24

Modelos para pronóstico

En los capítulos anteriores hemos sustituido números en los problemas, sin considerar de dónde provienen. Por ejemplo, en el ejemplo de Giapetto LP (ejemplo 1 del capítulo 3), se supuso que el costo variable de la producción de un tren es de 21 dólares. En realidad, tendríamos que estimar el costo de producción de un tren. Es posible hacer lo anterior mediante el método de regresión lineal simple explicado en la sección 24.6.

En los capítulos 15 y 16 se utiliza la teoría de inventarios para determinar las cantidades de producción y los puntos en los que se deben hacer nuevos pedidos. Para usar los modelos de los capítulos 15 y 16, se necesita la capacidad para pronosticar la demanda de un producto. Los métodos de extrapolación y de suavización, que se usan para pronosticar la demanda futura de un producto, se estudian en las secciones 24.1 a 24.5.

Para mostrar otra forma de cómo podemos usar los "buenos pronósticos", suponga que deseamos utilizar los modelos de colas del capítulo 20 con el fin de determinar de qué modo el número de cajeros en un banco debe variar según el día de la semana y la hora del día. Si queremos resolver el problema, necesitamos determinar cómo depende de la hora del día y del día de la semana la tasa a la cual los clientes llegan al banco. Por ejemplo, si supiéramos que más de la mitad de los clientes llega durante la hora del almuerzo (de mediodía a 1 PM), eso tendría un efecto importante en la forma de asignar de manera óptima el personal.

Dos importantes tipos de métodos de predicción, métodos de extrapolación y métodos de predicción causal, se tratan en este capítulo. Los métodos de extrapolación se estudian en las secciones 24.1 a 24.5; estos métodos se utilizan para pronosticar los valores futuros de series de tiempos a partir de valores anteriores de una serie temporal. Para ejemplificarlo, considere las ventas mensuales de Lowland Appliance Company de televisores, aparatos para tocar discos compactos (DC) y acondicionadores de aire (AA) de los últimos 24 meses que se proporcionan en la tabla 1. En un método de pronóstico por extrapolación, se supone que los patrones anteriores y las tendencias en las ventas continuarán en los meses futuros. Por lo tanto, la información anterior acerca de las ventas de aparatos (y ninguna otra información) se utilizan para generar los pronósticos para las ventas de aparatos durante los meses futuros. Los métodos de extrapolación (a diferencia de los métodos de predicción causal que se explican en las secciones 24.6 a 24.8) no toman en cuenta qué "ocasionó" los datos anteriores; simplemente se asume que las tendencias y los patrones anteriores continuarán en el futuro.

Los métodos de pronóstico causales pretenden pronosticar los valores futuros de una variable (llamada variable dependiente) con ayuda de la información anterior, a fin de estimar la relación entre la variable dependiente y una o más variables independientes. Por ejemplo, la compañía Lowland podría tratar de pronosticar las ventas futuras por mes de acondicionadores de aire a partir de los datos anteriores, para determinar cuál es la relación entre las ventas de acondicionadores de aire y las variables independientes, como precio, publicidad y el mes del año. Los métodos de pronóstico causal se analizan en las secciones 24.6 a 24.8.

24.1 Métodos de pronóstico con promedio móvil

Sean $x_1, x_2, \ldots, x_t, \ldots$ los valores observados de una serie de tiempo, donde x_t es el valor de la serie de tiempo observada durante el periodo t. Uno de los métodos de pronóstico más comunes es el método del promedio móvil. Se define $f_{t,1}$ como el periodo de

TABLA 1
Ventas de Lowland

Mes	Ventas de TV	Ventas de DC	Ventas de AA	Mes	Ventas de TV	Ventas de DC	Ventas de AC
1	30	40	13	13	38	79	36
2	32	47	7	14	30	82	21
3	30	50	23	15	35	80	47
4	39	49	32	16	30	85	81
5	33	56	58	17	34	94	112
6	34	53	60	18	40	89	139
7	34	55	90	19	36	96	230
8	38	63	93	20	32	100	201
9	36	68	63	21	40	100	122
10	39	65	39	22	36	105	84
11	30	72	37	23	40	108	74
12	36	69	29	24	34	110	62

pronóstico para el periodo $t + 1$ hecho después de observar x_t. En cuanto al método promedio móvil.

$$\begin{aligned} f_{t,1} &= \text{promedio de las últimas } N \text{ observaciones} \\ &= \text{promedio de } x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-N+1} \end{aligned}$$

donde N es un parámetro dado.

Para ejemplificar el método del promedio móvil, escogemos $N = 3$ y aplicamos el método para pronosticar las ventas de TV de los primeros 6 meses de datos proporcionados en la tabla 1. Los resultados se presentan en la tabla 2. En cuanto a los meses 1 a 3, todavía no hemos observado tres meses de información, de modo que (para $N = 3$) no podemos desarrollar un pronóstico de promedio móvil para las ventas de estos meses. Por lo que se refiere al mes 4, encontramos el pronóstico $f_{3,1} = \frac{30+32+30}{3} = 30.67$. Para el mes 5, el pronóstico es $f_{4,1} = \frac{32+30+39}{3} = 33.67$. Para el mes 6, el pronóstico es $f_{5,1} = \frac{30+39+33}{3} = 34$.

Obsérvese que a partir de uno de los periodos al siguiente, el pronóstico se "desplaza" para "reemplazar" a la observación "más antigua" en el promedio por la observación más reciente.

Elección de *N*

¿Cómo se elige N, el número de periodos usados para calcular el promedio móvil? Para dar respuesta a esta pregunta, es necesario definir una medida de precisión en el pronóstico. Se usa la **desviación absoluta media** (mean absolute deviation, MAD) como medida de la precisión del pronóstico. Antes de definir la MAD, se requiere definir el concepto de un **error de pronóstico**. Dado el pronóstico para x_t, se define e_t como el error de pronóstico de x_t, el cual se obtiene con

$$e_t = x_t - (\text{pronóstico para } x_t)$$

De acuerdo con la tabla 2, encontramos $e_4 = 39 - 30.67 = 8.33$, $e_5 = 33 - 33.67 = -0.67$, y $e_6 = 34 - 34 = 0$. La MAD es simplemente el promedio de los valores absolutos de todas las e_t. Por lo tanto, para los periodos 1 a 6, el pronóstico por promedio móvil proporciona una MAD de

$$\text{MAD} = \frac{|e_4| + |e_5| + |e_6|}{3} = \frac{8.33 + 0.67 + 0}{3} = 3$$

Por consiguiente, en promedio, el pronóstico para las ventas de televisores es erróneo por tres televisores por mes.

TABLA 2
Pronóstico con promedio móvil ($N = 3$) para las ventas de TV

Mes	Ventas reales	Ventas pronosticadas
1	30	—
2	32	—
3	30	—
4	39	$\frac{30+32+30}{3}$
5	33	$\frac{32+30+39}{3}$
6	34	$\frac{30+39+33}{3}$

Pretendemos pronosticar las ventas de los meses siguientes de televisores como un promedio de ventas reales de los últimos N meses. ¿Qué valor de N minimizará el error absoluto medio (obtenido al promediar el error real incurrido durante cada mes)? Trataremos con $N = 1, 2, \ldots, 12$.

Empezaremos con una explicación de la función de Excel DESREF (OFFSET, en el paquete en inglés). Esta función permite escoger un intervalo de celdas relacionado con una ubicación dada en la hoja de cálculo. La sintaxis de la función DESREF se indica enseguida:

DESREF(ref, filas, columnas, alto, ancho)

- **Ref** es la celda a partir de la cual usted establece las referencias de renglón y columnas.
- **Filas** ayuda a ubicar la esquina superior izquierda del intervalo de DESREF. Filas se mide por la cantidad de renglones hacia arriba y hacia abajo (hacia arriba es negativo y hacia abajo es positivo) a partir de la referencia de las celdas.
- **Columnas** ayuda a ubicar la esquina superior izquierda del intervalo de DESREF. Las columnas se miden por la cantidad de ellas a la izquierda o a la derecha (a la izquierda es negativo y a la derecha es positivo) a partir de la referencia de las celdas.
- **Alto** es el número de renglones en el intervalo seleccionado.
- **Ancho** es la cantidad de columnas en el intervalo seleccionado.

Offsetexample.xls

El archivo Offsetexample.xls contiene algunos ejemplos de cómo trabaja la función OFFSET (DESREF, en el paquete en español). Vea la figura 1. Lo curioso de la función OFFSET es que se puede copiar como cualquier fórmula. En la sección siguiente se muestra la efectividad real de la función OFFSET.

Tvsales.xls

El trabajo siguiente se encuentra en el archivo Tvsales.xls. Inicia con la estimación de un pronóstico en el mes 13, porque ése es el primer mes en el cual ya se cuenta con 12 meses de información previa. Véanse figuras 2 y 3.

Paso 1 Al copiar la fórmula siguiente de C17 a C18:C28,

=AVERAGE(OFFSET(B17,-D3,0,D3,1))

se obtiene el promedio de los últimos D3 meses de datos.

- B17 asegura que definimos el intervalo relativo a la celda directamente a la izquierda de la celda donde entra la fórmula.
- -D3 asegura que el intervalo empieza D3 renglones por arriba del renglón donde entra la fórmula.
- El 0 da la certeza de que el intervalo de OFFSET siempre está en la columna B.
- D3 asegura que promediamos las últimas observaciones D3.
- 1 ofrece la certeza de que el intervalo de OFFSET incluye una sola columna.

FIGURA 1

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		Offset examples									
4											
5											
6		1	2	3	4			1	2	3	4
7		5	6	7	8			5	6	7	8
8		9	10	11	12			9	10	11	12
9											
10	=SUM(OFFSET(B7,-1,1,2,1))	8					=SUM(OFFSET(H6,0,1,3,2))	39			
11											
12											
13											
14		1	2	3	4						
15		5	6	7	8						
16		9	10	11	12						
17											
18	=SUM(OFFSET(E16,-2,-3,2,3))	24									
19											

	A	B
4	Month	TV Sales Actual
5	1.00	30
6	2.00	32
7	3.00	30
8	4.00	39
9	5.00	33
10	6.00	34
11	7.00	34
12	8.00	38
13	9.00	36
14	10.00	39
15	11.00	30
16	12.00	36
17	13.00	38
18	14.00	30
19	15.00	35
20	16.00	30
21	17.00	34
22	18.00	40
23	19.00	36
24	20.00	32
25	21.00	40
26	22.00	36
27	23.00	40
28	24.00	34

FIGURA 2

	A	B	C	D	E	F	G	H
2				# OF PERIODS				
3				1				
4	Month	TV Sales Actual	Moving average forecast	Abs error	MAD	5		
5	1.00	30						
6	2.00	32						
7	3.00	30						
8	4.00	39					# of periods	5
9	5.00	33					1	5
10	6.00	34					2	3.666666667
11	7.00	34					3	3.361111111
12	8.00	38					4	3.333333333
13	9.00	36					5	3.016666667
14	10.00	39					6	3.111111111
15	11.00	30					7	3.226190476
16	12.00	36					8	3.21875
17	13.00	38	36	2			9	3.055555556
18	14.00	30	38	8			10	3.083333333
19	15.00	35	30	5			11	3.045454545
20	16.00	30	35	5			12	3.111111111
21	17.00	34	30	4			Min	3.016666667
22	18.00	40	34	6			best #	5
23	19.00	36	40	4				
24	20.00	32	36	4				
25	21.00	40	32	8				
26	22.00	36	40	4				
27	23.00	40	36	4				
28	24.00	34	40	6				

FIGURA 3

Paso 2 Al copiar la fórmula siguiente desde D17 a D18:D28

=ABS(B17-C17)

se calcula el valor absoluto del error en el pronóstico de cada mes (con base en un promedio móvil del mes D3).

Paso 3 En la celda F4 se calcula el promedio de los errores absolutos (llamado a menudo MAD) mediante la fórmula

=AVERAGE(D17:D28)

Paso 4 Se escribe el número de prueba de los periodos para el promedio móvil (1 a 12) en G9:G20, y, en la celda H8, se escribe el MAD con la fórmula

=F4

Paso 5 Después de seleccionar el intervalo de la tabla G8:H20 y elegir en una dirección una tabla de datos con la columna de entrada de la celda D3 se encuentra que un promedio móvil de 5 periodos genera el MAD más pequeño (3.02).

Paso 6 Se obtiene el MAD mínimo en la celda H21 mediante la fórmula

=MIN(H9:H20)

Paso 7 Al escribir la fórmula siguiente en la celda H22

=MATCH(H21,H9:H21,0)

se estima el número de periodos (5) que genera el MAD más pequeño.

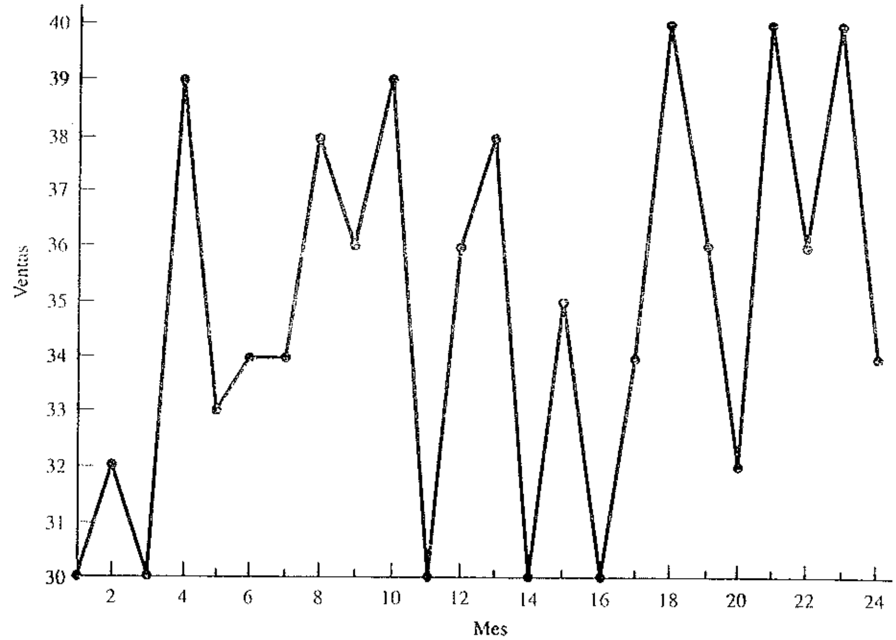


FIGURA 4
Ventas de TV

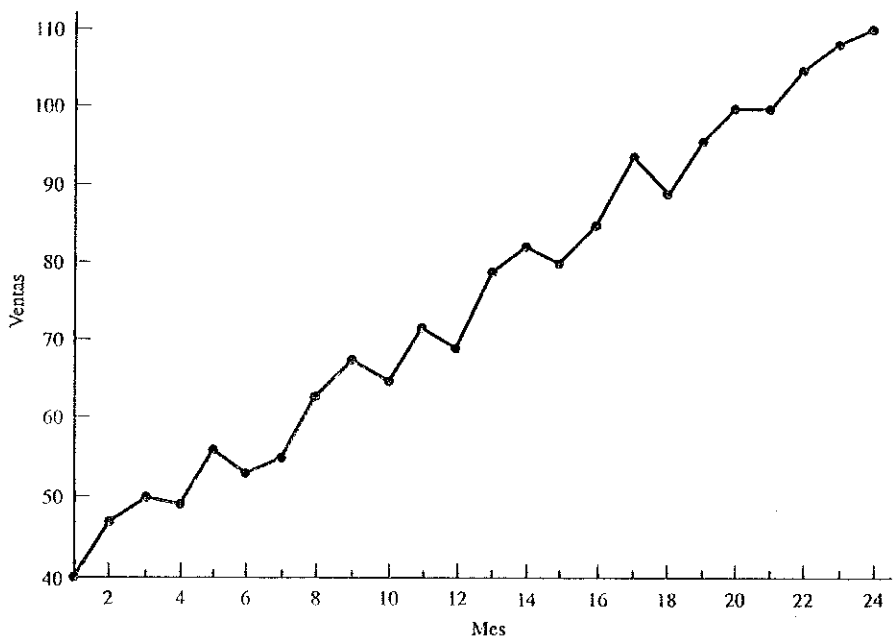


FIGURA 5
Ventas de DC

Los pronósticos por promedios móviles son efectivos en una serie de tiempo que fluctúa respecto a **un nivel base** constante. De acuerdo con la figura 4, parece que las ventas mensuales de televisores fluctúan respecto a un nivel base de 35. Con más formalidad, los pronósticos promedio móvil funcionan mejor si

$$x_t = b + \varepsilon_t \tag{1}$$

donde b es el nivel base para la serie y ε_t es la variación aleatoria en el periodo t respecto al nivel base.

De acuerdo con las figuras 5 y 6, observamos que las ventas de los aparatos que tocan DC y acondicionadores de aire no están muy bien descritas en la ecuación (1). Según la figura 5, vemos que hay una **tendencia** ascendente en las ventas de DC, de modo que no fluctúan respecto a un nivel base. En la figura 6 encontramos que las ventas de acondicionadores de aire muestran **estacionalidad**: los picos y los valles de la serie se repiten a intervalos regulares de 12 meses. Además, en la figura 6 también se observa que las

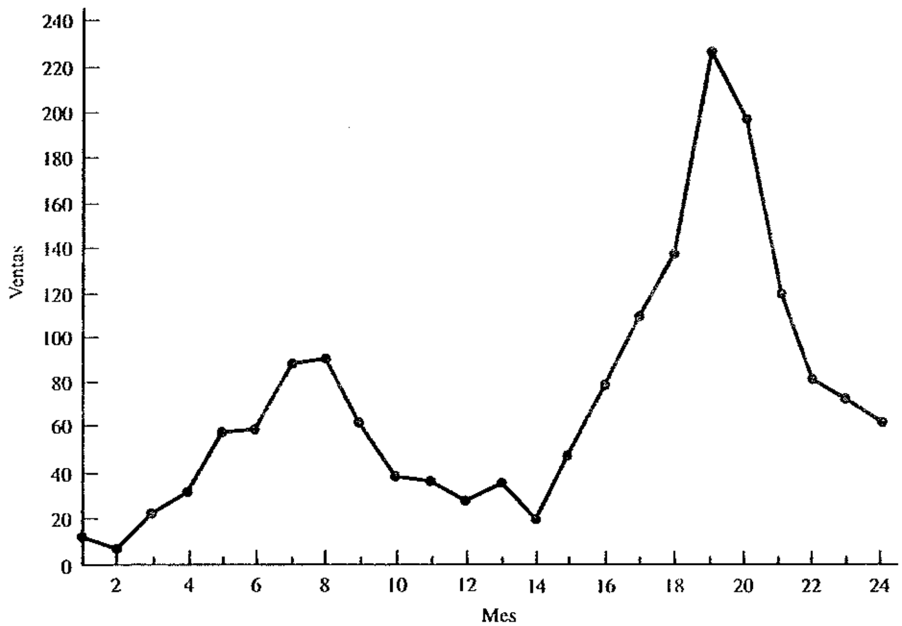


FIGURA 6
Ventas de acondicionadores de aire

ventas de los acondicionadores de aire siguen una tendencia ascendente. El método del promedio móvil ofrece por lo regular pronósticos pobres en situaciones donde está presente una tendencia, o estacionalidad, o ambas. Esta sección finaliza con la advertencia de que una serie de tiempo podría mostrar **comportamiento cíclico,** además de tendencia y estacionalidad. Por ejemplo, las ventas de automóviles se apegan a menudo al ciclo financiero de la economía nacional. El comportamiento cíclico es mucho más irregular que un patrón estacional y es, con frecuencia, difícil de detectar.

24.2 Suavizamiento exponencial simple

Si una serie de tiempo fluctúa respecto a un nivel base, se podría utilizar el **suavizamiento exponencial simple** con el fin de obtener buenos pronósticos para valores futuros de la serie. Con el objeto de describir el suavizamiento exponencial simple, sea A_t = promedio suavizado de una serie de tiempo después de observar x_t. Después de observar x_t, A_t es el pronóstico para el valor de la serie de tiempo durante cualquier periodo futuro. La ecuación clave en un suavizamiento exponencial simple es

$$A_t = \alpha x_t + (1 - \alpha)A_{t-1} \tag{2}$$

En (2), α es una **constante de suavizamiento** que satisface $0 < \alpha < 1$. Para comenzar el procedimiento de pronóstico, debemos contar con un valor (antes de observar x_1) para A_0. Por lo regular, hacemos que A_0 sea el valor observado para el periodo que precede inmediatamente al periodo 1. Al igual que con los pronósticos por promedios móviles, hacemos que $f_{t,k}$ represente el pronóstico para x_{t+k} hecho al final del periodo t. Entonces,

$$A_t = f_{t,k} \tag{3}$$

Si suponemos que tratamos de pronosticar un periodo más adelante, el error al pronosticar x_t (que se escribe de nuevo como e_t) se obtiene con

$$e_t = x_t - f_{t-1,1} = x_t - A_{t-1} \tag{4}$$

Para entender mejor (2), se utiliza (4) para reescribir (2) como

$$A_t = A_{t-1} + \alpha(x_t - A_{t-1}) = A_{t-1} + \alpha e_t$$

Por consiguiente, el nuevo pronóstico $A_t = f_{t,}$ 1 es igual al pronóstico antiguo (A_{t-1}) más una fracción del error del periodo t (e_t). De aquí se infiere que si "damos un exceso de predicción" a x_t, disminuye el pronóstico, y si "disminuimos" la predicción de x_t, elevamos el pronóstico. En cuanto a valores grandes de la constante de suavizamiento α, se da más peso a la observación más reciente (véase la observación 3 al final de la sección).

Ejemplificamos el suavizamiento exponencial simple (con $\alpha = 0.1$) para los primeros seis meses de las ventas de televisores. Los resultados se proporcionan en la tabla 3. Suponemos que se vendieron 32 televisores en el último mes, de modo que empezamos el procedimiento con $A_0 = 32$. Enseguida se proporcionan algunos ejemplos de los cálculos:

$$\begin{aligned}
A_t &= 0.1x_1 + 0.9A_0 = 0.1(30) + 0.9(32) = 31.8\\
f_{0,1} &= A_0 = 32\\
e_1 &= x_1 - A_0 = 30 - 32 = -2\\
f_{1,1} &= A_1 = 31.8\\
e_2 &= x_2 - A_1 = 32 - 31.8 = 0.2\\
A_2 &= 0.1x_2 + 0.9A_1 = 0.1(32) + 0.9(31.8) = 31.82
\end{aligned}$$

Por lo que se refiere a los meses 1 a 6, el MAD del pronóstico se estima con

$$\begin{aligned}
\text{MAD} &= \frac{|-2| + |0.2| + |-1.82\,| + |7.36| + |0.63| + |1.56|}{6}\\
&= 2.26
\end{aligned}$$

En cuanto al periodo completo de 24 meses, es posible determinar (mediante el uso de una tabla de datos en una dirección) el valor de una α que genera el MAD más bajo. Los resultados se ofrecen en la tabla 4. Al parecer, un valor de α entre 0.20 y 0.30 genera el MAD más bajo.

OBSERVACIONES

1 Como $\alpha < 1$, el suavizamiento exponencial "elimina" las variaciones en una serie de tiempo no dando el peso total a la última observación.

2 Si $\alpha = \frac{2}{N+1}$, el suavizamiento exponencial simple (con parámetro de suavizamiento α) y un pronóstico con promedio móvil de N periodos producirán pronósticos similares. Por ejemplo, $\alpha = 0.33$ equivale casi a un promedio móvil de cinco periodos.

3 Para entender por qué el método del suavizamiento *exponencial* recibe este nombre, considere (2) para $t - 1$:

$$A_{t-1} = \alpha x_{t-1} + (1 - \alpha)A_{t-2} \tag{5}$$

Al sustituir (5) en (2) se tiene

$$\begin{aligned}
A_t &= \alpha x_t + (1 - \alpha)[\alpha x_{t-1} + (1 - \alpha)A_{t-2}]\\
&= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2 A_{t-2}
\end{aligned} \tag{6}$$

Obsérvese que

$$A_{t-2} = \alpha x_{t-2} + (1 - \alpha)A_{t-3} \tag{7}$$

TABLA 3
Suavizamiento exponencial simple para las ventas de televisores ($\alpha = .1$)

Mes	Ventas reales	Pronóstico	A_t	e_t
1	30	32	31.8	−2.00
2	32	31.8	31.82	0.20
3	30	31.82	31.64	−1.82
4	39	31.64	32.37	7.36
5	33	32.37	32.44	0.63
6	34	32.44	32.60	1.56

TABLA 4
MAD para las ventas de TV

α	MAD
0.05	3.20
0.10	3.04
0.15	2.94
0.20	2.89
0.25	2.88
0.30	2.90
0.35	2.94
0.40	2.98
0.45	3.05
0.50	3.13

Al sustituir (7) en (6) se llega a

$$A_t = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + (1 - \alpha)^3 A_{t-3}$$

Repitiendo este proceso se obtiene

$$A_t = \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \cdots + \alpha(1 - \alpha)^k x_{t-k} + \cdots \tag{8}$$

Como $\alpha + \alpha(1 - \alpha) + \alpha(1 - \alpha)^2 + \cdots = 1$, la ecuación (8) muestra que si nos regresamos un número "infinito" de periodos, el promedio actual suavizado es un promedio ponderado de todas las observaciones anteriores. El peso dado a la observación de k periodos en el pasado declina en forma exponencial (con un factor de $(1 - \alpha)$). A medida que el valor de α es mayor, se da más peso a las observaciones más recientes. Por ejemplo, para $\alpha = 0.2$, las tres observaciones más recientes tienen 49% del peso (20, 16 y 13%), en tanto que para $\alpha = 0.5$, las tres observaciones más recientes tienen 88% del peso (50, 25 y 13%).

4 Por lo regular, se escoge, en la práctica, a α igual a 0.10, 0.30 o 0.50. Si el valor de α que minimiza el MAD sobrepasa 0.5, entonces tendencia, o estacionalidad o variación cíclica está presente, por lo que no se recomienda el suavizamiento exponencial simple como técnica de pronóstico. En estos casos, el método de Holt (suavizamiento exponencial con tendencia estudiada en la sección 24.3), o el método de Winter (suavizamiento exponencial con tendencia y estacionalidad, tratado en la sección 24.4) proporcionan probablemente los mejores pronósticos.

5 Aun cuando una serie de tiempo no fluctúa respecto a un nivel base constante, el suavizamiento exponencial simple podría ofrecer un buen pronóstico. Si $x_t = m_t + \varepsilon_t$ y $m_t = m_{t-1} + \delta_t$, donde ε_t y δ_t son términos de error independientes, cada uno con media 0, entonces el suavizamiento exponencial simple proporcionará un buen pronóstico. De aquí se infiere que si la demanda media (m_t) de un producto es generada en forma aleatoria sobre el tiempo, el suavizamiento exponencial simple proporcionará un buen pronóstico de la demanda del producto.

24.3 Método de Holt: suavizamiento exponencial con tendencia

Si somos de la opinión que una serie de tiempo muestra una tendencia lineal (y ninguna estacionalidad), entonces, el **método de Holt** ofrece, con frecuencia, un buen pronóstico. Al final del t-ésimo periodo, el método de Holt genera una estimación del nivel de base (L_t) y de la tendencia por periodo (T_t) de la serie. Por ejemplo, suponga que $L_{20} = 20$ y $T_{20} = 2$. Esto quiere decir que después de observar x_{20}, opinamos que el nivel base de la serie es 20 y que el nivel base se incrementa dos unidades por periodo. Por lo tanto, estimamos que, cinco periodos a partir de ahora, el nivel base de la serie será igual a 30.

Después de observar x_t, las ecuaciones (9) y (10) se utilizan para actualizar las estimaciones de la base y la tendencia, α y β son constantes de suavizamiento, cada una entre 0 y 1.

$$L_t = \alpha x_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \tag{9}$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \tag{10}$$

Si deseamos calcular L_t, tomamos un promedio ponderado de las dos cantidades siguientes:

1 x_t, que es una estimación del nivel base del periodo t a partir del periodo actual.

2 $L_{t-1} + T_{t-1}$, que es una estimación del nivel base del periodo t según datos anteriores.

Para calcular T_t consideramos un promedio ponderado de las dos cantidades siguientes:

1 Una estimación de la tendencia a partir del periodo actual dada por el incremento en la base suavizada a partir del periodo $t - 1$ al periodo t.

2 T_{t-1}, que es la estimación anterior de la tendencia.

Al igual que en el caso anterior, definimos $f_{t,k}$ como el pronóstico para x_{t+k} hecha al final del periodo t. Entonces,

$$f_{t,k} = L_t + kT_t \tag{11}$$

Para empezar con el método de Holt, necesitamos una estimación inicial (llámela L_0) de la base y una estimación inicial (llámela T_0) de la tendencia. Podríamos hacer a T_0 igual al incremento promedio mensual de la serie de tiempo durante al año anterior, y a L_0 igual a la observación del último mes.

De acuerdo con la figura 5, es evidente que las ventas del aparato que toca DC muestra una tendencia ascendente, pero ningún patrón estacional obvio está presente. Por lo tanto, el método de Holt debe ofrecer un buen pronóstico. Supongamos que las ventas de los aparatos que tocan CD durante cada uno de los últimos 12 meses son 4, 6, 8, 10, 14, 18, 20, 22, 24, 28, 31 y 34. Entonces,

$$T_0 = \frac{(6-4) + (8-6) + (10-8) + \cdots + (34-31)}{11}$$

$$= \frac{34-4}{11} = 2.73$$

Luego estimamos que $L_0 = 34$.

Si aplicamos el método de Holt a los primeros seis meses de ventas (usando $\alpha = 0.30$ y $\beta = 0.10$) obtenemos los resultados que se presentan en la tabla 5. Enseguida se muestran algunos ejemplos de los cálculos:

$$L_1 = 0.30x_1 + 0.70(L_0 + T_0) = 0.3(40) + 0.7(34 + 2.73) = 37.71$$
$$T_1 = 0.1(L_1 - L_0) + 0.9T_0 = 0.1(37.71 - 34) + 0.9(2.73) = 2.83$$
$$f_{1,1} = L_1 + T_1 = 37.71 + 2.83 = 40.54$$
$$e_2 = x_2 - f_{1,1} = 47 - 40.54 = 6.46$$

TABLA 5
Método de Holt para las ventas de CD (usando $\alpha = 0.30$, $\beta = 0.10$)

Mes	Ventas	L_t	T_t	$f_{t-1,1}$ $(L_{t-1} + T_{t-1})$	e_t $(x_t - f_{t-1,1})$
1	40	37.71	2.83	36.73	3.27
2	47	42.48	3.02	40.54	6.46
3	50	46.85	3.16	45.50	4.50
4	49	49.70	3.13	50.01	−1.01
5	56	53.78	3.22	52.83	3.17
6	53	55.80	3.10	57.00	−4.00

Por lo que se refiere a los primeros seis meses de ventas de los aparatos que tocan DC, tenemos

$$\text{MAD} = \frac{3.27 + 6.46 + 4.5 + 1.01 + 3.17 + 4.00}{6} = 3.74$$

En el caso del periodo completo de 24 meses, MAD = 2.85.

Para un ejemplo más de la ecuación (11), suponga que deseamos hacer un pronóstico al final del mes 6 para las ventas del mes 10 de DC. De acuerdo con (11), encontramos que $f_{6,4} = L_6 + 4T_6 = 55.80 + 4(3.10) = 68.2$. Al tratar varias combinaciones de α y β podríamos determinar los valores de α y β que minimizan el MAD. Si ambos valores no son menores de 0.5, entonces la estacionalidad o el comportamiento cíclico podría estar presente, por lo cual se tendría que aplicar otro método de pronóstico.

En resumen, el método de Holt proporciona un buen pronóstico para una serie con tendencia lineal. Tal serie se podría modelar como $x_t = a + bt + \varepsilon_t$, donde

a = nivel base en inicio del periodo 1

b = tendencia por periodo

ε_t = término de error para el periodo t

Es posible utilizar una versión multiplicativa del método de Holt (véase problema 15) para generar un buen pronóstico para una serie de la forma $x_t = ab^t\varepsilon_t$. Aquí, el valor de b representa el porcentaje de crecimiento en el nivel base de la serie durante cada periodo. Por lo tanto, $b = 1.1$ implica que el nivel base de la serie se incrementa 10% por periodo. En este modelo, ε_t es un factor de error aleatorio con una media de 1.

Organización de una hoja de cálculo del método de Holt

Holt.xls

La figura 7 (que se encuentra en el archivo Holt.xls) ilustra una aplicación del método de Holt. En las columnas B y C están escritos los 24 meses de ventas de aparatos para DC obtenidas de la tabla 1. En las celdas D4 y E4 están L_0 y T_0. Los valores de ensayo de alfa y beta aparecen en las celdas E2 y F2. En la celda D5 se calcula L_1 luego de escribir la fórmula =E$2*C5+(1−E$2)*(D4+E4). En la celda E5 se calcula T_1 con la fórmula =F$2*(D5−D4)+(1−F$2)*E4. En la celda F5 se estima $f_{0,1}$ mediante la fórmula =D4+E4. En la celda G5 se determina e_1 con la fórmula =C5−F5. En la celda H5 se encuentra $|e_1|$ con la fórmula =ABS(G5). Al copiar las fórmulas desde el intervalo C5:H5 hasta el intervalo C5:H28 se completa la aplicación del método de Holt. La fórmula =AVERAGE(H5:H28) en la celda G2 calcula el MAD (2.85) para los 24 meses.

Podemos usar una tabla de datos en Excel para estimar los valores de α y β que generan un MAD pequeño. Escribimos los valores posibles para α en un intervalo de celdas B31:B39, y los valores para β en el intervalo C30:K30. Escribimos una fórmula para determinar el MAD (=G2) en la celda B30. Mediante el comando DATATABLE (TABLA, en el paquete en español), elegimos el intervalo de la tabla B30:K39. Luego seleccionamos la celda E2 como la celda de los datos entrada de las columnas y la celda F2 como la celda de los renglones de entrada. Esto ocasiona que los valores en B1:B39 entren a E2 y que los valores en C30:K30 entren en la celda F2. Después de escoger OK (Aceptar), Excel calcula el MAD para cada combinación de α y β en la tabla. Observamos que de las combinaciones presentes, $\alpha = .10$ y $\beta = .40$ ofrece el MAD más bajo (2.70). Si queremos un MAD todavía más bajo, podríamos explorar los valores de α y β cercanos a .10 y .40, respectivamente, generando otra tabla. A propósito, F9 calcula una vez más la última tabla que usted elaboró en su hoja de cálculo.

FIGURA 7

Método de Holt

	B	C	D	E	F	G	H	I	J	K
1		HOLT	METHOD	ALPHA	BETA	MAD=				
2				0.3	0.1	2.84686937				
3	MONTH	CD SALES	Lt	Tt	f(t-1,1)	et	|et|			
4	0		34	2.73						
5	1	40	37.711	2.8281	36.73	3.27	3.27			
6	2	47	42.47737	3.021927	40.5391	6.4609	6.4609			
7	3	50	46.8495079	3.15694809	45.499297	4.500703	4.500703			
8	4	49	49.7045192	3.12675441	50.006456	-1.006456	1.00645599			
9	5	56	53.7818915	3.2218162	52.8312736	3.1687264	3.1687264			
10	6	53	55.8025954	3.10170497	57.0037077	-4.0037077	4.00370772			
11	7	55	57.7330103	2.98457596	58.9043004	-3.9043004	3.90430038			
12	8	63	61.4023104	3.05304837	60.7175862	2.28241378	2.28241378			
13	9	68	65.5187511	3.15938761	64.4553587	3.54464127	3.54464127			
14	10	65	67.5746971	3.04904345	68.6781387	-3.6781387	3.67813872			
15	11	72	71.0366184	3.09033123	70.6237406	1.37625945	1.37625945			
16	12	69	72.5888647	2.93652274	74.1269496	-5.1269496	5.12694962			
17	13	79	76.5677712	3.04076112	75.5253875	3.47461252	3.47461252			
18	14	82	80.3259726	3.11250515	79.6085324	2.39146765	2.39146765			
19	15	80	82.4069345	3.00935081	83.4384778	-3.4384778	3.4384778			
20	16	85	85.2913997	2.99686226	85.4162853	-0.4162853	0.41628527			
21	17	94	90.0017834	3.1682144	88.2882619	5.71173805	5.71173805			
22	18	89	91.9189984	3.04311447	93.1699978	-4.1699978	4.16999776			
23	19	96	95.273479	3.07425108	94.9621129	1.0378871	1.0378871			
24	20	100	98.8434111	3.12381918	98.3477301	1.65226989	1.65226989			
25	21	100	101.377061	3.06480227	101.96723	-1.9672303	1.96723025			
26	22	105	104.609304	3.08154636	104.441863	0.55813656	0.55813656			
27	23	108	107.783596	3.09082084	107.690851	0.30914923	0.30914923			
28	24	110	110.612091	3.06458835	110.874416	-0.8744164	0.87441638			
29					BETA					
30	2.84686937	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
31	0.1	2.85549229	2.79917857	2.73752333	2.70190155	2.74959079	2.79516165	2.83287197	2.86298631	2.92313769
32	0.2	2.76620734	2.73280037	2.76089108	2.78733151	2.84068253	2.92107922	2.97420106	2.98283649	2.99339222
33	0.3	2.84686937	2.87216928	2.90902939	2.95265285	2.98980036	3.04270174	3.13130108	3.21889032	3.2935572
34	0.4	2.96069374	3.0030147	3.05465632	3.110774	3.17038284	3.23467306	3.30196817	3.35141111	3.38544721
35	0.5	3.0707409	3.1289783	3.19166282	3.25002095	3.31027756	3.36404349	3.42489473	3.50405565	3.58298443
36	0.6	3.19398831	3.26198014	3.33253292	3.39877183	3.48348934	3.58574625	3.68496609	3.78162671	3.87646977
37	0.7	3.31493308	3.39345284	3.47449093	3.59818192	3.71792621	3.83504256	3.95075646	4.06889148	4.21031682
38	0.8	3.43163015	3.52755966	3.66950469	3.80654215	3.940503	4.07309251	4.26812614	4.4709749	4.67844324
39	0.9	3.53843635	3.69071551	3.844996	4.02099412	4.2282208	4.45607616	4.68692747	4.9520137	5.24579003

24.4 Método de Winter: suavizamiento exponencial con estacionalidad

El apropiadamente llamado **método de Winter** (método de invierno) se utiliza para pronosticar series temporales en las cuales están presentes la tendencia y la estacionalidad. Como ya se mencionó, las ventas de acondicionadores de aire muestran una tendencia ascendente y estacionalidad, lo cual se ilustra en la figura 6, por lo que el método de Winter es el candidato lógico para predecir estas ventas.

Para explicar el método de Winter se requieren dos definiciones. Sea c = número de periodos en la duración del patrón estacional ($c = 4$ para la información trimestral y $c = 12$ si los datos son mensuales). Sea s_t una estimación de un factor multiplicativo estacional para el mes t, obtenido después de observar x_t. Por ejemplo, suponga que el mes 7 es julio y $s_7 = 2$. Luego de observar las ventas de acondicionadores de aire del mes 7, opinamos que las ventas de acondicionadores de julio (con todos los otros factores iguales) serán el doble de las ventas esperadas durante un mes promedio. Si el mes 24 es diciembre y $s_{24} = 0.4$, entonces, después de observar las ventas del mes 24, pronosticaremos que las ventas de acondicionadores en diciembre serán 40% de las ventas esperadas durante un mes promedio. En lo siguiente, L_t y T_t significan lo mismo que en el método de Holt. Cada periodo, L_t, T_t y s_t están actualizados (en ese orden) por medio de las ecuaciones (12) a (14). Una vez más, α, β, y γ son constantes de suavizamiento, cada una de las cuales está entre 0 y 1.

$$L_t = \alpha \frac{x_t}{s_{t-c}} + (1 - \alpha)(L_{t-1} + T_{t-1}) \tag{12}$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \tag{13}$$

$$s_t = \gamma \frac{x_t}{L_t} + (1 - \gamma)s_{t-c} \tag{14}$$

Mediante la ecuación (12) se actualiza la estimación de la base de la serie, tomando un promedio ponderado de las dos cantidades siguientes:

1 $L_{t-1} + T_{t-1}$, que es la estimación del nivel base antes de observar x_t

2 La observación desestacionalizada $\frac{x_t}{s_{t-c}}$, la cual es una estimación de la base obtenida a partir del periodo actual

La ecuación (13) es idéntica a la ecuación (10) de T_t usada para actualizar la tendencia en el método de Holt.

Por otro lado, la ecuación (14) actualiza la estacionalidad del mes t considerando un promedio ponderado de las dos cantidades siguientes:

1 La estimación más reciente de la estacionalidad del mes t (s_{t-c})

2 $\frac{x_t}{L_t}$, la cual es una estimación de la estacionalidad del mes t, obtenida a partir del mes actual.

Al final del periodo t, la predicción ($f_{t,k}$) por el mes $t + k$ se obtiene con

$$f_{t,k} = (L_t + kT_t)s_{t+k-c} \tag{15}$$

Por lo tanto, para pronosticar el valor de la serie durante el periodo $t + k$ se multiplica la estimación de la base ($L_t + kT_t$) del periodo $t + k$ por la estimación más reciente del factor de estacionalidad (s_{t+k-c}) del mes ($t + k$).

Inicio del método de Winter

Para obtener buenos pronósticos por medio del método de Winter, se requiere obtener buenas estimaciones iniciales de la base, tendencia y todos los factores estacionales. Sean

$$\begin{aligned}
L_0 &= \text{estimación de la base al inicio del mes 1} \\
T_0 &= \text{estimación de la tendencia al inicio del mes 1} \\
s_{-11} &= \text{estimación del factor estacional de enero al inicio del mes 1} \\
s_{-10} &= \text{estimación del factor estacional de febrero al inicio del mes 1} \\
&\vdots \\
s_0 &= \text{estimación del factor estacional de diciembre al inicio del mes 1}
\end{aligned} \tag{16}$$

Se dispone de una gran diversidad de métodos para estimar los parámetros en (16). Escogemos un método simple que requiere dos años de información. Suponga que las ventas de los últimos dos años (por mes) fueron como se señala:

Año −2: 4, 3, 10, 14, 25, 26, 38, 40, 28, 17, 16, 13

Año −1: 9, 6, 18, 27, 48, 50, 75, 77, 52, 33, 31, 24

Ventas totales durante el año − 2 = 234

Ventas totales durante el año − 1 = 450

La estimación de T_0 se efectúa mediante

$$T_0 = \frac{\begin{pmatrix}\text{Ventas promedio mensuales} \\ \text{durante el año } -1\end{pmatrix} - \begin{pmatrix}\text{Ventas promedio mensuales} \\ \text{durante el año } -2\end{pmatrix}}{12}$$

o bien,

$$T_0 = \frac{\frac{450}{12} - \frac{234}{12}}{12} = 1.5$$

Determinamos primero la demanda promedio mensual durante el año $-1(\frac{450}{12})$ para estimar L_0. Esto estima la base que hay a mitad del año −1 (mes 6.5 del año −1). Si lo que se desea es la estimación hasta el fin del mes 12 del año −1, se suma $(12 - 6.5)T_0 = 5.5T_0$. Por lo tanto, la estimación de $L_0 = 37.5 + 5.5(1.5) = 45.75$.

Con el objeto de estimar el factor de estacionalidad de un mes específico (por ejemplo, enero $= s_{-11}$), se obtiene una estimación de la estacionalidad de enero para el año −2 y el año −1, y se promedian. En el año −2, la demanda promedio mensual fue $\frac{234}{12} = 19.5$; en enero del año −2, se vendieron 4 acondicionadores de aire. Por lo tanto,

$$\text{Estimación de la estacionalidad de enero del año } -2 = \frac{4}{19.5} = 0.205$$

De igual manera,

$$\text{Estimación de la estacionalidad de enero del año } -1 = \frac{9}{37.5} = 0.240$$

Por último, se obtiene $s_{-11} = \frac{0.205+0.24}{2} = 0.22$. De modo similar se llega a

$$s_{-10} = 0.16, \quad s_{-9} = 0.50, \quad s_{-8} = 0.72, \quad s_{-7} = 1.28, \quad s_{-6} = 1.33,$$
$$s_{-5} = 1.97, \quad s_{-4} = 2.05, \quad s_{-3} = 1.41, \quad s_{-2} = 0.88, \quad s_{-1} = 0.82, \quad s_0 = 0.65$$

A manera de comprobación, las estimaciones del factor estacional inicial deben promediar 1.

Antes de mostrar cómo se utilizan las ecuaciones (12) a (14), se explica la aplicación de (15) en los pronósticos. Al inicio del mes 1, el pronóstico de las ventas de acondicionadores de aire para el mes 1 es

$$f_{0,1} = (L_0 + T_0)s_{0+1-12} = (45.75 + 1.5)0.22 = 10.40$$

Al empezar el mes 1, el pronóstico para las ventas de acondicionadores de aire para el mes 7 es

$$f_{0,7} = (L_0 + 7T_0)s_{0+7-12} = (45.75 + 7(1.5))1.97 = 110.81$$

Para $\alpha = 0.5$, $\beta = 0.4$, $\gamma = 0.6$ al aplicar el método de Winter a los primeros 12 meses de los datos de las ventas de acondicionadores de aire se llega a los resultados que se presentan en la tabla 6.

Los cálculos para llegar a conocer L_1, T_1 y s_1 se ilustran enseguida:

$$L_1 = 0.5\left(\frac{x_1}{s_{-11}}\right) + 0.5(L_0 + T_0) = 0.5\left(\frac{13}{0.22}\right) + 0.5(45.75 + 1.5) = 53.17$$

$$T_1 = 0.4(L_1 - L_0) + 0.6T_0 = 0.4(53.17 - 45.75) + 0.6(1.5) = 3.87$$

$$s_1 = 0.6\left(\frac{x_1}{L_1}\right) + 0.4s_{-11} = 0.6\left(\frac{13}{53.17}\right) + 0.4(0.22) = 0.23$$

TABLA 6

Método de Winter para los acondicionadores de aire ($\alpha = 0.5$, $\beta = 0.4$, $\gamma = 0.6$)

Mes	Ventas	L_t	T_t	s_t	$f_{t-1,1}$	Error
1	13	53.17	3.87	0.23	10.40	2.60
2	7	50.39	1.21	0.15	9.13	−2.13
3	23	48.80	0.09	0.48	25.80	−2.80
4	32	46.67	−0.80	0.70	35.20	−3.20
5	58	45.59	−0.91	1.28	58.71	−0.71
6	60	44.90	−0.82	1.33	59.42	0.58
7	90	44.88	−0.50	1.99	86.82	3.18
8	93	44.87	−0.30	2.06	90.97	2.03
9	63	44.62	−0.28	1.41	62.84	0.16
10	39	44.33	−0.29	0.88	39.02	−0.02
11	37	44.58	−0.07	0.83	36.12	0.88
12	29	44.56	−0.05	0.65	28.93	0.07

Por lo tanto, al terminar el mes 1, el pronóstico para (por ejemplo) las ventas de acondicionadores en el mes 7 es $f_{1,6} = (L_1 + 6T_1)s_{1+6-12} = (53.17 + 6(3.87))1.97 = 150.49$. El pronóstico para el mes 7 hecho al final del mes 1 sobrepasa al pronóstico hecho al principio del mes 1 para el mes 7, porque las ventas del mes 1 fueron más altas que las pronosticadas.

Por lo que se refiere a la información que comprende 24 meses, los cálculos realizados en hoja de cálculo señalan que MAD = 10.48.

OBSERVACIONES

1 Como se utilizan tres constantes de suavización en el método de Winter, es un problema que consume mucho tiempo determinar la combinación de los valores de α, β y γ que generan el menor MAD. El uso de una hoja de cálculo para aplicar el método de Winter se estudia en el problema de repaso 3. El Solver que viene con Excel ayuda a determinar los valores de α, β y γ. Utilice Solver sólo para hallar los valores de los parámetros que minimizan MAD.

2 Aunque los valores de α y β que minimizan MAD no deben ser mayores que 0.5 (como en el método de Holt), es común, por lo que toca al mejor valor de γ, que ésta sea mayor que 0.5. Esto se debe a que, para los datos mensuales, cada factor estacional mensual está actualizado sólo durante $\frac{1}{12}$ de todos los periodos. Puesto que los factores de estacionalidad se actualizan tan pocas veces, necesitaríamos dar más peso a cada observación, de modo que $\gamma > 0.5$ no es imposible.

3 En la figura 8 se ilustra que los pronósticos de las ventas de acondicionadores de aire (para $\alpha = 0.5$, $\beta = 0.4$ y $\gamma = 0.6$) no son inferiores a las ventas reales de estos aparatos. La concordancia entre ventas pronosticadas y reales es muy buena, excepto en los meses 15 y 17. Los pronósticos para estos meses son mucho muy altos. Quizá se contrataron nuevos agentes de ventas durante estos dos meses, lo que ocasionaría que las ventas son menores que lo previsto.

Precisión en los pronósticos

Podríamos utilizar MAD para estimar s_e = desviación estándar de los errores de pronóstico en cualquier modelo de pronóstico en el cual los errores en el pronóstico están distribuidos normalmente. La relación entre MAD y s_e se expresa mediante la fórmula (17)

$$s_e = 1.25 \text{ MAD} \tag{17}$$

Si se supone que los errores siguen una distribución normal, entonces sabemos que alrededor de 68% de los pronósticos deben estar dentro de la s_e del valor real, y que casi 95% de los pronósticos deben encontrarse dentro de $2s_e$ del valor real. Por lo tanto, en cuanto a los pronósticos de las ventas de acondicionadores de aire, encontramos que $s_e = 1.25(10.48) = 13.10$. De modo que esperaríamos que, con respecto a 0.68(24) = 16 de los 24 meses, nuestros pronósticos sobre las ventas serían inexactos cuando mucho 13.10 acondicionadores, y para 0.95(24) = 23 de 24 meses, nuestros pronósticos serían inexactos cuando mucho 2(13.10) = 26.2 acondicionadores. En realidad, nuestros pronósticos para las ventas de los acondicionadores de aire son precisos dentro de 13.10 durante 17 meses y precisos dentro de 26.2 durante 22 meses.

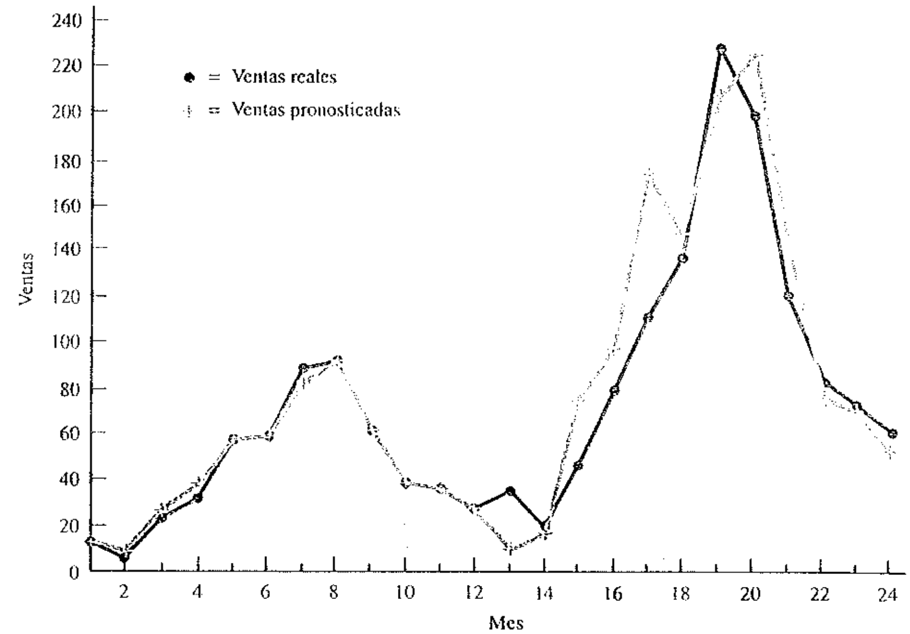


FIGURA 8
Pronósticos sobre las ventas de acondicionadores de aire

Observamos que, en la mayor parte de las situaciones donde se requiere un pronóstico, saber algo relacionado con la precisión probable del pronóstico es casi tan importante como el pronóstico real. Por lo tanto, ¡esta pequeña subsección es muy importante!

PROBLEMAS

Grupo A

1 Un suavizamiento exponencial simple (con $\alpha = 0.2$) se usa para pronosticar las ventas mensuales de cerveza en la vinatería Gordon's. Después de observar la demanda de abril, la demanda pronosticada para mayo es 4000 envases de cerveza.

a A principios de mayo, ¿cuál es el pronóstico para las ventas de cerveza en julio?

b La demanda real durante mayo y junio es como sigue: mayo, 4500 latas de cerveza; junio, 3500 latas de cerveza. Después de observar la demanda de junio, ¿cuál es el pronóstico para la demanda de julio?

c La demanda durante mayo y junio promedia $\frac{4500+3500}{2} =$ 4000 latas por mes. Esto es lo mismo que el pronóstico para las ventas mensuales antes de que observáramos los datos de mayo y junio. Todavía después de observar la demanda de cerveza de mayo y de junio, el pronóstico para la demanda de julio decreció desde lo que era al final de abril. ¿Por qué?

2 Estamos haciendo un pronóstico de las ventas trimestrales de bebidas carbonatadas en la vinatería de Gordon's mediante el método de Winter. Contamos con la siguiente información:

Factores de estacionalidad: otoño = 0.8, primavera = 1.2, invierno = 0.7, verano = 1.3
Estimación de la base actual = 400 envases por trimestre
Estimación de la tendencia actual = 40 envases por trimestre
$\alpha = 0.2 \quad \beta = 0.3 \quad \gamma = 0.5$

Ahora se observa la venta de 650 envases durante el trimestre del verano.

a Utilice esta información para actualizar las estimaciones de base, tendencia y estacionalidad.

b Después de observar la demanda del verano, pronostique la demanda para el trimestre del otoño y del trimestre del invierno.

3 Estamos utilizando el método de Winter e información mensual para pronosticar el PIB. (Todos lo números están en miles de millones de dólares). Al final de enero de 2005, $L_t =$ 600 y $T_t = 5$. Contamos con las estacionalidades siguientes: enero, 0.80; febrero, 0.85; diciembre, 1.2. Durante febrero de 2005, el PIB está en el nivel de 630. ¿Cuál es la predicción para el nivel de diciembre de 2005 del PIB al final de febrero? Utilice $\alpha = \beta = \gamma = 0.5$.

4 Mediante el método de Holt queremos predecir las ventas mensuales de videocaseteras en Highland Appliance. A fines de octubre de 2005, $L_t = 200$ y $T_t = 10$. Durante noviembre de 2005, se vendieron 230 videocaseteras. Al final de noviembre, MAD = 25 y estamos 95% seguros de que las ventas de videocaseteras para diciembre de 2005 estarán entre _____ y _______. Utilice $\alpha = \beta = 0.5$.

5 Mediante suavizamiento exponencial simple queremos predecir las ventas mensuales de las rasuradoras eléctricas en la tienda Hook's. Al final de octubre de 2006, el pronóstico de las ventas de diciembre de 2006 era de 40. En noviembre se vendieron 50 rasuradoras y 45 durante diciembre. Suponga que $\alpha = 0.50$. Al final de diciembre de 2006, ¿cuál es el pronóstico para el número total de rasuradoras que se venderán durante marzo y abril de 2007?

6 Usamos suavizamiento exponencial simple para predecir las ventas mensuales de autos en la Ford de Bloomington. La compañía cree que las ventas no manifiestan tendencia o estacionalidad, por eso esta técnica ha proporcionado pronósticos satisfactorios en la mayor parte. No obstante, la compañía observa cada marzo que las ventas tienden a sobrepasar el pronóstico por suavizamiento exponencial simple (A_{Feb}) en 200. Suponga que al final de febrero de 2004, $A_t = 600$. Durante marzo de 2004, se vendieron 900 automóviles.

a Si $\alpha = 0.3$ determine (al final de marzo de 2004) un pronóstico para las ventas de automóviles de abril de 2004.

b Suponga que al final de marzo, MAD = 60. Estamos 95% seguros de que las ventas de abril estarán entre ______ y ______.

7 La unión de crédito de la universidad abre de lunes a sábado. El método de Winter se utiliza para predecir ($\alpha = \beta = \gamma = 0.5$) la cantidad de clientes que llega al banco cada día. Después de incorporar los arribos de octubre 16, $L_t = 200$ clientes, $T_t = 1$ cliente y las estacionalidades son como se indican: lunes, 0.90; martes, 0.70; miércoles, 0.80; jueves, 1.1; viernes, 1.2; sábado, 1.3. Esto quiere decir que un lunes representativo, la cantidad de clientes es 90% de los clientes que llegan al banco en un día promedio. El martes, 17 de octubre, llegan 182 clientes al banco. Al cerrar el negocio el 17 de octubre, dé un pronóstico del número de clientes que llegarán el día 25 de octubre al banco.

8 Se utiliza el método de Holt (suavizamiento exponencial con tendencia y sin estacionalidad) para predecir las ventas de automóviles por semana en la Ford TOD. En la actualidad, se estima que la base es de 50 automóviles por semana, y la tendencia se estima en 6 automóviles por semana. Durante la semana que corre, se han vendido 30 automóviles. Después de observar las ventas actuales de la semana, pronostique la cantidad de automóviles que se venderán durante la semana que inicia tres semanas después de la finalización de la semana presente. Utilice $\alpha = \beta = 0.3$.

9 Se aplica el método de Winter para pronosticar (con $\alpha = 0.2$, $\beta = 0.1$ y $\gamma = 0.5$) el número de clientes atendidos cada día en el Last National Bank. El banco abre de lunes a viernes. En la actualidad se han estimado las estacionalidades presentes: lunes, 0.80; martes, 0.90; miércoles, 0.95; jueves, 1.10; viernes, 1.25. Una estacionalidad de 0.80 para el lunes quiere decir que en un lunes, la cantidad de clientes atendidos en el banco tiende a ser de 80% de promedio. En la actualidad, se estima que la base es 20 clientes, y que la tendencia es igual 1 cliente. Después de observar que en lunes se atienden a 30 clientes en el banco, pronostique la cantidad de clientes atendidos el miércoles.

10 Se nos ha asignado el trabajo de pronosticar la cantidad de motores para avión que pide cada mes la Engine Company. Al final de febrero, el pronóstico es que se pedirán 100 motores durante abril. Durante marzo se pedirán 120 motores.

a Con $\alpha = 0.3$ dé (a fines de marzo) un pronóstico para el número de pedidos hechos durante abril. Conteste la misma pregunta para mayo.

b Suponga que a fines de marzo, MAD = 16. A fines de marzo tenemos 68% de seguridad de que los pedidos de abril estarán entre _______ y ______.

11 El método de Winter se aplica para predecir las ventas al menudeo trimestrales de Estados Unidos (en miles de millones de dólares). Al final del primer trimestre, $L_t = 300$, $T_t = 30$, y los índices estacionales son como se indica a continuación: trimestre 1, 0.90; trimestre 2, 0.95; trimestre 3, 0.95; trimestre 4, 1.20. Durante el segundo trimestre, las ventas al menudeo son de 360 miles de millones. Suponga $\alpha = 0.2$, $\beta = 0.4$ y $\gamma = 0.5$.

a Al final del segundo trimestre, elabore un pronóstico para las ventas al menudeo durante el cuarto trimestre del año.

b Al finalizar el segundo trimestre dé un pronóstico para el segundo trimestre del año próximo.

Grupo B

12 Se utiliza suavizamiento exponencial simple con $\alpha = 0.3$ para pronosticar las ventas de radios de Lowland Appliance. Los pronósticos se hacen con base en la información mensual. Después de observar las ventas de los radios en agosto, el pronóstico para septiembre es de 100 radios.

a Durante septiembre se vendieron 120 radios. Después de observar las ventas de septiembre, ¿cuál es el pronóstico para las ventas de radios en octubre? ¿Y para noviembre?

b Resulta que las ventas de junio registraron 10 radios vendidos. Pero en realidad se vendieron 100 radios en junio. Luego de corregir el error, ¿cuál sería el pronóstico para la venta de radios de octubre?

13 En la explicación del método de Winter, una estacionalidad mensual de, por ejemplo, 0.80 para enero significa que, durante enero, se espera que las ventas de acondicionadores de aire sean 80% de las ventas esperadas en un mes promedio. Un método opcional para modelar la estacionalidad es hacer que el factor de estacionalidad para cada mes represente qué tan lejos por encima del promedio estarán las ventas de acondicionadores de aire durante el mes actual. Por ejemplo, si $s_{enero} = -50$, entonces se espera que las ventas de estos aparatos durante enero sean 50 menos que las ventas durante un mes promedio. Si $s_{julio} = 90$, entonces se espera que las ventas de acondicionadores durante julio sean 90 más que las ventas de aparatos durante un mes promedio. Sean

s_t = la estacionalidad para el mes t después de que se ha observado la demanda del mes t. Sea

L_t = la estimación de la base después de que se observa la demanda del mes t

T_t = la estimación de la tendencia después de observar la demanda del mes t

Entonces, las ecuaciones para el método de Winter dadas en el texto se modifican como se indica (* significa multiplicación)

$$L_t = \alpha * (I) + (1 - \alpha) * (L_{t-1} + T_{t-1})$$
$$T_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * T_{t-1}$$
$$s_t = \gamma * (II) + (1 - \gamma) * s_{t-12}$$

a ¿Qué deben ser I y II?

b Suponga que el mes 13 es enero, $L_{12} = 30$, $T_{12} = -3$, $s_1 = -50$, y $s_2 = -20$. Sea $\alpha = \gamma = \beta = 0.5$. Suponga que se vendieron 12 acondicionadores de aire durante el mes 13. Al final del mes 13, ¿cuál es el pronóstico sobre las ventas de acondicionadores de aire durante el mes 14?

14 En el método de Winter se supone una estacionalidad multiplicativa y una tendencia aditiva. Por ejemplo, una tendencia de 5 quiere decir que la base se incrementa 5 unidades por periodo. Suponga que hay en realidad una tendencia multiplicativa. Entonces (ignorando la estacionalidad) si la estimación actual de la base es de 50 y la estimación actual de la tendencia es de 1.2, pronosticaríamos que la demanda se incrementaría 20% por periodo. Si ignoramos la estacionalidad, pronosticaríamos, por lo tanto, la demanda del periodo siguiente como 50(1.2) y la demanda de dos periodos en el futuro como $50(1.2)^2$.

Si queremos usar una tendencia multiplicativa en el método de Winter, debemos usar las ecuaciones siguientes:

$$L_t = \alpha * \left(\frac{x_t}{s_{t-c}}\right) + (1 - \alpha) * (I)$$
$$T_t = \beta * (II) + (1 - \beta) * T_{t-1}$$
$$s_t = \gamma * \left(\frac{x_t}{L_t}\right) + (1 - \gamma)* s_{t-12}$$

a Determine qué deben ser *I* y *II*.

b Suponga que estamos trabajando con información mensual y que el mes 12 es diciembre, el mes 13 es enero, y así sucesivamente. Asimismo, suponga que $L_{12} = 100$, $T_{12} = 1.2$, $s_1 = 0.90$, $s_2 = 0.70$, y $s_3 = 0.95$. Suponga que $s_{13} = 200$. Al final del mes 13, ¿cuál es el pronóstico para x_{15}? Suponga $\alpha = \beta = \gamma = 0.5$.

15 En el método de Holt se supone una tendencia aditiva. Por ejemplo, una tendencia de 5 significa que la base se incrementará 5 unidades por periodo. Suponga que, en realidad, hay una tendencia multiplicativa. Por lo tanto, si la estimación actual de la base es 50 y la estimación actual de la tendencia es 1.2, pronosticaríamos un incremento en la demanda de 20% por periodo. De modo que pronosticaríamos la demanda del siguiente periodo como 50(1.2) y la demanda dos periodos en el futuro como $50(1.2)^2$. Si queremos utilizar una tendencia multiplicativa en el método de Holt debemos usar las ecuaciones siguientes:

$$L_t = \alpha * (x_t) + (1 - \alpha) * (I)$$
$$T_t = \beta * (II) + (1 - \beta) * T_{t-1}$$

a Determine qué deben ser *I* y *II*.

b Suponga que estamos trabajando con información mensual y que el mes 12 es diciembre, el mes 13 es enero y así sucesivamente. Asimismo, suponga que $L_{12} = 100$ y $T_{12} = 1.2$, y, además, que $x_{13} = 200$. Al final del mes 13, ¿cuál es el pronóstico para x_{15}? Suponga que $\alpha = \beta = 0.5$.

16 Es posible utilizar una versión del suavizamiento exponencial simple para predecir el resultado de los eventos deportivos. Para ejemplificarlo, considere el futbol americano profesional. Primero suponemos que todos los partidos se juegan en un campo neutral. El día anterior al juego, suponemos que cada equipo tiene una calificación. Por ejemplo, si la calificación de los Osos es +10 y la de los Bengalíes es de +6, pronosticaríamos que los Osos vencen a los Bengalíes por 10 − 6 = 4 puntos. Suponga que los Osos juegan contra los Bengalíes y ganan por 20 puntos. Por lo que se refiere a estas observaciones, estaríamos dando un "subpronóstico" para el desempeño de los Osos de 20 − 4 = 16 puntos. La mejor α para el futbol profesional es 0.10. Después del juego aumentamos, por lo tanto, la calificación de los Osos en 16(0.1) = 1.6, y disminuimos la calificación de los Bengalíes en 1.6 puntos. En un nuevo encuentro, los Osos serían los favoritos por (10 + 1.6) − (6 − 1.6) = 7.2 puntos.

a ¿De qué modo se relaciona este enfoque con la ecuación $A_t = A_{t-1} + \alpha(e_t)$?

b Suponga que la ventaja por jugar en casa en el futbol americano profesional es de 3 puntos; es decir, los equipos locales tienden a superar en el puntaje a los equipos visitantes por un promedio de 3 puntos por juego. ¿Cómo podría la ventaja de jugar en casa ser incorporado en el sistema?

c ¿Cómo podríamos determinar la mejor α para el futbol americano profesional?

d ¿Cómo podríamos determinar calificaciones para cada equipo al iniciar la temporada?

e Suponga que pretendemos aplicar el método del inciso anterior para predecir lo que sucede en el futbol profesional americano (calendario de 16 partidos), futbol americano colegial (calendario de 11 partidos), basquetbol colegial (calendario de 30 juegos) y basquetbol profesional (calendario de 82 partidos). ¿Qué deporte tendría la α mínima óptima? ¿Qué deporte tendría la α máxima óptima?

f ¿Por qué probablemente este enfoque generaría pronósticos malos para las ligas mayores de beisbol?

24.5 Pronósticos *ad hoc*

Suponga que desea determinar cuántos cajeros deben estar trabajando cada día en un banco con el fin de brindar un servicio adecuado. Con objeto de usar los modelos de colas del capítulo 20 para contestar esta pregunta, necesita ser capaz de pronosticar la cantidad de clientes que llegarán diario al banco. El gerente del banco opina que el mes del año y el día de la semana influyen en el número de clientes que llegan al banco. (El banco está abierto de lunes a sábado, excepto los días feriados.) ¿Será posible desarrollar un modelo sencillo de pronóstico para ayudar al banco a pronosticar la cantidad de clientes que acudirán diario?

La cantidad de clientes que llegó al banco diario durante el año anterior se presenta en la tabla 7. Hemos usado 1 = lunes, 2 = martes, . . . , 6 = sábado y 7 = domingo para designar los días de la semana. Una "Y" en la columna AH quiere decir que ese día es el día posterior a que el banco estuviera cerrado por ser día feriado.

Sea x_t = número de clientes que llegan al banco el día t. Planteamos que $x_t = B \times DW_t \times M_t \times \varepsilon_t$, donde

B = nivel base del tránsito de clientes que corresponde a un día promedio
DW_t = factor del día de la semana que corresponde al día de la semana en el cual cae t.
M_t = factor del mes que corresponde al mes durante el cual ocurre el día t
ε_t = término del error aleatorio cuyo valor promedio es igual a 1.

TABLA 7
Clientes que llegan al banco

Mes	Día M	Día W	Cliente	AH	Pronóstico
1	1	1			
1	2	2	431	Y	399.13
1	3	3	271		415.88
1	4	4	362		416.51
1	5	5	696		560.10
1	6	6	315		356.32
1	7	7			
1	8	1	330		493.98
1	9	2	352		399.13
1	10	3	606		415.88
1	11	4	550		416.51
1	12	5	626		560.10
1	13	6	392		356.32
1	14	7			
1	15	1	540		493.98
1	16	2	474		399.13
1	17	3	457		415.88
1	18	4	401		416.51
1	19	5	691		560.10
1	20	6	388		356.32
1	21	7			
1	22	1	533		493.98
1	23	2	384		399.13
1	24	3	360		415.88
1	25	4	515		416.51
1	26	5	325		560.10
1	27	6	412		356.32
1	28	7			
1	29	1	592		493.98
1	30	2	366		399.13
1	31	3	512		415.88
2	1	4	476		425.33
2	2	5	531		571.97
2	3	6	303		363.87
2	4	7			
2	5	1	474		504.45
2	6	2	255		407.58
2	7	3	282		424.69
2	8	4	321		425.33
2	9	5	416		571.97
2	10	6	257		363.87
2	11	7			
2	12	1	638		504.45
2	13	2	506		407.58
2	14	3	420		424.69
2	15	4	459		425.33
2	16	5	515		571.97

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	AH	Pronóstico
2	17	6	501		363.87
2	18	7			
2	19	1	556		504.45
2	20	2	510		407.58
2	21	3	436		424.69
2	22	4	512		425.33
2	23	5	547		571.97
2	24	6	319		363.87
2	25	7			
2	26	1	637		504.45
2	27	2	474		407.58
2	28	3	487		424.69
2	29	4	402		425.33
3	1	5	778		574.26
3	2	6	374		365.32
3	3	7			
3	4	1	544		506.46
3	5	2	485		409.21
3	6	3	361		426.39
3	7	4	315		427.03
3	8	5	423		574.26
3	9	6	357		365.32
3	10	7			
3	11	1	649		506.46
3	12	2	351		409.21
3	13	3	405		426.39
3	14	4	404		427.03
3	15	5	483		574.26
3	16	6	411		365.32
3	17	7			
3	18	1	309		506.46
3	19	2	453		409.21
3	20	3	515		426.39
3	21	4	380		427.03
3	22	5	426		574.26
3	23	6	427		365.32
3	24	7			
3	25	1	489		506.46
3	26	2	341		409.21
3	27	3	471		426.39
3	28	4	517		427.03
3	29	5	647		574.26
3	30	6	415		365.32
3	31	7			
4	1	1	363		483.02
4	2	2	337		390.27
4	3	3	314		406.65

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	AH	Pronóstico
4	4	4	465		407.26
4	5	5	584		547.67
4	6	6	313		348.41
4	7	7			
4	8	1	376		483.02
4	9	2	292		390.27
4	10	3	484		406.65
4	11	4	227		407.26
4	12	5	496		547.67
4	13	6	395		348.41
4	14	7			
4	15	1	625		483.02
4	16	2	430		390.27
4	17	3	454		406.65
4	18	4	372		407.26
4	19	5	455		547.67
4	20	6	253		348.41
4	21	7			
4	22	1	432		483.02
4	23	2	469		390.27
4	24	3	392		406.65
4	25	4	467		407.26
4	26	5	684		547.67
4	27	6	349		348.41
4	28	7			
4	29	1	750		483.02
4	30	2	409		390.27
5	1	3	348		373.31
5	2	4	230		373.88
5	3	5	630		502.78
5	4	6	358		319.85
5	5	7			
5	6	1	269		443.43
5	7	2	107		358.27
5	8	3	360		373.31
5	9	4	208		373.88
5	10	5	547		502.78
5	11	6	325		319.85
5	12	7			
5	13	1	473		443.43
5	14	2	337		358.27
5	15	3	317		373.31
5	16	4	341		373.88
5	17	5	338		502.78
5	18	6	369		319.85
5	19	7			
5	20	1	618		443.43

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	AH	Pronóstico
5	21	2	458		358.27
5	22	3	457		373.31
5	23	4	572		373.88
5	24	5	668		502.78
5	25	6	318		319.85
5	26	7			
5	27	1	300		443.43
5	28	2	469		358.27
5	29	3	434		373.31
5	30	4	419		373.88
5	31	5			
6	1	6	432	Y	354.08
6	2	7			
6	3	1	463		490.89
6	4	2	457		396.62
6	5	3	273		413.27
6	6	4	327		413.90
6	7	5	554		556.60
6	8	6	256		354.08
6	9	7			
6	10	1	465		490.89
6	11	2	479		396.62
6	12	3	437		413.27
6	13	4	585		413.90
6	14	5	616		556.60
6	15	6	318		354.08
6	16	7			
6	17	1	724		490.89
6	18	2	390		396.62
6	19	3	550		413.27
6	20	4	266		413.90
6	21	5	410		556.60
6	22	6	303		354.08
6	23	7			
6	24	1	514		490.89
6	25	2	353		396.62
6	26	3	397		413.27
6	27	4	539		413.90
6	28	5	411		556.60
6	29	6	413		354.08
6	30	7			
7	1	1	583		484.44
7	2	2	477		391.42
7	3	3	410		407.85
7	4	4			
7	5	5	615	Y	549.29
7	6	6	288		349.44

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	AH	Pronóstico
7	7	7			
7	8	1	478		484.44
7	9	2	298		391.42
7	10	3	253		407.85
7	11	4	366		408.46
7	12	5	410		549.29
7	13	6	270		349.44
7	14	7			
7	15	1	541		484.44
7	16	2	331		391.42
7	17	3	318		407.85
7	18	4	441		408.46
7	19	5	651		549.29
7	20	6	300		349.44
7	21	7			
7	22	1	608		484.44
7	23	2	401		391.42
7	24	3	390		407.85
7	25	4	391		408.46
7	26	5	619		549.29
7	27	6	391		349.44
7	28	7			
7	29	1	413		484.44
7	30	2	474		391.42
7	31	3	503		407.85
8	1	4	267		418.33
8	2	5	619		562.56
8	3	6	370		357.88
8	4	7			
8	5	1	406		496.15
8	6	2	432		400.87
8	7	3	333		417.70
8	8	4	327		418.33
8	9	5	647		562.56
8	10	6	407		357.88
8	11	7			
8	12	1	396		496.15
8	13	2	664		400.87
8	14	3	508		417.70
8	15	4	519		418.33
8	16	5	555		562.56
8	17	6	365		357.88
8	18	7			
8	19	1	492		496.15
8	20	2	420		400.87
8	21	3	360		417.70
8	22	4	469		418.33

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	All	Pronóstico
8	23	5	488		562.56
8	24	6	326		357.88
8	25	7			
8	26	1	465		496.15
8	27	2	384		400.87
8	28	3	280		417.70
8	29	4	292		418.33
8	30	5	649		562.56
8	31	6	493		357.88
9	1	7			
9	2	1			
9	3	2	459	Y	391.76
9	4	3	353		408.21
9	5	4	287		408.82
9	6	5	471		549.77
9	7	6	266		349.74
9	8	7			
9	9	1	505		484.87
9	10	2	528		391.76
9	11	3	342		408.21
9	12	4	551		408.82
9	13	5	525		549.77
9	14	6	304		349.74
9	15	7			
9	16	1	479		484.87
9	17	2	258		391.76
9	18	3	263		408.21
9	19	4	450		408.82
9	20	5	540		549.77
9	21	6	297		349.74
9	22	7			
9	23	1	399		484.87
9	24	2	264		391.76
9	25	3	479		408.21
9	26	4	459		408.82
9	27	5	915		549.77
9	28	6	247		349.74
9	29	7			
9	30	1	725		484.87
10	1	2	197		390.39
10	2	3	326		406.78
10	3	4	374		407.39
10	4	5	477		547.85
10	5	6	367		348.52
10	6	7			
10	7	1	317		483.17
10	8	2	205		390.39

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	AH	Pronóstico
10	9	3	519		406.78
10	10	4	483		407.39
10	11	5	489		547.85
10	12	6	345		348.52
10	13	7			
10	14	1	660		483.17
10	15	2	262		390.39
10	16	3	395		406.78
10	17	4	522		407.39
10	18	5	582		547.85
10	19	6	335		348.52
10	20	7			
10	21	1	503		483.17
10	22	2	396		390.39
10	23	3	548		406.78
10	24	4	471		407.39
10	25	5	528		547.85
10	26	6	344		348.52
10	27	7			
10	28	1	419		483.17
10	29	2	429		390.39
10	30	3	609		406.78
10	31	4	519		407.39
11	1	5	674		596.31
11	2	6	352		379.35
11	3	7			
11	4	1	360		525.91
11	5	2	500		424.92
11	6	3	339		442.76
11	7	4	326		443.43
11	8	5	459		596.31
11	9	6	255		379.35
11	10	7			
11	11	1	432		525.91
11	12	2	527		424.92
11	13	3	394		442.76
11	14	4	424		443.43
11	15	5	388		596.31
11	16	6	356		379.35
11	17	7			
11	18	1	635		525.91
11	19	2	309		424.92
11	20	3	613		442.76
11	21	4	580		443.43
11	22	5	627		596.31
11	23	6	514		379.35
11	24	7			

(Continúa)

TABLA 7
(Continuación)

Mes	Día M	Día W	Cliente	All	Pronóstico
11	25	1	686		525.91
11	26	2	452		424.92
11	27	3	384		442.76
11	28	4			
11	29	5	701	Y	596.31
11	30	6	425		379.35
12	1	7			
12	2	1	291		510.06
12	3	2	407		412.12
12	4	3	458		429.42
12	5	4	243		430.06
12	6	5	449		578.34
12	7	6	315		367.91
12	8	7			
12	9	1	633		510.06
12	10	2	429		412.12
12	11	3	375		429.42
12	12	4	540		430.06
12	13	5	615		578.34
12	14	6	455		367.91
12	15	7			
12	16	1	385		510.06
12	17	2	472		412.12
12	18	3	576		429.42
12	19	4	321		430.06
12	20	5	679		578.34
12	22	7			
12	23	1	407		510.06
12	24	2	328		412.12
12	25	3			
12	26	4	491	Y	430.06
12	27	5	586		578.34
12	28	6	367		367.91
12	29	7			
12	30	1	707		510.06
12	31	2	400		412.12

Para empezar, estimamos B = número promedio de clientes por día en que el banco está abierto = 438.33. Enseguida se ilustra la estimación de DW_t mediante

$$DW_t \text{ para el lunes} = \frac{\text{número promedio de clientes en los lunes en que el banco está abierto}}{B}$$

$$= \frac{492.07}{438.33} = 1.122$$

De igual manera, tenemos

$$\begin{aligned}
DW_t \text{ para el martes } &= 0.907\\
DW_t \text{ para el miércoles } &= 0.945\\
DW_t \text{ para el jueves } &= 0.947\\
DW_t \text{ para el viernes } &= 1.273\\
DW_t \text{ para el sábado } &= 0.809
\end{aligned}$$

Si deseamos estimar M_t (digamos, para mayo), escribimos

$$M_t \text{ para mayo} = \frac{\text{número promedio de llegadas en el día de mayo en que el banco está abierto}}{B}$$

$$= \frac{395}{438.33} = 0.901$$

Determinamos, de modo igual, que M_t para los meses restantes es:

$$\begin{aligned}
M_t \text{ para enero } &= 1.004\\
M_t \text{ para febrero } &= 1.025\\
M_t \text{ para marzo } &= 1.029\\
M_t \text{ para abril } &= 0.982\\
M_t \text{ para junio } &= 0.998\\
M_t \text{ para julio } &= 0.984\\
M_t \text{ para agosto } &= 1.008\\
M_t \text{ para septiembre } &= 0.985\\
M_t \text{ para octubre } &= 0.982\\
M_t \text{ para noviembre } &= 1.069\\
M_t \text{ para diciembre } &= 1.037
\end{aligned}$$

Con el fin de mostrar cómo se generaron los pronósticos de la tabla 7, considere cómo obtendríamos un pronóstico del número de clientes que entran al banco el jueves 1 de febrero del presente año. Suponiendo que ε_t es igual a su valor promedio de 1, pronosticaríamos que entrarían $B \times (DW_t$ para el jueves$) \times (M_t$ para febrero$) = 438.33(0.947)(1.025) =$ 425.48 clientes. (La diferencia entre los resultados impresos de la tabla se debe al redondeo de los valores DW_t y M_t.) Para pronosticar las llegadas de los clientes para un día futuro (por ejemplo, sábado 8 de febrero del año próximo), tendríamos $B \times (DW_t$ para el sábado$) \times (M_t$ para febrero$) = 438.33\ (0.809)(1.025) = 363.47$ clientes.

En cuanto a los datos dados en la tabla 7, el modelo simple generó un MAD de 79.1. Si este método se utilizara para generar pronósticos para el año venidero, es probable que el MAD fuera mayor que 79.1. Esto es así porque tenemos que ajustar los parámetros a los datos anteriores; no hay garantía de que los datos futuros "sepan" que deben seguir el mismo patrón que los datos anteriores. Asimismo, no hemos considerado de si está presente o no una tendencia ascendente en los datos (véase problema 3).

Suponga que el gerente del banco observa que el día siguiente a un día feriado el tráfico en el banco es superior al modelo predicho. La información de la tabla 8 señala que en verdad éste es el caso. ¿Cómo podemos aprovechar estos datos para obtener pronósticos más precisos acerca de la clientela para los días posteriores a los días feriados? De acuerdo con la tabla 8, el valor promedio de Real/Pronosticado para los días posteriores a los días feriados es 1.15. Por lo tanto, para cualquier día después de un día feriado obtenemos un nuevo pronóstico multiplicando simplemente nuestro pronóstico anterior por 1.15.

TABLA 8

Tráfico en el banco el día siguiente a un día feriado

Día siguiente a un día feriado	Real	Pronóstico (redondeado)	Real/Pronóstico
2 de enero	431	399	1.08
1 de junio	432	354	1.22
5 de julio	615	549	1.12
3 de septiembre	459	392	1.17
29 de noviembre	701	596	1.18
26 de diciembre	491	430	1.14

PROBLEMAS

Grupo A

1 Suponga que el banco es una unión de crédito de una universidad y que las llegadas al banco son mayores a lo usual en los días cuando los profesores de la universidad reciben su pago. Suponiendo que los profesores universitarios reciben su pago el primer día de la semana de cada mes, ¿cómo podemos incorporar este hecho en el procedimiento de pronóstico explicado en esta sección?

2 Suponga otra vez que el banco es una unión de crédito de una universidad, pero ahora el personal recibe su pago cada dos viernes. De nuevo, el tráfico en el banco es mucho mayor en los días de pago que lo acostumbrado. ¿Cómo podríamos incorporar este hecho en el procedimiento de pronóstico que se explica en esta sección?

3 Suponga que la cantidad de clientes que llega al banco está aumentando alrededor de 20% por año. ¿Cómo podríamos incorporar este hecho en el procedimiento de pronóstico que se explica en esta sección?

24.6 Regresión lineal simple

A menudo tratamos de predecir el valor de una variable (llamada **variable dependiente**) a partir del valor de otra variable (**variable independiente**). Algunos ejemplos son

Variable dependiente	Variable independiente
Ventas de productos	Precio del producto
Ventas de automóviles	Tasa de interés
Costo de producción total	Unidades fabricadas

Si la variable dependiente y la variable independiente se relacionan de un modo lineal, entonces se puede aplicar la regresión lineal simple para estimar esta relación. En la sección 24.7 se trata cómo estimar relaciones no lineales.

Recordemos el problema de Giapetto (ejemplo 1 del capítulo 3) con el fin de ilustrar la regresión lineal simple. Para plantear este problema necesitamos determinar el costo de producción de un soldado y el costo de producción de un tren. Supongamos que deseamos estimar el costo de producción de un tren. Para hacerlo, hemos observado durante diez semanas el número de trenes producidos cada semana y el costo total de producción de aquellos trenes. Esta información se proporciona en la tabla 9.

Los datos de la tabla 9 se grafican en la figura 9. Obsérvese que, al parecer, hay una fuerte relación lineal entre x_i (número de trenes producidos durante la semana i) y y_i (costo de producción de los trenes hechos durante la semana i). La recta trazada en la figura 9 parece, en un modo que se precisa más adelante, apegarse a la relación lineal entre unidades producidas y costos de producción. Pronto se explica cómo se escoge esta recta.

TABLA 9

Información de los costos por semana de los trenes

Semana	Trenes fabricados	Costo de producción de los trenes
1	10	\$257.40
2	20	\$601.60
3	30	\$782.00
4	40	\$765.40
5	45	\$895.50
6	50	\$1133.00
7	60	\$1152.80
8	55	\$1132.70
9	70	\$1459.20
10	40	\$970.10

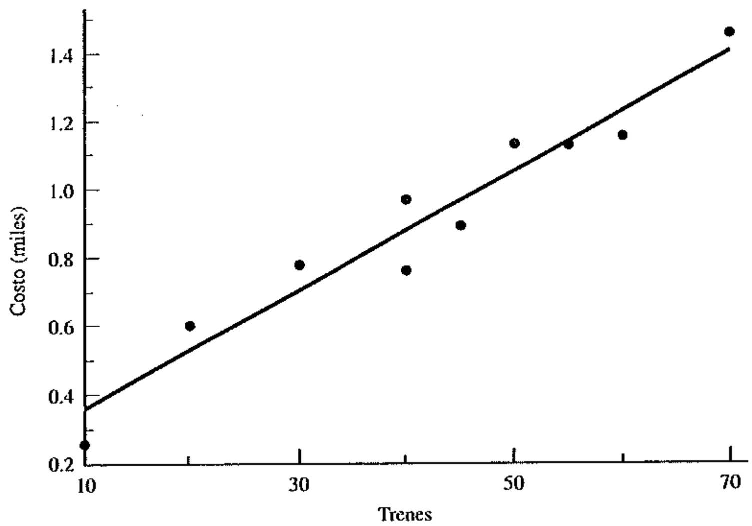


FIGURA 9
Gráfica de dispersión del costo de producción de los trenes

Para iniciar modelamos la relación lineal entre x_i y y_i mediante la ecuación siguiente

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

donde ε_i es un término de error que representa el hecho de que en una semana durante la cual se producen x_i trenes, el costo de producción podría no ser siempre igual a $\beta_0 + \beta_1 x_i$. Si $\varepsilon_i > 0$, el costo de producción de x_i trenes durante la semana i será mayor que $\beta_0 + \beta_1 x_i$, en tanto que si $\varepsilon_i < 0$, el costo de producción de x_i trenes durante la semana i será menor que $\beta_0 + \beta_1 x_i$. Sin embargo, esperamos que ε_i alcance un promedio de 0, de modo que el costo esperado durante una semana en la cual x_i trenes se fabrican es $\beta_0 + \beta_1 x_i$.

Los valores verdaderos de β_0 y β_1 son desconocidos. Supongamos que estimamos β_0 por medio de $\hat{\beta}_0$ y determinamos β_1 por medio de $\hat{\beta}_1$. Entonces las predicciones para y_i (puesto que el valor promedio de $\varepsilon_i = 0$) se obtienen mediante $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

Suponga que tenemos puntos dato de la forma $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$. ¿Cómo debemos escoger los valores de $\hat{\beta}_0$ y $\hat{\beta}_1$ que generan buenas estimaciones de β_0 y β_1? Seleccionamos valores de $\hat{\beta}_0$ y $\hat{\beta}_1$ que hagan las predicciones $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ cercanas a los puntos dato reales (x_i, y_i). Para formalizar esta idea, defina e_i = error o residuo para los puntos dato i = (costo real y_i) − (costo pronosticado $\hat{y}_i$) $= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$. Ahora escogemos $\hat{\beta}_0$ y $\hat{\beta}_1$ para que minimicen

$$F(\hat{\beta}_0, \hat{\beta}_1) = \sum e_i^2 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Los valores $\hat{\beta}_0$ y $\hat{\beta}_1$ que minimizan $F(\hat{\beta}_0, \hat{\beta}_1)$ se denominan las **estimaciones de mínimos cuadrados** de β_0 y β_1. Como se explicó en el ejemplo 19 del capítulo 11, determinamos $\hat{\beta}_0$ y $\hat{\beta}_1$ estableciendo

$$\frac{\partial F}{\partial \hat{\beta}_0} = \frac{\partial F}{\partial \hat{\beta}_1} = 0$$

Los valores resultantes de $\hat{\beta}_0$ y $\hat{\beta}_1$ se obtienen mediante

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \tag{18}$$

donde $\bar{x}$ = valor promedio de todas las x_i y $\bar{y}$ = valor promedio de todas las y_i.

A la ecuación $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ se le llama **recta de regresión de mínimos cuadrados.** En esencia, si la recta de mínimos cuadrados se ajusta muy bien a los puntos (en un sentido que se precisa más adelante), entonces usaremos $\hat{\beta}_0 + \hat{\beta}_1 x_i$ como nuestra la predicción para y_i.

La recta de mínimos cuadrados se determina, por lo regular, por medio de computadora. Excel, Minitab y muchos otros paquetes muy conocidos proporcionan $\hat{\beta}_0$ y $\hat{\beta}_1$. Sin embargo, por cuestiones de completitud, los cálculos necesarios para estimar $\hat{\beta}_0$ y $\hat{\beta}_1$ para los datos de la tabla 9 se dan en la tabla 10, donde usamos

$$\bar{x} = \frac{\sum x_i}{10} = 42 \qquad \text{y} \qquad \bar{y} = \frac{\sum y_i}{10} = 914.97$$

De acuerdo con la tabla 10 (la cual se puede organizar en una hoja de cálculo), determinamos que $\sum(x_i - \bar{x})(y_i - \bar{y}) = 53\,756.6$ y $\sum(x_i - \bar{x})^2 = 3010$. Con la ecuación (18) encontramos que

$$\hat{\beta}_1 = \frac{53756.6}{3010} = 17.86 \qquad \text{y} \qquad \hat{\beta}_0 = 914.97 - (17.86)42 = 164.88$$

La recta de mínimos cuadrados es $\hat{y} = 164.88 + 17.86x$. Por lo tanto, estimamos que cada tren extra genera un costo variable de $\hat{\beta}_1 = \$17.86$.

Los pronósticos y los errores para las diez semanas se dan en la tabla 11. Con el fin de ilustrar los cálculos, considere el primer punto (10 257.4). El costo pronosticado es $\hat{y}_1 = 164.88 + 17.86(10) = 343.5$ y el error es $e_1 = 257.4 - 343.5 = -86.1$.

Todas las rectas de mínimos cuadrados tienen dos propiedades:

1 Pasa por el punto $(\bar{x}, \bar{y})$. Por lo tanto, durante una semana en la cual Giapetto fabricara $\bar{x} = 42$ trenes, pronosticaríamos que la producción de estos trenes costaría 914.97 dólares.

TABLA 10
Cálculos de $\hat{\beta}_0$ y $\hat{\beta}_1$ para los datos de los costos de los trenes

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
10	257.4	−32	−657.57	21 042.24	1024
20	601.6	−22	−313.37	6 894.14	484
30	782.0	−12	−132.97	1 595.64	144
40	765.4	−2	−149.57	299.14	4
45	895.5	3	−19.47	−58.41	9
50	1 133.0	8	218.03	1 744.24	64
60	1 152.8	18	237.83	4 280.94	324
55	1 132.7	13	217.73	2 830.49	169
70	1 459.2	28	544.23	15 238.44	784
40	970.1	−2	55.13	−110.26	4

TABLA 11
Cálculo de los errores

x_i	y_i	$\hat{y}_i$	e_i
10	257.4	343.5	−86.1
20	601.6	522.1	79.5
30	782.0	700.7	81.3
40	765.4	879.3	−113.9
45	895.5	968.5	−73.0
50	1 133.0	1 057.8	75.2
60	1 152.8	1 236.4	−83.6
55	1 132.7	1 147.1	−14.4
70	1 459.2	1 415	44.2
40	970.1	879.3	90.8

2 $\sum e_i = 0$. La recta de mínimos cuadrados "divide" a los puntos dato en el sentido de que la suma de las distancias verticales desde los puntos que se localizan por arriba de la recta de mínimos cuadrados hasta dicha recta, es igual a la suma de las distancias verticales desde los puntos que se encuentran abajo de la misma recta.

¿Cómo lograr un buen ajuste?

¿De qué modo determinar qué tan bien se ajusta la recta de mínimos cuadrados a nuestros datos? Para contestar a esta pregunta necesitamos analizar tres componentes de la variación: **suma del total de cuadrados** (*sum of squares total*, SST), **suma de los errores cuadráticos** (*sum of squares error*, SSE) y **suma de los cuadrados de la regresión** (*sum of squares regression*, SSR). La suma total de cuadrados se obtiene con $\text{SST} = \sum(y_i - \bar{y})^2$. SST mide la variación total de y_i respecto a su media $\bar{y}$. La suma de los errores cuadráticos es $\text{SSE} = \sum(y_i - \hat{y}_i)^2 = \sum e_i^2$. Si la recta de mínimos cuadrados pasa por todos los puntos dato, SSE = 0. Por lo tanto, una SSE pequeña indicaría que la recta de mínimos cuadrado se ajusta muy bien a los datos. Definimos la suma de los cuadrados de la regresión como $\text{SSR} = \sum(\hat{y}_i - \bar{y})^2$. Se puede demostrar que

$$\text{SST} = \text{SSR} + \text{SSE} \tag{19}$$

Obsérvese que SST es una función sólo de los valores de y. Si hay un buen ajuste, SSE tendrá un valor pequeño, de modo que (19) muestra que SSR tendrá un valor grande en un buen ajuste. Podríamos definir, con más formalidad, el **coeficiente de determinación** (R^2) para y mediante:

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \text{porcentaje de variación en } y \text{ explicado por } x$$

En forma equivalente, la ecuación (19) permite expresar

$$1 - R^2 = \frac{\text{SSE}}{\text{SST}} = \text{porcentaje de variación en } y \text{ no explicado por } x$$

Según los resultados que da la computadora, vemos que SST = 1 021 762 y SSE = 61 705. Entonces, con la ecuación (19) se obtiene SSR = SST − SSE = 960 057. Por lo tanto, encontramos que $R^2 = \frac{960\,057}{1\,021\,762} = 0.94$. Esto significa que la cantidad de trenes fabricados durante una semana explica el 94% de la variación en el costo semanal de la producción de trenes. Todos los otros factores combinados explican a lo más 6% de la variación en el costo semanal, de modo que podemos tener la seguridad de que la relación lineal entre x y y es fuerte.

Una medida de la asociación lineal entre x y y es la **correlación lineal de la muestra** r_{xy}. Una correlación de la muestra cercana a +1 indica una fuerte relación lineal positiva entre x y y; una correlación de la muestra próxima a -1 significa una fuerte relación lineal negativa entre x y y; y una correlación de la muestra cercana a 0 quiere decir que existe una relación lineal débil entre x y y.

A propósito, si $\hat{\beta}_1 \geq 0$, entonces r_{xy} es igual a $+\sqrt{R^2}$, en tanto que si $\hat{\beta}_1 \leq 0$, la correlación de la muestra entre x y y se obtiene con $-\sqrt{R^2}$. Por lo tanto, en el ejemplo de los costos, $r_{xy} = \sqrt{0.94} = 0.97$, lo cual significa que existe una estrecha relación lineal entre x y y.

Precisión del pronóstico

Una medida de la precisión de los pronósticos derivados de la regresión se obtiene mediante el **error estándar de la estimación** (s_e). Si hacemos n = número de observaciones, s_e es

$$s_e = \sqrt{\frac{\text{SSE}}{n-2}}$$

Para los datos del ejemplo,

$$s_e = \sqrt{\frac{61\,705}{10-2}} = 87.8$$

Por lo regular se cumple[†] que alrededor de 68% de los valores de y están a no menos de s_e del valor pronosticado $\hat{y}$, y 95% de los valores de y estarán dentro de $2s_e$ del valor pronosticado $\hat{y}$. En el ejemplo presente, contamos con que 68% de las estimaciones de nuestros costos están cerca de 87.80 dólares del costo verdadero, y 95% estarán cerca de 175.60 dólares. En realidad, por lo que toca a 80% de los puntos, el costo real está cerca de s_e del costo pronosticado, y para 100% de los puntos, el costo real está a no menos de $2s_e$ del costo pronosticado.

Cualquier observación para la cual y no está dentro de $2s_e$ de $\hat{y}$ se denomina **valor atípico.** Los valores atípicos representan puntos raros, y se les debe examinar con mucho cuidado. Naturalmente, si un valor atípico es el resultado de un error al introducir los datos, se debe corregir. Si un valor atípico es de alguna manera extraño respecto a los puntos restantes, sería mejor omitirlo, y volver a estimar la recta de mínimos cuadrados. Puesto que todos los errores son menores que $2s_e$ en valor absoluto, no hay valores atípicos en nuestro ejemplo.

Pruebas *t* en regresión

La significancia de una relación lineal se prueba mediante una **prueba *t*.** Para probar H_0: $\beta_1 = 0$ (no existe relación lineal significativa entre x y y) contra H_a: $\beta_1 \neq 0$ (existe una relación lineal significativa entre x y y) en un nivel de significancia α, la variable t se calcula mediante

$$t = \frac{\hat{\beta}_1}{\text{EE}(\hat{\beta}_1)}$$

[†]De hecho, alrededor de 68% de los puntos debe estar dentro de

$$s_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

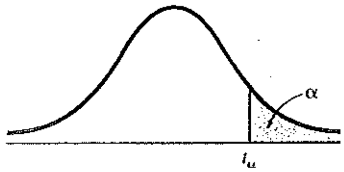
de $\hat{y}$, y 95% de los puntos debe estar dentro de

$$2s_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$$

de $\hat{y}$.

TABLA 12

Puntos del porcentaje de la distribución t †



df	$a = 0.1$	$a = 0.05$	$a = 0.025$	$a = 0.01$	$a = 0.005$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
240	1.285	1.651	1.970	2.342	2.596
inf.	1.282	1.645	1.960	2.326	2.576

†Calculada por P. J. Hildebrand. Reimpreso con autorización de PWS-KENT Publishing Company.

$EE(\hat{\beta}_1)$ mide la incertidumbre de la estimación de β_1; por lo regular se le encuentra en los resultados que proporciona la computadora. Rechazamos H_0 si $|t| \geq t_{(\alpha/2, n-2)}$, donde $t_{(\alpha/2,n-2)}$ se obtiene de la tabla 12. Para el ejemplo de los costos, $EE(\hat{\beta}_1) = 1.6$ (tomado de los resultados impresos de computadora), de modo que $t = \frac{17.86}{1.6} = 11.16$. Con $\alpha = 0.05$, tenemos que $t_{(.025,8)} = 2.306$, por lo que rechazamos H_0 y concluimos de nuevo que hay una fuerte relación lineal entre x y y.

Suposiciones que sustentan el modelo de regresión lineal simple

El análisis estadístico del modelo de regresión lineal simple requiere las siguientes suposiciones.

Suposición 1

La varianza del término de error no debe depender del valor de la variable independiente x. Esta suposición recibe el nombre de **homocedasticidad.** Si la varianza del término de error depende de x, entonces se dice que está presente la **heterocedasticidad**. Para ver si se cumple la suposición de la homocedasticidad, se grafican los errores en el eje de las y y el valor de x en el eje de las x. Una situación donde la suposición de homocedasticidad se satisface se ilustra en la figura 10; en la figura se manifiesta que no existe tendencia a que el tamaño de los errores dependan de x. En cambio, la magnitud de los errores tiende a aumentar a medida que se incrementa x en la figura 11. Éste es un ejemplo de heterocedasticidad. La heterocedasticidad se elimina, con frecuencia, usando $\ln y$, o bien, $y^{1/2}$ como variable dependiente.

Suposición 2

Los errores tienen distribución normal. Esta suposición no es esencial, de modo que no se trata más.

Suposición 3

Los errores deben ser independientes. Esta suposición se incumple a menudo cuando los datos se reúnen (como en el caso del ejemplo) sobre el tiempo. La independencia de los errores implica que conocer el valor de un error no nos dice nada respecto al valor del siguiente error (o cualquier otro).

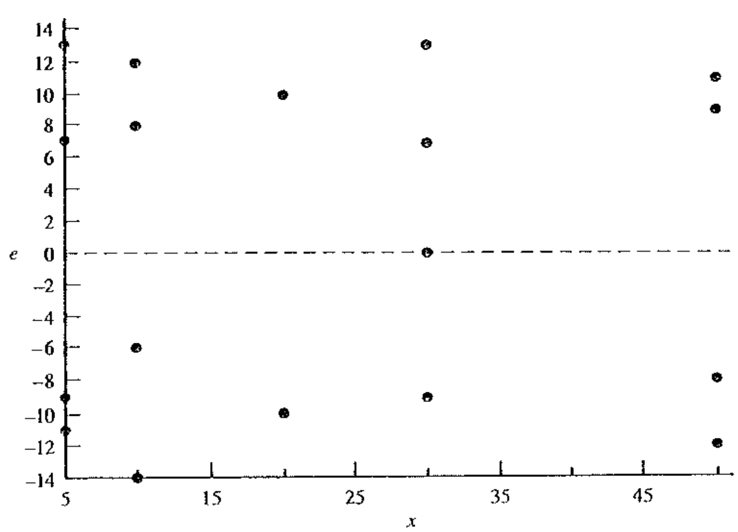


FIGURA 10
Homocedasticidad

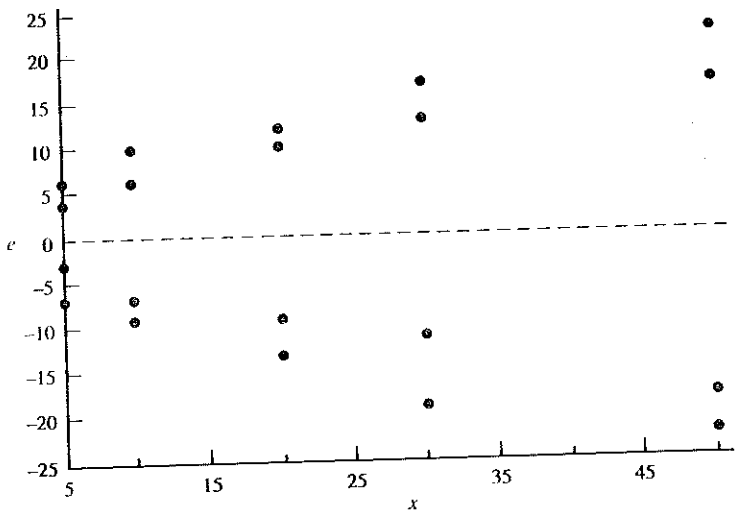


FIGURA 11
Heterocedasticidad

La validez de esta suposición se comprueba graficando los errores en una sucesión de series temporales. En la figura 12 se observa que los errores tenían los signos siguientes: $+ + + + + + - - - - - -$. Esta sucesión de errores expresa el patrón siguiente: a un error positivo (que corresponde a una subpredicción del valor real de y) le sigue otro error positivo, a un error negativo (que corresponde a una sobrepredicción del valor real de y) le sigue, por lo regular, otro error negativo. Este patrón manifiesta que errores sucesivos no son independientes; a esto se le conoce como **autocorrelación positiva**. En otras palabras, la autocorrelación positiva indica que los errores sucesivos tienen una correlación lineal positiva y no son linealmente independientes. Si la sucesión de errores en una sucesión de tiempo se parece a la figura 13, tenemos entonces una **autocorrelación negativa**. En este caso, la sucesión de errores es $+ - + - + - + - + - + -$. Esto significa que hay una tendencia a que a un error positivo le siga uno negativo, y viceversa. La conclusión es que errores sucesivos tienen una relación lineal negativa y no son independientes. La siguiente sucesión de errores: $+ + - + + - + - + + + -$. se muestra en la figura 14. No se observa ningún patrón obvio, por lo que, al parecer, se cumple la suposición de independencia. Obsérvese que los errores "alcanzan un promedio" de 0, de modo que contaríamos con que alrededor de la mitad de los errores sean positivos y la mitad sean negativos. Por lo tanto, si no hay un patrón en los errores, esperaríamos que los errores cambien de signo alrededor de la mitad de las veces. Esta observación nos permite formalizar la explicación anterior como sigue:

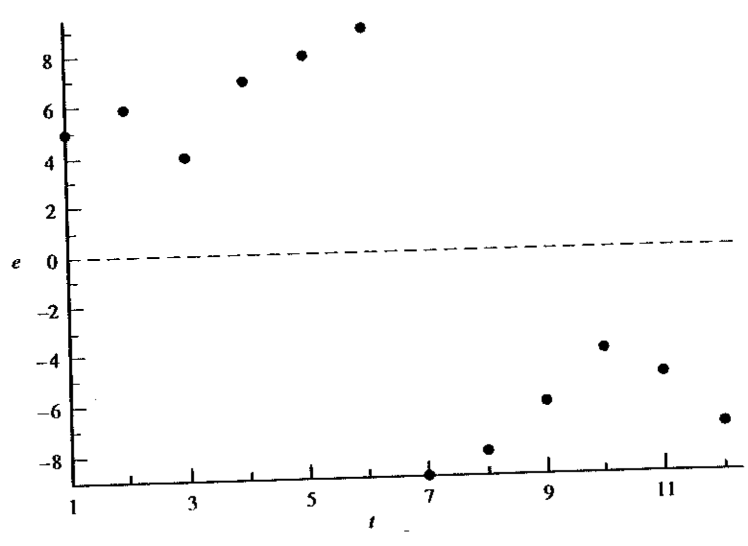


FIGURA 12
Autocorrelación positiva

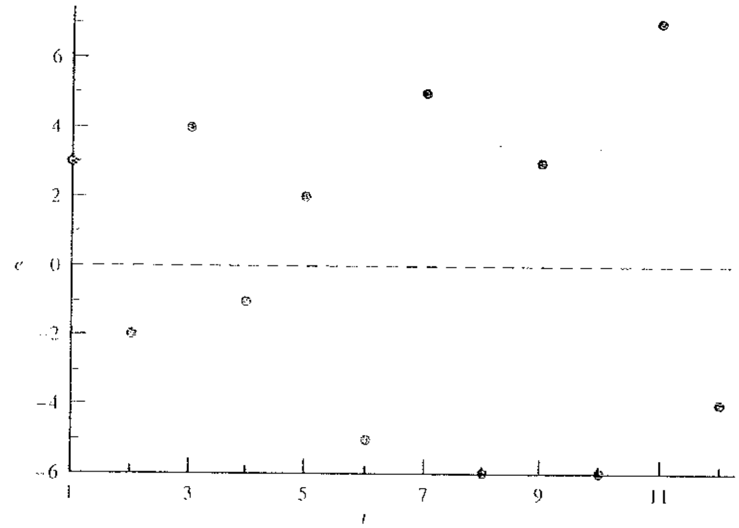


FIGURA 13
Autocorrelación negativa

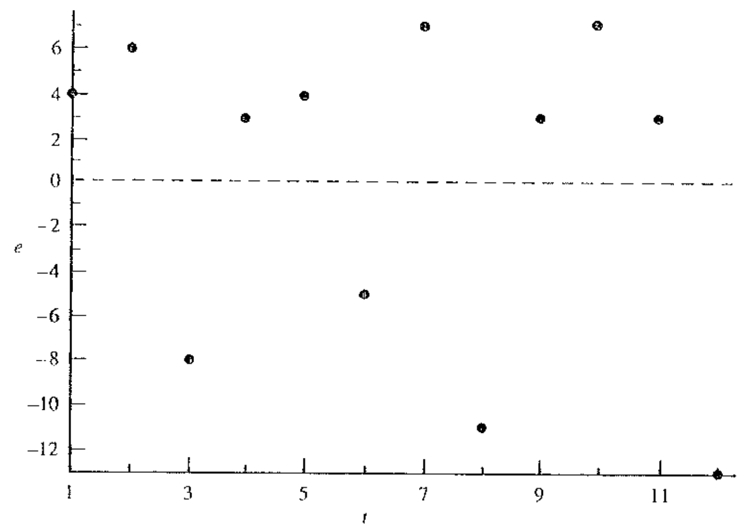


FIGURA 14
Ninguna autocorrelación

1 Si es muy raro que los errores cambien de signo (mucho menos que la mitad de las veces), entonces quizá incumplan la suposición de independencia, y probablemente esté presente la autocorrelación positiva.

2 Si los errores cambian de signo con mucha frecuencia (mucho más que la mitad de las veces), quizá incumplan la suposición de independencia, por lo que es probable que esté presente la autocorrelación negativa.

3 Si los errores cambian de signo alrededor de la mitad de las veces, es probable que satisfagan la suposición de independencia.

Si está presente la autocorrelación positiva o negativa, la corrección de la autocorrelación dará por resultado, con frecuencia, pronósticos mucho más precisos. Refiérase a las páginas 215 a 221 de Pindyck y Rubinfeld (1989) si desea más detalles.

Ejecución de regresiones con ayuda de Excel

Cost.xls

En la figura 15 (archivo Cost.xls) se ilustra el modo de ejecutar una regresión con Excel. Los datos de la tabla 9 se han escrito en el intervalo de celdas A2:B11, y luego, en Tools (Herramientas), seleccione Data Analysis (Análisis de datos), y ahí, Regression (Regre-

	A	B	C	D	E	F	G	H
1	trains	cost	yihatr	e				
2	10	257.4	9.32060032	248.0794	COST	REGRESSION		
3	20	601.6	18.6412006	582.958799	EXAMPLE			
4	30	782	27.961801	754.038199				
5	40	765.4	37.2824013	728.117599				
6	45	895.5	41.9427014	853.557299				
7	50	1133	46.6030016	1086.397				
8	60	1152.8	55.9236019	1096.8764				
9	55	1132.7	51.2633018	1081.4367				
10	70	1459.2	65.2442022	1393.9558				
11	40	970.1	37.2824013	932.817599				
12								
13								
14								
15		SUMMARY OUTPUT						
16								
17		*Regression Statistics*						
18		Multiple R	0.9693343					
19		R Square	0.9396089					
20		Adjusted R Squ	0.93206					
21		Standard Error	87.824643					
22		Observations	10					
23								
24		ANOVA						
25			*df*	*SS*	*MS*	*F*	*Significance F*	
26		Regression	1	960057.16	960057.16	124.4698887	3.72837E-06	
27		Residual	8	61705.344	7713.168			
28		Total	9	1021762.5				
29								
30			*Coefficients*	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
31		Intercept	164.87791	72.743329	2.2665708	0.053174264	-2.8686199	332.62443
32		trains	17.859336	1.6007855	11.156607	3.72837E-06	14.16791514	21.550756
33								
34								
35								
36								

FIGURA 15

sión).† Llene el cuadro de diálogo como se indica en la figura 16. En el *Y* range (el rango *Y*) B1:B11 se encuentra el nombre de la variable dependiente y los valores de la misma. En el X range (rango *X*) A1:A11 se encuentran el nombre de la variable independiente y los valores de la variable independiente. Como el primer renglón de los rangos X y Y llevan rótulos (labels), marque el cuadro de Labels. Marque la celda B15 como la esquina superior izquierda de Output Range (Rango de salida). No marque el cuadro de Residuals (Residuos). Si lo hiciera, obtendría el valor pronosticado y los residuos para cada observación. Los resultados de la regresión se presentan en la figura 15.

Examinemos qué significan los números importantes. (Se omite el análisis de las partes que son irrelevantes para nuestro estudio de la regresión).

R Square (R cuadrada) Es $r^2 = .939609$.

Multiple R (R múltiple) Es la raíz cuadrada de r^2, cuyo signo es el mismo que el de la pendiente de la recta de regresión.

Standard Error (Error estándar) Es $s_e = 87.82$.

Observations (Observaciones) Es la cantidad de puntos (10).

†Si no se muestra el Analysis Tool Pak (Análisis de datos) cuando selecciona Tools Data (Herramientas), vaya a Tools Add Ins (Complementos) y marque los cuadros de Analysis Tool Pak y Analysis Tool Pak Vba (Herramientas para análisis y Herramientas para análisis-VBA).

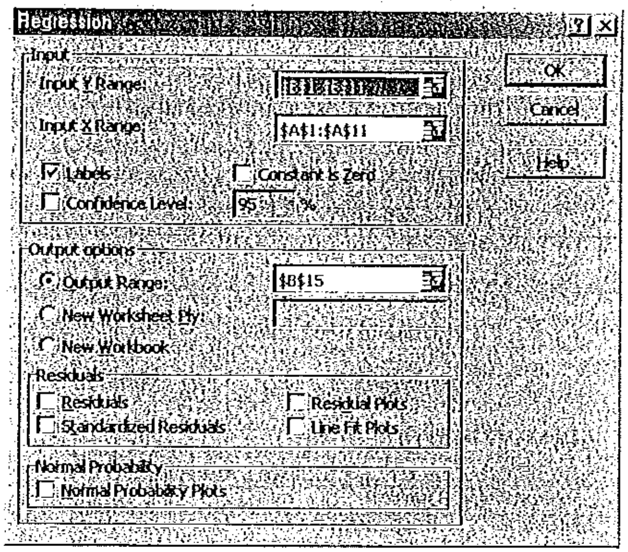


FIGURA 16

SS column (columna SS) La entrada de regresión (96 057.16) es SSR. La entrada del residuo (61 705.34) es SSE. La entrada del Total (1 021 762.5) es SST.

Coefficients column (columna de coeficientes) La entrada de la ordenada al origen (164.88) da el valor de $\hat{\beta}_0 = 164.88$, y la entrada de los trenes (17.86) da el valor de $\hat{\beta}_1 = 17.86$.

t stat (variable t) Proporciona la variable t observada (coeficiente/error estándar) para la ordenada al origen y la variable Trenes.

Standard Error column (columna del error estándar) La entrada de la ordenada al origen proporciona el error estándar $\hat{\beta}_0 = 72.74$, y la entrada de Trenes da el error estándar $\hat{\beta}_1 = 1.60$. El coeficiente dividido entre el error estándar es la variable t para la ordenada al origen o la pendiente (tabulada en la columna siguiente).

P-value (valor p) Para la ordenada al origen y la pendiente, éste da Probabilidad($|t_{n-2}| \geq$ |variable t observada|. Por ejemplo, si el valor p para los Trenes es menor que α, rechazamos H_0: $\beta_1 = 0$; si no es así, aceptamos $\beta_1 = 0$. Para $\alpha = .05$, rechazamos $\beta_1 = 0$. Para un valor de p de .05, no está muy definido si aceptamos o no la hipótesis $\beta_0 = 0$.

Se obtiene $\hat{y}_1$ en la celda C2 mediante la fórmula =D$14+A2*C$20. Determinamos e_1 en la celda D2 con la fórmula =B2−C2. Al copiar desde el intervalo C2:D2 a C2:D11 generamos pronósticos y errores para todas las observaciones.

Cómo obtener un diagrama de dispersión con Excel

Para generar un diagrama de dispersión, sea el rango X el intervalo donde está la variable independiente. Entonces, el intervalo donde está la variable dependiente es el rango Y. Luego marque X-Y (dispersión) en el cuadro que sale al seleccionar el icono de gráficas.

24.7 Ajuste de relaciones no lineales

Una gráfica de puntos de la forma (x_i, y_i) a menudo no es una función lineal de x. En estos casos, la gráfica podría indicar que hay una relación no lineal entre x y y. Por ejemplo,

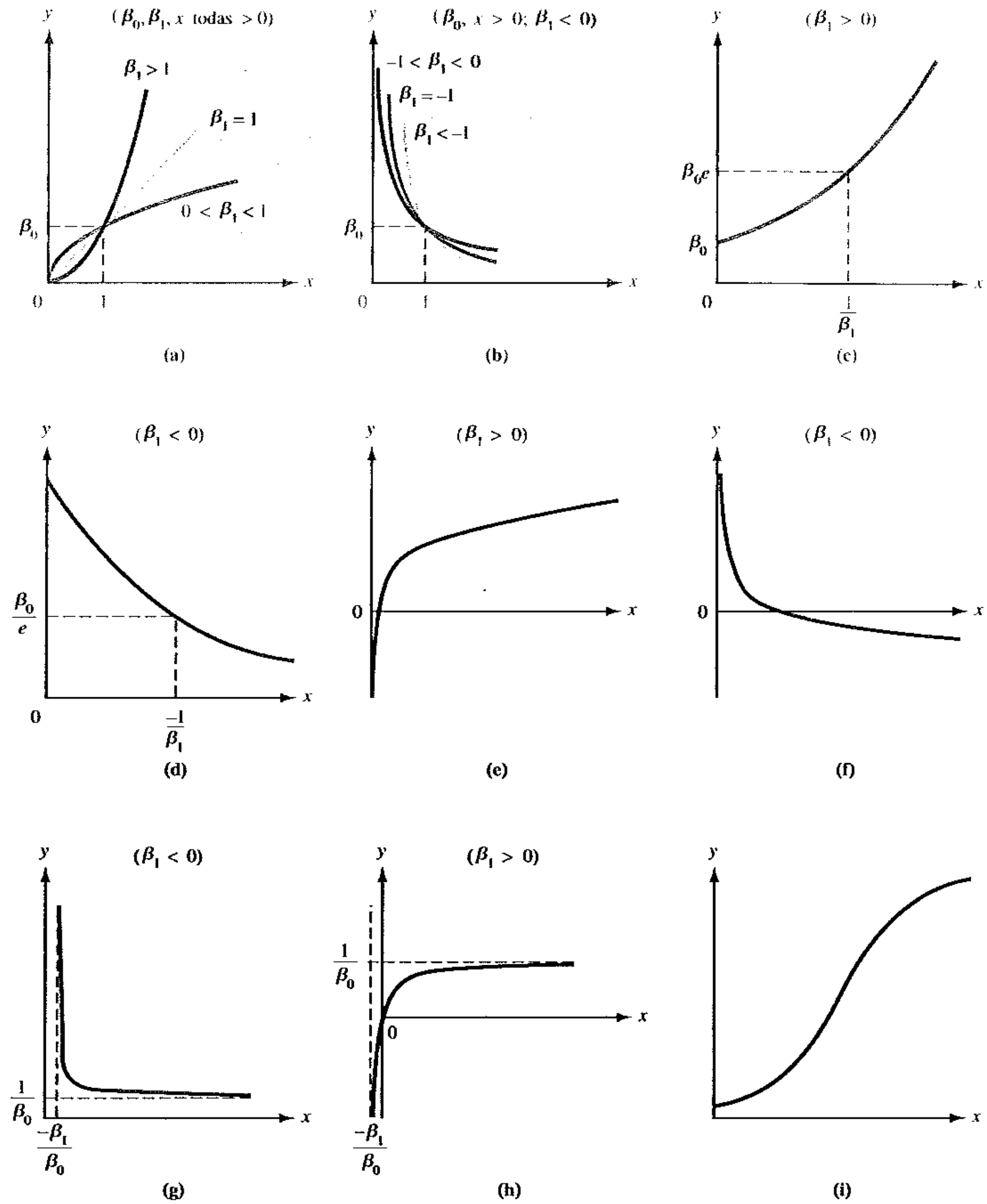


FIGURA 17
Gráficas de funciones linealizables†

†Reimpreso con autorización de C. Daniel y F. Wood, Fitting Functions to Data. Copyright 1980, John Wiley and Sons.

si el diagrama de (x_i, y_i) se ve como algunas de las partes (a) a (i) de la figura 17, entonces hay una relación no lineal entre x y y.

El procedimiento que sigue es uno de tantos que se puede usar para estimar una relación no lineal.

Paso 1 Localice los puntos y determine qué parte de la figura 17 se ajusta mejor a los datos. Supongamos, por ejemplo, que los datos se parecen a la parte (c).

Paso 2 La segunda columna de la tabla 13 proporciona la relación funcional entre x y y. Por lo que toca a la parte (c), podría ser $y = \beta_0 \exp(\beta_1 x)$.

Paso 3 Transforme cada punto dato de acuerdo con las reglas de la tercera columna de la tabla 13. Por lo tanto, si la parte (c) de la figura es relevante, transforme cada valor de y en $\ln y$ y cada valor de x en x. Dada la relación en la segunda columna de la tabla 13, los da-

TABLA 13

Modo de ajustar una relación no lineal

Si la gráfica se ve como una parte de la figura 17	Se tiene la relación funcional	Transforme (x_i, y_i) en	Estimación de la relación funcional
(a) o (b)	$y = \beta_0 x^{\beta_1}$	$(\ln x_i, \ln y_i)$	$\hat{y} = \exp(\hat{\beta} + s_e^2/2)x^{\hat{\beta}_1}$
(c) o (d)	$y = \beta_0 \exp(\beta_1 x)$	$(x_i, \ln y_i)$	$\hat{y} = \exp(\hat{\beta}_0 + \hat{\beta}_1 x + s_e^2/2)$
(e) o (f)	$y = \beta_0 + \beta_1(\ln x)$	$(\ln x_i, y_i)$	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 (\ln x)$
(g) o (h)	$y = \dfrac{x}{\beta_0 x + \beta_1}$	$\left(\dfrac{1}{x_i}, \dfrac{1}{y_i}\right)$	$\hat{y} = \dfrac{x}{\hat{\beta}_0 x + \hat{\beta}_1}$
(i)	$y = \exp\left(\beta_0 + \dfrac{\beta_1}{x}\right)$	$\left(\dfrac{1}{x_i}, \ln y_i\right)$	$\hat{y} = \exp(\hat{\beta}_0 + \dfrac{\hat{\beta}_1}{x} + s_e^2/2)$

tos transformados de la misma tabla deben, al graficarse, indicar una relación de línea recta. Para la parte (c), por ejemplo, si $y = \beta_0 \exp(\beta_1 x)$, entonces, al calcular los logaritmos naturales de ambos miembros se tiene $\ln y = \ln(\beta_0) + \beta_1 x$, de modo que en realidad sí existe una relación lineal entre x y $\ln y$.

Paso 4 Estime la recta de regresión de mínimos cuadrados para los datos transformados. Si $\hat{\beta}_0$ es la ordenada al origen de la recta de mínimos cuadrados (para los datos transformados), $\hat{\beta}_1$ es la pendiente de la recta de mínimos cuadrados (para los datos transformados) y s_e es el error estándar de la estimación de la regresión, entonces lea la relación estimada en la columna final de la tabla 13. Por lo tanto, si la parte (c) fuera relevante, estimaría que $\hat{y} = \exp(\hat{\beta}_0 + \hat{\beta}_1 x + s_e^2/2)$.

Para ejemplificar la idea, suponga que queremos predecir las ventas de videocaseteras en una tienda de enseres domésticos. Las ventas de los últimos 24 meses se proporcionan en la tabla 14 y se grafican en la figura 18 (donde cada punto indica las ventas reales). Usaremos x = número del mes como variable independiente. En la figura 18 se observa una relación en forma de "S" entre x = número del mes y y = ventas durante el mes (como la parte (i) de la figura 17). De modo que, según la tabla 13,

$$y = \exp\left(\beta_0 + \frac{\beta_1}{x}\right)$$

Al recorrer la tercera columna de la tabla 13, estimamos la recta de regresión de mínimos cuadrados para los puntos $(\frac{1}{1}, \ln 23)$, $(\frac{1}{2}, \ln 156)$, . . . , $(\frac{1}{24}, \ln 3495)$. Encontramos $\hat{\beta}_0 = 8.387$, $s_e = .276$ y $\hat{\beta}_1 = -5.788$. La relación estimada entre x y y la obtenemos de acuerdo con la última columna de la tabla 13:

$$\hat{y} = \exp\left(8.387 + .5(.276)^2 - \frac{5.788}{x}\right)$$
$$= \exp\left(8.425 - \frac{5.788}{x}\right)$$

Para ilustrar el uso de esta fórmula en la predicción de ventas, supongamos que queremos pronosticar las ventas de videocaseteras durante el mes 26. Para $x = 26$, podríamos pronosticar que se venderían

$$y = \exp\left(8.425 - \frac{5.788}{26}\right) = 3\,649.3 \text{ videocaseteras}$$

OBSERVACIONES

1 Para la regresión en los puntos $(\frac{1}{1}, \ln 23)$, $(\frac{1}{2}, \ln 156)$, . . . , $R^2 = 0.95$. Esto significa que 95% de la variación en $\ln y$ se explica por la variación en $\frac{1}{x}$. Infortunadamente, esto no nos dice nada respecto a qué tan precisas serán probablemente nuestras predicciones de las ventas reales (y). Para esto, calculamos las ventas pronosticadas para cada mes, y e_i = (ventas reales del mes i) − (ventas pronosticadas

TABLA 14
Ventas de videocaseteras

Mes	Venta de videocaseteras
1	23
2	156
3	330
4	482
5	1209
6	1756
7	2000
8	2512
9	2366
10	2942
11	2872
12	2937
13	3136
14	3241
15	3149
16	3524
17	3542
18	3312
19	3547
20	3376
21	3375
22	3403
23	3697
24	3495

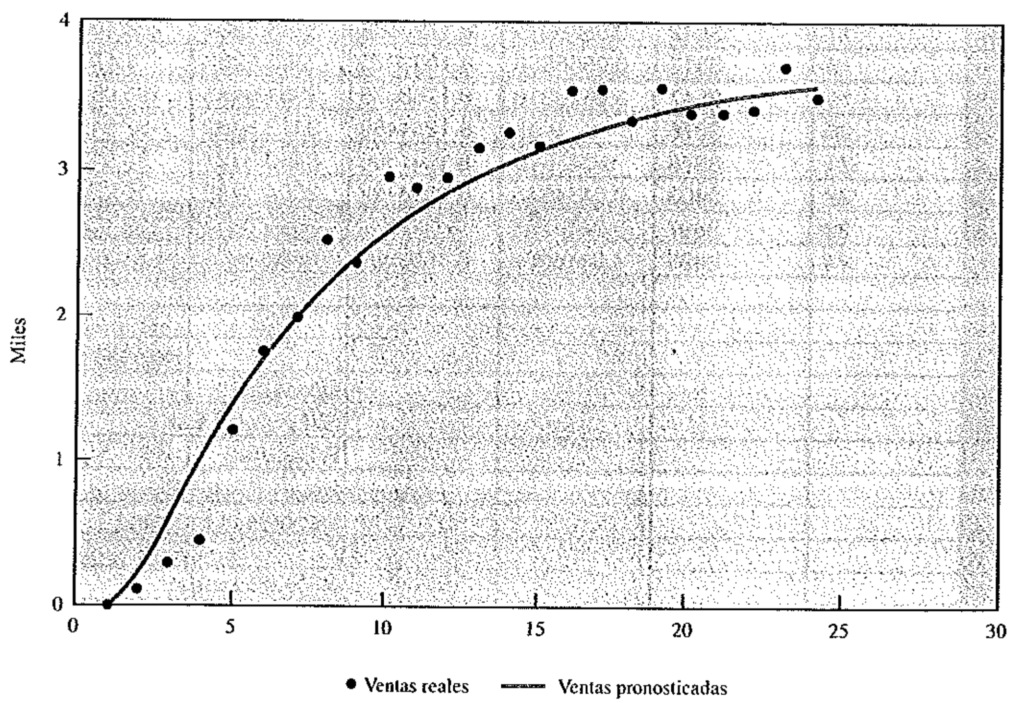


FIGURA 18
Ventas de videocaseteras

del mes i). Al promediar $|e_i|$ para los 24 meses, encontramos que el MAD de las predicciones es 170.3. Al aplicar la ecuación (17) estimaríamos que la desviación estándar de los pronósticos es 1.25(170.3) = 212.88. Por lo tanto, 95% de las veces esperaríamos que las predicciones de las ventas de videocaseteras fueran precisas dentro de 2(212.88) = 425.76 videocaseteras.

2 Si, en forma errónea, hubiéramos tratado de ajustar una recta a estos datos, habríamos obtenido $s_e = 546$, de modo que ajustar la curva en forma de "S" mejora en gran medida los pronósticos.

Uso de una hoja de cálculo para ajustar una relación no lineal

VCR.xls En la figura 19 (archivo VCR.xls) se ilustra el modo de usar una hoja de cálculo para ajustar una curva a los datos en la tabla 14. Primero se introducen los datos en el intervalo de celdas A2:B26. Las variables transformadas 1/MONTH y LNSALES se generan en las co-

FIGURA 19

	A	B	C	D	E	F	G	H
1		MAD=	170.2927		VCR	EXAMPLE		
2	MONTH	SALES	1/MONTH	LNSALES	PREDICT	ERROR	ABSERR	
3	1	23	1	3.1354942	13.969974	9.0300255	9.0300255	
4	2	156	0.5	5.049856	252.37307	-96.37307	96.373067	
5	3	330	0.3333333	5.7990927	662.20451	-332.2045	332.20451	
6	4	482	0.25	6.1779441	1072.6714	-590.6714	590.6714	
7	5	1209	0.2	7.0975489	1432.6874	-223.6874	223.6874	
8	6	1756	0.1666667	7.4707938	1737.5658	18.434166	18.434166	
9	7	2000	0.1428571	7.6009025	1994.303	5.6969922	5.6969922	
10	8	2512	0.125	7.8288345	2211.4573	300.54274	300.54274	
11	9	2366	0.1111111	7.768956	2396.5747	-30.57475	30.574747	
12	10	2942	0.1	7.9868449	2555.765	386.23499	386.23499	
13	11	2872	0.0909091	7.9627639	2693.8455	178.15445	178.15445	
14	12	2937	0.0833333	7.9851439	2814.5945	122.4055	122.4055	
15	13	3136	0.0769231	8.0507034	2920.9846	215.01543	215.01543	
16	14	3241	0.0714286	8.0836372	3015.3711	225.62887	225.62887	
17	15	3149	0.0666667	8.0548402	3099.6364	49.363637	49.363637	
18	16	3524	0.0625	8.167352	3175.2979	348.70211	348.70211	
19	17	3542	0.0588235	8.1724468	3243.5904	298.40965	298.40965	
20	18	3312	0.0555556	8.1053075	3305.5268	6.47315	6.47315	
21	19	3547	0.0526316	8.1738575	3361.9455	185.05446	185.05446	
22	20	3376	0.05	8.1244469	3413.5452	-37.54522	37.545218	
23	21	3375	0.047619	8.1241506	3460.9127	-85.91273	85.912726	
24	22	3403	0.0454545	8.1324127	3504.5442	-101.5442	101.54423	
25	23	3697	0.0434783	8.215277	3544.8619	152.13809	152.13809	
26	24	3495	0.0416667	8.1590887	3582.2271	-87.22711	87.227114	
27								
28								
29						Regression Output:		
30					Constant			8.3867886
31					Std Err of Y Est			0.2761082
32					R Squared			0.9527748
33					No. of Observations			24
34					Degrees of Freedom			22
35								
36					X Coefficient(s)		-5.787996	
37					Std Err of Coef.		0.2747317	
38								
39								
40								
41								

lumnas C y D. La fórmula 1/A3 se escribe en C3. La fórmula =LN(B3) se escribe en D3. Al copiar desde C3:D3 a C3:D26 se generan los valores transformados de xi y y_i. Enseguida se ejecuta una regresión con C3:C26 del rango X y D3:D26 del rango Y. Los resultados de esta regresión se utilizan para predecir las ventas de videocaseteras de cada mes. La fórmula =EXP(C$47+C$48/A3+.5(C$37^2)) se escribe en la celda E3 para generar un pronóstico para las ventas de videocaseteras del mes 1. En la celda F3 se determina e_1 con la fórmula =B3−E3. En la celda G3 se estima $|e_1|$ con la fórmula =ABS(F3). Al copiar del intervalo E3:G3 a E3:G26 se generan las predicciones y los errores para los 24 meses. El MAD para los 24 meses se calcula en la celda C1 con la fórmula =AVERAGE(G3:G26).

Cómo utilizar la curva de tendencia de Excel (Trend Curve)

La curva de tendencia de Excel facilita el ajuste de una ecuación a un conjunto de datos. Después de generar un diagrama de dispersión X-Y, dé clic en los puntos de la gráfica hasta que se vuelvan dorados. Luego seleccione Chart Add Trendline (Agregar línea de tendencia). Véase figura 20.

- Al elegir Lineal se genera la recta que mejor se ajusta a los puntos.
- Escoja Logarítmica si el diagrama de dispersión se ve como (e) o (f) en la figura 17. Entonces Excel proporciona la ecuación del mejor ajuste de la forma $y = \beta_0 + \beta_1(\ln(x))$.
- Escoja Potencial si el diagrama de dispersión se ve como (a) o (b) de la figura 17. Excel da entonces la ecuación de mejor ajuste de la forma $y = \beta_0 x^{\beta_1}$.
- Elija Exponencial si el diagrama se parece a (c) o (d) de la figura 17. Luego Excel proporciona la ecuación del mejor ajuste de la forma $y = \beta_0 e^{\beta_1 x}$.
- Al seleccionar Polinomial del orden n ($n = 1, 2, 3, 4, 5, \text{o } 6$) se llega a la ecuación de mejor ajuste de la forma $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n$.

Antes de tener Trend Curve ajuste la curva, seleccione Options y marque los cuadros de Display Equation on Chart (Presentar ecuación en el gráfico), así como Display R^2 on Chart (Presentar el valor R cuadrado en el gráfico). El valor de R^2 mostrado es la R^2 asociada con la regresión lineal basada en la (x_i, y_i) transformada, listada en la tercera columna de la tabla 13. Por lo que se refiere a las opciones Lineal, Polinomial y Logarítmica, al elegir Intercept = 0 (ordenada al origen = 0) se fija $\beta_0 = 0$.

Cost.xls

En la figura 21 se muestran los resultados obtenidos con Trend Curve cuando se aplicó para encontrar la recta del mejor ajuste para los datos de la hoja de cálculo Cost.xls. Obsérvese que R^2 y las estimaciones de la ecuación corresponden con lo obtenido mediante las Analysis Tool Pak (Herramientas de análisis).

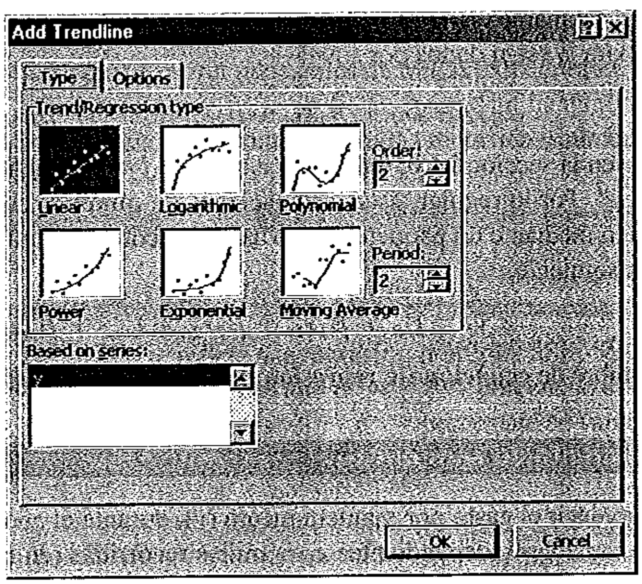


FIGURA 20

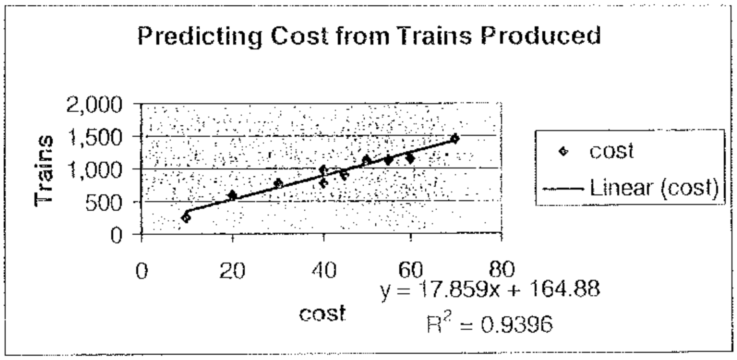


FIGURA 21

24.8 Regresión múltiple

En muchas ocasiones, más de una variable independiente podría ser útil para pronosticar el valor de una variable dependiente. Entonces se aplica la **regresión múltiple**. Por ejemplo, con el objeto de predecir las ventas mensuales de una cadena nacional que comercializa bocadillos de pollo, podríamos pensar en usar las variables independientes siguientes: ingreso nacional, precio del pollo, dólares gastados en publicidad durante el mes actual y dólares gastados en publicidad en los meses anteriores.

Suponga que estamos usando k variables independientes para predecir la variable dependiente y y hay n puntos dato de la forma $(y_i, x_{1i}, x_{2i}, \ldots, x_{ki})$, donde x_{ji} = valor de la j-ésima variable independiente para el i-ésimo punto dato y y_i = valor de la variable dependiente para el i-ésimo punto dato. La relación entre y y las k variables independientes se modela mediante la siguiente expresión en la regresión múltiple

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

donde ε_i es un término de error con media 0, que representa el hecho de que el valor real de y_i podría no ser igual a $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}$. Por otro lado, se podría pensar que β_j es el incremento en y si el valor de la j-ésima variable independiente se incrementa una unidad y todas las otras variables independientes se mantienen constantes. Por lo tanto, β_j es análogo a $\frac{\partial y}{\partial x_j}$, donde x_j es la j-ésima variable independiente.

Estimación de las β_i

Suponga que estimamos β_i ($i = 0, 1, 2, \ldots, k$) usando $\hat{\beta}_i$. Entonces los pronósticos o estimaciones para y_i están dados por

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}$$

Al igual que en la sección 24.6, definimos $e_i = y_i - \hat{y}_i$ y escogemos $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ para minimizar Σe_i^2. Por lo regular, estas estimaciones de mínimos cuadrados de $\beta_0, \beta_1, \ldots, \beta_k$ se obtienen mediante un paquete que se usa en la computadora, como Minitab o Excel. La ecuación siguiente,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_k x_{ki}$$

recibe el nombre de **ecuación de regresión de mínimos cuadrados**.

EJEMPLO 1 Mantenimiento de camiones

Deseamos predecir el gasto por mantenimiento (y), durante el año actual, de un camión, a partir de las variables independientes x_1 = millas recorridas (en miles) durante el presen-

TABLA 15

Datos del mantenimiento del camión

y	x_1	x_2
$832	6	8
$733	7	7
$647	9	6
$553	11	5
$467	13	4
$373	15	3
$283	17	2
$189	18	1
$96	19	0

TABLA 16

Resultados por computadora para el ejemplo 1

Variable	Constante	Error estándar de estimación	Valor t
Constante	17.73846	31.0271	0.57171
x_1	4.061538	1.56742	2.59123
x_2	98.50769	2.756428	35.73744

Error estándar de estimación = 2.106157

te año, y x_2 = antigüedad del camión (en años) a principios del año actual. Contamos con la información de la tabla 15.

Solución El resultado que proporciona la computadora se presenta en la tabla 16. Al recorrer la columna Coeficiente y redondear a dos cifras decimales, obtenemos $\hat{\beta}_0 = 17.74$, $\hat{\beta}_1 = 4.06$ y $\hat{\beta}_2 = 98.51$. Por lo tanto, pronosticaríamos un costo de mantenimiento anual para un camión a partir de

$$\hat{y} = 17.74 + 4.06x_1 + 98.51x_2 \tag{20}$$

En el caso de un vehículo de cinco años de antigüedad que recorrió 10 000 millas durante un año, pronosticaríamos costos de mantenimiento anual de $17.74 + 4.06(10) + 98.51(5) = \$550.89$.

De acuerdo con la ecuación (20), concluimos que (conservando constante la edad del vehículo) recorrer unos miles de millas extra durante un año incrementa los costos de mantenimiento anual en $\hat{\beta}_1 = \$4.06$, y que un incremento de un año en la edad del vehículo (conservando constantes las millas recorridas) aumenta los costos de mantenimiento anual en $\hat{\beta}_2 = \$98.51$.

Bondad de ajuste revisado

Por lo que se refiere a la regresión múltiple, SSR, SSE y SST se definen en la sección 24.6. También encontramos que $R^2 = \frac{\text{SSR}}{\text{SST}}$ = porcentaje de variación en y explicado por las k variables independientes y $1 - R^2$ = porcentaje de variación en y no explicada por las k variables independientes. Si definimos el error estándar de la estimación como

$$s_e = \sqrt{\frac{\text{SSE}}{(n - k - 1)}}$$

entonces (como en la sección 24.6) contamos con que aproximadamente 68% de los valores y estén dentro de s_e de $\hat{y}$ y alrededor de 95% de los valores y estén dentro de $2s_e$ de $\hat{y}$. Ya se vio en la tabla 16 que $s_e = 2.106$. Por lo tanto, 95% del tiempo esperamos que los pronósticos de los gastos al año de mantenimiento del camión sean precisos dentro de 4.21 dólares.

Prueba de hipótesis

Si hemos incluido las variables independientes $x_1, x_2, \ldots, x_k$ en una regresión múltiple, queremos probar a menudo

H_0: $\beta_i = 0$ (x_i no tiene un efecto significativo sobre y cuando las otras variables independientes están incluidos en la ecuación de regresión)

contra

H_a: $\beta_i \neq 0$ (x_i no tiene un efecto significativo sobre y cuando las otras variables independientes están comprendidas en la ecuación de regresión)

Para probar estas hipótesis, estimamos

$$t = \frac{\hat{\beta}_i}{\mathrm{EE}(\hat{\beta}_i)}$$

donde $\mathrm{EE}(\hat{\beta}_i)$ mide la cantidad de incertidumbre presente en la estimación de β_i. $\mathrm{EE}(\hat{\beta}_i)$ (y con frecuencia la variable t) se toma de los resultados de la computadora. Como un nivel de significancia α, rechazamos H_0 si $|t| > t_{(\alpha/2, n-k-1)}$. De acuerdo con la tabla 16, encontramos que (t para x_1) = 2.59 y (t para x_2) = 35.74. Suponga que $\alpha = 0.05$. Puesto que $t_{(.025, 9-2-1)} = 2.447$, rechazamos H_0 para cada variable independiente, y concluimos que tanto como las millas recorridas como la edad del vehículo tienen un importante efecto en el costo de mantenimiento anual.

Por lo regular, las variables incluidas en una ecuación de regresión deben tener estadísticas t significativas. ($\alpha = 0.10$ o $\alpha = 0.05$ son niveles que se usan habitualmente en el análisis de regresión). Si una variable independiente tiene una t no significativa, eliminamos por lo general la variable independiente de la ecuación, y obtenemos nuevas estimaciones de mínimos cuadrados. Con el fin de ilustrar este concepto, suponga que contamos con los datos de la tabla 17 sobre ventas del restaurante Bloomington Happy Chicken durante los últimos 20 años (archivo Chicken.xls). (POP = población dentro de 10 millas del restaurante Happy Chicken, AD = miles de dólares gastados en publicidad durante el año actual, LAGAD = miles de dólares gastados en publicidad durante el año anterior y SALES = ventas en miles de dólares).

Chicken.xls

Pretendemos estimar el modelo

$$\text{SALES} = \beta_0 + \beta_1\text{YEAR} + \beta_2\text{POP} + \beta_3\text{AD} + \beta_4\text{LAGAD} + \varepsilon$$

Usamos YEAR como una variable independiente con la esperanza de mejorar una posible tendencia ascendente en las ventas. LAGAD se utiliza como variable independiente porque opinamos que la publicidad del año anterior podría afectar las ventas del presente año. Obtenemos la siguiente ecuación de regresión estimada (la estadística t para cada variable independiente está entre paréntesis):

$$\widehat{\text{SALES}} = 10\,951.51 + \underset{(1.91)}{169.51\ \text{YEAR}} - \underset{(-.70)}{.059\ \text{POP}} + \underset{(13.84)}{122.38\ \text{AD}} + \underset{(28.92)}{276.93\ \text{LAGAD}} \qquad (21)$$

No podemos usar datos del primer año, porque LAGAD está indefinida. Como $t_{(.05, 19-4-1)} = 1.761$, entonces todas las variables independientes excepto POP son significativas para $\alpha = 0.10$. Por lo tanto, YEAR, AD y LAGAD tienen, al parecer, un efecto

TABLA 17

Datos de las ventas de Happy Chicken

Year	POP	AD	LAGAD	Sales
1	96 020	30	—	13 000
2	102 558	20	30	15 713
3	101 792	15	20	12 937
4	104 347	25	15	12 872
5	106 180	30	25	16 227
6	106 562	15	30	15 388
7	105 209	25	15	13 180
8	109 185	35	25	17 199
9	109 976	40	35	20 674
10	110 659	20	40	20 350
11	111 844	25	20	14 444
12	111 576	35	25	17 530
13	113 784	5	35	16 711
14	112 482	12	5	19 715
15	116 487	16	12	12 248
16	117 316	21	16	13 856
17	117 830	22	21	15 285
18	118 148	24	22	15 620
19	118 481	26	24	17 158
20	121 069	28	26	17 800

importante en las ventas. Después de eliminar la variable no significativa POP de la ecuación, obtenemos la ecuación de regresión estimada siguiente:

$$\widehat{\text{SALES}} = 5150.94 + \underset{(8.29)}{108.58}\,\text{YEAR} + \underset{(14.10)}{121.59}\,\text{AD} + \underset{(31.72)}{274.30}\,\text{LAGAD}$$

Todas las variables independientes son significativas. Asimismo, encontramos que $R^2 = 0.99$ y $s_e = 309$. Por lo tanto, estamos razonablemente satisfechos con esta ecuación y esperamos que 95% del tiempo, nuestros pronósticos de las ventas estén dentro de 618 000 dólares de las ventas reales.

Elección de la mejor ecuación de regresión

¿De qué modo podemos escoger entre varias ecuaciones de regresión que tienen diferentes conjuntos de variables independientes? Queremos escoger, por lo regular, la ecuación con el valor mínimo de s_e, ya que proporcionará los pronósticos más exactos. Asimismo, queremos que la t sea significativa para todas las variables de la ecuación. Estos dos objetivos podrían entrar en conflicto, en cuyo caso es difícil determinar cuál es la "mejor" ecuación. Si se cuenta con los resultados obtenidos mediante computadora, y éstos contienen la estadística C_p, entonces la regresión elegida debe tener un valor C_p cercano a (número de variables independientes en la ecuación) + 1. Por ejemplo, si una regresión con tres variables independientes tiene $C_p = 80$, entonces podemos estar seguros de que no es una "buena" regresión. En realidad, si C_p de una regresión es mucho más grande que p, esto significa que ha sido omitida por lo menos una variable importante en la regresión. (Véase un análisis acerca de C_p en Daniel y Wood (1980)).

Multicolinealidad

Si una ecuación de regresión estimada contiene dos o más variables independientes que manifiestan una fuerte relación lineal, se dice que hay **multicolinealidad**. Una fuerte relación lineal entre algunas de las variables independientes podría hacer que las estimaciones por computadora de la β_i sean poco confiables. En ciertas circunstancias, la multicolinealidad hasta puede ocasionar que una variable que debe tener β_i positiva tenga una $\hat{\beta}_i$ sustancialmente menor que 0. El ejemplo del Happy Chicken ilustra la multicolinealidad. Se empieza con YEAR y POP como variables independientes, y como POP y YEAR se incrementan en el tiempo, contaríamos con que existe una fuerte relación lineal positiva entre ellas. De hecho, la correlación entre YEAR y POP es 0.98. Para ver que las estimaciones de β_{YEAR} y β_{POP} son de poco fiar, obsérvese que en (21), $\hat{\beta}_{\text{POP}} < 0$, lo cual indica que un aumento en la cantidad de clientes cercanos a Happy Chicken disminuye las ventas. Esta anomalía se debe a la multicolinealidad. La fuerte relación entre YEAR y POP dificulta la estimación exacta de β_{POP} y de β_{YEAR} mediante computadora. Después de eliminar POP de la ecuación estimada, desaparece la multicolinealidad porque ya no hay una relación fuerte entre cualquiera de las variables independientes.

A propósito, si existe una relación lineal *exacta* entre dos o más variables independientes, entonces hay una cantidad infinita de combinaciones de β_i que minimizará la suma de los errores cuadrados, y la mayor parte de paquetes de computadora imprimen un mensaje de error. Por ejemplo, si hacemos que x_1 = gasto de los consumidores de E.U. durante un año, x_2 = inversión en E. U. durante un año, x_3 = gasto gubernamental de E.U. durante un año y x_4 = ingreso nacional de E.U. durante un año, es bien sabido que $x_4 = x_1 + x_2 + x_3$. En este caso no podemos utilizar x_1, x_2, x_3 y x_4 como variables independientes; por lo menos se debe eliminar una de la ecuación.

Variables ficticias

Una variable independiente no cuantitativa (cualitativa) llega a influir, con frecuencia, en la variable dependiente, por ejemplo,

Variable dependiente	**Variable independiente categórica**
Salario del empleado	Raza del empleado
Gastos de los consumidores en el año	Si es año de guerra o de paz
Clientes que van al banco un día dado	Día de la semana
Ventas de acondicionadores de aire en un mes dado	Mes del año

La variable independiente no asume, en cada una de estas situaciones, un valor numérico, pero se podría clasificar en alguna de c categorías. Es decir, por lo que toca a las ventas mensuales de acondicionadores de aire, $c = 12$, y en el ejemplo de los gastos de los consumidores, $c = 2$.

Sean los valores posibles de la variable categórica valor 1, valor 2, . . . , valor c. Para modelar el efecto de una variable categórica sobre una variable dependiente, definamos $c + 1$ **variables ficticias** como se indica enseguida:

$x_1 = 1$ Si la observación toma el valor 1 de la variable categórica

$x_1 = 0$ Si no sucede así

$x_2 = 1$ Si la observación toma un valor 2 de la variable categórica

$x_2 = 0$ Si no sucede así

$\vdots$

$x_{c-1} = 1$ Si la observación toma el valor $c - 1$ de la variable categórica

$x_{c-1} = 0$ Si no sucede así

TABLA 18
Clientes en la unión de crédito de la universidad

Número del día	Día de la semana	¿Día de pago de la universidad?	Clientes que llegan a la unión de crédito
1	Lunes	No	515
12	Martes	No	360
18	Miércoles	Sí	548
23	Miércoles	No	386
24	Jueves	No	440
46	Lunes	Sí	687
48	Miércoles	No	350
52	Martes	No	430
54	Jueves	No	370
55	Viernes	No	496
70	Viernes	No	506
81	Lunes	No	509
89	Jueves	Sí	508
104	Jueves	No	396
106	Lunes	No	600
108	Miércoles	No	266
122	Martes	No	360
130	Viernes	Sí	521
152	Martes	No	398

Ahora incluimos $x_1, x_2, \ldots, x_{c-1}$ (junto con otras variables independientes pertinentes) en la ecuación estimada de regresión.

Para ilustrar el uso de las variables ficticias, suponga que pretendemos pronosticar el número de clientes que entran todos los días a la unión de crédito de la universidad. El gerente del banco opina que el tráfico en el banco depende del día de la semana (la unión de crédito está abierta de lunes a viernes) y de si es día de pago para los empleados de una universidad. Contamos con el número de personas que entran al banco durante 18 días elegidos al azar. (El día 1 es el día actual, el día 6 es dentro de una semana, y así sucesivamente). La información pertinente es encuentra en la tabla 18.

En esta situación, hay dos variables categóricas que interesan: el día de la semana ($c = 5$) y si es día de pago o no ($c = 2$). En cuanto al día de la semana definimos como valor 1 = lunes, valor 2 = martes, valor 3 = miércoles, valor 4 = jueves y valor 5 = viernes. Entonces, hacemos

$x_1 = 1$	Si el día es un lunes	$x_1 = 0$	Si no es así
$x_2 = 1$	Si el día es un martes	$x_2 = 0$	Si no es así
$x_3 = 1$	Si el día es un miércoles	$x_3 = 0$	Si no es así
$x_4 = 1$	Si el día es un jueves	$x_4 = 0$	Si no es así

Por lo que se refiere a si es día de pago o no, definimos $c - 1 = 1$ variables ficticias. Si hacemos que valor 1 = día de pago y valor 2 = no es día de pago, definimos

$$x_5 = 1 \quad \text{Si el día es día de pago} \quad x_5 = 0 \quad \text{Si no es así}$$

Para explicar una posible tendencia en la cantidad de clientes incluimos T = número del día como una variable independiente. Para ilustrar cómo codificar esta información en la computadora, "codificamos" las dos últimas observaciones:

T	x_1	x_2	x_3	x_4	x_5	Clientes
130	0	0	0	0	1	521
152	0	1	0	0	0	398

Obtenemos la siguiente ecuación de regresión estimada (valor de t entre paréntesis)

$$\hat{y} = 496.1 - 0.36T + 71.1x_1 - 78.5x_2 - 122.5x_3 - 74.8x_4 + 127.1x_5$$
$$(-1.29) \quad (1.87) \quad (-2.04) \quad (-3.17) \quad (-1.99) \quad (4.43)$$

Para $\alpha = 0.10$ encontramos que todas las variables independientes, excepto T, son significativas. Después de eliminar T de esta ecuación estimada, llegamos a la ecuación siguiente:

$$\hat{y} = 466.2 + 80.4x_1 - 79.2x_2 - 109.8x_3 - 68.8x_4 + 124.3x_5$$
$$(2.1) \quad (-2.0) \quad -(2.86) \quad (-1.79) \quad (4.24) \tag{22}$$

Para $\alpha = 0.10$, todas las variables independientes son significativas, de modo que esta ecuación es, al parecer satisfactoria. Encontramos, asimismo, que $R^2 = 0.85$ y $s_e = 48.8$. Por lo tanto, 95% de todos los días, el pronóstico para el número de clientes que llega a la unión de crédito debe ser exacto dentro de $2(48.8) = 97.6$ clientes.

Interpretación de los coeficientes de las variables ficticias

¿Cuál es la interpretación de los coeficientes de las variables ficticias? Con el objeto de ejemplificar esta cuestión, determinemos cómo afecta al tráfico de la unión de crédito si el día es de pago o no. En un día de pago, $x_5 = 1$, y predecimos que $466.2 + 80.4x_1 - 79.2x_2 - 109.8x_3 - 68.8x_4 + 124.3$ clientes llegarán a la unión de crédito. En un día que no es día de pago, $x_5 = 0$, y el pronóstico es que $466.2 + 80.4x_1 - 79.2x_2 - 109.8x_3 - 68.8x_4$ clientes llegarán al banco. Si restamos, tenemos que en un día de pago el pronóstico es (con todas las otras cuestiones iguales) que $\hat{\beta}_5 = 124.3$ clientes más acudirán al banco que en un día que no es día de pago.

Para ver cómo el día de la semana influye en el tráfico de la unión de crédito, observe que la ecuación (22) genera un pronóstico diferente para cada día de la semana. Para el lunes, $x_1 = 1$, $x_2 = x_3 = x_4 = 0$, y obtenemos $\hat{y} = 466.2 + 80.4 + 124.3x_5 = 546.6 + 124.3x_5$. Para el martes, $x_2 = 1$, $x_1 = x_3 = x_4 = 0$, y obtenemos $\hat{y} = 466.2 - 79.2 + 124.3x_5 = 387 + 124.3x_5$. Para el miércoles, $x_3 = 1, x_2 = x_1 = x_4 = 0$, y obtenemos $\hat{y} = 466.2 - 109.8 + 124.3x_5 = 356.4 + 124.3x_5$. Para el jueves, $x_4 = 1, x_1 = x_3 = x_2 = 0$, y obtenemos $\hat{y} = 466.2 - 68.8 + 124.3x_5 = 397.4 + 124.3x_5$. Para el viernes, $x_1 = x_2 = x_3 = x_4 = 0$, y obtenemos $\hat{y} = 466.2 + 124.3x_5$. Por lo tanto, observamos que (con todos los otros aspectos iguales) el tráfico de la unión de crédito es más pesado en los lunes, en el segundo lugar está el viernes, en tercer lugar está el jueves, en cuarto lugar está el martes y el quinto lugar en tráfico lo tiene el miércoles.

Modelos multiplicativos

A menudo creemos que existe una relación de la siguiente forma:

$$Y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \cdots x_k^{\beta_k} \tag{23}$$

Con el fin de estimar tal relación, calculamos los logaritmos de ambos miembros de la ecuación (23). Entonces,

$$\ln Y = \ln \beta_0 + \beta_1(\ln x_1) + \beta_2(\ln x_2) + \cdots + \beta_k(\ln x_k)$$

Por lo tanto, para estimar (23), ejecutamos una regresión múltiple en la cual la variable dependiente es $\ln Y$ y las variables independientes son $\ln x_1, \ln x_2, \ldots, \ln x_k$. Para ilustrar la idea, suponga que deseamos determinar cómo los costos de operación anual de una com-

TABLA 19
Datos de las sucursales de la compañía de seguros

Sucursal	Costo de operación anual	Número de pólizas de seguros para casas	Número de pólizas de seguros para automóviles
1	$124 000	400	1 200
2	$71 000	350	360
3	$136 000	600	800
4	$219 000	800	1 800
5	$230 000	900	1 600
6	$75 000	200	1 000
7	$56 000	120	900
8	$110 000	340	1 100
9	$120 000	490	900
10	$144 000	700	800

pañía de seguros dependen del número de pólizas de seguros a casas y automóviles que se han logrado. En la tabla 19 se proporciona la información pertinente para diez sucursales de la compañía de seguros (archivo Branch.xls).

Branch.xls

Para establecer el modelo $Y = \beta_0 x_1^{\beta_1} x_2^{\beta_2}$, donde Y = costo de operación anual, x_1 = pólizas de seguros para casa y x_2 = pólizas para asegurar automóviles, introduciríamos a la computadora los puntos de la forma (ln 124 000, ln 400, ln 1200) y así sucesivamente. Las estimaciones de mínimos cuadrados obtenidas son

$$\begin{aligned} \text{Estimación del término constante } &= 5.339 \\ \text{Estimación para } \beta_1 &= 0.583 \\ \text{Estimación para } \beta_2 &= 0.409 \end{aligned}$$

Esta regresión genera una R^2 de 0.998, lo cual indica un muy buen ajuste. La estimación del término constante es una estimación de ln β_0, de modo que la estimación real de β_0 es $e^{5.339} = 208.3$, y estimamos que $Y = 208.3x_1^{0.583}x_2^{0.409}$. Para ilustrar el uso de esta ecuación, pronosticamos el costo de operación anual para una sucursal de la aseguradora redactando 500 pólizas para casa y 1200 pólizas para automóviles. Para esta sucursal obtenemos $Y = 208.3(500)^{0.583}(1200)^{0.409}$ = $141 767.

Heterocedasticidad y autocorrelación en la regresión múltiple

Cuando de grafican los errores en sucesiones de series de tiempo se podría verificar (según se explica en la sección 24.6) si los errores en una regresión múltiple son independientes. Si está presente la autocorrelación y los errores al parecer no son independientes, entonces la corrección por autocorrelación generará mejores pronósticos.

Al graficar los errores (en el eje de las y) contra el valor pronosticado de y (en el eje de las x) es posible determinar si está presente la homocedasticidad o la heterocedasticidad. Si la que existe es la homocedasticidad, entonces, la gráfica no debe mostrar un patrón obvio (es decir, la gráfica debe ser similar a la de la figura 10), en tanto que si la que está presente es la heterocedasticidad, se debe observar, en la gráfica, un patrón obvio que indica que los errores dependen, por alguna razón, del valor pronosticado de y (quizá como en la figura 11). Si existe heterocedasticidad, las pruebas t explicadas en esta sección no son válidas.

Organización de la regresión múltiple en una hoja de cálculo

Credit.xls

En las figuras 22 y 23 (archivo Credit.xls) se muestra la regresión ejecutada con los datos de la tabla 18. Se introduce la cuenta de clientes de cada día en el intervalo A3:A21, y el número del día en las celdas B3:B21. En las celdas H3:H21 se escribe el día de la semana para cada observación (1 = Monday (lunes), . . . , 5 = Friday (viernes). Una variable ficticia en las celdas G3:G21 indica si es día de pago o no. Luego se utiliza el enunciado =IF para generar las variables ficticias para el día de la semana. La fórmula =IF(H3=1,1,0) se escribe en la celda C3. De esta manera se coloca un 1 en C3, lo cual señala que la primera observación es en Monday (lunes). En la celda D3 se escribe la fórmula =IF(H3=2,1,0); en la celda E3, =IF(H3=3,1,0) y en la celda F3, =IF(H3=4,1,0). Al copiar del intervalo C3:F3 hasta C3:F21 se generan los valores de las variables ficticias de todas las observaciones. Para ejecutar la regresión, seleccione el rango Y de A2:A21 y el rango X de B2:G21. La regresión se podría interpretar de la manera siguiente:

Ordenada al origen Es $\hat{\beta}_0 = 496.0857$.

Error estándar Es $s_e = 48.84517$.

R cuadrado Es $R^2 = .845927$. Esto quiere decir que juntas, todas las variables independientes en la regresión, explican 84.6% de la variación en la cantidad de clientes que llega cada día.

Observaciones Es la cantidad de puntos datos (19).

GL totales Son los grados de libertad ($n - k - 1 = 19 - 6 - 1$) usados para la prueba t de H_0: $\beta_i = 0$ contra H_1: $\beta \neq 0$.

Coeficientes Esta columna proporciona el coeficiente de las variables independientes que se encuentren en la ecuación de mínimos cuadrados. Por ejemplo, $\hat{\beta}_T = -0.36222$.

Error estándar Este renglón da EE $\hat{\beta}_i$. para cada variable independiente. Por ejemplo, EE $\hat{\beta}_T = 0.279852$. El coeficiente X dividido entre el error estándar del coeficiente da el valor de t para probar H_0: $\beta_i = 0$ contra H_1: $\beta_i \neq 0$.

Valor de t Esto da el valor de la estadística t observada (coeficiente/error estándar) para la ordenada al origen y todas las variables independientes. Por ejemplo, t para lunes es 1.87.

Columna del error estándar La ordenada al origen da el error estándar $\hat{\beta}_0 = 37.66$, y los datos de los coeficientes proporcionan el error estándar para cada una de las variables independientes. Por ejemplo, error estándar $\hat{\beta}_{\text{Monday}} = 38.07$. La entrada del coeficiente

	A	B	C	D	E	F	G	H
1		CREDIT	UNION	EXAMPLE				
2	CUSTOMEI	DAY#	MON	TUES	WED	THUR	PAYDAY?	DAYWK
3	515	1	1	0	0	0	0	1
4	360	12	0	1	0	0	0	2
5	548	18	0	0	1	0	1	3
6	386	23	0	0	1	0	0	3
7	440	24	0	0	0	1	0	4
8	687	46	1	0	0	0	1	1
9	350	48	0	0	1	0	0	3
10	430	52	0	1	0	0	0	2
11	370	54	0	0	0	1	0	4
12	496	55	0	0	0	0	0	5
13	506	70	0	0	0	0	0	5
14	509	81	1	0	0	0	0	1
15	508	89	0	0	0	1	1	4
16	396	104	0	0	0	1	0	4
17	600	106	1	0	0	0	0	1
18	266	108	0	0	1	0	0	3
19	360	122	0	1	0	0	0	2
20	521	130	0	0	0	0	1	5
21	398	152	0	1	0	0	0	2

FIGURA 22
Ejemplo de la unión de crédito

	B	C	D	E	F	G
22	SUMMARY OUTPUT					
23						
24	*Regression Statistics*					
25	Multiple R	0.9197432				
26	R Square	0.8459275				
27	Adjusted R Square	0.7478813				
28	Standard Error	51.017121				
29	Observations	19				
30						
31	SUMMARY OUTPUT					
32						
33	*Regression Statistics*					
34	Multiple R	0.9197432				
35	R Square	0.8459275				
36	Adjusted R Square	0.7688912				
37	Standard Error	48.845174				
38	Observations	19				
39						
40	ANOVA					
41		*df*	*SS*	*MS*	*F*	*ignificance*
42	Regression	6	157192.73	26198.789	10.980899	0.0002823
43	Residual	12	28630.212	2385.851		
44	Total	18	185822.95			
45						
46		*Coefficients*	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*
47	Intercept	496.08565	37.660289	13.172646	1.7E-08	414.03093
48	DAY#	-0.362218	0.2798522	-1.294318	0.2199096	-0.971963
49	MON	71.076945	38.076099	1.8667076	0.0865578	-11.88375
50	TUES	-78.47826	38.509376	-2.0379	0.0642282	-162.383
51	WED-	122.5236	38.6517	-3.16994	0.0080705	-206.7384
52	THUR	-74.82254	37.669971	-1.986265	0.070328	-156.8984
53	PAYDAY?	127.10854	28.682168	4.4316224	0.0008188	64.615462
54						

FIGURA 23

dividida entre la entrada del error estándar genera la estadística t para la ordenada al origen o pendiente (tabulada en la siguiente columna).

Valor *P* Por lo que toca a la ordenada al origen y cada variable independiente de una regresión con k variables independientes, éste da la probabilidad($|t_{n-k-1}| \geq |$, estadística t observada$|$). Por ejemplo, si el valor p para Wednesday (miércoles) es menor que α, rechazamos H_0: $\beta_{\text{Wednesday}} = 0$; si no es así, aceptamos $\beta_{\text{Wednesday}} = 0$. Para $\alpha = .05$, rechazamos $\beta_{\text{Wednesday}} = 0$.

Las Herramientas de Análisis de datos para la regresión son capaces de manejar un máximo de 15 variables independientes. Los datos de las variables independientes deben estar en columnas adyacentes.

PROBLEMAS

Grupo A

1 Para los años 1961 a 1970, el rendimiento anual de las acciones de General Motors y el rendimiento sobre el índice bursátil de Standard and Poor fueron los que se presentan en la tabla 20 (archivo Beta.xls).

a Sea Y = rendimiento de las acciones de General Motors durante un año y X = rendimiento de índice de Standard and Poor durante un año. La teoría financiera sugiere que $Y = \beta_0 + \beta_1 X + \varepsilon$, donde β_1 se llama **beta** para General Motors. Proporcione una interpretación de la beta de unas acciones (en este caso, General Motors), y utilice los datos de la tabla 20 para estimar la beta de General Motors.

b ¿El índice de Standard and Poor parece tener un efecto significativo (para $\alpha = 0.05$) sobre el rendimiento de las acciones de General Motors?

c ¿Qué porcentaje de la variación en el rendimiento en las acciones de General Motors se explica por la variación en el índice de Standard and Poor?

d ¿Qué porcentaje de la variación en el rendimiento en las acciones de General Motors no se explica por la variación en el índice de Standard and Poor?

e Durante un año en el cual el índice de Standard and Poor se incrementó en 15%, ¿qué se podría predecir en

TABLA 20

Año	Rendimiento de las acciones de General Motors	Índice de rendimiento para Standard and Poor
1961	12%	21%
1962	2%	−3%
1963	38%	15%
1964	26%	20%
1965	18%	12%
1966	−10%	0%
1967	0%	10%
1968	9%	10%
1969	−2%	2%
1970	−1%	−15%

cuanto al rendimiento de las acciones de General Motors?

2 Pretendemos determinar las horas de mano de obra requeridas para fabricar una unidad de un producto. Tenemos la información de la tabla 21 (archivo Learn.xls). Por ejemplo, la segunda unidad producida requirió 517 horas de mano de obra, y la unidad número 600 que se produjo necesitó 34 horas de mano de obra.

a Trate de determinar una relación entre la cantidad de unidades ya producidas y las horas de mano de obra necesarias para fabricar la unidad siguiente. ¿Por qué esta relación se llama **curva de aprendizaje**?

b ¿Cuántas horas de mano de obra se requerirían para fabricar la unidad número 800?

c Estamos 95% seguros de que la predicción del inciso (b) es exacta dentro de ________ horas.

3 Las ventas trimestrales de una tienda de departamentos en un periodo de seis años se proporcionan en la tabla 22 (archivo Sales.xls).

a Utilice la regresión múltiple para desarrollar un modelo que se pueda aplicar para predecir las ventas trimestrales. (*Sugerencia*: utilice variables ficticias y una variable independiente para el número de los trimestres (trimestre 1, trimestre 2, . . . , trimestre 24).

b Si se hace Y_t = ventas durante el trimestre número t, analice cómo ajustar el modelo siguiente a los datos en la tabla 22:

$$Y_t = \beta_0\beta_1^t\beta_2^{x_2}\beta_3^{x_3}\beta_4^{x_4}$$

donde $x_2 = 1$ si t es el primer trimestre, $x_3 = 1$ si t es el trimestre segundo y $x_4 = 1$ si t es el cuarto trimestre. (*Sugerencia:* calcule los logaritmos de ambos miembros).

c Interprete la respuesta del inciso (b).

d ¿Qué modelo parece proporcionar mejores predicciones para las ventas?

TABLA 21

Producción acumulativa	Horas de mano de obra necesarias para la última unidad
1	715
2	517
10	239
20	174
40	126
60	104
100	82
150	68
200	59
300	47
500	37
600	34

TABLA 22

Año	Trimestre	Ventas (millones de dólares)
1984	1	$50 147
1984	2	$49 325
1984	3	$57 048
1984	4	$76 781
1985	1	$48 617
1985	2	$50 898
1985	3	$58 517
1985	4	$77 691
1986	1	$50 862
1986	2	$53 028
1986	3	$58 849
1986	4	$79 660
1987	1	$51 640
1987	2	$54 119
1987	3	$65 681
1987	4	$85 175
1988	1	$56 405
1988	2	$60 031
1988	3	$71 486
1988	4	$92 183
1989	1	$60 800
1989	2	$64 900
1989	3	$76 997
1989	4	$103 337

4 Con el objeto de determinar cuánto influye el precio sobre las ventas, una compañía cambió el precio de un producto durante un periodo de 20 semanas. El precio de cada semana y la cantidad de unidades vendidas se proporcionan en la tabla 23 (archivo Price.xls). Desarrolle un modelo para relacionar las ventas con el precio.

5 Confederate Express Service desea determinar qué tanto sus costos de embarque mensuales dependen del número de unidades embarcadas durante un mes. El número de unidades embarcadas y el costo total de embarque de los últimos quince meses se proporcionan en la tabla 24 (archivo Ship.xls).

TABLA 23

Precio	Unidades vendidas
$1	1145
$2	788
$3	617
$4	394
$5	275
$6	319
$7	289
$8	241
$9	259
$10	176
$11	179
$12	232
$13	183
$14	181
$15	222
$16	212
$17	186
$18	110
$19	183
$20	172

TABLA 24

Mes	Unidades embarcadas	Costo total de embarque
1	300	$1060
2	400	$1380
3	500	$1640
4	200	$740
5	300	$1060
6	350	$1190
7	460	$1520
8	480	$1580
9	120	$540
10	760	$2420
11	580	$2200
12	340	$1470
13	120	$790
14	100	$720
15	500	$1960

a Determine una relación entre las unidades embarcadas y los costos de embarque mensuales.

b Grafique los errores de los pronósticos en orden de la serie de tiempo. ¿Hay un patrón poco común?

c Se nos dice que hubo una huelga de transportistas durante los meses 11 a 15, y nuestra opinión es que este asunto influyó en los costos de embarque. ¿Cómo se puede modificar la respuesta al inciso (a) para que explique los efectos de la huelga?

Después de explicar los efectos de la huelga, ¿desaparece el patrón poco común del inciso (b)?

Grupo B

6 En el ejemplo 1 ejecutamos una regresión con sólo x_1 (millas recorridas) como una variable independiente. Encontramos que el coeficiente de x_1 es -51.68 en esta regresión. Al parecer, esto indica (al contrario de lo que podríamos esperar) que aumentar las millas recorridas ocasionará menores costos de mantenimiento. Explique este resultado. (*Sugerencia*: estime la correlación entre x_1 y x_2.)

7 Suponga que queremos ajustar una curva a unos datos, y la parte (i) de la figura 17 es relevante. Explique la razón de que los puntos de la forma $(\frac{1}{x_i}, \ln y_i)$, al ser graficados, deben indicar una relación de línea recta.

8 Considere la regresión en la cual estimamos el costo de administrar una compañía de seguros en función del número de pólizas de seguro para casas y automóviles. Si hubiera un aumento de 1% en la cantidad de pólizas de seguro para automóviles, ¿cuál sería el porcentaje que daríamos como pronóstico en que se incrementarían los costos totales?

9 En el ejemplo en el cual pronosticamos la cantidad de clientes que llegan a la unión de crédito, suponga que utilizamos cinco (en lugar de cuatro) variables ficticias para representar los días de la semana. ¿Qué problemas habrían surgido?

RESUMEN Pronóstico con promedio móvil

$$f_{t,1} = \text{promedio de las últimas } N \text{ observaciones}$$
$$e_t = x_t - (\text{pronóstico para } x_t)$$
$$\text{MAD} = \text{valor promedio de } |e_t|$$

Elija N que minimice MAD.

Suavizamiento exponencial simple

A_t = promedio suavizado al final del periodo t

$= f_{t,k}$ = pronóstico para el periodo $t + k$ efectuado al final del periodo t

$$A_t = \alpha x_t + (1 - \alpha)A_{t-1}$$

Elija α que minimice a MAD.

Método de Holt

El método de Holt se utiliza cuando hay una tendencia presente, pero no estacionalidad.

L_t = estimación de la base al final del periodo t

T_t = estimación de la tendencia por periodo al final del periodo t

$$L_t = \alpha x_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$f_{t,k} = L_t + kT_t$$

Método de Winter

Este método se aplica cuando opinamos que tanto tendencia como estacionalidad podrían estar presentes.

s_t = estimación para el factor estacional del mes t

$$L_t = \frac{\alpha x_t}{s_{t-c}} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$s_t = \frac{\gamma x_t}{L_t} + (1 - \gamma)s_{t-c}$$

$$f_{t,k} = (L_t + kT_t)s_{t+k-c}$$

Por lo que se refiere a todos los métodos de extrapolación, contamos con que 68% de nuestros pronósticos estará dentro de $s_e = 1.25$ MAD del valor real, y 95% de los pronósticos estarán dentro de $2s_e$ del valor real.

Regresión lineal simple

Dados los puntos dato $(x_1, y_1), \ldots, (x_n, y_n)$, estimamos una relación lineal entre x y y mediante $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, donde

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad \text{y} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$R^2 = \frac{\text{SSR}}{\text{SST}}$$ = porcentaje de variación en y explicado por x

r_{xy} = correlación lineal de la muestra entre x y y
(r_{xy} indica la fuerza de la relación lineal entre x y y)

$$s_e = \sqrt{\frac{\text{SSE}}{n - 2}}$$

Esperamos que 68% de los pronósticos estén dentro de s_e del valor real y 95% de las predicciones, dentro de $2s_e$ del valor real.

Una estadística t que sea mayor que $t_{(\alpha/2,n-2)}$ en valor absoluto es prueba (en el nivel de significancia α) que hay una relación lineal significativa entre x y y.

Ajuste de una relación no lineal

Paso 1 Grafique los puntos y determine qué parte de la figura 17 se ajusta mejor a los datos.

Paso 2 La segunda columna de la tabla 13 representa la relación funcional entre x y y.

Paso 3 Transforme cada punto dato de acuerdo con las reglas de la tercera columna de la tabla 13.

Paso 4 Estime la recta de regresión de mínimos cuadrados para los datos transformados. Si $\hat{\beta}_0$ es la ordenada al origen de la recta de mínimos cuadrados (para los datos transformados), y $\hat{\beta}_1$ es la pendiente de la recta de mínimos cuadrados (para los datos transformados), entonces obtenga la relación estimada de la columna final de la tabla 13.

Regresión múltiple

Se utiliza cuando se requiere más de una variable independiente para predecir y.

$$R^2 = \text{porcentaje de variación en } y \text{ explicado por las variables independientes}$$

Rechace H_0: $\beta_i = 0$ en el nivel de significancia α si (t para x_i) $\geq t_{(\alpha/2,n-k-1)}$, donde k es el número de variables independientes que se utilizan para predecir y.

Si hay una fuerte relación lineal entre dos o más variables independientes, entonces $\hat{\beta}_i$ podría ser una estimación poco confiable para β_i. En estos casos, decimos que existe **multicolinealidad.**

Si se cree que una variable independiente no cuantitativa (cualitativa), como el día de la semana o el mes del año, influye en una variable dependiente, se podrían usar variables ficticias para modelar el efecto de la variable independiente cualitativa sobre la variable dependiente. Si la variable cualitativa es capaz de asumir c valores, utilice sólo $c - 1$ variables ficticias.

PROBLEMAS DE REPASO

Grupo A

1 La información de la tabla 25 se relaciona con las ventas de cerdos (archivo Pork.xls). El precio está en dólares por cada cien libras vendidas, la cantidad vendida está en miles de millones de dólares, el ingreso per cápita está en dólares, la población estadounidense está en millones y el Producto Interno Bruto está en miles de millones de dólares.

a Utilice estos datos para desarrollar una ecuación de regresión que se pueda usar para pronosticar la cantidad de cerdos vendida en periodos futuros. ¿Es un problema la autocorrelación, heterocedasticidad o la multicolinealidad?

b Suponga que durante cada uno de los dos trimestres siguientes, precio = 45 dólares, población de E.U. = 240, PIB = 2620 e ingreso per cápita = 10 000 dólares. Pronostique la cantidad de cerdos vendidos durante cada uno de los dos trimestres próximos.

c Contamos con que 68% de las veces nuestras predicciones de las ventas de cerdo sean exactas dentro de ____________.

d Mediante el método de Winter, desarrolle un pronóstico para las ventas de cerdo durante los dos trimestres siguientes. (Use los dos primeros años para empezar.)

2 Vamos a pronosticar las ventas para una cadena de moteles de acuerdo con la información de la tabla 26 (archivo Motel.xls).

a Utilice esta información y la regresión múltiple para predecir las ventas de la cadena de moteles durante los cuatro trimestres siguientes. Suponga que la publicidad durante cada uno de los cuatro trimestres siguientes, es de 50 000 dólares.

TABLA 25

Trimestre	Año	Precio del cerdo	Cantidad vendida	Ingreso per cápita	Población de EE.UU.	PIB
1	1975	39.35	30.44	8255	212	1549
2	1975	46.11	29.23	8671	213	1589
3	1975	58.83	25.12	8583	214	1629
4	1975	52.2	28.35	8649	215	1669
1	1976	47.99	28.95	8775	216	1718
2	1976	49.19	27.83	8812	217	1768
3	1976	43.88	29.53	8884	218	1818
4	1976	34.25	35.9	8967	219	1868
1	1977	39.08	32.94	9036	220	1918
2	1977	40.87	31.86	9125	221	1978
3	1977	43.85	30.74	9280	222	2038
4	1977	41.38	34.99	9399	223	2098
1	1978	47.44	32.43	9487	224	2148
2	1978	47.84	32.65	9530	225	2218
3	1978	48.52	31.58	9622	226	2288
4	1978	50.05	35.40	9732	227	2338
1	1979	51.98	33.98	9813	228	2398
2	1979	48.04	37.58	9778	229	2448
3	1979	38.52	38.59	9809	230	2478
4	1979	36.39	43.47	9867	231	2508
1	1980	36.31	41.24	9958	232	2539
2	1980	31.18	43.00	9805	235	2598
3	1980	46.23	37.57	9882	235	2598

TABLA 26

Trimestre	Clientes potenciales (miles)	Publicidad (miles de dólares)	Estación	Ventas (millones)
1	100	30	Invierno	1200
2	105	20	Primavera	880
3	111	15	Verano	1800
4	117	40	Otoño	1050
5	122	10	Invierno	1700
6	128	50	Primavera	350
7	135	5	Verano	2500
8	142	40	Otoño	760
9	149	20	Invierno	2300
10	156	10	Primavera	1000
11	164	60	Verano	1570
12	172	5	Otoño	2430
13	181	35	Invierno	1320
14	190	15	Primavera	1400
15	200	70	Verano	1890
16	210	25	Otoño	3200
17	221	30	Invierno	2200
18	232	60	Primavera	1440
19	243	80	Verano	4000
20	264	60	Otoño	4100

TABLA 27

Año	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
1965	38	44	53	49	54	57	51	58	48	44	42	37
1966	42	43	53	49	49	40	40	36	29	31	26	23
1967	29	32	41	44	49	47	46	47	43	45	34	31
1968	35	43	46	46	43	41	44	47	41	40	32	32
1969	34	40	43	42	43	44	39	40	33	32	31	28
1970	34	29	36	42	43	44	44	48	45	44	40	37
1971	45	49	62	62	58	59	64	62	50	52	50	44
1972	51	56	60	65	64	63	63	72	61	65	51	47

b Mediante el método de Holt para pronosticar las ventas de la cadena de moteles durante los cuatro trimestres siguientes.

c Utilice suavizamiento exponencial simple para predecir las ventas de la cadena de moteles en los cuatro trimestres siguientes.

d Determine por medio del método de Winter los pronósticos para las ventas de la cadena de moteles en los cuatro trimestres siguientes.

e ¿Cuáles pronósticos esperaríamos que fueran los más confiables? (*Sugerencia*: utilice la publicidad, retrasada un periodo, como una variable independiente).

3 En la tabla 27 se proporcionan los siguientes datos sobre las ventas mensuales de casas en E.U. (en miles de casas) para 1965-1972.

a Utilice los años 1965 y 1966 para iniciar los parámetros para el método de Winter. Luego determine los valores de α, β y γ que generan un MAD (para 1967 a 1972) de menos de 3.5. (*Sugerencia*: podría ser necesario usar $\alpha > 0.5$).

b Contaríamos que 68% de los pronósticos sean exactos dentro de ______ y 95% de los mismos dentro de ________.

c Compruebe para ver si los datos son consistentes con la respuesta del inciso (b).

d Aunque no hemos analizado la autocorrelación para los métodos de suavizamiento, los buenos pronósticos que provienen de los métodos de suavizamiento no deben mostrar autocorrelación. ¿Los errores de pronóstico para este problema muestran autocorrelación?

e Se ha establecido que si sólo son factores importantes la tendencia y la estacionalidad, entonces α debe ser cuando mucho 0.5. Explique por qué este problema requiere $\alpha > 0.5$.

f A finales de diciembre de 1972, ¿cuál es el pronóstico para la venta de casas durante los primeros tres meses de 1973?

Nota: Esta tarea es cosa fácil con una hoja de cálculo. La hoja de cálculo podría ser como la tabla 28. En B14 se escribe =A$3*A14/D2+(1-A$3)*(B13+C13). Inserte fórmulas análogas en C14, D14, E14. Recuerde que el pronóstico debe hacerse antes de "ver" A14. Escriba en F14 =A14-E14. En G14 introduzca =ABS(F14). Copie de B14:G14 a ??. Para calcular MAD, promedie los errores absolutos de cada mes (renglones 14 a 85)

4 Si x es la variable independiente y y es la variable dependiente, determine la recta de mínimos cuadrados para los tres puntos dato siguientes:

x	y
1	2
4	5
7	2

TABLA 28

Renglón	A	B	C	D	E	F	G
1	VENTAS	BASE	TENDENCIA	ESTACIONALIDAD	PRONÓSTICO	ERRORES	ERRORES ABSOLUTOS
2	ALPHA	BETA	GAMMA	S-11			
3	.1	.2	.3	S-10			
.							
.							
.							
13		LO	TO	SO			
14	29						
15	32						
.							
.							
85	47						

TABLA 29

x	y
0	100
1	130
2	170
3	200
4	260
5	300
6	305
7	330
8	380

5 Pretendemos pronosticar el número de veces que se usa el cajero automático en función del tiempo. Los datos se proporcionan en la tabla 29. En este caso, x = número de años después de 1980, y y = número de veces que se usan al mes los cajeros (en millones) durante el año dado. La ecuación de regresión estimada es $\hat{y} = 102.3 + 34.8x$. Se nos dice que SST = 74 100, SSE = 1 298 y $EE(\hat{\beta}_1) = 1.76$.

a Pruebe H_0: $\beta_1 = 0$ contra H_a: $\beta_1 \neq 0$ para $\alpha = 0.05$. Interprete el resultado.

b Encuentre la correlación entre x y y.

c El dato de 1987 es un dato atípico.

d Si las tendencias actuales continúan, ¿cuál es la probabilidad aproximada de que, durante 1990, haya más de 470 millones de transacciones por mes en cajeros automáticos? (*Sugerencia*: apóyese en el hecho de que los errores están distribuidos en forma normal.)

6 Carboco coloca revestimientos metálicos en las hojas de los impulsores de los jets. A medida que el revestimiento es más duro, es más alta la calidad del mismo. El revestimiento se coloca mediante una pistola F, que contienen gas a presión, sobre la hoja del impulsor. Carboco puede controlar la temperatura y la presión del gas en la pistola; asimismo, puede controlar la humedad en el ambiente. Para ver cómo influyen la presión del gas, la temperatura y la humedad en la dureza, los ingenieros de Carboco han ejecutado una regresión para la cual la variable dependiente es

HARD = dureza de un revestimiento
y las variables independientes son
HUMIDI = humedad ambiente
TEMP = 1 si el nivel de temperatura es alto
= 0 si el nivel de temperatura es bajo
PRESS = 1 si la presión del gas es alta
= 0 si la presión del gas es baja
T*P = producto de TEMP y PRESS

Suponemos que la temperatura y la presión del gas tienen sólo dos niveles posibles, bajo y alto. Los datos relevantes se proporcionan en la tabla 30 (archivo Temp.xls).

a Si se ignoran las consideraciones de heterocedasticidad, multicolinealidad y autocorrelación, ¿Qué ecuación se debe usar para predecir la dureza explique.

b ¿Qué combinación de presión de gas y de temperatura establecerán la dureza máxima?

c Explique cómo el cambiar la temperatura desde un nivel bajo a un nivel alto afecta la dureza. ¡Sea específico!

TABLA 30

Humidi	Temp	Press	Hard
40	1	0	148
60	1	0	209
50	1	0	177
70	1	0	208
80	1	0	262
60	1	1	248
65	1	1	253
70	1	1	263
35	1	1	184
45	1	1	220
70	0	0	129
28	0	0	53
49	0	0	98
89	0	0	170
90	0	0	172
34	0	1	80
56	0	1	90
77	0	1	151
23	0	1	58
56	0	1	107

7 Hemos sido asignados para determinar cómo el costo de producción total a la semana de Widgetco depende de la cantidad de productos fabricados en la semana. Se ha propuesto el modelo siguiente:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$$

donde X = número de productos fabricados durante la semana y Y = costo total de producción en la semana. Para 15 semanas de datos, encontramos que SSR = 215 475 y SST = 229 228. Para este modelo obtenemos la siguiente ecuación de regresión estimada (el valor de t para cada coeficiente está entre paréntesis):

$$\hat{y} = -29.7 + \underset{(0.78)}{19.8X} - \underset{(0.62)}{0.39X^2} + \underset{(1.25)}{0.005X^3}$$

a Para $\alpha = 0.10$, pruebe H_0: $\beta_i = 0$ contra H_a: $\beta_i \neq 0$ ($i = 1, 2, 3$).

b Determine R^2 para este modelo. ¿Cómo puede el valor alto de R^2 reconciliarse con la respuesta del inciso (a)?

8 Sea Y_t = ventas durante el mes t (en miles de dólares) de un estudio fotográfico (SALES en la tabla 31) y P_t = precio cargado a los retratos durante el mes t (PRICE). Utilice una computadora para ajustar el modelo siguiente a los datos de la tabla 31 (archivo Portrait.xls):

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 P_t + \varepsilon_t$$

Por lo tanto, las ventas del último mes y el precio del mes actual son variables independientes.

a Si el precio de un retrato durante el mes 21 es 10 dólares, ¿cuál sería el pronóstico para las ventas del mes 21?

b ¿Parece haber un problema con la autocorrelación, heterocedasticidad o la multicolinealidad?