez obtenida una serie estacionaria, el analista debe identificar la forma del modelo a utilizar.

Este paso se logra mediante la comparación de los coeficientes de autocorrelación y de autocorrelación parcial de los datos a ajustar con las correspondientes distribuciones de los

diversos modelos ARIMA. Para ayudar en la selección de un modelo apropiado, en las figs. 10.2, 10.3 y 10.4, se muestran las distribuciones teóricas para los modelos ARIMA más comunes.

Como puede verse, cada modelo tiene un conjunto único de autocorrelaciones y de autocorrelaciones parciales, y el analista debe ser capaz de poder ubicar los coeficientes correspondientes a los datos, a una de las distribuciones teóricas.

Aun cuando en general no será posible hacer coincidir exactamente los datos con las distribuciones teóricas, se pueden efectuar pruebas durante la etapa 2 para determinar si el modelo es adecuado. Entonces, si el modelo no es satisfactorio, se puede intentar un modelo alternativo. Después de un poco de práctica, el analista se hará más apto en la identificación de un modelo adecuado.

En términos generales, el analista debe identificar las autocorrelaciones que caen exponencialmente a cero. Si las autocorrelaciones descienden exponencialmente a cero, el proceso indicado es el AR; si son las autocorrelaciones parciales las que descienden a cero, entonces el proceso indicado es el MA; y, si tanto los coeficientes de autocorrelación como los coeficientes de autocorrelación parcial descienden a cero, el indicado es un proceso mixto ARIMA. El analista puede determinar el orden de los procesos AR y/o MA contando el número de coeficientes de autocorrelación y de autocorrelación parcial que son diferentes de cero en forma significativa.

Etapla 2: Estimación del modelo y prueba de su adecuación

1. Una vez seleccionado un modelo tentativo, se deben estimar los parámetros para ese modelo.

Por ejemplo, suponga que se eligió un modelo ARIMA(1, 1). La fórmula matemática y la fórmula de pronóstico del modelo respectivamente son

$$Y_t = \phi_1 Y_{t-1} + \epsilon_t - w_1 \epsilon_{t-1}$$

y

$$\hat{Y}_t = \phi_1 Y_{t-1} - w_1 \epsilon_{t-1}$$

Para utilizar la ecuación de pronóstico, el analista debe calcular los valores para ϕ_1 y w_1 . Estos cálculos se realizan mediante el programa de cómputo Box-Jenkins, utilizando como criterio de selección de valores óptimos el error medio cuadrado.

Suponga que los valores calculados de ϕ_1 y w_1 fueron .25 y .5. Ahora, el modelo tentativo es

$$\hat{Y}_t = .25 Y_{t-1} - .5 \epsilon_{t-1}$$

2. Antes de usar el modelo para pronosticar, el analista debe verificar si es adecuado.

Este paso se realiza revisando los términos de error, $\epsilon_t = Y_t - \hat{Y}_t$, para asegurarse de que son aleatorios. Esta verificación puede hacerse revisando que las autocorrelaciones de los términos de error para estar seguros de que no son diferentes de cero en forma significativa. Si algunos retrasos de orden menor o estacionales son diferentes de cero de manera significativa, entonces el modelo resulta inadecuado. El analista debe regresar a la etapa 1, paso 2, seleccionar un modelo alternativo y después continuar el análisis.

También se puede revisar lo adecuado del modelo haciendo una prueba de ji cuadrada (χ^2), conocida como la estadística Q de Box–Pierce, sobre las autocorrelaciones de los residuos. La estadística de prueba es

$$Q = (N - d) \sum_{k=1}^m r_k^2 \quad (10.4)$$

La cual está distribuida aproximadamente como una distribución de ji cuadrada con $k - p - q$ grados de libertad. En esta ecuación

N = longitud de la serie histórica

k = primeras k autocorrelaciones que se verifican

m = número máximo de retrasos verificados

r_k = función de autocorrelación de la muestra del k -ésimo término de residuo

d = grado de diferenciación para obtener una serie estacionaria

Si el valor calculado de Q es mayor que la χ^2 para $k - p - q$ grados de libertad, entonces se debe considerar que el modelo es inadecuado. El analista debe regresar a la etapa 1, paso 2, seleccionar otro modelo y continuar el análisis hasta encontrar un modelo satisfactorio.

Ninguna de estas dos pruebas de adecuación debe considerarse como la última palabra, aunque en general deben utilizarse juntas, además de un juicio considerable por parte del analista. Por ejemplo si se pueden explicar algunas desviaciones mayores por circunstancias no usuales, se pueden ignorar estas desviaciones si el resto del modelo es adecuado.

Es posible que se puedan juzgar dos o más modelos como aproximadamente iguales, aunque ninguno de ellos se ajuste exactamente a los datos. En este caso, debe prevalecer el principio de parsimonia y elegir el modelo más sencillo.

Etapas 3: Pronóstico con el modelo

1. Una vez que se encontró un modelo adecuado, se pueden realizar pronósticos para uno o varios periodos a futuro.

También se pueden formular intervalos de confianza sobre estas estimaciones. En general, entre más a futuro se pronostica, mayor será el intervalo de confianza. Estos pronósticos e intervalos de confianza se calculan mediante el programa de Box–Jenkins a solicitud del analista.

2. Al haber más datos disponibles, se puede utilizar el mismo modelo para revisar los pronósticos, seleccionando otro periodo de origen.
3. Si la serie parece cambiar a través del tiempo, pudiera ser necesario recalcular los parámetros, o incluso desarrollar un modelo nuevo por completo.

Si se aprecian pequeñas diferencias en los errores de pronóstico, pudieran indicar que es necesario recalcular los parámetros, y el analista deberá regresar a la etapa 2, paso 1. Cuando se aprecian grandes diferencias en la dimensión de los errores de pronóstico, pudieran indicar que se requiere un modelo completamente nuevo, y el analista deberá regresar a la etapa 1, paso 2, o inclusive a la etapa 1, paso 1 y repetir el proceso de ajustar un nuevo modelo a la serie histórica.

La siguiente notación se utiliza con frecuencia en las técnicas de Box-Jenkins. Se identifica un modelo como $ARIMA(p, d, q)$, en donde p es el orden del término autorregresivo, d es el nivel de diferenciación y q es el orden del término del promedio móvil. En la práctica, cuando no hay diferenciación, el modelo apropiado es $ARIMA(p, q)$.

Ejemplo 10.3

La Cameron Consulting Corporation se especializa en portafolios de servicios de inversión. A Lynn Stephens, el analista de la compañía, se le asignó la tarea de desarrollar técnicas de pronóstico más complejas para pronosticar los promedios del Dow-Jones. Lynn asistió recientemente a un taller sobre la metodología de Box-Jenkins y decidió probar la técnica sobre el Índice de Transporte. La Tabla 10.2 representa los 55 promedios de cierre diario del Índice de Transporte durante los meses de verano.

TABLA 10.2 PROMEDIOS DIARIOS AL CIERRE DEL ÍNDICE DE TRANSPORTE

PERIODO	PROMEDIO AL CIERRE	PERIODO	PROMEDIO AL CIERRE	PERIODO	PROMEDIO AL CIERRE	PERIODO	PROMEDIO AL CIERRE
1	222.34	15	223.07	29	246.74	43	249.76
2	222.24	16	225.36	30	248.73	44	251.66
3	221.17	17	227.60	31	248.83	45	253.41
4	218.88	18	226.82	32	248.78	46	252.04
5	220.05	19	229.69	33	249.61	47	248.78
6	219.61	20	229.30	34	249.90	48	247.76
7	216.40	21	228.96	35	246.45	49	249.27
8	217.33	22	229.99	36	247.57	50	247.95
9	219.69	23	233.05	37	247.76	51	251.41
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		

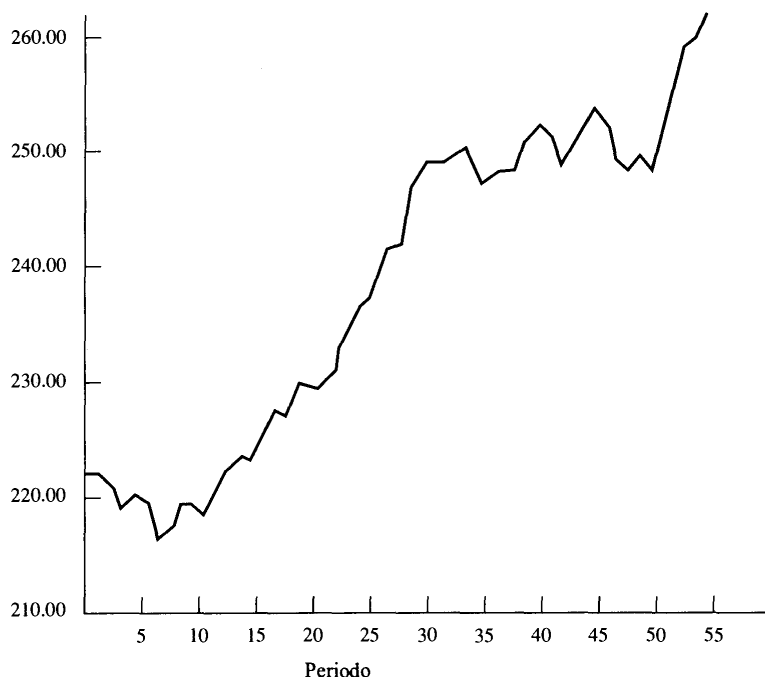


Figura 10.5 Promedios diarios al cierre del índice de transporte.

Lynn comienza el análisis observando la gráfica de los datos que se muestra en la fig. 10.5. Parecía haber una tendencia en los datos. Su primer paso en la identificación de un modelo tentativo fue observar los coeficientes de autocorrelación de los datos, que se presentan en la fig. 10.6. Al ver que las 12 primeras autocorrelaciones parecían descender a cero, reconoció que su apreciación original era correcta y que estos datos tenían una tendencia y deberían diferenciarse. Lynn tomó en cuenta la tendencia ascendente en los datos mediante la aplicación de una diferenciación no estacional.

La gráfica de las autocorrelaciones de los datos diferenciados que se presenta en la fig. 10.7 mostró que los datos eran estacionarios. Sin embargo, no parecía existir ningún patrón aparente. Una revisión de las autocorrelaciones parciales confirma la ausencia de un patrón. Es evidente que sólo existe una tendencia en los datos del Índice de Transporte. Se puede emplear la metodología de Box-Jenkins para modelar los datos; sin embargo, otros métodos serían menos costosos y posiblemente más precisos.

La estimación de la constante general es .725. Un examen de la función de autocorrelación de los residuos mostró que el modelo era adecuado. Lynn confirmó esta observación mediante la prueba de estadística ji cuadrada. El valor calculado de ji cuadrada es menor que el valor de la tabla para 17 grados de libertad a un nivel de significación de .05 ($17.06 < 27.6$).

Normalmente, las Y se expresan como desviaciones de su media ($Y' = Y - \bar{Y}$); no obstante, cuando se usa un término constante, se modelan los valores reales de Y .

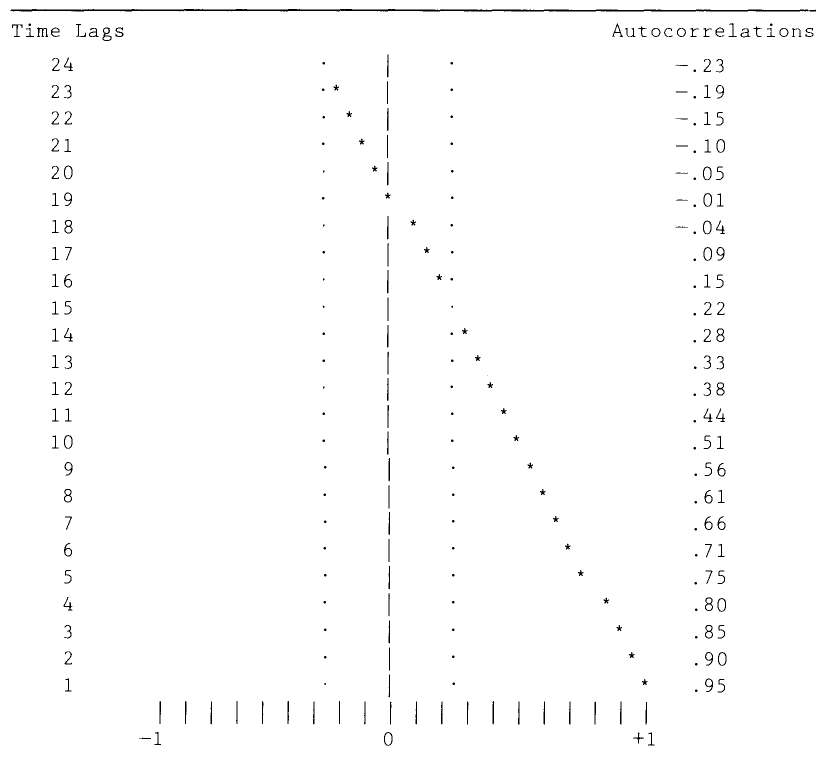


Figura 10.6 Autocorrelación de los datos sin procesar de la tabla 10.2.

El modelo de pronóstico con un término constante es

$$\hat{Y}_t = .725 + Y_{t-1}$$

Los pronósticos para los periodos 56 y 57 son

$$\hat{y}_{56} = .725 + 261.49$$

$$\hat{y}_{56} = 262.22$$

y

$$\hat{y}_{57} = .725 + 262.22$$

$$\hat{y}_{57} = 262.94$$

Nótese que el pronóstico se incrementa en .725 en cada periodo.

Lynn decidió que la autocorrelación marginalmente alta (.19) en el retraso 1 de la fig. 10.7 indicaba la necesidad de intentar un modelo ARIMA(0, 1, 1). Observe que en la metodología de Box-Jenkins siempre se ejecuta primero el modelo más sencillo. De nuevo, la revisión de la función de autocorrelación de residuos mostró que el modelo era adecuado. Lynn probó de nuevo la estadística ji cuadrada. El valor calculado de ji cuadrada es menor

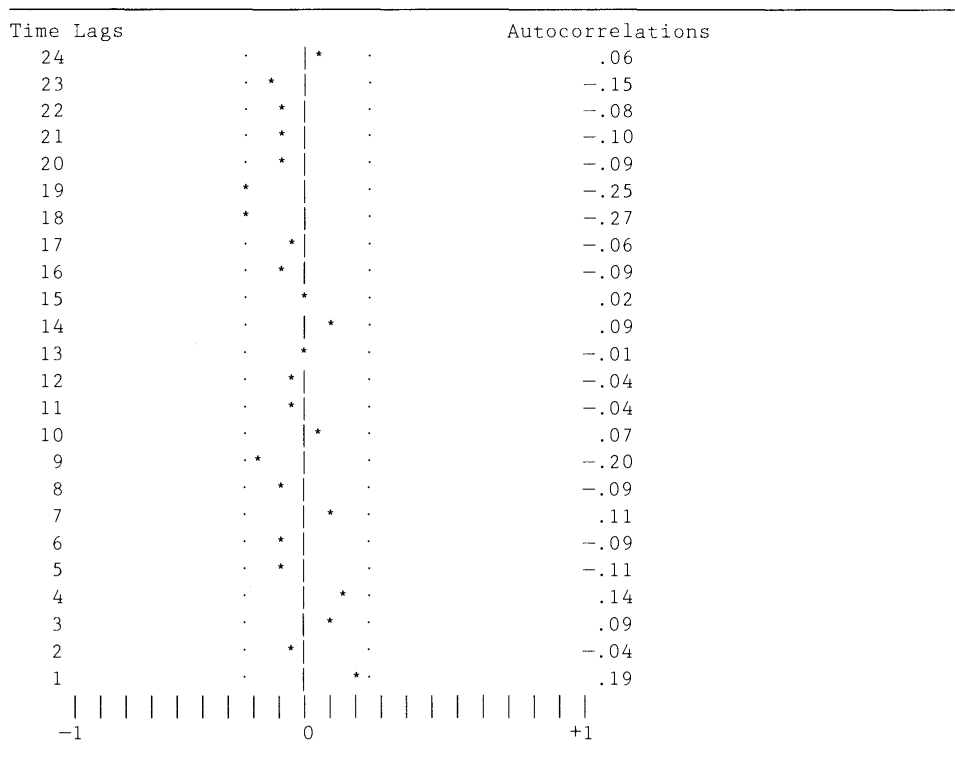


Figura 10.7 Autocorrelaciones de los datos diferenciados no estacionales de la tabla 10.2.

que el valor de la tabla para 15 grados de libertad a un nivel de significación de .05 (13.8 < 25.0).

Los comandos de Minitab para resolver este ejemplo son

```
MTB > READ 'TRANS.DAT' C1
      55 ROWS READ

C1
  222.34  222.24  221.17  218.88  .  .  .

MTB > ACF C1

ACF OF C1

      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +---+---+---+---+---+---+---+---+---+---+
1  0.951
2  0.900
3  0.847
```

4	0.795	XXXXXXXXXXXXXXXXXXXXX
5	0.748	XXXXXXXXXXXXXXXXXXXXX
6	0.707	XXXXXXXXXXXXXXXXXXXXX
7	0.658	XXXXXXXXXXXXXXXXXXXXX
8	0.609	XXXXXXXXXXXXXXXXXXXXX
9	0.561	XXXXXXXXXXXXXXXXXXXXX
10	0.505	XXXXXXXXXXXXXXXXXXXXX
11	0.440	XXXXXXXXXXXXXXXXXXXXX
12	0.381	XXXXXXXXXXXXXXXXXXXXX
13	0.328	XXXXXXXXXXXXX
14	0.277	XXXXXXXXXX
15	0.215	XXXXXXX
16	0.152	XXXXX
17	0.093	XXX

El correlograma muestra que los datos son no estacionarios (con tendencia) y requieren ser diferenciados.

```
MTB > DIFFERENCE 1 C1, STORE IN C2
```

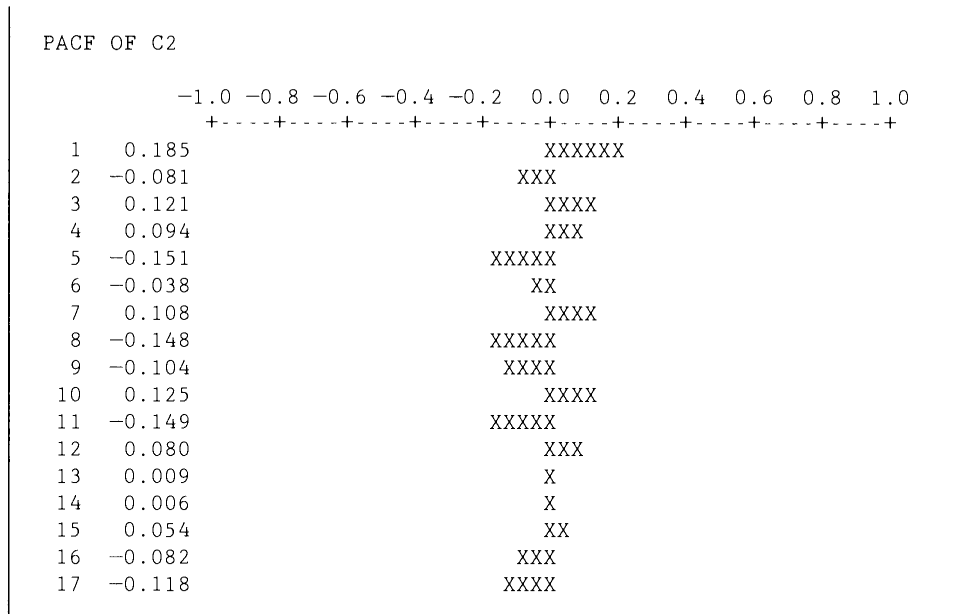
A continuación, se calculan y exhiben los coeficientes de autocorrelación y de autocorrelación parcial.

```
MTB > ACF C2

ACF OF C2
```

	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
	+---+---+---+---+---+---+---+---+---+---+										
1	0.185					XXXXXX					
2	-0.044					XX					
3	0.093					XXX					
4	0.136					XXXX					
5	-0.110					XXXX					
6	-0.095					XXX					
7	0.110					XXXX					
8	-0.088					XXX					
9	-0.196					XXXXXX					
10	0.069					XXX					
11	-0.042					XX					
12	-0.044					XX					
13	-0.009					X					
14	0.095					XXX					
15	0.025					XX					
16	-0.086					XXX					
17	-0.056					XX					

```
MTB > PACF C2
```



- El comando ACF se usa para calcular los coeficientes de autocorrelación para los datos de transporte en C1.
- El comando DIFFERENCE se emplea para tomar la primera diferenciación de la columna C1 y almacenarla en C2.
- El comando ACF se utiliza para calcular los coeficientes de autocorrelación de los datos de la primera diferenciación en C2.
- El comando PACF se usa para calcular los coeficientes de autocorrelación parcial para los datos de la primera diferenciación en C2.

A continuación se ejecuta el modelo ARIMA apropiado. El análisis previo indicó que los datos de la primera diferenciación seguían un camino aleatorio ligeramente ascendente. Los comandos de Minitab correspondientes son

```
MTB > ARIMA(0 1 0) C1;
SUBC> CONSTANT;
SUBC> FORECAST 3 DAYS AHEAD STORE IN C3, & CONF LIMITS
CONT> IN C4 C5.
```

```
Estimates at each iteration
Iteration      SSE      Parameters
0             191.847    0.825
1             191.308    0.730
2             191.307    0.725
3             191.307    0.725
```

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
Constant	0.7248	0.2581	2.81

Differencing: 1 regular difference

No. of obs.: Original series 55, after differencing 54

Residuals: SS = 191.307 (backforecasts excluded)

MS	DF
3.610	53

Modified Box-Pierce chisquare statistic

Lag	12	24	36	48
Chisquare	9.5 (DF=12)	27.0 (DF=24)	33.7 (DF=36)	49.8 (DF=48)

Forecasts from period 55

Period	Forecast	95 Percent Limits		Actual
		Lower	Upper	
56	262.215	258.490	265.939	
57	262.940	257.672	268.207	
58	263.664	257.213	270.115	

- El comando ARIMA(0 1 0)C1 se usa para ejecutar un modelo sin el término AR o MA para los datos de la primera diferenciación. El comando ARIMA(0 0 0)C2 produciría el mismo resultado.
- El subcomando CONSTANT se emplea para ejecutar el modelo con un término constante.
- El subcomando FORECAST se utiliza para almacenar tres pronósticos en C3 y colocar límites de confianza en C4 y C5.
- La salida muestra que la estimación final del parámetro constante es .7248, la proporción t para este parámetro es 2.81 y el EMC (error medio cuadrado) es igual a 3.61.

A continuación, se ejecuta el modelo de promedio móvil de primer orden. Los comandos de Minitab son

```
MTB > ARIMA(0 1 1)C1;
SUBC> CONSTANT.
```

Estimates at each iteration

Iteration	SSE	Parameters	
0	200.915	0.100	0.825
1	188.362	-0.050	0.787
2	183.584	-0.190	0.741

3	183.383	-0.219	0.731
4	183.377	-0.224	0.730
5	183.377	-0.225	0.729
6	183.377	-0.226	0.730

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
MA 1	-0.2256	0.1366	-1.65
Constant	0.7296	0.3132	2.33

Differencing: 1 regular difference

No. of obs.: Original series 55, after differencing 54

Residuals: SS = 183.359 (backforecasts excluded)

MS = 3.526 DF = 52

Modified Box-Pierce chisquare statistic

Lag	12	24	36	48
Chisquare	9.6(DF=11)	23.7(DF=23)	29.6(DF=35)	42.4(DF=47)

Note que la proporción t para el término MA(1) no es significativa al probarla al nivel de significación de 0.1. Por esta razón y también porque es más sencillo, se considera mejor el modelo con términos autorregresivos y promedio móvil de orden cero.

Ejemplo 10.4

Jim White, analista de Atron Corporation, tenía una serie histórica de lecturas para un proceso que requería de un pronóstico. Los datos se muestran en la tabla 10.3 y se grafican en la fig. 10.8. Jim pensó que la metodología de Box-Jenkins funcionaría mejor para estos datos.

Jim comenzó el primer paso en la identificación de un modelo tentativo observando las autocorrelaciones de los datos que se muestran en la fig. 10.9, para determinar si la serie era estacionaria. Las autocorrelaciones para los retrasos inferiores caen hacia cero con rapidez, de modo que Jim concluyó que la serie era estacionaria.

TABLA 10.3 LECTURAS DE UN PROCESO DE ATRON CORPORATION

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	

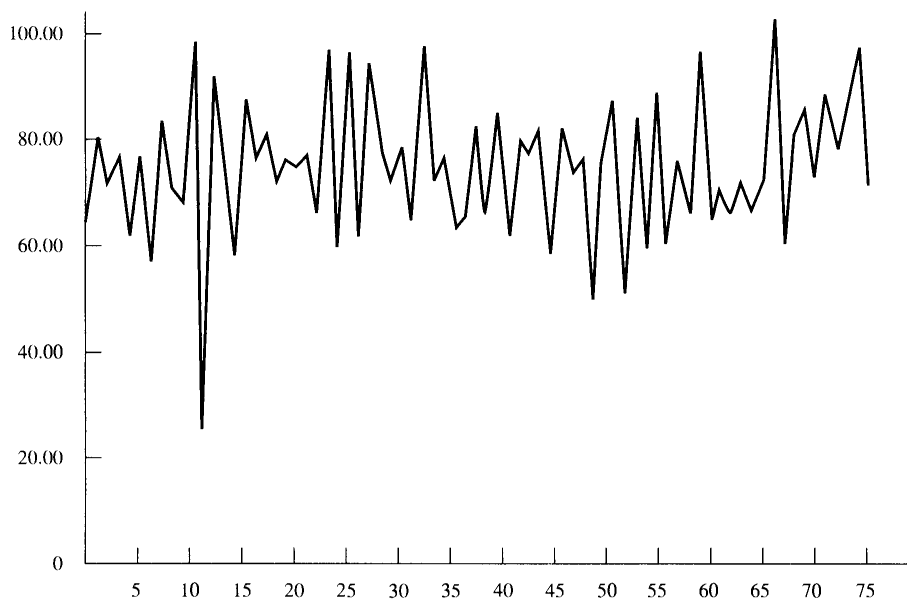


Figura 10.8 Lecturas de un proceso de Atron Corporation.

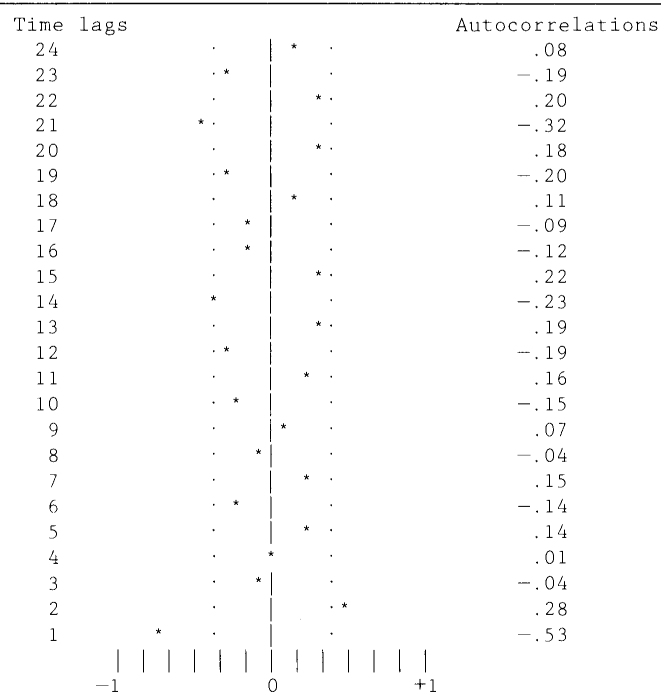


Figura 10.9 Autocorrelaciones de los datos sin procesar de la tabla 10.3.

Los dos primeros coeficientes de autocorrelación de la muestra ($-.53$ y $.28$) son significativos a un nivel del 95% ya que se ubican fuera del rango.

$$0 \pm 1.96 \left(\frac{1}{\sqrt{75}} \right)$$

$$0 \pm 1.96(.115)$$

$$0 \pm .225$$

El coeficiente de autocorrelación de la muestra en el retraso 21 ($-.32$) también es significativo, aunque este resultado se debe probablemente a un desvío que se presenta en los retrasos los superiores.

En los coeficientes de autocorrelación parcial que aparecen en la fig. 10.10, Jim observó que el primero era significativo, aunque ninguna de las demás autocorrelaciones parciales se aproximó a ser significativa.

Los dos modelos son sugeridos mediante esta revisión de los coeficientes de autocorrelación y de autocorrelación parcial. El parcial en el retraso 2 es cero (véase fig. 10.10). El significado de este parcial es la clave de cuál de los modelos funcionará mejor. Como los parciales en los retrasos 3, 4 y 5 no son cercanos a cero, Jim piensa que el coeficiente de autocorrelación parcial de cero en el retraso 2 ocurrió debido a un error de muestreo casual. Si de hecho los parciales descienden a cero, entonces se sugiere el modelo MA(2) mediante los dos coeficientes significativos de autocorrelación en los retrasos 1 y 2 de la fig. 10.9. Sin

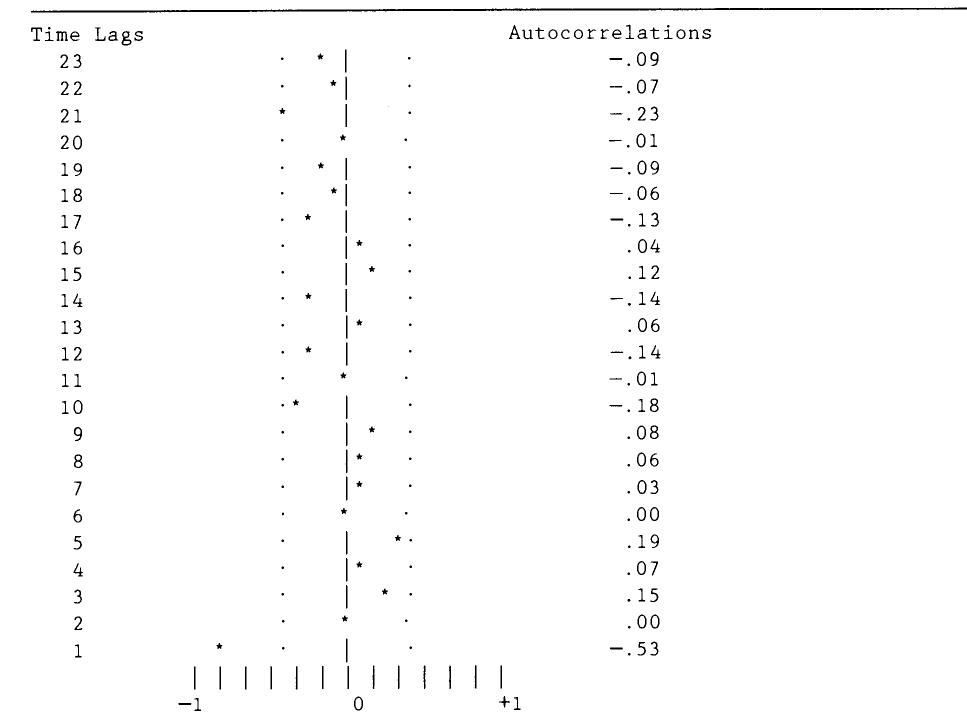


Figura 10.10 Autocorrelaciones parciales de los datos sin procesar de la tabla 10.3.

Tabla 10.4 MODELOS MA(2) Y AR(1) PARA EL EJEMPLO 10.4

MA(2) Model

Estimates at each iteration

Iteration	SSE	Parameters		
7	9786.2	0.567	-0.356	75.410

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
MA 1	0.5667	0.1107	5.12
MA 2	-0.3560	0.1146	-3.11
Constant	75.410	1.061	71.08
Mean	75.410	1.061	

No. of obs.: 75

Residuals: SS = 9724.97 (backforecasts excluded)
MS = 135.07 DF = 72

Modified Box-Pierce chisquare statistic

Lag	12	24	36	48
Chisquare	7.0(DF=10)	23.8(DF=22)	31.8(DF=34)	46.9(DF=46)

Forecasts from period 75

Period	Forecast	95 Percent 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	

AR(1) Model

Estimates at each iteration

Iteration	SSE	Parameters	
7	10113.5	-0.538	115.842

Relative change in each estimate less than 0.0010

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
AR 1	-0.5379	0.0986	-5.46
Constant	115.842	1.356	85.44
Mean	75.3267	0.8816	

No. of obs.: 75

Residuals: SS = 10065.2 (backforecasts excluded)
MS = 137.9 DF = 72

Modified Box-Pierce chisquare statistic

Lag	12	24	36	48
Chisquare	9.3(DF=11)	29.9(DF=23)	37.3(DF=35)	58.3(DF=47)

Forecasts from period 75

Period	Forecast	95 Percent Limits	
		Lower	Upper
76	77.116	54.097	100.135
77	74.364	48.226	100.502
78	75.844	48.871	102.817

embargo, si las autocorrelaciones parciales realmente cayeron a cero en el retraso 2, entonces pudiera ser apropiado un modelo AR(1).

Se usará Minitab para analizar los datos mediante ambos modelos. Observe que se incluye un término constante en los modelos debido a que los valores de la variable dependiente no se han expresado como desviaciones de su media. Si los datos hubieran sido diferenciados, no se usaría el término constante. Los comandos de Minitab se muestran en la sección del Paquete de cómputo minitab, al final del capítulo.

El Cuadro 10.4 muestra los resultados de la ejecución de los modelos MA(2) y AR(1). Ambos modelos parecen ser adecuados. Las proporciones t para ambos parámetros del modelo MA(2) son significativos al nivel de .01. El parámetro para el modelo AR(1) es significativo a un nivel de significación de .01. La estadística modificada de Box-Pierce para χ^2 cuadrada es también significativa para todos los retrasos dados. Esta prueba muestra que los residuos no son diferentes de cero en forma significativa. El EMC para ambos modelos es similar. Por lo tanto, es probable que el modelo AR(1) sea la mejor elección, pues sólo tiene un parámetro, comparado con los dos que utiliza el MA(2), y es el más sencillo de los dos modelos. Los parámetros estimados para el modelo son 115.84 para la constante y $-.5379$ para el parámetro autorregresivo.

Los coeficientes de autocorrelación de los residuos se muestran en la fig.10.11. Sólo un coeficiente es significativo y fácilmente podría haber ocurrido por casualidad. Jim confirma

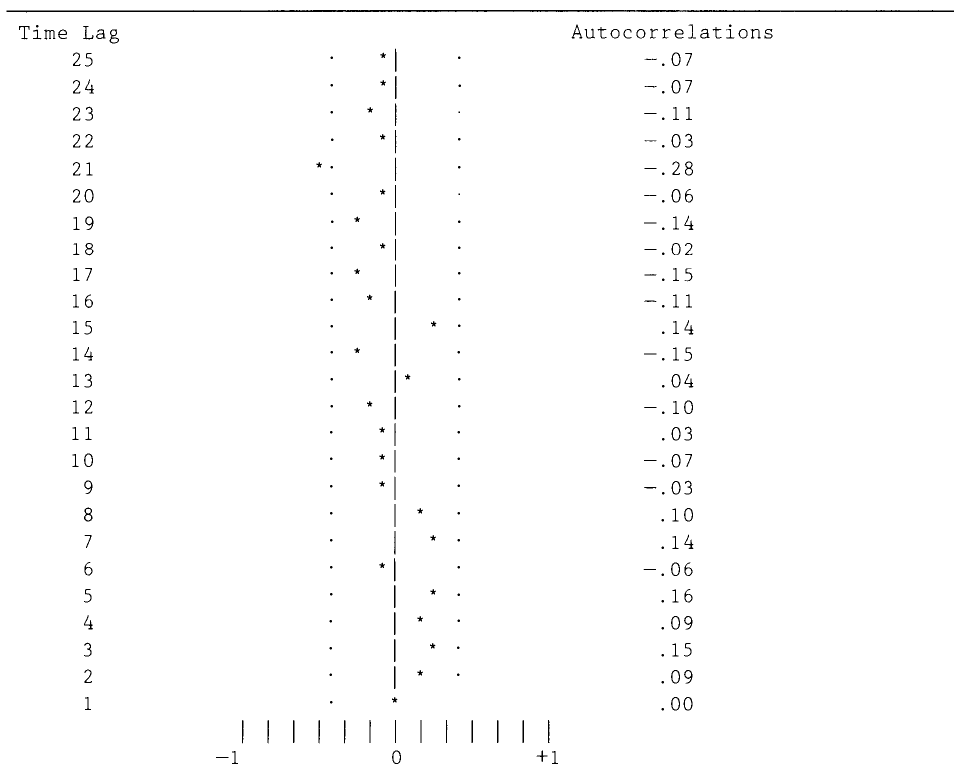


Figura 10.11 Coeficientes de autocorrelación de los residuos del modelo AR(1) ajustado con una constante.

esta observación probando la estadística ji cuadrada para 25 periodos de retraso. El valor calculado de ji cuadrada es menor que el valor de la tabla para 23 grados de libertad a un nivel de significación de .05 ($30.31 < 35.2$). El modelo resulta adecuado.

El modelo de pronóstico AR(1) con un término constante es

$$\hat{Y}_t = 115.84 + (-.5379)Y_{t-1}$$

El pronóstico para el periodo 76 es

$$\hat{y}_{76} = 115.84 + (-.5379)(72)$$

$$\hat{y}_{76} = 77.11$$

El pronóstico para el periodo 77 es

$$\hat{y}_{77} = 115.84 + (-.5379)(77.11)$$

$$\hat{y}_{77} = 74.36$$

Ejemplo 10.5

Jim White quedó complacido con los resultados de su pronóstico para la serie histórica de lecturas del proceso mostrado en la tabla 10.3. Decidió utilizar la metodología Box-Jenkins para intentar pronosticar los errores en un proceso de control de calidad a su cargo. Los datos se presentan en la tabla 10.5 y se grafican en la fig. 10.12.

Jim comenzó el primer paso con la identificación de un modelo tentativo observando las autocorrelaciones de los datos que aparecen en la fig. 10.13, para determinar si la serie es

TABLA 10.5 ERRORES DEL CONTROL DE CALIDAD DE ATRON

-.23	-.2	-1.93	-.97	.1
.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	.7	.8	1.34
-1.28	-.04	.69	-1.95	-1.83
0	.9	-.24	2.61	.31
-.63	1.79	.34	.59	1.13
.08	-.37	.6	.71	-.87
-1.3	.4	.15	-.84	1.45
1.48	-1.19	-.02	-.11	-1.95
-.28	.98	.46	1.27	-.51
-.79	-1.51	-.54	-.8	-.41
1.86	.9	.89	-.76	.49
.07	-1.56	1.07	1.58	1.54
.09	2.18	.2	-.38	-.96

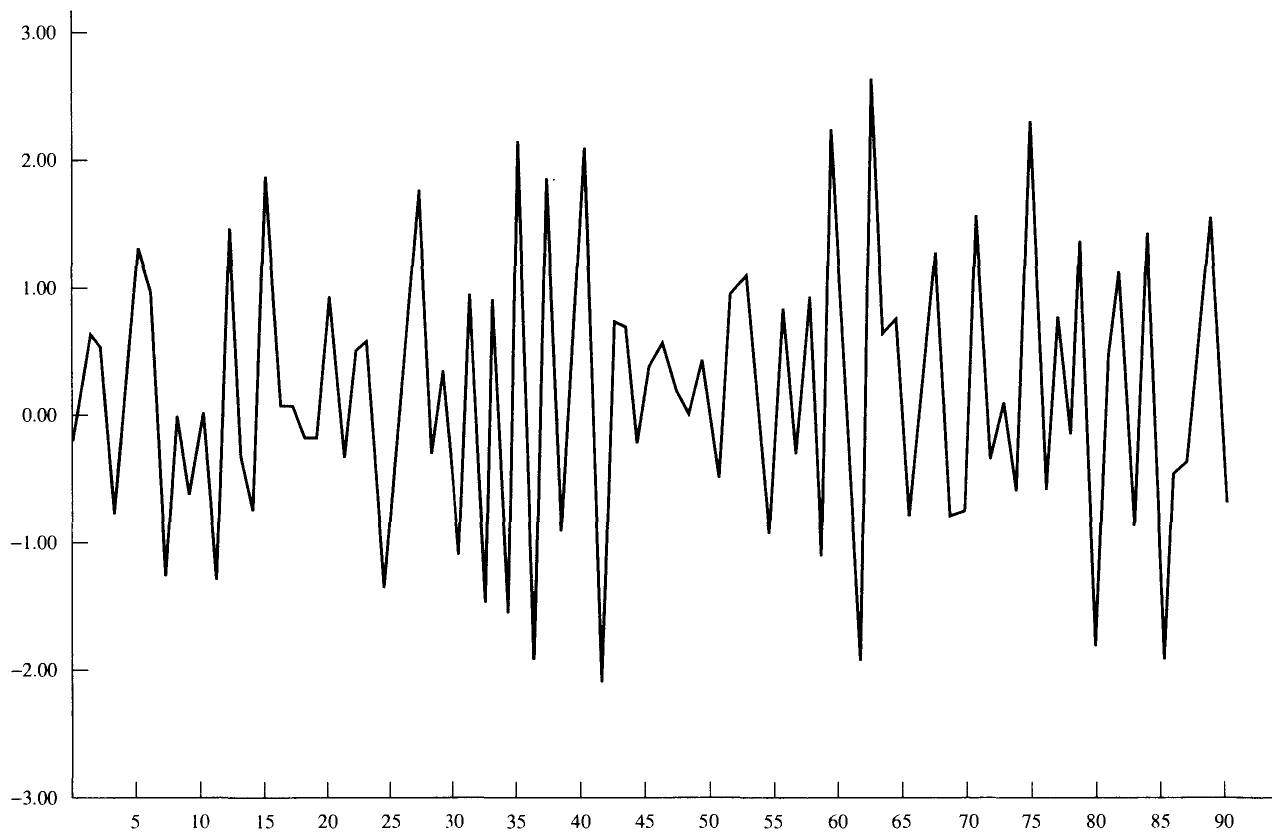


Figura 10.12 Errores del control de calidad de Atron.

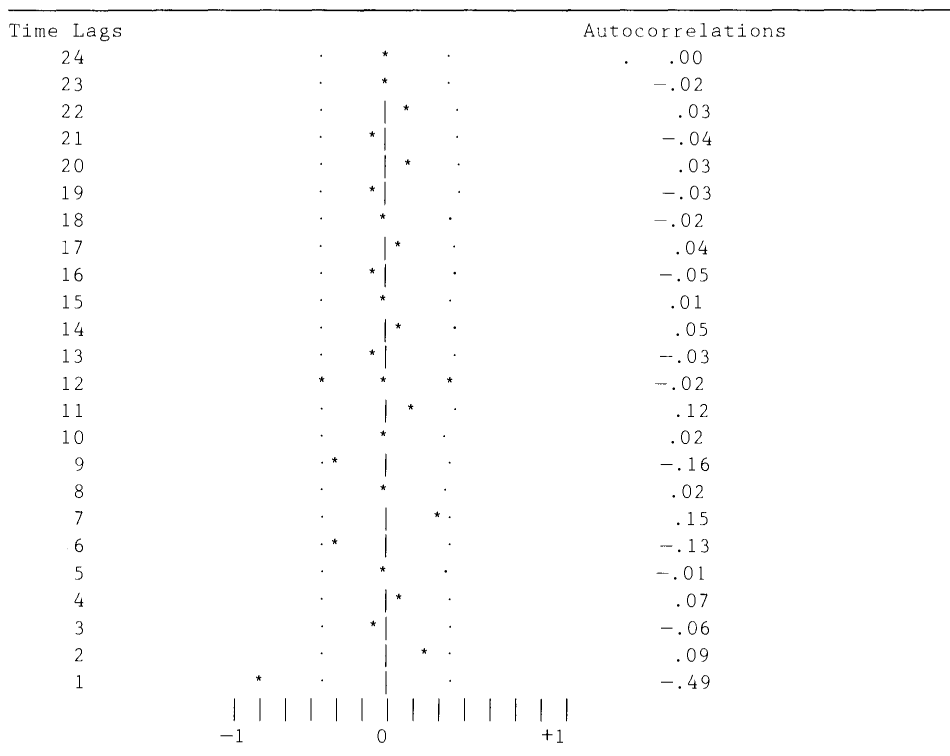


Figura 10.13 Autocorrelaciones de los datos, sin procesar, de la tabla 10.5.

estacionaria. Sólo uno de los coeficientes de autocorrelación es diferente de cero en forma significativa a un nivel de significación del 95%, así que Jim concluye que la serie es estacionaria.

La fig. 10.13 muestra que sólo el primer coeficiente de autocorrelación de la muestra (-.49), es diferente de cero al nivel del 95%. Como los coeficientes de autocorrelación se recortan después del periodo de retraso 1, se sugiere un modelo MA(1).

A partir de la muestra de coeficientes de autocorrelación parcial de la fig. 10.14, Jim observó que la primera y segunda autocorrelaciones parciales eran significativas. Este patrón semeja al de un proceso en el que los parciales teóricos descienden hacia cero, sugiriendo de nuevo un modelo MA(1). A continuación, Jim usó Minitab para estimar los parámetros del modelo.

```
MTB > ARIMA(0 0 1)C1,C5;
SUBC> CONSTANT;
SUBC> FORECAST 3 C2, C3 C4.
```

Estimates at each iteration

Iteration	SSE	Parameters
6	74.3130	0.588 0.153

Relative change in each estimate less than 0.0010

Final Estimates of Parameters				
Type	Estimate	St. Dev.	t-ratio	
MA 1	0.5880	0.0862	6.82	
Constant	0.15256	0.04012	3.80	
Mean	0.15256	0.04012		
No. of obs.: 90				
Residuals: SS = 74.3056 (backforecasts excluded)				
MS = 0.8444 DF = 88				
Modified Box-Pierce chisquare statistic				
Lag	12	24	36	48
Chisquare	9.1(DF=11)	10.9(DF=23)	17.2(DF=35)	31.3(DF=47)
Forecasts from period 90				
		95 Percent Limits		
Period	Forecast	Lower	Upper	Actual
91	0.27810	-1.52331	2.07952	
92	0.15256	-1.93716	2.24227	
93	0.15256	-1.93716	2.24227	

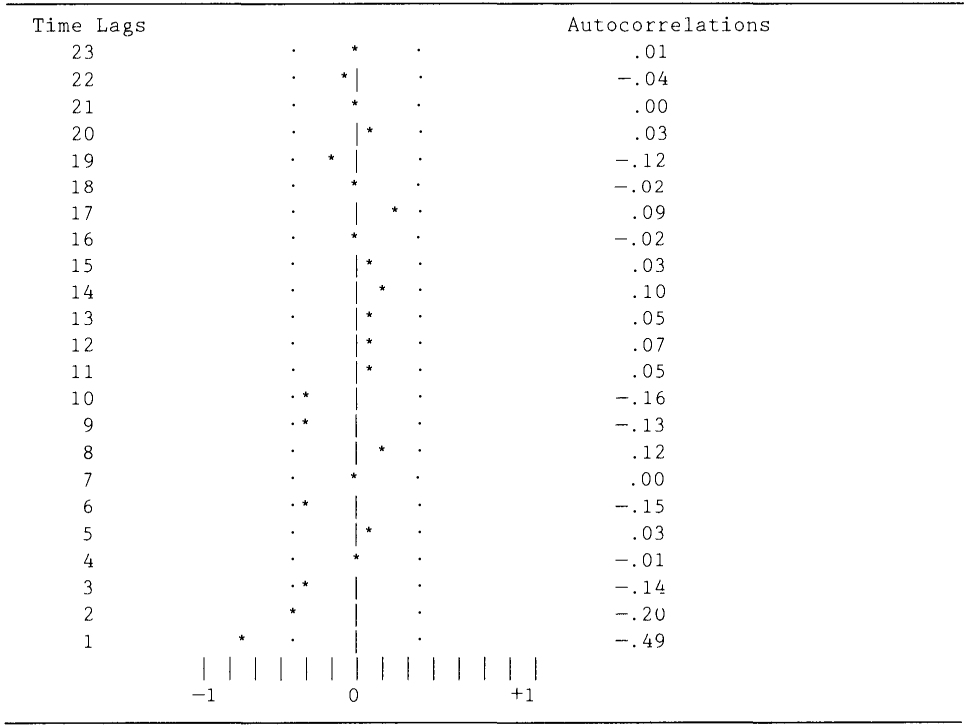


Figura 10.14 Autocorrelaciones parciales de los datos, sin procesar, de la tabla 10.5.

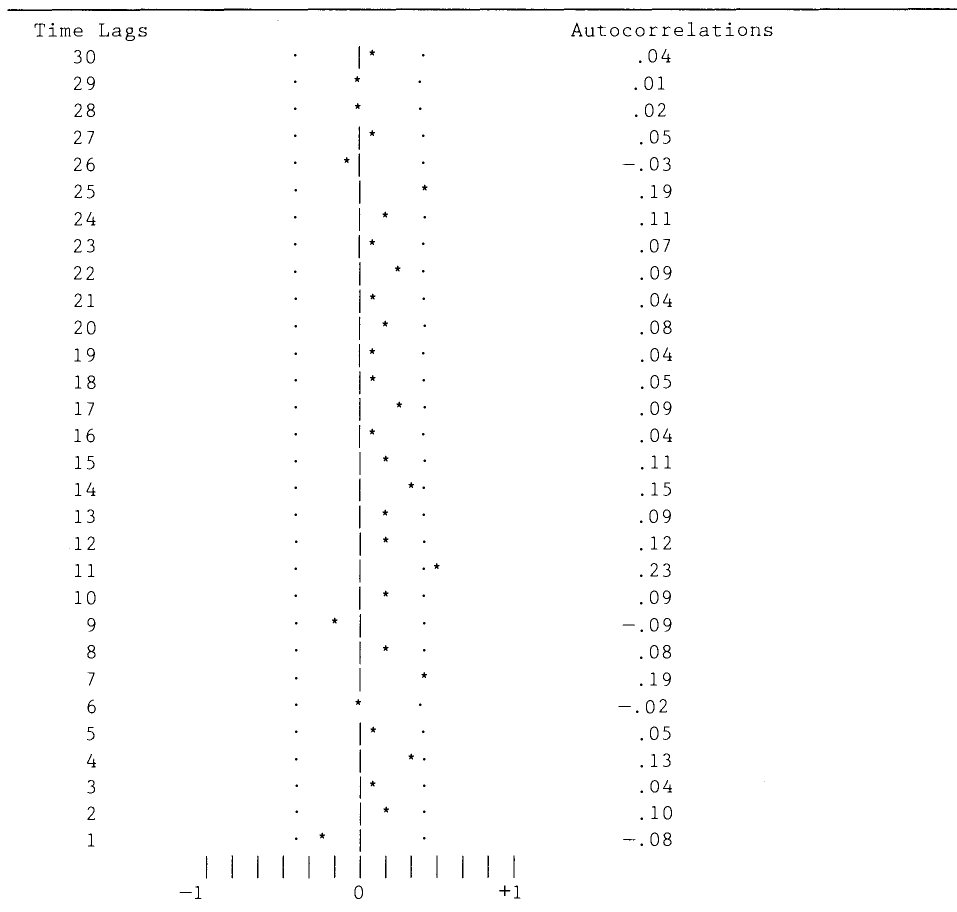


Figura 10.15 Coeficientes de autocorrelación de los residuos del modelo MA(1).

- El comando ARIMA(0 0 1)C1, C5 se usa para ejecutar un modelo MA(1) para los datos en C1 y almacena los resultados en C5.

Se estima que el parámetro para el modelo MA(1) es .588 y la constante es .15256. Las proporciones t son significativas para ambos. Los coeficientes de autocorrelación de los residuos para este modelo se almacenaron en la columna 5. Una revisión de los coeficientes de autocorrelación de estos residuos se muestra en la fig.10.15 e indica que son aleatorios. Esta observación se confirmó al probar la estadística modificada de Box-Pierce para ji cuadrada, para varios periodos.

El modelo de pronóstico MA(1) es

$$\hat{Y}_t = .15256 - (.588)e_{t-1}$$

$$\hat{y}_{91} = .15256 - (.588)(-.21353) = .27816$$

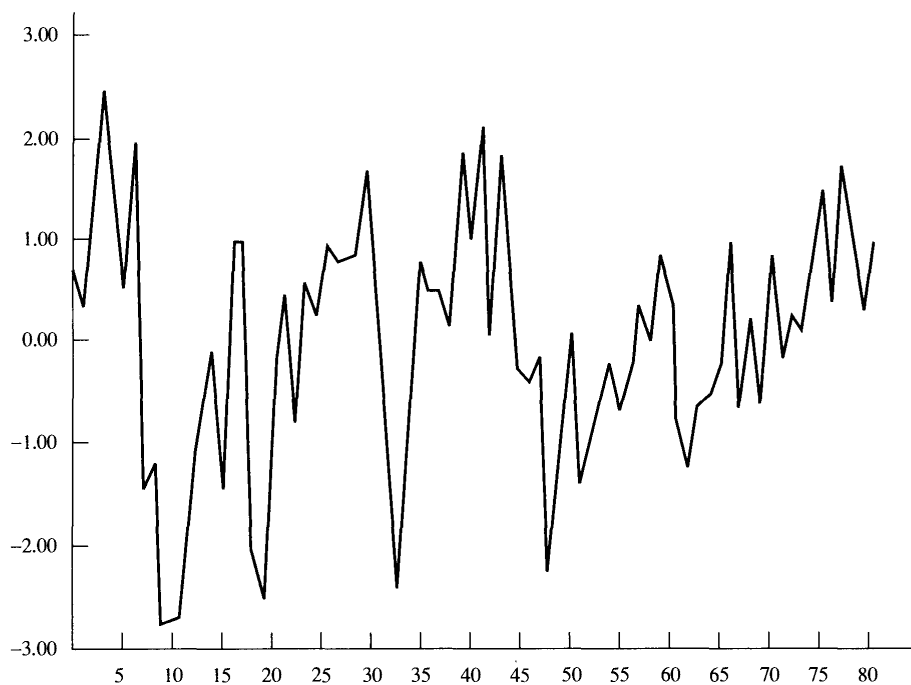
Observe que el pronóstico para el periodo 92 es la constante .15256 ya que se desconoce el error para el periodo 91.

**Ejemplo
10.6**

A Jim White, analista de Atron Corporation, le agradó el procedimiento de pronóstico de Box-Jenkins para los errores en el proceso de control de calidad expuesto en el ejemplo 10.5. En una conferencia, Jim se encontró a su viejo amigo Ed Jones y le platicó el éxito obtenido. Ed tenía una aplicación similar y decidió intentar el método de Box-Jenkins. Los datos se muestran en la tabla 10.6. La fig. 10.16 presenta una gráfica de los datos.

TABLA 10.6 ERRORES DEL CONTROL DE CALIDAD DE ED JONES

.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

**Figura 10.16** Errores del control de calidad de Ed Jones.

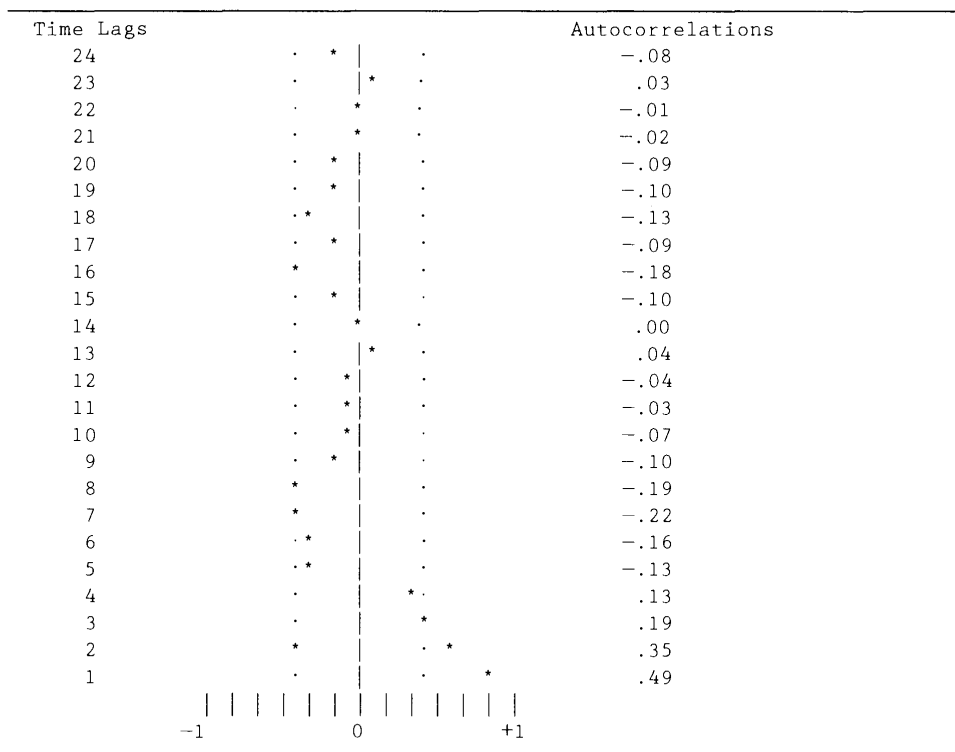


Figura 10.17 Autocorrelaciones de los datos en bruto de la tabla 10.6.

Ed comenzó el primer paso con la identificación de un modelo tentativo observando las autocorrelaciones de los datos que aparecen en la fig. 10.17, para determinar si la serie es estacionaria. Las autocorrelaciones de los retrasos inferiores descienden hacia cero después del tercer retraso, de modo que Ed concluye que la serie histórica es estacionaria.

Los dos primeros coeficientes de autocorrelación de la muestra (.49 y .35) son significativos a un nivel de 95%, ya que se ubican fuera del límite.

$$0 \pm 1.96 \left(\frac{1}{\sqrt{80}} \right)$$

$$0 \pm 1.96(.112)$$

$$0 \pm .219$$

El coeficiente de autocorrelación de la muestra en el periodo de retraso 7 (-.22) también es significativo, aunque este resultado es probable que se deba a la casualidad. Los coeficientes de autocorrelación descienden hacia cero, sugiriendo un modelo AR.

A partir de los coeficientes de autocorrelación parciales de la muestra que se presentan en la fig. 10.18, Ed observó que el primero era significativo. Ya que las autocorrelaciones

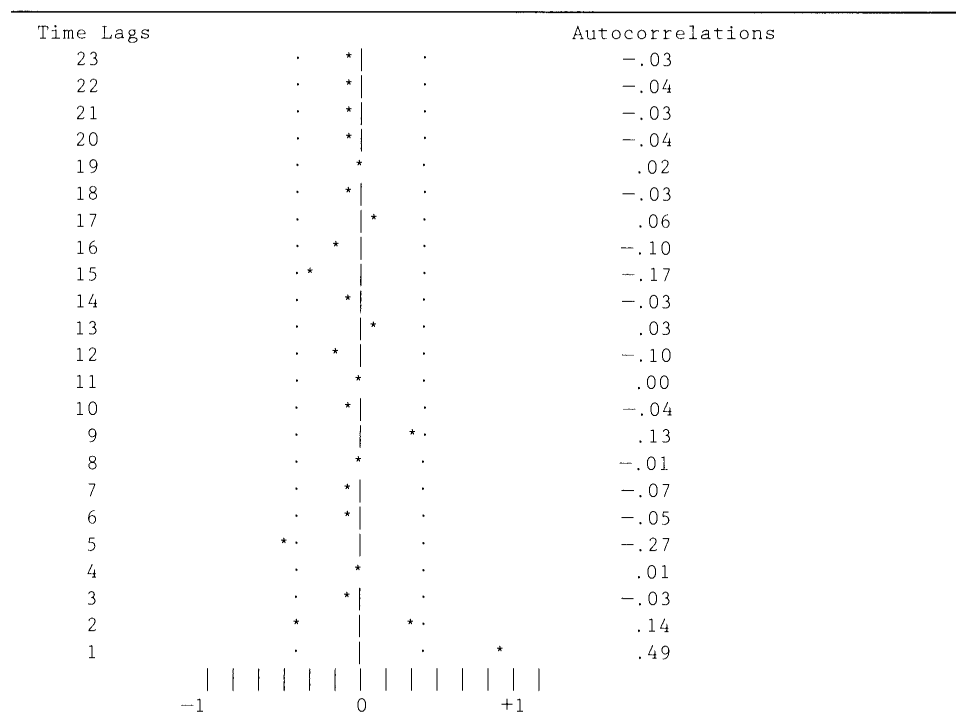


Figura 10.18 Autocorrelaciones parciales de los datos en bruto de la tabla 10.6.

parciales caen hacia cero después del primer periodo de retraso y como se puede considerar que los coeficientes de autocorrelación descienden exponencialmente hacia cero, el proceso indicado es del tipo AR(1). No obstante, la autocorrelación parcial del segundo retraso es .14, lo que podría indicar que los parciales descienden en forma exponencial hacia cero. Como existe esta posibilidad, se podría también probar un modelo ARIMA(1, 0, 1).

A continuación, Ed utiliza un programa de cómputo de Box-Jenkins para estimar el(los) parámetro(s) del modelo. Debido a que los valores de la variable independiente no han sido expresados como desviaciones de su media, normalmente se incluye un término constante en el modelo. Sin embargo, como la media de Y es en extremo pequeña (casi cero) comparada con la desviación estándar, en este modelo no se incluye un término constante. Se estima que el parámetro para el modelo AR(1) es .50084. Una revisión de los coeficientes de autocorrelación de los residuos de la fig. 10.19 indica que son aleatorios, con una suma de los residuos cuadrados igual a 1.1. Ed confirma esta observación probando la estadística de ji cuadrada para 26 periodos retrasados. El valor calculado de ji cuadrada es menor que el valor de la tabla para 25 grados de libertad ($k - p - q = 26 - 1 - 0 = 25$) a un nivel de significación de .05 ($17.43 < 37.6$). El modelo resulta adecuado.

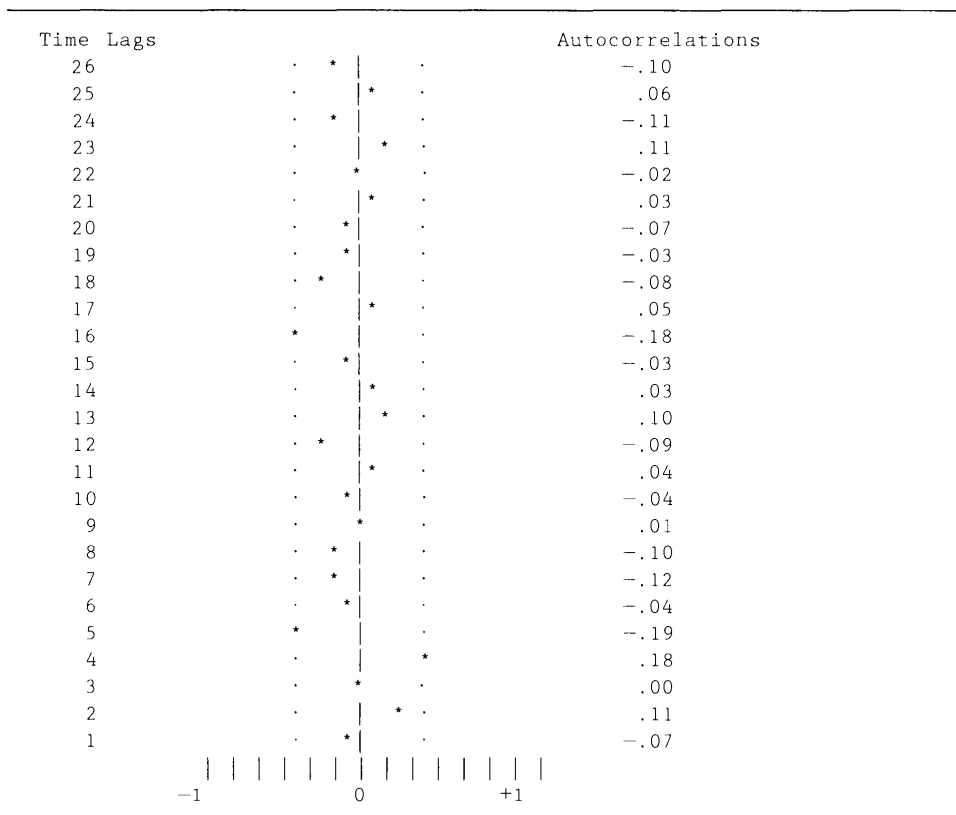


Figura 10.19 Coeficientes de autocorrelación de los residuos del modelo AR(1).

El modelo de pronóstico ARIMA(1 0 1) es

$$\hat{Y}_t = .50084Y_{t-1}$$

Los pronósticos para los periodos 81 y 82 son

$$\begin{aligned} \hat{y}_{81} &= (.50084)(1.06) & \text{y} & \hat{y}_{82} = (.50084)(.53) \\ \hat{y}_{81} &= .53 & & \hat{y}_{82} = .265 \end{aligned}$$

Ejemplo 10.7

Jill Blake, analista de la ISC Corporation, necesita encontrar un método para pronosticar la oferta al cierre de las acciones de ISC. Los datos se muestran en la tabla 10.7. Jill considera que la metodología de Box-Jenkins es la más apropiada para estos datos. La fig. 10.20 muestra una gráfica de los datos.

Jill comenzó el primer paso con la identificación de un modelo tentativo observando las autocorrelaciones de los datos que aparecen en fig. 10.21, para de determinar si la serie es estacionaria. Las autocorrelaciones de los retrasos inferiores descienden rápidamente hacia cero, así que Jill concluye que la serie histórica es estacionaria.

TABLA 10.7 CIERRE DE ACCIONES DE LA ISC CORPORATION

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

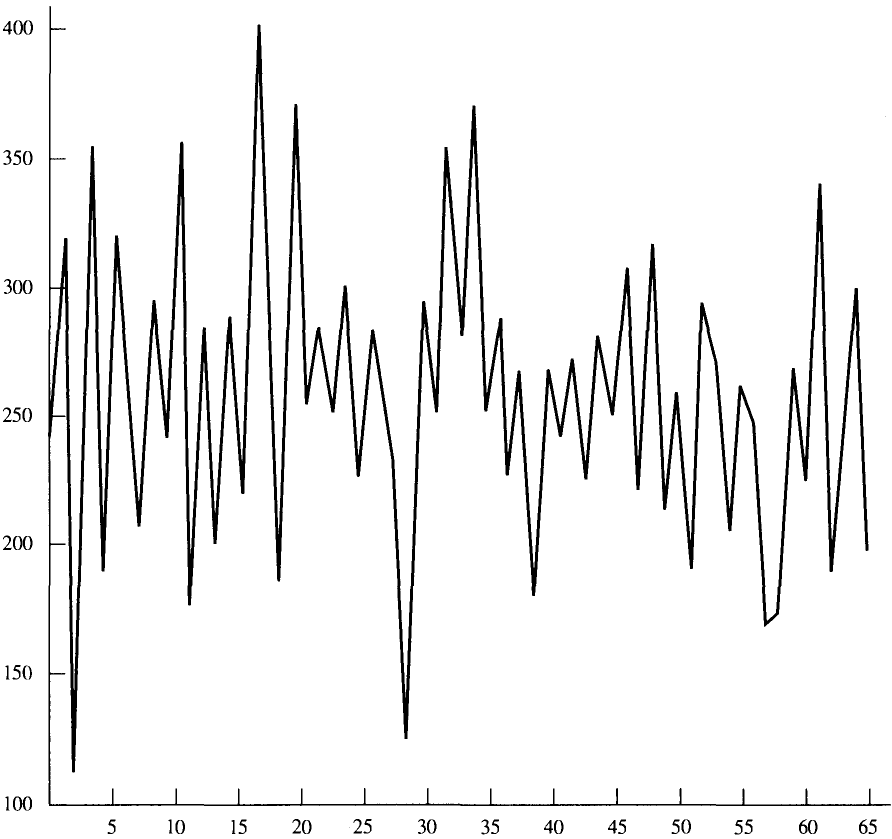


Figura 10.20 Cierre de acciones de la ISC Corporation.

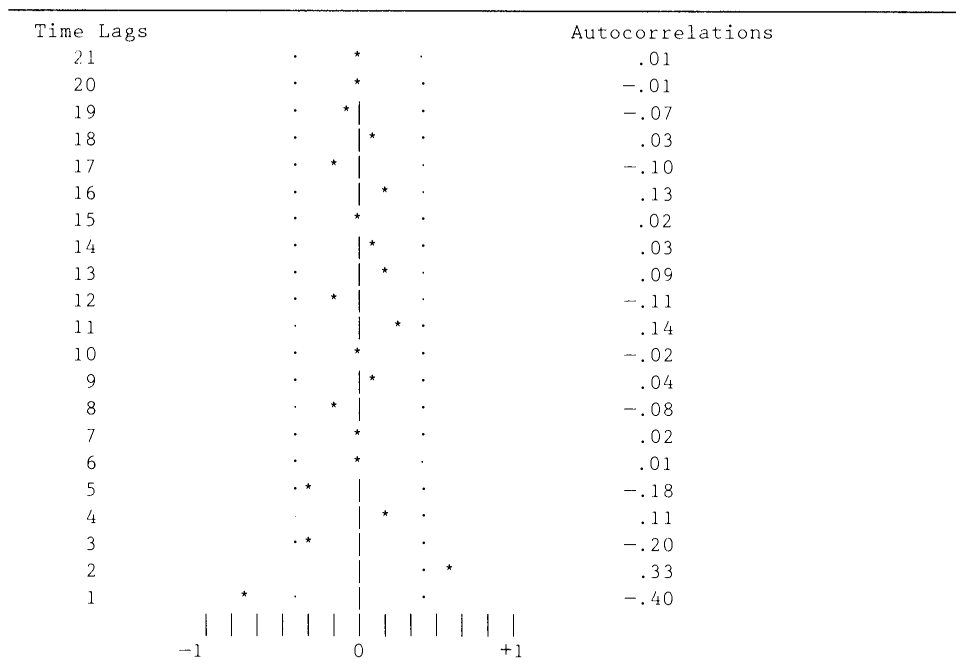


Figura 10.21 Autocorrelaciones de los datos sin procesar de la tabla 10.7.

Los dos primeros coeficientes de autocorrelación de la muestra ($-.40$ y $.33$) son significativos a un nivel de 95%, y el resto parece descender exponencialmente hacia cero.

A partir de los coeficientes de autocorrelación parcial de la muestra, que aparecen en la fig. 10.22, Jill observó que el primero era significativo y el segundo estaba cerca de serlo. Sin embargo el tercero desciende hacia cero. Como las autocorrelaciones parciales caen hacia cero después del segundo periodo de retraso y se puede considerar que los coeficientes de autocorrelación descienden a cero en forma exponencial, se debe considerar un modelo AR(2).

A continuación, Jill utilizó un programa de cómputo de Box-Jenkins para estimar el(los) parámetro(s) del modelo. Ya que los valores de las variables independientes no fueron expresados como desviaciones de su media, se debe incluir en el modelo un término constante. Los parámetros estimados para el modelo AR(1) son 361.47 para la constante general y .40152 para el parámetro autorregresivo. Ninguno de los coeficientes de autocorrelación de los residuos es diferente de cero en forma significativa y la suma de los residuos cuadrados es igual a 2,851.4. El modelo parece ser adecuado. Jill confirma esta observación probando la estadística de ji cuadrada para 24 periodos retrasados. La ji cuadrada calculada es menor que el valor de la tabla para 22 grados de libertad a un nivel de significación de .05 ($16.45 < 33.92$).

A continuación, Jill usa el programa de cómputo para estimar los parámetros del modelo ARIMA(2, 0, 0). Se estima que los parámetros son $-.32218$ y $.20809$ con una constante general igual a 286.48. Ninguno de los coeficientes de autocorrelación de los residuos es diferente de cero de manera significativa y la suma de los residuos cuadrados es 2,719.9. Este modelo también parece ser adecuado. Jill confirma esta observación utilizando la estadística de ji

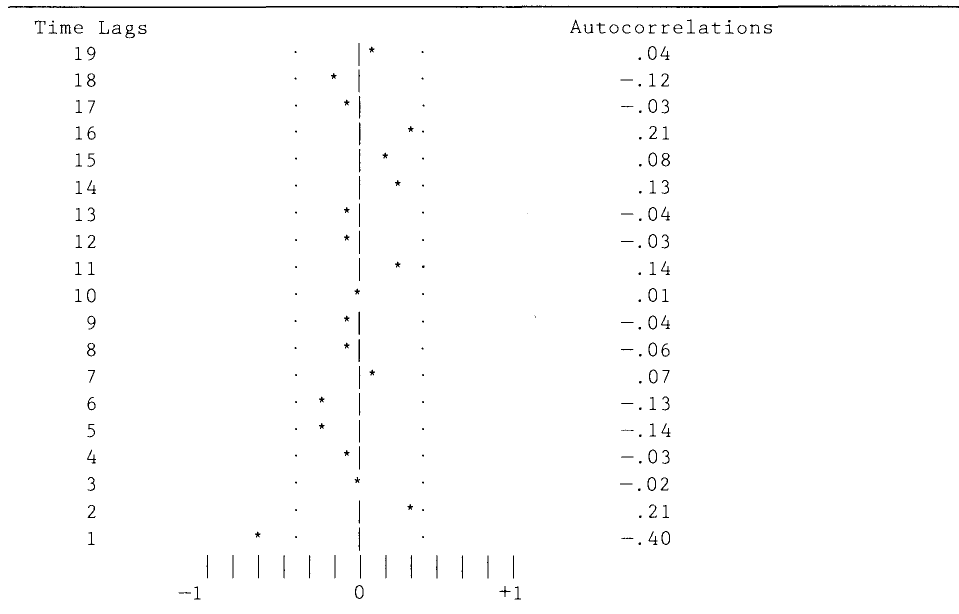


Figura 10.22 Autocorrelaciones parciales de los datos sin procesar de la tabla 10.7.

cuadrada para 24 periodos retrasados. El valor calculado de ji cuadrada es menor que el valor de la tabla para 21 grados de libertad a un nivel de significación de .05 ($13.76 < 32.57$). Ya que la suma de los residuos cuadrados es considerablemente menor se debe emplear el modelo ARIMA (2, 0, 0).

El modelo de pronóstico ARIMA (2, 0, 0) es

$$\hat{Y}_t = 268.48 + (-.32218)Y_{t-1} + .20809Y_{t-2}$$

Los pronósticos para los periodos 66 y 67 son

$$\hat{y}_{66} = 268.48 + (-.32218)(195) + .20809(300) = 268.08$$

$$\hat{y}_{67} = 268.48 + (-.32218)(268.08) + .20809(195) = 222.69$$

UN ANÁLISIS ESTACIONAL

Los patrones estacionales incorporan una mayor dificultad al pronóstico con modelos de Box-Jenkins debido que además del patrón entre periodo y periodo existe un patrón repetitivo más largo que ocurre cada L -ésimo periodo, en donde L es la longitud de la estacionalidad. Para estimar un patrón estacional, se deben incluir en el modelo parámetros estacionales. Al igual que en los modelos no estacionales, pueden ser procesos autorregresivos o de promedio móvil. Los ejemplos 10.8 y 10.9 aportarán indicaciones de cómo resolver una aplicación estacional.

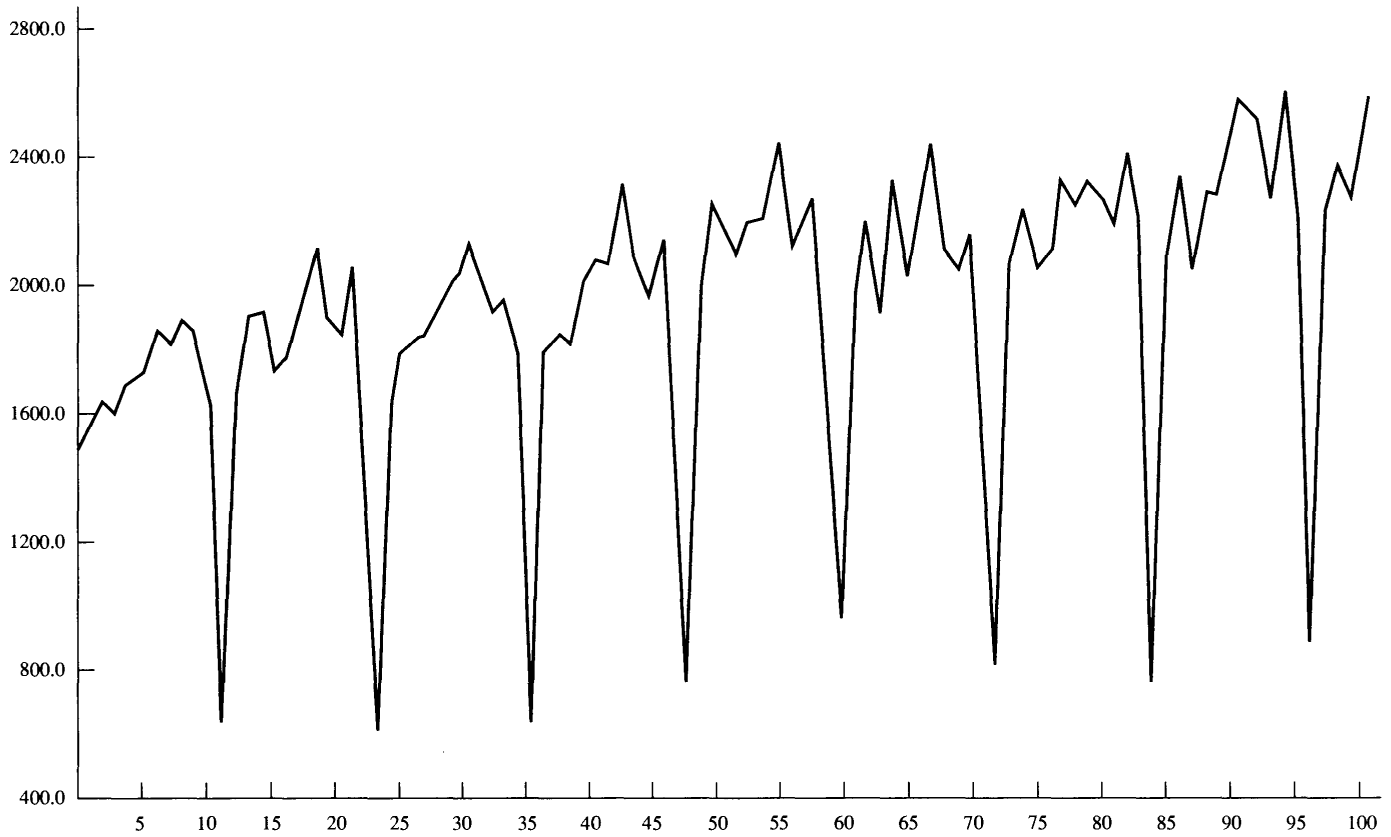


Figura 10.23 Ventas de Keytron Corporation.

**Ejemplo
10.8**

A Kathy Fisher se le ha asignado la responsabilidad de pronosticar las ventas de Keytron Corporation. La compañía empezó hace 10 años y Kathy pudo encontrar 115 meses de datos de ventas. Los datos comienzan en enero de 1986 y se muestran en la tabla 10.8. Kathy examinó la gráfica de datos de la fig. 1.23 y decidió probar la metodología de Box-Jenkins.

Kathy inició el primer paso con la identificación de un modelo tentativo observando las autocorrelaciones de los datos que se presentan en la fig. 10.24 para determinar si la serie era estacionaria. En su forma original, los datos no son estacionarios. Muchas de las correlaciones son diferentes de cero en forma significativa (los retrasos 1, 3, 9, 11, 12, 13, 23, 24 y 25). Se indica un ligero patrón de tendencia debido a que todas las correlaciones, excepto cinco, son positivas. Kathy decide hacer una primera diferencia de los datos.

La figura 10.25 muestra las autocorrelaciones de la primera diferenciación de los datos originales. El ligero patrón de tendencia ha desaparecido. Se ha logrado una serie estacionaria con una base de periodo a periodo. Sin embargo, existen muy altas autocorrelaciones en los retrasos 12, 24 y 36. Kathy concluye que existe una tendencia entre periodos estacionales sucesivos (separados por un periodo de desfase de 12 meses). Si el coeficiente de autocorrelación hubiera caído a cero en el retraso 24, Kathy habría concluido que los datos de la primera diferencia eran estacionarios.

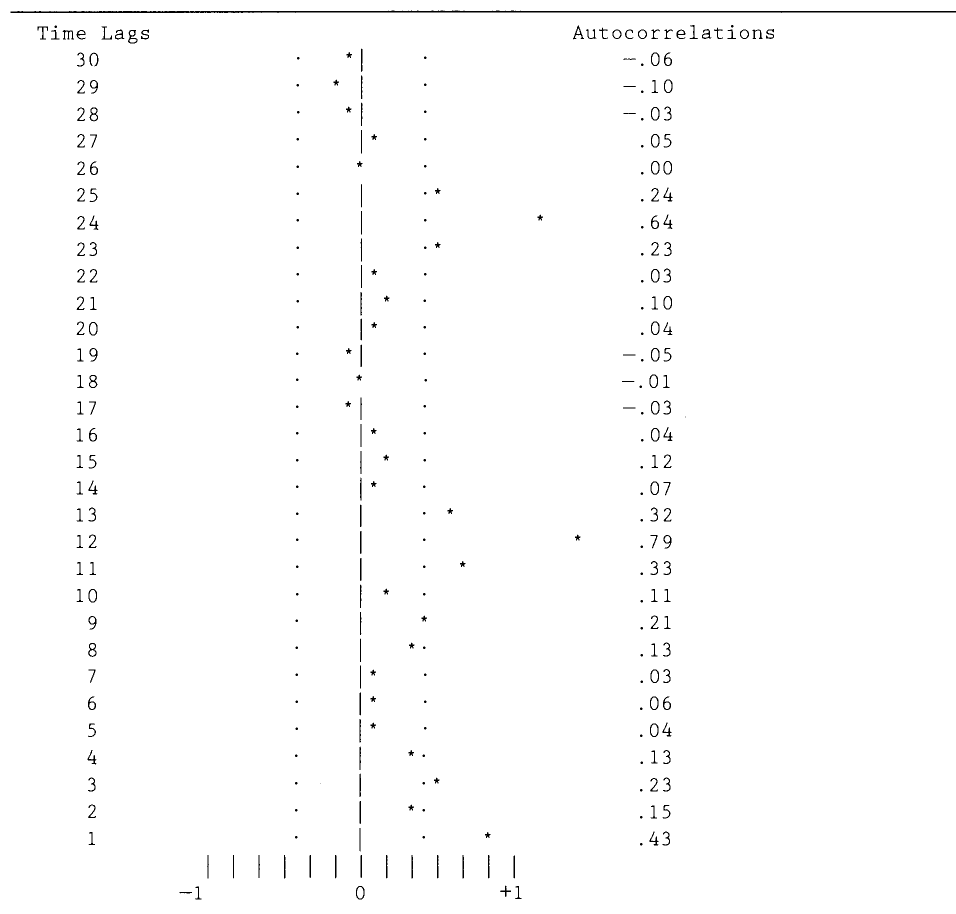


Figura 10.24 Autocorrelaciones de los datos sin procesar de la tabla 10.8.

TABLA 10.8 VENTAS DE KEYTRON CORPORATION

1,736.8	1,627.6	1,895.4	1,768.0	2,191.8	2,035.8	2,056.6
1,297.4	1,575.6	1,822.6	1,840.8	2,202.2	2,152.8	2,340.0
559.0	1,682.2	2,054.0	1,804.4	2,449.2	1,708.2	2,033.2
1,455.6	1,710.8	1,544.4	2,007.2	2,090.4	806.0	2,288.0
1,526.2	1,853.8	600.6	2,067.0	2,184.0	2,028.0	2,275.0
1,419.6	1,788.8	1,604.2	2,048.8	2,267.2	2,236.0	2,581.8
1,484.6	1,822.4	1,796.6	2,314.0	1,705.6	2,028.0	2,540.2
1,651.0	1,838.2	1,822.6	2,072.2	962.0	2,100.8	2,519.4
1,661.4	1,635.4	1,835.6	1,952.6	1,929.2	2,327.0	2,267.2
1,851.2	618.8	1,944.8	2,134.6	2,202.2	2,225.6	2,615.6
1,617.2	1,593.8	2,009.8	1,799.2	1,903.2	2,321.8	2,163.2
1,614.6	1,898.0	2,116.4	756.6	2,337.4	2,275.0	899.6
1,757.6	1,911.0	1,994.2	1,890.2	2,022.8	2,171.0	2,210.0
1,302.6	1,695.0	1,895.4	2,256.8	2,225.6	2,431.0	2,376.4
572.0	1,757.6	1,947.4	2,111.2	2,441.4	2,165.8	2,259.4
1,458.6	1,944.8	1,770.6	2,080.0	2,113.8	780.0	2,584.4
1,567.8	2,108.6	626.6				

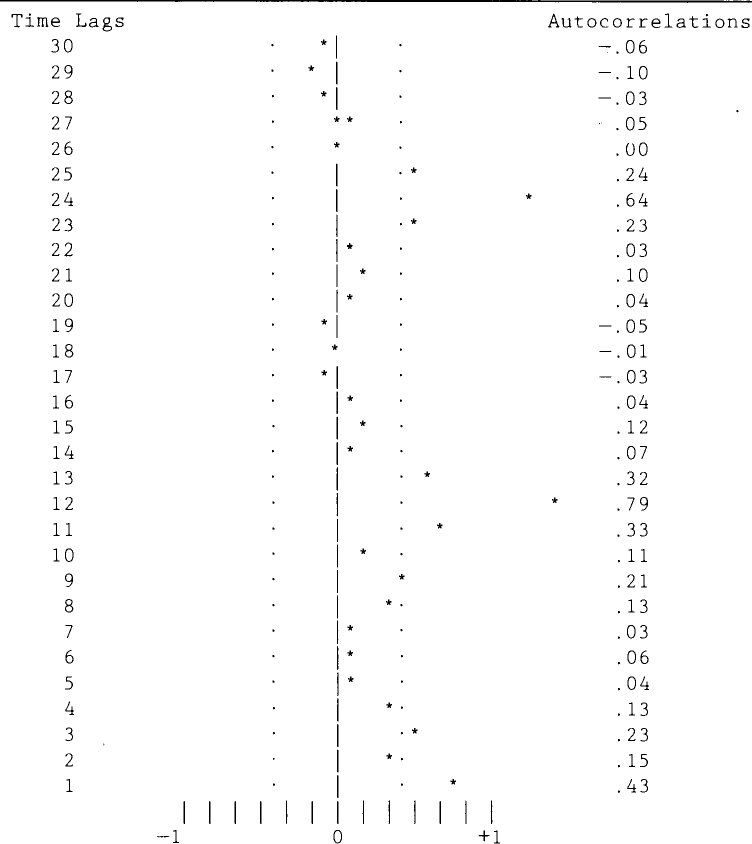


Figura 10.25 Autocorrelaciones de la primera diferencia de los datos de la tabla 10.8.

Para eliminar la tendencia para los retrasos 12, 24 y 36 se emplea la diferencia de largo plazo (diferencias cuya longitud está separada por 12 periodos). La fig. 10.26 muestra las autocorrelaciones para los datos después de la primera diferencia y de la diferencia de largo plazo (separada por un periodo de retraso de 12 meses). Sólo las autocorrelaciones en los periodos retrasados 1, 11, 12 y 13 son diferentes de cero en forma significativa. El resto de las autocorrelaciones está disperso en forma aleatoria alrededor de cero y la serie parece ser estacionaria.

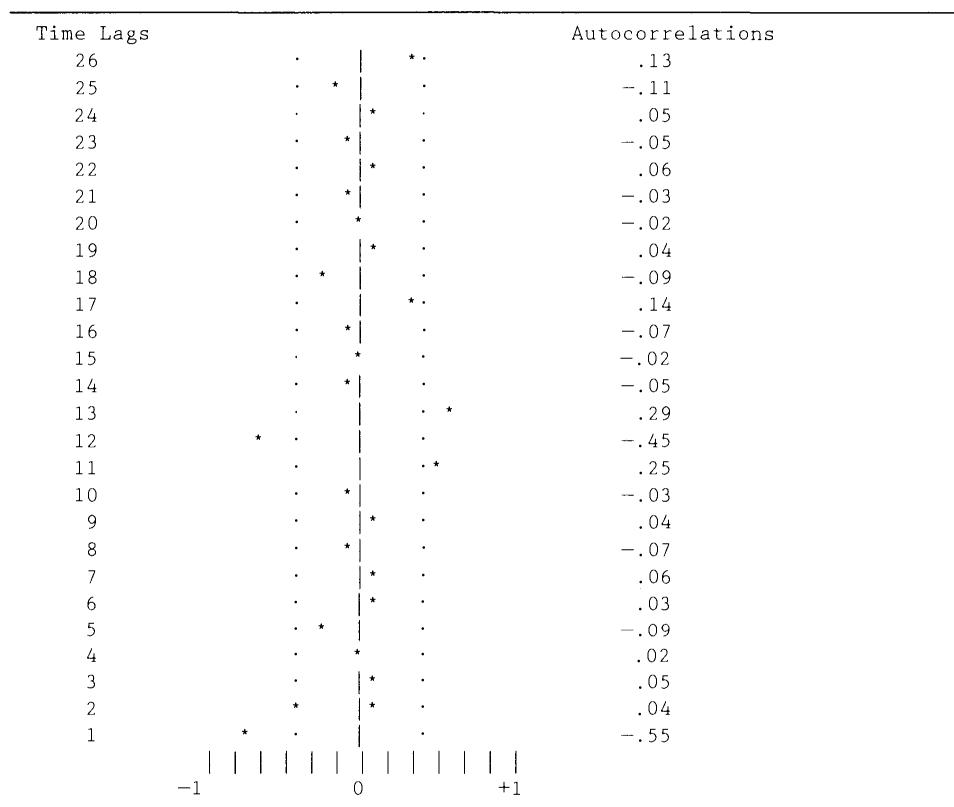


Figura 10.26 Autocorrelaciones de diferencia corta y larga de los datos de la tabla 10.8.

Kathy examina las autocorrelaciones de la fig. 10.26 y las autocorrelaciones parciales de la fig. 10.27 para identificar el modelo apropiado. Primero, se concentra en el patrón no estacional. La fig. 10.26 muestra que las autocorrelaciones descienden a cero después del primer coeficiente significativo en el periodo de retraso 1 ($-.55$). La fig. 10.27 muestra una caída exponencial de las autocorrelaciones parciales no estacionales, indicando un modelo IMA(1) para el patrón no estacional.

Ahora Kathy examina el patrón estacional. La fig. 10.26 muestra que las autocorrelaciones descienden hacia cero (.05 en el periodo retrasado 24) después del primer coeficiente significativo en el periodo de retraso 12 ($-.45$). La fig. 10.27 muestra la disminución de la

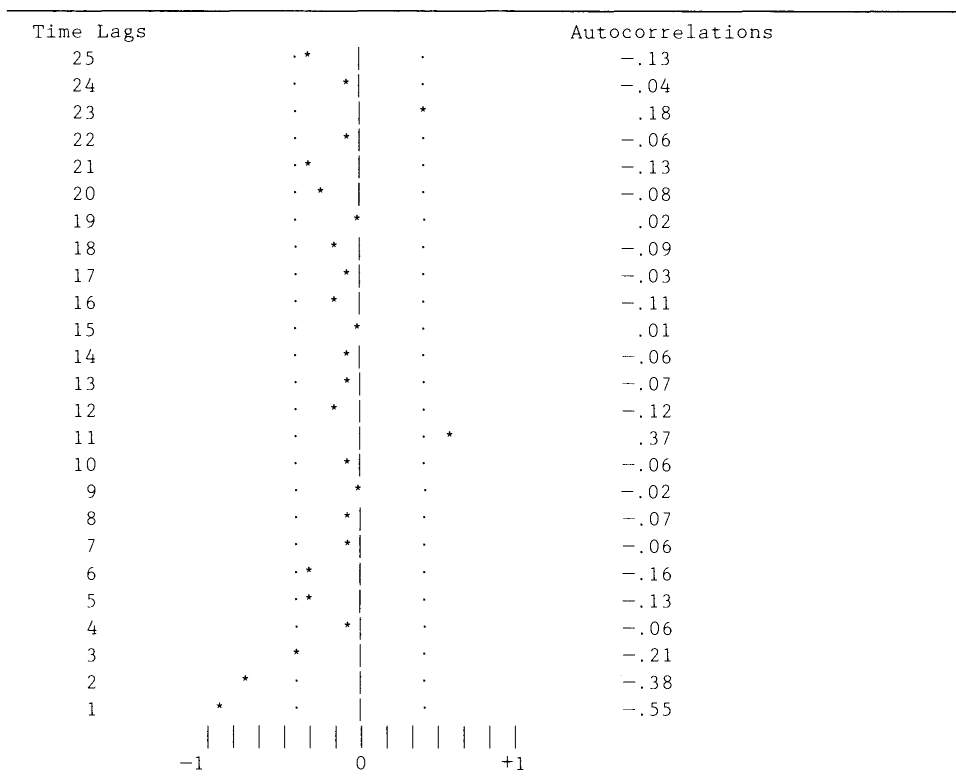


Figura 10.27 Autocorrelaciones parciales de diferenciaciones corta y larga de los datos de la tabla 10.8.

autocorrelación parcial estacional, indicando un modelo IMA(1) para el patrón estacional. Hubiera sido deseable examinar los coeficientes de los retrasos 36, 48 y 60, pero no fue posible.

A continuación, Kathy utiliza el programa de cómputo de Box-Jenkins para estimar el(los) parámetro(s) del modelo. Se estima que los parámetros para el modelo ARIMA(0, 1, 1)(0, 1, 1) son .77948 y .61031. Un examen de los coeficientes de autocorrelación de la fig. 10.28 indica que son aleatorios. Kathy confirma esta observación efectuando la prueba estadística de ji cuadrada para 34 periodos retrasados. El valor calculado de ji cuadrada es menor que el valor de la tabla a un nivel de significación de .05 ($11.82 < 43.8$). El modelo resulta adecuado.

El modelo de pronóstico ARIMA(0, 1, 1)(0, 1, 1) es

$$\hat{Y}_t = Y_{t-1} + Y_{t-2} - Y_{t-13} + \epsilon_t - .77948e_{t-1} - .61031e_{t-12} + .47572e_{t-13}$$

$$\hat{y}_{116} = y_{115} + y_{104} - y_{103} + e_{116} - .77948e_{115} - .61031e_{104} + .47572e_{103}$$

El pronóstico para el periodo 116 es

$$\begin{aligned}\hat{y}_{116} &= 2,584.4 + 2,275 - 2,288 - .77948(181.48) \\ &\quad - .61031(-36.43) + .47572(34.35)\end{aligned}$$

$$\hat{y}_{116} = 2,468.5$$

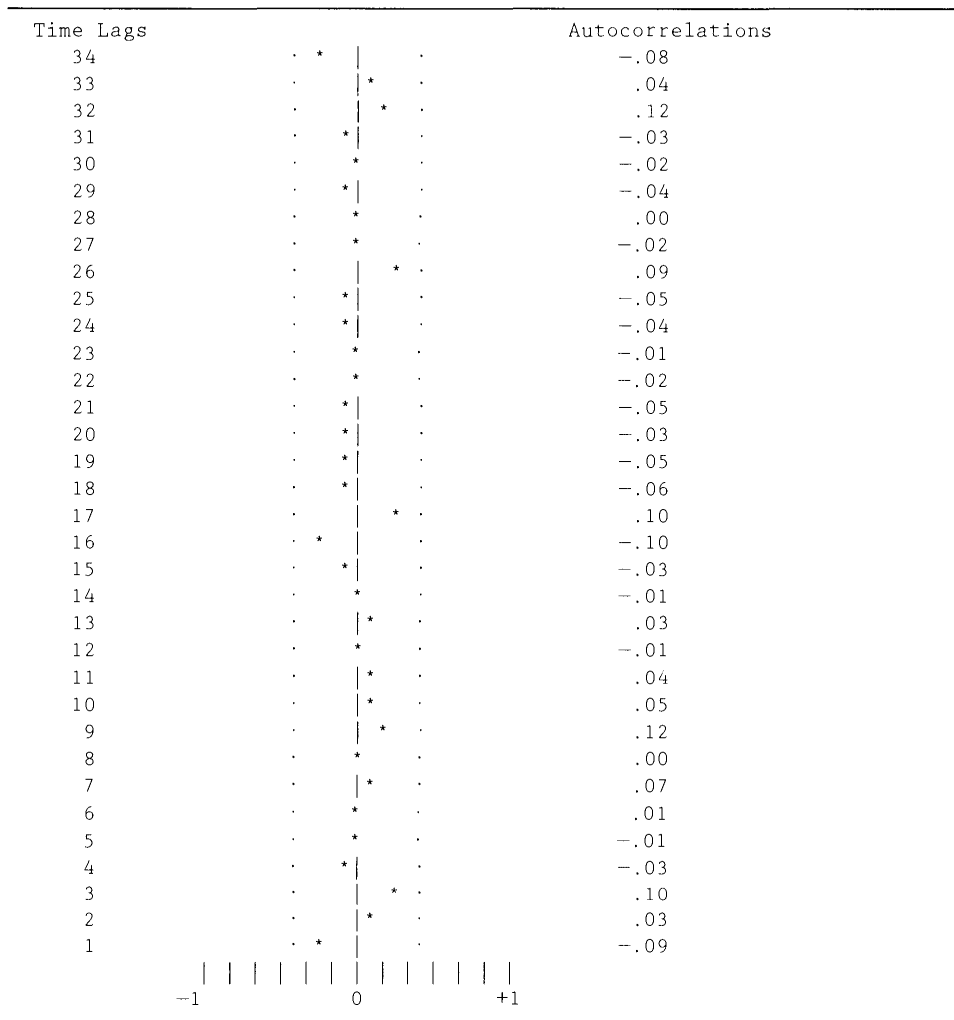


Figura 10.28 Coeficientes de autocorrelación de los residuos del modelo ARIMA(0, 1, 1)(0, 1, 1).

El pronóstico para el periodo 117 es

$$\hat{y}_{117} = 2,599.8$$

Ejemplo 10.9

En el ejemplo 4.4, Perkin Kendell, analista de Outboard Marine Corporation, utilizó un análisis de autocorrelación para concluir que la ventas trimestrales eran estacionales. Perkin decidió pronosticar las ventas de 1994 mediante la metodología de Box-Jenkins. Los datos se graficaron en la figura 4.7.

Perkin observa primero para ver si las autocorrelaciones de los datos son estacionarias. Los comandos de Minitab son

```

MTB > READ 'OUTBOARD.DAT' C1
      40 ROWS READ

C1
  147.6   251.8   273.1   249.1   .   .   .

MTB > ACF C1

ACF OF C1

      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +---+---+---+---+---+---+---+---+---+---+
  1   0.421                                XXXXXXXXXXXX
  2   0.194                                XXXXXX
  3   0.312                                XXXXXXXXXXXX
  4   0.733                                XXXXXXXXXXXXXXXXXXXX
  5   0.177                                XXXXX
  6  -0.119                                XXXX
  7  -0.052                                XX
  8   0.297                                XXXXXXXX
  9  -0.193                                XXXXXX
 10  -0.442                                XXXXXXXXXXXX
 11  -0.347                                XXXXXXXXXXXX
 12   0.028                                XX
 13  -0.361                                XXXXXXXXXXXX
 14  -0.509                                XXXXXXXXXXXXXXXXXXXX
 15  -0.386                                XXXXXXXXXXXX
 16   0.014                                X

```

Los datos no son estacionarios en su forma original. En el ejemplo 4.4. Perkin determinó que cualquier coeficiente de autocorrelación mayor que .31 o menor que $-.31$ era diferente de cero en forma significativa, utilizando un nivel de significación de .05. Muchos de los periodos de retraso tienen autocorrelaciones diferentes de cero de manera significativa (retrasos 1, 2, 4, 10, 11, 13, 14 y 15). Ya que las primeras cinco autocorrelaciones son positivas, pudiera existir una ligera tendencia.

Perkin decide hacer, utilizando Minitab, una primera diferencia de los datos y examinar los coeficientes de autocorrelación:

```

MTB > DIFFERENCE 1 C1 STORE C2
MTB > ACF C2

ACF OF C2

      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +---+---+---+---+---+---+---+---+---+---+
  1  -0.275                                XXXXXXXX
  2  -0.307                                XXXXXXXXXXXX
  3  -0.282                                XXXXXXXX

```


4	0.836	XXXXXXXXXXXXXXXXXXXXXXX
5	-0.200	XXXXXX
6	-0.316	XXXXXXXXXX
7	-0.254	XXXXXXX
8	0.712	XXXXXXXXXXXXXXXXXXXXXXX
9	-0.177	XXXXXX
10	-0.288	XXXXXXXXXX
11	-0.260	XXXXXXXXXX
12	0.640	XXXXXXXXXXXXXXXXXXXXXXX
13	-0.153	XXXXXX
14	-0.217	XXXXXXX
15	-0.242	XXXXXXXXXX
16	0.540	XXXXXXXXXXXXXXXXXXXXXXX

Las autocorrelaciones para los datos de la primera diferencia muestran que se eliminó la tendencia y se logró una serie estacionaria en una base de periodo a periodo. Sin embargo, como hay altas autocorrelaciones en los retrasos 4 (.836), 8 (.712), 12 (.64) y 16 (.54), no se ha logrado una serie estacionaria sobre una base estacional. Si la autocorrelación hubiera caído hacia cero en el retraso 8, Perkin hubiera podido concluir que los datos de la primera diferencia eran estacionarios.

Para eliminar la tendencia estacional, se emplea la diferenciación de largo plazo (diferencias cuya longitud está separada por cuatro periodos). La columna 2 contiene los datos de la primera diferencia y la columna 3 contiene los datos de la primera diferencia con una diferencia de largo plazo (separada por un periodo de retraso de 4 trimestres). A continuación, se calculan las autocorrelaciones para esta serie (C3):

MTB > DIFFERENCE 4 C2 C3			
MTB > PRINT CI -C3			
ROW	C1	C2	C3
1	147.6	*	*
2	251.8	104.200	*
3	273.1	21.300	*
4	249.1	-24.000	*
5	139.3	-109.800	*
6	221.2	81.900	-22.3000
7	260.2	39.000	17.7000
8	259.5	-0.700	23.3000
9	140.5	-119.000	-9.2000
10	245.5	105.000	23.1000
11	298.8	53.300	14.3000
12	287.0	-11.800	-11.1000
13	168.8	-118.200	0.8000
14	322.6	153.800	48.8000
15	393.5	70.900	17.6000
16	404.3	10.800	22.6000
17	259.7	-144.600	-26.4000
18	401.1	141.400	-12.4000
19	464.6	63.500	-7.4000
20	479.7	15.100	4.3000

21	264.4	-215.300	-70.7000
22	402.6	138.200	-3.2000
23	411.3	8.700	-54.8000
24	385.9	-25.400	-40.5000
25	232.7	-153.200	62.1000
26	309.2	76.500	-61.7000
27	310.7	1.500	-7.2000
28	293.0	-17.700	7.7000
29	205.1	-87.900	65.3000
30	234.4	29.300	-47.2000
31	285.4	51.000	49.5000
32	258.7	-26.700	-9.0000
33	193.2	-65.500	22.4000
34	263.7	70.500	41.2000
35	292.5	28.800	-22.2000
36	315.2	22.700	49.4000
37	178.3	-136.900	-71.4000
38	274.5	96.200	25.7000
39	295.4	20.900	-7.9000
40	311.8	16.400	-6.3000

MTB > ACF C3

ACF OF C3

	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
	+---+---+---+---+---+---+---+---+---+---+										
1	-0.341					XXXXXXXXXX					
2	0.239						XXXXXXX				
3	0.051						XX				
4	0.041						XX				
5	0.053						XX				
6	-0.085					XXX					
7	0.035						XX				
8	-0.386					XXXXXXXXXX					
9	0.175						XXXXX				
10	-0.348					XXXXXXXXXX					
11	0.147						XXXXX				
12	-0.284					XXXXXXXXX					
13	-0.055						XX				
14	0.073						XXX				
15	-0.010						X				

Sólo las autocorrelaciones en los periodos retrasados 1, 2, 8 y 10 son diferentes de cero en forma significativa. El resto de las autocorrelaciones estan dispersas de manera aleatoria alrededor de cero y la serie parece ser estacionaria.

Perkin ejecuta a continuación las autocorrelaciones parciales para identificar el modelo apropiado:

```
MTB > PACF C3
```

```
PACF OF C3
```

	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
	+-----+-----+-----+-----+-----+-----+-----+-----+-----+										
1	-0.341					XXXXXXXXXX					
2	0.139						XXXX				
3	0.194						XXXXXX				
4	0.092						XXX				
5	0.043						XX				
6	-0.117					XXXX					
7	-0.073					XXX					
8	-0.457			XXXXXXXXXXXX							
9	-0.106			XXXX							
10	-0.221			XXXXXXXXXX							
11	0.136					XXXX					
12	-0.093					XXX					
13	-0.140					XXXX					
14	0.018					X					
15	0.147					XXXXX					

Perkin se concentra primero en el patrón no estacional. Advierte que los coeficientes de autocorrelación caen cerca de cero después del segundo periodo retrasado y que las autocorrelaciones parciales muestran una disminución gradual. La autocorrelación en el periodo retrasado 2 parece ser crucial para la determinación del modelo. Si este coeficiente es significativo, resultaría apropiado un modelo MA(2). Si no es significativo, el adecuado pudiera ser un modelo ARIMA(1, 1).

A continuación, Perkin pone su atención en el patrón estacional. Como el retraso 4 no es significativo, resulta difícil decir si sería apropiado un modelo MA o uno AR. Decide probar ambos. El mejor modelo comprende un término MA(2) para el no estacional y un término MA(1) para el estacional. Los comandos de Minitab para ejecutar el mejor modelo son

```
MTB > ARIMA (0 1 2) (0 1 1) 4 C1 PUT C4;
SUBC> FORECASTS 4.
```

```
Estimates at each iteration
```

Iteration	SSE	Parameters		
0	44598.9	0.100	0.100	0.100
1	40342.5	0.160	-0.050	0.125
2	37801.7	0.253	-0.200	0.092
3	36716.7	0.329	-0.279	0.002
4	35963.1	0.354	-0.293	-0.099
5	35388.2	0.376	-0.308	-0.192
6	35046.1	0.398	-0.320	-0.265
7	34878.4	0.417	-0.329	-0.315
8	34804.6	0.431	-0.335	-0.347
9	34773.5	0.440	-0.339	-0.368

```

10      34760.6      0.447   -0.342   -0.382
11      34755.3      0.451   -0.344   -0.390
12      34753.1      0.453   -0.345   -0.395
13      34752.2      0.455   -0.345   -0.399
14      34751.9      0.456   -0.346   -0.401
15      34751.7      0.457   -0.346   -0.403
16      34751.6      0.457   -0.346   -0.404
17      34751.6      0.457   -0.346   -0.404
18      34751.6      0.458   -0.347   -0.405
Relative change in each estimate less than 0.0010

```

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
MA 1	0.4576	0.1659	2.76
MA 2	-0.3465	0.1683	-2.06
SMA 4	-0.4045	0.1831	-2.21

Differencing: 1 regular, 1 seasonal of order 4

No. of obs.: Original series 40, after differencing 35

Residuals: SS = 34351.0 (backforecasts excluded)
MS = 1073.5 DF = 32

Modified Box-Pierce chisquare statistic

Lag	12	24	36	48
Chisquare	14.5(DF= 9)	25.0(DF=21)	* (DF= *)	* (DF= *)

Forecasts from period 40

Period	Forecast	95 Percent Limits		Actual
		Lower	Upper	
41	154.756	90.526	218.986	
42	248.895	175.824	321.966	
43	273.166	180.431	365.900	
44	277.728	168.825	386.632	

- El comando ARIMA(0 1 2)(0 1 1) 4 C1 PUT C4 se usa para ejecutar el modelo. La primera parte (0 1 2) hace una primera diferencia de los datos no estacionales y ajusta un modelo MA(2). La segunda parte (0 1 1) 4 crea una diferencia estacional de cuatro órdenes y ajusta un modelo MA(1). Los residuos se almacenan en la columna 4.

A continuación, Perkin desarrolla un correlograma de las autocorrelaciones de los residuos.

```
MTB > ACF C4
```

```
ACF OF C4
```

```

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
          +-----+-----+-----+-----+-----+-----+
1    0.027                                     XX
2   -0.018                                     X
3    0.117                                     XXXX
4   -0.097                                     XXX

```

5	0.108	XXXX
6	0.028	XX
7	-0.263	XXXXXXXX
8	-0.283	XXXXXXXX
9	0.006	X
10	-0.230	XXXXXXXX
11	0.026	XX
12	-0.239	XXXXXXXX
13	-0.196	XXXXXX
14	0.205	XXXXXX
15	0.034	XX

Una revisión de estos coeficientes de autocorrelación de los residuos indica que son aleatorios. Perkin confirma esta observación probando la estadística de ji cuadrada para 12 y 24 periodos retrasados. El valor calculado de ji cuadrada es menor que el valor de la tabla a un nivel de significación de .05 para ambos periodos ($14.5 < 16.9$) y ($25 < 32.7$). También son significativas las proporciones t para cada uno de los parámetros estimados. El modelo resulta adecuado. Se estima que los parámetros son 0.4576 para el término MA(1), -0.3465 para el término MA(2) y -0.4045 para el término estacional MA(1).

El enfoque de Box-Jenkins al análisis de series de tiempo es una muy poderosa herramienta para proporcionar pronósticos de corto plazo más precisos. Combina las fuerzas de los métodos autorregresivo y de promedio móvil sin hacer suposiciones acerca del número de términos en la ecuación de pronóstico o sobre las intercorrelaciones de sus coeficientes. Además, se proporciona una prueba estadística para determinar lo adecuado del modelo ajustado, así como un medio para construir intervalos de confianza sobre los pronósticos.

Sin embargo, el enfoque de Box-Jenkins tiene sus desventajas. Algunas de ellas son las siguientes.

1. Se requiere una cantidad de datos relativamente grande. Debe reconocerse si los datos son estacionales, como en los datos con estacionalidad anual, entonces las observaciones mensuales para un año constituyen por lo general 1 punto de datos y no 12. Makridakis y colegas estiman que 72 puntos de datos es el mínimo requerimiento para un uso confiable del método de Box-Jenkins, asumiendo un patrón estacional de 12 meses de duración. En muchos casos de este tipo, simplemente no hay disponibles datos históricos suficientes para utilizar el método de Box-Jenkins, y se deberá utilizar algún otro método de series de tiempo.

En general, se considera aplicable el método de Box-Jenkins cuando los datos ocurren en un periodo relativamente corto. Algunas de las aplicaciones apropiadas para este método serían el análisis de precios de acciones sobre una base diaria o semanal de datos, y los datos de un proceso químico o de manufactura en el que se pudieran tomar muestras a menudo.

2. No hay formas fáciles de actualizar los parámetros del modelo al haber disponibles nuevos datos, como las hay en los modelos directos de atenuación. El modelo debe reajustarse periódicamente o, peor aún, debe desarrollarse un modelo nuevo.
3. La construcción de un modelo satisfactorio requiere de una gran inversión del tiempo del analista y en otros recursos. Los costos del desarrollo del modelo, el tiempo de ejecución en la computadora y los requerimientos de almacenamiento son sustancialmente mayores para los modelos de Box-Jenkins que para las técnicas más tradicionales como la atenuación.²

APLICACIÓN EN LA ADMINISTRACION

Los pronósticos son uno de los insumos más importantes de que disponen los administradores para ayudarse en el proceso de toma de decisiones.

Prácticamente, cada decisión importante de operación depende en algún grado de un pronóstico. La acumulación de inventario se relaciona con el pronóstico de la demanda esperada; el departamento de producción requiere de calendarizar sus necesidades de personal y de pedidos de materias primas para uno o dos meses siguientes; el departamento de finanzas debe adecuar el financiamiento de corto plazo para el próximo trimestre; el departamento de personal debe determinar los requerimientos de contratación y descanso. La lista de aplicaciones de los pronósticos es muy larga.

Los ejecutivos están conscientes de la importancia de los pronósticos. De hecho, se invierte una gran cantidad de tiempo en el estudio de las tendencias en la economía, en los asuntos políticos y en cómo los sucesos pudieran afectar la demanda de productos y/o servicios. Un aspecto de particular interés es la importancia que los ejecutivos asignan a los métodos cuantitativos de pronóstico comparada con sus propias opiniones. Este aspecto es particularmente sensible cuando se encuentran involucrados sucesos que tienen un impacto significativo sobre la demanda. Un problema con los métodos cuantitativos de pronóstico es que dependen de datos históricos. Por esta razón son probablemente los menos

² Para un resumen de los requerimientos relativos de tiempos de analista y computadora, requerimientos mínimos de datos, aplicaciones, etc., para diversos métodos de pronóstico, consúltese S. Makridakis *et al.*, "An interactive Forecasting System", *American Statistician*, vol 28 (1974), p. 157. Para un resumen de algunas de las desventajas de la metodología de Box-Jenkins, se remite al lector a D. C. Montgomery y L. A. Johnson, *Forecasting and Time Series Analysis* (New York: McGraw-Hill, 1976). Los niveles de uso de diversas técnicas de pronóstico de demanda se muestran en S. G. Taylor, "The APICS Process Industry Survey: Implications for Education and Research", *Proceedings and Abstracts of The American Institute for Decision Sciences Eighth Annual Meeting, Western Regional Conference, 1979*, pp. 202-204.

efectivos para detectar los giros que con frecuencia dan por resultado agudos cambios, hacia arriba o hacia abajo, en la demanda. La computadora ha auxiliado a los administradores en automatizar la sensibilidad a los cambios agudos en la demanda a través de sistemas adaptables y modelos causales.

Un tipo de problema que enfrentan con frecuencia los administradores de empresas es la necesidad de preparar pronósticos de corto plazo para varios elementos diferentes. Un ejemplo típico es el del gerente que debe calendarizar la producción con base en algún pronóstico de demanda para varios cientos de productos diferentes en una línea de producción. A las técnicas que se emplean con más frecuencia en esta situación se les conoce como métodos de atenuación.

La atenuación exponencial es una técnica popular para el pronóstico a corto plazo. Sus mayores ventajas son el bajo costo y la simplicidad. Por lo regular no es tan precisa como otros métodos más complejos, como las técnicas autorregresivas. Sin embargo, cuando se requieren pronósticos de inventarios que contienen miles de elementos, los métodos de atenuación son a menudo el único enfoque razonable.

El análisis de series de tiempo depende de la suposición de que hay componentes regulares y repetitivos que interactúan para producir una serie total. La tarea del analista consiste en examinar e identificar cada uno de estos componentes. Una vez que los pronosticadores tienen algún conocimiento de cada una de las partes, están en una mejor posición para hacer señalamientos sobre el valor esperado de la serie total en algún periodo futuro. Las series de tiempo son apropiadas para pronosticar variables que fluctúan en cierto patrón estable a través del tiempo.

Los modelos autorregresivos se emplean con variables económicas a fin de contabilizar las correlaciones entre observaciones adyacentes en una serie histórica.

La técnica de filtración adaptable es a la vez sencilla y económica. Proporciona la característica única de ajustar los parámetros de manera automática. Es claro que resulta más poderosa que los métodos de atenuación, aunque también es más complicada. La filtración adaptable llena una brecha entre las técnicas de pronóstico ya que puede manejar en forma efectiva cualquier patrón de datos.

El enfoque de Box–Jenkins es una herramienta muy poderosa para proveer pronósticos más precisos de corto plazo. En general se le considera más aplicable cuando los datos ocurren en un periodo relativamente corto. Los administradores deben tener en cuenta que la construcción de un modelo satisfactorio de Box–Jenkins requiere de una alta inversión en tiempo de análisis y en los recursos de cómputo de la empresa.

Aplicaciones apropiadas a la metodología de Box–Jenkins son el análisis de precios de acciones sobre una base diaria o el análisis de datos de procesos químicos o de manufactura en los que se pueden tomar muestras con frecuencia. La metodología de Box–Jenkins también se ha utilizado para

Estimar un cambio de precio en la estructura de la industria telefónica de EUA

Investigar la relación entre la concentración de amoníaco, la proporción del flujo y la temperatura en los ríos

Analizar la competencia entre rutas aéreas y ferroviarias

Pronosticar el empleo

Analizar un gran número de series de tiempo de consumo de energía para una empresa
 Analizar el efecto de las promociones en las ventas de productos de consumo
 Pronosticar diferentes categorías de asegura de la calidad de un producto.

GLOSARIO

Técnicas de Box-Jenkins Las técnicas de Box-Jenkins aplican métodos autorregresivos y de promedio móvil a problemas de pronóstico de series de tiempo.

FÓRMULAS CLAVE

Box-Jenkins: modelo autorregresivo

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \quad (10.1)$$

Box-Jenkins: modelo de promedio móvil

$$Y_t = w_0 + \epsilon_t - w_1 \epsilon_{t-1} - w_2 \epsilon_{t-2} + \cdots + w_q \epsilon_{t-q} \quad (10.2)$$

Box-Jenkins: modelo ARIMA

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t - w_1 \epsilon_{t-1} - w_2 \epsilon_{t-2} - \cdots - w_q \epsilon_{t-q} \quad (10.3)$$

Estadística Q de Box-Pierce
$$Q = (N - d) \sum_{k=1}^m r_k^2 \quad (10.4)$$

PROBLEMAS

1. a. Calcule un intervalo de confianza de 95% para los coeficientes de autocorrelación de una muestra de 100 observaciones de datos aleatorios.
 b. ¿Qué conclusión se puede obtener si todos los coeficientes de autocorrelación se encuentran dentro del intervalo y no muestran un patrón particular?
 c. ¿Qué conclusión se puede obtener si los primeros cinco coeficientes de autocorrelación son diferentes de cero en forma significativa y el patrón desciende a cero en forma gradual?
 d. ¿A qué conclusión se puede llegar si r_4 , r_8 y r_{12} son diferentes de cero en forma significativa?
2. Suponga que el siguiente modelo de serie histórica se ajustó a los datos históricos y se ha verificado que es un modelo adecuado.

$$Y_t = 35 + e_t + .25e_{t-1} - .3e_{t-2}$$

Las primeras cuatro observaciones son $y_1 = 32.5$, $y_2 = 36.6$, $y_3 = 33.3$ y $y_4 = 31.9$. Calcule los pronósticos para los periodos 5, 6 y 7 a partir del origen 4. Haga e_{t-1} y e_{t-2} igual a cero para el periodo 1.

3. Se ajustó un modelo de serie histórica dando por resultado

$$\hat{Y}_t = 50 + .45y_{t-1} + e_t$$

Suponga que en el periodo $t = 50$ la observación es $y_{50} = 100$.

- Determine los pronósticos para los periodos 51, 52 y 53 a partir del origen 50.
 - Suponga que el valor observado de y_{51} es 90. Revise los pronósticos de los periodos 52 y 53.
 - Suponga que la estimación de la varianza del término de error $\hat{\sigma}^2$ es 1.2. Calcule un intervalo de confianza del 95% sobre la estimación del periodo 51.
4. Complete la información faltante en el siguiente cuadro, indicando si la distribución teórica de las autocorrelaciones y de las autocorrelaciones parciales descende o cae a cero para estos modelos.

MODEL	AUTOCORRELATIONS	PARTIAL-AUTOCORRELATIONS
AR		
MA		
ARIMA		

- Dadas las gráficas en las figs. P5a a la P5c de las autocorrelaciones y autocorrelaciones parciales de los datos originales de alguna serie, identifique un modelo potencial para cada una.
- Después de ajustar un modelo MA(1) a la primera diferenciación de 24 observaciones de una serie histórica, en la fig. P.6, se muestran las autocorrelaciones de los diez primeros residuos.

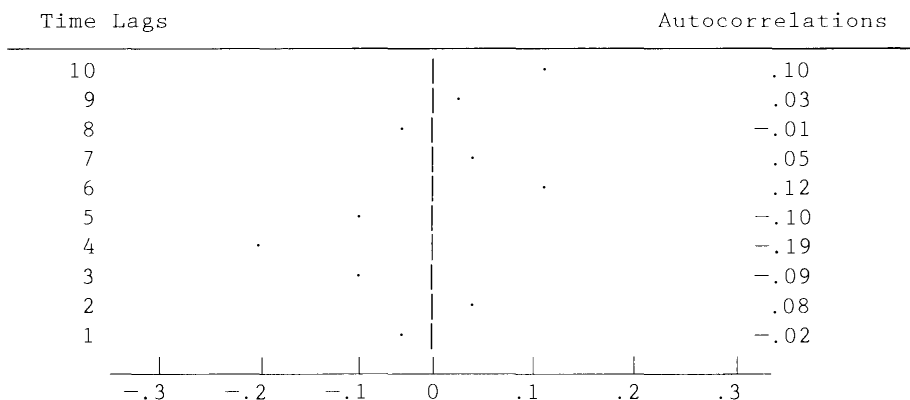


Figura P.6.

- ¿Qué le dice la inspección de los coeficientes de autocorrelación de los residuos con respecto a lo adecuado del modelo?
- Calcule la estadística de ji cuadrada para estos datos y haga una prueba para determinar si el modelo es adecuado con base en la prueba de ji cuadrada.

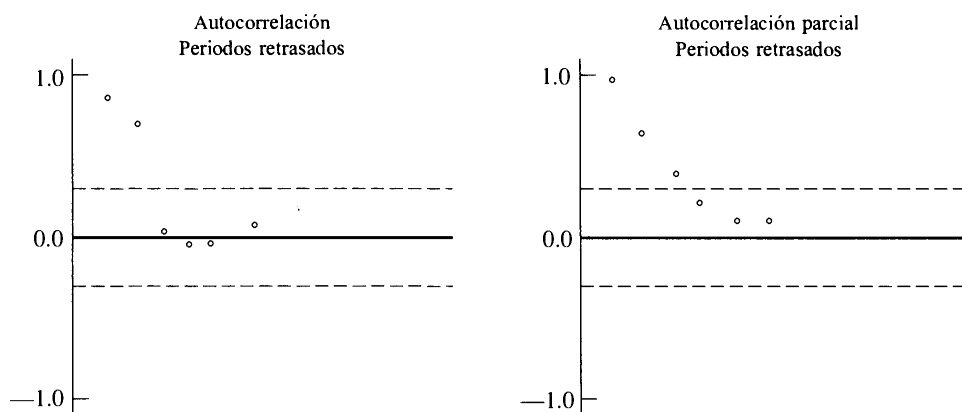


Figura P.5a.

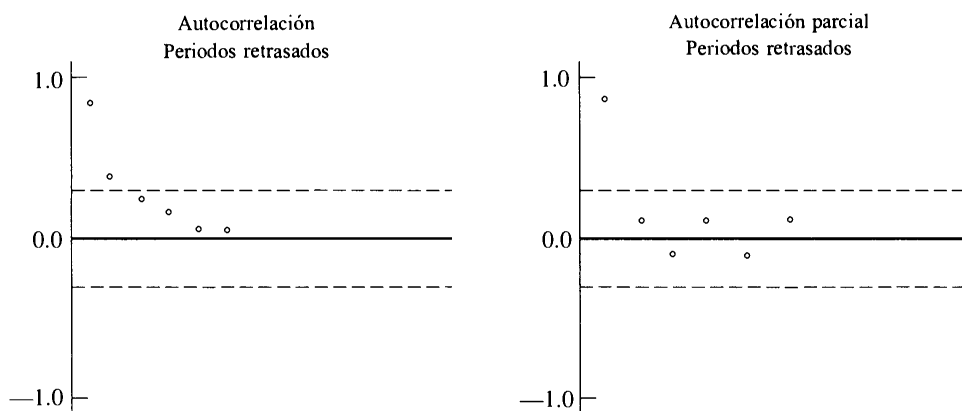


Figura P.5b.

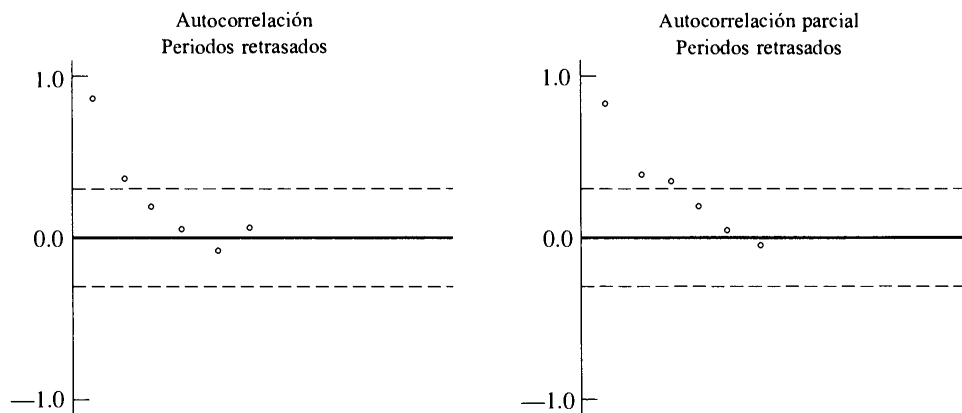


Figura P.5c.

7. Chips Bakery ha tenido problemas en el pronóstico de la demanda de su pan integral y le gustaría que usted le ayudara. En los cuadros P7a al P7d se presentan los datos de demanda semanal y las autocorrelaciones para los datos sin procesar y sus diversas diferencias.

TABLA P.7A DEMANDA DE VENTAS SEMANALES (EN MILES) DE PAN INTEGRAL

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 FUNCIÓN DE AUTOCORRELACIÓN DE LOS DATOS SIN PROCESAR

RETRASO	AUTOCORRELACIÓN	RETRASO	AUTOCORRELACIÓN
1	.94	7	.59
2	.88	8	.53
3	.82	9	.48
4	.77	10	.43
5	.71	11	.38
6	.65	12	.32

TABLA P.7C FUNCIÓN DE AUTOCORRELACIÓN DE LA PRIMERA DIFERENCIA DE LA SERIE

RETRASO	AUTOCORRELACIÓN	RETRASO	AUTOCORRELACIÓN
1	-.40	7	.20
2	-.29	8	-.03
3	.17	9	-.03
4	.21	10	-.23
5	-.22	11	.21
6	-.05	12	.14

TABLA P.7D FUNCIÓN DE AUTOCORRELACIÓN DE LA SEGUNDA DIFERENCIA DE LA SERIE

RETRASO	AUTOCORRELACIÓN	RETRASO	AUTOCORRELACIÓN
1	-.53	7	.16
2	-.10	8	-.05
3	.11	9	.06
4	.18	10	-.23
5	-.20	11	.16
6	-.04	12	.13

- Inspeccione las gráficas y sugiera un modelo apropiado para el pronóstico de demanda de ventas semanales. ¿Por qué se decidió por este modelo?
 - ¿Cuál es la ecuación para el pronóstico de demanda de ventas semanales del pan integral?
 - Realice las pruebas necesarias de diagnóstico para determinar si el modelo elegido es adecuado o no.
 - Utilizando el programa de Box-Jenkins, pronostique la demanda para las siguientes cuatro semanas, comenzando con la semana 53 y construya intervalos de confianza del 95% sobre estos pronósticos.
8. Los datos de la tabla P.8 representan cotizaciones semanales de acciones de IBM.

TABLA P.8

PERIODO	IBM	PERIODO	IBM	PERIODO	IBM	PERIODO	IBM
Ene. 6	267	Abr. 7	241	Jul. 7	258	Oct. 6	279
13	267	14	244	14	259	13	287
20	268	21	254	21	268	20	276
27	264	28	262	28	276	27	273
Feb. 3	263	May 5	261	Ago. 4	285	Nov. 3	270
10	260	12	265	11	288	10	264
17	256	19	261	18	295	17	261
24	256	26	261	25	297	24	268
Mar. 2	252	Jun. 2	257	Sep. 1	292	Dic. 1	270
10	245	9	268	8	299	8	276
17	243	16	270	15	294	15	274
24	240	23	266	22	284	22	284
31	238	30	259	29	277	29	304

- Obtenga gráficas de los datos, autocorrelaciones y autocorrelaciones parciales mediante el programa de Box-Jenkins; utilice después esta información para sugerir un modelo o modelos apropiado(s) de pronóstico.
- ¿Es la serie estacionaria? ¿Qué corrección recomendaría si los datos no son estacionarios?

- c. Utilice el programa de Box–Jenkins para calcular valores mejorados de los parámetros.
 - d. Realice pruebas de diagnóstico para determinar si el modelo es adecuado.
 - e. Después de encontrar un modelo satisfactorio, haga los pronósticos de enero para la primera semana del próximo año.
9. Los datos de la tabla P.9 corresponden a 150 días de las cotizaciones al cierre de acciones de la DEF Corporation. Determine el modelo apropiado de Box–Jenkins y pronostique 5 días a futuro. ¿Qué tan preciso es su pronóstico?

TABLA P.9

PERIODO	DEF	PERIODO	DEF	PERIODO	DEF
1	136.0	51	140.8	101	136.0
2	132.8	52	135.2	102	137.6
3	130.4	53	133.6	103	131.2
4	128.8	54	134.4	104	136.0
5	136.8	55	134.4	105	136.0
6	135.2	56	137.6	106	133.6
7	134.4	57	134.4	107	129.6
8	139.2	58	140.8	108	132.8
9	136.8	59	137.6	109	135.2
10	136.0	60	132.8	110	132.0
11	133.6	61	136.8	111	132.8
12	139.2	62	135.2	112	132.8
13	137.6	63	132.8	113	136.0
14	139.2	64	144.0	114	136.8
15	139.2	65	137.6	115	136.8
16	136.0	66	138.4	116	133.6
17	138.4	67	136.0	117	134.4
18	137.6	68	135.2	118	130.4
19	139.2	69	138.4	119	132.8
20	134.4	70	134.4	120	134.4
21	136.8	71	138.4	121	135.2
22	139.2	72	139.2	122	136.8
23	139.2	73	141.6	123	134.4
24	140.0	74	134.4	124	136.0
25	139.2	75	135.2	125	137.6
26	140.8	76	136.0	126	138.4
27	139.2	77	135.2	127	137.6
28	138.4	78	136.0	128	138.4
29	136.0	79	132.8	129	137.6
30	142.4	80	133.6	130	137.6
31	140.0	81	134.4	131	140.0
32	144.8	82	133.6	132	135.2
33	140.0	83	131.2	133	135.2
34	139.2	84	132.0	134	135.2
35	139.2	85	131.2	135	136.0
36	136.8	86	132.8	136	132.0
37	140.8	87	132.0	137	133.6
38	141.6	88	133.6	138	134.4
39	139.2	89	131.2	139	133.6

TABLA P.9 (continuación)

PERIODO	DEF	PERIODO	DEF	PERIODO	DEF
40	142.4	90	131.2	140	133.6
41	140.8	91	129.6	141	132.8
42	140.0	92	131.2	142	132.0
43	132.0	93	130.4	143	136.0
44	142.4	94	131.2	144	133.6
45	138.4	95	136.0	145	133.6
46	138.4	96	135.2	146	135.2
47	136.8	97	136.8	147	139.2
48	139.2	98	136.8	148	136.8
49	135.2	99	133.6	149	136.0
50	138.4	100	135.2	150	134.4

10. Los datos de la tabla P.10 son el número de accidentes automovilísticos semanales para los años de 1984 y 1985. Determine el modelo apropiado de Box-Jenkins y pronostique los accidentes para la semana 91. ¿Qué tan preciso es su pronóstico?

TABLA P.10

PERIODO	OBSERVACIÓN	PERIODO	OBSERVACIÓN	PERIODO	OBSERVACIÓN
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

11. Utilice la metodología de Box–Jenkins para modelar los datos de ventas al menudeo en EUA, que se presentaron en el caso de estudio 4.1.

CASO DE ESTUDIO 10.1

VENTAS DE RESTAURANTE

Este caso se refiere a los datos y a la situación del caso de estudio de ventas de restaurante del cap. 7.

Jim Price terminó un curso sobre pronósticos y está ansioso de aplicar la metodología de Box–Jenkins a los datos de ventas del restaurante. Estos datos, que se muestran en la tabla 10.9(a), comienzan con la semana que termina el domingo 4 de enero de 1981 y continúan hasta la semana que termina el domingo 26 de diciembre de 1982. La tabla 10.9(b) contiene nuevos datos para la semana que termina el 2 de enero de 1983 hasta la semana que termina el 30 de octubre de 1983.

TABLA 10.9A VENTAS DE RESTAURANTE: DATOS ANTERIORES

TÉRMINO DE LA SEMANA	VENTAS	TÉRMINO DE LA SEMANA	VENTAS	TÉRMINO DE LA SEMANA	VENTAS
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
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 10.9B VENTAS DE RESTAURANTE: DATOS NUEVOS

TÉRMINO DE LA SEMANA	VENTAS	TÉRMINO DE LA SEMANA	VENTAS	TÉRMINO DE LA SEMANA	VENTAS
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 modelo de Box-Jenkins apropiado para utilizar sobre los datos originales?
2. ¿Cuáles son sus pronósticos para las primeras cuatro semanas de enero de 1983?
3. ¿Cómo son estos pronósticos comparados con las ventas reales?
4. ¿Cómo resulta su modelo de Box-Jenkins comparado con los modelos de regresión utilizados en el cap. 7?
5. ¿Utilizaría el mismo modelo de Box-Jenkins si se combinaran los datos nuevos con los anteriores?

CASO DE ESTUDIO 10.2

MR. TUX

John Mosby decidió intentar el método de Box-Jenkins para pronosticar sus ventas mensuales. Él entiende que este procedimiento es mucho más complejo que los métodos más sencillos que ha estado intentando. Sabe también que es posible obtener pronósticos más precisos mediante este avanzado método. Además, tiene acceso al programa de cómputo Sybil/Runner que puede ejecutar el método de Box-Jenkins. y como ya tiene los datos reunidos y almacenados en su computadora, decide probarlo.

John decide utilizar todo su conjunto de datos, que consiste en 96 meses de datos de ventas. En los mensajes introductorios de la pantalla de su computadora, al activar el programa de Box-Jenkins, John lee que el programa puede modelar un elemento estacional en series de tiempo de datos, así como utilizar tanto los promedios móviles como un modelo autorregresivo.

En el caso de estudio 4.3, John ejecutó un correlograma y determinó que sus datos tenían una tendencia. Hizo una primera diferenciación de los datos y encontró estacionalidad. Como la autocorrelación en el periodo retrasado 24 fue significativa, John decidió que los datos de la primera diferencia no eran estacionarios. Para eliminar la tendencia estacional, realizó una diferencia de largo plazo (diferencia la estacionalidad separada por 12 meses). Los coeficientes de autocorrelación se muestran en la fig. 10.29 y los coeficientes de autocorrelación parcial en la fig. 10.30.

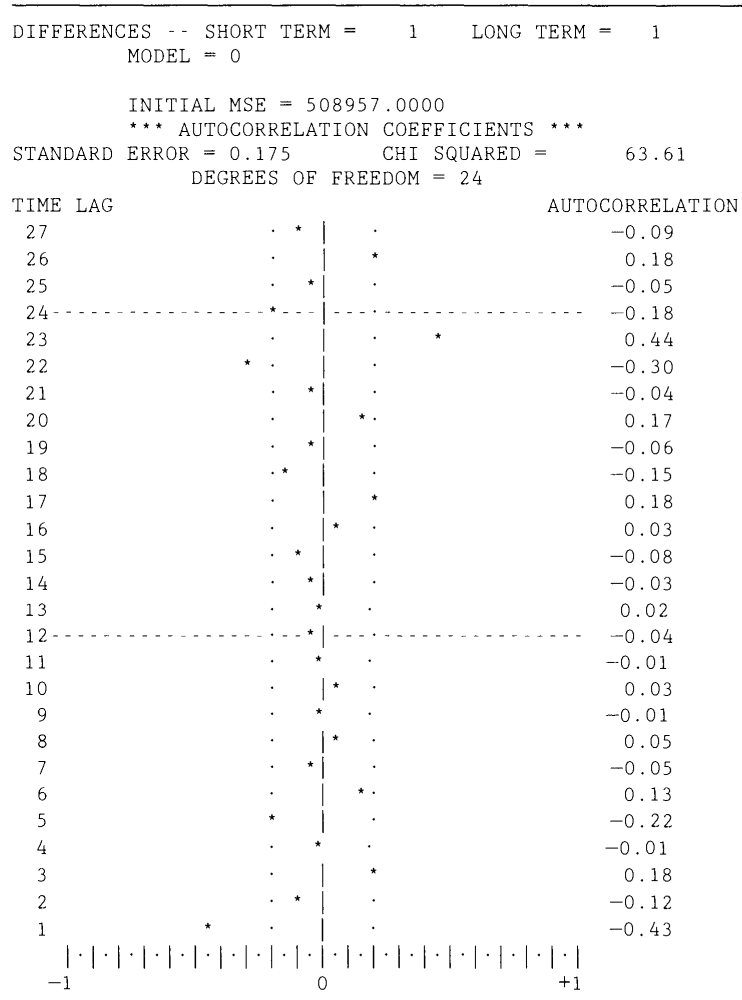


Figura 10.29 Coeficientes de autocorrelación de los datos con diferencia de corto y largo plazos para el caso de estudio 10.2.

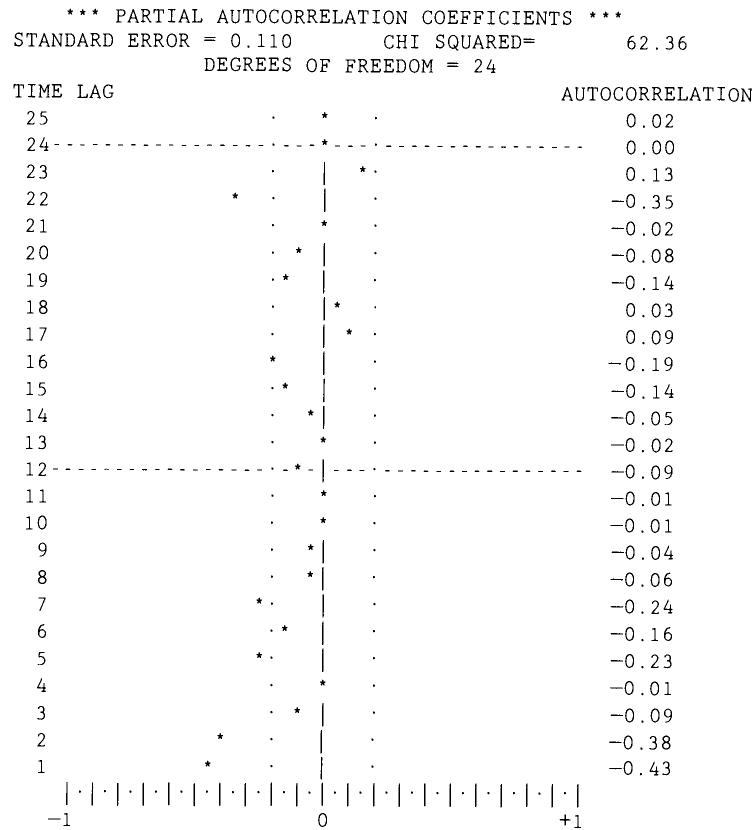


Figura 10.30 Coeficientes de autocorrelación parcial de los datos con diferencias de corto y largo plazos para el caso de estudio 10.2.

TAREA

1. Considerando la complejidad del método de Box-Jenkins, si John encuentra un buen modelo, ¿podrá explicarlo a su banco y al resto de su equipo de administración?
2. ¿Qué modelo(s) emplearía usted para ajustar el término no estacional?
3. ¿Qué modelo(s) emplearía usted para ajustar el término estacional?
4. Si tiene usted un programa de Box-Jenkins, intente encontrar un buen modelo.

CASO DE ESTUDIO 10.3

CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se describió en los caps. 1 (caso de estudio 1.2) y 3 (caso de estudio 3.4).

El director ejecutivo Marv Harinshfeger llegó a la conclusión de que la variable más importante que requería pronosticar CCC era el número de clientes por atender en el resto de 1993. Marv proporcionó 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 caso de estudio 4.3).

Hasta ahora, Dorothy estaba muy contenta con su trabajo de pronósticos. Sin embargo, usted no está por completo satisfecho y decide que es el momento de realmente impresionarla. Usted le informa que debería intentar una de las más poderosas técnicas utilizadas en los pronósticos, los modelos de Box–Jenkins. Dorothy nunca escuchó hablar de Box–Jenkins, pero está dispuesta a dejar que usted lo intente.

TAREA

Escriba un memorando para Dorothy que explique la metodología de Box–Jenkins. Desarrolle un modelo utilizando esta metodología y pronostique el resto de 1993. Escriba un segundo memorando para Dorothy que contenga los resultados de este análisis.

CASO DE ESTUDIO 10.4

LYDIA E. PINKHAM MEDICINE COMPANY

La Lydia E. Pinkham Medicine Company era una empresa familiar cuyo ingreso principal provenía de la venta del Compuesto Vegetal Lydia Pinkham. Quizá los estudiantes podrían usar hoy algo de ese compuesto para aliviar el estrés; desafortunadamente ya no se vende. La etiqueta tenía la imagen de Lydia Pinkham y el compuesto se comercializaba hacia las mujeres. La publicidad del compuesto incluía esta invitación: "... escriba libre y ampliamente a la Sra. Pinkham a Lynn, Mass., y asegure el consejo que ofrece sin cargo a todas las mujeres. Este consejo ha hecho brillar el Sol en muchos hogares a los que el nerviosismo y la irritabilidad casi habían destruido". De hecho, la compañía aseguraba que cada carta era respondida por una empleada. La mujeres *escribieron* a la Sra. Pinkham. Entre los comentarios estaba el siguiente: "Sin el Compuesto vegetal de Lydia Pinkham, a estas alturas ya estaría muerta, o aún peor, loca . . . ya me había dado por vencida; como había intentado tantas cosas, creí que ya nada me haría bien. Pero gracias a su medicina, estoy ahora sana y fuerte; de hecho soy una persona completamente nueva." Este y otros testimonios se reprodujeron en publicidad impresa del compuesto.

La naturaleza única de la compañía (un producto dominante que representaba la mayor parte de las ventas de la empresa, la ausencia de personal de ventas y una gran proporción de las utilidades invertidas en publicidad), así como la disponibilidad de datos tanto en ventas como en publicidad, condujeron en 1943 al Comité de Determinación de Precios del Departamento Nacional de Investigación Económica (NBER National Bureau of Economic Research) a recomendar que se sometieran los datos

Este caso es una contribución de Susan C. White, Northwestern State University, Natchitoches, Louisiana. Para mayor información, consúltese Susan C. White, "Predicting Time Series with Neural Networks versus Statistical Models: The Lydia Pinkham Data", *Proceedings of the 24 th Annual Conference of the Decision Sciences Institute, Southwest Region*, 1993, 108-110.

a un análisis minucioso. Durante varios años no se emprendió la investigación.³ Los analistas habían estudiado los datos mediante modelos causales que incluían los datos de publicidad y otras variables económicas (similares a los presentados en el cap. 7). Sin embargo, varios investigadores sugirieron que los enfoques de Box-Jenkins utilizando sólo los datos de ventas ofrecían predicciones comparables, e incluso superiores, al compararlas con los enfoques causales.⁴ Los datos de ventas eran atractivos de estudiar por dos razones:

1. El producto en sí no había cambiado durante la duración de la serie, es decir, no había cambios en la serie derivados de cambios en el producto.
2. No había cambios en la fuerza de ventas durante la duración de la serie, y la proporción de utilidades gastadas en publicidad se mantenía mas o menos constante. Por ello, no había cambios en los datos generados por promociones especiales u otros fenómenos de comercialización.

Es típico que los datos reales no presenten esta “claridad” en términos del producto y la continuidad en la comercialización.

Entonces, la tarea a realizar consistía en determinar qué modelo de Box-Jenkins sería “mejor” para los datos, mismos que se presentan al final de este caso. El modelo se desarrollaría con los datos de 1907 a 1948 y se probaría con los datos de 1949 a 1960 que se presentan en la tabla 10.10.

k	FAC	FACP
1	0.978	0.978
2	0.949	-0.168
3	0.916	-0.113
4	0.876	-0.138
5	0.836	0.049
6	0.795	-0.071
7	0.757	0.111
8	0.725	0.069
9	0.700	0.135
10	0.682	0.056

IDENTIFICACIÓN DEL MODELO

Se empleó Timeslab⁵ para examinar los datos de 1907 a 1948; los datos de 1949 a 1969 se utilizaron para examinar la capacidad de pronóstico del modelo seleccionado. Las pruebas preliminares sugirieron que los datos eran estacionarios (esto es, no hay una tendencia aparente), así que no se utiliza la diferenciación. Después de examinar las autocorrelaciones y las autocorrelaciones parciales (abajo se muestran los datos de estos coeficientes para 10 periodos), se determinó que un modelo AR era el más apropiado para los datos. Para seleccionar el orden del modelo AR se utilizó el criterio de información de Akaike (AIC). (Esta herramienta excede el alcance de este libro.) Dicho criterio condujo a la elección

³ Kristian S. Palda, *The Measurement of Cumulative Advertising Effects* (Englewood Cliffs, N. J.: Prentice-Hall, 1964).

⁴ Véase P. W. Kyle, “Lydia Pnkham Revisited: A Box-Jenkins Approach”, *Journal of Advertising Research*, 18, 2 (1978), 31-39; y Thomas H. Naylor, Terry G. Seaks y D. W. Wichern, “Box-Jenkins Methods: An Alternative to Econometric Models”, *International Statistical Review*, 40, 2 (1972), 123-137.

⁵ H. Joseph Newton, *TIMESLAB: A Time Series Analysis Laboratory* (Wadsworth & Brook/Cole Publishing Company, 1988).

TABLA 10.10 DATOS DE LYDIA E. PINKHAM MEDICINE COMPANY

AÑO	VENTAS (MILES DE DÓLARES CORRIENTES)	AÑO	VENTAS (MILES DE DÓLARES CORRIENTES)	AÑOS	VENTAS (MILES DE DÓLARES CORRIENTES)
1907	1016	1925	3438	1943	2602
1908	921	1926	2917	1944	2518
1909	934	1927	2359	1945	2637
1910	976	1928	2240	1946	2177
1911	930	1929	2196	1947	1920
1912	1052	1930	2111	1948	1910
1913	1184	1931	1806	1949	1984
1914	1089	1932	1644	1950	1787
1915	1087	1933	1814	1951	1689
1916	1154	1934	1770	1952	1866
1917	1330	1935	1518	1953	1896
1918	1980	1936	1103	1954	1684
1919	2223	1937	1266	1955	1633
1920	2203	1938	1473	1956	1657
1921	2514	1939	1423	1957	1569
1922	2726	1940	1767	1958	1390
1923	3185	1941	2161	1959	1397
1924	3352	1942	2336	1960	1289

de un modelo AR(2) para los datos. Los coeficientes de autocorrelación y de autocorrelación parcial parecen indicar que debería seleccionarse un modelo AR(1) (véase pregunta 2).

ESTIMACIÓN DEL MODELO Y PRUEBA DE SU SUFICIENCIA

Se utilizó Timeslab para estimar los parámetros del modelo AR(2) utilizando los datos de 1907 a 1948. El modelo tentativo de pronóstico fue:

$$\hat{Y}_t = 1.302 Y_{t-1} - 0.309 Y_{t-2}$$

Los términos de error resultantes se probaron utilizando la prueba Q con $k = 10$. El valor calculado de la estadística de la prueba es 2.285 con un valor de p de 0.971; esto indica que el modelo es adecuado. Además, se probaron los coeficientes individuales bajo la hipótesis nula de que el coeficiente verdadero era cero. El valor de p para el primer coeficiente fue de 0.0001; el valor de p para el segundo coeficiente fue de 0.0424. En ambos casos se rechaza la hipótesis nula de que el coeficiente es cero a un nivel de significación del 5%.

PRONÓSTICO MEDIANTE EL MODELO

El último paso en el análisis del modelo AR(2) para estos datos fue pronosticar un periodo a futuro, con los valores de 1949 a 1960. (Por ejemplo, para el pronóstico de 1959 se emplearon los datos hasta 1958.) Los resultados fueron

AÑO	REAL	PRONÓSTICO
1949	1984	1893.2
1950	1787	1992.6
1951	1689	1713.3
1952	1866	1646.6
1953	1896	1907.3
1954	1684	1891.6
1955	1633	1606.4
1956	1657	1605.5
1957	1569	1652.5
1958	1390	1530.5
1959	1387	1324.7
1960	1289	1376.1

Además de los pronósticos, se calcularon algunas mediciones de error tanto para los datos de estimación (1907–1948) como para los datos de prueba (1949–1960). Estos fueron

DATOS	DAM	PEMA	EMC
Estimación	206.3	12.00%	80502.8
Prueba	100.9	6.14%	15337.3

Así los pronósticos del modelo AR(2) tienen una media de desviación absoluta de \$100,900, una media de porcentaje de error absoluto de 6.14% y un error medio cuadrado de 15,337.3. Estas cifras se comparan favorablemente con las mediciones de error de los modelos causales desarrollados por otros investigadores.

RESUMEN Y CONCLUSIONES

Se ajustó un modelo AR(2) parsimonioso (número mínimo de términos o parámetros) a los datos de Lydia Pinkham para los años de 1907 a 1948. Este modelo produjo pronósticos de un paso bastante precisos para los años de 1949 a 1960.

PREGUNTAS

1. Después de concluido este análisis, se tuvieron disponibles los datos de ventas de 1961: \$1,426. ¿Cuál es el pronóstico del modelo para 1961?. Si se agregara este año al conjunto de datos de prueba, ¿cuáles serían los ajustes en el DAM, PEMA y EMC?
2. Ajuste a los datos un modelo AR(1) y compare los resultados con los del modelo AR(2). ¿Cuál de ellos *piensa* que sea mejor? ¿Por qué?
3. Estos datos son atractivos debido a la naturaleza única (no cambiante) del producto y la comercialización para la duración de los datos. ¿Qué factores pudieran afectar los datos de ventas anuales de automóviles y de cobre durante este mismo periodo? ¿Por qué?

Para una lectura adicional sobre la Lydia E. Pinkham Medicine Company y sobre Lydia Pinkham, consúltese Sarah Stage, *Female Complaints: Lydia Pinkham and the Business of Women's Medicine* (New York: Norton, 1979).

CASO DE ESTUDIO 10.5

DEMANDA DIARIA DE PERMISOS PARA PESCA DE TRUCHA

Para conservar los recursos naturales y proporcionar recreación a turistas y a sus ciudadanos, el estado de Missouri tiene funcionando 15 áreas de administración de truchas. El Missouri Department of Conservation cría anualmente aproximadamente 3 millones de truchas. Después de 15 meses estas truchas se liberan en lagos y corrientes. Cuatro de las funciones principales son “parques de truchas”, que tienen corrientes brotantes que se surten diariamente durante la temporada que va del 1 de marzo al 31 de octubre (245 días). Este caso describe el desarrollo y aplicación de un modelo para pronosticar la demanda diaria de truchas en el parque de truchas más grande del estado, Bennett Spring.⁶ El modelo ha sido utilizado por el administrador del criadero para la planeación diaria de liberación de peces. A continuación se exponen los resultados de su uso durante la temporada de 1979.

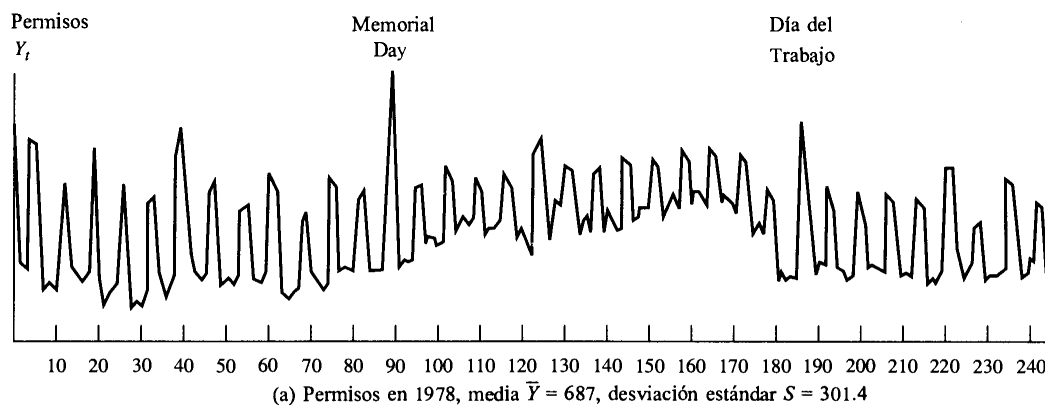


Figura 10.31a Permisos vendidos en 1977 y 1978 por día.

Para pescar en los parques de truchas de Missouri, una persona debe tener una licencia y comprar un permiso diario de pesca de trucha por \$1.50. Este permiso concede al pescador un límite de 5 truchas por día. Cada uno de los cuatro parques de truchas de Missouri tiene un administrador responsable de pronosticar las ventas de permisos de manera que puedan liberarse un número adecuado de truchas. Para aclimatar éstas a la corriente y al shock del manejo y transportación, las truchas se liberan la noche anterior al siguiente día del periodo de pesca. La política del Departamento de Conservación consiste en liberar, en promedio, 2.25 peces por cada permiso vendido. Además, al liberar los peces sólo se tiene disponible el pronóstico del número de permisos vendidos. En consecuencia, la calidad de la pesca en cualquier día depende de la precisión del pronóstico. Si no se liberan suficientes peces, declina la calidad y cantidad de la pesca. Aunque una declinación del número de pescadores pudiera no ser un problema serio, los ingresos del criadero y la economía local dependen mucho del comercio con pescadores y turistas. En contraste, si se liberan demasiados peces, los costos por funcionamiento se harán excesivos.

La venta de permisos vendidos del día se conoce antes de liberar las truchas por la tarde. El administrador del criadero emplea esta información para hacer el pronóstico de un día a futuro. Como

Este caso es una contribución de Stephen A. De Lurgio, University of Missouri, Kansas City.

⁶ Stephen A. De Lurgio, “Forecasting Daily Demand at a State Trout Park Using ARIMA/Box-Jenkins Methodology”, *Proceedings of the American Institute for Decision Sciences Conference -1980*, Las Vegas, Nev., pp. 237-239.

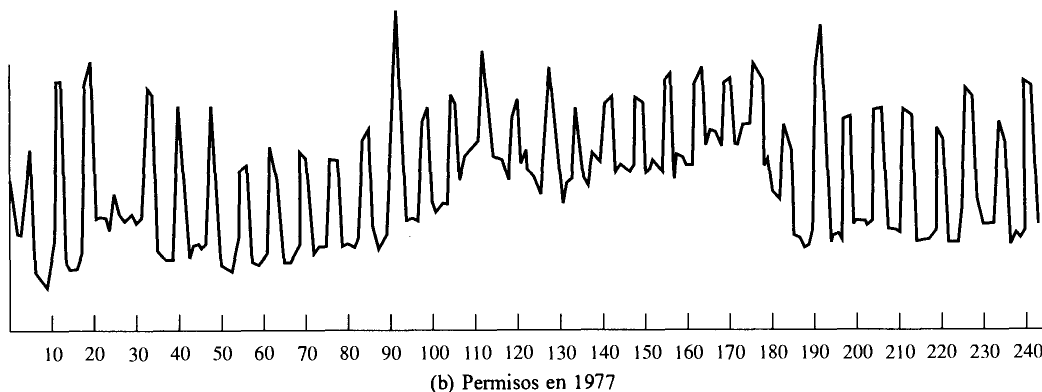


Figura 10.31b Permisos vendidos en 1977 y 1978 por día.

podría esperarse, la venta de los permisos es estacional con respecto tanto al día de la semana como al mes del año. Las figuras 10.31(a) y (b) ilustran la demanda de permisos en 1977 y 1978. Observe el alto grado de asociación entre las dos series. Mediante el uso de estos datos, se desarrolla un modelo de pronóstico.

IDENTIFICACIÓN DEL MODELO

Las gráficas de las figs. 10.31(a) y (b) revelan una estacionalidad diaria lo mismo que una anual [p. ej., las ventas para un día dado de la semana, periodo t , está altamente correlacionado con el del mismo día (no fecha) del año anterior, es decir $t - 364$]. Esta alta correlación entre semana y semana y entre un año y otro sugiere tres posibles enfoques al modelo de la serie: 1) modelos con parámetros estacionales semanales (v. gr., diferenciación de 7 días o índices estacionales), 2) aquellos con diferenciación anual y parámetros (p. ej., diferenciación de 364 días o parámetros) y, (3) alguna combinación de los dos anteriores.

Para modelar el proceso estocástico de una serie histórica, es necesario tener una serie estacionaria en nivel y varianza. Lo estacionario en el nivel se logra a través de diferenciación y lo estacionario en la varianza a través del uso de transformación de potencias (p. ej., logaritmos y exponenciación). Las figs. 10.32(a) y (b) ilustran los valores de autocorrelación y los patrones para la venta de permisos en 1978. Como se puede apreciar, la serie tiene una estacionalidad pronunciada de 7 días; de hecho, la serie no es estacionaria con respecto a los retrasos de 7 días. De ahí que sea necesaria una diferencia de 7 días para lograr una serie estacionaria. Aunque no se ilustra aquí, las autocorrelaciones en los retrasos 364 y 728 confirman que la serie no es estacionaria con respecto a los retrasos de 364 días. Por lo tanto, una diferencia de 364 días es la alternativa para lograr una serie estacionaria. Ya que con tomar una diferenciación ya sea de 7 o de 364 días se logra una serie estacionaria, no es necesario tomar ambas. Es decir, existen dos enfoques alternativos para modelar la serie: modelar el comportamiento estocástico utilizando una diferenciación de 7 días, o modelar con una diferenciación anual, pero no ambas.

Debido a que la serie diferenciada de forma anual arroja el mejor modelo, sólo se desarrollarán aquí los modelos con diferenciación anual. La efectividad de la diferenciación anual es bastante evidente en la asociación de las gráficas (a) y (b) de la fig. 10.31. Esta asociación se confirma por el alto coeficiente de determinación (R^2) de Z .

$$Z_t = Y_t - Y_{t-364}$$

AUTOCORRELATION FUNCTION
 DATA - BENNET SPRINGS DAILY DATA 1978
 DIFFERENCING - ORIGINAL SERIES IS YOUR DATA.
 DIFFERENCES BELOW ARE OF ORDER 1

ORIGINAL SERIES
 MEAN OF THE SERIES =0.63993E+03
 ST. DEV. OF SERIES =0.30138E+03
 NUMBER OF OBSERVATIONS = 245

1- 7	0.56	0.06	-0.04	-0.05	0.01	0.43	0.76
ST.E.	0.06	0.08	0.08	0.08	0.08	0.08	0.09
8- 14	0.42	0.01	-0.05	-0.04	0.03	0.40	0.70
ST.E.	0.11	0.12	0.12	0.12	0.12	0.12	0.13
15- 21	0.38	0.00	-0.05	-0.04	0.01	0.36	0.64
ST.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.15
22- 28	0.33	-0.04	-0.10	-0.10	-0.05	0.31	0.60
ST.E.	0.16	0.16	0.16	0.16	0.16	0.16	0.16
29- 35	0.29	-0.08	-0.13	-0.12	-0.07	0.27	0.56
ST.E.	0.17	0.18	0.18	0.18	0.18	0.18	0.18
36- 42	0.26	-0.10	-0.16	-0.16	-0.12	0.19	0.46
ST.E.	0.19	0.19	0.19	0.19	0.19	0.19	0.19
43- 49	0.19	-0.12	-0.17	-0.16	-0.11	0.18	0.43
ST.E.	0.19	0.19	0.19	0.20	0.20	0.20	0.20
50	0.18						
ST.E.	0.20						

MEAN DIVIDED BY ST. ERROR = 0.33235E+02
 TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE
 0.10935E+04 SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE
 WITH 50 DEGREES OF FREEDOM

DIFFERENCE 1
 MEAN OF THE SERIES = -.96311E+00
 ST. DEV. OF SERIES =0.28340E+03
 NUMBER OF OBSERVATIONS = 244

1- 7	0.07	-0.45	-0.10	-0.09	-0.04	0.09	0.77
ST.E.	0.06	0.06	0.08	0.08	0.08	0.09	0.09
8- 14	0.07	-0.38	-0.07	-0.08	-0.35	0.08	0.70
ST.E.	0.11	0.11	0.12	0.12	0.12	0.12	0.12
15- 21	0.06	-0.36	-0.06	-0.06	-0.33	0.07	0.68
ST.E.	0.14	0.14	0.14	0.14	0.14	0.14	0.14
22- 28	0.06	-0.34	-0.07	-0.06	-0.37	0.08	0.69
ST.E.	0.16	0.16	0.16	0.16	0.16	0.16	0.16
29- 35	0.06	-0.34	-0.07	-0.05	-0.34	0.06	0.67
ST.E.	0.18	0.18	0.18	0.18	0.18	0.18	0.18
36- 42	0.07	-0.33	-0.07	-0.05	-0.31	0.04	0.61
ST.E.	0.19	0.19	0.19	0.19	0.19	0.20	0.20
43- 49	0.06	-0.29	-0.07	-0.04	-0.29	0.05	0.57
ST.E.	0.20	0.20	0.20	0.20	0.20	0.21	0.21
50	0.06						
ST.E.	0.21						

MEAN DIVIDED BY ST. ERROR = 0.53086E-01
 TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE
 0.12280E+04 SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE
 WITH 50 DEGREES OF FREEDOM

Figura 10.32(a) Valores de autocorrelación de la serie original.



Figura 10.32(b) Patrones de autocorrelación de la serie original.

La desviación estándar de las ventas diarias de permisos para 1978 es 301.4, mientras que la desviación estándar de la serie diferenciada anualmente Z_t es 116.7. Por lo tanto, el coeficiente de determinación es

$$\begin{aligned} R^2 &= 1 - \frac{(Y_t - Y_{t-364})^2/(n-1)}{(Y_t - \bar{Y}_t)^2/(n-1)} \\ &= 1 - \frac{85^2}{301.4^2} = 92\% \end{aligned}$$

Al desarrollar modelos potenciales, resultó evidente que por lo menos había dos diferentes series representadas por la estación de 245 días. La fig. 10.33 ilustra la gráfica de las diferencias anuales (Z_t). Como ahí se muestra, la demanda durante los primeros 60 días del año es mucho más variable que más avanzado el año. Esta gran variabilidad es, en parte, el resultado de una más alta variabilidad el clima en los meses de marzo y abril. Las temperaturas son menores y más variables, mientras que la lluvia y la nieve son mayores durante este periodo. Además, en la primavera, acuden relativamente más pescadores vecinos al parque estatal, mientras que en el verano hay más pescadores que son turistas de ciudades y estados cercanos. El número de observaciones extremas durante marzo y abril confunde la identificación del modelo; en consecuencia, se decidió estudiar dos modelos, uno para los 60 días de primavera y otro para los 185 días del verano. A continuación se expondrá el modelo de verano. El modelo para primavera se abordará brevemente más adelante.

La fig. 10.34 muestra las funciones de autocorrelación (FAC) y de autocorrelación parcial (FACP) para la serie diferenciada en forma anual de los meses de verano. Como ahí se muestra, el patrón de autocorrelación confirma que la serie es estacionaria con respecto al nivel (p. ej., las FAC descienden rápidamente). La declinación exponencial de las autocorrelaciones en los retrasos 1 y 5 es indicativa de un proceso autorregresivo de primer orden. Para confirmar esta identificación, considere los patrones de las FACP que aparecen en la fig. 10.35. Las FACP de un modelo autorregresivo de primer orden deberían tener un solo pico en el retraso 1. La fig. 10.35 no muestra este pico en el retraso 1. Ninguno de los otros retrasos tiene una FACP estadísticamente significativa. Por lo tanto, esta identificación parece ser correcta.

Para lograr una serie estacionaria se obtuvo una diferenciación anual de la serie original. Si la media de estas diferencias (Z_t) es estadísticamente diferente de cero, entonces la serie ha experimentado una tendencia; en contraste, si la media de la serie diferenciada no es estadísticamente diferente de cero, entonces la serie ha experimentado un giro estacional. Se realiza una prueba de significación estadística sobre la serie diferenciada en forma anual. Como se muestra en la fig. 10.34, la media de la serie diferenciada anualmente (Z_t) es 47.246 con una desviación estándar de 84.7. Esta media está a 7.54 errores estándar de una media de cero sin tendencia.

$$\begin{aligned} Z &= \frac{47.24 - .0}{\frac{84.7}{\sqrt{183}}} \\ &= 7.54 \end{aligned}$$

En consecuencia, el modelo tentativo que se ha identificado es un modelo autorregresivo de primer orden para una serie diferenciada anualmente que parece poseer una tendencia:

$$\begin{aligned} Z_t &= \phi_1 Z_{t-1} + \phi_0 + \epsilon_t \\ Y_t - Y_{t-364} &= \phi_1 (Y_{t-1} - Y_{t-365}) + \phi_0 + \epsilon_t \end{aligned}$$



Figura 10.33 Datos diferenciados anualmente ($Z_t = Y_t - Y_{t-364}$), Media $Z = 47$, Desviación estándar $S = 85$.

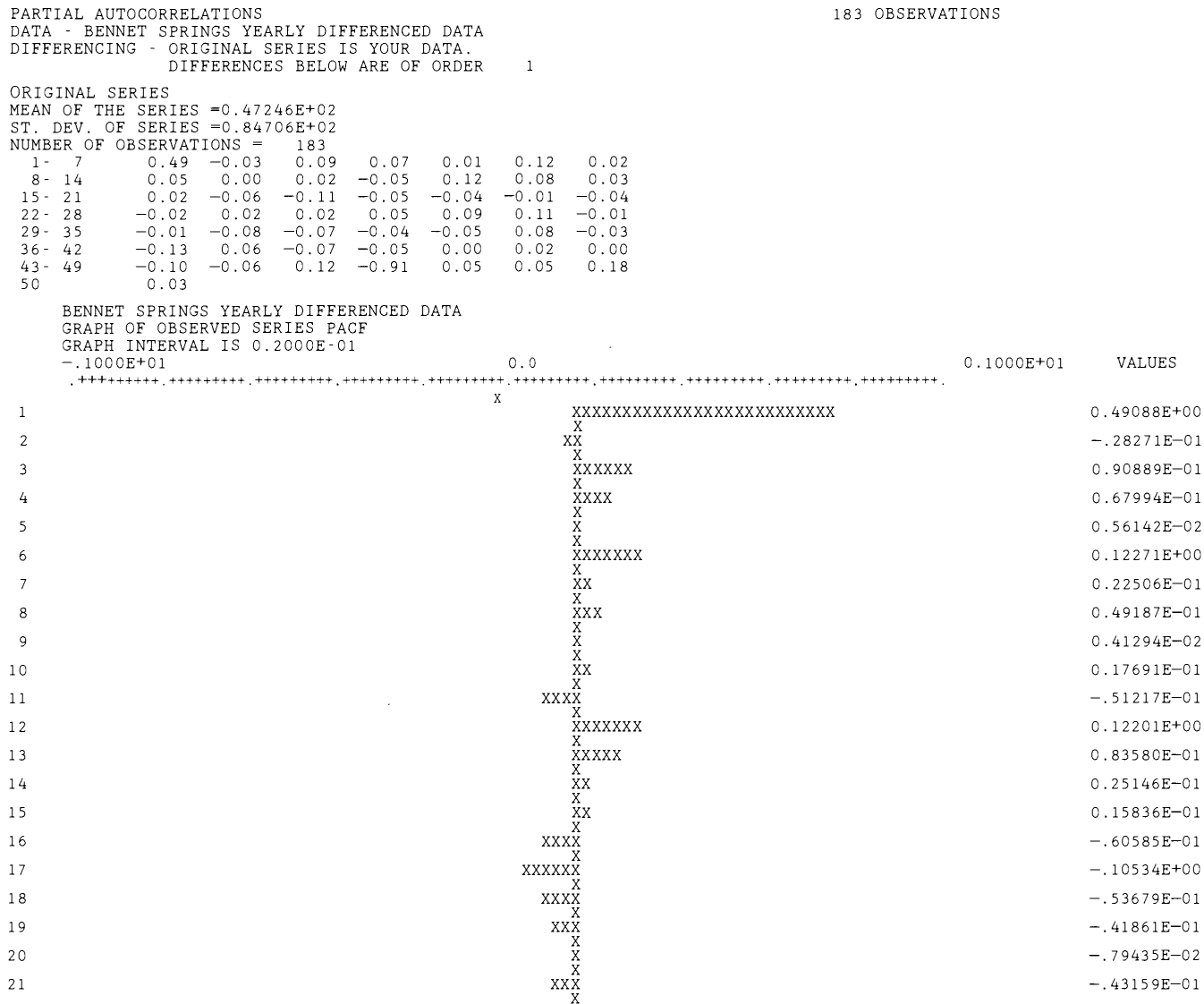


Figura 10.35 Autocorrelaciones parciales de los datos diferenciados anualmente.

en donde

ϕ_1 = el parámetro autorregresivo de primer orden

ϕ_0 = un parámetro de tendencia

ϵ_t = el error en el ajuste del modelo

ESTIMACIÓN

Mediante los programas Pack⁷, se estimaron las estadísticas y el intervalo de confianza del 95% de la tabla 10.11.

El modelo estimado (de aquí en adelante denominado modelo 1) es

$$Y_t - Y_{t-364} = .4969(Y_{t-1} - Y_{t-365}) + 49.7 + \epsilon_t$$

Si se escribe este modelo en forma de pronóstico es atrayente y perspicaz.

$$\hat{Y}_t = Y_{t-364} + .4969(Y_{t-1} - Y_{t-365}) + 49.7$$

La demanda pronosticada en el periodo t es igual a la demanda del mismo día del año anterior más .4969 veces el cambio anual en la demanda experimentado en el periodo $t - 1$ más un incremento anual (tendencia) de 49.7. Este resultado parece muy lógico considerando el tipo de estacionalidad de la serie.

TABLA 10.11 ESTADÍSTICAS Y PARÁMETROS DEL MODELO 1

PARÁMETRO	VALOR	NIVEL DE CONFIANZA DE 95%	
		LIMITE INFERIOR	LIMITE SUPERIOR
ϕ_1	.4969	.3666	.6272
ϕ_0	49.7	26.41	70.17
Error estándar de los residuos $\epsilon_t = 74.22$			
Número de residuos = 182			

DIAGNÓSTICO

El procedimiento de modelado Box–Jenkins ARIMA es iterativo. Después de identificar y estimar un modelo, el analista está obligado a diagnosticar el modelo usando pruebas estadísticas bien definidas. Estas pruebas incluyen 1) una prueba estadística de los parámetros del modelo, 2) el análisis de las FAC y FACP de los residuos, 3) una prueba de la bondad del ajuste de ji cuadrada sobre los coeficientes de autocorrelación y, 4) un análisis de la gráfica de los resultados para mayor confirmación de la aleatoriedad y lo estacionario en nivel y varianza. Si cualquiera de estas pruebas indica un modelo inadecuado, entonces se identifica un nuevo modelo, se estima y se diagnostica. Hagamos estas pruebas sobre el modelo 1.

⁷ David J. Pack, *A computer Program for the Analysis of Time Series Models Using the Box–Jenkins Philosophy* (Hatboro, Pa: Automatic Forecasting Systems, 1978).

Pruebas de diagnóstico

Prueba estadística de los parámetros del sistema. Debido a que los intervalos de confianza de 95% de la tabla 10.11 no incluyen al cero, ambos parámetros de este modelo son estadísticamente diferentes de cero. En consecuencia, el modelo parece ser adecuado con respecto a los parámetros individuales.

Las FAC y FACP de los residuos. Las FAC de los residuos del modelo están dadas en la fig. 10.36. Como se puede ver, las FAC en los retrasos inferiores y superiores son todas insignificantes estadísticamente. (p. ej., han demostrado no ser en forma significativa estadísticamente diferentes de cero). No hay FAC que sean mayores a dos errores estándar del cero. De ahí que la serie parece no tener interferencias. Aunque no se muestran, las FACP confirman esta conclusión.

Por otra parte, debe reconocerse que en una gráfica de 50 FAC de una serie sin interferencias cabría esperar que algunas FAC estuvieran a más de dos errores estándar del cero, sólo por casualidad. Para esta serie, si se presentara un solo pico en los retrasos 1, 2, 3, 7 o 14, uno podría preocuparse debido a que sería una indicación de que los residuos pudieran no estar libres de interferencia y que por ello el modelo resultara inadecuado. No obstante, si se presentaran algunos picos en otros retrasos, podría inferirse que son consecuencia de un error aleatorio. La estructura del modelo y la serie nos guía en la decisión de si unas FAC altas son motivo de preocupación.

AUTOCORRELATION FUNCTION				182 OBSERVATIONS			
DATA - THE ESTIMATED RESIDUALS - MODEL 1							
ORIGINAL SERIES							
MEAN OF THE SERIES =0.43597E-04							
ST. DEV. OF SERIES =0.73815E+02							
NUMBER OF OBSERVATIONS = 182							
1- 7	0.02	-0.07	0.02	0.08	-0.02	0.09	0.06
ST.E.	0.07	0.07	0.07	0.07	0.08	0.08	0.08
8- 14	0.05	0.03	0.08	-0.08	0.04	0.12	0.07
ST.E.	0.08	0.08	0.08	0.08	0.08	0.08	0.08
15- 21	0.05	0.05	-0.05	-0.05	-0.03	0.03	-0.02
ST.E.	0.08	0.08	0.08	0.08	0.08	0.08	0.08
22- 28	-0.02	-0.01	-0.02	-0.01	0.03	0.11	0.05
ST.E.	0.08	0.08	0.08	0.08	0.08	0.08	0.08
29- 35	0.04	-0.03	-0.06	-0.05	-0.08	0.06	0.07
ST.E.	0.08	0.08	0.08	0.08	0.08	0.08	0.08
36- 42	-0.13	0.01	-0.01	-0.06	-0.03	0.05	0.07
ST.E.	0.08	0.08	0.08	0.08	0.08	0.08	0.08
43- 49	-0.06	-0.14	0.02	-0.07	0.01	-0.06	0.11
ST.E.	0.08	0.08	0.09	0.09	0.09	0.09	0.09
50	0.08						
ST.E.	0.09						
MEAN DIVIDED BY ST. ERROR = 0.79678E-05							
TO TEST WHETHER THIS SERIES IS WHITE NOISE, THE VALUE							
0.30010E+02 SHOULD BE COMPARED WITH A CHI-SQUARE VARIABLE							
WITH 43 DEGREES OF FREEDOM							

Figura 10.36 Autocorrelaciones de los residuos del modelo 1.

Prueba de la bondad del ajuste de ji cuadrada. Anteriormente se confirmó que ningún pico individual excede del cero más de dos errores estándar. Sin embargo, se pueden tener varios picos estadísticamente insignificantes que son, en forma individual, congruentes con la suposición de no interferencia, pero que en conjunto estos picos son una fuerte indicación de un comportamiento no aleatorio (con interferencia). Por fortuna, hay disponible la prueba de la bondad del ajuste de ji cuadrada, para probar la significación de un grupo de autocorrelaciones. Como se muestra en la fig. 10.36, el valor de ji cuadrada para las FAC es de 30.00 a 43 grados de libertad. Este valor no es estadísticamente diferente de cero en forma significativa. Esto es, el valor de ji cuadrada para 43 grados de libertad de un modelo sin interferencia tiene una posibilidad entre el 90 y 95% de tener un valor mayor o igual a 30.00; en cuarenta grados de libertad el valor de ji cuadrada tendría que exceder 51.8 antes de alcanzar un nivel de significación de .10. En consecuencia, la serie ha demostrado no ser significativamente diferente de una poca interferencia.

Una gráfica de residuos. Ya que no resulta reveladora para estos datos, no se ilustra una gráfica de los residuos. El análisis visual de los residuos es un paso muy importante en el diagnóstico de cualquier modelo de pronóstico. La gráfica de residuos para este modelo es congruente con la hipótesis de poca interferencia.

PRECISIÓN DEL PRONÓSTICO

La fuerza de los métodos de construcción ARIMA radica en su habilidad para modelar el proceso estocástico fundamental de la serie. Cuando este proceso estocástico permanece estable en el futuro, se pueden obtener pronósticos muy precisos. Exploremos la precisión del modelo así como la comparación de su pronóstico con los de la administración.

En la tabla 10.12 se muestran las precisiones de pronóstico del modelo y de la administración utilizando el pronóstico de un periodo a futuro durante 1979. El primer pronóstico (mayo) se hizo con los parámetros ajustados a los datos de 1977–1978. La media de los residuos ϵ_t y el error estándar de los residuos se obtienen del pronóstico de un periodo a futuro (p. ej., el pronóstico de mayo 13, conociendo la demanda de mayo 12). Después de obtener todos los datos de mayo, se revaloran los parámetros usando una serie diferenciada de mayo–octubre de 1978 hasta mayo de 1979. Entonces se

TABLA 10.12 PRONÓSTICO ADAPTABLE (RODANTE) PARA 1979

MES	PARÁMETROS DEL MODELO		PRONÓSTICO DEL MODELO 1		PRONÓSTICO DEL ADMINISTRADOR	
	α_0	α_1	ϵ_t	RSE	ϵ_t	ESR
Mayo	42.6	.5373	−9.3	89.9	−20.8	81.8
Junio	38.9	.5067	9.8	66.8	−7.5	84.2
Julio	35.7	.5033	−9.1	90.3	−12.7	76.0
Agosto	37.5	.4623	10.0	59.4	−27.7	77.0
		Media	.5	76.6	−17.2	80.0

hacen pronósticos para junio y se repite el procedimiento. Como puede verse en la tabla 10.12, el modelo fue bastante preciso; de hecho, fué preciso en el pronóstico de cuatro meses de la temporada de 1979 ($ESR = 76.6$), como lo fue en el ajuste de los datos de 1978 ($ESR = 74.22$).

Como se muestra en la tabla 10.12, si el administrador hubiera usado el modelo 1 en vez de sus estimaciones subjetivas, su pronóstico habría sido un poco más preciso. Hubiera experimentado un error medio cercano a cero y un menor ESR. Sin embargo, existe una pequeña diferencia entre los pronósticos del administrador y del modelo debido a que el administrador usa los mismos datos que el modelo para llegar a su pronóstico. Esto es, compara el periodo $t - 1$ con el periodo $t - 365$ y ajusta el periodo $t - 364$ para llegar al pronóstico del periodo t . Aunque las diferencias entre los pronósticos del administrador y los efectuados por el modelo no son grandes, la efectividad y parsimonia del modelo son muy evidentes.

IMPLEMENTACIÓN

El modelo 1 es muy sencillo de implementar. Los datos necesarios para pronosticar la demanda se han obtenido rutinariamente por cinco años anteriores a 1979. En la última parte de 1979, el administrador utilizó el pronóstico del modelo 1 junto con sus propias estimaciones para obtener la demanda esperada del día siguiente. Aproximadamente cada cuatro semanas se actualizan los parámetros del modelo mediante programas Pack.

Como se muestra en la tabla 10.12, los parámetros del modelo han variado poco. Si se revaloran los parámetros sólo una vez al año, la precisión del pronóstico sólo se afectaría ligeramente.

RESUMEN Y CONCLUSIONES

El propósito de modelar consiste en proporcionar a un gerente operativo un modelo parsimonioso. Como se expuso aquí, los procedimientos de Box-Jenkins si representan un enfoque efectivo en esta aplicación. A través del uso del modelo, se pueden mejorar la precisión y facilidad con que se elaboran los pronósticos. El administrador ha encontrado que el modelo ARIMA es útil y fácil de implementar. Que estos modelos se desempeñen ligeramente mejor que un administrador experimentado es la evidencia de su efectividad en el mejoramiento de los pronósticos. Es de dudar que otros modelos de pronóstico hubieran convergido tan rápido en un modelo sencillo y preciso como los procedimientos del ARIMA.

APÉNDICE – MODELOS DE INICIO DE PRIMAVERA

Para esta serie y ubicación geográfica, la influencia del clima en la venta de permisos fue mucho mayor en los primeros 60 a 90 días de la serie que en los meses de verano. Con datos de temperatura y precipitación pluvial, se utilizaron varios métodos, incluyendo el ARIMA de variación múltiple y los de regresión múltiple, para pronosticar la demanda de permisos para pesca de trucha durante los meses de marzo, abril y mayo. Aunque se identificaron modelos estadísticamente adecuados, sus estadísticas de error (p. ej., precisión) fueron ya sea inferiores a las del administrador o, si fueron más precisas, eran tan complejas como para considerarlas como alternativas no deseables para el personal operativo. A la fecha, se continúan investigando estos modelos.

PREGUNTAS

1. Dadas las siguientes demandas actuales, pronostique la demanda para el periodo t , utilizando el modelo 1.

$$y_{t-1} = 700$$

$$y_{t-364} = 732$$

$$y_{t-365} = 670$$

2. Exponga las razones por las que los pronósticos del administrador son casi tan precisos que los del modelo durante los meses de verano y por qué su pronóstico pudiera ser mucho más preciso durante la temporada inicial de primavera.
3. Escriba el modelo 1 en forma de pronóstico para pronosticar dos periodos a futuro.
4. La venta de permisos es altamente estacional. Identifique tantas causas fundamentales de esta estacionalidad como le sea posible.
5. En estos datos ¿por qué cree usted que los modelos de diferenciación anual se desempeñen mejor de manera significativa que los modelos de siete periodos (p. ej., aquéllos que modelan una estacionalidad de siete días)?
6. Al calcular 20 FAC para una serie con poca interferencia, ¿cuál es la probabilidad de que una o más FAC estén distantes más de dos errores estándar del valor hipotético de cero?

VERSIÓN ESTUDIANTIL DE FORECAST PLUS

Se accesa el paquete de pronóstico seleccionando la opción 3 del menú principal. Queremos probar el programa Box-Jenkins, opción 13. La pantalla exhibe lo siguiente

```

*** Box-Jenkins Analysis Menu ***

1) Identification
2) Estimation/forecasting
3) Return to Forecasting Menu

```

Nuestra elección es 1, para identificar el modelo apropiado a probar. La pantalla muestra

```

*** Box-Jenkins Analysis: Identification ***

Transformation to the original data series:      7
(1= Reciprocal, 2=1/Sqrt, 3=1/4th root,
4=Log, 5=4th root, 6=Sqrt, 7=none)

Degree of Regular Differencing:                  0

Degree of Seasonal Differencing:                 0

```

```

Length of Seasonal Period:                12

Plot of Autocorrelation Function:         Y
Plot of Partial Autocorrelation Function: Y
Number of lags to print in ACFs:         24

Press ENTER to accept:

```

En el cap. 4 analizamos los coeficientes de autocorrelación para estos datos y se concluyó que había presente una ligera tendencia. Las figs. 10.37 y 10.38 muestran las autocorrelaciones y las autocorrelaciones parciales al hacer una primera diferenciación de los datos (se usa un 1 para una diferenciación regular).

NUMBER OF CLIENTS - BOX-JENKINS ANALYSIS: IDENTIFICATION

AUTOCORRELATION FUNCTION

98 Observations in the Working Series
Mean of the Working Series = .088667
Standard Deviation of the Working Series = 26.69835
Degree of Regular Differencing = 1

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.31	-3.04	:	***[****:]	:
2	-0.08	-0.74	:	[**:]	:
3	-0.09	-0.86	:	[**:]	:
4	0.18	1.57	:	[****:]	:
5	-0.04	-0.38	:	[*:]	:
6	-0.14	-1.24	:	[****:]	:
7	0.04	0.34	:	[*:]	:
8	0.11	0.92	:	[****:]	:
9	-0.11	-0.96	:	[****:]	:
10	-0.08	-0.70	:	[**:]	:
11	0.01	0.06	:	[*:]	:
12	0.16	1.30	:	[****:]	:
13	0.08	0.66	:	[**:]	:
14	-0.11	-0.93	:	[****:]	:
15	-0.15	-1.23	:	[****:]	:
16	0.16	1.30	:	[****:]	:
17	0.08	0.65	:	[**:]	:
18	-0.17	-1.37	:	[****:]	:
19	-0.05	-0.41	:	[*:]	:
20	0.06	0.45	:	[*:]	:
21	0.15	1.14	:	[****:]	:
22	-0.26	-1.95	:	[*****:]	:
23	-0.02	-0.14	:	[*:]	:
24	0.31	2.25	:	[*****]*	:

[] = Estimated Two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 24 Degrees of Freedom = 50.5
Probability - .001

Figura 10.37 Coeficientes de autocorrelación de la primera diferenciación del número de clientes, para el caso de estudio 10.3.

NUMBER OF CLIENTS - BOX-JENKINS ANALYSIS: IDENTIFICATION

PARTIAL AUTOCORRELATION FUNCTION

98 Observations in the Working Series

Mean of the Working Series = .088667

Standard Deviation of the Working Series = 26.69835

Degree of Regular Differencing = 1

Lag	Value	T-Value	-1.0	0.0	+1.0
1	-0.31	-3.04	:	***[****:]	:
2	-0.19	-1.92	:	[****:]	:
3	-0.21	-2.10	:	[****:]	:
4	0.06	0.58	:	[:*]	:
5	0.01	0.05	:	[:]	:
6	-0.15	-0.50	:	[****:]	:
7	-0.05	-0.47	:	[*:]	:
8	0.05	0.51	:	[:*]	:
9	-0.10	-1.00	:	[***:]	:
10	-0.12	-1.21	:	[***:]	:
11	-0.10	-1.03	:	[***:]	:
12	0.04	0.41	:	[:*]	:
13	0.19	1.88	:	[:****]	:
14	0.06	0.58	:	[:*]	:
15	-0.19	-1.89	:	[****:]	:
16	-0.00	-0.03	:	[:]	:
17	0.12	1.18	:	[:****]	:
18	-0.09	-0.86	:	[**:]	:
19	-0.10	-0.97	:	[**:]	:
20	-0.10	-1.03	:	[***:]	:
21	0.09	0.90	:	[**:]	:
22	-0.08	-0.80	:	[**:]	:
23	-0.13	-1.25	:	[****:]	:
24	0.19	1.84	:	[:****]	:

[] = Estimated Two-Standard Error Limits

Figura 10.38 Coeficientes de autocorrelación parcial de la primera diferenciación del número de clientes, para el caso de estudio 10.3.

Como las autocorrelaciones ya sea que caen o descienden después del primer coeficiente de autocorrelación y, el coeficiente de autocorrelación cae a cero, deben funcionar los modelos MA(1) o ARIMA(1,1).

En el menú de análisis de Box-Jenkins se selecciona la opción de estimación/pronóstico (2). La pantalla exhibe

```

*** Box-Jenkins Analysis: Estimation ***

Transformation to the original data series:          7
(1= Reciprocal, 2=1/Sqrt, 3=1/4th root,
4=Log, 5=4th root, 6=Sqrt, 7=none]

Degree of Regular Differencing:                      1

```

```

Degree of Seasonal Differencing:          0
Length of Seasonal Period:               12

Regular Autoregressive Terms:            0
Regular Moving Average Terms:            0
Seasonal Autoregressive Terms:           0
Seasonal Moving Average Terms:           0
Include constant term (Y/N/A):           A

Number of Backforecasts:                 0
Maximum Iterations for Parameter Estimates: 50
Convergence Tolerance:                   .001
Initial estimates for data (A) or user (U): A
Plot of Residual Autocorrelation Function: Y
Number of lags to print in ACF:          24

```

Press ENTER to accept:

El modelo MA(1) es el mejor. Se emplea un 1 para los términos regulares de promedio móvil. La salida se muestra en la fig. 10.39.

Por último, el programa le preguntará si desea almacenar los pronósticos de los residuos

Do you want to save the forecasts of residuals (y/n)?

NUMBER OF CLIENTS - BOX-JENKINS ANALYSIS

PARAMETER ESTIMATES AND STATISTICS

98 Observations Used for Parameter Estimates
Degree of Regular Differencing = 1

Sum of Squared Errors = 59075.64
Residual Variance = 602.8126

<u>Parameter</u>	<u>Value</u>	<u>T-Statistic</u>
MA(1)	0.4678	5.20

NUMBER OF CLIENTS - BOX-JENKINS ANALYSIS

RESIDUAL AUTOCORRELATION FUNCTION

97 Observations in the Residual Series
Mean of the Residual Series = .221538
Standard Deviation of the Residual Series = 24.56271

Lag	Value	T-Value	-1.0	0.0	+1.0
:	:	:	:	:	:
1	0.03	0.25	:	[:*]	:
2	-0.12	-1.16	:	[***:]	:

Figura 10.39 Modelo MA(1) para el número de clientes del caso de estudio 10.3.

3	-0.10	-1.00	:	[***:]
4	0.11	1.09	:	[:***]
5	-0.06	-0.61	:	[**:]
6	-0.18	-1.67	:	[****:]
7	-0.00	-0.01	:	[:]
8	0.05	0.45	:	[:*]
9	-0.14	-1.32	:	[****:]
10	-0.13	-1.17	:	[***:]
11	0.05	0.43	:	[:*]
12	0.24	2.17	:	[:****]
13	0.12	1.05	:	[:***]
14	-0.13	-1.08	:	[***:]
15	-0.16	-1.36	:	[****:]
16	0.13	1.04	:	[:***]
17	0.05	0.37	:	[:*]
18	-0.20	-1.60	:	[*****:]
19	-0.12	-0.92	:	[***:]
20	0.07	0.52	:	[:**]
21	0.09	0.67	:	[:**]
22	-0.23	-1.78	:	[*****:]
23	0.06	0.41	:	[:*]
24	0.35	2.59	:	[:*****]**

[] = Estimated two-Standard Error Limits

Box-Pierce Chi-Square Statistic with 23 Degrees of Freedom = 48
 Probability = .000

Figura 10.39 (Continúa)

PAQUETE DE CÓMPUTO MINITAB

El problema. En el ejemplo 10.4, Jim White, analista de Atron Corporation, desea usar el método de Box-Jenkins para pronosticar las lecturas de un proceso en particular. Jim desea ejecutar los modelos MA(2) y AR(1).

La solución mediante Minitab

```
MTB > READ 'ATRON.DAT' C1
      75 ROWS READ

C1
  60.0    81.0    72.0    78.0    .    .    .

MTB > ACF C1

ACF OF C1

      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +-----+-----+-----+-----+-----+-----+
1  -0.528                                XXXXXXXXXXXXXXX
2   0.282                                XXXXXXXX
```

```

3  -0.038      XX
4   0.008      X
5   0.144     XXXXX
6  -0.137     XXXX
7   0.147     XXXXX
8  -0.036      XX
9   0.068     XXX
10 -0.150     XXXXX
11  0.158     XXXXX
12 -0.189     XXXXXX
13  0.187     XXXXXX
14 -0.234     XXXXXX
15  0.221     XXXXXXXX
16 -0.120     XXXX
17 -0.086     XXX
18  0.105     XXXX

```

MTB > PACF C1

PACF OF C1

```

      -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
      +---+---+---+---+---+---+---+---+---+---+
1  -0.528      XXXXXXXXXXXXXXXX
2   0.003      X
3   0.155      XXXXX
4   0.065      XXX
5   0.189      XXXXXX
6   0.002      X
7   0.026      XX
8   0.060      XXX
9   0.084      XXX
10 -0.184     XXXXXX
11 -0.009      X
12 -0.143     XXXXX
13  0.057      XX
14 -0.135     XXXX
15  0.123     XXXX
16  0.038      XX
17 -0.129     XXXX
18 -0.055     XX

```

Los comandos de MINITAB para probar el modelo MA(2) son

```

MTB > ARIMA(0 0 2)C1;
SUBC> CONSTANT;
SUBC> FORECAST 3 PERIODS AHEAD STORE IN C3. & CONF LIMITS
CONT> IN C4 AND C5
Estimates at each iteration

```



```

Iteration      SSE      Parameters
      7      9786.2      0.567   -0.356   75.410
Relative change in each estimate less than  0.0010

```

Final Estimates of Parameters

Type	Estimate	St. Dev.	t-ratio
MA 1	0.5667	0.1107	5.12
MA 2	-0.3560	0.1146	-3.11
Constant	75.410	1.061	71.08
Mean	75.410	1.061	

```

No. of obs:   75
Residuals:    SS = 9724.97 (backforecasts excluded)
               MS = 135.07  DF = 72

```

Modified Box-Pierce chisquare statistic

Lag	12	24	36	48
Chisquare	7.0(DF=10)	23.8(DF=22)	31.8(DF=34)	46.9(DF=46)

Forecasts from period 75

Period	Forecast	95 Percent 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	

```
MTB > STOP
```

- El comando ARIMA(0 0 2)C1 se usa para ejecutar un modelo MA(2) para los datos almacenados en C1.
- El subcomando CONSTANT se emplea para ejecutar el modelo con un término constante.
- El subcomando FORECAST se utiliza para almacenar tres pronósticos en C3 y colocar los límites de confianza en C4 y C5.
- La salida muestra la estimación final del parámetro constante, 75.41, el coeficiente del término MA(1), .5667 y el coeficiente del término MA(2), -.3560. Las proporciones t para estos parámetros son 71.08, 5.12 y -3.11, respectivamente.

Los comandos de Minitab para probar el modelo AR(1) son

```

MTB > ARIMA(1 0 0)C1;
SUBC> CONSTANT.

Estimates at each iteration
Iteration      SSE      Parameters

```

```

      7      10113.5    -0.538   115.842
Relative change in each estimate less than  0.0010

Final Estimates of Parameters
Type      Estimate      St. Dev.   t-ratio
AR      1      -0.5379      0.0986     -5.46
Constant 115.842      1.356      85.44
Mean      75.3267      0.8816

No. of obs.: 75
Residuals:  SS = 10065.2  (backforecasts excluded)
              MS =   137.9  DF = 73

Modified Box-Pierce chisquare statistic
Lag          12          24          36
Chisquare    9.3(DF=11)  29.9(DF=23)  37.3(DF=35)
48
58.3(DF=47)

Forecasts from period 75
              95 Percent Limits
Period      Forecast      Lower      Upper
76          77.116      54.097     100.135
77          74.364      48.226     100.502
78          75.844      48.871     102.817

```

PAQUETE DE CÓMPUTO SAS

El problema. En el ejemplo 10.4, Jim White, analista de Atron Corporation, desea usar el método de Box-Jenkins para pronosticar las lecturas de un proceso en particular. Jim desea ejecutar los modelos MA(2) y AR(1).

```

SAS Solution

$ SAS/CC=PRINT
OPTIONS PAGESIZE=60 LINESIZE=80;
TITLE "BOX-JENKINS MODELS FOR EXAMPLE 10.4";
DATA WHITE;
  INFILE 'ATRON.DAT';
  INPUT Y :4.1;
PROC ARIMA DATA=WHITE;
  IDENTIFY VAR=Y;
  ESTIMATE P=1;
  FORECAST LEAD=5;
  ESTIMATE Q=2;
  FORECAST LEAD=5;
ENDSAS;

```

- El comando TITLE da nombre a esta ejecución de SAS.
- La declaración DATA asigna un nombre a los datos.
- La declaración INFILE identifica un archivo externo, ATRON.DAT, que se desea leer con una declaración INPUT.
- La declaración INPUT nombra la variable Y e indica que el formato a leer es 4.1.
- El comando PROC ARIMA modela un valor en respuesta a una serie histórica como una combinación lineal de sus propios valores anteriores.
- El subcomando IDENTIFY indica que la variable a analizar es Y .
- El subcomando ESTIMATE solicita una estimación del modelo especificado por P y Q . P especifica los grados de la parte autorregresiva del modelo, y Q da los grados de la parte de promedio móvil del modelo. El primer subcomando ESTIMATE indica un modelo AR(1) y el segundo un modelo MA(2).
- El subcomando FORECAST se emplea para generar valores de pronóstico para una serie histórica utilizando las estimaciones de parámetros producidas por el comando previo ESTIMATE. LEAD especifica el número de valores de pronósticos de paso múltiple a calcular. La salida de cómputo para esta ejecución de SAS se muestra en la fig. 10.40.

```

BOX-JENKINS MODELS FOR EXAMPLE 10.4

ARIMA Procedure Identification of Model

Name of variable = Y.

Mean of working series =      75.24
Standard deviation      = 13.72598
Number of observations =      75

Autocorrelations

Lag  Covariance  Correlation-1  9  8  7  6  5  4  3  2  1  0  1  2  3  4  5  6  7  8  9  1
0    188.402    1.00000  :          :*****:
1   -99.565968  -0.52848  :          :*****:
2    53.049264   0.28157  :          :*****:
3   -7.222704  -0.03834  :          :*****:
4    1.495728   0.00794  :          :*****:
5    27.179760   0.14426  :          :*****:
6   -25.886208  -0.13740  :          :*****:
7    27.643824   0.14673  :          :*****:
8   -6.778944  -0.03598  :          :*****:
9    12.782688   0.06785  :          :*****:
10  -28.340880  -0.15043  :          :*****:
11   29.810352   0.15823  :          :*****:

```

Figura 10.40 Modelos AR(1) y MA(2) para el ejemplo 10.4.

```

12 -35.537616 -0.18863 : . ****. . :
13 35.224416 0.18696 : . : ****. :
14 -44.065152 -0.23389 : . ****. . :
15 41.576880 0.22068 : . : ****. :
16 -22.644288 -0.12019 : . ** : . :
17 -16.208256 -0.08603 : . ** : . :
18 19.846176 0.10534 : . : ** :

```

“.” marks two standard errors

Inverse Autocorrelations

```

Lag Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1 0.40194 : . : ****. :
2 -0.05296 : . * : . :
3 -0.16432 : . *** : . :
4 -0.15553 : . *** : . :
5 -0.17276 : . *** : . :
6 -0.05491 : . * : . :
7 -0.10161 : . ** : . :
8 -0.08906 : . ** : . :
9 0.01222 : . : . :
10 0.14989 : . : *** :
11 0.08808 : . : ** :
12 0.05877 : . : * :
13 -0.01142 : . : . :
14 -0.05127 : . : * :
15 -0.10505 : . : ** :
16 0.07156 : . : * :
17 0.12525 : . : *** :
18 0.03722 : . : * :

```

ARIMA Procedure

Partial Autocorrelations

```

Lag Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
1 -0.52848 : . : ****. :
2 0.00318 : . : . :
3 0.15495 : . : *** :
4 0.06502 : . : * :
5 0.18938 : . : ****. :
6 0.00201 : . : . :
7 0.02572 : . : * :
8 0.06007 : . : * :
9 0.08414 : . : ** :
10 -0.18429 : . : ****. :
11 -0.00943 : . : . :
12 -0.14288 : . : *** :
13 0.05708 : . : * :
14 -0.13511 : . : *** :
15 0.12298 : . : ** :
16 0.03789 : . : * :
17 -0.12886 : . : *** :
18 -0.05518 : . : * :

```

Autocorrelation Check for White Noise

To	Chi	Autocorrelations								
Lag	Square	DF	Prob							
6	31.49	6	0.000	-0.528	0.282	-0.038	0.008	0.144	-0.137	
12	41.36	12	0.000	0.147	-0.036	0.068	-0.150	0.158	-0.189	
18	57.76	18	0.000	0.187	-0.234	0.221	-0.120	-0.086	0.105	

Figura 10.40 (Continúa)

ARIMA Procedure Estimation of AR(1) Parameters

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	75.30138	0.89556	84.08	0
AR1,1	-0.52889	0.09938	-5.32	1

Constant Estimate = 115.12749

Variance Estimate = 139.455214

Std Error Estimate = 11.8091157

AIC = 585.144392*

SBC = 589.779368*

Number of Residuals= 75

* Does not include log determinant.

Correlations of the Estimates

Parameter	MU	AR1,1
MU	1.000	-0.003
AR1,1	-0.003	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	5.38	5	0.372	0.004	0.089	0.147	0.094	0.164	-0.029
12	9.27	11	0.597	0.136	0.094	-0.016	-0.096	0.032	-0.083
18	16.17	17	0.512	0.021	-0.118	0.130	-0.115	-0.161	-0.021
24	29.63	23	0.160	-0.139	-0.059	-0.282	-0.016	-0.137	-0.066

Model for Variable Y

Estimated Mean = 75.3013812

Autoregressive Factors

Factor 1: $1 + 0.52889 B^{**}(1)$

ARIMA Procedure Forecast Using AR(1) Model

Forecasts for variable Y

Obs	Forecast	Std Error	Lower 95%	Upper 95%
76	77.0474	11.8091	53.9020	100.1929
77	74.3779	13.3590	48.1947	100.5612
78	75.7898	13.7614	48.8180	102.7616
79	75.0431	13.8719	47.8547	102.2314
80	75.4380	13.9026	48.1894	102.6866

ARIMA Procedure Estimation of MA(2) Parameters

Figura 10.40 (Continúa)

Conditional Least Squares Estimation

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	75.38035	1.06124	71.03	0
MA1,1	0.55463	0.11157	4.97	1
MA1,2	-0.34081	0.11584	-2.94	2

Constant Estimate = 75.3803455

Variance Estimate = 137.163733

Std Error Estimate = 11.7116921

AIC = 584.867281*

SBC = 591.819745*

Number of Residuals= 75

* Does not include log determinant.

Correlations of the Estimates

Parameter	MU	MA1,1	MA1,2
MU	1.000	-0.002	-0.047
MA1,1	-0.002	1.000	-0.395
MA1,2	-0.047	-0.395	1.000

Autocorrelation Check of Residuals

To	Chi	Autocorrelations							
Lag	Square	DF	Prob						
6	2.60	4	0.627	0.007	0.031	-0.008	0.084	0.152	-0.018
12	7.16	10	0.710	0.146	0.075	-0.003	-0.117	0.043	-0.098
18	12.93	16	0.678	0.041	-0.098	0.136	-0.105	-0.136	0.010
24	23.98	22	0.348	-0.108	-0.026	-0.270	0.005	-0.130	-0.033

Model for variable Y

Estimated Mean = 75.3803455

Moving Average Factors

Factor 1: $1 - 0.55463 B^{**}(1) + 0.34081 B^{**}(2)$

ARIMA Procedure Forecast Using MA(2) Model

Forecasts for variable Y

Obs	Forecast	Std Error	Lower 95%	Upper 95%
76	80.3867	11.7117	57.4322	103.3412
77	77.9901	13.3925	51.7414	104.2389
78	75.3803	13.9746	47.9906	102.7701
79	75.3803	13.9746	47.9906	102.7701
80	75.3803	13.9746	47.9906	102.7701

Figura 10.41 Modelos AR(1) y MA(2) para el ejemplo 10.4.

BIBLIOGRAFÍA

- BOX, G. y JENKINS, G. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1976.
- DE LURGIO, S.A. "Forecasting Daily Demand at a State Trout Park Using ARIMA/Box-Jenkins Methodology." *Proceedings of the American Institute for Decision Sciences Conference, 1980*. Las Vegas, Nev., pp. 237-239.
- JENKINS, G. y MCLEOD, G. *Case Studies in Time Analysis*. Lancaster, UK: Gwilym Jenkins & Partners Ltd., 1982.
- LIBERT, G. "The M-. Competition with a Fully Automatic Box-Jenkins Procedure." *Journal of Forecasting* 3 (3) (1984): 325-328.
- LUSK, E. J. y NEVES, J. S. "A Comparative ARIMA Analysis of the 111 Series of the Makridakis Competition." *Journal of Forecasting* 3 (3) (1984): 329-332.
- MAKRIDAKIS, S. y WHEELWRIGHT, S. C. *Interactive Forecasting*, Palo Alto, Calif.: Scientific Press, 1977.
- MAKRIDAKIS, S. *et al.* "An Interactive Forecasting System." *American Statistician* 28 (1974): 153-158.
- MONTGOMERY, D. S. y JOHNSON, L. A. *Forecasting and Time Series Analysis*, New York: McGraw-Hill, 1976.
- O'DONOVAN, T. M. *Short Term Forecasting: An Introduction to the Box-Jenkins Approach*. New York: John Wiley & Sons, 1983.
- PACK, D.J. "In Defense of ARIMA Modeling". *International Journal of Forecasting* 6 (2) (1990): 211-218.
- QUENOUILLE, M. H. "The Joint Distribution of Serial Correlation Coefficients." *Annals of Mathematical Statistics* 20 (1949): 561-571.

Elementos de juicio en los pronósticos

ACTIVIDAD

Todas las técnicas de pronóstico cubiertas en este libro comprenden la manipulación de datos históricos para producir predicciones o pronósticos de importantes variables de interés. Las exposiciones en los capítulos previos se refirieron a análisis complejos de datos y quizá se implicó que el juicio del pronosticador no estaba comprendido. De hecho, el uso del buen juicio es un componente esencial de toda buena técnica de pronóstico. Se requiere de buen juicio para decidir sobre los datos que son relevantes al problema y para interpretar los resultados del proceso de análisis de datos, y en ocasiones constituye una parte principal del propio análisis.

En el cap. 3, se describieron métodos para identificar y recolectar datos relevantes al pronóstico. En este capítulo se exponen algunos de los elementos importantes de pronóstico que son complementos o sustituciones de la manipulación metódica de dichos datos históricos. Además, se describen diversos elementos relativos a la administración del proceso de pronóstico.

PRONÓSTICOS DE JUICIO

Pronósticos de juicio

En muchas situaciones de pronóstico sólo se emplea el análisis de datos históricos para generar el pronóstico final; no se inyecta en el proceso el juicio u opinión del analista. Este libro se refiere principalmente a estas técnicas de pronóstico y, como resultado, a los pronósticos de corto y mediano plazos. Tales pronósticos son la preocupación esencial de la mayoría de los niveles de administración en una organización y están asociados con la mayor parte de las decisiones que deben efectuarse.

Estos procedimientos de pronóstico dependen de la manipulación de datos históricos y asumen un pasado y un futuro indistinguibles, con excepción de las variables específicas identificadas que afectan la posibilidad de sucesos futuros. Esta suposición excluye un cambio sustantivo en la base tecnológica de la sociedad, una suposición de que la sugerencia

de muchos desarrollos recientes es errónea. Considérese, por ejemplo, la introducción y proliferación de computadoras personales de alta velocidad y bajo costo.

En algunas situaciones de pronóstico, el analista complementa el proceso de análisis de datos después de considerar las circunstancias no usuales de la situación, o después de reconocer que el pasado histórico no es un predictor preciso del futuro. La cantidad de juicio que se inyecta en el proceso tiende a incrementarse en la medida en que los datos históricos sean reducidos en número o se juzgue que son parcialmente irrelevantes. En el caso extremo, pudiera ser la opinión del analista que ninguno de los datos históricos es directamente relevante en el proceso de pronóstico. Bajo estas condiciones, se emplean pronósticos basados exclusivamente en las opiniones de “expertos” para formular el pronóstico o escenario para el futuro.

De manera interesante, la investigación muestra que, cuando hay disponibles datos históricos, la modificación de los pronósticos producida por medios analíticos tiende a reducir la precisión de los pronósticos.¹ Este descubrimiento pudiera atribuirse a alguna desviación por parte del pronosticador, debido quizá a una tendencia a ser demasiado optimista o a subestimar la incertidumbre futura. Se ha demostrado también que el uso de un componente de juicio en el proceso de pronóstico tiende a incrementar el costo del mismo.²

Cuando hay pocos o ningún dato histórico disponible que ayude en el proceso de pronóstico, se debe depender del juicio, si es que se desean pronósticos o predicciones sobre el futuro. Como estas situaciones surgen a menudo, en especial en la alta dirección, se han desarrollado técnicas para mejorar la precisión de dichos pronósticos mediante la utilización del juicio ejecutivo disponible para una mayor ventaja. Vale la pena tomar en consideración estas técnicas, ya que con frecuencia los ejecutivos consideran su propio juicio superior a otros métodos de predicción del futuro. Como establece Makridakis, “La gente prefiere hacer pronósticos de juicio. Creen que su conocimiento del producto, del mercado y del cliente, así como su visión e información interior le proporciona una habilidad única para pronosticar a juicio.”³

A continuación se presentan varias preguntas, cada una de las cuales sugiere mas bien el uso de la imaginación y “actividad cerebral intensa” que la completa dependencia de la recolección y manipulación de datos históricos. Para cada una de estas preguntas, alguna de las técnicas de pronóstico que se abordan en este capítulo pudiera proporcionar valiosos puntos de vista sobre el ambiente de operación futuro de la empresa, al equipo de administración de una compañía.

- ¿Cuál será la distribución de edades en EUA en el año 2025?
- ¿Hasta qué grado trabajarán en su hogar los ciudadanos estadounidenses dentro de 25 años?
- ¿Qué ciudades serán los mayores centros de población y negocios en 20 años?

¹ S. Makridakis, “The Art and Science of Forecasting”, *International Journal of Forecasting*, Vol. 2 (1986), p. 45.

² *Ibid.*

³ *Ibid.*, p. 63.

- ¿Hasta qué grado dependerá EUA de otros países para la manufactura de productos de consumo claves?
- ¿Hasta qué grado serán populares las compras desde el hogar utilizando la televisión o la computadora dentro de 20 años?
- ¿De qué tiempo de ocio dispondrá el estadounidense promedio durante el siguiente siglo?
- ¿Comenzará EUA a retirarse de sus compromisos alrededor del mundo durante los próximos 25 años? De ser así, ¿en qué forma afectará esto a los negocios en EUA?

En ocasiones las técnicas expuestas en este capítulo se les denomina *pronósticos de juicio*, ya que el juicio es el principal o único componente del proceso. A algunas de ellas se les conoce también como métodos de *pronóstico tecnológico* puesto que con frecuencia se relacionan con la proyección de los efectos del cambio tecnológico en el futuro incierto. El resto de este capítulo esboza algunos de los métodos de pronóstico de uso más común en donde el juicio del pronosticador es el ingrediente principal.⁴

Curvas de crecimiento

El método de pronóstico de curva de crecimiento se concentra en los cambios a largo plazo de una variable de interés y proyecta esta variable al futuro sin tomar en cuenta la tecnología que pudiera dar por resultado los cambios predecidos. Por ejemplo, el costo por transacción de computadora se pudiera proyectar al próximo siglo con base en el pasado histórico de esta variable. En este proceso no se considera la tecnología necesaria para producir los cambios proyectados.

En forma similar se podría proyectar el costo por milla del viaje en automóvil, así como el uso del automóvil por la población estadounidense. De nueva cuenta, no se considera la tecnología comprendida en la producción de automóviles ni el combustible necesarios para los cambios predecidos.

Como por lo regular las proyecciones de curvas de crecimiento comprenden pronósticos de largo plazo, generalmente se usan varias relaciones curvas entre el tiempo y la variable de interés. La curva exponencial es a menudo la base de las proyecciones de curvas de crecimiento, ya que esta curva comprende mejoras siempre decrecientes a través del tiempo.

Aunque pudieran esperarse imprecisiones al utilizar las curvas de crecimiento para predecir el futuro, este método puede ser de gran beneficio para la administración ya que concentra su atención en los aspectos de largo plazo de los negocios de una compañía. Es típico que la alta dirección use las curvas de crecimiento a menudo en retiros ejecutivos o en sesiones a puerta cerrada.

⁴ Para una descripción organizada de muchos métodos de pronóstico, incluyendo los de juicio, consúltese David M. Georgoff y Robert G. Murdick, "Manager's Guide to Forecasting." *Harvard Business Review*, Vol. 1 (enero-febrero, 1986), pp. 110-120.

El método Delphi

Cuando en una reunión de expertos se les pregunta sobre el futuro, la dinámica del grupo puede a veces distorsionar el proceso y dar por resultado un consenso que pudiera no haber sido pensado cuidadosamente por todos los participantes. El método Delphi, utilizado primero en un proyecto en la década de 1950 de la RAND Corporation, financiado por la Air Force, pretende eliminar de las deliberaciones de los pronosticadores el aspecto de la dinámica de grupo. En la primera ronda del método, los expertos responden por escrito las preguntas propuestas por el equipo de investigación. Después, este equipo resume los comentarios de los participantes y se les envía por correo. Entonces, los participantes pueden leer las reacciones de los demás y, ya sea, defender sus puntos de vista originales, o modificarlos con base en las opiniones de los demás.

Este proceso continúa durante dos o tres rondas más hasta que los investigadores están satisfechos con los diversos puntos de vista desarrollados y cuidadosamente considerados. Entonces, se puede invitar a los participantes a reunirse para compartir y debatir sus puntos de vista. Al concluir este proceso, el grupo de investigación podría tener un muy buen punto de vista sobre el futuro y comenzar a planear la posición de su organización de acuerdo con éste.

Variantes del método Delphi incluyen la eliminación de la conferencia del grupo al final del proceso escrito. Es probable que fuera en extremo difícil reunir a diversos expertos connotados, de modo que se puede concluir el proceso después de dos o tres rondas por escrito.

Otra variante es el método Delphi en “tiempo–real”, en la que los expertos se encuentran en la misma ubicación, pero separados durante la fase escrita. Cada experto recibe, en pocos minutos, una retroalimentación de los puntos de vista de los otros; de esta manera, se pueden efectuar varias rondas escritas antes de reunir a los expertos para la fase de discusión.

Una variante más comprende el uso de computadoras en tiempo–real conectadas por teléfono. La fase escrita se puede llevar a cabo en un breve periodo aunque los participantes estén en ubicaciones diferentes. Para concluir el proceso se puede emplear una conferencia de seguimiento utilizando las computadoras o el teléfono.

La ventaja del método Delphi consiste en que se puede pedir a connotados expertos que consideren cuidadosamente el tema de interés y replicar concienzudamente a los puntos de vista de los demás sin la interferencia de la dinámica de grupo. El resultado, si el proceso se maneja con cuidado, pudiera ser un buen consenso del futuro junto con las alternativas de diversos escenarios.⁵

Formulación de escenarios

La formulación de escenarios comprende la definición de las particularidades de un futuro incierto mediante la escritura de un “guión” del medio ambiente de una organización para

⁵ Para una descripción detallada del método Delphi, véanse: Frederick J. Parente y Janet K. Anderson-Parente, “Delphi Inquiry Systems”, en *Judgmental Forecasting*, George Wright y Peter Ayton, eds. (New York: John Wiley & Sons, 1987), pp. 129-156.

varios años a futuro. La nueva tecnología, los cambios en la población y la variación en la demanda por parte del consumidor se encuentran entre los factores que se consideran y entrelazan en esta especulación para estimular la imaginación de la alta dirección.

Por lo regular se escribe el escenario más probable junto con uno o más escenarios menos probables, pero posibles. Al considerar la posición de la compañía para cada uno de estos ambientes futuros, la alta dirección está en una mejor posición para reaccionar a los cambios reales del medio empresarial según se vayan presentando y para reconocer las implicaciones de largo plazo de agudos cambios que de otra forma pasarían desapercibidos. De esta manera, la organización está en mejor posición de mantener su rentabilidad a largo plazo en vez de concentrarse en las utilidades de corto plazo e ignorar el medio tecnológico cambiante en el que opera.

El proceso de formulación de escenarios es a menudo seguido por una fase de discusión, a veces por un grupo distinto al que desarrolló el escenario. Se puede emplear la discusión entre los grupos para defender y modificar los puntos de vista de modo que se desarrollen un sólido consenso y escenarios alternativos. Por ejemplo, el equipo de planeación de una compañía podría desarrollar los escenarios y discutirlos con el equipo de la alta dirección. Incluso si no se demuestra después que ninguno de los escenarios es verdadero, este proceso alienta la previsión de largo plazo del equipo de alta dirección y lo prepara mejor para reconocer y reaccionar a los cambios importantes del medio.

COMBINACIÓN DE PRONÓSTICOS

Una rama de desarrollo del estudio de los pronósticos comprende la combinación de dos o más métodos de pronóstico para producir un pronóstico final. Un ejemplar del *International Journal of Forecasting* (Vol. 5, Núm. 4, 1989) contiene una sección especial sobre esta nueva técnica. Algunas partes de los resúmenes de tres artículos en este ejemplar ilustran la naturaleza del desarrollo de pronósticos combinados:

Una investigación sobre 200 estudios demuestra que la combinación de pronósticos produce ganancias consistentes aunque modestas en la precisión. Sin embargo, esta investigación no define bien las condiciones bajo las cuales es más efectiva la combinación, ni cuántos métodos deben combinarse en cada situación.⁶

El volumen de la investigación sobre combinación de pronósticos es sustancial. No obstante, se sabe poco relativamente sobre cuándo y cómo los administradores combinan pronósticos. Elementos gerenciales importantes que requieren un mayor estudio incluyen el ajuste gerencial de pronósticos cuantitativos, el uso de sistemas expertos en la combinación de pronósticos y el análisis de costos de la combinación de pronósticos.⁷

⁶ J. Scott Armstrong, "Combining Forecasts: The End of the Beginning or the Beginning of the End?" *International Journal of Forecasting*, Vol. 5, Núm. 4 (1989), pp. 585-592.

⁷ Essam Mahoud, "Combining Forecasts: Some Managerial Issues", *International Journal of Forecasting*, Vol. 5, Núm. 4 (1989), pp. 599-600.

A través de los años, se ha acumulado una considerable literatura respecto a la combinación de pronósticos. La primera conclusión de esta línea de investigación es que se puede mejorar sustancialmente la precisión de los pronósticos por medio de la combinación de varios pronósticos individuales . . . Este documento proporciona una revisión y anotación de esa literatura . . .⁸

En los años venideros, es posible que se lleve a cabo una mayor investigación sobre las ventajas de combinar pronósticos, junto con las técnicas para hacerlo. El objetivo de tales combinaciones será el de desarrollar pronósticos precisos cuyo costo sea efectivo.

LOS PRONÓSTICOS Y LAS REDES NEURALES

Los métodos convencionales de pronóstico, como los que se describen en este libro, dependen de datos históricos para desarrollar un modelo y utilizarlo para proyectar a futuro las variables de interés. Estos valores proyectados se convierten en los pronósticos que se utilizan para desarrollar planes para el negocio. En estos modelos, se asume que el futuro será exactamente como el pasado, excepto para aquellas variables específicas reconocidas por el modelo.

En ocasiones, los modelos convencionales hacen suposiciones sobre la forma de la distribución de la población, suposiciones que pudieran o no estar sujetas a verificación. Por ejemplo, los intervalos de estimación que usan un modelo de regresión, asumen que la población fundamental sigue una distribución normal.

El campo de desarrollo de la inteligencia artificial pretende duplicar el proceso del cerebro humano y el sistema nervioso mediante la computadora. Aunque este campo se origina en la biología y en la psicología, está avanzando con rapidez hacia otras áreas, incluyendo los negocios y la economía. Los tres principales impulsos en la inteligencia artificial son el procesamiento del lenguaje, la robótica y las redes neurales artificiales. Este último campo posee las aplicaciones más comerciales, incluyendo a los pronósticos.

En las redes neurales, se programan muchos ejemplos en la computadora, ejemplos que captan el nivel completo de relaciones anteriores entre todas las variables que pudieran afectar los resultados de las variables dependientes. Después, el programa de redes neurales asimila estos ejemplos e intenta desarrollar las relaciones fundamentales “aprendiendo” al avanzar el proceso. Este proceso de aprendizaje, también denominado *entrenamiento*, es análogo al aprendizaje de capacitación del ser humano en el trabajo.

Algunos investigadores de pronósticos han apreciado la similitud entre las técnicas de redes neurales y los métodos convencionales de pronóstico que pretenden encontrar variables que predigan con éxito la variable dependiente. La ventaja teórica de las redes neurales como herramienta de pronóstico es que no se necesita especificar la relación por adelantado, ya que el método comprende el aprendizaje sobre las relaciones utilizando los

⁸ Robert T. Clemen, “Combining Forecasts: A Review and Annotated Bibliography”, *International Journal of Forecasting*, Vol. 5, Núm. 4 (1989), pp. 559-583.

ejemplos dados. Además, las redes neurales no requieren ninguna suposición sobre la distribución fundamental de la población y, a diferencia de muchos métodos convencionales de pronóstico, pueden funcionar con datos incompletos.

A una aplicación de red neural exitosa se le denomina en ocasiones *enlace compatible*. Esto significa que el programa de red neural puede rápida y fácilmente reemplazar el modelo existente, digamos un análisis de regresión, sin perder continuidad. Se puede lograr a veces una mejora en el desempeño, como pronósticos más precisos, con un impacto mínimo en las operaciones actuales. Las redes neurales son en especial valiosas en los casos en que faltan insumos o están altamente correlacionados, o cuando los sistemas no son lineales en extremo.

A continuación se presentan algunos ejemplos del uso exitoso de redes neurales en casos prácticos. Estos ejemplos fueron proporcionados por NeuralWare, Inc., un proveedor comercial de programas de redes neurales (Penn Center West, Building IV, Suite 227, Pittsburgh, PA 15276).

- Una planta de Kodak en Texas redujo sus costos en 3 millones de dólares, manteniendo el rendimiento y calidad del producto. Comenzaron reuniendo datos operativos históricos, que se utilizaron en una red neural para predecir la calidad del producto en función de los diversos parámetros de proceso.
- Uno de los problemas clásicos en la interpretación de señales sísmicas para la exploración petrolera consiste en encontrar la primera evidencia de una onda de choque en los registros de los sensores. Los métodos tradicionales de proceso de señales no han tenido mucho éxito en identificar la primera evidencia: sólo el 30% de un algoritmo. Varias compañías petroleras han utilizado con éxito una red neural para identificar la primera onda, Amoco reporta una tasa del 95%. Esto tuvo un gran impacto en su capacidad para mejorar el proceso de datos sísmicos para el modelo de embalse.
- Un fabricante alemán de motores de ventiladores eléctricos automotrices empleaba inspectores maestros para determinar si un motor terminado era muy ruidoso o presentaba un sonido “curioso”. A una red neural le tomó una semana para aprender de un inspector maestro y actualmente realiza el trabajo de pruebas de ruido. El impacto financiero es un producto uniforme de mayor calidad con pocas quejas por parte de los clientes. Los inspectores maestros fueron relevados de este trabajo de gran presión para concentrarlos en otras tareas de mejoramiento de calidad en la fábrica.
- Desde 1989, una compañía de préstamos ha utilizado una red neural para determinar si los clientes son sujetos de un crédito instantáneo.
- Varias compañías han descubierto que se pueden usar redes neurales para desarrollar indicadores útiles en el comercio. Los resultados indican que los indicadores comerciales de las redes neurales, acoplados con estrategias de comercio apropiadas, pueden arrojar utilidades en forma congruente en el mercado.

- La Air Force está empleando una red neural para pronosticar las fallas de componentes en sus aviones. Se reunieron diversos datos sobre cada avión y se entrenó a la red neural para predecir la posibilidad de fallas específicas en aviones específicos. Los resultados se usan para el inventario de refacciones y el mantenimiento preventivo.
- Se ha desarrollado un producto comercial para predecir los resultados de las carreras de caballos. El sistema requiere que el usuario alimente información específica sobre los caballos en una carrera. Acierta el 80% de las veces que decide las probabilidades de la apuesta al favorito.

Con ejemplos como éstos que actualmente disfrutan de éxito en la práctica, no es difícil imaginar a un sistema de red neural prediciendo con éxito el volumen mensual de ventas, la tasa de abstencionismo de los empleados o la tasa líder de interés. Liberadas de las restricciones de confinamiento por adelantado de relaciones específicas y de la suposición de las distribuciones de la población, estas redes tienen un gran potencial de pronósticos precisos y su uso está ganando una amplia aceptación.

RESUMEN DE PRONÓSTICOS DE JUICIO

El peligro en el uso de la mayoría de los métodos de pronóstico que se exponen en este texto consiste en que comprenden la manipulación de datos históricos para generar el pronóstico. Esta práctica es un procedimiento válido si el pronóstico es para el futuro cercano, pero se hace cada vez más sospechosa al incrementarse el horizonte del pronóstico. Los métodos que se abordan en este capítulo son valiosas adiciones al arsenal del pronosticador cuando se considera la particular preocupación de la alta dirección: la predicción a largo plazo del medio ambiente de la compañía.

Diversos autores han fascinado al público en general con sus especulaciones sobre las tendencias de largo plazo actualmente evidentes en nuestra sociedad y el futuro que auguran. Entre los más provocativos se encuentran

Alvin Toffler, *El shock del futuro* (New York: Random House, 1970).

Alvin Toffler, *La tercera ola* (New York: Bantam Books, 1980).

John Naisbitt, *Megatrends* (New York: Warner Books, 1982).

John Naisbitt y Patricia Aburdene, *Megatrends 2000* (New York:

William Morrow, 1990).

Estos autores poseen una habilidad única para especular sobre el futuro en formas provocativas, sin embargo, las necesidades específicas de pronóstico de un negocio requieren de procedimientos más formales que arrojen proyecciones plausibles de la conformación a largo plazo de su ambiente competitivo. Este capítulo se refiere a algunos de los métodos que pueden emplearse para formular tales proyecciones.

OTROS ELEMENTOS DE JUICIO EN LOS PRONÓSTICOS

Se han desarrollado diversas técnicas para ayudar a quienes toman decisiones en el proceso de evaluación de información sobre el futuro incierto y tomar las mejores decisiones. Con frecuencia, estas técnicas se abordan en libros de texto y artículos periodísticos bajo el título general de “toma de decisiones”, o en ocasiones “toma de decisiones bajo incertidumbre”. Cuando los administradores se enfrentan a la tarea de tomar decisiones de frente a lo incierto, su habilidad de pronosticar se convierte en un elemento crítico en el proceso de toma de decisiones.

Cuando se ha concluido el análisis de datos históricos, quien toma las decisiones debe formular un juicio con respecto a las alteraciones en el curso de acción de la empresa. En otras palabras, el analista debe enlazar los resultados del proceso de pronóstico dentro de los procedimientos de toma de decisiones existentes en la empresa. A continuación se exponen brevemente algunos de los elementos de la teoría de decisiones que con frecuencia son relevantes en esta etapa del proceso.

Quienes toman decisiones usan a menudo el concepto de *valor esperado*, ya sea en forma explícita o implícita. Este concepto comprende el valor promedio que puede tomar una variable numérica aleatoria a lo largo de varias pruebas. En la tabla 11.1 se presenta una variable aleatoria discreta X en una distribución de probabilidad; se muestra cada valor posible que puede asumir X junto con la probabilidad de cada uno.

El *valor esperado* de una variable aleatoria discreta es el valor promedio que asume la variable aleatoria durante un gran número de observaciones.

Nótese que la suma de probabilidades en la tabla 11.1 es 1.00 o 100%, lo que significa que se ha identificado cada valor posible que puede asumir X . Suponga que la tabla 11.1 representa el número de contratos mayores que una empresa firmará durante el siguiente año fiscal. La pregunta que responde el valor esperado es: ¿Cuántos pedidos nuevos pueden

TABLA 11.1 DISTRIBUCIÓN DE PROBABILIDAD DEL VALOR ESPERADO

X	$P(X)$
1	.10
2	.20
3	.25
4	.30
5	.15
	1.00

esperarse en promedio, si es válida la distribución de probabilidad de la tabla 11.1? La ecuación 11.1 se usa para calcular el valor esperado de una distribución de probabilidad como la que se muestra en la tabla 11.1.

$$E(X) = \sum X[P(X)] \quad (11.1)$$

en donde

$E(X)$ = valor esperado

X = valores que puede asumir la variable aleatoria

$P(X)$ = probabilidad de ocurrencia de cada X

Se puede calcular el valor esperado de la tabla 11.1 mediante la ecuación 11.1 de la manera siguiente:

$$\begin{aligned} E(X) &= 1(.10) + 2(.20) + 3(.25) + 4(.30) + 5(.15) \\ &= 3.2 \end{aligned}$$

El valor esperado de la distribución de probabilidad de la tabla 11.1 es 3.2. Si en este ejemplo X representa el número de contratos mayores para el próximo año fiscal, se podría decir que, en promedio, si se repitiera muchas veces esta misma distribución, se firmarían 3.2 nuevos contratos. Nótese que no es posible el valor 3.2 en cualquier año; sólo son posibles los valores enteros (1, 2, 3, 4, 5). No obstante, el valor 3.2 representa el valor *promedio* de la distribución a lo largo de muchos intentos. Es frecuente que quienes toman decisiones se interesen en dichos valores esperados y los utilicen como su mejor expectativa de variables numéricas críticas en la planeación del futuro incierto.

Todos los que toman decisiones utilizan los conceptos de la *teoría de la utilidad* sobre una base diaria, tanto para decisiones personales como de la organización. Sin embargo, pocos se dan cuenta de que la teoría de la utilidad se ha estudiado formalmente y que su uso puede mejorar el proceso de toma de decisiones. La teoría de la utilidad reconoce formalmente las cantidades en dólares que no pueden juzgarse simplemente observando sus valores. Se debe considerar también en la empresa el impacto de diversas cantidades en dólares.

Considérense las tablas 11.1 y 11.2. En la tabla 11.1 se calculó que el valor esperado

TABLA 11.2 DISTRIBUCIÓN DE PROBABILIDAD DE LA TEORÍA DE LA UTILIDAD

X	$P(X)$
-10	.15
-5	.10
0	.25
1	.25
19.8	.25
	1.00

sería 3.2. También el valor esperado de la tabla 11.2 es 3.2, calculado mediante la ecuación 11.1, como sigue:

$$\begin{aligned} E(X) &= -10(.15) + -5(.10) + 0(.25) + 1(.25) + 19.8(.25) \\ &= 3.2 \end{aligned}$$

Suponga ahora que los valores de X tanto en la tabla 11.1 como en la 11.2 representan las posibles utilidades netas de una compañía en millones de dólares. El equipo de administración de la compañía debe decidir si desean embarcarse en el proyecto reflejado por las posibles utilidades de la tabla 11.1, o si prefieren la imagen de utilidades que se refleja en la tabla 11.2. Si sólo se emplea el concepto de valor esperado, ambos proyectos serían idénticos en atractivo, ya que presentan el mismo valor esperado, 3.2 millones de dólares.

Sin embargo, nótese la vasta diferencia entre los dos cuadros en términos de riesgo y retribución. En la tabla 11.1 es imposible incurrir en una pérdida. Incluso en la peor situación posible, habría una utilidad positiva de 1 millón de dólares. Por otra parte, la mayor utilidad posible es sólo de 5 millones de dólares. En la tabla 11.2 hay distintas posibilidades de perder una gran cantidad de dinero. Por otra parte, es posible una gran utilidad de 19.8 millones de dólares. Una compañía conservadora podría elegir la inversión que se refleja en la tabla 11.1, mientras que una compañía agresiva escogería la imagen de utilidad/pérdida de la tabla 11.2. Los procedimientos de la teoría de decisiones permiten a quienes las toman reflejar formalmente sus actitudes hacia diferentes cantidades de dinero en el proceso de toma de decisiones.

La *teoría de decisión* se refiere formalmente a los elementos que comprenden la función de toma de decisiones de los líderes empresariales. A menudo, los valores esperados y la teoría de decisiones son entrelazadas dentro de esta más general consideración de toma de decisiones. También, se emplea el *diagrama de árbol* para ayudar a quienes toman decisiones a visualizar una situación compleja y a tomar decisiones racionales. En la fig. 11.1 se muestra uno de estos diagramas de árbol.

La fig. 11.1 refleja la incertidumbre que existe sobre la naturaleza de las ventas futuras e incorpora la decisión entre construir una nueva planta o reparar la vieja. El problema consiste en que si la compañía supiera que habría una demanda alta, sería mejor construir la nueva planta; por otra parte, si se supiera que habría una baja demanda, las utilidades serían mayores reparando la planta vieja. Incluso en este sencillo ejemplo, se puede apreciar el beneficio del diagrama de árbol: permite a quienes toman decisiones observar las diversas opciones disponibles, identificar aquellas incertidumbres que están más allá del control de la empresa y determinar de manera explícita los costos, utilidades y probabilidades de eventos futuros. En situaciones más complicadas, son evidentes los beneficios del diagrama de árbol y de la teoría formal de decisiones.

El *teorema de Bayes* es un concepto estadístico diseñado para revisar las probabilidades preliminares con base en evidencia de muestra. Este concepto es a menudo aplicable en situaciones en las que se determina la estimación de probabilidades de eventos futuros desconocidos y luego se modifican después de reunir evidencia de muestra. Un ejemplo sería el concepto de mercado de prueba utilizado por muchos fabricantes de

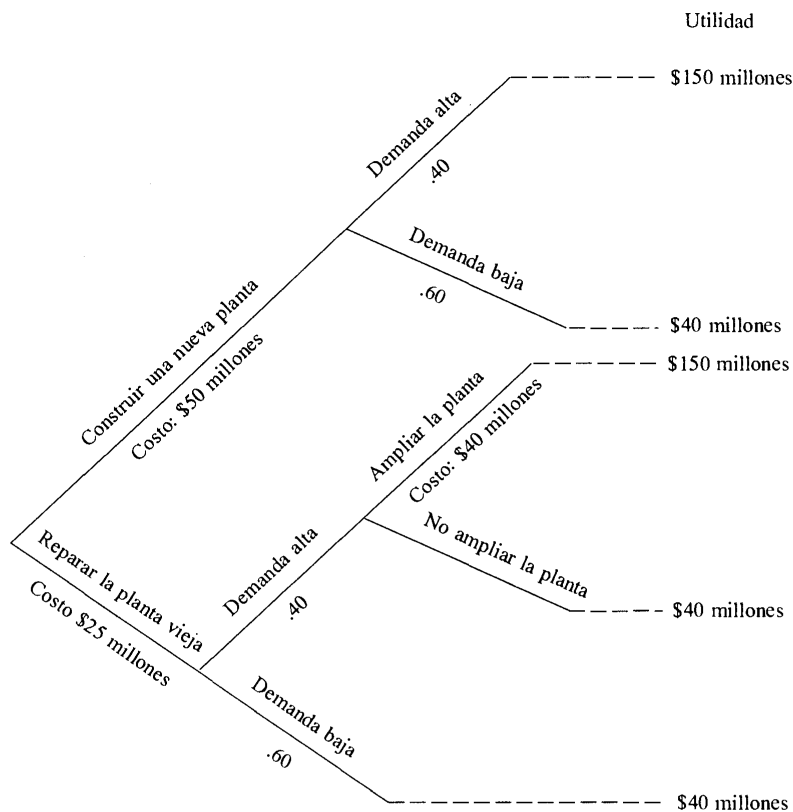


Figura 11.1 Diagrama de árbol.

productos de consumo. Una compañía como éstas podría estimar como muy alta la probabilidad de la aceptación del público de un nuevo producto. Sin embargo, antes de arriesgar los millones de dólares que requiere una campaña nacional, se debe emprender un mercado de prueba en áreas que la compañía reconozca como buenos mercados representativos. Entonces, los resultados de esta prueba de mercado se emplean para modificar las estimaciones originales del éxito del producto y se toma una decisión sobre la introducción del producto a nivel nacional.

La fig. 11.2 muestra una aplicación del teorema de Bayes en una situación de mercado de prueba. La administración de una gran compañía de productos de consumo necesita decidir si introduce un nuevo producto a nivel nacional. Estiman que su nuevo producto tiene un 50% de probabilidad de ventas altas en el mercado nacional. Están considerando utilizar un mercado de prueba para determinar si pueden pronosticar mejor las ventas altas o bajas del nuevo producto. La fig. 11.2 presenta un árbol de decisión de los resultados del mercado de prueba.

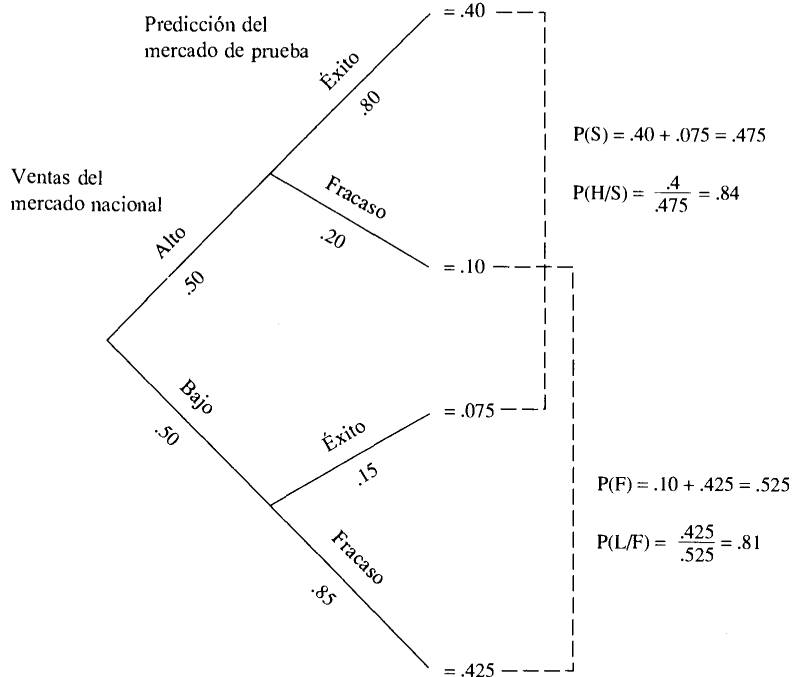


Figura 11.2 Decisión de Bayes para la introducción de un nuevo producto.

La experiencia anterior muestra que cuando se introdujo un nuevo producto y se alcanzaron ventas altas, el mercado de prueba predijo con éxito estas ventas altas el 80% de las veces. También la experiencia anterior muestra que cuando se introdujo un nuevo producto y las ventas fueron bajas, el mercado de prueba predijo con éxito las ventas bajas el 85% de las veces. Se puede utilizar el teorema de Bayes para calcular la probabilidad de que un nuevo producto tenga éxito en el mercado nacional bajo diferentes condiciones del mercado de prueba. La fig. 11.2 muestra el cálculo de dos probabilidades de particular interés.⁹

- Si la prueba de mercado predice el éxito, hay un 84% de probabilidades de que el producto tenga ventas altas a nivel nacional.
- Si la prueba de mercado predice el fracaso, hay un 81% de probabilidades de que el producto tenga ventas bajas a nivel nacional.

En este ejemplo, una prueba de mercado ayudaría a la administración a decidir si se introduce o no el producto a nivel nacional. La precisión del mercado de prueba es suficiente

⁹ La propuesta general del teorema de Bayes para n eventos $A_1, A_2, \dots, A_n$ es

$$P(A_1 | B) = \frac{P(A_1)P(B | A_1)}{P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + \dots + P(A_n)P(B | A_n)}$$

para modificar la probabilidad de ventas altas del valor previo a la prueba de 50% ya sea a 84% (éxito) o a 81% (fracaso); estas probabilidades se calcularon usando el teorema de Bayes. Será mucho más fácil tomar la decisión sobre la introducción del producto de lo que hubiera sido sin la prueba de mercado, ya que la probabilidad de éxito cambiará de manera significativa del valor previo a la prueba del 50%.

En la ley de Bayes, se incorpora el procedimiento estadístico formal para modificar las probabilidades previas con base en nueva evidencia. Se puede encontrar una exposición de este procedimiento en libros de estadísticas administrativas, inferencia estadística y toma de decisiones estructurada.¹⁰

ADMINISTRACIÓN DEL PROCESO DE PRONÓSTICO

En el cap. 1 se presentó el tema de los pronósticos con sugerencias sobre su utilidad dentro de la organización moderna. La necesidad de pronosticar los requerimientos de las organizaciones modernas de formular decisiones oportunas al enfrentar la incertidumbre. Este proceso de realizar adivinaciones sustentadas sobre la incertidumbre futura (pronósticos) comprende un proceso racional de extender al futuro los datos históricos y experiencias. Ahora, con el beneficio de los capítulos intermedios que cubren numerosas técnicas de pronóstico, es posible revisar, mediante estas técnicas, las nociones básicas del cap. 1, para ilustrar los puntos importantes.

El proceso de pronóstico

Un punto clave que se abordó en el cap. 1 es la importancia del uso del buen juicio administrativo junto con técnicas cuantitativas para producir buenos pronósticos. El buen juicio administrativo es de hecho importante y se puede ahora ilustrar mediante las diversas técnicas de pronóstico expuestas en los capítulos previos. Si los pronósticos han de ser precisos, comprendidos y utilizados por quienes toman decisiones en la empresa, siempre es necesaria una mezcla juiciosa de técnicas cuantitativas y sentido común.

El análisis de series de tiempo (descomposición) es un buen ejemplo de una técnica que comprende la necesidad del uso del buen juicio junto con el análisis de la historia previa. Una compañía podría pronosticar la historia anterior de ventas unitarias mediante un programa de análisis de series de tiempo que utilizara datos mensuales y arrojará un análisis de tendencia, de variación cíclica, de variación estacional y movimientos irregulares. Un recomposición de estos cuatro factores en meses futuros comprendería un juicio considerable con respecto a los elementos cíclicos e irregulares y, si se formularan estos elementos, se produciría un pronóstico utilizable de ventas unitarias. Al

¹⁰ Véase, por ejemplo, John E. Hanke y Arthur G. Reitsch, *Understanding Business Statistics*, 2a. ed. (Homewood, Ill.: Richard D. Irwin, Inc., 1994) cap. 18.

utilizar el método de descomposición es importante decidir cuántos periodos históricos incluir en el análisis.

Siempre que se lleve a cabo un pronóstico, el análisis de regresión comprende la necesidad del uso del juicio junto con las técnicas estadísticas. Si se conduce una regresión múltiple utilizando como variable dependiente la tasa de desempeño en el puesto de los empleados y como variables de predicción el examen inicial de aptitud y la edad, se podría obtener un valor de R cuadrada de 75%. Además los valores t de las variables de predicción podrían ser ambos significativos junto con el valor de regresión F . Entonces, el pronosticador se ve tentado a medir las dos variables de predicción para cada aspirante al puesto, para predecir el desempeño en el puesto. Sin embargo, es necesario considerar tres preguntas adicionales. Primero, ¿es suficiente la explicación del 75% de la varianza para fines de pronóstico? Quizá el juicio intuitivo sobre la contratación deseable de una persona sea un método superior, o quizá se requiera de mayor precisión en el proceso de pronóstico y debieran sondearse otras variables de predicción. Segundo, ¿se puede asumir que futuros aspirantes son idénticos en esencia a los muestreados en el estudio de regresión? Si difieren en cualquier forma sustancial, el modelo de pronóstico no será válido. Por último, ¿se justifica el costo del proceso de pronóstico en términos del beneficio recibido? La prueba de la empresa podría ser costosa, en especial si se adquiere de una agencia externa de pruebas y debe justificarse mediante los beneficios del pronóstico.

Las técnicas de Box-Jenkins expuestas en este texto ilustran un problema común de pronóstico mencionado en el cap. 1. Estos procedimientos son a menudo métodos superiores de pronóstico, que producen menores errores de pronóstico en muchas situaciones complejas. Su desventaja consiste en que se requiere de cierta capacidad por parte del usuario. Si el proceso que genera el pronóstico es un total misterio para quien toma las decisiones, el pronóstico pudiera ser descartado en la administración de la organización sin importar qué tan preciso sea.

Las pruebas de chi cuadrada expuestas en este libro ilustran la extensión del análisis de datos dentro de áreas de datos no numéricas. La opinión de la gente es por lo regular el insumo de estas técnicas y permiten a las organizaciones probar las reacciones del público hacia nuevos productos, campañas publicitarias o imagen de la compañía. Existen muchas otras técnicas diseñadas para datos cualitativos, aunque las pruebas de chi cuadrada se usan comúnmente para analizar tales datos, en especial en los programas de tabulación de datos que hay ahora disponibles para las computadoras personales.

La regresión de series de tiempo de datos se presenta con frecuencia en las organizaciones en que se lleva a cabo un registro de medición del desempeño con una base semanal, mensual o trimestral. Como la correlación es un problema común en dichos estudios, es vital una comprensión de esta condición y de su cura si se espera que los resultados de tales análisis sean válidos para el proceso de toma de decisiones. Por desgracia, a menudo falta esa comprensión; este defecto se ha vuelto un problema creciente con el advenimiento de programas de cómputo económicos para análisis de regresión.

Los aspectos de corto, mediano y largo plazos de las técnicas de pronóstico y su relación con los diferentes niveles de administración en una empresa se pueden ilustrar con el análisis de series de tiempo y el pronóstico tecnológico. La primera y segunda líneas de administración de una empresa pudieran interesarse en un análisis de una serie

TABLA 11.3 MODELOS DE PRONÓSTICO

MÉTODO	DESCRIPCIÓN	APLICACIONES	COSTO	¿SE REQUIERE COMPUTADORA?	CAPÍTULO
Modelos casuales de pronóstico					
Análisis de regresión	Pronóstico exploratorio; asume una relación de causa y efecto entre la entrada a un sistema y su salida	Pronóstico de corto y mediano plazos de productos y servicios existentes; estrategias de mercado, producción y sencilla planeación	De bajo a medio	Por lo regular	6
Regresión múltiple	Pronóstico exploratorio; asume una relación de causa y efecto entre más de una entrada a un sistema y su salida	Igual que el anterior	De bajo a medio	Sí	7
Modelos de pronóstico de series históricas					
Método de descomposición	Pronóstico exploratorio; asume una relación de causa y efecto entre el tiempo y la salida de un sistema; el sistema se descompone en sus componentes	Pronóstico de mediano plazo, para planeación de una nueva planta y equipo, financiamiento, desarrollo de nuevos productos y nuevos métodos de ensamble, pronósticos de corto plazo para personal, publicidad, inventarios, financiamiento y planeación de producción	De bajo a medio	Sí	8
Promedios móviles	Para eliminar la aleatoriedad en una serie histórica; pronósticos basados en proyección de datos de series históricas atenuados mediante un promedio móvil	Pronósticos de corto plazo para operaciones como inventarios, calendarización, control, precios y promociones especiales; se usa para calcular los componentes tanto cíclico como estacional para el método de descomposición de corto plazo	Bajo	No	5 y 8
Atenuación exponencial	Similar a los promedios móviles pero con valores ponderados exponencialmente, dando más peso a los datos más recientes	Pronósticos de corto plazo para operaciones como inventarios, calendarización, control, precios y promociones especiales	Bajo	Sí	5
Modelos autorregresivos	Se emplean con variables económicas para considerar las relaciones entre observaciones adyacentes en una serie histórica	Pronósticos de corto y mediano plazos para datos económicos ordenados en una serie histórica; precios, inventarios, producción, acciones y ventas	Medio	Sí	9 y 10
Técnicas de Box-Jenkins	No asume ningún patrón particular en los datos históricos de la serie a pronosticar; usa un enfoque iterativo de identificación de un posible modelo útil a partir de tipos generales de modelos	Igual que el anterior	Alto	Sí	10

histórica de las ventas unitarias mensuales reunidas durante los últimos cuatro años. Mediante el uso del juicio con respecto al futuro del componente cíclico de esta serie, se podrían pronosticar las ventas para el siguiente año fiscal y utilizarse para calendarizar la producción mensual de la fábrica. Los administradores de nivel medio pudieran emplear el mismo programa de serie histórica para analizar los datos de ventas unitarias anuales durante los últimos ocho años y pronosticarlas para cinco años a futuro. En este caso, se podría ignorar el componente cíclico en un intento de planear las necesidades de gasto de capital para la fábrica durante este periodo de cinco años. Al mismo tiempo, la alta dirección podría comprometerse en el pronóstico tecnológico utilizando el método Delphi junto con la formulación de escenarios. Su finalidad sería evaluar la posición actual de la compañía en el mercado y buscar los cambios tecnológicos o sociales que amenazaran su nicho de mercado durante los próximos veinte años, u ofrecer oportunidades que no son evidentes en la operación diaria.

Las técnicas de análisis de datos expuestas en este libro se resumen en la tabla 11.3. Este cuadro proporciona descripciones, señala aplicaciones, estima niveles de costo e indica si es necesaria una capacidad de cómputo para implementar cada técnica. Se hace también referencia a los capítulos en que se aborda cada técnica.

MONITOREO DE LOS PRONÓSTICOS

La recolección de datos y la selección de una técnica de pronóstico aceptable son apenas los primeros pasos en un proceso efectivo y dinámico de pronóstico. En este libro se han descrito diversos pasos en el proceso de pronóstico, haciendo énfasis en el aprendizaje de las técnicas comúnmente empleadas en los pronósticos reales. En la fig. 11.3 se resumen los pasos clave del proceso de pronóstico.

La recolección y revisión de datos históricos apropiados se describió previamente en este libro (caps. 3 y 4), junto con las consideraciones en la selección de una técnica o modelo de pronóstico. Como sugiere la fig. 11.3, el siguiente paso consiste por lo regular en pronosticar periodos históricos anteriores cuyos valores reales se conocen. Los errores resultantes se pueden resumir en varias formas, como se expone en el cap. 4 y, este proceso se continúa hasta encontrar una técnica con una proporción aceptable de beneficio–costo. Entonces, se usa el modelo para pronosticar periodos a futuro y los resultados se incorporan al proceso de toma de decisiones de la empresa.

De vez en cuando es necesario hacer una pausa en el proceso de pronóstico y reconsiderar los procedimientos que se están utilizando. Los pasos usuales son los siguientes:

1. Se descartan los valores históricos más antiguos que se están empleando en el proceso de pronóstico y se agregan al banco de datos los valores reales más recientes.
2. Después de esta actualización de datos, se recalculan los parámetros utilizados en el modelo de pronóstico. Por ejemplo, el(los) valor(es) óptimo(s) de la(s) constante(s) de ponderación empleado(s) en la atenuación exponencial pudiera(n) cambiar, tal vez en forma considerable, al agregar valores de datos más recientes. O los coeficientes en un análisis de regresión pudieran cambiar al ajustar diferentes valores de datos en una ecuación.

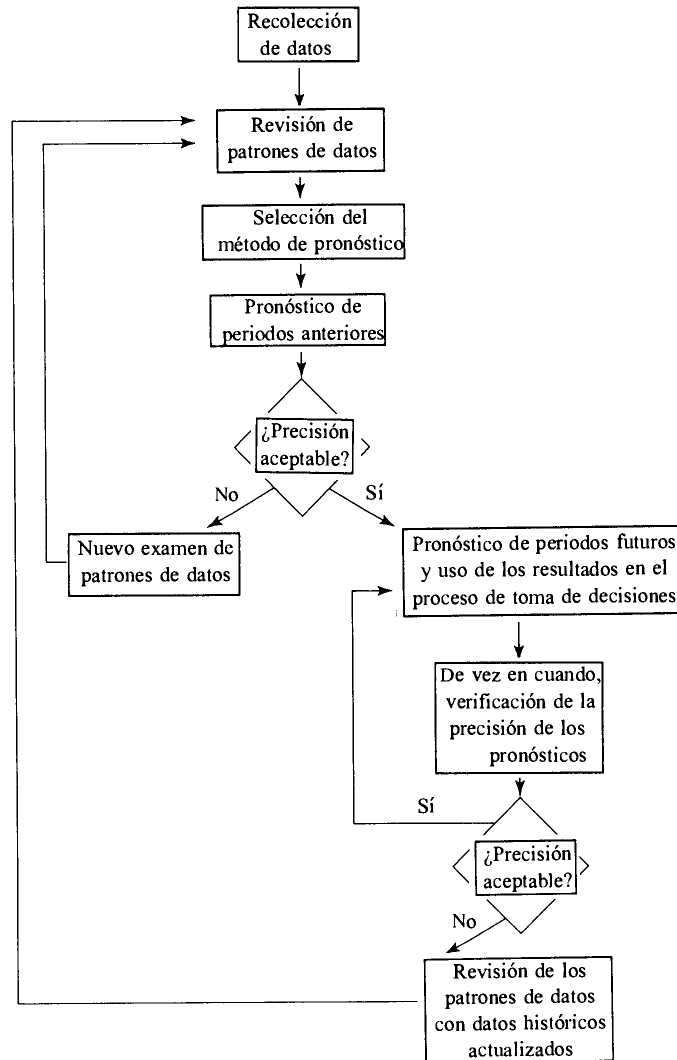


Figura 11.3 Proceso de pronóstico.

3. Se examina el modelo de pronóstico con nuevos parámetros para una adecuada precisión. Si se juzga que esta precisión es suficiente, el moldeo se usa como antes hasta el siguiente periodo de actualización. Si la precisión del pronóstico se juzga inadecuada o marginal, se pueden revisar los patrones en los nuevos datos con la posibilidad de elegir un nuevo procedimiento de pronóstico. Este proceso continúa hasta que se juzga adecuada la precisión del modelo elegido, de acuerdo con la precisión de pronóstico de periodos históricos anteriores.

El proceso anterior se resume en el diagrama de flujo de la fig. 11.3 y constituye el tipo de ciclo de retroalimentación que se encuentra comúnmente en diseños de sistemas de todo tipo.

En ocasiones, se monitorean los pronósticos de manera constante mediante una señal de rastreo, concepto que se abordó en el cap. 5 después del material sobre el método simple de atenuación exponencial. La idea consiste en establecer límites dentro de los cuales se espera que se ubiquen los errores generados por el pronóstico, si el proceso de pronóstico es adecuado. Mientras que los errores se ubiquen dentro de estos límites aceptables, el proceso de pronóstico continúa. Tan pronto como un error cae fuera del margen aceptable, la atención de la administración se enfoca en el proceso de pronóstico y se retoman los pasos de actualización y revisión antes señalados. Este concepto se ilustra en el ejemplo 11.1.

Ejemplo 11.1

Sue Bradley es la responsable del proceso de pronóstico de las ventas mensuales en dólares de su compañía. Sue eligió un modelo de pronóstico que presenta una tasa de error aceptable para sus superiores. Específicamente, el error estándar de su proceso de pronóstico es de \$935; es decir, el pronóstico y los valores reales de ventas mensuales en dólares están típicamente separados en \$935.

Sue asume que los errores de pronóstico presentan una distribución normal con una media de cero y una desviación estándar de \$935. Hace esta suposición después de revisar una gráfica de errores anteriores de pronóstico y encuentra que siguen una curva en forma de campana alrededor de cero. Utilizando un nivel de confianza del 95%, establece los siguientes límites dentro de los cuales espera que se ubique el error de pronóstico de cada mes:

$$0 \pm (1.96)(935)$$

$$0 \pm 1833$$

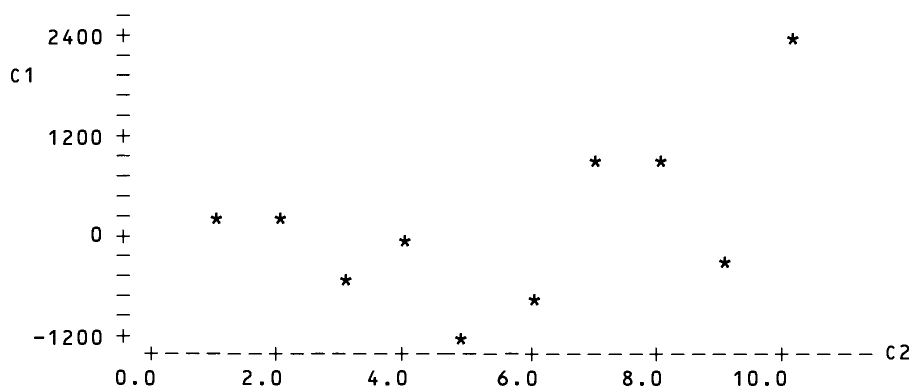


Figura 11.4 Gráfica de errores de datos.

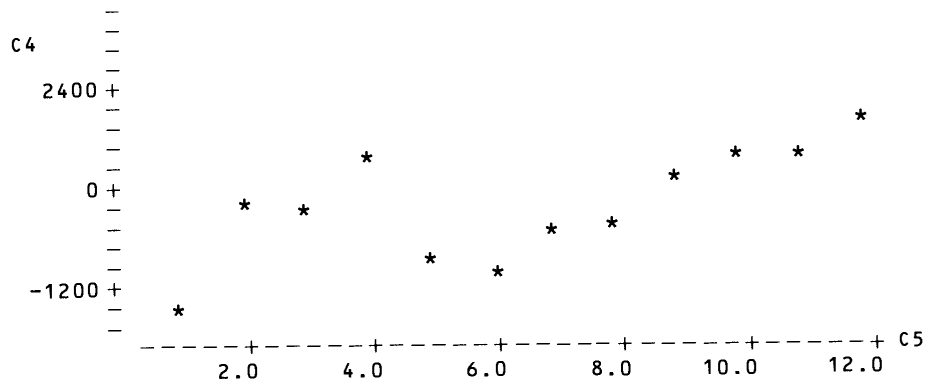


Figura 11.5 Gráfica de errores de datos.

De este modo, Sue espera que cada pronóstico mensual se encuentre dentro de \$1 833 del valor real para el mes, con una confianza del 95 por ciento. Si es así, el procedimiento de pronóstico continuará sin su atención. Pero si el error de pronóstico fuera mayor de \$1 833, ella revisaría ambos parámetros en la técnica de pronóstico elegida e incluso consideraría el uso de otra técnica.

Para monitorear con más facilidad los errores de pronóstico, Sue diseña una gráfica para llevar un registro de los mismos. Durante el transcurso de varios meses, Sue encuentra dos gráficas que hacen que examine de cerca sus procedimientos de pronóstico. La primera, que se muestra en la fig. 11.4, muestra errores de pronóstico que aparentan una distribución normal, hasta el periodo más reciente. Este error fuera de tolerancia conduce a Sue a actualizar los parámetros en su modelo de pronóstico, después de actualizar su base de datos mediante la agregación de valores recientes y la eliminación del mismo número de valores más antiguos. Algún tiempo después, una segunda gráfica de error, fig. 11.5, hace que de nuevo Sue revise su proceso de pronóstico. Aún cuando ninguno de los valores ha excedido sus límites de tolerancia, Sue aprecia que los errores recientes no parecen tener una distribución normal. De hecho, los errores aumentan en una dirección positiva y es obvio que pronto estarán fuera de control. Sue actualiza su base de datos y, después de examinar cuidadosamente los patrones de datos, elige una nueva técnica de pronóstico.

REVISIÓN DE LOS PASOS DE PRONÓSTICO

En el cap. 1 se señaló en varias ocasiones que si se esperaba obtener un pronóstico preciso, debería ejercitarse en forma constante el buen juicio junto con el análisis cuantitativo. Se enlistaron diversas preguntas que deberían surgir si la administración del proceso de pronóstico se conducía de manera adecuada. A continuación repetimos estas preguntas.

- ¿Por qué se requiere de los pronósticos?
- ¿Quién usará el pronóstico y cuáles son sus requerimientos específicos?

- ¿Qué nivel de detalle o agregación se requiere y cuál es el horizonte de tiempo apropiado?
- ¿Qué datos hay disponibles y serán éstos suficientes para generar el pronóstico requerido?
- ¿Cuál será el costo del pronóstico?
- ¿Qué tan preciso se puede esperar que sea el pronóstico?
- ¿Se hará el pronóstico a tiempo para ayudar en el proceso de toma de decisiones?
- ¿Comprende el pronosticador con claridad cómo se utilizará el pronóstico en la organización?
- ¿Hay disponible un proceso de retroalimentación para evaluar el pronóstico después de efectuado y para evaluar de acuerdo con ello el proceso de pronóstico?

El proceso real de pronóstico puede comenzar cuando se formulan y responden adecuadamente las preguntas anteriores. En el cap. 1 se listaron los pasos a seguir en dicho proceso.

Recolección de datos

Reducción y condensación de datos

Construcción del modelo

Extrapolación del modelo (el pronóstico real)

A esta lista de pasos pudiera agregarse otra: retroalimentación una vez que el proceso de pronóstico está en marcha, para determinar si se obtuvo la precisión suficiente y si la administración encuentra útil y de costo efectivo el pronóstico en el proceso de toma de decisiones.

Aunque el tema principal de este libro ha sido la construcción de modelos o la selección de un método apropiado de pronóstico, los autores esperan haber enfatizado también la importancia del proceso de pronóstico. Las preguntas anteriores son importantes en todas las situaciones de pronóstico y se deben considerar si se espera obtener resultados útiles. Nosotros recomendamos especialmente el estudio de los casos al final de este capítulo, ya que están diseñados para enfatizar los aspectos de juicio de los pronósticos.

LA RESPONSABILIDAD EN LOS PRONÓSTICOS

La ubicación del proceso de pronóstico dentro de una empresa varía dependiendo del tamaño de ésta, la importancia conferida a los pronósticos formales y la naturaleza del estilo de administración de la empresa. La responsabilidad de los pronósticos recae en alguna parte de la continuidad entre un departamento de pronóstico por separado y los pronósticos dentro de pequeñas unidades administrativas sin relación a otros esfuerzos dentro de la empresa.

Los equipos de pronóstico son más comunes en las grandes organizaciones que en las pequeñas. Las grandes empresas pueden sufragar la contratación de los expertos requeridos para pronósticos complejos y pueden equipar a su personal con una moderna capacidad de cómputo y programas. La ventaja de este esfuerzo centralizado consiste en que esta opinión experta está disponible para todas las unidades de la organización. La desventaja es que es muy difícil lograr la coordinación entre el personal de pronósticos y los administradores de línea. El personal de pronósticos se pudiera encontrar ocupando más tiempo en negociar con los usuarios y explicar su función que realmente pronosticando.

En el otro extremo se encuentra la ubicación del proceso de pronóstico dentro de cada unidad de la empresa sin haber coordinación o cooperación entre las unidades. La ventaja de este proceso es que no hay malentendidos entre quienes pronostican y quienes usan los pronósticos: son las mismas personas. Los pronósticos generados en estas condiciones tienden a ser aceptados y utilizados en el proceso de toma de decisiones. La desventaja es que pudiera ser difícil lograr la complejidad y por ende la precisión de los pronósticos, debido a que los recursos y programas de cómputo no pueden repartirse entre muchos usuarios. Por lo regular es difícil persuadir a la alta dirección para adquirir equipo y programas cuando ellos saben que existen en otros puntos de la compañía.

Muchas organizaciones intentan ubicar la responsabilidad de los pronósticos en un punto medio entre los extremos antes mencionados. Se podría asignar un pequeño grupo de pronosticadores a subunidades dentro de la empresa para atender las necesidades de diversas áreas funcionales. La tarea de dicho grupo comprendería la adecuada coordinación entre usuarios así como la generación de pronósticos precisos. En ocasiones esta responsabilidad de los pronósticos se combina con otras funciones como las de apoyo estadístico o computacional.

El advenimiento de pequeñas computadoras económicas y programas de cómputo de pronóstico ha generado una tendencia a desplazar hacia abajo, en las organizaciones, la función de pronóstico. Hoy en día es posible que los administradores tengan acceso a herramientas complejas de pronóstico por una fracción del costo de dicha capacidad hace cinco o diez años. Sin embargo, el conocimiento requerido para usar adecuadamente esta capacidad no se incluye en el paquete de equipo o de programas; la necesidad de comprender el uso adecuado de las técnicas de pronóstico se ha incrementado al desplazarse la capacidad de cómputo de las manos de los "expertos" a las de los usuarios en una organización.

COSTO DE LOS PRONÓSTICOS

Los costos obvios comprendidos en la creación de pronósticos son el equipo y programas de cómputo y el equipo de personal. Pero hay costos adicionales no tan obvios debidos al gasto de la compañía en tiempo de personal así como en dinero. Se deben considerar como costos del pronóstico el tiempo empleado por personal asalariado en la recolección de datos para el proceso de pronóstico, el monitoreo del proceso y la interpretación de resultados. Este costo debe sopesarse frente a los beneficios recibidos si es que se van a alcanzar decisiones racionales con respecto a la utilidad de los pronósticos resultantes.

Una alternativa para la producción interna de pronósticos es utilizar consultores para este fin. Esta práctica es en especial atractiva si la necesidad de un pronóstico es un requerimiento de única vez en lugar de uno continuo. Además un requerimiento de pronóstico que rebase la capacidad técnica del personal de una empresa sugiere el empleo de consultores profesionales. Esta contratación externa de asistencia en pronósticos hace que la identificación del costo sea un asunto fácil.

ADMINISTRACIÓN EN LOS PRONÓSTICOS

Si es que los pronósticos generados dentro de una empresa han de convertirse en aspectos importantes del proceso de toma de decisiones, existen varios factores que son consideraciones relevantes. Primero, debe reconocerse que los administradores efectivos están interesados en resultados prácticos y útiles. En general, los pronosticadores deben cubrir los requerimientos de estos administradores; deben proporcionar respuestas a la serie de preguntas planteadas anteriormente en este capítulo.

Segundo, los pronósticos deben ser lo suficientemente precisos para ser útiles. Los buenos administradores no se quedarán con un proceso de pronóstico prolongado, sin importar su complejidad, si no se generan resultados precisos.

Por último, se deben reconocer en los procesos de pronóstico los instintos de beneficio-costos de todo buen administrador. La habilidad de analizar las situaciones en términos de costo incurrido frente a beneficio obtenido es la clave de un proceso administrativo efectivo y debe ser reconocido por el personal de pronósticos. Esta situación crea a menudo dificultades entre el pronosticador y el usuario. El pronosticador debe siempre tener presente que el resultado final del proceso de pronóstico es generar un producto cuyos beneficios al proceso administrativo excedan el costo de generarlo.

LOS PRONÓSTICOS Y EL SISTEMA DE INFORMACIÓN ADMINISTRATIVA

En los últimos años, los sistemas de información administrativa de las empresas modernas han crecido en complejidad y utilidad. Su principal beneficio en el proceso de pronóstico comprende su enorme capacidad para recolectar y registrar datos a través de la organización. El pronosticador debe resistirse a la tentación de reunir datos para los modelos de pronóstico en uso, si dichos datos ya fueron recolectados y registrados en el sistema de información administrativa de la compañía.

Es importante la dependencia en los bancos de datos existentes aun cuando los datos disponibles no estén precisamente en el formato o secuencia de tiempo requeridos por el pronosticador. Antes de desechar los datos ya registrados y recolectar otros nuevos, deben considerarse modificaciones en el modelo de pronóstico o en los datos disponibles. Desde luego, este consejo asume que el proceso de recolección de datos implica un gasto considerable en tiempo. Si los datos necesarios para un modelo de pronóstico son fáciles de obtener en el formato correcto, esto sería preferible a utilizar los datos reunidos anteriormente que no tienen la forma adecuada o no están actualizados.

Una ventaja adicional de utilizar los datos disponibles en el sistema de información administrativa de la empresa consiste en que el proceso de pronóstico se convierte en un componente de este sistema. Como tal, ingresa a la red de distribución y toma de decisiones ya establecida por el sistema y se incorpora con mayor facilidad al proceso de toma de decisiones de la compañía. Este proceso contrasta con un proceso de pronóstico que pretende infiltrarse en los procesos de toma de decisiones que ya emplean los administradores de la compañía.

EL FUTURO DE LOS PRONÓSTICOS

Como se mencionó en el cap. 1, los pronósticos siguen ganando importancia en las organizaciones modernas debido a la creciente complejidad del mundo de los negocios junto con la disponibilidad de equipo y programas de cómputo de menor costo y mayor capacidad. Para cualquiera resulta obvia la continua competencia en las áreas de computadoras pequeñas y programas; menos obvias pudieran ser las tendencias de largo plazo que lentamente cambian el maquillaje de la escena de los negocios y el juicio en las sutiles pero poderosas presiones en las formas de operación de los empresarios.

Considérense algunas de las “megatendencias” identificadas por Naisbitt en su libro con ese título.¹¹ Estas fuerzas tienen una particular importancia para el mundo empresarial y nos hacen pensar en la necesidad de pronosticar utilizando programas y equipo modernos.

Sociedad industrial hacia sociedad de información

Economía nacional hacia economía mundial

Previsión de corto plazo hacia previsión de largo plazo

Centralización hacia descentralización

Opciones de tipo y/o hacia opciones múltiples

Estas tendencias subrayan la importancia de la continuidad del desarrollo administrativo por parte de los empresarios, que deben tratar con los complejos elementos que enfrentan sus empresas durante los próximos años. En particular enfatizan la necesidad de desarrollar métodos cada vez más complejos para enfrentar las incertidumbres de eventos futuros. En otras palabras, hacen énfasis en la importancia creciente de combinar el buen juicio y los métodos sofisticados de manipulación de datos en buenos pronósticos empresariales.

Al comenzar a manifestarse en la escena de los negocios las tendencias de Naisbitt, o cierta combinación de ellas, la atención se enfoca con mayor agudeza en la habilidad de los líderes empresariales para reaccionar rápida y redituablemente a los eventos cambiantes. La pregunta básica en los negocios ¿qué pasará después? adquirirá una mayor importancia; a esta pregunta es a la que se refieren los pronósticos empresariales.

¹¹ John Naisbitt, *Megatrends* (New York: Warner Books, 1982).

GLOSARIO

Valor esperado El valor esperado de una variable aleatoria discreta es el valor promedio que asume la variable aleatoria durante un gran número de observaciones.

FÓRMULA CLAVE

$$\text{Fórmula del valor esperado} \quad E(X) = \sum X[P(X)] \quad (11.1)$$

CASO DE ESTUDIO 11.1

BOUNDARY ELECTRONICS

Boundary Electronics es un gran proveedor de productos electrónicos para uso doméstico. Entre sus ventas mayores están las videocaseteras para el hogar y los sistemas de televisión por satélite. Debido a que los negocios de la compañía crecieron con rapidez, Guy Peterson, presidente de Boundary, está preocupado por un cambio en las condiciones del mercado que pudiera alterar su patrón de ventas.

Al preguntar a sus administradores sobre el futuro de la compañía, Guy descubrió dos cosas. La primera, que la mayoría de sus gerentes están muy ocupados pensando en los problemas cotidianos para cubrir la creciente demanda que no piensan gran cosa en el futuro a largo plazo. Y segunda, las opiniones que sí tienen varían de las muy optimistas a las muy pesimistas. Como presidente de la compañía, Guy siente que tiene la obligación de considerar seriamente el ambiente futuro de su compañía.

Después de meditar el asunto, Guy planea un retiro en sábado con seis miembros de su equipo de altos directivos. Renta un salón de reuniones en un hotel local y hace arreglos para el almuerzo y pausas para café durante el día. Cuando se reúne el grupo el sábado por la mañana hace una presentación del tema del día e instruye después a cada persona para que elabore una descripción de una o dos páginas del ambiente operativo de la compañía durante los próximos 20 años para cada una de las situaciones siguientes:

1. El medio operativo de la compañía continuará en esencia como hasta ahora. Los productos que demandará el mercado serán modificaciones de productos actuales y no intervendrá tecnología nueva.
2. Grandes cambios tecnológicos harán obsoleta la línea actual de productos de la compañía. Habrá que desarrollar nuevos productos para cubrir la demanda en el tiempo libre de la población estadounidense.
3. Entre estos dos extremos, ¿cuál es el escenario más probable para el ambiente operativo de la compañía?

Guy concede una hora a los miembros del equipo para desarrollar los escenarios para cada una de estas tres situaciones. Durante esa hora, Guy piensa en el resto del día y en lo que sucederá. Espera que sus gerentes desarrollen algunas ideas provocativas y que en las discusiones subsecuentes resulten vivas e interesantes. Además de obtener ideas para su propio uso, Guy espera que el ejercicio de ese día ayude

a sus gerentes a ver más allá de los problemas y oportunidades inmediatos de la compañía y les dé una visión de más largo plazo de la misma.

PREGUNTAS

1. ¿Qué proceso cree usted que debe utilizar Guy después de la hora de actividades escritas?
2. Dados sus objetivos, ¿hay algún otro enfoque que pudiera haber utilizado Guy?
3. ¿Cree usted que Guy logrará sus objetivos con esta reunión sabatina?

CASO DE ESTUDIO 11.2

GOLDEN GARDENS RESTAURANT

Sue y Bill Golden decidieron abrir un restaurante en una ciudad del medio oeste. Emplearon más de un año en investigar el área y visitando restaurantes de precio medio a alto. Definitivamente piensan que hay sitio para otro restaurante y encontraron un buen lugar disponible a buen precio.

Además, tienen contactos con varios chefs de primera línea y creen poder atraer a uno de ellos a su nuevo restaurante. Sus investigaciones con los bancos locales los convencieron de que dados sus propios recursos y su experiencia en el negocio de restaurantes, habrá financiamiento disponible con facilidad.

El único punto que aún preocupa a los Golden es la atmósfera o decoración de su restaurante. Realizaron ya entrevistas a tres grupos diferentes de residentes del área que comen fuera con regularidad sin llegar a un consenso en esta materia. Han platicado el asunto entre ellos ampliamente y ahora piensan que algunas otras opiniones serían valiosas.

Después de leer sobre algunas de las técnicas utilizadas en los pronósticos de juicio, creen que alguna de ellas podría ayudarles a decidir sobre la atmósfera para su nuevo restaurante. Han identificado varios de sus amigos y socios que estarían dispuestos a ayudarles pero no están seguros de utilizar sus habilidades.

PREGUNTAS

1. ¿Qué método sugeriría usted a los Golden para utilizar la experiencia de sus amigos en la decisión sobre la atmósfera y decoración de su nuevo restaurante?
2. ¿Existen otros métodos que hayan pasado por alto en su intento de investigar este asunto?

CASO DE ESTUDIO 11.3

BUSBY ASSOCIATES

Jill Tilson era una recién graduada de una escuela universitaria de administración cuando entró a trabajar con Busby Associates, un exportador en grande de equipo agrícola. Durante el proceso de contratación,

el presidente de Busby observó un curso de pronósticos en el currículo de Jill y decidió iniciar el empleo de Jill con un proyecto de pronóstico que había estado discutiendo muchas veces con los altos ejecutivos de la empresa.

El presidente de Busby pensaba que había una fuerte relación entre las ventas de exportación de la empresa y las cifras nacionales de exportación. Las cifras nacionales estaban fácilmente disponibles de fuentes gubernamentales, así que el proyecto de Jill era pronosticar una buena y representativa variable de exportación. Si este esfuerzo tenía éxito, el presidente de Busby creía que tendría una poderosa herramienta para pronosticar sus propias ventas de exportación. Jill localizó la copia más reciente del *Survey of Current Business* en una biblioteca local y registró las cifras trimestrales de exportación de bienes de consumo en miles de millones de dólares. Ella pensó que ésta era una buena variable representativa de las exportaciones nacionales totales. Anticipando la posibilidad de pronosticar mediante el análisis de regresión, registró también los valores de otras cuatro variables que pensó podrían tener una buena correlación con su variable dependiente. Concluyó con valores de cuatro variables para 14 trimestres.

Después calculó tres variables adicionales a partir de los valores de su variable dependiente: el cambio en Y , el porcentaje de cambio en Y y Y retrasada un periodo. Así, al comenzar a pensar en las diversas formas de pronosticar su variable, había reunido los datos que se presentan en la tabla 11.4.

Jill capturó sus datos en un programa de cómputo que efectuaba análisis de regresión y calculó la matriz de correlación para sus siete variables. Después de revisar esta matriz, eligió tres regresiones con una variable de predicción y seis regresiones con dos variables de predicción. Ejecutó después estas regresiones y eligió la que consideró la mejor: utilizaba un predictor (Y retrasada un periodo) con los siguientes resultados:

$$r \text{ cuadrada} = .98389, \quad t = 25.9, \quad F = 671.6, \quad \text{Durbin-Watson} = 2.18$$

La tabla de Durbin-Watson que Jill utilizó comenzaba con $n = 15$ y su tamaño de muestra era sólo de 13. De modo que interpoló los valores críticos de DW y estimó que el límite superior para la prueba de autocorrelación era 1.35. Como su estadística de DW era 2.18, concluyó que no había presente autocorrelación en su regresión. Ella pensó que había encontrado una buena variable de predicción (Y retrasada un periodo).

Jill se dio cuenta de que su tamaño de muestra era muy reducido: 13 trimestres. Volvió al *Survey of Current Business* para reunir más puntos de datos y se desilusionó al encontrar que los años que le interesaban en la definición de su variable dependiente cambiaban, resultando en una serie inconsistente. Esto es, la serie tuvo un medio salto hacia arriba durante el periodo final de la década de 1980 que ella estaba estudiando.

Jill comentó este problema con su jefe y se acordó que se podría emplear como variable dependiente las exportaciones totales de mercancías en vez de las exportaciones de bienes de consumo. Jill encontró que esta variable se mantenía congruente durante varios ejemplares del *Survey of Current Business* y que podía recolectar varios años de datos. Jill reunió los datos que se muestran en la tabla 11.5, retrasó los datos un periodo y ejecutó de nuevo un análisis de regresión utilizando como variable de predicción a Y retrasada un periodo.

En esta ocasión obtuvo nuevamente buenas estadísticas en su reporte de regresión, con excepción de la estadística de Durbin-Watson. Este valor era de .96 y ella concluyó que había presente autocorrelación, ya que este valor está por debajo del valor inferior de la tabla para la prueba de Durbin-Watson. Jill probó otras ejecuciones agregando el número de periodo y el cambio en Y como variables de predicción, pero no pudo encontrar una estadística de Durbin-Watson lo suficientemente

TABLA 11.4 SERIE HISTÓRICA DE DATOS TRIMESTRALES

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, exportaciones, miles de millones de dólares

2: Ahorro personal bruto, miles de millones de dólares

3: Ingreso nacional de comercio al menudeo, miles de millones de dólares

4: Índices de precios de ponderación fija para compras de la defensa nacional, equipo militar, 1982 = 100

5: Cambio en la variable dependiente respecto al periodo previo

6: Porcentaje de cambio en la variable dependiente respecto al periodo previo

7: Variable dependiente retrasada un periodo

Fuente de las variables 1 a 4: *Survey of Current Business*, Vol. 70, Núm. 7 (Julio, 1990), U.S. Department of Commerce.

alta como para concluir que se había eliminado la autocorrelación. Jill decidió entonces buscar otras técnicas para pronosticar su nueva variable independiente: exportaciones totales de mercancías. Utilizó la serie histórica de datos que se presenta en la columna Y de la tabla 11.5.

Entre los programas de cómputo disponibles en Busby estaba Sibyl/Runner, un programa de pronóstico que maneja una variedad de métodos de pronóstico. Jill preparó sus datos para utilizarlos con este programa y comenzó a trabajar a través de varias rutinas de pronóstico. Entre ellas había un procedimiento para graficar los datos sin procesar. En la fig. 11.6 se muestra la gráfica trimestral de exportaciones totales de mercancías para los años de 1984 al segundo trimestre de 1990.

Después de estudiar la fig. 11.6, Jill decidió usar sólo los últimos 16 puntos de datos en su esfuerzo de pronóstico. Su razonamiento fue que comenzando en el periodo 9, la serie había mostrado un

TABLA 11.5 SERIE HISTÓRICA DE DATOS TRIMESTRALES: EXPORTACIONES TOTALES DE MERCANCÍAS (MILES DE MILLONES DE DÓLARES)

	PERIODO	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*, varios ejemplares.

incremento relativamente estable, mientras que antes de ese periodo presentaba crecimiento y declinación. El programa Sibyl/Runner le advirtió sobre el uso de un número reducido de puntos de datos (16), pero ella continuó después de ignorar esta advertencia.

Jill comenzó su análisis llamando la rutina del programa de Sibyl, utilizando sólo los últimos 16 puntos de datos. El primer análisis producido por el programa indicó que había cierto patrón en los datos, pero no era estacionalidad. El programa utilizó diversos métodos de pronóstico para pronosticar los valores de muestra e imprimió una tabla con los porcentajes de error de pronóstico para cada método y cada periodo.

Después de estudiar estos errores de pronóstico, Jill eligió tres técnicas de pronóstico para un mayor análisis. Los tres que escogió tenían el porcentaje más bajo de errores de pronóstico y fueron

Atenuación exponencial simple

Atenuación exponencial de Brown, la cual puede acomodar una tendencia en los datos

Atenuación exponencial cuadrática, que permite una línea curva de tendencia a través de los datos

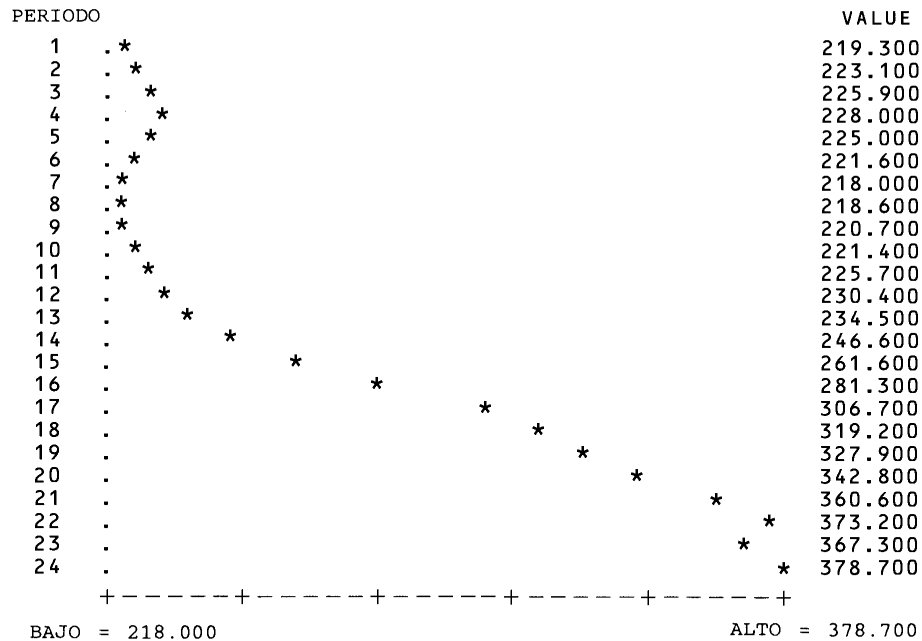


Figura 11.6 Gráfica de valores de datos trimestrales: exportaciones totales de mercancías, del primer trimestre de 1984 al segundo trimestre de 1990.

Después, pidió los pronósticos de sus datos utilizando cada técnica, con los resultados siguientes:

TÉCNICA	ERROR MEDIO CUADRADO	CONSTANTE ÓPTIMA DE ATENUACIÓN
Exponencial simple	185.5	.999
Exponencial de Brown	81.0	.722
Exponencial cuadrática	87.7	.485

Jill advirtió que la constante óptima de atenuación utilizando la atenuación exponencial simple era casi 1.00 (.999). Aparentemente, a fin de rastrear a través de los datos de manera óptima, el programa estaba básicamente utilizando un valor de datos para predecir el siguiente. Esto es equivalente a usar el método no formal simple para pronosticar.

Como la atenuación exponencial de Brown produjo el error mínimo cuadrado más reducido, Jill pensó que debería emplear ese método. Ella revisó más adelante la impresión para este método y encontró que el porcentaje de error medio absoluto (PEMA) era 2.39% y el porcentaje medio de error (PME) o desviación era .34%. Jill consideró satisfactorios estos valores y decidió usar el método de atenuación exponencial de Brown para pronosticar valores futuros.

Utilizando el método que eligió, solicitó pronósticos para los siguientes cuatro periodos después del final de sus datos. Los valores de pronóstico se presentan a continuación:

PERIODO	PRONÓSTICO
17	401.27
18	407.30
19	413.34
20	419.38

Jill se dio cuenta de que al transcurrir cada trimestre habría disponible un nuevo valor real de las exportaciones totales de mercancías y se podrían actualizar los pronósticos de periodos futuros.

Jill se reunió después con su jefe para discutir sus resultados. Ella indicó que pensaba que tenía una buena forma de pronosticar la variable nacional exportaciones totales de mercancías, mediante el uso de atenuación exponencial con ajustes de tendencia. Su jefe le pidió que explicara este método, lo cual hizo. Su siguiente asignación era utilizar los datos reales para verificar la corazonada del presidente de Busby: que las exportaciones de Busby tenían una buena correlación con las exportaciones nacionales. Si ella podía establecer este vínculo, Busby tendría un buen método para pronosticar sus exportaciones y utilizar éstos para planear operaciones futuras.

PREGUNTAS

1. Jill no consideró la combinación de los pronósticos generados por los tres métodos que analizó ¿Cómo le iría si lo hiciera? ¿Cuáles serían las ventajas y desventajas de tal acción?
2. La constante óptima de atenuación utilizada por la atenuación exponencial de Brown fue .722. Al llegar más datos en los siguientes trimestres, Jill probablemente ejecutaría sus datos para ver si este valor cambia. ¿Con qué frecuencia cree usted que deba hacerse esto?
3. Es posible que la elección del método de pronóstico pudiera cambiar a otra técnica al agregar los datos de nuevos trimestres a la base de datos. ¿Debe Jill reejecutar de vez en cuando su análisis por completo para verificar esto? ¿Con qué frecuencia se debe hacer?

CASO DE ESTUDIO 11.4

MR. TUX

El propietario de varias tiendas de renta de ropa Mr. Tux, John Mosby, ha intentado diversas formas de pronosticar su variable más importante, ventas mensuales en dólares. Sus esfuerzos se describen al final de cada capítulo de este libro. Después de concluir estos intentos de pronóstico, John decide que es hora de resumir sus esfuerzos e intentar llegar a un método de pronosticar el futuro. Él está consciente de que debe actualizar tanto sus datos como su método de pronóstico en algún momento futuro, pero por el momento necesita elegir una forma de pronosticar los siguientes meses.

Para comenzar, John resume los resultados de los métodos que ha intentado hasta ahora. A continuación se presentan los resultados de cada método, junto con el capítulo en el que se expusieron sus esfuerzos.

- *Capítulo 2:* Podría funcionar el uso del promedio anual para pronosticar las ventas anuales futuras, pero como John observó una tendencia creciente, necesita de una forma más efectiva de extender estos promedios en el futuro. Además, John está muy preocupado por el efecto estacional, pues

sabe que sus ventas varían considerablemente por mes. Sus esfuerzos mediante promedios anuales no rinden frutos.

- *Capítulo 3:* Junto con su asistente, Virginia Perot, John superó los problemas en la recolección de datos, existentes antes del comienzo de su estudio. Él no pensó que fuera tan difícil obtener los datos históricos de sus ventas mensuales, pero está contento de tener un conjunto completo de datos con los cuales trabajar.
- *Capítulo 4:* Mediante el programa Sibyl/Runner, John estableció que en sus datos existían tanto una tendencia como un efecto estacional. Aunque antes de empezar él sabía que estos elementos estaban ahí, quedó complacido al ver que su programa de cómputo lo establecía en forma estadística. El programa indicó también que varios coeficientes de autocorrelación estaban fuera de los límites de control, indicando a John que se necesitaba reflejar en su modelo final de pronóstico los efectos tanto estacional como de tendencia.
- *Capítulo 5:* Cuando John utilizó la atenuación exponencial, incluyendo métodos que tomaron en cuenta los factores estacional y de tendencia, las mediciones de error resultantes no fueron satisfactorias. John se dio cuenta de que estas mediciones, como el error promedio y el porcentaje promedio de error, eran el resultado de predecir valores anteriores de su variable. Pero como eran demasiado altos, él no quiso emplear estas técnicas para predecir el futuro desconocido.
- *Capítulo 6:* La siguiente técnica que intentó John fue el análisis de regresión simple, utilizando el número de periodo como variable independiente o de predicción. Él razonó que esta variable tomaría en cuenta el factor de tendencia que él sabía tenían sus datos. Con un valor de r cuadrada de 49%, este método demostró no ser satisfactorio.
- *Capítulo 7:* A continuación, John probó la regresión múltiple utilizando tanto el número de periodo para reflejar la tendencia, como un conjunto de variables ficticias para contabilizar el efecto estacional (meses). Su valor de R cuadrada de 88% fue una mejora considerable con respecto a su regresión simple, pero la computadora reportó un porcentaje de error medio (PEMA) de 72.71%. Con este tipo de error, John decidió no usar la regresión múltiple.
- *Capítulo 8:* Finalmente, John obtuvo algunos resultados alentadores mediante el método de descomposición. Pudo mostrar a su banco los índices estacionales y hacer los convenios deseados sobre los pagos de su préstamo. También generó pronósticos para los meses siguientes usando los componentes de tendencia, cíclico, estacional e irregular. Sin embargo, le quedó un sentimiento de inquietud de que no tenía una medición precisa en su método de pronóstico. Pensó en utilizar sus pronósticos y llevar un registro de su precisión mes a mes al ir llegando los datos reales.
- *Capítulo 9:* El modelo autorregresivo que probó John dio por resultado un valor de R cuadrada de 91%. John estaba muy complacido con el resultado hasta que notó que unos de sus valores t era 0.3, dando por resultado que no se rechazara una hipótesis nula. Aunque pensó que sabía por qué había resultado este valor bajo, le pareció que este modelo no iba a funcionar como un método permanente de pronóstico en su forma actual.
- *Capítulo 10:* El método de Box–Jenkins preocupó a John desde un principio, ya que no lo comprendía en su totalidad. Él sabía que tendría que explicar cualquier pronóstico que obtuviera a los inversionistas y banqueros en su intento por obtener capital para expansión, de modo que quería un método de pronóstico que fuera a la vez preciso y comprensible. No estaba seguro de que los reportes de Box–Jenkins calificaran en ninguna de estas categorías.

Al pensar en estos esfuerzos, John se dio cuenta de que el tiempo había transcurrido y que pronto debía generar los pronósticos de sus ingresos mensuales. Disponía de un tiempo limitado para intentar modificaciones a los métodos utilizados y poder pensar en la combinación de dos o más métodos. Pero no tenía tiempo para adquirir nuevos programas de cómputo e intentar métodos completamente diferentes. Al preguntarse qué debería hacer, miró de nuevo uno de sus lemas favoritos que tenía enmarcado en la pared de su oficina: “Hay que hacer algo, aunque esté mal.”

TAREA

Suponga que se le contrata para ayudar a John Mosby con su problema de pronósticos. Redáctele un memorando que resuma sus esfuerzos a la fecha y recomiéndele un curso de acción. Tenga presente que John debe desarrollar pronósticos de sus ventas mensuales con rapidez, que sean lo más precisos posible y que pueda emplear en su exposición a los inversionistas. Sea específico en su recomendación al propietario de Mr. Tux.

CASO DE ESTUDIO 11.5

CONSUMER CREDIT COUNSELING

La operación de Consumer Credit Counseling (CCC) se describió en los caps. 1 (Caso de estudio 1.2) y 3 (Caso de estudio 3.4).

El director ejecutivo, Marv Harinshfeger, llegó a la conclusión de que la variable más importante que CCC requería pronosticar era el número de nuevos clientes por atender en el resto de 1993. Marv proporcionó 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 Caso de estudio 4.3).

Dorothy, con su ayuda, intentó diversas formas de pronosticar la variable más importante. Estos esfuerzos se describen al final de cada capítulo. Concluidos estos intentos, Dorothy decide que es tiempo de resumir estos esfuerzos e intentar llegar a un método para pronosticar el resto del año.

TAREA

Suponga que Dorothy lo asigna a usted para ayudarlo con este problema de pronóstico. Escríbale un reporte que recomiende un curso de acción. Tenga presente que Marv debe desarrollar el pronóstico de número de clientes atendidos que sea lo más preciso posible y que él pueda usar en la toma diaria de decisiones de la organización. Sea específico sobre su recomendación a Dorothy y Marv. Recuerde considerar los aspectos expuestos en este capítulo, como el costo.

CASO DE ESTUDIO 11.6

NUEVA VISITA A LYDIA E. PINKHAM MEDICINE COMPANY

Este caso demuestra una aplicación real del uso de redes neurales en el pronóstico de series de tiempo de datos. Comprendemos que no se han proporcionado a los estudiantes los antecedentes para entender

Este caso es una contribución de Susan C. White, Northwestern State University, Natchitoches, Louisiana. Para mayor información, consúltese 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.

completamente este caso. Sin embargo, pensamos que se derivarán beneficios de la experiencia de este caso real.

La empresa Lydia E. Pinkham y el Compuesto Vegetal de Lydia Pinkham se presentaron en el Caso de estudio 10.4.

Durante el final de la década de 1980 y principios de la de 1990, se han hecho muchos intentos de utilizar redes para pronosticar series de tiempo de datos. La mayoría de los trabajos han sido en el campo de la energía, ya que estas compañías requieren de pronósticos precisos de la demanda por hora de su producto. Sin embargo algunas investigaciones se han enfocado en series de tiempo de negocios más tradicionales como series micro y macroeconómicas, datos demográficos y datos específicos de empresas. Prácticamente todo este trabajo ha utilizado una red de alimentación hacia adelante entrenada mediante retropropagación. Este caso de estudio empleará este tipo de red para pronosticar los datos de ventas de Lydia Pinkham. Los pronósticos resultantes se compararán con los del modelo AR(2) presentado en el Caso de estudio 10.4.

La fig. 11.7 describe una red neural de alimentación hacia adelante 2-4-1 —la red utilizada para este estudio. El primer número en el conjunto 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} . (El uso de dos periodos previos para pronosticar el periodo actual es congruente con el modelo AR(2); así, tanto el modelo AR(2) como el modelo de red neural usan la misma “información” en el cálculo de predicciones un paso adelante.) El 4 indica el número de nodos, o unidades de proceso, en el *estrato oculto*. (Se le denomina “oculto” ya que no está directamente conectado con el “mundo exterior”, como lo están los estratos de entrada y salida.) El número de nodos en el estrato oculto se elige, en cierto sentido, de manera arbitraria: pocos nodos ocultos limitan la capacidad de la red para “ajustar” los datos y demasiados nodos ocultos hace que la red memorice los datos de entrenamiento (o estimación). La memorización conduce a un muy pobre desempeño sobre la muestra de prueba. En este caso, el número de nodos en el estrato oculto es simplemente el doble del número de entradas. Por último, el único nodo de salida da el pronóstico un paso adelante, o Y_t .

La red neural calcula su salida de la siguiente manera: cada uno de los arcos de conexión entre los nodos de dos estratos adyacentes tiene un peso ponderado asociado. Cada nodo en el estrato oculto calcula una suma ponderada de sus entradas. (Los nodos del estrato de entrada simplemente pasan las entradas al estrato oculto.) Esta suma ponderada se “transforma” entonces de cierta forma, tal que $Y = 1/(1 + e^{-x})$, en donde Y pasa al estrato de salida, en donde cada nodo calcula de nuevo una suma ponderada. Esta suma ponderada final es la salida de la red. La red se entrena ajustando todas las ponderaciones conectadas en una forma iterativa.

EL MODELO DE RED NEURAL

La red neural se entrenó utilizando BrainMaker,¹² un paquete comercial de simulación disponible para redes neurales basadas en PC. Para este estudio, el tamaño del paso se fijó en 0.500 y la tolerancia del entrenamiento a 0.001. (Otros paquetes podrían requerir que el usuario especifique una tasa de aprendizaje y un término de momentum.) Como en el Caso de estudio 10.4, se usan las primeras 42 observaciones para entrenar la red; las 12 últimas se emplean para evaluar el desempeño de la red en la predicción de un paso adelante. Un problema con el uso de las redes neurales para pronosticar series de tiempo de datos, está en determinar qué tanto entrenar a la red; una red sobreentrenada tiende a memorizar los datos de entrenamiento y tiene un desempeño pobre sobre los datos de prueba. Por lo

¹² BrainMaker, California Scientific Software, 1990.

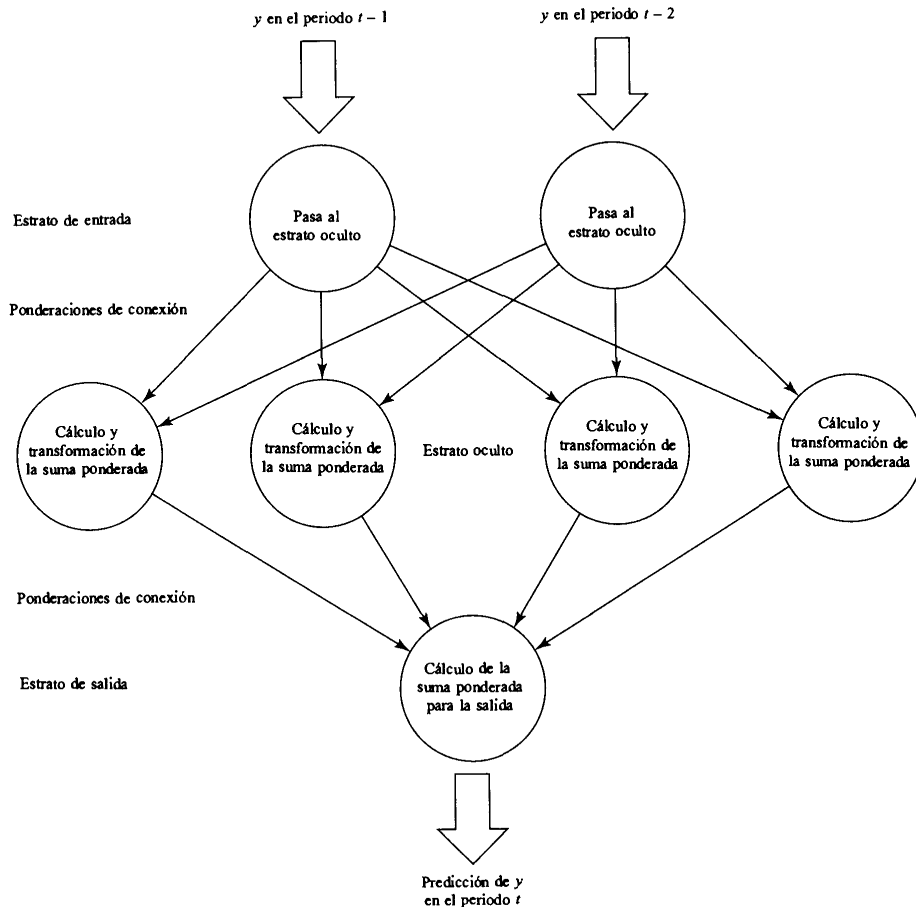


Figura 11.7 Una red 2-4-1 de alimentación hacia adelante.

tanto, algunos investigadores han sugerido simplemente suspender “temprano” el entrenamiento —antes de que la red neural haya memorizado los datos.¹³ Para determinar el efecto del entrenamiento, se entrenó a la red para 10,000, 20 000, 50,000, 75,000 y 100,000 iteraciones. (Una iteración es la presentación de una observación; esto representa 250, 500, 1 250, 1 875 y 2 500 pasadas a través de un conjunto completo de entrenamiento, respectivamente.) Esto permite al analista evaluar el posible impacto del sobreentrenamiento. (El trabajo se realizó en una PC con un procesador Intel 386SX-20, y el “tiempo” de entrenamiento para 100 000 iteraciones fue de aproximadamente 20 minutos.)

¹³ N. Morgan y H. Bourlard, “Generalization and Parameter Estimation in Feedforward Nets: Some Experiments”, en D.S. Touretzky, ed. *Neural Information Processing Systems 2*, 1990, 630-637.

TABLA 11.6 RESULTADOS DE LA RED NEURAL DE LYDIA E. PINKHAM MEDICINE

ITERACIONES	DATOS ESTIMADOS			DATOS DE PRUEBA		
	DAM	PEMA	EMC	DAM	PEMA	EMC
10,000	189.7	11.06%	51923.0	182.2	11.26%	43441.1
20,000	189.8	10.80%	53926.3	169.6	10.18%	37737.4
50,000	195.0	11.22%	54311.3	176.3	10.51%	41529.7
75,000	196.0	11.40%	54424.2	178.1	10.66%	42758.1
100,000	196.5	11.53%	54725.5	181.9	11.02%	43985.4

RESULTADOS

En la tabla 11.6 se presentan el DAM, PEMA y EMC para los diversos modelos de red neural. Éstos no se comparan en forma favorable con el modelo AR(2). El autor de este caso está actualmente experimentando con un tipo diferente de red neural –una red neural de base radial– la cual produce resultados comparables a los del modelo AR(2).

Para una lectura adicional sobre redes neurales, véase Tarun Khanna, *Foundations of Neural Networks* (Reading, MA.: Addison–Wesley Publishing Company, 1990).

PREGUNTAS

1. Encuentre un artículo que describa una aplicación de redes neurales a los pronósticos de series de tiempo. ¿Qué método utilizaron los autores y qué tanto éxito tuvo?
2. Si tiene usted acceso a un simulador de red neural, trate de encontrar una red que produzca mejores pronósticos que la red 2–4–1 presentada aquí.
3. ¿Por qué se vislumbra a las redes neurales como una alternativa viable a los otros métodos de pronóstico expuestos en este libro?

BIBLIOGRAFÍA

- ADAMS, F. G. *The Business Forecasting Revolution*. New York: Oxford University Press, 1986.
- GEORGOFF, D. M. y MARDICK, R. G. "Manager's Guide to Forecasting." *Harvard Business Review* 1 (1986): 110–120.
- HOGARTH, R. M. y MAKRIDAKIS, S. "Forecasting and Planning: An Evaluation." *Management Science* 27 (2) (febrero, 1981): 115–138.
- MAKRIDAKIS, S. "The Art and Science of Forecasting." *International Journal of Forecasting* 2 (1986): 15–39.
- REID, R. A. "The Forecasting Process: Guidelines for the MIS Manager." *Journal of Systems Management* (Noviembre, 1986): 33–37.
- WRIGHT, G. y AYTON, P. EDS. *Judgmental Forecasting*. New York: John Wiley & Sons, 1987.

APÉNDICES

Apéndice A Derivaciones

Apéndice B Gráficas de proporción o semilogarítmicas

Apéndice C Tablas

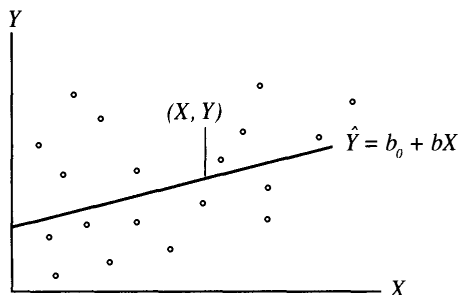
Apéndice D Datos para el Caso de estudio 7.1

Apéndice E Conjuntos de datos y base de datos

Derivaciones

Derivación de correlación

$$\begin{aligned}
 r &= \frac{\sum Z_X Z_Y}{N} = \sum \frac{[(X - \mu_X)/\sigma_X][(Y - \mu_Y)/\sigma_Y]}{N} \\
 &= \frac{\sum (X - \mu_X)(Y - \mu_Y)}{\sqrt{(\sum X^2/N) - (\sum X/N)^2} \sqrt{[(\sum Y^2/N) - (\sum Y/N)^2]/N}} \\
 &= \frac{\sum (X - \mu_X)(Y - \mu_Y)}{\sqrt{[N \sum X^2 - (\sum X)^2]/N^2} \sqrt{[N \sum Y^2 - (\sum Y)^2]/(N^2/N)}} \\
 &= \frac{N \sum (X - \mu_X)(Y - \mu_Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \\
 &= \frac{N \sum (XY - Y\mu_X - X\mu_Y + \mu_X\mu_Y)}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \\
 &= \frac{N[\sum XY - (\sum X \sum Y/N) - (\sum X \sum Y/N) + N(\sum X/N)(\sum Y/N)]}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \\
 &= \frac{N[\sum XY - (\sum X \sum Y/N) - (\sum X \sum Y/N) + (\sum X \sum Y/N)]}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \\
 &= \frac{N[\sum XY - (\sum X \sum Y/N)]}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}} \\
 &= \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}
 \end{aligned}$$

Derivación de mínimos cuadrados

$$\begin{aligned}
 d &= Y - \hat{Y} \\
 &= Y - (b_0 + bX) \\
 d^2 &= [Y - (b_0 + bX)]^2 \\
 \Sigma d^2 &= \Sigma [Y - (b_0 + bX)]^2 \\
 &= \Sigma (Y - b_0 - bX)^2
 \end{aligned}$$

Derivadas parciales

$$\begin{aligned}
 \frac{\delta \Sigma}{\delta b} &= 2 \Sigma (Y - bX - b_0)(-X) & \frac{\delta \Sigma}{\delta b_0} &= 2 \Sigma (Y - bX - b_0)(-1) \\
 &= 2 \Sigma (-XY + bX^2 + b_0X) & &= 2 \Sigma (-Y + bX + b_0)
 \end{aligned}$$

Para obtener los mínimos, igualar los parciales a cero

$$\begin{aligned}
 \frac{\delta \Sigma}{\delta b} = 0: \quad 2 \Sigma (-XY + bX^2 + b_0X) &= 0 & \frac{\delta \Sigma}{\delta b_0} = 0: \quad 2 \Sigma (-Y + bX + b_0) &= 0 \\
 \Sigma (-XY + bX^2 + b_0X) &= 0 & \Sigma (-Y + bX + b_0) &= 0 \\
 -\Sigma XY + b_0 \Sigma X + b \Sigma X^2 &= 0 & -\Sigma Y + Nb_0 + b \Sigma X &= 0
 \end{aligned}$$

Encontrar una b_0 y b tales que Σd^2 sea un mínimo

$$\begin{aligned}
 b_0 \Sigma X + b \Sigma X^2 &= \Sigma XY & \times N \\
 Nb_0 + b \Sigma X &= \Sigma Y & \times \Sigma X \\
 Nb_0 \Sigma X + Nb \Sigma X^2 &= N \Sigma XY \\
 Nb_0 \Sigma X + b(\Sigma X)^2 &= \Sigma X \Sigma Y & \text{resta} \\
 \hline
 Nb \Sigma X^2 - b(\Sigma X)^2 &= N \Sigma XY - \Sigma X \Sigma Y \\
 b[N \Sigma X^2 - (\Sigma X)^2] &= N \Sigma XY - \Sigma X \Sigma Y \\
 b &= \frac{N \Sigma XY - \Sigma X \Sigma Y}{N \Sigma X^2 - (\Sigma X)^2} & \text{fórmula de la pendiente}
 \end{aligned}$$

$$Nb_0 + b \Sigma X = \Sigma Y$$

$$Nb_0 = \Sigma Y - b \Sigma X$$

$$b_0 = \frac{\Sigma Y}{N} - \frac{b \Sigma X}{N} = \bar{Y} - b\bar{X} \quad \text{fórmula de la intersección con } Y$$

CONVERSIÓN DE VALORES DE TENDENCIA ANUAL A MENSUAL

Pasos para convertir la ecuación de tendencia de anual a mensual:

1. Se deben convertir la pendiente y la constante de valores anuales a mensuales. Para ello, divida b_0 y b_1 entre 12. Este paso debe omitirse cuando los datos sean acumulativos o cuando las cifras anuales sean porcentajes. En otras palabras, para que este paso sea aplicable, los datos mensuales deben sumar los totales anuales. La ecuación anual para el registro de automóviles nuevos es

$$\hat{Y} = 8.043 + .06368X$$

en donde 1 unidad de X representa 1 año, y $X = 0$ representa el 1 de julio de 1959.

$$\hat{Y} = 8.043/12 + .06368/12(X)$$

$$\hat{Y} = .67025 + .0053066X$$

2. Los datos anuales y mensuales deben registrarse en el mismo tipo de unidades. Si están así, omita este paso. En caso contrario, convierta los datos anuales a mensuales desplazando el punto decimal según corresponda. Como los datos anuales están registrados en millones de registros y los mensuales en miles, desplace el punto decimal tres posiciones hacia la derecha.

$$\hat{Y} = 670.25 + 5.3066X$$

3. Se debe convertir el valor de X de anual a mensual dividiéndolo entre 12. En la ecuación, la X se refiere a unidades en términos de años. Convierta X de modo que represente meses dividiéndola entre 12.

$$\hat{Y} = 670.25 + 5.3066/12(X)$$

$$\hat{Y} = 670.25 + .4422X$$

4. Se debe cambiar el origen a la mitad del primer mes de datos, por lo regular el 15 de enero. Para moverlo del 1 de julio de 1959 al 15 de enero de 1986, se deben sumar 317.5 meses al valor codificado de X . La ecuación mensual final es

$$\hat{Y} = 670.25 + .4422(X + 317.5)$$

$$\hat{Y} = 670.63 + .4422X + 140.399$$

$$\hat{Y} = 810.649 + .4422X$$

en donde

Y = valores mensuales de tendencia esperados

X = valores mensuales con el 0 ubicado en el 15 de enero de 1986

Gráficas de proporción o semilogarítmicas

ACTIVIDAD

Cuando una gráfica con una escala aritmética vertical no demuestra ser útil, se encuentran muchos tipos de problemas. Como un ejemplo típico de esta situación, considérese el caso de Alexander Furniture Company. Esta empresa ha estado fabricando una línea de muebles que nosotros llamaremos línea de producto A. En 1990 comenzó a producir una línea adicional de muebles, línea de producto B, que difería de la anterior tanto en estilo como en margen de precios. En el periodo de 1990 a 1994, ambas líneas de productos han incrementado sus ventas, como se muestra en la fig. B.1.

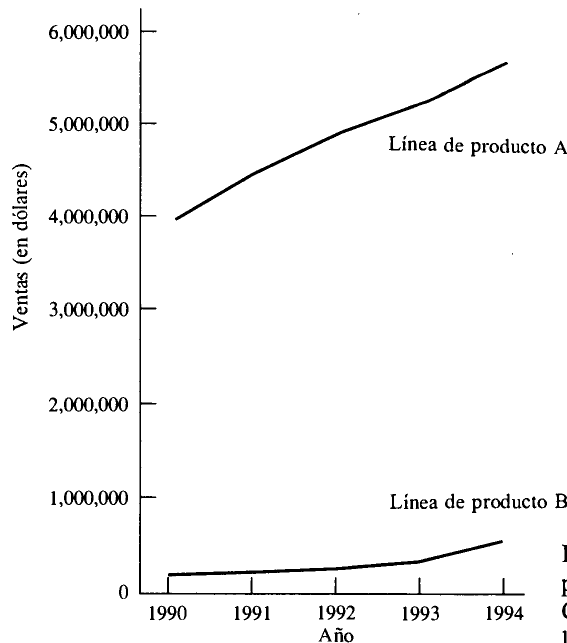


Figura B.1 Ventas de dos líneas de productos de la Alexander Furniture Company, 1990-1994; gráfica de línea aritmética.

Esta gráfica, que utiliza una escala aritmética vertical, ofrece la impresión distintiva de que la línea de producto A se desempeñó mucho mejor en este periodo que la línea de producto B. Esta impresión surge debido a que el incremento *absoluto* en las ventas de la línea de producto A fue mucho mayor que el de la línea de producto B, de ahí que la curva de la línea A suba de manera más acentuada en la gráfica.

La administración de la compañía no tenía expectativas de que el volumen de ventas de la nueva línea se igualara al de la línea ya establecida, ni tampoco de que el incremento en las ventas de la nueva línea igualara o superara al de la línea establecida. Por ello, el interés no se centró en una comparación de las ventas *absolutas* o del incremento *absoluto*, sino más bien en una comparación del incremento *relativo* en las ventas de ambas líneas. Específicamente, la pregunta que se formuló fue si las ventas de la línea B se incrementaron más rápidamente que las de la línea A. Se puede responder a esta pregunta mediante el cálculo de lazos relativos para cada línea de producto y graficándolos en papel gráfico aritmético. Una gráfica con una escala proporcional o logarítmica también responderá a esta pregunta –y sin ningún cálculo intermedio.

Antes de continuar con la exposición relativa a la Alexander Furniture Company, abordaremos las características de la escala proporcional o logarítmica.

ESCALA PROPORCIONAL O LOGARÍTMICA

En la fig. B.2 se muestra una gráfica cuya escala horizontal, o de tiempo, es una escala aritmética convencional. Sin embargo, la escala vertical es una *escala proporcional* o

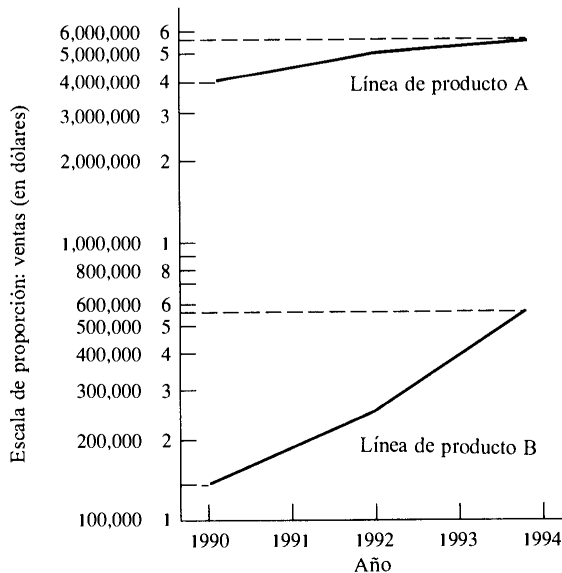


Figura B.2 Ventas de dos líneas de productos de la Alexander Furniture Company, 1990-1994; gráfica de línea proporcional o semilogarítmica.

logarítmica, es decir, una escala en la que intervalos iguales representan diferencias iguales en los logaritmos o números. Dicha gráfica se denomina *gráfica proporcional o semilogarítmica*.

Nótese los espacios graduados de las marcas en la escala vertical. En ella se muestran dos conjuntos de números. Los números interiores, 1, 2, etc., no son parte, propiamente dicha, de la escala, pero corresponden a los números guía que se encuentran en el papel gráfico semilogarítmico. Por lo regular estos números guía no se reproducen en la gráfica, pero los incluimos aquí para mostrar cómo se construye una escala proporcional.

El número inferior guía es 1 y sirve como el punto de partida de la escala. Cualquier número positivo de la escala diferente de cero debe colocarse aquí. En el caso de la Alexander Furniture Company el volumen de ventas más pequeño fue de \$135 000; por ello, es más conveniente iniciar la línea de base de la escala con \$100 000. Aunque podríamos comenzar la línea de base de la escala con cualquier valor positivo diferente de cero, por lo regular es más sencillo empezar con 1, 10 o algún múltiplo de 10, ya que entonces las principales divisiones horizontales señaladas con marcas representarán números redondos. En el número guía opuesto 2, colocamos el valor que es el doble del de la línea de base, o sea \$200 000. En forma similar, escribimos \$300 000 que corresponden al número guía 3 y así sucesivamente. Al número guía 1 que sigue al número 9 se le asigna un valor de 10 veces el de la línea de base; en este caso se le da un valor de \$1 000 000. Nótese que las distancias verticales entre los números guía, que venían haciéndose más pequeñas hasta este punto, ahora se hacen más grandes. De hecho, es la misma secuencia de graduaciones que inicia de nuevo. Cada secuencia completa de graduaciones se denomina *ciclo*. Al etiquetar el siguiente ciclo, se procede igual que antes. Debemos escribir \$2 000 000 que corresponden al número guía 2 del segundo ciclo, ya que este número es dos veces el valor de la base del segundo ciclo. Nótese que el tope del primer ciclo se convierte ahora en la base del segundo ciclo. Procediendo de esta forma, llenaremos la escala vertical de la fig. B.2. Se puede apreciar que el segundo ciclo no está completo, pues no hay necesidad de llevar la escala hasta \$1 000 000.

El papel gráfico semilogarítmico se encuentra disponible en uno, dos, tres o incluso más ciclos. Desde luego, el número de ciclos depende del alcance de los datos. Si se graficara una serie en la que el valor más alto es ocho veces el valor más pequeño, por lo regular se usaría papel de un ciclo, ya que sabemos que el tope de un ciclo tiene 10 veces el valor de la línea de base. Por otra parte, una serie cuyo valor más alto fuera, por decir, 12 veces el más pequeño, requeriría de un papel gráfico semilogarítmico con por lo menos dos ciclos.

Una vez etiquetada la escala vertical, se produce la graficación de la serie en la forma usual. Es común que el papel gráfico semilogarítmico tenga un rayado entre los números guía de la escala vertical para ayudar a la graficación. Si, por decir, hay 20 espacios entre los números guía 1 y 2 en el papel gráfico empleado en la fig. B.2, cada uno de los espacios representará \$5 000 en el primer ciclo y \$50 000 en el segundo. Estas líneas de rayado adicional en la escala vertical permiten graficar los datos con mucha precisión.

INTERPRETACIÓN DE LA ESCALA PROPORCIONAL O LOGARÍTMICA

Después de considerar la mecánica de etiquetado de la escala vertical y de graficar los puntos, podemos ahora discutir las formas de interpretar una gráfica semilogarítmica. Primero, compare la distancia en la escala vertical de la fig. B.2 entre \$100 000 y \$300 000 y entre \$300 000 y \$900 000. Nótese que en cada instancia se cubrió la misma distancia en la escala vertical, sin importar el hecho de que en el primer caso representa un incremento de \$200 000, mientras que en el segundo representa un incremento de \$600 000. Nótese también que el incremento *relativo* en cada caso es el mismo, es decir 200%. Si se mide en la escala vertical cualquier otro caso en el que se haya un incremento de 200%, se encontrará que la distancia requerida para un incremento de 200% en la escala vertical de la fig. B.2 es siempre la misma.

En general, la escala proporcional tiene las siguientes características importantes:

- Porcentajes iguales de cambio requieren de distancias verticales iguales en la escala proporcional.
- Entre mayor sea el cambio relativo, mayor será la distancia vertical requerida.

Ahora regresaremos a nuestra discusión de las escalas de las líneas de los productos A y B de la Alexander Furniture Company. En la fig. B.2, se superpusieron a la gráfica líneas horizontales de guiones para ayudar a comparar las distancias requeridas para reflejar el incremento en las ventas entre 1990 y 1994 para cada una de las dos líneas de producto. Como la distancia es mayor para la línea del producto B que para la línea del producto A, concluimos que la línea del producto más reciente experimentó un mayor incremento *relativo* en las ventas en estos cinco años.

Para interpretar la gráfica proporcional, no es necesaria la medición, en esta forma, de la distancia vertical. Nótese en la fig. B.2 que la curva de la línea del producto B está más inclinada que la de la línea del producto A, ya que la curva de B tiene que cubrir una mayor distancia vertical durante el mismo periodo. De ahí que podamos decir simplemente lo siguiente:

- Entre más inclinada sea la curva en una gráfica proporcional, mayor será el porcentaje de cambio.

Este resultado significa que sólo se necesita comparar las pendientes de las curvas para obtener conclusiones sobre los cambios relativos. Si las dos líneas de la fig. B.2 hubieran sido paralelas, habrían indicado una proporción de cambio igual, por supuesto porque dos líneas rectas paralelas tienen la misma pendiente.

Además de comparar los cambios relativos entre dos periodos entre varias series sobre una gráfica proporcional, podemos también estudiar los cambios relativos entre un periodo y otro en cualquier serie dada. Por ejemplo, la fig. B.3(a) presenta una gráfica proporcional en la que la serie está representada por una línea recta. Es evidente de inmediato que el porcentaje de cambio entre uno y otro periodo de esta serie es constante, ya que una línea

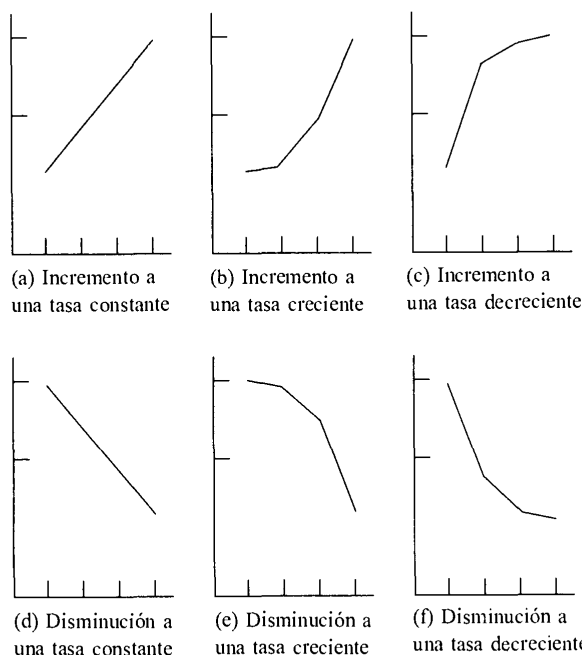


Figura B.3 Ejemplos de diferentes patrones sobre papel proporcional.

recta tiene una pendiente constante. Mientras que una línea recta en una gráfica aritmética indica una *cantidad* constante de cambio, una línea recta en una gráfica proporcional indica una *proporción* o *tasa* constante de cambio.

La fig. B.3(b) presenta en una gráfica proporcional una serie cuya tasa de incremento se hace cada vez más grande al graficar cada punto sucesivo. Se sabe este resultado porque la curva se hace cada vez más inclinada. En breve, estableceremos que esta serie aumenta en una tasa creciente.

En forma similar, podemos fácilmente apreciar que la serie en la fig. B.3(c) aumenta en una tasa decreciente ya que la curva se hace cada vez menos inclinada. La fig. B.3(d) presenta una serie que disminuye a una tasa constante, mientras que las figs. B.3(e) y (f) presentan dos series que disminuyen a tasas creciente y decreciente, respectivamente.

Hagamos ahora una Interpretación más completa de la fig. B.2. Aunque las ventas de ambas líneas de productos han estado aumentando, las ventas de la nueva línea de producto lo ha hecho más rápido. De hecho, las ventas de la línea del producto B han estado aumentando a una tasa creciente, mientras que las ventas de la línea del producto A han aumentado a una tasa decreciente. De ahí que, aunque las ventas de la nueva línea aún están muy por debajo de las de la línea ya establecida, y aunque los incrementos absolutos en ventas sean menores, la nueva línea de producto ha tenido éxito. Por lo tanto, las gráficas proporcionales proporcionan respuestas expeditas a las preguntas de la administración con respecto al incremento *relativo* en las ventas de las dos líneas de producto.

Tablas

- Tabla C.1 Términos individuales de la Distribución Binomial
- Tabla C.2 Áreas de Distribución de probabilidad normal estándar
- Tabla C.3 Valores críticos de t
- Tabla C.4 Valores críticos de χ^2 cuadrada
- Tabla C.5 Dígitos aleatorios
- Tabla C.6 Distribución F
- Tabla C.7 Límites de prueba de Durbin-Watson

TABLA C.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	.0025
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000	.4950	.4800	.4550	.4200	.3750	.3200	.2550	.1800	.0950
3	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500	.3025	.3600	.4225	.4900	.5625	.6400	.7225	.8100	.9025
	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250	.0911	.0640	.0429	.0270	.0156	.0900	.0034	.0010	.0001
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750	.3341	.2880	.2389	.1890	.1406	.0960	.0574	.0270	.0071
4	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
	0		.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625	.0410	.0256	.0150	.0081	.0039	.0016	.0005	.0001	.0000
5	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
	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
6	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0078	.0503	.0313	.0185	.0102	.0053	.0024	.0010	.0003	.0001	.0000	.0000
	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
	4	.0000	.0004	.0022	.0064	.0146	.0283	.0488	.0768	.1128	.1562	.2059	.2592	.3124	.3601	.3955	.4096	.3915	.3281	.2036
7	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
	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
8	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
	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
	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734	.2388	.1935	.1442	.0972	.0577	.0287	.0109	.0026	.0002
9	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
	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

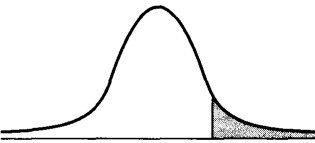
<i>n</i>	<i>x</i>	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039	.0017	.0007	.0002	.0001	.0000	.0000	.0000	.0000	.0000
	1	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0313	.0164	.0079	.0033	.0012	.0004	.0001	.0000	.0000	.0000
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094	.0703	.0413	.0217	.0100	.0038	.0011	.0002	.0000	.0000
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188	.1719	.1239	.0808	.0467	.0231	.0092	.0026	.0004	.0000
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734	.2627	.2322	.1875	.1361	.0865	.0459	.0185	.0046	.0004
	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188	.2568	.2787	.2786	.2541	.2076	.1468	.0839	.0331	.0054
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094	.1569	.2090	.2587	.2965	.3115	.2936	.2376	.1488	.0515
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312	.0548	.0896	.1373	.1977	.2670	.3355	.3847	.3826	.2793
9	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039	.0084	.0168	.0319	.0576	.1001	.1678	.2725	.4305	.6634
	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
10	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
	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
11	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
	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

(continuación)

TABLA C.1 (Continuación)

<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
12	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
	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	.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

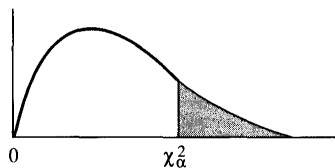
Fuente: Tabla A, págs. 464-466 en *Business Statistics: Concepts and Applications* de William J. Stevenson, Copyright © 1978 por William J. Stevenson. Reproducida con permiso de Harper & Row, Publishers, Inc.

TABLA C.3 VALORES CRÍTICOS DE t


t_{α}

df	$t_{.995}$	$t_{.950}$	$t_{.925}$	$t_{.910}$	$t_{.905}$
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." Calculada por Maxine Merrington, *Biometrika*, vol. 32 (1941), p. 300. Reproducida con permiso del Profesor D. V. Lindley.

TABLA C.4 VALORES CRÍTICOS DE χ^2 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

(continuación)

TABLA C.4 (Continuación)

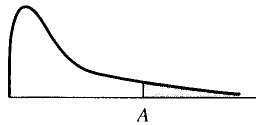
df	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99147	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8381
4	7.77944	9.48773	11.1433	13.2767	14.8602
5	9.23635	11.0705	12.8325	15.0863	16.7496
6	10.6446	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 Percentage Points of the χ^2 -Distribution." *Biometrika*, vol. 32 (1941), pp. 188-189, por Catherine M. Thompson. Reproducida con permiso del Profesor D. V. Lindley.

TABLA C.5 DÍGITOS ALEATORIOS

43732	52254	51717	24199	14995	28638	94266	95896	97286	93363
29127	93840	32774	55120	65026	42329	24853	29925	76811	81401
24907	74544	66673	00700	66710	66969	74990	20032	21995	06036
27618	67022	10133	91336	55075	03262	96546	49329	25175	18575
76254	64180	39786	34653	87041	62316	40460	13053	81241	04385
62416	36939	55843	27845	49480	77704	47938	49743	45798	81296
31941	66753	21574	01290	78304	10121	25145	44925	96389	02748
48769	40172	91480	49345	40787	26343	44517	27111	69002	07130
19841	14663	09283	61166	78039	33309	94009	13456	49850	07814
95278	45022	96058	47206	45136	06897	13029	98610	47895	29255
11681	17274	17775	70451	22664	67014	88052	07139	86031	41752
19662	09277	24043	30468	25419	44660	52122	77683	89932	61867
38220	45565	14942	08320	43174	92076	52890	98982	51549	24199
74342	90018	21144	33405	63152	95923	77259	09132	55290	49080
36330	30763	76197	40481	45306	40321	67829	49329	59366	84654
55667	24316	35987	50597	08340	19788	50319	57122	43216	85841
97044	03505	01390	48719	77194	08143	85905	95243	18460	81857
10801	41372	68587	75813	97859	89824	00856	68893	89724	13555
62086	15756	82269	62301	54394	88005	69419	92167	81404	27619
82969	15207	17095	99636	32773	53706	01064	71431	60025	76456
92689	66288	41679	28175	43057	31307	79854	99889	65340	67466
82257	95475	34860	82583	88431	28935	87509	20727	37989	73978
03500	89229	23013	34269	86323	82028	09026	03845	47049	09033
49935	10627	06590	00319	21022	67060	97351	83563	49386	68421
90394	48453	98568	59114	67484	92490	15912	17007	06152	46270
46528	28504	62341	69676	36687	62032	57678	78816	59456	60820
61139	84737	07313	96815	80079	28473	00893	54263	84568	89126
72052	17047	82703	98378	61551	52642	01676	82279	78996	97089
96412	05437	41920	79190	96446	75572	86149	04486	64642	99840
38542	80332	61559	51540	27508	77623	52532	76913	03934	25756
13240	10094	36445	78755	02259	79075	86304	44848	05617	19112
73520	51141	98860	57952	91325	24661	21656	40584	88869	66593
80523	22661	47316	22278	41056	46455	72563	09140	32256	48198
46746	30636	21103	65113	40794	62196	72234	26648	48913	34334
93786	00483	52986	55922	30830	40750	63223	63371	65047	20933
16286	09146	89517	87223	79385	41937	91686	50357	70316	09026
46721	17036	59388	17337	96097	60336	81050	02307	69823	10816
52053	30051	42444	54844	77985	70091	60464	79822	63536	15688
87528	93477	21244	58252	75052	19270	62520	47603	61395	25861
77360	08067	93429	68336	45621	75564	73052	00972	10065	97831
46689	76298	22771	73816	08640	77702	13451	51694	53359	83400
01401	68433	58746	01648	68134	22848	67255	66420	70730	78826
53605	54676	09068	58018	61071	32171	47150	68624	56213	75608
01524	52265	96518	45107	36343	93236	84759	25889	43255	98284
12742	23345	74617	19338	29434	91944	62796	37134	18666	31430

TABLA C.6 DISTRIBUCIÓN *F*



Por ejemplo, el valor de la escala *F* para $\delta_2 = 3$, $\delta_1 = 10$ correspondiente al área .01 en la parte derecha de la cola es 6.55.

El valor *F* correspondiente al área .05 en la parte derecha de la cola aparece en caracteres normales.

El valor *F* correspondiente al área .01 en la parte derecha de la cola aparece en negritas.

δ_1 , Grados de libertad del denominador	δ_2 , grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	10
1	161 4,052	200 4,999	216 5,403	225 5,625	230 5,764	234 5,859	237 5,928	239 5,981	241 6,022	242 6,056
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.36	19.37 99.37	19.38 99.39	19.39 99.40
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.29	4.78 10.15	4.74 10.05
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.03	2.96 4.69	2.85 4.46	2.77 4.28	2.70 4.14	2.65 4.03	2.60 3.94
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.70 4.14	2.64 4.00	2.59 3.89	2.55 3.80
16	4.49 8.53	3.36 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69
17	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.62 3.93	2.55 3.79	2.50 3.68	2.45 3.59

TABLA C.6 (Continuación)

δ_1 , Grados de libertad del denominador	δ_2 , grados de libertad del numerador									
	1	2	3	4	5	6	7	8	9	10
18	4.41 8.28	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.85	2.51 3.71	2.46 3.60	2.41 3.51
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: Reproducido mediante permiso de *Statistical Methods*, 7a. ed., por George W. Snedecor y William C. Cochran. Copyright © 1980 de la Iowa State University Press, Ames, Iowa.

TABLA C.7 LÍMITES DE PRUEBA DE DURBIN-WATSON

Nivel de significación $\alpha = .05$										
n	$p-1=1$		$p-1=2$		$p-1=3$		$p-1=4$		$p-1=5$	
	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
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

TABLA C.7 (Continuación)

Nivel de significación $\alpha = .01$										
n	$p - 1 = 1$		$p - 1 = 2$		$p - 1 = 3$		$p - 1 = 4$		$p - 1 = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
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
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

Fuente: Reproducido con permiso de J. Durbin y G. S. Watson. "Testing for Serial Correlation in Least Squares regression-II", *Biometrika*, vol. 38 (1951), pp. 159-178.

Datos para el Caso de estudio 7.1

14.75	0.00	0.00	2.01	12.57	10.00	20.50
14.50	0.00	0.00	2.53	12.57	10.00	20.00
14.13	1.00	0.00	2.10	12.57	10.00	20.00
14.63	0.00	1.00	4.13	12.14	30.00	20.00
14.00	1.00	0.00	2.10	12.57	10.00	20.00
13.38	0.00	1.00	3.97	12.57	10.00	20.00
14.57	0.00	1.00	3.27	12.14	30.00	20.00
13.88	1.00	0.00	3.50	13.19	10.00	19.50
15.38	0.00	0.00	2.85	13.19	10.00	19.50
15.63	0.00	0.00	1.81	13.12	10.00	18.50
15.88	1.00	0.00	2.72	12.69	30.00	18.50
15.00	1.00	0.00	2.43	13.12	10.00	18.00
16.13	0.00	0.00	3.27	12.69	30.00	18.00
15.25	0.00	1.00	3.13	12.69	30.00	17.50
16.00	0.00	0.00	2.55	13.68	10.00	17.00
16.25	0.00	0.00	2.08	13.68	10.00	17.50
17.38	0.00	0.00	2.12	13.20	30.00	17.50
16.35	1.00	0.00	3.40	14.10	10.00	19.00
17.00	1.00	0.00	2.63	13.60	30.00	19.00
16.00	0.00	1.00	2.61	14.10	10.00	19.50
16.63	1.00	0.00	2.06	14.10	10.00	19.50
16.38	0.00	0.00	2.08	14.10	10.00	20.00
16.75	1.00	0.00	2.09	13.60	30.00	20.00
15.13	0.00	1.00	4.29	12.69	30.00	20.00
16.00	1.00	0.00	2.50	12.96	30.00	20.00
14.50	0.00	1.00	3.32	13.47	10.00	20.00
16.25	0.00	0.00	2.95	12.96	30.00	20.00
16.88	0.00	0.00	1.85	14.28	10.00	20.50
17.38	0.00	0.00	1.55	13.59	30.00	20.50
16.00	0.00	1.00	3.33	14.28	10.00	20.50
16.75	1.00	0.00	2.77	14.94	10.00	20.50

17.13	0.00	0.00	2.18	14.94	10.00	20.50
17.50	0.00	1.00	4.21	14.67	30.00	20.50
17.00	1.00	0.00	2.66	15.32	10.00	19.50
16.75	0.00	1.00	3.58	15.32	10.00	19.50
17.20	0.00	1.00	2.96	15.32	10.00	19.50
18.75	0.00	0.00	1.93	15.32	10.00	19.50
17.50	0.00	1.00	2.57	14.68	30.00	19.00
17.50	0.00	0.00	3.18	15.15	10.00	18.00
18.00	0.00	0.00	1.93	15.15	10.00	18.00
15.63	0.00	0.00	2.20	13.39	10.00	17.00
14.75	1.00	0.00	2.21	13.39	10.00	17.00
15.25	0.00	1.00	3.24	13.35	30.00	16.50
15.75	1.00	0.00	2.35	13.35	30.00	16.50
15.25	1.00	0.00	2.11	13.39	10.00	16.50
15.75	1.00	0.00	2.80	13.35	30.00	16.50
15.63	0.00	0.00	1.95	13.39	10.00	16.50
16.13	0.00	0.00	2.80	13.39	10.00	16.50
15.75	1.00	0.00	4.00	13.35	30.00	16.00
16.13	0.00	0.00	2.81	13.50	10.00	15.80
16.25	1.00	0.00	3.38	13.50	30.00	15.80
16.00	0.00	0.00	2.57	13.50	10.00	15.80
15.88	0.00	1.00	3.96	13.50	30.00	15.80
16.50	1.00	0.00	2.67	13.50	30.00	15.80
16.38	1.00	0.00	3.05	13.50	30.00	15.80
12.50	1.00	0.00	2.36	10.60	30.00	15.30
12.25	1.00	0.00	2.54	10.60	30.00	15.30
14.25	1.00	0.00	2.20	12.13	30.00	15.30
15.00	1.00	0.00	3.03	12.13	30.00	15.80
15.25	1.00	0.00	3.24	12.13	30.00	16.50
16.00	0.00	0.00	1.95	12.34	30.00	17.80
14.88	1.00	0.00	2.86	12.34	30.00	17.80
14.75	1.00	0.00	2.64	12.34	30.00	19.00
15.50	1.00	0.00	2.23	11.40	30.00	20.00
13.75	1.00	0.00	2.24	11.40	30.00	19.50
11.30	1.00	0.00	3.24	11.36	30.00	17.50
12.38	1.00	0.00	1.95	11.36	30.00	17.50
12.15	1.00	0.00	2.32	11.36	30.00	14.50
11.75	1.00	0.00	2.45	9.81	30.00	13.00
12.38	1.00	0.00	1.88	9.81	30.00	13.00
12.63	0.00	0.00	1.76	9.81	30.00	13.00
11.13	1.00	0.00	1.99	9.81	30.00	12.50
11.38	0.00	0.00	2.20	9.78	10.00	12.50
11.88	1.00	0.00	2.14	9.81	30.00	12.00
11.75	1.00	0.00	2.61	9.81	30.00	12.00
13.63	0.00	0.00	1.84	10.24	30.00	11.00
13.88	0.00	0.00	1.62	11.00	30.00	11.00
13.00	1.00	0.00	3.56	11.00	30.00	11.00
12.00	1.00	0.00	2.65	11.10	10.00	11.00
13.13	1.00	0.00	2.65	11.00	30.00	11.00
14.27	0.00	0.00	1.80	11.34	30.00	12.30

14.63	0.00	0.00	1.69	11.34	30.00	12.30
15.25	0.00	0.00	1.88	11.34	30.00	12.20
14.25	1.00	0.00	2.77	11.34	30.00	12.30
13.52	1.00	0.00	2.22	11.75	10.00	13.50
14.63	1.00	0.00	2.42	11.59	30.00	13.50
14.75	0.00	0.00	1.77	11.39	30.00	13.50
14.00	0.00	0.00	2.22	11.75	10.00	13.50
14.50	0.00	0.00	2.99	11.59	30.00	13.50
14.25	0.00	0.00	2.22	11.75	10.00	13.50
14.63	0.00	0.00	1.93	11.75	10.00	14.50
13.30	1.00	0.00	3.35	12.68	10.00	15.50
14.50	0.00	0.00	2.21	12.68	10.00	17.00

Conjuntos de datos y base de datos

Este apéndice contiene 17 conjuntos de series de datos y ocho variables financieras medidas cada una para 266 corporaciones de EUA en los últimos años.¹

¹ Estamos en deuda con la Dra. Lynn Stephens de la Eastern Washington University por proporcionarnos estos últimos datos.

APÉNDICE E CONJUNTOS DE DATOS

SALIDAS AL EXTRANJERO DE CIUDADANOS ESTADUNIDENSES (EN MILES)		PRODUCCIÓN EN GRANJAS CASILLEROS DE HUEVO (EN MILLONES)
1961	2,020	173.4
	2,292	176.6
	2,588	176.4
	2,841	181.2
	3,341	182.5
1966	3,814	184.7
	4,334	194.8
	4,820	192.5
	5,767	191.5
	6,499	190.1
1971	7,059	193.6
	8,312	192.3
	8,758	183.5
	8,306	182.3
	8,177	179.5
1976	7,755	179.2
	8,330	179.5
	9,245	186.5
	9,939	192.3
	9,971	193.6
1981	9,978	194.0
	10,275	193.7
	12,258	189.4
	13,909	189.5
	14,768	190.0
1986	14,357	190.0
	16,425	193.2
	17,209	193.5
1991	17,603	186.8
	19,022	188.6
	17,579	192.2
		195.9

PRINCIPALES ELECTRODOMÉSTICOS, EMBARQUES INDUSTRIALES, CONGELADORES (EN MILES).

AÑO	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
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)

AÑO	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
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	

PRESCRIPCIONES NUEVAS

AÑO	1984	1985	1986	1987	1988	1989	1990
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 DIARIOS POR AUTOBÚS PARA 146 DÍAS EN LA RUTA PIKE STREET, SEATTLE, WASHINGTON.

350	339	351	364	369	331	331	340	346	341	357	398	381	367	383	375
353	361	375	371	373	366	382	429	406	403	429	425	427	409	402	409
419	404	429	463	428	449	444	467	474	463	432	453	462	456	474	514
489	475	492	525	527	533	527	522	526	513	564	599	572	587	599	601
611	620	579	582	592	581	630	663	638	631	645	682	601	595	521	521
516	496	538	575	537	534	542	538	547	540	526	548	555	545	594	643
625	616	640	625	637	634	621	641	654	649	662	699	672	704	700	711
715	718	652	664	695	704	733	772	716	712	732	755	761	748	748	750
744	731	782	810	777	816	840	868	872	811	810	762	634	626	649	697
657	549														

Los datos se leen de izquierda a derecha.

ESTADÍSTICAS DE OCUPACIÓN MENSUAL DEL MOTEL NINE.

AÑO	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
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 CELULOSA (MILES DE TONELADAS CORTAS).

AÑO	1988	1989	1990	1991	1992
Enero	5276	5466	5480	5579	5681
Febrero	4919	4821	4988	5033	5254
Marzo	5227	5307	5272	5319	5525
Abril	5003	5193	5254	5285	5489
Mayo	5004	5087	5156	5188	5320
Junio	4949	5102	5175	5197	5417
Julio	5219	5399	5490	5604	5711
Agosto	5265	5301	5366	5357	5419
Septiembre	5007	5063	5200	5133	5318
Octubre	5079	5295	5410	5483	5409
Noviembre	4974	5021	5123	5186	5310
Diciembre	5282	4991	5428	5472	5458

ÍNDICE DE PRECIO DE LA EDUCACIÓN SUPERIOR (1983=100)	NÚMERO DE GRANJAS EN EUA (EN MILES)

1970	39.5	
	42.1	
	44.3	
	46.7	
	49.9	
1975	54.3	2521
	57.9	2497
	61.7	2456
	65.8	2436
	70.6	2437
1980	77.5	2440
	85.9	2440
	94.0	2407
	100.0	2379
	104.7	2334
1985	110.5	2293
	115.6	2250
	120.3	2213
	125.8	2197
	133.1	2171
1990		2140
		2105

PRECIO PROMEDIO DE *WELLHEAD,
GAS NATURAL (CENTAVOS POR
MILLÓN DE PIES CUADRADOS)

PRECIO PROMEDIO ANUAL DE
ACCIONES COMUNES DE
MOODY'S ELECTRIC UTILITY

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
		126.97

PROMEDIO ANUAL DE PAGO EN
TRABAJOS DE FERROCARRIL
(CENTAVOS POR HORA)

1970	403.0
	451.9
	503.4
	550.6
	575.9
1975	631.3
	690.3
	740.0
	840.2
	922.7
1980	1021.0
	1114.0
	1226.0
	1466.8
	1384.6
1985	1429.9
	1474.7
	1511.0
	1564.0
	1576.2
1990	1582.7
	1681.9

RENDIMIENTO TRIMESTRAL PROMEDIO

DE BONOS DE LA TESORERÍA A 90 DÍAS

PROMEDIO DE LA TASA LÍDER

AÑO	TRIMESTRE	(EN PORCENTAJE)	(EN PORCENTAJE)
1985	1	8.18	10.54
	2	7.46	10.20
	3	7.10	9.50
	4	7.17	9.50
1986	1	6.89	9.37
	2	6.13	8.61
	3	5.53	7.85
	4	5.34	7.50
1987	1	5.53	7.50
	2	5.73	8.05
	3	6.03	8.40
	4	6.00	8.87
1988	1	5.76	8.59
	2	6.23	8.78
	3	6.99	9.71
	4	7.70	10.18
1989	1	8.53	10.98
	2	8.44	11.36
	3	7.85	10.66
	4	7.64	10.50
1990	1	7.76	10.04
	2	7.77	10.00
	3	7.49	10.00
	4	7.02	10.00
1991	1	6.05	9.19
	2	5.59	8.67
	3	5.41	8.40
	4	4.58	7.60

MILES DE CARGAS DE PRODUCTOS FORESTALES EN TRANSPORTES DE CARGA

AÑO	1987	1988	1989	1990	1991
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

PRECIOS INDUSTRIALES/PROPORCIÓN DE INGRESOS

AÑO	TRIMESTRE	P/I
1986	1	17.56
	2	18.09
	3	16.47
	4	18.73
1987	1	22.16
	2	21.80
	3	20.82
	4	14.09
1988	1	13.61
	2	13.29
	3	12.50
	4	12.35
1989	1	12.46
	2	13.30
	3	14.73
	4	15.40
1990	1	15.8
	2	17.0
	3	14.8
	4	15.9
1991	1	18.8
	2	20.4
	3	23.7
	4	29.2
1992	1	28.4
	2	26.8
	3	26.0
	4	26.0

BASE DE DATOS

590

NÚMERO DE COMPAÑÍA	VENTAS (MILLONES)	EMPLEADOS (MILES)	GASTO DE CAPITAL (MILLONES)	GASTOS INTANGIBLES (MILLONES)	COSTO DE BIENES VENDIDOS (MILLONES)	GASTOS DE NÓMINA Y RELACIONADOS (MILLONES)	GASTOS DE PUBLICIDAD (MILLONES)	GASTOS EN INVESTIGACIÓN Y DESARROLLO (MILLONES)
1	3221.8008	42.0000	147.9000	30.6000	2285.2007	599.7998	118.3000	28.0000
2	1690.6001	20.9050	93.0000	29.1000	1057.2002	343.2000	114.9000	8.9000
3	2197.2764	39.0000	66.8670	55.8600	1387.0679	661.3997	95.5680	11.1820
4	2357.8206	23.3000	59.5560	69.6080	1743.7952	25.6320	51.9170	8.5000
5	8129.0000	35.0000	297.0000	29.0000	7423.0000	1178.0000	12.8000	9.2530
6	11851.0000	23.0000	394.0000	20.0000	10942.0000	2556.0000	11.6530	14.6000
7	323.8606	3.9000	2.5900	4.2880	233.5300	22.8350	3.5290	30.7320
8	660.4856	8.3780	10.9840	3.3720	582.2649	25.6250	44.9990	64.8730
9	4351.1601	50.9120	102.7080	217.0920	4156.8671	12.8360	66.2640	8.7790
10	985.8357	5.5000	16.6010	29.5900	874.1287	19.5000	112.3860	18.3650
11	3802.5581	39.6000	206.1020	157.3520	2997.2703	518.0000	139.7290	16.4130
12	2576.0464	22.6000	50.6690	47.0790	1885.9053	349.4910	48.8170	9.5000
13	106.0160	28.0000	1.3120	42.0000	84.6590	35.5550	22.9370	8.7330
14	5669.8945	46.8810	103.0000	31.1000	4424.3007	785.0000	141.3000	18.5000
15	319.6570	2.8940	4.5770	2.2090	246.6980	42.8370	87.0000	1.1000
16	511.7217	10.1000	19.5600	27.0000	286.2288	48.9990	1.8700	23.6520
17	884.6189	22.8010	58.0940	33.0000	467.4436	36.5000	16.0350	29.6320
18	166.3750	2.3000	3.9510	5.2890	111.0310	31.0000	4.0230	38.5420
19	59.1310	18.0000	1.1400	14.5000	43.7430	26.3210	90.3250	56.9820
20	136.6970	3.1000	2.0090	18.4930	105.3300	15.8880	46.3000	8.6330
21	767.8799	8.1000	37.4250	18.0560	519.3948	112.1350	21.8470	2.7860
22	61.3280	1.1390	1.3880	26.3250	35.2020	17.3140	2.4270	88.5230
23	445.6387	5.8000	18.9780	12.6000	213.2880	12.1000	62.8060	1.4600
24	2259.6316	16.0270	228.7270	27.3350	1696.3772	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	32.000
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	4337.9140	36.1000	306.0220	101.4290	2344.1631	36.1000	270.2576	16.1100

34	271.0076	2.0780	27.1230	6.5030	134.3790	35.7730	20.2540	87.4830
35	209.4520	2.9980	14.4690	14.6060	176.4890	2.0780	1.8970	714.9990
36	62.4180	3.8000	3.7390	7.6680	34.4700	2.9980	44.0500	121.3300
37	4300.0000	95.5000	412.2886	157.6030	2108.5503	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	3267.9551	31.9790	502.0398	45.6140	2517.7566	754.8977	3.9510	57.7210
44	2745.7439	43.9680	251.0340	16.1110	1638.7969	45.0000	161.2000	108.1480
45	2609.0000	33.2000	248.0001	10.0000	1874.0000	564.0000	18.0000	83.0000
46	1677.6016	11.6440	284.6089	87.4830	1185.9717	24.4530	6.4840	36.1310
47	6887.6210	53.5000	1075.1719	84.0390	4721.9570	1375.7996	44.0700	231.4690
48	10584.1990	132.1400	714.2002	22.6000	7353.5000	3204.2688	93.4000	377.1001
49	2912.7644	45.8540	195.2680	45.6430	2189.5293	879.6548	14.9460	66.0560
50	4309.5820	66.8000	275.3079	67.3120	2913.9036	993.3997	1.6590	40.5470
51	1946.4766	24.4530	121.3300	6.2920	1403.4976	546.0508	35.2020	40.0810
52	9254.1171	151.2000	1431.0906	121.3300	6187.7851	2125.2012	95.9510	334.8057
53	5018.6914	62.8510	479.8997	1.6240	3478.0989	1318.0999	9.2530	144.3000
54	1510.7798	15.3000	207.9320	63.5190	1157.2117	13.9700	27.6660	39.7150
55	1560.0750	22.7000	162.5190	61.9380	1188.9126	18.4340	19.3190	24.7010
56	2794.0000	37.4000	256.0999	7.3000	1928.4988	780.7996	18.3650	70.1000
57	921.3689	13.9700	61.9380	18.4340	597.7000	45.1640	19.2020	22.6500
58	1253.5430	13.0580	66.4310	13.9700	806.6758	236.5000	32.0000	48.6510
59	1328.1138	13.1160	201.1960	31.2730	851.8938	1.1550	31.2730	33.5620
60	1314.6299	27.3460	36.9330	43.0750	569.7327	6.4690	174.4610	42.1160
61	7869.6914	113.3710	687.7998	90.2000	5580.5976	1931.5005	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	1422.4507	16.5000	100.4700	69.8820	1060.5420	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	1445.0166	27.0030	84.1490	99.9080	786.8777	1.9360	39.8650	76.1870
71	3062.6316	49.6190	67.6310	83.1580	1446.5227	668.9910	243.0450	74.5240
72	2450.4285	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
74	6.8030	5.1000	9.5230	1.4590	2.3980	12.2490	.7230	11.9490

(continuación)

BASE DE DATOS (continuación)

NÚMERO DE COMPAÑÍA	VENTAS (MILLONES)	EMPLEADOS (MILES)	GASTO DE CAPITAL (MILLONES)	GASTOS INTANGIBLES (MILLONES)	COSTO DE BIENES VENDIDOS (MILLONES)	GASTOS DE NÓMINA Y RELACIONADOS (MILLONES)	GASTOS DE PUBLICIDAD (MILLONES)	GASTOS EN INVESTIGACIÓN Y DESARROLLO (MILLONES)
75	1852.0896	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	6.730	11.8950
77	1981.4397	28.7000	155.8530	141.2700	668.7720	634.0596	55.2940	161.3500
78	2362.1326	40.7000	110.1000	99.8430	1055.4187	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	1082.4927	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	1112.0386	16.5890	74.5790	44.6610	378.7710	7.3000	91.2150	77.3130
84	1515.8816	37.0000	108.0460	52.3290	758.5320	469.9229	74.5950	61.8300
85	1328.5508	19.9200	44.6810	6.2850	566.2200	323.7090	36.9560	115.5890
86	2878.4956	58.0000	182.2670	348.1426	1247.2339	1.1500	391.6277	85.3970
87	4312.0507	56.6000	169.2950	66.9970	2672.3262	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	2014.7056	31.0000	74.7910		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	1710.4700	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	0.4570	27.1500
100	1451.6687	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	1132.3499	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	1581.8760	20.8580	142.2810	5.8820	1280.1670	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	1497.0076	7.4000	131.9400	6.0210	1098.2969	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	4701.1210	7.5620	353.2830	48.6920	3707.6846	4.9110	3.1490	6.4800
124	1167.8340	2.8100	48.6920	8.4580	1017.6038	2.3500	7.0620	59.2350
125	12298.3980	50.7000	1221.8008	10.4000	9285.7109	1016.5000	13.1160	64.6000
126	439.4727	1.9020	65.1100	39.8650	263.8108	51.1480	27.3460	31.2730
127	29127.0030	108.7000	1897.0005	9.9040	20032.0000	78.7700	16.5000	86.0000
128	1993.6624	8.0000	43.4190	45.7820	1755.5662	3.5730	31.1370	43.0750
129	4660.8945	18.1000	636.1238	28.4900	3675.6895	440.7996	3.4000	11.6000
130	976.4578	8.8280	14.8590	55.2940	879.3516	91.8000	15.3440	90.2000
131	3834.9324	6.6610	316.7156	68.2690	3557.4734	7.4050	2.8250	69.8820
132	9535.7382	42.7800	1107.3838	75.7000	7075.1875	971.0000	6.5480	29.7730
133	657.7776	1.2640	56.1460	36.8860	565.0176	14.4700	27.0030	4.2120
134	100.4570	43.0750	44.0680	7.1610	72.7830	22.0310	49.6110	83.1580
135	60334.5110	130.0000	4186.9296	40.6150	45999.0070	3405.0000	32.6000	290.0000
136	2150.0000	90.2110	311.7000	91.2150	1460.7996	57.4030	1.3040	25.1000
137	18069.0000	58.3000	1680.0000	74.5900	13442.0000	1345.0000	25.4000	88.8250
138	109.7380	69.8870	32.2560	36.9560	97.0130	2.5200	4.9030	6.7300
139	592.7710	3.2520	123.7680	3.8770	420.3206	67.3300	28.7000	1.4590
140	4642.3945	14.3280	353.5999	33.5620	4085.0989	324.0000	40.7000	25.0000
141	2072.4412	11.1480	270.1846	42.1160	1640.8118	1.2400	5.5500	4.9810
142	4509.3828	13.3540	502.2720	1.6500	2838.0845	236.4540	2.0370	12.8000
143	34736.0030	207.7000	1760.7100	2.7200	26053.9060	20.9400	3.7000	16.7160
144	1191.0337	4.2070	255.6150	1.9000	865.6477	82.6730	.2670	99.8430
145	312.7300	4.2120	76.5000	6.4200	452.4130	17.0050	17.9000	52.1780
146	1553.1077	9.1500	343.9539	23.6410	988.8760	185.6600	12.5840	9.7580
147	6997.7734	30.0080	956.1719	11.2330	4886.8125	720.5000	17.1000	58.2460

(continuación)

BASE DE DATOS (continuación)

NÚMERO DE COMPAÑÍA	VENTAS (MILLONES)	EMPLEADOS (MILES)	GASTO DE CAPITAL (MILLONES)	GASTOS INTANGIBLES (MILLONES)	COSTO DE BIENES VENDIDOS (MILLONES)	GASTOS DE NÓMINA Y RELACIONADOS (MILLONES)	GASTOS DE PUBLICIDAD (MILLONES)	GASTOS EN INVESTIGACIÓN Y DESARROLLO (MILLONES)
148	513.1880	5.1420	41.9800	41.9800	375.3599	25.0200	11.3330	11.9490
149	28085.0030	94.8000	2913.0000	32.5600	20632.0000	2344.0000	89.0000	231.0000
150	11062.8980	34.9740	1774.3904	43.0250	8259.7656	1051.0000	16.5890	114.0000
151	23232.4060	37.5750	1049.6729	90.2110	19964.6050	994.0000	37.0000	89.7370
152	14961.5000	47.0110	1744.0364	69.8870	10046.0000	1126.7310	19.9200	80.3250
153	5197.7070	24.1450	762.2510	4.2120	3336.7566	431.9976	7.7130	15.0520
154	7428.2343	33.7210	601.1216	6.7310	5714.3085	9.7320	58.0000	21.0000
155	28607.5030	67.8410	1344.3777	10.4000	24787.6050	1572.7996	56.6000	52.0000
156	87.6100	6.7310	12.7120	39.8650	74.5510	31.5580	31.0000	44.6610
157	1165.6736	3.5310	26.6780	9.9040	1035.7129	6.6000	18.5170	2.4490
158	567.3650	1.5420	97.4910	28.9400	480.5110	23.5230	3.300	52.3290
159	5954.9414	16.2970	732.0000	55.9240	4540.4609	444.8997	31.7000	18.5000
160	368.0940	2.3150	15.0860	2.7160	319.4939	10.6050	3.4800	6.2850
161	751.7327	6.2550	51.1060	13.5380	606.8318	3.5230	6.8000	9.9000
162	895.4087	10.9000	145.5140	9.3840	681.9656	26.3250	39.0000	30.6000
163	1063.2908	16.1790	51.1480	25.7670	746.2820	12.6000	16.6980	14.6320
164	1306.0867	19.3970	78.7700	2.7490	1021.4856	435.2998	23.3000	13.2830
165	140.4440	1.9190	3.5730	55.8600	122.3210	27.3350	35.0000	29.1000
166	4357.2812	52.1400	110.4470	12.0830	3540.9612	1235.0000	3.9000	55.8600
167	263.9048	3.7000	7.4050	27.2080	203.3440	2.8080	8.3780	3.2500
168	6184.8945	94.5000	398.2000	69.6080	5224.0000	2550.0000	50.9120	37.1000
169	257.6509	3.3640	14.4730	7.5700	190.4190	8.3710	5.5000	69.6080
170	50.5150	52.5350	29.1000	29.0000	18.0560	43.5920	39.6000	29.0000
171	419.6470	4.3020	22.0310	20.0000	341.5906	135.6000	22.6000	20.0000
172	1227.4490	20.0000	57.4030	4.2880	999.7520	27.8920	28.0000	9.0000
173	779.3450	8.8000	22.0670	3.3700	678.4258	229.1270	46.8810	4.2880
174	72.1760	1.3000	2.5210	29.5900	50.9650	24.8290	2.8940	3.3700
175	3248.0076	36.0620	263.6167	19.4460	2710.3455	974.3379	10.1000	29.5900
176	921.1270	12.6590	67.3340	10.5250	771.0059	23.5420	22.8010	157.3520
177	711.9827	12.5120	133.3850	45.0790	653.8069	351.4700	2.3000	45.0790
178	72.4110	1.0250	1.2400	42.0000	60.0820	93.0000	18.0000	42.0000
179	297.5686	4.1520	20.9420	.8990	248.7160	123.1000	3.1000	31.1000

180	677.8489	6.0700	17.0050	31.1000	613.3047	169.2570	8.1000	2.2090
181	582.6238	1.4000	25.0290	2.2090	474.3450	66.8670	1.1390	27.0000
182	3750.4109	38.1700	120.8280	27.0000	3240.7886	1132.6216	5.8000	33.0000
183	88.8070	2.3330	9.7320	7.2110	66.6540	59.5500	16.0270	5.2890
184	306.9397	2.8000	31.5880	33.0000	220.4980	2.5900	8.7000	14.5000
185	331.7366	5.2000	6.6000	11.0250	295.3848	10.9840	4.0000	18.4930
186	546.9500	8.9000	23.5230	5.2890	439.0479	16.0010	2.1000	18.0560
187	7.5910	30.6000		14.5000	5.0480	50.6690	5.0000	26.3250
188	3479.4573	41.3940	170.3720	18.4930	3100.5391	1177.5999	1.5760	15.0500
189	485.6138	6.6580	58.6750	3.5250	335.3318	42.0000	93.0000	1.4320
190	123.2280	2.0450	10.6050	5.2550	96.6630	20.9000	66.8670	12.6000
191	488.2327	4.6500	20.4800	1.1111	402.8457	402.8400	77.0101	22.2426
192	100.7820	1.7030	2.4430	1.6800	88.7960	4.0000	21.0000	28.3032
193	165.7970	4.7660	3.2790	88.0003	120.1080	2.0000	4.0008	18.2022
194	274.8440	3.5500	21.7900	2.9530	213.1860	3.0000	3.7521	24.2628
195	11049.5000	166.8480	667.7998	55.5000	9955.3984	4485.1953	22.0007	52.5000
196	1154.8477	14.4190	32.2360	4.0800	1037.4727	424.4556	21.1234	30.3234
197	578.7107	11.4920	26.3000	8.0141	433.8230	1.0111	12.3456	5.3300
198	124.5440	1.8000	4.6280	1.9850	101.5300	23.6630	78.9101	36.3840
199	3711.2029	63.4000	303.3838	4.5720	2729.9280	22.0222	91.0111	33.0000
200	124.8600	2.0000	5.2240	2.3200	79.7770	51.0000	21.3141	2.6500
201	2466.0000	26.8650	161.7000	2.0202	2028.7996	18.4021	3.2000	14.9000
202	2829.2991	36.2000	156.8000	27.1000	2261.0000	930.2000	51.1617	25.2000
203	814.8196	14.8000	48.5520	16.0111	622.9507	204.9000	18.1920	1.4150
204	4051.7996	46.0000	349.5999	2.6000	3036.5999	1215.2996	21.2223	56.6000
205	67.0390	28.0000	3.5010	2.5170	54.9070	66.5620	24.2526	42.4446
206	240.5670	4.0000	5.5670	1.3220	184.1350	61.6900	2.5860	3.0470
207	45.2140	2.0000	1.4110	18.1010	38.0970	62.3201	27.2829	48.5052
208	69.9520	81.0000	33.3333	8.0033	65.4570	52.3302	30.3132	54.5658
209	54.5490	1.1270	1.7720	17.7200	42.5990	42.4444	33.3435	60.6264
210	317.4480	5.7840	12.6650	11.0330	254.1990	80.1010	36.3738	66.6870
211	847.9927	24.0000	85.0240	19.7930	664.9578	34.1021	39.4041	10.4000
212	467.9546	4.8450	13.1650	2.3810	400.5806	4.0999	42.4344	1.0011
213	126.6750	14.0007	7.7490	14.1441	109.6830	50.6410	45.4647	1.0022
214	85.7290	49.0000	2.1610	49.4949	72.8400	9.9901	48.4950	1.0033
215	680.7666	8.2200	19.2340	77.7878	578.8528	9.8175	51.5253	1.3090
216	211.3230	1.5670	4.8350	15.6180	171.4130	65.0000	54.5556	1.8201
217	254.3030	3.1000	2.7620	2.3570	205.8410	42.4381	57.5859	2.0880
218	1396.8108	29.4160	79.9820	28.2626	1000.2886	3.8107	16.1580	3.4510
219	3981.0000	52.9000	188.3000	70.3000	3120.5999	1085.7996	75.8000	37.5000

(continuación)

BASE DE DATOS (continuación)

NÚMERO DE COMPAÑÍA	VENTAS (MILLONES)	EMPLEADOS (MILES)	GASTO DE CAPITAL (MILLONES)	GASTOS INTANGIBLES (MILLONES)	COSTO DE BIENES VENDIDOS (MILLONES)	GASTOS DE NÓMINA Y RELACIONADOS (MILLONES)	GASTOS DE PUBLICIDAD (MILLONES)	GASTOS EN INVESTIGACIÓN Y DESARROLLO (MILLONES)
220	3943.0990	56.5320	259.5000	49.9000	3352.3008	1275.7002	60.6162	42.3000
221	1260.2349	17.2880	103.0320	11.4810	1055.9436	12.0000	63.6465	2.1133
222	973.2527	9.8850	25.4530	5.5580	848.7227	4.0877	66.6768	3.3210
223	19.9060	18.0002	5.6666	1.4100	16.5170	3.3876	69.7071	4.2242
224	66.8260	1.3200	6.1110	88.1388	48.9480	4.5222	72.7374	5.6210
225	178.7460	2.1980	5.5430	138.0000	138.5690	43.4350	75.7677	6.2155
226	26.7510	1.0560	8.8888	211.0113	17.9930	18.1111	78.7980	7.2102
227	20.5750	43.1111	7.7777	82.1003	13.9720	14.2222	81.8283	8.9712
228	51.5960	18.5216	1.6940	1.1620	38.8190	88.9922	81.0077	24.2601
229	106.1150	2.6000	4.6850	9.9210	64.0500	12.4243	77.0222	23.2810
230	8.5160	14.2421	12.0818	12.1402	5.9500	7.8246	22.4443	24.8588
231	308.8916	5.7000	15.8370	13.1402	144.7340	42.4444	47.8877	2.7060
232	753.8069	16.8750	37.4620	3.6210	491.1160	210.0050	16.4370	4.9340
233	41.2960	1.1080	2.5820	12.1213	28.1320	81.8118	12.5456	24.5293
234	145.6840	3.4190	13.3250	1.0087	105.1630	51.7100	51.8196	1.8480
235	51.3130	1.0000	1.5700	8.0025	35.9730	43.4400	21.4231	59.6085
236	21.4070	12.5358	18.7842	5.5554	12.9550	12.8888	37.8286	64.8821
237	585.6597	8.2000	56.0530	80.9960	359.8350	77.9999	13.6920	8.9610
238	516.7239	10.3000	17.9320	9.3610	376.4170	1.1007	5.6670	5.6000
239	316.8147	7.0000	3.9360	12.1314	267.2456	2.0008	86.8686	76.7686
240	509.7000	10.0000	27.0360	15.1617	375.3457	179.9240	85.8686	3.6080

241	341.3887	7.1270	7.1570	8.1819	287.6907	9.0007	86.8888	86.7795
242	33.0660	1.0872	1.9540	9.2021	24.0720	12.7210	83.1111	95.9594
243	200.5920	4.0000	5.3320	20.0290	153.5480	7.6660	82.2222	94.9388
244	184.5810	4.0500	7.2780	10.3570	142.7160	8.7770	22.6665	1.0790
245	217.7520	4.0880	7.3840	10.1146	179.1020	78.3910	44.6621	89.9012
246	386.8118	7.4040	18.4880	47.1213	302.5586	2.9990	18.1716	3.8620
247	69.1530	12.1212	1.6190	48.1415	54.4310	11.3410	15.1413	13.8125
248	81.4650	1.6220	4.1890	16.4950	70.5080	4.4555	12.1110	47.8552
249	329.5518	6.0040	12.2520	8.0540	269.6377	12.1417	9.8765	51.9185
250	36.3870	133.0000	12.7246	51.5355	27.7690	21.8283	4.3210	54.3321
251	344.7937	7.5000	24.7400	57.5982	205.0610	92.9395	8.1234	4.8200
252	22.8030	84.1000	2.1060	83.4952	10.6830	96.9899	5.6788	43.8388
253	196.3030	5.4660	5.9730	99.9242	142.1520	97.9294	12.4582	2.2710
254	31.5660	13.7124	8.1264	10.1115	22.3750	95.0092	14.5220	66.7274
255	108.8580	1.7000	1.2870	92.4445	45.9130	92.6666	1.4330	53.5422
256	83.6260	1.2320	4.1220	55.6677	45.0950	92.5555	13.5620	22.5673
257	390.8726	6.1660	17.3310	40.5880	296.8577	58.2130	18.0000	10.0000
258	363.9839	7.0160	11.2700	11.5610	234.6320	2.7310	6.4860	86.0000
259	52.2620	.4420	5.1030	1.1500	43.5110	12.1440	44.0700	16.0000
260	228.6110	5.6500	1.8370	41.5600	161.4700	20.5400	14.9460	3.0730
261	60.8250	1.5000	1.4910	45.3100	41.6820	1.8970	1.6590	30.7300
262	16.6890	40.5000	57.6000	9.8540	9.8450	18.3700	35.2020	63.5300
263	39.8290	62.1000	3.9900	1.5610	32.6580	4.9080	9.2530	30.0000
264	28.9020	93.4000	1.1040	36.5000	23.1410	2.5900	18.3650	10.0000
265	8.7410	27.0000	55.6000	32.1400	6.3700	59.3100	27.6600	56.6660
266	61.9446	7.0000	35.0470	43.2000	432.3777	160.6660	19.2020	6.3530

Índice

- Abacus Scientific Software Company, 10
- Aburdene, P., 527
- Ace Personnel Department (caso de estudio), 243-44
- Add-a-Stat, 10
- Administradores:
 - Véase también* Pronóstico empresarial, aplicación en la administración
 - ventas sobre pronósticos, 542
 - y el enfoque Box-Jenkins, 479, 480
 - y el monitoreo de pronósticos, 536-539
 - y el proceso de pronóstico, 533-536
 - y la necesidad de los pronósticos, 2-3, 73-74, 176-177, 289, 290-91, 318, 351-353, 400-401, 478-480, 533
 - y la responsabilidad del pronóstico, 540-541
- Ajuste de la curva S, 118
- Alcam Electronics (caso de estudio), 56-58
- American Statistics Index*, 67
- Análisis de regresión simple vs. múltiple, 255
 - aplicaciones en la administración, 230-231. *Véase también* Análisis de regresión
- Análisis de regresión, 203-233, 255-291, 318, 374-402, 534
 - aplicación en la administración, 230-231, 289-291
 - coeficiente de determinación en, 213-216, 232
 - error estándar del pronóstico para, 210-232
 - estimación del error estándar para, 208-209, 232
 - línea de regresión definida, 205
 - método de mínimos cuadrados en, 205-207, 232
 - muestra de ecuación de regresión para, 206
 - muestra de línea de regresión en, 207
 - por cambios de porcentaje, 383, 387-392
 - predicción de y , 209-213
 - primeras diferencias en, 396-398
 - problemas con la autocorrelación en, 377-380
 - prueba de hipótesis en, 220, 226
 - regresión por pasos para, 279-284
 - residuales en, 207, 217
 - selección de la ecuación de regresión para, 274-284
 - sobre datos de series de tiempo, 374-401
 - suma de residuales cuadrados para, 217-219
 - suma total de cuadrados para, 217, 233
 - suposiciones en, 211, 226, 261, 271
 - todas las regresiones posibles, procedimiento en, 277-279
 - Véase también* Correlación, Regresión Múltiple, Regresión de datos de series de tiempo
 - y las transformaciones de variables, 226-230
- Análisis de series de tiempo, 98-100, 118, 119, 318-353, 379-402, 479, 533
 - aplicaciones en la administración, 351-353
 - Census II, método de descomposición de, 350-351
 - componente aleatorio en, 322, 353
 - datos ajustados en forma estacional en, 346-347, 353
 - descomposición en, 98, 99, 319-20, 349
 - estacionarias, 432
 - heteroscedasticidad en, 374-377
 - índice de precios en, 320-322
 - juicios en, 63, 319
 - método de mínimos cuadrados en, 324-327, 393-396
 - modelo multiplicativo en, 254-295
 - regresión de, 374-401
 - tendencia en, 319, 322-330
 - tendencia estacional en, 342-346
 - tendencia no lineal en, 285-287
 - valores de tendencia mensuales o trimestrales en, 300-304, 311
 - variación cíclica en, 319, 330-338
 - variación estacional en, 320-338, 347
 - variaciones cíclicas e irregulares de corto plazo en, 347-349
 - y deflación de precios, 321-322
 - y el análisis, 336
 - y el modelo aditivo, 284-287
 - y la correlación serial, 377-400
 - y la curva de Gompertz, 329-330
 - y la síntesis, 337
 - y la suma total de cuadrados, 233
 - y los indicadores empresariales, 332-335, 353
- Anderson, Parente, Janet, 523
- Applied Decisions Systems, 9
- Argyl Food Products (caso de estudio), 78-81
- Armstrong, J. Scott, 524
- Atenuación exponencial cuadrática, 118
- Atenuación exponencial de Brown, 116, 163-168

- Atenuación exponencial doble (Método de Brown), 163-168
 - fórmula para la, 164-165, 178
- Atenuación lineal exponencial, 118, 163-168
- Audits and Surveys, Inc., 68
- Autobox, 9
- Autocast II, 9
- Autocorrelación parcial, 436-437
- Autocorrelación, 98, 100-114, 122-125, 377-400
 - fórmulas para, 100, 101, 125, 126
 - parcial, 436-437
 - problemas de, 377-380
 - prueba de Durbin-Watson para, 380-383, 578-579
 - solución de problemas para, 383-384, 387-400
 - y el análisis de regresión, 211
 - y el enfoque iterativo, 398-400
 - y la regresión de datos de series de tiempo, 377-400
 - y la técnica Box-Jenkins, 439-465
 - y la tendencia, 108
 - y los modelos de autorregresión, 392-393, 436-437
 - y mínimos cuadrados generalizados, 393-396
 - y primeras diferencias, 396-398
- Autocorrelación, coeficiente de, *véase* Coeficiente de autocorrelación
- Bayes, teorema de, 530-533
- Beaumont, *et al.*, 10
- Berhard, Arnold and Company, 68
- Bondad del ajuste, prueba de, 37-39. *Véase también* Prueba de Ji-cuadrada
- Boundary Electronics (caso de estudio), 544-545
- Box-Jenkins, técnica, 1, 108, 115, 117, 118, 431-480, 534
 - aplicación de la metodología en la, 439-465
 - aplicación en la administración, 479-480
 - desventajas de la, 477-478
 - etapa 1 (identificación del modelo), de la, 439-440
 - etapa 2 (estimación y prueba del modelo), de la, 440-441
 - etapa 3 (pronóstico con), de la, 441-465
 - modelo autorregresivo de promedio móvil (ARIMA), en la, 439-465
 - modelo de autorregresión (AR) en la, 436-437
 - modelo de promedio móvil (MA), en la, 438
 - y el análisis estacional, 465-478
 - y las autocorrelaciones parciales, 436-437
- Box-Pierce, prueba de, 453, 480
- Brown, Método de (atenuación exponencial doble), 163-168
- Bureau of Economic Activity, 64
- Burns, Arthur F., 533
- Busby Associates (caso de estudio), 545-550
- Business and Economic Forecasting*, 9
- Business Forecast Systems, Inc., 9, 10
- Business Week*, 289
- Butcher Products, Inc. (caso de estudio), 241-243
- Cambios de porcentaje, regresión de, 383-387-392
- Capacitación (en redes neurales), 525
- Carbone, Robert, 9
- Casos de estudio:
 - Ace Personal Department, 234-244
 - Alcam Electronics, 55-57
 - Argyl Food Products, 76-79
 - Boundary Electronics, 544-545
 - Busby Associates, 545-50
 - Butcher Products, Inc., 241-243
 - Consumer Credit Counseling, 12, 80-82, 135-36, 186-87, 245-246, 307-308, 365, 416-417, 490-491, 552
 - Demanda diaria de permisos para pesca de trucha, 495-507
 - El Mercado de Valores, 299-302
 - Estudio de factibilidad de mercado, 82-89
 - Golden Gardens Restaurant, 545
 - Índice de actividad empresarial para el Condado de Spokane, 409-415
 - La Compañía de Su Elección, 408-409
 - Lydia E. Pinkham Medicine Company, 491-494, 552-555
 - Mr. Tux, 11-12, 57-58, 80, 134-135, 185-186, 244-245, 305-306, 360-365, 415-416, 488-490, 550-552
 - Murphy Brothers Furniture, 131-134
 - Proyección de ingreso quinquenal para Downtown Radiology, 187-196
 - Small Engine Doctor, 358-360
 - Solar Alternative Company, 184-185
 - Solomon's Jewelry, 79
 - Tiger Transports, 239-241
 - Ventas de restaurante, 302-305, 487-488
- Causalidad vs. correlación, 46
- Censo, datos del, 64-65
- Census II, método de descomposición, 117, 118, 350-351
- CIS/Index to Publications of the United States Government*, 65
- Cochrane, 399-402
- Coeficiente de autocorrelación, 100-114, 122, 125, 126, 432
- Coeficiente de correlación, 43-47, 48, 50, 216, 220
 - con autocorrelación, 100-112
 - fórmula para el, 45, 50
 - y el coeficiente de determinación, 216
 - y la relación negativa, 44
- Coeficiente de determinación, 213-216
 - fórmula para el, 45, 48
 - vs. coeficiente de correlación, 216
- Coeficiente de regresión, 223, 260-61, 291, 396
- Coeficientes de regresión, 223, 260-261, 291, 396
- Coeficientes netos de regresión, 260-261, 291
- Colinealidad, 256, 272-272, 291
- Combinación de pronósticos, 524, 525
- Componente aleatorio (en análisis de series de tiempo), 99-100, 125, 320, 353
- Componente cíclico, definición de, 99, 117, 125, 319
- Componente estacional o series, 99, 111-114, 116-117, 125
 - atenuación exponencial ajustada para, 172-176
 - en el análisis de series de tiempo, 320, 338-347, 342-346
 - y la autocorrelación, 11-114
 - y la regresión, 284-87, 291
 - y la técnica de Box-Jenkins, 465-478
 - y los métodos no formales, 150
- Compuserve, Inc., 66
- Computat, 66
- Computadoras:
 - como cambio tecnológico, 118, 177, 478, 521-541
 - y la organización de la función de pronóstico, 3
 - y la relevancia de los datos, 3
- Computer Software Consultants, Inc., 10
- Conexión compatible, 526
- Confiabilidad (de los datos), 61
- Constante de atenuación, 159-160
- Construcción de modelos, 6-7
- Consultores, pronóstico mediante, 542
- Contingencia, tabla de prueba de, *véase* Pruebas de Ji-cuadrada
- Control de producción, 48
- Correlación serial de primer orden, 377-378
 - estimación de para, 398-400
- Correlación serial, 277, 400-401
 - prueba de Durbin-Watson para, 380-383
 - soluciones para, 383-384, 387-392, 392-393, 393-396
 - y los datos de series de tiempo, 377-400
- Correlación, 42, 89, 203, 318. *Véase también* Regresión múltiple,
 - Análisis de regresión, Análisis de series de tiempo
 - aplicación en la administración, 48, 230
 - coeficiente de, 43-47, 48, 204, 398
 - fórmula para el, 45, 50
 - diagramas de dispersión de, 40-42, 205
 - vs. causalidad, 46
 - y autocorrelación, 98, 100-114, 122-124, 125, 377-400. *Véase también* Autocorrelación

- Correlogramas, 103, 125
- Costos de la pronóstico, 541-542
- County and City Data Book*, 65
- Creative Computing*, "Choosing and Using Business Forecasting Software", 10, 10n
- Cuestionario por correo, 67
- Cuestionario por correo, 69
- Curvas de crecimiento, Gompertz (logística), 116, 118, 329, 330, 522
- Chase Econometrics, 9
- Daisy, 9
- Daniells, Lorna, 65n
- Datos cualitativos (no numéricos), 4
 - pruebas de Ji-cuadrada para, 534
 - técnicas de pronóstico en, 97, 117, 520-527
 - variables de prueba para, 266
- Datos del Gobierno de Estados Unidos de América, 63, 64, 65, 73, 74
- Datos estacionales ajustados, 346-349, 353
- Datos no numéricos, véase Datos cualitativos
- Datos primarios, 67-73-75
 - encuesta, 67-69
- Datos propios, 66-68
- Datos secundarios, 53, 64-67, 69-73
- Datos, fuentes de:
 - bases de datos computarizadas, 64
 - comerciales, 66
 - fuentes externas, 64-66
 - primarios, 53-54, 63, 74, 65
 - propios, 65-66
 - publicaciones, 64, 65, 66
 - secundarios, 65, 66-69, 71-75
 - sistemas como MIS, 542
 - tipos de, 62-63, 344-345
- Datos, recolección de, 6, 61, 73, 540, 563
- Datos, reducción o condensación de, 6, 540
- Datos:
 - acumulación de, 6, 61, 73, 540, 563
 - aplicabilidad, 118
 - consistencia, 61
 - criterios de utilidad, 611
 - encuesta, 67-69. Véase también Encuestas
 - integridad, 73, 118
 - oportunidad, 61
 - precisión, 61
 - relevancia, 61, 73
 - series de tiempo, fuentes de, 62, 318
 - validez, 73
- Decision Support Modeling, 9
- Decisiones, teoría de, 530-533
- Deflación de precios, 321-322, 353
- Demanda diaria de permisos para pesca de trucha (caso de estudio), 495-507
- Demografía, datos sobre, 64-65
- Derivación de correlación, 559
- Derivación de mínimos cuadrados, 560
- Derivaciones:
 - correlación, 559
 - mínimos cuadrados, 560
- Descomposición en el análisis de series de tiempo, 98, 99, 319-320, 349
 - clásica, 99, 117, 118
 - método Census II para, 117, 118, 350-351
- Desestacionalización de datos, 176, 346-347
- Desfase, indicadores de, 335
- Desfase, variables, 392. Véase también Modelos autorregresivos
- Desviación estándar, 23
- Desviación media absoluta (MAD), 120, 122, 126
- Desviaciones, 263n
- Diagrama de árbol, 530-531
- Diagramas de dispersión, 39-43, 48
 - línea de regresión en, 180-185
- Diferenciación (primera diferencia), 108-110, 384, 396-398, 402, 439
- Distribución binomial, 22-28, 47
 - aplicación en la administración, 47
 - fórmula para, 22, 49
 - reglas para, 27
 - términos individuales de, 568-570
- Distribución continua, 27
- Distribución de Poisson, 39, 48
- Distribución normal, 28, 47, 209, 571
 - tabla de áreas para, 571
 - y la prueba de Ji-cuadrada, 39
- Distribuciones de probabilidad, 26-39
- Distribuciones:
 - binomial, 27-28, 47, 568-570
 - de Poisson, 39, 48
 - de probabilidad, 21-30
 - muestral, 30
 - normal, 28, 47, 571
 - t, 30, 572
- Dólares constantes, 320. Véase también Poder de compra y Deflación de precios
- Dow Jones News Retrieval Service, 66
- Draper and Smith, 279, 279n
- Dun and Bradstreet, Inc., 66
- Durbin-Watson, prueba de, 380-383, 401, 578-579
- Easy Forecaster Plus I o II, 9
- Econometric Models and Economic Forecasts*, 398
- Econometrics Software, Inc., 9
- Econometrics Toolkit (ET), 9
- Economic Report of the President*, 65
- Economics Software Program (ESP), 9
- Ecuación de regresión, 205, 206, 232
- Ecuaciones normales, 259
- Ejemplos de cómputo:
 - para calcular autocorrelaciones y desarrollar correlogramas, 103, 104, 107, 108, 114, 123, 124, 132, 143
 - para calcular diferencias, 109, 143, 397-398
 - para calcular índices estacionales, 371-373
 - para calcular un coeficiente de correlación, 242-43, 246-49, 250-51
 - para desviación estándar, 59-60
 - para determinar la estacionalidad, 114-145
 - para el análisis de regresión con estacionalidad, 286-287
 - para el análisis de regresión, 224, 240, 242-243, 246-249, 250-252, 252-253
 - para el análisis de series de tiempo, 326-331
 - para el procedimiento de regresión por pasos, 314-315
 - para encontrar índices cíclicos, 331
 - para generar números aleatorios, 95-96
 - para imprimir la relación X-Y, 227, 228, 230
 - para imprimir un diagrama de dispersión, 267
 - para la atenuación exponencial doble, 196-199, 199-202
 - para la media, 58-59
 - para la regresión de datos de series de tiempo, 326, 376, 385-386, 387
 - para la varianza, 58-59
 - para localizar una variable perdida, 385-386, 417-421
 - para mostrar colinealidad, 274
 - para obtener residuales, 167, 306
 - para regresión múltiple, 265, 269, 301, 302, 308-312, 312-314-315-317
 - para transformar variables, 227
- Véase también ForeCalc, Forecaster Plus, Lotus, Minitab, SAS, Spreadsheet Forecaster, TSP
- y el modelo autorregresivo, 394-421-424
- y la administración de datos, 89-93
- y la correlación serial, 385-386, 387, 388-390, 392, 428-430

- y la descomposición Censal X11, 366-367, 367-371
- y las técnicas Box-Jenkins, 444, 445-453, 456-458, 460-462, 464-477
- Ejemplos de conveniencia, 72
- El Mercado de Valores (caso de estudio), 299-302
- Employment and Earnings*, 65
- Empresarial, ciclo e indicadores empresariales, 332-335
- Encuestas:
 - datos para, 67-69
 - diseño del muestreo para, 68-69
 - tamaño de la muestra para, 69
- Enfoque iterativo, 384, 398-400, 431
- Entra basura-sale basura (GIGO), 61
- Entrevista personal, 68
- Entrevista telefónica, 68
- Error del pronóstico (error estándar del pronóstico), 210, 232. *Véase*
- Error estándar de la distribución muestral, 30
- Error estándar de la estimación, 208-209, 232, 263-269, 291
 - en la regresión múltiple, 263-264, 291
- Error estándar del coeficiente de regresión, 223
- Error estándar del pronóstico, 209
- Error medio cuadrado (MSE), 120, 122, 126, 163, 167, 171, 175
- Error, términos de, 119-122
 - en la prueba de Durbin-Watson, 380-381
 - en la técnica Box-Jenkins, 442
 - Véase también* Residuales
- Errores de pronóstico. *véase* Errores en el pronóstico
- Errores en el pronóstico, 7, 119-122, 126, 147-149, 152, 207, 210, 374, 378
 - error de especificación en el modelo (en la regresión), 384-387
 - medición de, 119-122
 - rastreo de, 161-163, 538
 - suposición de independencia en la regresión, 211
 - tipos de, 33
 - y la fórmula para el error de pronóstico (residual), 119, 126
- ESP (Economics Software Program), 9
- Estadística de Ji-cuadrada (Prueba de la Bondad del Ajuste), 37-39, 48, 534
 - como prueba Box-Pierce, 441
 - fórmula para, 37-49
 - y los grados de libertad, 38-49
- Estadística descriptiva, 22-25, 47, 48
 - aplicación en la administración, 22, 47
- Estadística descriptiva, 22-25, 47-48
- Estadísticas de ingreso, 65
- Estimación, 31-33, 47
- Estudio de factibilidad de mercado (caso de estudio), 82-89
- Estudios de segmentación de mercado, análisis de regresión en, 289-290
- Expert Choice, 9
- Express Easycast, 9
- Extrapolación de modelos, 6, 7, 62, 540
- F, estadística y prueba, 222, 233, 281
 - tabla de distribución para, 576-577
- Federal Reserve Bulletin*, 62, 65, 318
- Filtración adaptativa, 118, 479
- Fleming Software, 10
- Fore Calc, 10
 - ejemplo de computadora para la atenuación exponencial, 199-202
- Forecast Master, 9
- Forecast Plus, 9
 - administración de datos, 91-95
 - atenuación exponencial doble, 196-199
 - autocorrelaciones y correlogramas, 136-139
 - Box Jenkins, 507-511
 - descomposición Censal X-11, 366-367
 - ejemplos de computadora para:
 - introducción a, 13-14
 - regresión múltiple, 308-312
 - regresión, 246-249
- Forecast Pro, 9
- Forecast!, 10
- Forecast/DSS, 10
- Forecaster, The*, 10
- Forecasting Methods and Applications*, 351
- Forecasting Studies*, 65
- Fortune Directory, The*, 66
- Framework, 10
- Futurcast: The Total Forecasting System, 9
- Futurion Associates, Inc., 9
- Generación de escenarios, 523-524
- Gobierno de los Estados Unidos de América, datos del, 63, 64, 65, 73, 74
- Golden Gardens Restaurant (caso de estudio), 545
- Gompertz, curvas de crecimiento de, 116, 118, 329-330, 522
- Grado de relación, 43
- Grados de libertad, 24-25, 48
 - y los valores t ,
- Gráficas de proporción, 562-566
- Gráficas semilogarítmicas, 562-566
- Graph N' Calc, 9
- Guide to Consumer Markets*, 66
- Gwilym Jenkins & Partners, Ltd., 9
- Handbook of Labor Statistics, 65
- Heteroscedasticidad, 271-272, 291, 374-377
- Hewlett Packard, 9
- Hipótesis nula, 33
- Indicadores cíclicos, 332-335
 - problemas con, 335
- Indicadores coincidentes, 335
- Indicadores económicos (empresariales), 117, 332-335, 346, 353
- Indicadores empresariales, 117, 332-335, 346, 353. *Véase también*
- Indicadores cíclicos e Indicadores económicos
- Indicadores estadísticos, 333-335
- Indicadores principales, 118, 334
- Índice de actividad empresarial para el Condado de Spokane (caso de estudio), 409-415
- Índices de precios, en el análisis de series de tiempo, 320-322
- Inferencia estadística, 21
- Inflación y el análisis de series de tiempo, 320-322
- Institute of Business Forecasting, 9
- International Journal of Forecasting*, 2n, 5, 10, 524, 524n
- Intervalo de confianza (estimación del intervalo), 31-33, 48
 - con autocorrelación, 379
 - en el modelo Box Jenkins, 439
 - en la regresión, 211, 212, 232
 - fórmula para una proporción en, 31, 49
- Intervalos, estimación de 31-33, 48, 209-213
- Intervalos, pronóstico de, 209-213
- Intex Solutions, Inc., 10
- Investigación de mercado, 72-73, 74. *Véase también* Encuestas, pruebas de mercado en
- Investigación exploratoria, 70
- Journal of Forecasting, 8n, 10n
- Juicio intuitivo, *véase* Juicio personal
- Juicios personales (subjetivos, intuitivos). *Véase* Pronóstico empresarial, aspectos de juicio en
- Kcating, K., 10n
- Kinnear, T.C. y J.R. Taylor, 66n
- La compañía de su elección (caso de estudio), 408-409
- Levenbach Associates, 9, 10

- Línea de regresión, 205, 232
 Lisa 7/7 Office System, 10
 Lotus 1-2-3, 10, 14
 Lotus, ejemplos de computadora para:
 autocorrelación, 139-141
 introducción, 14-16, 96-97
 Lotus-FORMAN, 10
 modelo autorregresivo, 421-424
 regresión múltiple, 312-314
 Lydia E. Pinkham Medicine Company (caso de estudio), 491-494, 552-555
- Macropredicción, 4-5
 Mahmoud, et al., 8,9
 Mahoud, E., 524n
 Makridakis, Spyros, 2n, 9, 351, 477, 521n
 Management Decision Systems, Inc., 9
 Market Facts, Inc., 66
 Market Research Corporation of America, 66
 Matriz de correlación, 256-257
 McGee, V.E., 351
 Media (estadística), 22, 47, 49, 151
 Medición numérica, véase Técnicas de pronóstico cuantitativas
 Megatrends 2000, 527, 543
 Megatrends, 527, 543
 Mentiras y muestras de juicio, 72
 Mercados de prueba, 72-73
 Método de Holt de atenuación lineal exponencial de dos parámetros, 116, 168-172, 179
 Método de los mínimos cuadrados, 205-207, 232, 324-327, 402,
 Véase también Enfoque iterativo
 con correlación serial, 393-396
 con el procedimiento de Cochran y Orcutt, 399-400
 en el análisis de regresión, 205-207
 en el análisis de series de tiempo, 324-327, 393-396
 en la regresión múltiple, 258
 fórmula para el, 206
 Método de mínimos cuadrados no lineales, 436-437
 Método Delphi y Rand Corporation, 523
 Método Delphi, 523
 Método residual, 330-332
 Métodos de atenuación exponencial, 115, 116, 118, 146, 158-176, 177, 479-535
 ajuste por tendencia (Método de Holt), 118, 168-172
 ajuste por tendencia y variación estacional (Modelo de Winter), 172-176
 aplicación de los, 115, 116, 118, 177
 aplicación en la administración, 177
 atenuación exponencial doble (Método de Brown), 163-168
 fórmulas para, 158, 178
 para pronóstico de corto plazo, 118, 177, 479
 y el rastreo, 161-163
 Métodos de atenuación, véase Métodos de atenuación exponencial
 Métodos de promedios de pronósticos, 22, 47, 115, 146, 151-158, 176, 177
 Métodos no formales, 115, 118, 146, 147-150, 177, 178
 aplicación en la administración, 177
 e indicadores estadísticos, 333-334
 método Census II, por, 350-351
 National Bureau of Economic Research (NBER), 333-334, 350
 Métodos para seleccionar la ecuación de regresión, 274-284
 regresión por pasos, 279-284
 todas las regresiones posibles, 277-279
- Micranal, 9
 Micro-BJ, 9
 Micro-Macro, entomo, 4
 Microcomputadoras, véase Computadoras
 Ming Telecomputing, Inc., 10
 Mínimos cuadrados generalizados, 393-396
- Mínimos cuadrados, método, véase Método de mínimos cuadrados
 Minitab, 8, 9
 Minitab, ejemplos de computadora para:
 análisis de regresión con estacionalidad, 286-287
 análisis de regresión, 242-243, 250-252
 análisis de series de tiempo, 326, 331, 385-87, 388-90
 análisis residual, 167
 autocorrelación, cálculo y desarrollo de correlogramas, 104, 141-143, 197, 108
 correlación serial, 385-387, 388-390
 desviación estándar, cálculo de la, 38-59
 diferenciación de datos, 109, 143, 397-398
 estacionalidad de muestreo, 112
 generación de variable perdida, 385-386, 387
 índice cíclico, para localizar, 331
 índices estacionales, cálculo de, 371-373
 introducción a, 16, 18
 media, cálculo de la, 58-59
 números aleatorios, generación de, 95-96
 para la regresión por pasos, 314-315
 procedimiento de Box-Jenkins, 445-452, 456-458, 472-477, 511-514
 regresión múltiple, 269
 relación *X-Y*, 227, 228-230
 transformación de variables, 227
 varianza, cálculo de la, 58-59
- Mitchell, Wesley, 333
 Modelo aditivo de series de tiempo, 284-87
 Modelo multiplicativo, 284, 349, 353
 Modelos autorregresivos de promedios móviles (ARIMA), 431, 432, 435, 439-478
 Modelos de autorregresión (AR), 384, 392-393, 401, 433, 436-437
 aplicación de, 439-465
 aplicación en la administración, 479
 como técnica de Box-Jenkins, 436-437
 Modelos de promedio móvil (MA), 118, 152-158, 176, 177, 178, 434, 438
 aplicación en la administración, 177
 Modelos econométricos, 117, 118, 187-288
 Monthly Labor Review, 62, 318
 Moody's Investors Service, 68
 Moore, Geoffrey H., 334n 336
 Morgenthau, Henry, 333
 Muestra, 22,23
 aleatoria, 30, 70-72
 bloque, 71-72
 coeficiente de correlación de, 44-45
 conveniencia (cuota), 72
 desviación estándar de, 23, 49
 estratificada, 71
 media de, 22, 23, 49
 mercados de prueba, 72-73
 no aleatoria, 72
 varianza de, 24, 49
- Muestra, tamaño de la (en encuestas), 70
 Muestrales, distribuciones, 30, 48
 error de muestreo de, 30
 Muestras a juicio, 72
 Muestras no aleatorias, 72
 Muestreo aleatorio simple, 71-72
 Muestreo aleatorio, 70-72, 75
 Muestreo de grupos, 71-72
 Muestreo estratificado, 71
 Muestreo por cuotas, 72
 Muestreo sistemático, 71
 Muestreo, diseño para encuesta, 68-69
 Multifit, 10
 Multiple Time Series Program (MTS), 9
 Murphy Brothers Furniture (caso de estudio), 131-134

- Naisbitt, John, 527, 543
National and Regional Economic Projection Series, The, 66
 National Family Opinion, 66
 Netter and Wasserman, 259
 News Retrieval Service, 66
 Nielsen, A.C., Company, 66
 Nivel de significancia, 34
 NPD Research, Inc., 66
 Nuametrics, 9
 Número índice, para la variación estacional, 339-342, 353
 Números aleatorios, 70-575
- Orcutt, 399-402
 Oxford University Press, 9
- Paquetes de cómputo para pronóstico, 8-10
 Parente, Frederick J., 523
PC Magazine, 10, 10n
 Pendiente, en la relación lineal, 42-43
 Pindyck, R.S., 398
 Población (estadística), 22, 23
 coeficiente de correlación de, 44, 50
 desviación estándar de, 23, 49
 media de, 22, 23, 49
 parámetros de, 23
 varianza de, 24, 49
 Poder de compra, 321
 Políticas de pronóstico, 11
 Porcentaje medio de error (MPE), 120-121, 122, 126, 172
 Porcentaje medio de error absoluto (MAPE) 120, 122, 126, 171
 Predicasts, Inc., 66
 Primera diferencia, 108-110, 384, 396-398, 402, 439
 Pro* Cast, 9
 Procedimiento de regresión por pasos, 279-284
 Programas de cómputo, 8-10
 Promedio lineal móvil, 116
 Promedio móvil doble, 154-158
 fórmula para, 156-178
 Promedio móvil simple, 115, 118, 176
 Promedio simple, 115, 151-152, 177, 178
 Promedio, véase Media
 Pronóstico a corto plazo, 45, 62, 117, 118, 318, 431, 477, 479, 534-536
 Pronóstico demográfico, importancia, 64-65
 Pronóstico empresarial, 1-7, 9
 a corto plazo, 4, 5, 62, 117, 118, 318, 431, 477, 479, 534-536
 a largo plazo, 4, 5, 62, 117, 118, 318, 522, 527, 534, 536
 aceptación en la administración de la, 5-6, 11, 542
 administración de la, 7-8, 533
 aplicaciones a la administración de la, 3, 7-8, 11, 46-47, 73-74, 124-125, 176-177, 239-231, 289-291, 351-353, 400-401, 478-480, 526-527, 533-536, 542
 aspectos de juicio de la, 2, 4, 5, 6-7, 15, 62-63, 70-72, 97, 150, 284, 319, 478-479, 520-540
 combinación del, 524-525
 conceptos básicos del, 6, 97-98. Véase también Conceptos específicos
 costo-beneficio del, 6, 11, 536, 542
 datos en el, 61-75, 97, 318, 478
 en las ventas al menudeo, 352
 error de medición en la, 7, 119-22, 1266, 147-49, 152, 207, 210, 74, 378. Véase también Residuales
 futuro del, 543
 generación de escenarios en el, 455
 historia del, 1
 intermedio, 118, 534, 535
 juicios personales en el. Véase Pronóstico empresarial, aspectos de juicio del
 lista de paquetes de cómputo para el, 8-10
 macropronósticos, 4, 5
 método Delphi en el, 523
 métodos de atenuación para el, 115, 166, 118, 146, 158-176, 479, 535.
 Véase también Métodos exponenciales de atenuación
 métodos de promedio para el, 22, 47, 115, 146, 151-158, 176-177
 métodos no formales para la, 62, 115, 118, 146, 147-150, 177, 178
 micropronósticos, 4, 5
 monitoreo de la, 536-539
 necesidad de la, 2-3, 73-74, 176-177, 289, 290-291, 318, 351-353, 400-401, 478-480, 533
 pasos en el, 6-7, 539-540
 política del, 11
 procedimiento de la curva de crecimiento en el, 116, 118, 329-330, 522
 proceso de, 533-539
 pronósticos macroeconómicos, 4, 5
 reglas para el, 11
 responsabilidad del, 540-541
 resumen de modelos de la, 118, 535
 selección y evaluación de métodos para el, 5-6, 48-49, 97-98, 114-118, 122, 146-147, 151, 154, 158, 163, 176, 177, 230-231, 289, 290-291, 318, 400-401, 477-478, 479-480, 521-522, 533-539 544
 subjetiva, Véase Pronóstico empresarial, aspectos de juicio de tecnológica, 522. Véase Pronóstico empresarial, aspectos de juicio de y los horizontes en el tiempo, 117, 146
 técnicas determinísticas en, 98. Véase también Regresión múltiple, Análisis de series de tiempo, Regresión de series de tiempo
 tipos de, 4
 y el análisis de regresión, 203-233, 255-291, 318, 374-402, 525.
 Véase también Análisis de regresión
 y el análisis de series de tiempo, 98, 100, 118, 119, 318-353, 379-402
 y el sistema de información administrativo (MIS), 542-543
 y la correlación, 39-43-203, 318. Véase también Correlación
 y la disponibilidad de computadoras y costos, 1, 3, 8, 61, 118, 177, 478, 521, 543
 y la regresión múltiple, 117, 118, 230, 255-291. Véase también Regresión múltiple
 y las redes neurales, 525-527
 y variables dependientes vs. variables independientes, 203, 230-231, 255, 318, 431. Véase también Autocorrelación, Colinealidad, Correlación serial
 Pronóstico en los negocios en los años 80, 65
 Pronóstico estacional (para una serie de tiempo), 349-350
 Pronóstico macroeconómico, 4-5
 Pronóstico por juicio, véase Pronóstico empresarial, aspectos de juicio de
 Pronóstico tecnológico, 522
 Pronóstico, 9
 Pronóstico, métodos de, véase Pronóstico empresarial
 Pronóstico, véase también Pronóstico empresarial
 Pronóstico:
 intervalo, 211-213, 232
 punto, 209
 selección de, 274-284
 variables de pronóstico, 255-256, 275
 Véase también Pronóstico empresarial e Inferencia estadística y la colinealidad, 256
 Pronósticos a largo plazo, 4, 5, 62, 117, 118, 318, 522, 534-436
 uso de, 14-16
 Proporción con el método de promedio móvil, 339-342
 Proyección de ingreso quinquenal para Downtown Radiology (caso de estudio), 187-196
 Proyección de ingreso quinquenal para Downtown Radiology (caso de estudio), 187
 Prueba de hipótesis, 33-36, 47-48, 220-226
 pasos en, 33
 Prueba *F* parcial, 279-280
 Pruebas:
 de hipótesis, 33-36, 47-48, 220-226

- de Ji-cuadrada, 37-39, 48, 534. Véase también Ji-cuadrada, pruebas
- Publicaciones:
 - datos secundarios en, 64-66
 - sobre el cambio tecnológico (libros), 527
 - sobre indicadores empresariales, 334
- Punto, estimación de, 31, 33, 48
- Punto, pronóstico de, 209
- Quattro Pro, 10, 14
- Quenouille, M.H., 105, 105n
- Rastreo, 161-163, 538
- RATS, 10
- Redes neurales (y los pronósticos), 525-527
- Regresión de cambios de porcentaje, 383, 387-392
- Regresión múltiple, 117, 118, 230, 255-291
 - aplicación en la administración, 230-231, 289-291
 - coeficientes netos de regresión en, 260-261, 291
 - ecuación para, 257-291
 - error estándar de estimación en, 263-264, 291
 - inferencia estadística en, 261
 - procedimiento de regresión por pasos en, 377-380
 - procedimiento de todas las regresiones posibles en, 277-279
 - regresión múltiple de series de tiempo, 374-401
 - problemas con la correlación serial en, 377-380
 - selección de la ecuación en, 274-284
 - sobrecajuste en, 289
 - suposición en, 261
 - validación del modelo para, 271
 - variables de pronóstico en, 256
 - variables de prueba en, 266-271, 291
 - Véase también Correlación, Análisis de regresión, Análisis de series de tiempo
 - vs. regresión simple, 255
 - y el pronóstico econométrica, 287-288
 - y la colinealidad, 256, 272-274, 291
 - y la heteroscedasticidad, 271-272, 291
 - y la matriz de correlación, 256-257
 - y la regresión estacional, 284-287
 - y las ecuaciones normales, 259
 - y los coeficientes de regresión, 260-261
 - y los residuales, 261-263. Véase también Residuales
- Relación curvada, 41, 226-230, 328-330, 522
 - curva de crecimiento de Gompertz para, 329-330, 522
- Relación lineal negativa perfecta, 41
- Relación lineal positiva perfecta, 41
- Relación lineal, 41, 42, 47, 203
 - como suposición del análisis de regresión, 211
- Relación negativa, 41, 43
- Relación positiva, 41
- Relaciones de variables, véase Correlación, Regresión múltiple, Análisis de datos de series de tiempo
- Relaciones no lineales, 227
- Residuales, 119, 122-123, 125, 207, 217, 261-263, 377
 - correlación de, 122-123
 - medición de error en
 - Véase también metodología de Box-Jenkins, Pronóstico empresarial,
- Rubinfeld, D.L., 398
- SAS (Statistical Analysis System), 8
 - ejemplos de computadora para:
 - análisis de regresión, 252, 253, 315-317
 - Box-Jenkins, 514-518
 - estacionalidad, 143-145
 - introducción, 18-20
 - modelo autorregresivo, 426-428
 - regresión múltiple, 315-317
- SCA, 10
- Scientific Computing Associates Corporation, 10
- Scientific Systems, Inc., 9
- Series de tiempo estacionarias, 108, 115, 125, 432, 439
- Series de tiempo no estacionarias, 108, 148. Véase también Tendencia
- Series de tiempo, 98-100, 125
 - autocorrelación en, 377-400
 - estacionarias vs. no estacionarias, 108, 115, 125, 432, 439
 - fuentes de datos de, 62
- Shazam, 9
- Shiskin, Julius, 334n, 336, 350
- Shock del Futuro, El*, 527
- Sistema MIS y pronóstico, 542-543
- Small Engine Doctor (caso de estudio), 358-360
- Smart Software Inc., 10
- Smartforecast II, 10
- Sobrecajuste, 289
- Software, véase Programas de cómputo
- Solar Alternative Company (caso de estudio), 184-185
- Solomon's Jewelry (caso de estudio), 79
- SORITEC, 10
- Sorites Group, 10
- Source, The*, 66
- Spreadsheet Forecaster, 10
- Spreadsheet Forecaster, ejemplo de computadora de la descomposición X-11, 367-371
- SPSSx (Statistical Package for the Social Sciences), 8,9
- Standard and Poor's Corporation Statistical Service, 66
- STAT-PACKETS, 10
- Statistical Abstract of the United States*, 62, 65, 318
- Statistical Analysis System (SAS), 8
 - ejemplos de computadora para:
 - análisis de regresión, 252-253, 315-317
 - Box-Jenkins, 514-518
 - estacionalidad, 143-145
 - introducción, 18-20
 - modelo autorregresivo, 426-428
 - regresión múltiple, 315-317
- Statistical Indicator Reports*, 334
- Statistical Indicators Association, 334
- Statistical Package for the Social Sciences (SPSSx), 8
- StatPac, 9
- StatPac, Inc., 10
- Statpro, 9
- Stratix, 9
- Suma de los residuales cuadrados, 217-219
- Suma total de cuadrados, 233
- Suposiciones:
 - de similitud pasado-futuro, 6
 - en el análisis de regresión, 211, 226, 261, 271
- Survey of Buying Power, The*, 66
- Survey of Current Business, The*, 52, 64, 318, 334-336
- Sybyl/Runner, 9
- Symphony, 10
- SYSTAT, 9
- t, distribución, 30, 122, 212, 572
- t, estadística, 281
 - también Pronóstico, errores en, Error, términos de y Residuales
- Tashman, L., 10n
- Técnicas cualitativas, véase Pronóstico empresarial, aspectos de juicio de
- Técnicas de pronóstico cuantitativas (numéricas), 2, 4, 5, 7, 97-98. Véase también Promedios, Promedios móviles, Atenuación exponencial, Regresión, Regresión Múltiple, Métodos no formales, Análisis de series de tiempo y Métodos Box-Jenkins
- Tendencia central, mediciones de, 24
- Tendencia estacional (en el análisis de series de tiempo), 342-346
- Tendencia no lineal, 328-330

- Tendencia, 99, 108, 116, 125, 322-330
 - atenuación exponencial ajustada para, 118, 168-172
 - curva de crecimiento de Gompertz para, 116, 118, 329-330, 522
 - definición de, 99
 - ejemplo de computadora para correlación serial, 428-430
 - en el análisis de series de tiempo, 319, 322-330, 338-339
 - estacional, 342-346
 - mensual o trimestral, 342-346
 - no lineal, 328-330
- TSP, 9
 - valor t y prueba, 50, 122, 212, 281, 572
 - y el método no formal, 148
- Tendencias futuras y necesidad de pronósticos, 527-543
- Tendencias, valores mensuales o trimestrales, 342-346
- Teorema del límite central, 29, 271
- Teoría de la utilidad, 529-530
- Tercera Ola, La*, 527
- Thompson, J. Walter Company, 66
- Tiger Transport (caso de estudio), 239-241
- Timestat, 10
- Timestat, Inc., 10
- Todas las regresiones posibles, procedimiento, 277-79
- Toffler, Alvin:
 - El Shock del Futuro*, 527
 - La Tercera Ola*, 527
- Valor esperado, 26-27, 48, 49, 528-529, 544
 - fórmula para el, 27-49
- Valores de tendencia cuatrimestral, 342-346
- Valores estandarizados, 28-29, 49
- Value Line Investment Survey*, 66
- Value Line, 66
- Variable aleatoria continua, 48-127
- Variable aleatoria discreta, 26, 48
- Variable aleatoria, 27, 49
- Variable:
 - aleatoria continua, 26, 48
 - aleatoria discreta, 26, 48
 - aleatoria, 26
 - de prueba, 266-271, 291
 - dependiente, 203
 - faltante, 378, 383-387
 - independiente, 203
 - predictor, 255-256, 275
- Variables clave faltantes, 378, 383-387
- Variables de indicador (de prueba), 266-271, 291
- Variables de prueba, 266-271, 291
- Variables, transformación de, 226-230
- Variaciones cíclicas de corto plazo (en el análisis de series de tiempo), 347-349
- Variaciones cíclicas, 319, 330-338
 - y la demografía, 352
 - y las fluctuaciones de corto plazo, 335
 - y los indicadores empresariales, 332-335, 346, 353
- Variaciones, véase Variaciones cíclicas; Fluctuaciones irregulares,
- Variación estacional
- Varianza, 263
 - fórmula para una muestra, 23
 - fórmula para una población, 23
 - y la heteroscedasticidad, 271-272, 291, 374-377
 - y la suma de residuales cuadrados, 217-219
 - y las desviaciones, 263
- Ventas de restaurante (caso de estudio) 302-305, 487-488
- Viches, R.P., 66n
- Visi On, 109
- VisTrend/Plot, 9
- Wadsworth Electronic Publishing Company, 10
- Wall Street Journal, The*, 289
- Wallonick Associates, 9
- Wheelwright, S.C., 351
- Whitcomb, R., 10n
- Wiley, John & Sons, 9, 10
- Winter, modelo de, 117, 172-176, 179
- X-11 variante del método Census II, programa de ajuste estacional, 9, 351
- Xtrapolator, 9
- Younger, 259
- YSTAT, 10
- Zikmund, W.G., 65n