

CAPÍTULO 1



Datos y estadísticas

CONTENIDO

LA ESTADÍSTICA EN LA PRÁCTICA: *BUSINESSWEEK*

1.1 APLICACIONES EN LOS NEGOCIOS Y EN LA ECONOMÍA
Contaduría
Finanzas
Marketing
Producción
Economía

1.2 DATOS
Elementos, variables y observaciones
Escala de medición

Datos cualitativos y cuantitativos
Datos de sección transversal y de series de tiempo

1.3 FUENTES DE DATOS
Fuentes existentes
Estudios estadísticos
Errores en la adquisición de datos

1.4 ESTADÍSTICA DESCRIPTIVA

1.5 INFERENCIA ESTADÍSTICA

1.6 LAS COMPUTADORAS Y EL ANÁLISIS ESTADÍSTICO



LA ESTADÍSTICA *en* LA PRÁCTICA

*BUSINESSWEEK** NUEVA YORK, NUEVA YORK

Con una circulación mundial de más de 1 millón de ejemplares, *BusinessWeek* es la revista más leída en el mundo. Más de 200 reporteros y editores especializados en 26 oficinas alrededor del mundo producen diversos artículos de interés para la comunidad interesada en los negocios y la economía. Junto a los artículos principales y los tópicos de actualidad, la revista presenta diversas secciones regulares sobre negocios internacionales, análisis económicos, procesamiento de la información y ciencia y tecnología. La información en las secciones regulares ayuda a los lectores a mantenerse al día de los avances y novedades y a evaluar el impacto de éstos en los negocios y en las condiciones económicas.

La mayor parte de los números de *BusinessWeek* contienen un artículo de fondo sobre algún tema de interés actual. Por ejemplo, el número del 6 de diciembre de 2004 contenía un reportaje especial sobre los precios de los artículos hechos en China; el número del 3 de enero de 2005 proporcionaba información acerca de dónde invertir en 2005 y el número del 4 de abril de 2005 proporcionaba una panorámica de *BusinessWeek 50*, un grupo diverso de empresas de alto desempeño. Además, la revista semanal *BusinessWeek Investor* proporciona artículos sobre el estado de la economía, que comprenden índices de producción, precios de las acciones de fondos mutualistas y tasas de interés.

BusinessWeek también usa métodos e información estadísticos en la administración de su propio negocio. Por ejemplo, una encuesta anual hecha a sus suscriptores le permitió tener datos demográficos sobre sus hábitos de lectura, compras probables, estilo de vida, etc. Los directivos de *BusinessWeek* usan resúmenes estadísticos obtenidos a partir de las encuestas para dar un mejor servicio a sus sus-

*Los autores agradecen a Charlene Trentham, Director de investigación de *BusinessWeek* por proporcionar este artículo para La estadística en la práctica.



BusinessWeek usa datos y resúmenes estadísticos en muchos de sus artículos. © Terri Millar/E-Visual Communications, Inc.

criptores y anunciantes. Mediante una encuesta reciente entre los suscriptores estadounidenses se supo que 90% de los suscriptores de *BusinessWeek* tienen una computadora personal en casa y que 64% de ellos realizan en el trabajo compras por computadora. Estas estadísticas indican a los directivos de *BusinessWeek* que los avances en computación serán de interés para sus suscriptores. Los resultados de la encuesta también le son proporcionados a sus anunciantes potenciales. Los elevados porcentajes de personas que tienen una computadora en casa y que realizan compras por computadora en el trabajo podría ser un incentivo para que los fabricantes de computadoras se anunciaran en *BusinessWeek*.

Este capítulo muestra los tipos de datos con que se cuenta en un análisis estadístico y describe cómo se obtienen los datos. Presenta la estadística descriptiva y la inferencia estadística como medios para convertir los datos en información estadística que tienen un significado y que es fácil de interpretar.

Con frecuencia aparece en los periódicos y revistas el siguiente tipo de información:

- La asociación de agentes inmobiliarios informó que la mediana del precio de venta de una casa en Estados Unidos es de \$215 000 (*The Wall Street Journal*, 16 de enero de 2006).
- Durante el Super Bowl de 2006 el costo promedio de un spot publicitario de 30 segundos en televisión fue de \$2.5 millones (*USA Today*, 27 de enero de 2007).

- En una encuesta de Jupiter Media se encontró que 31% de los hombres adultos ven más de 10 horas de televisión a la semana. Entre las mujeres sólo 26% (*The Wall Street Journal*, 26 de enero de 2004).
- General Motors, uno de los líderes automotrices en descuentos en efectivo da, en promedio, \$4300 de incentivo en efectivo por vehículo (*USA Today*, 27 de enero de 2006).
- Más de 40% de los directivos de Marriott Internacional ascienden por escalafón (*Fortune*, 20 de enero de 2003).
- Los Yankees de Nueva York tienen la nómina más alta dentro de la liga mayor de béisbol. En el año 2005 la nómina del equipo fue de \$208 306 817, siendo la mediana por jugador de \$5 833 334 (*USA Today*, febrero 2006).
- El promedio industrial Dow Jones cerró en 11 577 (*Barron's*, 6 de mayo de 2006).

A los datos numéricos de las frases anteriores se les llama estadísticas. En este sentido el término *estadística* se refiere a datos numéricos, tales como promedios, medianas, porcentajes y números índices que ayudan a entender una gran variedad de negocios y situaciones económicas. Sin embargo, como se verá, el campo de la estadística es mucho más que datos numéricos. En un sentido amplio, la **estadística** se define como el arte y la ciencia de reunir datos, analizarlos, presentarlos e interpretarlos. Especialmente en los negocios y en la economía, la información obtenida al reunir datos, analizarlos, presentarlos e interpretarlos proporciona a directivos, administradores y personas que deben tomar decisiones una mejor comprensión del negocio o entorno económico, permitiéndoles así tomar mejores decisiones con base en mejor información. En este libro se hace hincapié en el uso de la estadística para la toma de decisiones en los negocios y en la economía.

El capítulo 1 empieza con algunos ejemplos de aplicaciones de la estadística en los negocios y en la economía. En la sección 1.2 se define el término *datos* y se introduce el concepto de conjunto de datos. En esta sección se introducen también términos clave como *variables* y *observaciones*, se muestra la diferencia entre datos cualitativos y cuantitativos y se ilustra el uso de datos transversales y de serie de tiempo. En la sección 1.3 se enseña a obtener datos de fuentes ya existentes o mediante encuestas y estudios experimentales diseñados para obtener datos nuevos. Se resalta también el papel tan importante que tiene ahora Internet en la obtención de datos. En las secciones 1.4 y 1.5 se describe el uso de los datos en la estadística descriptiva y para hacer inferencias estadísticas.

1.1

Aplicaciones en los negocios y en la economía

En el entorno mundial actual de los negocios y de la economía, todo mundo tiene acceso a enormes cantidades de información estadística. Los directivos y los encargados de tomar decisiones que tienen éxito entienden la información y saben usarla de manera eficiente. En esta sección se proporcionan ejemplos que ilustran algunos de los usos de la estadística en los negocios y en la economía.

Contaduría

Las empresas de contadores públicos al realizar auditorías para sus clientes emplean procedimientos de muestreo estadístico. Por ejemplo, suponga que una empresa de contadores desea determinar si las cantidades en cuentas por cobrar que aparecen en la hoja de balance del cliente representan la verdadera cantidad en cuentas por cobrar. Por lo general, el gran número de cuentas por cobrar hace que su revisión tome demasiado tiempo y sea muy costosa. Lo que se hace en estos casos es que el personal encargado de la auditoría selecciona un subconjunto de las cuentas al que se le llama muestra. Después de revisar la exactitud de las cuentas tomadas en la muestra (muestreadas) los auditores concluyen si la cantidad en cuentas por cobrar que aparece en la hoja de balance del cliente es aceptable.

Finanzas

Los analistas financieros emplean una diversidad de información estadística como guía para sus recomendaciones de inversión. En el caso de acciones, el analista revisa diferentes datos financieros como la relación precio/ganancia y el rendimiento de los dividendos. Al comparar la información sobre una determinada acción con la información sobre el promedio en el mercado de acciones, el analista empieza a obtener conclusiones para saber si una determinada acción está sobre o subvaluada. Por ejemplo, *Barron's* (12 de septiembre de 2005) informa que la relación promedio precio/ganancia de 30 acciones del promedio industrial Dow Jones fue 16.5. La relación precio/ganancia de JPMorgan es 11.8. En este caso la información estadística sobre las relaciones precio/ganancia indican un menor precio en comparación con la ganancia para JPMorgan que el promedio en las acciones Dow Jones. Por tanto el analista financiero concluye que JPMorgan está subvaluada. Ésta y otras informaciones acerca de JPMorgan ayudarán al analista a comprar, vender o a recomendar mantener las acciones.

Marketing

Escáneres electrónicos en las cajas de los comercios minoristas recogen datos para diversas aplicaciones en la investigación de mercado. Por ejemplo, proveedores de datos como ACNielsen e Information Research Inc. compran estos datos a las tiendas de abarrotes, los procesan y luego venden los resúmenes estadísticos a los fabricantes; quienes gastan cientos de miles de dólares por producto para obtener este tipo de datos. Los fabricantes también compran datos y resúmenes estadísticos sobre actividades promocionales como precios o *displays* promocionales. Los administradores de marca revisan estas estadísticas y las propias de las actividades promocionales para analizar la relación entre una actividad promocional y las ventas. Estos análisis suelen resultar útiles para establecer futuras estrategias de marketing para diversos productos.

Producción

La importancia que se le da actualmente a la calidad hace del control de calidad una aplicación importante de la estadística a la producción. Para vigilar el resultado de los procesos de producción se usan diversas gráficas de control estadístico de calidad. En particular, para vigilar los resultados promedio se emplea una gráfica \bar{x} -barra. Suponga, por ejemplo, que una máquina llena botellas con 12 onzas de algún refresco. Periódicamente un empleado del área de producción toma una muestra de botellas y mide el contenido promedio de refresco. Este promedio o valor \bar{x} -barra se marca como un punto en una gráfica \bar{x} -barra. Si este punto queda arriba del límite de control superior de la gráfica, hay un exceso en el llenado, y si queda debajo del límite de control inferior de la gráfica hay falta de llenado. Se dice que el proceso está “bajo control” y puede continuar, siempre que los valores \bar{x} -barra se encuentren entre los límites de control inferior y superior. Con una interpretación adecuada, una gráfica de \bar{x} -barra ayuda a determinar si es necesario hacer algún ajuste o corrección a un proceso de producción.

Economía

Los economistas suelen hacer pronósticos acerca del futuro de la economía o sobre algunos aspectos de la misma. Usan una variedad de información estadística para hacer sus pronósticos. Por ejemplo, para pronosticar las tasas de inflación, emplean información estadística sobre indicadores como el índice de precios al consumidor, la tasa de desempleo y la utilización de la capacidad de producción. Estos indicadores estadísticos se utilizan en modelos computarizados de pronósticos que predicen las tasas de inflación.

Aplicaciones de la estadística como las descritas en esta sección integran este libro. Dichos ejemplos proporcionan una visión general de la diversidad de las aplicaciones estadísticas. Como complemento de estos ejemplos, profesionales en los campos de los negocios y de la economía proporcionan los artículos de *La estadística en la práctica* que se encuentran al principio de cada capítulo, en los que se presenta el material que se estudiará en el capítulo. Las aplicaciones en *La estadística en la práctica* muestran su importancia en diversas situaciones de los negocios y la economía.

1.2 Datos

Datos son hechos/informaciones y cifras que se recogen, analizan y resumen para su presentación e interpretación. A todos los datos reunidos para un determinado estudio se les llama **conjunto de datos** para el estudio. La tabla 1.1 muestra un conjunto de datos que contiene información sobre 25 empresas que forman parte del S&P 500. El S&P 500 consta de 500 empresas elegidas por Standard & Poor's. Estas empresas representan 76% de la capitalización de mercado de todas las acciones de Estados Unidos. Las acciones de S&P 500 son estrechamente observadas por los inversionistas y por los analistas de Wall Street.

TABLA 1.1 CONJUNTO DE DATOS DE 25 EMPRESAS S&P 500

Empresa	Bolsa de valores	Denominación abreviada Ticker	Posición en <i>BusinessWeek</i>	Precio por acción (\$)	Ganancia por acción (\$)
Abbott Laboratories	N	ABT	90	46	2.02
Altria Group	N	MO	148	66	4.57
Apollo Group	NQ	APOL	174	74	0.90
Bank of New York	N	BK	305	30	1.85
Bristol-Myers Squibb	N	BMJ	346	26	1.21
Cincinnati Financial	NQ	CINF	161	45	2.73
Comcast	NQ	CMCSA	296	32	0.43
Deere	N	DE	36	71	5.77
eBay	NQ	EBAY	19	43	0.57
Federated Dept. Stores	N	FD	353	56	3.86
Hasbro	N	HAS	373	21	0.96
IBM	N	IBM	216	93	4.94
International Paper	N	IP	370	37	0.98
Knight-Ridder	N	KRI	397	66	4.13
Manor Care	N	HCR	285	34	1.90
Medtronic	N	MDT	53	52	1.79
National Semiconductor	N	NSM	155	20	1.03
Novellus Systems	NQ	NVLS	386	30	1.06
Pitney Bowes	N	PBI	339	46	2.05
Pulte Homes	N	PHM	12	78	7.67
SBC Communications	N	SBC	371	24	1.52
St. Paul Travelers	N	STA	264	38	1.53
Teradyne	N	TER	412	15	0.84
UnitedHealth Group	N	UNH	5	91	3.94
Wells Fargo	N	WFC	159	59	4.09

Fuente: Business Week (4 de abril de 2005).

Elementos, variables y observaciones

Elementos son las entidades de las que se obtienen los datos. En el conjunto de datos de la tabla 1.1, cada acción de una empresa es un elemento; los nombres de los elementos aparecen en la primera columna. Como se tienen 25 acciones, el conjunto de datos contiene 25 elementos.

Una **variable** es una característica de los elementos que es de interés. El conjunto de datos de la tabla 1.1 contiene las cinco variables siguientes:

- *Bolsa de valores (mercado bursátil)*: Dónde se comercializa (cotiza) la acción: N (Bolsa de Nueva York) y NQ (Mercado Nacional Nasdaq).
- *Ticker (denominación abreviada)*: Abreviación usada para identificar la acción en la lista de la bolsa
- *Posición en BusinessWeek*: Número del 1 al 500 que indica la fortaleza de la empresa.
- *Precio por acción (\$)*: El precio de cierre (28 de febrero de 2005).
- *Ganancia por acción (\$)*: Las ganancias por acción en los últimos 12 meses.

Los valores encontrados para cada variable en cada uno de los elementos constituyen los datos. Al conjunto de mediciones obtenidas para un determinado elemento se le llama **observación**. Volviendo a la tabla 1.1, el conjunto de mediciones para la primera observación (Abbott Laboratories) es N, ABT, 90, 46 y 2.02. El conjunto de mediciones para la segunda observación (Altria Group) es N, MO, 148, 66 y 4.57, etc. Un conjunto de datos que tiene 25 elementos contiene 25 observaciones.

Escalas de medición

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

Cuando el dato de una variable es una etiqueta o un nombre que identifica un atributo de un elemento, se considera que la escala de medición es una **escala nominal**. Por ejemplo, en relación con la tabla 1.1 la escala de medición para la variable bolsa de valores (mercado bursátil) es nominal porque N y NQ son etiquetas que se usan para indicar dónde cotiza la acción de la empresa. Cuando la escala de medición es nominal, se usa un código o una etiqueta no numérica. Por ejemplo, para facilitar la recolección de los datos y para guardarlos en una base de datos en una computadora puede emplearse un código numérico en el que 1 denote la Bolsa de Nueva York y 2 el Mercado Nacional Nasdaq. En este caso los números 1 y 2 son las etiquetas empleadas para identificar dónde cotizan las acciones. La escala de medición es nominal aun cuando los datos aparezcan como valores numéricos.

Una escala de medición para una variable es **ordinal** si los datos muestran las propiedades de los datos nominales y además tiene sentido el orden o jerarquía de los datos. Por ejemplo, una empresa automovilística (Eastside Automotive) envía a sus clientes cuestionarios para obtener información sobre su servicio de reparación. Cada cliente evalúa el servicio de reparación como excelente, bueno o malo. Como los datos obtenidos son las etiquetas excelente, bueno o malo, tienen las propiedades de los datos nominales, pero además pueden ser ordenados o jerarquizados en relación con la calidad del servicio. Un dato excelente indica el mejor servicio, seguido por bueno y, por último, malo. Por lo que la escala de medición es ordinal. Observe que los datos ordinales también son registrados mediante un código numérico. Por ejemplo, en la tabla 1.1 la posición de los datos en *BusinessWeek* es un dato ordinal. Da una jerarquía del 1 al 500 de acuerdo con la evaluación de *BusinessWeek* sobre la fortaleza de la empresa.

Una escala de medición para una variable es una **escala de intervalo** si los datos tienen las características de los datos ordinales y el intervalo entre valores se expresa en términos de una unidad de medición fija. Los datos de intervalo siempre son numéricos. Las calificaciones en una prueba de aptitudes escolares son un ejemplo de datos de intervalo. Por ejemplo, las ca-

lificaciones obtenidas por tres alumnos en la prueba de matemáticas con 620, 550 y 470, pueden ser ordenadas en orden de mejor a peor. Además las diferencias entre las calificaciones tienen significado. Por ejemplo, el estudiante 1 obtuvo $620 - 550 = 70$ puntos más que el estudiante 2 mientras que el estudiante 2 obtuvo $550 - 470 = 80$ puntos más que el estudiante tres.

Una variable tiene una **escala de razón** si los datos tienen todas las propiedades de los datos de intervalo y la proporción entre dos valores tiene significado. Variables como distancia, altura, peso y tiempo usan la escala de razón en la medición. Esta escala requiere que se tenga el valor cero para indicar que en este punto no existe la variable. Por ejemplo, considere el costo de un automóvil. El valor cero para el costo indica que el automóvil no cuesta, que es gratis. Además, si se compara el costo de un automóvil de \$30 000, con el costo de otro automóvil, \$15 000, la propiedad de razón muestra que $\$30\,000/\$15\,000 = 2$: el primer automóvil cuesta el doble del costo del segundo.

Datos cualitativos y cuantitativos

A los datos cualitativos se les suele llamar datos categóricos.

Los datos también son clasificados en cualitativos y cuantitativos. Los **datos cualitativos** comprenden etiquetas o nombres que se usan para identificar un atributo de cada elemento. Los datos cualitativos emplean la escala nominal o la ordinal y pueden ser numéricos o no. Los **datos cuantitativos** requieren valores numéricos que indiquen cuánto o cuántos. Los datos cuantitativos se obtienen usando las escalas de medición de intervalo o de razón.

El método estadístico adecuado para resumir los datos depende de si los datos son cualitativos o cuantitativos.

Una **variable cualitativa** es una variable con datos cualitativos. El análisis estadístico adecuado para una determinada variable depende de si la variable es cualitativa o cuantitativa. Si la variable es cualitativa, el análisis estadístico es bastante limitado. Tales datos se resumen contando el número de observaciones o calculando la proporción de observaciones en cada categoría cualitativa. Sin embargo, aun cuando para los datos cualitativos se use un código numérico, las operaciones aritméticas de adición, sustracción, multiplicación o división no tienen sentido. En la sección 2.1 se ven las formas de resumir datos cualitativos.

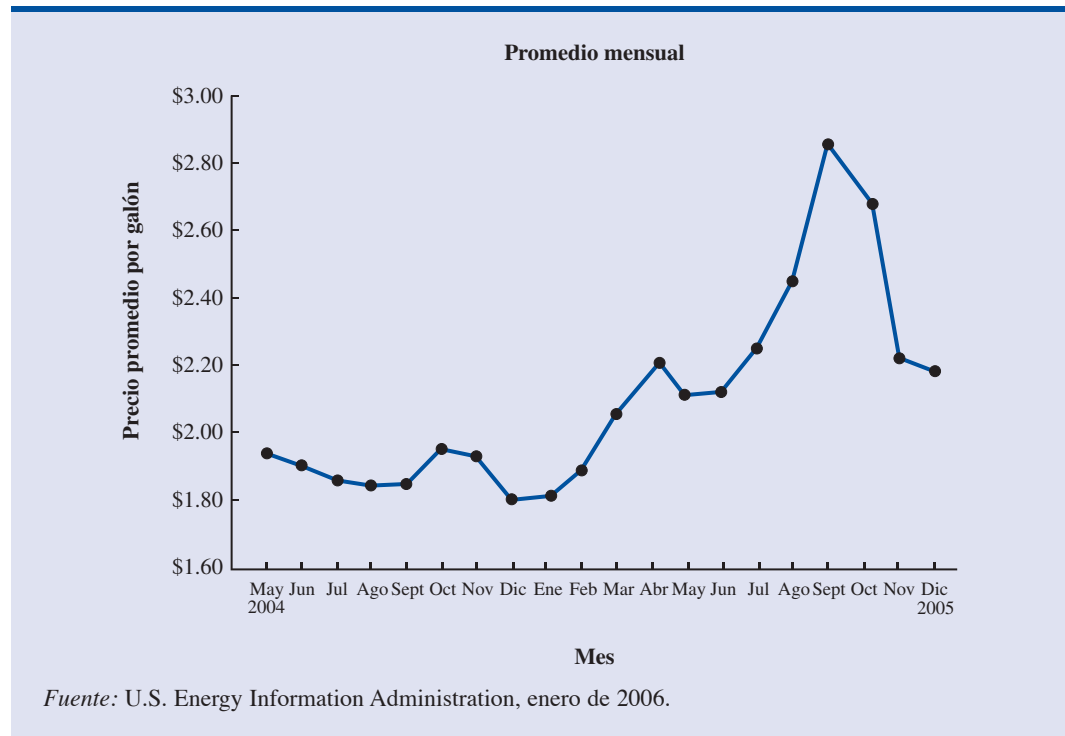
Por otro lado, las operaciones aritméticas sí tienen sentido en las variables cuantitativas. Por ejemplo, cuando se tienen variables cuantitativas, los datos se pueden sumar y luego dividir entre el número de observaciones para calcular el valor promedio. Este promedio suele ser útil y fácil de interpretar. En general hay más alternativas para el análisis estadístico cuando se tienen datos cuantitativos. La sección 2.2 y el capítulo 3 proporcionan condiciones para resumir datos cuantitativos.

Datos de sección transversal y de series de tiempo

Para los propósitos del análisis estadístico la distinción entre datos transversales y datos de series de tiempo es importante. **Datos de sección transversal** son los obtenidos en el mismo o aproximadamente el mismo momento (punto en el tiempo). Los datos de la tabla 1.1 son datos transversales porque describen las cinco variables de las 25 empresas del 25 S&P en un mismo momento. Los **datos de series de tiempo** son datos obtenidos a lo largo de varios periodos. Por ejemplo, la figura 1.1 presenta una gráfica de los precios promedio por galón de gasolina normal en las ciudades de Estados Unidos. En la gráfica se observa que los precios son bastantes estables entre \$1.80 y \$2.00 desde mayo de 2004 hasta febrero de 2005. Después el precio de la gasolina se vuelve volátil. Se eleva en forma notable culminando en un agudo pico en septiembre de 2005.

En las publicaciones sobre negocios y economía se encuentran con frecuencia gráficas de series de tiempo. Estas gráficas ayudan a los analistas a entender lo que ocurrió en el pasado, a identificar cualquier tendencia en el transcurso del tiempo y a proyectar niveles futuros para la series de tiempo. Las gráficas de datos de series de tiempo toman formas diversas como se muestra en la figura 1.2. Con un poco de estudio, estas gráficas suelen ser fáciles de entender y de interpretar.

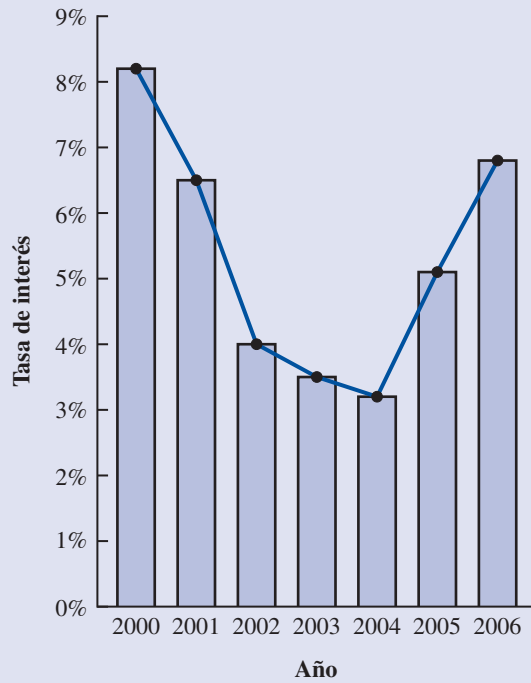
FIGURA 1.1 PRECIO PROMEDIO POR GALÓN DE GASOLINA NORMAL EN LAS CIUDADES DE ESTADOS UNIDOS



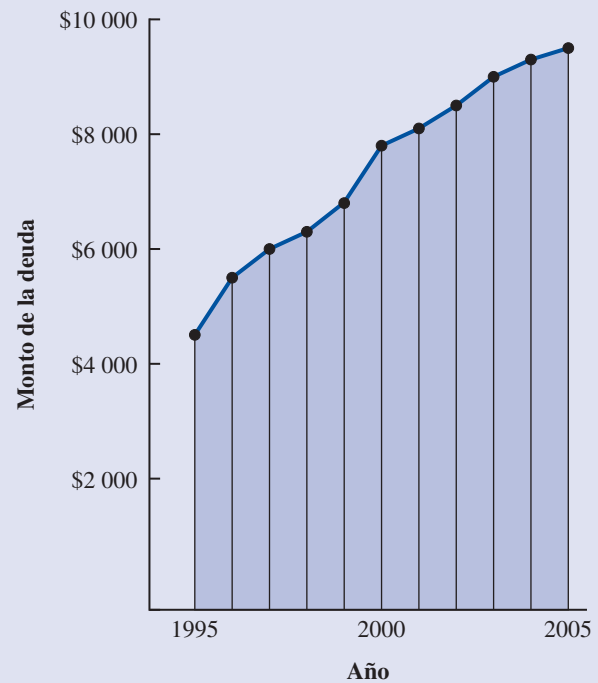
Por ejemplo, la gráfica (A) de la figura 1.2, muestra las tasas de interés en Stafford Loans para los estudiantes entre el año 2000 y el 2006. Después del año 2000 las tasas de interés disminuyen y llegan al nivel más bajo, 3.2%, en el año 2004. Pero, después de este año se observa un marcado aumento en estas tasas de interés, y llegan a 6.8% en el año 2006. El Departamento de Educación de Estados Unidos estima que más de 50% de los estudiantes terminan sus estudios con una deuda; esta creciente tasa de interés es una gran carga financiera para muchos estudiantes recién egresados.

En la gráfica (B) se observa un inquietante aumento en el adeudo promedio por hogar en tarjetas de crédito durante un periodo de 10 años, de 1995 a 2005. Advierta cómo en la serie de tiempo se nota un aumento anual casi constante en el adeudo promedio por hogar en tarjetas de crédito que va de \$4500 en 1995 a \$9500 en 2005. En 2005 un adeudo promedio de 10 000 no parece lejano. La mayor parte de las empresas de tarjetas de crédito ofrecen tasas de interés iniciales relativamente bajas. Sin embargo, después de este periodo inicial, tasas de interés anuales del 18%, 20% y más son frecuentes. Estas tasas dificultan a los hogares pagar los adeudos de las tarjetas de crédito.

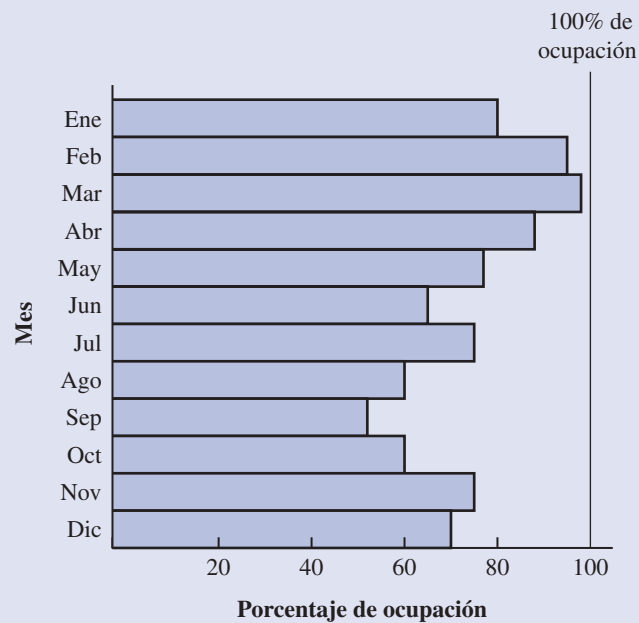
En la gráfica (C) se observan las tasas de ocupación en los hoteles de Florida del sur durante un año. Observe que la forma de esta gráfica es diferente a (A) y (B); en esta gráfica el tiempo en meses se encuentra en el eje vertical y no en el horizontal. Las tasas de ocupación más altas, 95% y 98%, se encuentran en los meses de febrero y marzo que es cuando el clima en Florida del sur es atractivo para los turistas. En efecto, de enero a abril es la estación de mayor ocupación en los hoteles de Florida del sur. Por otro lado, las tasas de ocupación más bajas se observan de agosto a octubre, siendo la menor ocupación en septiembre. Las temperaturas demasiado elevadas y la estación de huracanes son las principales razones de la caída de la ocupación en este periodo.

FIGURA 1.2 DIVERSAS GRÁFICAS DE DATOS DE SERIES DE TIEMPO

(A) Tasas de interés en los Stafford Loans para estudiantes



(B) Adeudo promedio en tarjetas de crédito por hogar



(C) Tasas de ocupación en hoteles de Florida del sur

Las series de tiempo y los pronósticos con series de tiempo se verán en el capítulo 16 cuando se estudien los métodos de pronóstico. Fuera del capítulo 16, los métodos estadísticos que se presentan en este libro son para datos de sección transversal y no para series de tiempo

NOTAS Y COMENTARIOS

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

1.3

Fuentes de datos

Los datos se obtienen de fuentes ya existentes o por medio de encuestas y estudios experimentales realizados con objeto de recolectar nuevos datos.

Fuentes existentes

En algunos casos los datos que se necesitan para una determinada aplicación ya existen. Las empresas cuentan con diversas bases de datos sobre sus empleados, clientes y operaciones de negocios. Datos sobre los salarios de los empleados, sus edades y los años de experiencia suelen obtenerse de los registros internos del personal. Otros registros internos contienen datos sobre ventas, gastos de publicidad, costos de distribución, inventario y cantidades de producción. La mayor parte de las empresas cuentan también con datos detallados de sus clientes. En la tabla 1.2 se muestran algunos de los datos obtenibles de los registros internos de las empresas.

De las organizaciones que se especializan en la recolección y almacenamiento de datos se obtienen cantidades importantes de datos económicos y de negocios. Las empresas disponen de estas fuentes externas de datos si los compran o mediante acuerdos de arrendamiento con opción de compra. Tres empresas que proporcionan amplios servicios de bases de datos a clientes son Dun & Bradstreet, Bloomberg y Dow Jones & Company. ACNielsen e Information Resources, Inc. han hecho un exitoso negocio recolectando y procesando datos que venden a publicistas y a fabricantes de productos.

TABLA 1.2 EJEMPLOS DE DATOS DISPONIBLES DE LOS REGISTROS DE EMPRESAS INTERNACIONALES

Fuente	Algunos de los datos disponibles
Registros sobre los empleados	Nombre, dirección, número de seguridad social, salario, días de vacaciones, días de enfermedad y bonos
Registros de producción	Parte o número de producto, cantidad producida, costo de mano de obra y costo de materiales
Registros de inventario	Parte o número de producto, cantidad de unidades disponibles, nivel de reaprovisionamiento, cantidad económica a ordenar y programa de descuento
Registros de ventas	Número del producto, volumen de ventas, volumen de ventas por región y volumen de ventas por tipo de cliente
Registros de créditos	Nombre del cliente, dirección, número de teléfono, crédito límite y cuentas por cobrar
Perfil de clientes	Edad, género, nivel de ingresos, número de miembros en la familia, dirección y preferencias

También se obtienen datos de diversas asociaciones industriales y de organizaciones de interés especial. La asociación Travel Industry Association of America cuenta con información relacionada con los viajes como número de turistas y gastos en viajes por estado. Estos datos interesan a empresas e individuos de la industria turística. El Graduate Management Admission Council cuenta con datos sobre calificaciones en exámenes, características de los estudiantes y programas de educación para administradores/directivos. La mayor parte de los datos de estas fuentes están a disposición de los usuarios calificados a un costo moderado.

La importancia de Internet como fuente de datos y de información estadística sigue creciendo. Casi todas las empresas cuentan con una página Web que proporciona información general acerca de la empresa así como datos sobre ventas, cantidad de empleados, cantidad de productos, precios de los productos y especificaciones de los productos. Además, muchas empresas se especializan ahora en proporcionar información a través de Internet. Con lo que uno puede tener acceso a cotizaciones de acciones, precios de comidas en restaurantes, datos de salarios y a una variedad casi infinita de información.

Las dependencias de los gobiernos son otra fuente importante de datos. Por ejemplo, el Departamento del Trabajo de Estados Unidos cuenta con una cantidad considerable de datos sobre tasas de empleo, tasas de salarios, magnitud de la fuerza laboral y pertenencia a sindicatos. En la tabla 1.3 se presentan algunas de las dependencias de gobierno junto con los datos que proporcionan. La mayor parte de las dependencias de los gobiernos que recolectan y procesan datos también los ponen a disposición a través de una página en la Web. Por ejemplo, la Oficina de Censos de Estados Unidos tiene una abundancia de datos en el sitio www.census.gov. En la figura 1.3 se muestra la página Web de la Oficina de Censos de Estados Unidos.

Estudios estadísticos

Algunas veces, los datos necesarios para una aplicación particular no se pueden obtener de las fuentes existentes. En tales casos los datos suelen conseguirse realizando un estudio estadístico. Dichos estudios se clasifican como *experimentales* u *observacionales*.

En los estudios experimentales se identifica primero la variable de interés. Después se ubica otra u otras variables que son controladas para lograr datos de cómo ésta influye sobre la variable de interés. Por ejemplo, a una empresa farmacéutica le interesa realizar un experimento para saber la forma en que un medicamento afecta la presión sanguínea. La variable que interesa en el estudio es la presión sanguínea. Otra variable es la dosis del nuevo medicamento que se espera tenga un efecto causal sobre la presión sanguínea. Para obtener estos datos acerca del nuevo medicamento, los investigadores eligen una muestra de individuos. La dosis del medicamento se controla dando diferentes dosis a distintos grupos de individuos. Antes y después se mide la pre-

El mayor estudio estadístico experimental jamás realizado se cree que es el experimento del Servicio de Salud Pública para la vacuna Salk contra la polio. Se eligieron casi 2 millones de niños de 1o., 2o. y 3er. grados en Estados Unidos.

TABLA 1.3 EJEMPLO DE LOS DATOS DISPONIBLES DE ALGUNAS DEPENDENCIAS GUBERNAMENTALES

Dependencia gubernamental	Algunos de los datos disponibles
Oficina de Censos www.census.gov	Datos poblacionales, número de hogares e ingresos de los hogares
Junta de la Reserva Federal www.federalreserve.gov	Datos sobre dinero en circulación, créditos a plazos, tasas de cambio y tasas de interés
Oficina de Administración y Presupuesto www.whitehouse.gov/omb	Datos sobre ingresos, gastos y deudas del gobierno federal
Departamento de Comercio www.doc.gov	Datos sobre las actividades comerciales, valor de los embarques por industria, nivel de ganancia por industria e industrias en crecimiento y en decremento
Oficina de Estadística Laboral www.bls.gov	Gasto de los consumidores, salarios por hora, tasa de desempleo y estadísticas internacionales

FIGURA 1.3 PÁGINA DE INICIO DEL SITIO WEB DE LA OFICINA DE CENSOS DE ESTADOS UNIDOS

Los estudios sobre fumadores y no fumadores son estudios observacionales porque los investigadores no determinan o controlan quién fuma y quién no.

sión sanguínea en cada grupo. El análisis estadístico de los datos experimentales ayuda a determinar el efecto del nuevo medicamento sobre la presión sanguínea.

En los estudios estadísticos no experimentales y observacionales, no se controlan las variables de interés. El tipo más usual de estudio observacional es quizá una encuesta. Por ejemplo, en una encuesta mediante entrevistas personales, primero se identifican las preguntas de la investigación. Después se presenta un cuestionario a los individuos de la muestra. Algunos restaurantes emplean estudios observacionales para obtener datos acerca de la opinión de sus clientes respecto a la calidad de los alimentos, del servicio, de la atmósfera, etc. En la figura 1.4 se presenta un cuestionario empleado por el restaurante Lobster Pot de Florida. Observe que en el cuestionario se pide a los clientes evaluar cinco variables: calidad de los alimentos, amabilidad en el servicio, prontitud en el servicio, limpieza y gestión. Las categorías para las respuestas de excelente, bueno, satisfactorio e insatisfactorio proporcionan datos ordinales que permiten a los directivos de Lobster Pot evaluar la calidad de operación del restaurante.

Los directivos que deseen emplear datos y análisis estadístico como ayuda en la toma de decisiones deben estar conscientes del tiempo y costo que requiere la obtención de los datos. Cuando es necesario obtener los datos en poco tiempo, es deseable el uso de fuentes de datos ya existentes. Si no es posible obtener con facilidad datos importantes de fuentes ya existentes, debe tomarse en cuenta el tiempo y el costo necesarios para obtener los datos. En todos los casos, las personas encargadas de tomar las decisiones deben considerar la contribución del análisis estadístico en el proceso de la toma de decisiones. El costo de la adquisición de datos y del subsiguiente análisis no deben exceder a los ahorros generados por el uso de esta información para tomar una decisión mejor.

Errores en la adquisición de datos

Los directivos siempre deben estar conscientes de la posibilidad de errores en los datos de los estudios estadísticos. Usar datos erróneos es peor que no usar ningún dato. Un error en la adquisición de datos se tiene siempre que el valor del dato obtenido no es igual al verdadero valor o al valor real que se hubiera obtenido con un procedimiento correcto. Estos errores ocurren de va-

FIGURA 1.4 CUESTIONARIO PARA CONOCER LA OPINIÓN DE LOS CLIENTES EMPLEADO EN EL RESTAURANTE THE LOBSTER POT DE REDINGTON SHORES, FLORIDA

TheLOBSTERPot

RESTAURANT

Nos alegramos de su visita al restaurante Lobster Pot y queremos estar seguros de que volverá. De manera que si tiene unos minutos le agradeceríamos mucho que nos llenara esta tarjeta. Sus comentarios y sugerencias son extremadamente importantes para nosotros. Gracias.

Nombre de la persona que lo atendió _____

	Excelente	Bueno	Satisfactorio	Insatisfactorio
Calidad de los alimentos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Amabilidad en el servicio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Prontitud en el servicio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Limpieza	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gestión	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Comentarios	_____			
¿Qué lo motivó a visitarnos?	_____			

Favor de depositarlo en el buzón de sugerencias que se encuentra a la entrada.

rias maneras. Por ejemplo, un entrevistador puede cometer un error de escritura, como una transposición al escribir la edad de una persona y en lugar de 24 años escribir 42 años, o en una entrevista, el entrevistado puede malinterpretar una pregunta y dar una respuesta incorrecta.

Los analistas de datos con experiencia tienen sumo cuidado tanto al recolectar los datos como al registrarlos para garantizar que no se cometan errores. Para comprobar la consistencia interna de los datos se emplean procedimientos especiales. Tales procedimientos indican al analista, por ejemplo, que debe revisar la consistencia de los datos cuando un entrevistado aparece con 22 años de edad pero informa tener 20 años de experiencia en el trabajo. El analista de datos también debe revisar datos que tengan valores inusualmente grande o pequeños, llamados observaciones atípicas, que son candidatos a posibles errores en los datos. En el capítulo 3 se muestran algunos de los métodos estadísticos útiles para identificar observaciones atípicas.

Los errores suelen presentarse durante la adquisición de datos. Emplear a ciegas cualquier dato que se tenga o valerse de datos que fueron adquiridos con poco cuidado da como resultado información desorientadora y malas decisiones. Así, tomar medidas para adquirir datos precisos ayuda a garantizar información confiable y valiosa para la toma de decisiones.

1.4

Estadística descriptiva

La mayor parte de la información estadística en periódicos, revistas, informes de empresas y otras publicaciones consta de datos que se resumen y presentan en una forma fácil de leer y de entender. A estos resúmenes de datos, que pueden ser tabulares, gráficos o numéricos se les conoce como **estadística descriptiva**.

TABLA 1.4 FRECUENCIAS Y FRECUENCIAS PORCENTUALES DE LA VARIABLE BOLSA DE VALORES

Bolsa de valores	Frecuencia	Frecuencia porcentual
Bolsa de Nueva York	20	80
Mercado Nacional Nasdaq	5	20
Totales	25	100

Vuelva al conjunto de datos de la tabla 1.1 que presenta 25 de las empresas de S&P 500. Los métodos de la estadística descriptiva pueden emplearse para resumir la información en este conjunto de datos. Por ejemplo, en la tabla 1.4 se presenta un resumen tabular de los datos de la variable bolsa de valores. Un resumen gráfico de los mismos datos, al que se le llama gráfica de barras aparece en la figura 1.5. Estos tipos de resúmenes, tabular y gráfico, permiten que los datos sean más fáciles de interpretar. Al revisar la tabla 1.4 y la figura 1.5 es fácil entender que la mayor parte de las acciones del conjunto de datos cotizan en la bolsa de Nueva York. Si emplea porcentajes: 80% cotizan en la bolsa de Nueva York y 20% en el Nasdaq.

En la figura 1.6 se presenta un resumen gráfico, llamado histograma, de los datos de la variable cuantitativa precio por acción. El histograma facilita ver que los precios por acción van de \$0 a \$100, con una mayor concentración entre \$20 y \$60.

Además de las presentaciones tabular y gráfica para resumir datos se emplea también la estadística descriptiva numérica. El estadístico descriptivo más común para resumir datos es el promedio o media. Mediante los datos de la variable ganancia por acción de las acciones S&P de la tabla 1.1, el promedio se calcula sumando las ganancias por acción de las 25 acciones y dividién-

FIGURA 1.5 GRÁFICA DE BARRAS DE LA VARIABLE BOLSA DE VALORES

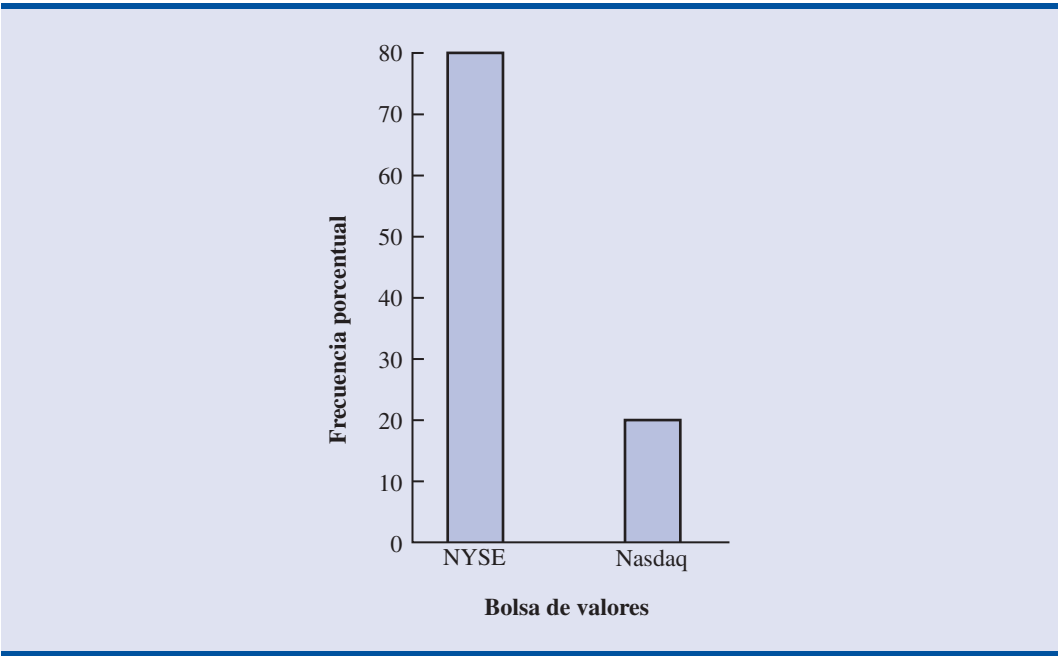
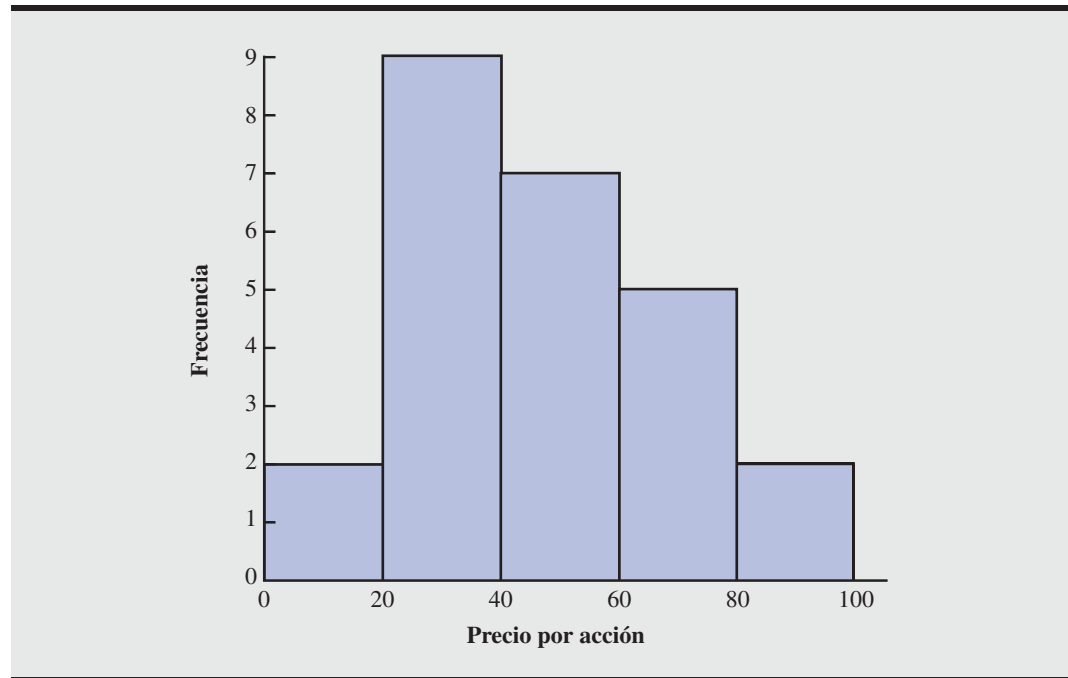


FIGURA 1.6 HISTOGRAMA DE LOS PRECIOS POR ACCIÓN DE 25 ACCIONES S&P

do entre 25. Al hacer esto se obtiene como ganancia promedio por acción \$2.49. Este promedio da una tendencia central, o posición central, de los datos de la variable.

En numerosos campos sigue creciendo el interés por los métodos estadísticos que son aplicables para elaborar y presentar estadísticas descriptivas. En los capítulos 2 y 3 se dedica la atención a los métodos tabulares, gráficos y numéricos de la estadística descriptiva.

1.5

Inferencia estadística

En muchas situaciones se requiere información acerca de grupos grandes de elementos (individuos, empresas, votantes, hogares, productos, clientes, etc.). Pero, debido al tiempo, costo y a otras consideraciones, sólo es posible recolectar los datos de una pequeña parte de este grupo. Al grupo grande de elementos en un determinado estudio se le llama **población** y al grupo pequeño **muestra**. En términos formales se emplean las definiciones siguientes.

POBLACIÓN

La población es el conjunto de todos los elementos de interés en un estudio determinado.

MUESTRA

La muestra es un subconjunto de la población.

El gobierno de Estados Unidos realiza un censo cada 10 años. Las empresas de investigación de mercado realizan estudios muestrales cada día.

Al proceso de realizar un estudio para recolectar datos de toda una población se le llama **censo**. Al proceso de efectuar un estudio para recolectar datos de una muestra se le llama **encuesta muestral**. Una de las principales contribuciones de la estadística es emplear datos de una muestra para hacer estimaciones y probar hipótesis acerca de las características de una población mediante un proceso al que se le conoce como **inferencia estadística**.

Como un ejemplo de inferencia estadística, considere un estudio realizado por Norris Electronics. Norris fabrica focos de alta intensidad que se emplean en diversos productos electrónicos. Con objeto de incrementar la vida útil de estos focos, el grupo de diseño del producto elaboró un filamento nuevo. En este caso, la población está definida por todos los focos que se produzcan con el filamento nuevo. Para evaluar las ventajas del filamento, se fabricaron 200 focos. Los datos recolectados de esta muestra dan el número de horas que duró cada foco hasta que se quemara el filamento. Véase la tabla 1.5.

Suponga que Norris desea usar estos datos muestrales para hacer una inferencia acerca del número de horas promedio de vida útil de todos los focos que se producen con el filamento nuevo. Al sumar los 200 valores de la tabla 1.5 y dividir la suma entre 200 se obtiene el promedio del tiempo de vida de los focos: 76 horas. Este resultado muestral sirve para estimar que el tiempo de vida promedio de los focos de la población es 76 horas. En la figura 1.7 se proporciona un resumen gráfico del proceso de inferencia estadística empleado por Norris Electronics.

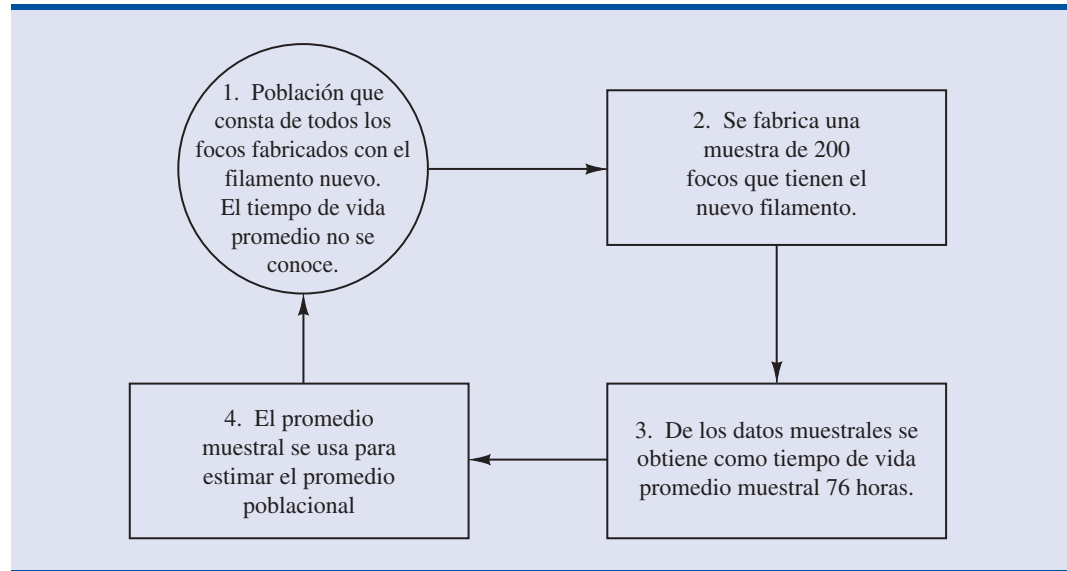
Siempre que un estadístico usa una muestra para estimar una característica poblacional que interesa, suele proporcionar información acerca de la calidad o precisión de la estimación. En el ejemplo de Norris, el estadístico puede informar que la estimación puntual del tiempo de vida promedio de la población de los nuevos focos es 76 horas con un margen de error de ± 4 horas. Entonces, el intervalo de estimación del tiempo de vida promedio de los focos fabricados con el nuevo filamento es de 72 a 80 horas. El estadístico también puede informar qué tan confiado está de que el intervalo de 72 a 80 horas contenga el promedio poblacional.

TABLA 1.5 HORAS DE DURACIÓN DE UNA MUESTRA DE 200 FOCOS DE NORRIS

107	73	68	97	76	79	94	59	98	57
54	65	71	70	84	88	62	61	79	98
66	62	79	86	68	74	61	82	65	98
62	116	65	88	64	79	78	79	77	86
74	85	73	80	68	78	89	72	58	69
92	78	88	77	103	88	63	68	88	81
75	90	62	89	71	71	74	70	74	70
65	81	75	62	94	71	85	84	83	63
81	62	79	83	93	61	65	62	92	65
83	70	70	81	77	72	84	67	59	58
78	66	66	94	77	63	66	75	68	76
90	78	71	101	78	43	59	67	61	71
96	75	64	76	72	77	74	65	82	86
66	86	96	89	81	71	85	99	59	92
68	72	77	60	87	84	75	77	51	45
85	67	87	80	84	93	69	76	89	75
83	68	72	67	92	89	82	96	77	102
74	91	76	83	66	68	61	73	72	76
73	77	79	94	63	59	62	71	81	65
73	63	63	89	82	64	85	92	64	73



FIGURA 1.7 PROCESO DE INFERENCIA ESTADÍSTICA EMPLEADO EN EL EJEMPLO DE NORRIS ELECTRONICS



1.6

Las computadoras y el análisis estadístico

Como en el análisis estadístico suelen emplearse grandes cantidades de datos, los analistas usan software para realizar estos trabajos. Por ejemplo, calcular el tiempo de vida promedio de los 200 focos del ejemplo de Norris Electronics (véase tabla 1.5) resultaría muy tedioso si no se contara con una computadora. Para facilitar el uso de una computadora, los conjuntos de datos de este libro se proporcionan en el disco compacto que viene con el libro. Un logotipo al margen izquierdo del texto identifica a estos conjuntos de datos. Los archivos de datos se encuentran en formatos para Minitab y para Excel. Además, en los apéndices de los capítulos aparecen las instrucciones para llevar a cabo los procedimientos estadísticos usando Minitab y Excel.

Resumen

La estadística es el arte y la ciencia de recolectar, analizar, presentar e interpretar datos. Casi todos los estudiantes de áreas relacionadas con los negocios o la economía necesitan tomar un curso de estadística. Este libro empezó describiendo las aplicaciones típicas de la estadística a los negocios y a la economía.

Los datos consisten en hechos/informaciones y cifras que se recolectan y analizan. Las cuatro escalas de medición que se usan para obtener datos sobre una determinada variable son nominal, ordinal, de intervalo y de razón. La escala de medición para una variable es nominal cuando los datos son etiquetas o nombres que se usan para identificar un atributo de un elemento. La escala es ordinal si los datos presentan las propiedades de los datos nominales y tiene sentido hablar del orden o jerarquía de los datos. La escala es de intervalo si los datos presentan las propiedades de los datos ordinales y los intervalos entre valores se expresan en términos de una unidad fija de medición. Por último, la escala de medición es de razón si los datos presentan las propiedades de los datos de intervalo y tiene sentido hablar de la razón entre dos valores.

Para los propósitos del análisis estadístico, los datos son clasificables en cuantitativos y cualitativos. Los datos cualitativos emplean etiquetas o nombres para identificar un atributo en cada elemento. Los datos cualitativos emplean las escalas de medición nominal u ordinal y pueden ser no numéricos o numéricos. Los datos cuantitativos son valores numéricos que indican cuánto o cuántos. Los datos cuantitativos emplean las escalas de medición de intervalo o de razón. Las operaciones aritméticas usuales sólo tienen sentido si los datos son cuantitativos. Por tanto, los cálculos estadísticos usados para datos cuantitativos no siempre son apropiados para datos cualitativos.

En las secciones 1.4 y 1.5 se introdujeron los temas de estadística descriptiva e inferencia estadística. Estadística descriptiva son los métodos tabulares, gráficos o numéricos que se usan para resumir datos. El proceso de la inferencia estadística emplea los datos obtenidos de una muestra para hacer estimaciones o probar hipótesis acerca de las características de la población. En la última sección del capítulo se indicó que las computadoras facilitan el análisis estadístico. Los conjuntos de datos grandes en los archivos de Minitab o de Excel se encuentran en el disco compacto que va con el libro.

Glosario

Estadística El arte y la ciencia de recolectar, analizar, presentar e interpretar datos.

Datos Los hechos y las cifras que se recolectan, analizan y resumen para su presentación e interpretación.

Conjunto de datos Todos los datos recolectados en un estudio determinado.

Elementos Entidades sobre las que se recolectan los datos.

Variable Una característica que interesa de un elemento.

Observación El conjunto de mediciones obtenidas de un elemento determinado.

Escala nominal Escala de medición de una variable cuando los datos son etiquetas o nombres que se emplean para identificar un atributo de un elemento. Los datos nominales pueden ser no numéricos o numéricos.

Escala ordinal Escala de medición de una variable cuando los datos presentan las propiedades de los datos nominales y el orden o jerarquía de los datos tiene sentido. Los datos ordinales pueden ser no numéricos o numéricos.

Escala de intervalo Escala de medición de una variable cuando los datos presentan las propiedades de los datos ordinales y los intervalos entre valores se expresan en términos de una unidad o medida fija. Los datos de intervalo siempre son numéricos.

Escala de razón Escala de medición de una variable cuando los datos presentan todas las propiedades de los datos de intervalo y la razón entre dos valores tiene sentido. Los datos de razón siempre son numéricos.

Datos cualitativos Etiquetas o nombres utilizados para identificar un atributo de cada elemento. Los datos cualitativos usan las escalas de medición nominal y ordinal y pueden ser no numéricos o numéricos.

Datos cuantitativos Valores numéricos que indican cuánto o cuántos de algo. Los datos cuantitativos se obtienen mediante la escala de intervalo o de razón.

Variable cualitativa Una variable con datos cualitativos.

Variable cuantitativa Una variable con datos cuantitativos.

Datos de sección transversal Datos recolectados en el mismo o aproximadamente en el mismo momento.

Datos de series de tiempo Datos recolectados a lo largo de varios periodos de tiempo.

Estadística descriptiva Resúmenes tabulares, gráficos o numéricos de datos.

Población Conjunto de todos los elementos que interesan en un estudio determinado.

Muestra Un subconjunto de la población.

Censo Un estudio para recolectar los datos de toda la población.

Encuesta muestral Un estudio para recolectar los datos de una muestra.

Inferencia estadística El proceso de emplear los datos obtenidos de una muestra para hacer estimaciones o probar hipótesis acerca de las características de la población.

Autoexamen

Autoexamen

1. Describa la diferencia entre estadística como dato numérico y estadística como disciplina o campo de estudio.
2. La revista *Condé Nast Traveler* realiza una encuesta anual entre sus suscriptores con objeto de determinar los mejores alojamientos del mundo. En la tabla 1.6 se presenta una muestra de nueve hoteles europeos (*Condé Nast Traveler*, enero de 2000). Los precios de una habitación doble estándar van de \$(precio más bajo) a \$\$\$\$ (precio más alto). La calificación general corresponde a la evaluación de habitaciones, servicio, restaurante, ubicación/atmósfera y áreas públicas; cuanto más alta sea la calificación general, mayor es el nivel de satisfacción.
 - a. ¿Cuántos elementos hay en este conjunto de datos?
 - b. ¿Cuántas variables hay en este conjunto de datos?
 - c. ¿Cuáles variables son cualitativas y cuáles cuantitativas?
 - d. ¿Qué tipo de escala de medición se usa para cada variable?
3. Vaya a la tabla 1.6.
 - a. ¿Cuál es el número promedio de habitaciones en los nueve hoteles?
 - b. Calcule la calificación general promedio.
 - c. ¿Qué porcentaje de los hoteles se encuentra en Inglaterra?
 - d. ¿En qué porcentaje de los hoteles el precio de la habitación es de \$\$?
4. Los equipos de sonido todo en uno, llamados minicomponentes, cuentan con sintonizador AM/FM, casetera doble, cargador para un disco compacto con bocinas separadas. En la tabla 1.7 se muestran los precios de menudeo, calidad de sonido, capacidad para discos compactos, sensibilidad y selectividad de la sintonización y cantidad de caseteras en los artículos de una muestra de 10 minicomponentes (*Consumer Report Buying Guide 2002*).
 - a. ¿Cuántos elementos contiene este conjunto de datos?
 - b. ¿Cuál es la población?
 - c. Calcule el precio promedio en la muestra.
 - d. Con los resultados del inciso c, estime el precio promedio para la población.
5. Considere el conjunto de datos de la muestra de los 10 minicomponentes que se muestra en la tabla 1.7.
 - a. ¿Cuántas variables hay en este conjunto de datos?
 - b. De estas variables, ¿cuáles son cualitativas y cuáles son cuantitativas?
 - c. ¿Cuál es la capacidad promedio de CD en la muestra?
 - d. ¿Qué porcentaje de los minicomponentes tienen una sintonización de FM buena o excelente?
 - e. ¿Qué porcentaje de los minicomponentes tienen dos caseteras?

TABLA 1.6 CALIFICACIONES PARA NUEVE LUGARES DONDE ALOJARSE EN EUROPA

Nombre del lugar	País	Precio de la habitación	Número de habitaciones	Calificación general
Graveteye Manor	Inglaterra	\$\$	18	83.6
Villa d'Este	Italia	\$\$\$\$	166	86.3
Hotel Prem	Alemania	\$	54	77.8
Hotel d'Europe	Francia	\$\$	47	76.8
Palace Luzern	Suiza	\$\$	326	80.9
Royal Crescent Hotel	Inglaterra	\$\$\$	45	73.7
Hotel Sacher	Austria	\$\$\$	120	85.5
Duc de Bourgogne	Bélgica	\$	10	76.9
Villa Gallici	Francia	\$\$	22	90.6

Fuente: *Condé Nast Traveler*, enero de 2000.

TABLA 1.7 UNA MUESTRA DE 10 MINICOMPONENTES

Marca y modelo	Precio (\$)	Calidad de sonido	Capacidad para CD	Sintonización FM	Caseteras
Aiwa NSX-AJ800	250	Buena	3	Regular	2
JVC FS-SD1000	500	Buena	1	Muy buena	0
JVC MX-G50	200	Muy buena	3	Excelente	2
Panasonic SC-PM11	170	Regular	5	Muy buena	1
RCA RS 1283	170	Buena	3	Mala	0
Sharp CD-BA2600	150	Buena	3	Buena	2
Sony CHC-CL1	300	Muy buena	3	Muy buena	1
Sony MHC-NX1	500	Buena	5	Excelente	2
Yamaha GX-505	400	Muy buena	3	Excelente	1
Yamaha MCR-E100	500	Muy buena	1	Excelente	0



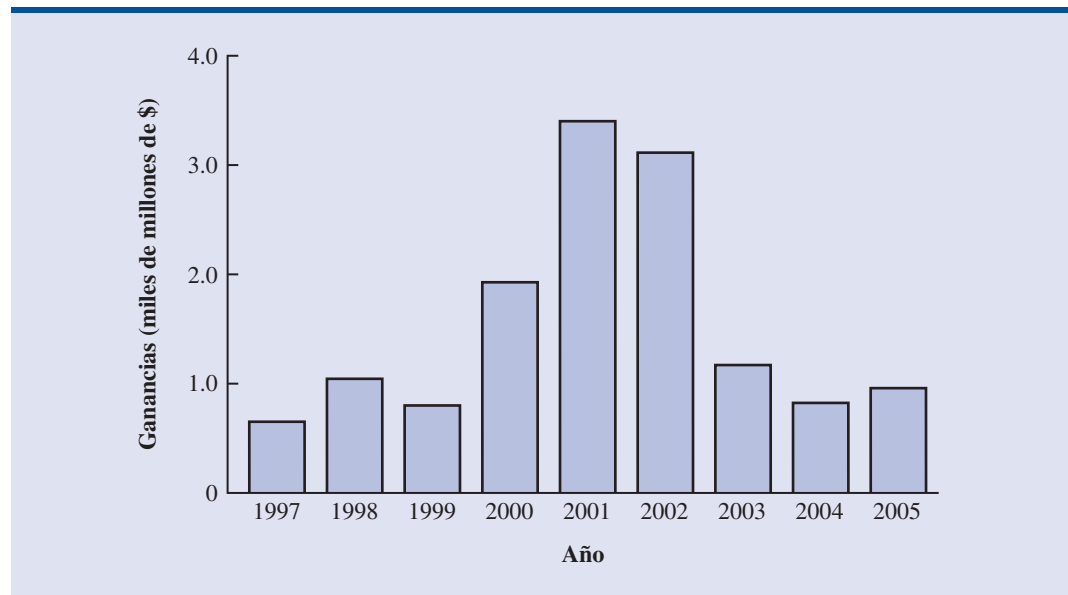
6. La Columbia House vende discos compactos a los miembros de su club de venta por correo. En una encuesta sobre música se les pidió a los nuevos miembros del club que llenaran un cuestionario con 11 preguntas. Algunas de las preguntas eran:
 - a. ¿Cuántos discos compactos has comprado en los últimos 12 meses?
 - b. ¿Eres miembro de algún club de venta de libros por correo (Sí o No)?
 - c. ¿Cuál es tu edad?
 - d. Incluyéndote a ti, de cuántas personas (adultos y niños) consta tu familia.
 - e. ¿Qué tipo de música te interesa comprar? Se presentaban quince categorías entre las que se encontraban rock pesado, rock ligero, música contemporánea para adultos, rap y rancheras. Responde si los datos que se obtienen con cada pregunta son cualitativos o cuantitativos.
7. El hotel Ritz Carlton emplea un cuestionario de opinión del cliente para obtener datos sobre la calidad de sus servicios de restaurante y entretenimiento (The Ritz-Carlton Hotel, Naples, Florida, febrero de 2006). Se les pidió a los clientes que evaluaran seis puntos: recibimiento, servicio, alimentos, menú, atención y atmósfera. Los datos registrados para cada factor fueron 1 para Pasadero, 2 Regular, 3 Bueno y 4 Excelente.
 - a. Las respuestas de los clientes proporcionan datos para seis variables. ¿Son estas variables cualitativas o cuantitativas?
 - b. ¿Qué escala de medición se usa?
8. La empresa Gallup realizó una encuesta telefónica empleando una muestra aleatoria nacional compuesta de 1005 adultos de 18 años o más. En la encuesta se les preguntó a los participantes “Cómo considera que es su salud física en este momento” (www.gallup.com, 7 de febrero de 2002). Las respuestas podían ser Excelente, Buena, Regular o Ninguna opinión.
 - a. ¿Cuál es el tamaño de la muestra de esta investigación?
 - b. ¿Son estos datos cualitativos o cuantitativos?
 - c. ¿Sería conveniente usar promedios o porcentajes para resumir los datos de estas preguntas?
 - d. De las personas que respondieron, 29% dijo que su salud era excelente. ¿Cuántos fueron los individuos que dieron esta respuesta?
9. El Departamento de Comercio informa haber recibido las siguientes solicitudes para concursar por el Malcolm Baldrige National Quality Award: 23 de empresas fabricantes grandes, 18 de empresas grandes de servicios y 30 de negocios pequeños.
 - a. ¿Es el tipo de empresa una variable cualitativa o cuantitativa?
 - b. ¿Qué porcentaje de las solicitudes venían de negocios pequeños?
10. En una encuesta de *The Wall Street Journal* (13 de octubre de 2003) se les hacen a los suscriptores 46 preguntas acerca de sus características e intereses. De cada una de las preguntas si-

guientes, ¿cuál proporciona datos cualitativos o cuantitativos e indica la escala de medición apropiada?

- a. ¿Cuál es su edad?
 - b. ¿Es usted hombre o mujer?
 - c. ¿Cuándo empezó a leer el *WSJ*? Preparatoria, universidad al comienzo de la carrera, a la mitad de la carrera, al final de la carrera o ya retirado.
 - d. ¿Cuánto tiempo hace que tiene su trabajo o cargo actual?
 - e. ¿Qué tipo de automóvil piensa comprarse la próxima vez que compre uno? Ocho categorías para las respuestas, entre las que se encontraban sedán, automóvil deportivo, miniván, etcétera.
11. Diga de cada una de las variables siguientes si es cualitativa o cuantitativa e indique la escala de medición a la que pertenece.
 - a. Ventas anuales.
 - b. Tamaño de los refrescos (pequeño, mediano, grande).
 - c. Clasificación como empleado (GS 1 a GS 18).
 - d. Ganancia por acción.
 - e. Modo de pago (al contado, cheque, tarjeta de crédito).
 12. La Oficina de Visitantes a Hawai recolecta datos de los visitantes. Entre las 16 preguntas hechas a los pasajeros de un vuelo de llegada en junio de 2003 estaban las siguientes.
 - Este viaje a Hawai es mi 1o., 2o., 3o., 4o. etc.
 - La principal razón de este viaje es: (10 categorías para escoger entre las que se encontraban vacaciones, luna de miel, una convención).
 - Dónde voy a alojarme: (11 categorías entre las que se encontraban hotel, departamento, parientes, acampar).
 - Total de días en Hawai
 - a. ¿Cuál es la población que se estudia?
 - b. ¿El uso de un cuestionario es una buena manera de tener información de los pasajeros en los vuelos de llegada?
 - c. Diga de cada una de las cuatro preguntas si los datos que suministra son cualitativos o cuantitativos.
 13. En la figura 1.8 se presenta una gráfica de barras que resume las ganancias de Volkswagen de los años 1997 a 2005 (*BusinessWeek*, 26 de diciembre de 2005).



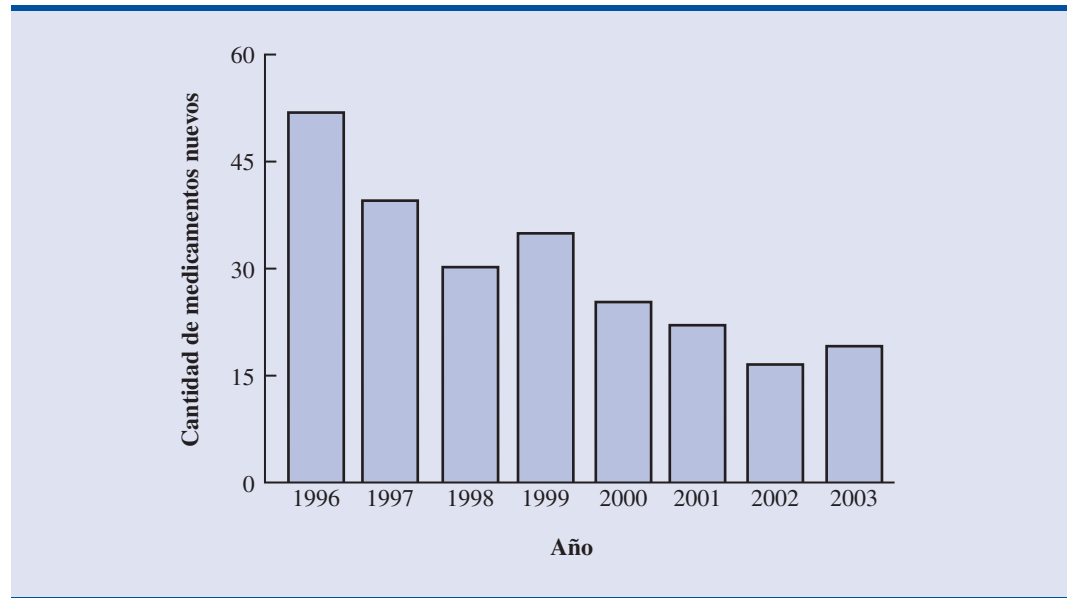
FIGURA 1.8 GANANCIAS DE VOLKSWAGEN



- a. ¿Estos son datos cualitativos o cuantitativos?
 - b. ¿Son datos de series de tiempo o datos de sección transversal?
 - c. ¿Cuál es la variable de interés?
 - d. Comente la tendencia en las ganancias de Volkswagen a lo largo del tiempo. El artículo de *BusinessWeek* (26 de diciembre de 2005) estimó las ganancias en 2006 en \$600 millones o \$0.6 mil millones. ¿Indica la figura si esta estimación parece ser razonable?
 - e. Un artículo similar que apareció en *BusinessWeek* el 23 de julio de 2001 sólo contaba con los datos de 1997 a 2000 junto con elevadas ganancias proyectadas para 2001. ¿Cómo era la perspectiva de las ganancias de Volkswagen en julio de 2001? En 2001, ¿parecía promotor invertir en Volkswagen? Explique.
 - f. ¿Qué advertencia sugiere esta gráfica acerca de la proyección de datos como los de las ganancias de Volkswagen hacia el futuro?
14. CSM Worldwide pronostica la producción mundial de todos los fabricantes de automóviles. Los datos siguientes de CSM muestran el pronóstico de la producción mundial para General Motors, Ford, DaimlerChrysler y Toyota para los años 2004 a 2007 (*USA Today*, 21 de diciembre de 2005). Estos datos están dados en millones de vehículos.

Fabricante	2004	2005	2006	2007
General Motors	8.9	9.0	8.9	8.8
Ford	7.8	7.7	7.8	7.9
DaimlerChrysler	4.1	4.2	4.3	4.6
Toyota	7.8	8.3	9.1	9.6

- a. Haga una gráfica de series de tiempo para los años 2004 a 2007 en la que se observe la cantidad de vehículos fabricados por cada empresa. Muestre las series de tiempo de los cuatro fabricantes en la misma gráfica.
 - b. General Motors ha sido sin discusión el principal fabricante de automóviles desde 1931. En esta gráfica de series de tiempo, ¿cuál es el mayor fabricante de automóviles? Explique.
 - c. Haga una gráfica que muestre los vehículos producidos por los fabricantes de automóviles usando los datos de 2007. ¿Está basada en datos de series de tiempo o en datos de sección transversal?
15. La Food and Drug Administration (FDA) da información sobre la cantidad de medicamentos aprobados en un periodo de ocho años (*The Wall Street Journal*, 12 de enero de 2004). En la figura 1.9 se presenta una gráfica de barras que resume el número de medicamentos nuevos aprobados cada año.
- a. ¿Estos datos son cualitativos o cuantitativos?
 - b. ¿Son datos de series de tiempo o son datos de sección transversal?
 - c. ¿Cuántos medicamentos fueron aprobados en 2003?
 - d. ¿En qué año se aprobaron menos medicamentos? ¿Cuántos fueron?
 - e. Presente un comentario sobre la tendencia en el número de medicamentos nuevos aprobados por la FDA en este periodo de ocho años.
16. El departamento de marketing de su empresa elabora un refresco dietético que dice captará una gran parte del mercado de adultos jóvenes.
- a. ¿Qué datos desearía ver antes de invertir una cantidad importante para introducir el nuevo producto en el mercado?
 - b. ¿Cómo esperaría que se obtuvieran los datos mencionados en el inciso a?
17. El directivo de una empresa grande recomienda un aumento de \$10 000 para evitar que un empleado se cambie a otra empresa. ¿Qué fuentes de datos internas y externas pueden usarse para decidir si es apropiado ese incremento de salario?

FIGURA 1.9 NÚMERO DE MEDICAMENTOS NUEVOS APROBADOS POR LA FDA

18. En una encuesta a 430 viajeros de negocios se encontró que 155 de ellos empleaban los servicios de un agente de viajes para la preparación de sus viajes (*USA Today*, 20 de noviembre de 2003).
 - a. Elabore una estadística descriptiva que sirva para estimar el porcentaje de viajeros de negocios que emplean un agente de viajes para preparar su viaje.
 - b. Con la encuesta se encontró que la manera más frecuente en que los viajeros de negocios hacen los preparativos de su viaje es mediante un sitio en línea. Si 4% de los viajeros de negocios encuestados hacen los preparativos de su viaje de esta manera, ¿cuántos de los 430 encuestados emplearon un sitio en línea?
 - c. Estos datos sobre cómo se hacen los preparativos, ¿son cualitativos o cuantitativos?
19. En un estudio sobre los suscriptores de *BusinessWeek* de Estados Unidos se recogen datos de una muestra de 2861 suscriptores. Cincuenta y nueve por ciento de los encuestados señalaron tener un ingreso de \$75 000 o más y 50% indicaron poseer una tarjeta de crédito de American Express.
 - a. ¿Cuál es la población de interés en este estudio?
 - b. ¿Es el ingreso anual un dato cualitativo o cuantitativo?
 - c. ¿Es la posesión de una tarjeta de crédito de American Express una variable cualitativa o cuantitativa?
 - d. ¿Hacer este estudio requiere datos de series de tiempo o de sección transversal?
 - e. Describa cualquier inferencia estadística posible para *BusinessWeek* con base en esta encuesta.
20. En una encuesta a 131 directores de inversión en Barron's se encontró lo siguiente (Barron's 28 de octubre de 2002):
 - De los dirigentes 43% se clasificaron como optimistas o muy optimistas sobre el mercado de acciones.
 - El rendimiento promedio esperado en los 12 meses siguientes en títulos de capital fue 11.2%.
 - La atención a la salud fue elegida por 21% como el sector con más probabilidad de ir a la cabeza del mercado en los próximos 12 meses.
 - Cuando se les preguntó cuánto tiempo se necesitaría para que las acciones de tecnología y telecomunicación recobraran un crecimiento sostenible, la respuesta promedio de los directivos fue 2.5 años.

- a. Cite dos estadísticas descriptivas.
 - b. Haga una inferencia sobre la población de todos los directivos de inversiones respecto al rendimiento promedio esperado en los títulos de capital durante los siguientes 12 meses.
 - c. Haga una inferencia acerca de la cantidad de tiempo que se necesitará para que las acciones de tecnología y telecomunicación recobren un crecimiento sostenible.
21. En una investigación médica que duró siete años se encontró que las mujeres cuyas madres habían tomado el medicamento DES durante el embarazo, tenían el doble de posibilidades de presentar anomalías en los tejidos que pudieran conducir a un cáncer, que aquellas cuyas madres no habían tomado este medicamento.
 - a. En este estudio se compararon dos poblaciones. ¿Cuáles son?
 - b. ¿Es posible pensar que los datos se obtuvieron mediante una encuesta o mediante un experimento?
 - c. De la población de las mujeres cuyas madres habían tomado el medicamento DES durante el embarazo, se encontró que en una muestra de 3980 mujeres 63 presentaban anomalías en tejidos que podrían conducir a un cáncer. Dé un estadístico descriptivo útil para estimar el número de mujeres, de cada 1000, de esta población que pueden presentar anomalías en los tejidos.
 - d. De la población de mujeres cuyas madres no tomaron el medicamento DES durante el embarazo, ¿cuál es el número estimado de mujeres, de cada 1000, que pueden presentar anomalías en los tejidos?
 - e. Estudios médicos a menudo utilizan muestras grandes (en este caso, 3980). ¿Por qué?
22. En otoño de 2003, Arnold Schwarzenegger disputó al gobernador Gray Davis la gobernatura de California. En una encuesta realizada entre los votantes registrados se encontró que Arnold Schwarzenegger iba a la cabeza con un porcentaje estimado de 54% (*Newsweek*, 8 de septiembre de 2003).
 - a. ¿Cuál fue la población en este estudio?
 - b. ¿Cuál fue la muestra en este estudio?
 - c. ¿Por qué se empleó una muestra en esta situación? Explique.
23. Nielsen Media Research realiza cada semana un sondeo entre los televidentes de Estados Unidos y publica datos tanto de índice de audiencia como de participación en el mercado. El índice de audiencia de Nielsen es el porcentaje de hogares que tienen televisión y que están viendo un programa, mientras que la participación de Nielsen es el porcentaje de hogares que están viendo un programa, entre los hogares que tiene la televisión en uso. Por ejemplo, los resultados de Nielsen Media Research para la Serie Mundial de Béisbol de 2003 entre los Yankees de Nueva York y los Marlins de Florida dieron un índice de audiencia de 12.8% y una participación de 22% (*Associated Press*, 27 de octubre de 2003). Por tanto, 12.8% de los hogares que tenían televisión estaban viendo la Serie Mundial y 22% de los hogares que estaban viendo la televisión, estaban viendo la Serie Mundial. A partir de los datos de índices de audiencia y de participación, Nielsen publica un ranking semanal de los programas de televisión así como un ranking semanal de las cuatro principales cadenas de televisión en Estados Unidos: ABC, CBS, NBC y Fox.
 - a. ¿Qué trata de medir Nielsen Media Research?
 - b. ¿Cuál es la población?
 - c. ¿Por qué se usaría una muestra en esta situación?
 - d. ¿Qué tipo de decisiones o de acciones están basadas en los rankings de Nielsen?
24. En una muestra con cinco calificaciones de los estudiantes en un determinado examen los datos fueron: 72, 65, 82, 90, 76. ¿Cuáles de las afirmaciones siguientes son correctas y cuáles deben cuestionarse como una generalización excesiva?
 - a. La calificación promedio de este examen en la muestra de las calificaciones de cinco estudiantes es 77.
 - b. La calificación promedio de todos los estudiantes en este examen es 77.
 - c. Una estimación para la calificación promedio de todos los estudiantes que hicieron el examen es 77.
 - d. Más de la mitad de los estudiantes que hicieron el examen tendrán calificaciones entre 70 y 85.
 - e. Si se incluyen en la muestra otros cinco estudiantes, sus calificaciones estarán entre 65 y 90.

TABLA 1.8 CONJUNTO DE DATOS DE 25 ACCIONES SHADOW

Empresa	Bolsa de valores	Denominación abreviada Symbol	Capacidad de mercado (millones de \$)	Relación precio/ganancia	Margen de ganancia bruta (%)
DeWolfe Companies	AMEX	DWL	36.4	8.4	36.7
North Coast Energy	OTC	NCEB	52.5	6.2	59.3
Hansen Natural Corp.	OTC	HANS	41.1	14.6	44.8
MarineMax, Inc.	NYSE	HZO	111.5	7.2	23.8
Nanometrics Incorporated	OTC	NANO	228.6	38.0	53.3
TeamStaff, Inc.	OTC	TSTF	92.1	33.5	4.1
Environmental Tectonics	AMEX	ETC	51.1	35.8	35.9
Measurement Specialties	AMEX	MSS	101.8	26.8	37.6
SEMCO Energy, Inc.	NYSE	SEN	193.4	18.7	23.6
Party City Corporation	OTC	PCTY	97.2	15.9	36.4
Embrex, Inc.	OTC	EMBX	136.5	18.9	59.5
Tech/Ops Sevcon, Inc.	AMEX	TO	23.2	20.7	35.7
ARCADIS NV	OTC	ARCAF	173.4	8.8	9.6
Qiao Xing Universal Tele.	OTC	XING	64.3	22.1	30.8
Energy West Incorporated	OTC	EWST	29.1	9.7	16.3
Barnwell Industries, Inc.	AMEX	BRN	27.3	7.4	73.4
Innodata Corporation	OTC	INOD	66.1	11.0	29.6
Medical Action Industries	OTC	MDCI	137.1	26.9	30.6
Instrumentarium Corp.	OTC	INMRY	240.9	3.6	52.1
Petroleum Development	OTC	PETD	95.9	6.1	19.4
Drexler Technology Corp.	OTC	DRXR	233.6	45.6	53.6
Gerber Childrenswear Inc.	NYSE	GCW	126.9	7.9	25.8
Gaiam, Inc.	OTC	GAIA	295.5	68.2	60.7
Artesian Resources Corp.	OTC	ARTNA	62.8	20.5	45.5
York Water Company	OTC	YORW	92.2	22.9	74.2



25. En la tabla 1.8 aparece un conjunto de datos con información sobre 25 de las acciones shadow vigiladas por la American Association of Individual Investors (aaii.com, febrero de 2002). Acciones shadow son acciones comunes de empresas pequeñas que no son estrechamente vigiladas por los analistas de Wall Street. Este conjunto de datos se encuentra también en el disco compacto que se incluye en este libro, en el archivo Shadow02.
- ¿Cuántas variables hay en este conjunto de datos?
 - ¿Qué variables son cualitativas y cuáles son cuantitativas?
 - Par la variable bolsa de valores muestre la frecuencia y la frecuencia porcentual de AMEX, NYSE y OTC. Construya una gráfica de barras como la de la figura 1.5.
 - Muestre la distribución de frecuencias del margen de ganancia bruta empleando cinco intervalos: 0–14.9, 15–29.9, 30–44.9, 45–59.9 y 60–74.9. Construya un histograma como el de la figura 1.6.
 - ¿Cuál es la proporción precio/ganancia promedio?