About the Authors

Dennis J. Sweeney. Dennis J. Sweeney is Professor of Quantitative Analysis and Founder of the Center for Productivity Improvement at the University of Cincinnati. Born in Des Moines, Iowa, he earned a B.S.B.A. degree from Drake University and his M.B.A. and D.B.A. degrees from Indiana University, where he was an NDEA Fellow. During 1978–79, Professor Sweeney worked in the management science group at Procter & Gamble; during 1981–82, he was a visiting professor at Duke University. Professor Sweeney served as Head of the Department of Quantitative Analysis and as Associate Dean of the College of Business Administration at the University of Cincinnati.

Professor Sweeney has published more than thirty articles and monographs in the area of management science and statistics. The National Science Foundation, IBM, Procter & Gamble, Federated Department Stores, Kroger, and Cincinnati Gas & Electric have funded his research, which has been published in *Management Science*, *Operations Research*, *Mathematical Programming*, *Decision Sciences*, and other journals.

Professor Sweeney has coauthored ten textbooks in the areas of statistics, management science, linear programming, and production and operations management.

Thomas A. Williams. Thomas A. Williams is Professor of Management Science in the College of Business at Rochester Institute of Technology. Born in Elmira, New York, he earned his B.S. degree at Clarkson University. He did his graduate work at Rensselaer Polytechnic Institute, where he received his M.S. and Ph.D. degrees.

Before joining the College of Business at RIT, Professor Williams served for seven years as a faculty member in the College of Business Administration at the University of Cincinnati, where he developed the undergraduate program in Information Systems and then served as its coordinator. At RIT he was the first chairman of the Decision Sciences Department. He teaches courses in management science and statistics, as well as graduate courses in regression and decision analysis.

Professor Williams is the coauthor of eleven textbooks in the areas of management science, statistics, production and operations management, and mathematics. He has been a consultant for numerous *Fortune* 500 companies and has worked on projects ranging from the use of data analysis to the development of large-scale regression models.

David R. Anderson. David R. Anderson is Professor of Quantitative Analysis in the College of Business Administration at the University of Cincinnati. Born in Grand Forks, North Dakota, he earned his B.S., M.S., and Ph.D. degrees from Purdue University. Professor Anderson has served as Head of the Department of Quantitative Analysis and Operations Management and as Associate Dean of the College of Business Administration. In addition, he was the coordinator of the College's first Executive Program.

At the University of Cincinnati, Professor Anderson has taught introductory statistics for business students as well as graduate-level courses in regression analysis, multivariate analysis, and management science. He has also taught statistical courses at the Department of Labor in Washington, D.C. He has been honored with nominations and awards for excellence in teaching and excellence in service to student organizations.

Professor Anderson has coauthored ten textbooks in the areas of statistics, management science, linear programming, and production and operations management. He is an active consultant in the field of sampling and statistical methods.

CHAPTER 1

Data and Statistics

CONTENTS

STATISTICS *in* PRACTICE

BUSINESSWEEK*
NEW YORK, NEW YORK

With a global circulation of more than 1 million, *BusinessWeek* is the most widely read business magazine in the world. More than 200 dedicated reporters and editors in 26 bureaus worldwide deliver a variety of articles of interest to the business and economic community. Along with feature articles on current topics, the magazine contains regular sections on International Business, Economic Analysis, Information Processing, and Science & Technology. Information in the feature articles and the regular sections helps readers stay abreast of current developments and assess the impact of those developments on business and economic conditions.

Most issues of *BusinessWeek* provide an in-depth report on a topic of current interest. Often, the in-depth reports contain statistical facts and summaries that help the reader understand the business and economic information. For example, the April 24, 2006, issue included a special report on the world's most innovative companies; the December 25, 2006, issue provided advice on where to invest in 2007; and the January 8, 2007, issue contained a feature article about business travel. In addition, the weekly *BusinessWeek Investor* provides statistics about the state of the economy, including production indexes, stock prices, mutual funds, and interest rates.

BusinessWeek also uses statistics and statistical information in managing its own business. For example, an annual survey of subscribers helps the company learn about subscriber demographics, reading habits, likely purchases, lifestyles, and so on. *BusinessWeek* managers use statistical summaries from the survey to provide better services to subscribers and advertisers. One recent North American subscriber survey indicated that 90% of *BusinessWeek* subscribers use a personal computer at home and that 64% of *BusinessWeek* subscribers are involved with computer purchases at work. Such statistics alert *BusinessWeek* managers to subscriber interest in articles about new developments in computers. The results of the survey are also made available to potential advertisers. The high percentage of subscribers using personal computers at home and the high percentage of subscribers involved with computer purchases at work would be an incentive for a computer manufacturer to consider advertising in *BusinessWeek*.

BusinessWeek uses statistical facts and summaries in many of its articles. © Terri Miller/E-Visual Communications, Inc.

In this chapter, we discuss the types of data available for statistical analysis and describe how the data are obtained. We introduce descriptive statistics and statistical inference as ways of converting data into meaningful and easily interpreted statistical information.

*The authors are indebted to Charlene Trentham, Research Manager at *BusinessWeek*, for providing this Statistics in Practice.

Frequently, we see the following types of statements in newspapers and magazines:

- The National Association of Realtors reported that the median selling price for a house in the United States was $222,600 (*The Wall Street Journal,* January 2, 2007).
- The average cost of a 30-second television commercial during the 2006 Super Bowl game was $2.5 million (*USA Today,* January 27, 2006).

- A Jupiter Media survey found 31% of adult males watch television 10 or more hours a week. For adult women it was 26% (*The Wall Street Journal,* January 26, 2004).
- General Motors, a leader in automotive cash rebates, provided an average cash incentive of $4300 per vehicle (*USA Today,* January 27, 2006).
- More than 40% of Marriott International managers work their way up through the ranks (*Fortune,* January 20, 2003).
- The New York Yankees have the highest payroll in major league baseball. In 2005, the team payroll was $208,306,817, with a median of $5,833,334 per player (*USA Today* Salary Database, February 2006).
- The Dow Jones Industrial Average closed at 13,265 (*Barron's,* May 5, 2007).

The numerical facts in the preceding statements ($222,600; $2.5 million; 31%; 26%; $4300; 40%; $5,833,334; and 13,265) are called statistics. In this usage, the term *statistics* refers to numerical facts such as averages, medians, percents, and index numbers that help us understand a variety of business and economic conditions. However, as you will see, the field, or subject, of statistics involves much more than numerical facts. In a broader sense, **statistics** is defined as the art and science of collecting, analyzing, presenting, and interpreting data. Particularly in business and economics, the information provided by collecting, analyzing, presenting, and interpreting data gives managers and decision makers a better understanding of the business and economic environment and thus enables them to make more informed and better decisions. In this text, we emphasize the use of statistics for business and economic decision making.

Chapter 1 begins with some illustrations of the applications of statistics in business and economics. In Section 1.2 we define the term *data* and introduce the concept of a data set. This section also introduces key terms such as *variables* and *observations,* discusses the difference between quantitative and qualitative data, and illustrates the uses of cross-sectional and time series data. Section 1.3 discusses how data can be obtained from existing sources or through surveys and experimental studies designed to obtain new data. The important role that the Internet now plays in obtaining data is also highlighted. The uses of data in developing descriptive statistics and in making statistical inferences are described in Sections 1.4 and 1.5.

1.1 Applications in Business and Economics

In today's global business and economic environment, anyone can access vast amounts of statistical information. The most successful managers and decision makers understand the information and know how to use it effectively. In this section, we provide examples that illustrate some of the uses of statistics in business and economics.

Accounting

Public accounting firms use statistical sampling procedures when conducting audits for their clients. For instance, suppose an accounting firm wants to determine whether the amount of accounts receivable shown on a client's balance sheet fairly represents the actual amount of accounts receivable. Usually the large number of individual accounts receivable makes reviewing and validating every account too time-consuming and expensive. As common practice in such situations, the audit staff selects a subset of the accounts called a sample. After reviewing the accuracy of the sampled accounts, the auditors draw a conclusion as to whether the accounts receivable amount shown on the client's balance sheet is acceptable.

Finance

Financial analysts use a variety of statistical information to guide their investment recommendations. In the case of stocks, the analysts review a variety of financial data including price/earnings ratios and dividend yields. By comparing the information for an individual stock with information about the stock market averages, a financial analyst can begin to draw a conclusion as to whether an individual stock is over- or underpriced. For example, *Barron's* (September 12, 2005) reported that the average price/earnings ratio for the 30 stocks in the Dow Jones Industrial Average was 16.5. JPMorgan showed a price/earnings ratio of 11.8. In this case, the statistical information on price/earnings ratios indicated a lower price in comparison to earnings for JPMorgan than the average for the Dow Jones stocks. Therefore, a financial analyst might conclude that JPMorgan was underpriced. This and other information about JPMorgan would help the analyst make a buy, sell, or hold recommendation for the stock.

Marketing

Electronic scanners at retail checkout counters collect data for a variety of marketing research applications. For example, data suppliers such as ACNielsen and Information Resources, Inc., purchase point-of-sale scanner data from grocery stores, process the data, and then sell statistical summaries of the data to manufacturers. Manufacturers spend hundreds of thousands of dollars per product category to obtain this type of scanner data. Manufacturers also purchase data and statistical summaries on promotional activities such as special pricing and the use of in-store displays. Brand managers can review the scanner statistics and the promotional activity statistics to gain a better understanding of the relationship between promotional activities and sales. Such analyses often prove helpful in establishing future marketing strategies for the various products.

Production

Today's emphasis on quality makes quality control an important application of statistics in production. A variety of statistical quality control charts are used to monitor the output of a production process. In particular, an x-bar chart can be used to monitor the average output. Suppose, for example, that a machine fills containers with 12 ounces of a soft drink. Periodically, a production worker selects a sample of containers and computes the average number of ounces in the sample. This average, or x-bar value, is plotted on an x-bar chart. A plotted value above the chart's upper control limit indicates overfilling, and a plotted value below the chart's lower control limit indicates underfilling. The process is termed "in control" and allowed to continue as long as the plotted x-bar values fall between the chart's upper and lower control limits. Properly interpreted, an x-bar chart can help determine when adjustments are necessary to correct a production process.

Economics

Economists frequently provide forecasts about the future of the economy or some aspect of it. They use a variety of statistical information in making such forecasts. For instance, in forecasting inflation rates, economists use statistical information on such indicators as the Producer Price Index, the unemployment rate, and manufacturing capacity utilization. Often these statistical indicators are entered into computerized forecasting models that predict inflation rates.

Applications of statistics such as those described in this section are an integral part of this text. Such examples provide an overview of the breadth of statistical applications. To supplement these examples, practitioners in the fields of business and economics provided chapter-opening Statistics in Practice articles that introduce the material covered in each chapter. The Statistics in Practice applications show the importance of statistics in a wide variety of business and economic situations.

1.2 Data

Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation. All the data collected in a particular study are referred to as the **data set** for the study. Table 1.1 shows a data set containing information for 25 companies that are part of the S&P 500. The S&P 500 is made up of 500 companies selected by Standard & Poor's. These companies account for 76% of the market capitalization of all U.S. stocks. S&P 500 stocks are closely followed by investors and Wall Street analysts.

TABLE 1.1 DATA SET FOR 25 S&P 500 COMPANIES

CD file BWS&P

Company	Exchange	Ticker	*BusinessWeek* Rank	Share Price ($)	Earnings per Share ($)
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	BMY	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

Source: BusinessWeek (April 4, 2005).

Elements, Variables, and Observations

Elements are the entities on which data are collected. For the data set in Table 1.1, each individual company's stock is an element; the element names appear in the first column. With 25 stocks, the data set contains 25 elements.

A **variable** is a characteristic of interest for the elements. The data set in Table 1.1 includes the following five variables:

- *Exchange:* Where the stock is traded—N (New York Stock Exchange) and NQ (Nasdaq National Market)
- *Ticker Symbol:* The abbreviation used to identify the stock on the exchange listing
- *BusinessWeek Rank:* A number from 1 to 500 that is a measure of company strength
- *Share Price ($):* The closing price (February 28, 2005)
- *Earnings per Share ($):* The earnings per share for the most recent 12 months

Measurements collected on each variable for every element in a study provide the data. The set of measurements obtained for a particular element is called an **observation**. Referring to Table 1.1, we see that the set of measurements for the first observation (Abbott Laboratories) is N, ABT, 90, 46, and 2.02. The set of measurements for the second observation (Altria Group) is N, MO, 148, 66, and 4.57, and so on. A data set with 25 elements contains 25 observations.

Scales of Measurement

Data collection requires one of the following scales of measurement: nominal, ordinal, interval, or ratio. The scale of measurement determines the amount of information contained in the data and indicates the most appropriate data summarization and statistical analyses.

When the data for a variable consist of labels or names used to identify an attribute of the element, the scale of measurement is considered a **nominal scale**. For example, referring to the data in Table 1.1, we see that the scale of measurement for the exchange variable is nominal because N and NQ are labels used to identify where the company's stock is traded. In cases where the scale of measurement is nominal, a numeric code as well as nonnumeric labels may be used. For example, to facilitate data collection and to prepare the data for entry into a computer database, we might use a numeric code by letting 1 denote the New York Stock Exchange and 2 denote the Nasdaq National Market. In this case the numeric values 1 and 2 provide the labels used to identify where the stock is traded. The scale of measurement is nominal even though the data appear as numeric values.

The scale of measurement for a variable is called an **ordinal scale** if the data exhibit the properties of nominal data and the order or rank of the data is meaningful. For example, Eastside Automotive sends customers a questionnaire designed to obtain data on the quality of its automotive repair service. Each customer provides a repair service rating of excellent, good, or poor. Because the data obtained are the labels—excellent, good, or poor—the data have the properties of nominal data. In addition, the data can be ranked, or ordered, with respect to the service quality. Data recorded as excellent indicate the best service, followed by good and then poor. Thus, the scale of measurement is ordinal. Note that the ordinal data can also be recorded using a numeric code. For example, the *BusinessWeek* rank for the data in Table 1.1 is ordinal data. It provides a rank from 1 to 500 based on *BusinessWeek*'s assessment of the company's strength.

The scale of measurement for a variable becomes an **interval scale** if the data show the properties of ordinal data and the interval between values is expressed in terms of a fixed

unit of measure. Interval data are always numeric. Scholastic Aptitude Test (SAT) scores are an example of interval-scaled data. For example, three students with SAT math scores of 620, 550, and 470 can be ranked or ordered in terms of best performance to poorest performance. In addition, the differences between the scores are meaningful. For instance, student 1 scored $620 - 550 = 70$ points more than student 2, while student 2 scored $550 - 470 = 80$ points more than student 3.

The scale of measurement for a variable is a **ratio scale** if the data have all the properties of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight, and time use the ratio scale of measurement. This scale requires that a zero value be included to indicate that nothing exists for the variable at the zero point. For example, consider the cost of an automobile. A zero value for the cost would indicate that the automobile has no cost and is free. In addition, if we compare the cost of $30,000 for one automobile to the cost of $15,000 for a second automobile, the ratio property shows that the first automobile is $30,000/$15,000 = 2 times, or twice, the cost of the second automobile.

Qualitative and Quantitative Data

Qualitative data are often referred to as categorical data.

Data can also be classified as either qualitative or quantitative. **Qualitative data** include labels or names used to identify an attribute of each element. Qualitative data use either the nominal or ordinal scale of measurement and may be nonnumeric or numeric. **Quantitative data** require numeric values that indicate how much or how many. Quantitative data are obtained using either the interval or ratio scale of measurement.

The statistical method appropriate for summarizing data depends upon whether the data are qualitative or quantitative.

A **qualitative variable** is a variable with qualitative data, and a **quantitative variable** is a variable with quantitative data. The statistical analysis appropriate for a particular variable depends upon whether the variable is qualitative or quantitative. If the variable is qualitative, the statistical analysis is rather limited. We can summarize qualitative data by counting the number of observations in each qualitative category or by computing the proportion of the observations in each qualitative category. However, even when the qualitative data use a numeric code, arithmetic operations such as addition, subtraction, multiplication, and division do not provide meaningful results. Section 2.1 discusses ways for summarizing qualitative data.

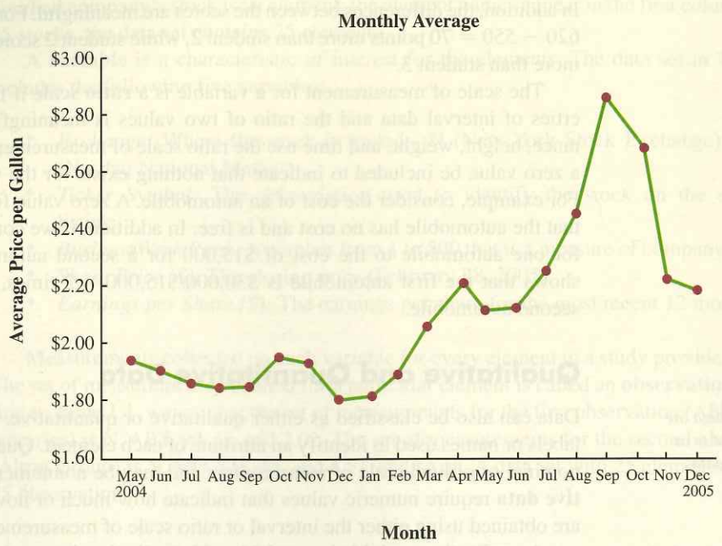
On the other hand, arithmetic operations often provide meaningful results for a quantitative variable. For example, for a quantitative variable, the data may be added and then divided by the number of observations to compute the average value. This average is usually meaningful and easily interpreted. In general, more alternatives for statistical analysis are possible when the data are quantitative. Section 2.2 and Chapter 3 provide ways of summarizing quantitative data.

Cross-Sectional and Time Series Data

For purposes of statistical analysis, distinguishing between cross-sectional data and time series data is important. **Cross-sectional data** are data collected at the same or approximately the same point in time. The data in Table 1.1 are cross-sectional because they describe the five variables for the 25 S&P 500 companies at the same point in time. **Time series data** are data collected over several time periods. For example, Figure 1.1 provides a graph of the U.S. city average price per gallon for unleaded regular gasoline. The graph shows gasoline price in a fairly stable band between $1.80 and $2.00 from May 2004 through February 2005. After that gasoline price became more volatile. It rose significantly, culminating with a sharp spike in September 2005.

Graphs of time series data are frequently found in business and economic publications. Such graphs help analysts understand what happened in the past, identify any trends over

FIGURE 1.1 U.S. CITY AVERAGE PRICE PER GALLON FOR CONVENTIONAL REGULAR GASOLINE



Source: U.S. Energy Information Administration, January 2006.

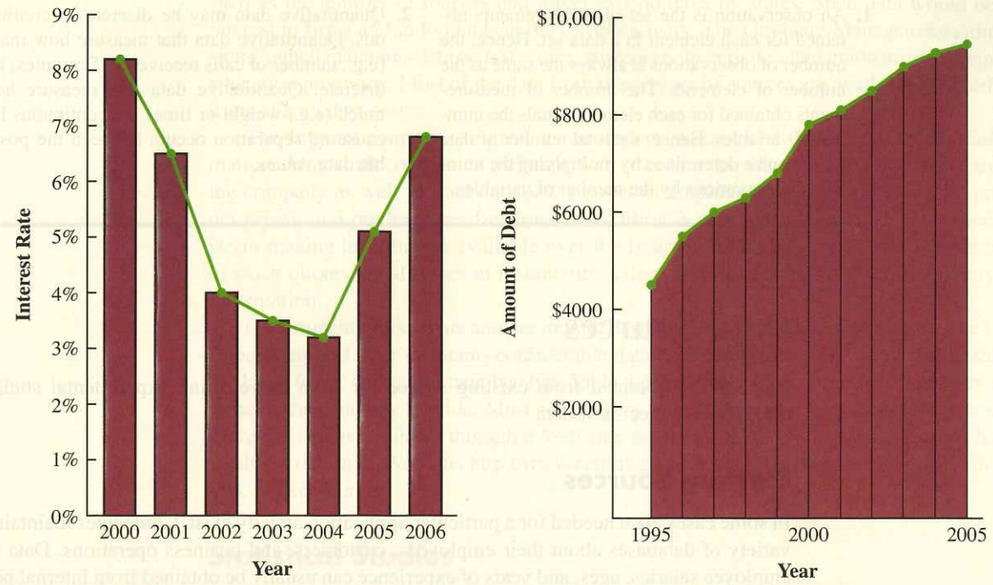
time, and project future levels for the time series. The graphs of time series data can take on a variety of forms, as shown in Figure 1.2. With a little study, these graphs are usually easy to understand and interpret.

For example, panel A in Figure 1.2 is a graph showing the interest rate for student Stafford Loans between 2000 and 2006. After 2000, the interest rate declined and reached its lowest level of 3.2% in 2004. However, after 2004, the interest rate for student loans showed a steep increase, reaching 6.8% in 2006. With the U.S. Department of Education estimating that more than 50% of undergraduate students graduate with debt, this increasing interest rate places a greater financial burden on many new college graduates.

The graph in panel B shows a rather disturbing increase in the average credit card debt per household over the 10-year period from 1995 to 2005. Notice how the time series shows an almost steady annual increase in the average credit card debt per household from $4500 in 1995 to $9500 in 2005. In 2005, an average credit card debt per household of $10,000 appeared not far off. Most credit card companies offer relatively low introductory interest rates. After this initial period, however, annual interest rates of 18%, 20%, or more are common. These rates make the credit card debt difficult for households to handle.

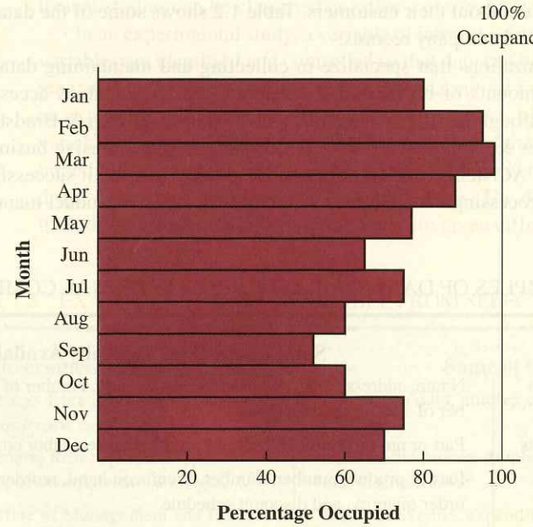
Panel C shows a graph of the time series for the occupancy rate of hotels in South Florida during a typical one-year period. Note that the form of the graph in panel C is different from the graphs in panels A and B, with the time in months shown on the vertical, rather than the horizontal axis. The highest occupancy rates of 95% to 98% occur during the months of February and March when the climate of South Florida is attractive to tourists. In fact, January to April is the typical high occupancy season for South Florida hotels. On the other hand, note the low occupancy rates in August to October; the lowest occupancy of 50% occurs in September. Higher temperatures and the hurricane season are the primary reasons for the drop in hotel occupancy during this period.

FIGURE 1.2 A VARIETY OF GRAPHS OF TIME SERIES DATA



(A) Interest Rate for Student Stafford Loans

(B) Average Credit Card Debt per Household



(C) Occupancy Rate of South Florida Hotels

NOTES AND COMMENTS

1. An observation is the set of measurements obtained for each element in a data set. Hence, the number of observations is always the same as the number of elements. The number of measurements obtained for each element equals the number of variables. Hence, the total number of data items can be determined by multiplying the number of observations by the number of variables.
2. Quantitative data may be discrete or continuous. Quantitative data that measure how many (e.g., number of calls received in 5 minutes) are discrete. Quantitative data that measure how much (e.g., weight or time) are continuous because no separation occurs between the possible data values.

1.3 Data Sources

Data can be obtained from existing sources or from surveys and experimental studies designed to collect new data.

Existing Sources

In some cases, data needed for a particular application already exist. Companies maintain a variety of databases about their employees, customers, and business operations. Data on employee salaries, ages, and years of experience can usually be obtained from internal personnel records. Other internal records contain data on sales, advertising expenditures, distribution costs, inventory levels, and production quantities. Most companies also maintain detailed data about their customers. Table 1.2 shows some of the data commonly available from internal company records.

Organizations that specialize in collecting and maintaining data make available substantial amounts of business and economic data. Companies access these external data sources through leasing arrangements or by purchase. Dun & Bradstreet, Bloomberg, and Dow Jones & Company are three firms that provide extensive business database services to clients. ACNielsen and Information Resources, Inc., built successful businesses collecting and processing data that they sell to advertisers and product manufacturers.

TABLE 1.2 EXAMPLES OF DATA AVAILABLE FROM INTERNAL COMPANY RECORDS

Source	Some of the Data Typically Available
Employee records	Name, address, social security number, salary, number of vacation days, number of sick days, and bonus
Production records	Part or product number, quantity produced, direct labor cost, and materials cost
Inventory records	Part or product number, number of units on hand, reorder level, economic order quantity, and discount schedule
Sales records	Product number, sales volume, sales volume by region, and sales volume by customer type
Credit records	Customer name, address, phone number, credit limit, and accounts receivable balance
Customer profile	Age, gender, income level, household size, address, and preferences

Data are also available from a variety of industry associations and special interest organizations. The Travel Industry Association of America maintains travel-related information such as the number of tourists and travel expenditures by states. Such data would be of interest to firms and individuals in the travel industry. The Graduate Management Admission Council maintains data on test scores, student characteristics, and graduate management education programs. Most of the data from these types of sources are available to qualified users at a modest cost.

The Internet continues to grow as an important source of data and statistical information. Almost all companies maintain Web sites that provide general information about the company as well as data on sales, number of employees, number of products, product prices, and product specifications. In addition, a number of companies now specialize in making information available over the Internet. As a result, one can obtain access to stock quotes, meal prices at restaurants, salary data, and an almost infinite variety of information.

Government agencies are another important source of existing data. For instance, the U.S. Department of Labor maintains considerable data on employment rates, wage rates, size of the labor force, and union membership. Table 1.3 lists selected governmental agencies and some of the data they provide. Most government agencies that collect and process data also make the results available through a Web site. For instance, the U.S. Census Bureau has a wealth of data at its Web site, http://www.census.gov. Figure 1.3 shows the homepage for the U.S. Census Bureau.

Statistical Studies

Sometimes the data needed for a particular application are not available through existing sources. In such cases, the data can often be obtained by conducting a statistical study. Statistical studies can be classified as either *experimental* or *observational.*

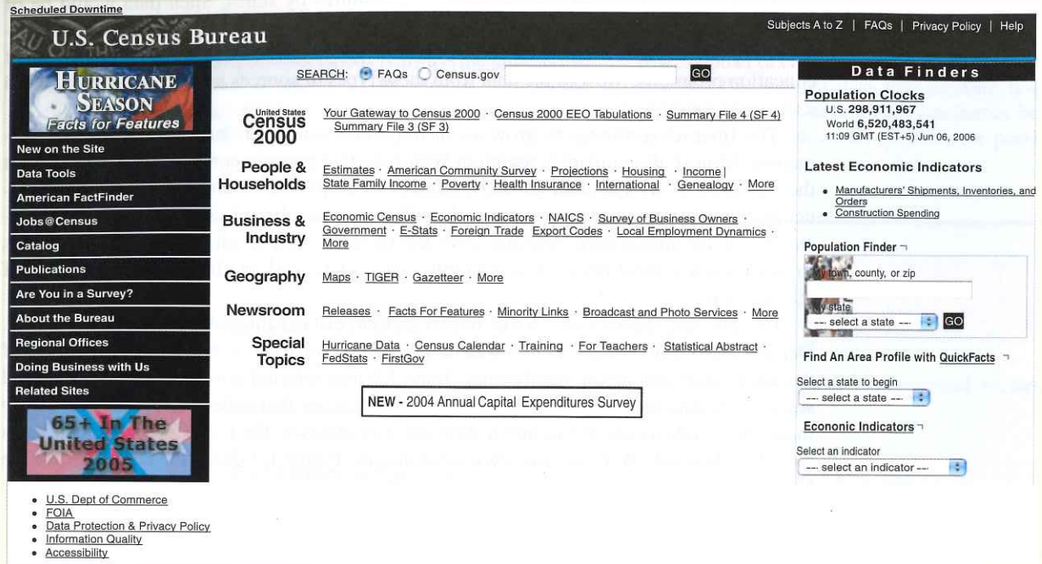
The largest experimental statistical study ever conducted is believed to be the 1954 Public Health Service experiment for the Salk polio vaccine. Nearly 2 million children in grades 1, 2, and 3 were selected from throughout the United States.

In an experimental study, a variable of interest is first identified. Then one or more other variables are identified and controlled so that data can be obtained about how they influence the variable of interest. For example, a pharmaceutical firm might be interested in conducting an experiment to learn about how a new drug affects blood pressure. Blood pressure is the variable of interest in the study. The dosage level of the new drug is another variable that is hoped to have a causal effect on blood pressure. To obtain data about the effect of the new drug, researchers select a sample of individuals. The dosage level of the new drug is controlled, as different groups of individuals are given different dosage levels. Before and after

TABLE 1.3 EXAMPLES OF DATA AVAILABLE FROM SELECTED GOVERNMENT AGENCIES

Government Agency	Some of the Data Available
Census Bureau *http://www.census.gov*	Population data, number of households, and household income
Federal Reserve Board *http://www.federalreserve.gov*	Data on the money supply, installment credit, exchange rates, and discount rates
Office of Management and Budget *http://www.whitehouse.gov/omb*	Data on revenue, expenditures, and debt of the federal government
Department of Commerce *http://www.doc.gov*	Data on business activity, value of shipments by industry, level of profits by industry, and growing and declining industries
Bureau of Labor Statistics *http://www.bls.gov*	Consumer spending, hourly earnings, unemployment rate, safety records, and international statistics

FIGURE 1.3 U.S. CENSUS BUREAU HOMEPAGE



Studies of smokers and nonsmokers are observational studies because researchers do not determine or control who will smoke and who will not smoke.

data on blood pressure are collected for each group. Statistical analysis of the experimental data can help determine how the new drug affects blood pressure.

Nonexperimental, or observational, statistical studies make no attempt to control the variables of interest. A survey is perhaps the most common type of observational study. For instance, in a personal interview survey, research questions are first identified. Then a questionnaire is designed and administered to a sample of individuals. Some restaurants use observational studies to obtain data about their customers' opinions of the quality of food, service, atmosphere, and so on. A questionnaire used by the Lobster Pot Restaurant in Redington Shores, Florida, is shown in Figure 1.4. Note that the customers completing the questionnaire are asked to provide ratings for five variables: food quality, friendliness of service, promptness of service, cleanliness, and management. The response categories of excellent, good, satisfactory, and unsatisfactory provide ordinal data that enable Lobster Pot's managers to assess the quality of the restaurant's operation.

Managers wanting to use data and statistical analysis as aids to decision making must be aware of the time and cost required to obtain the data. The use of existing data sources is desirable when data must be obtained in a relatively short period of time. If important data are not readily available from an existing source, the additional time and cost involved in obtaining the data must be taken into account. In all cases, the decision maker should consider the contribution of the statistical analysis to the decision-making process. The cost of data acquisition and the subsequent statistical analysis should not exceed the savings generated by using the information to make a better decision.

Data Acquisition Errors

Managers should always be aware of the possibility of data errors in statistical studies. Using erroneous data can be worse than not using any data at all. An error in data acquisition occurs whenever the data value obtained is not equal to the true or actual value that would be obtained with a correct procedure. Such errors can occur in a number of ways.

FIGURE 1.4 CUSTOMER OPINION QUESTIONNAIRE USED BY THE LOBSTER POT RESTAURANT, REDINGTON SHORES, FLORIDA

The LOBSTER Pot
RESTAURANT

We are happy you stopped by the Lobster Pot Restaurant and want to make sure you will come back. So, if you have a little time, we will really appreciate it if you will fill out this card. Your comments and suggestions are extremely important to us. Thank you!

Server's Name _____

	Excellent	Good	Satisfactory	Unsatisfactory
Food Quality	❑	❑	❑	❑
Friendly Service	❑	❑	❑	❑
Prompt Service	❑	❑	❑	❑
Cleanliness	❑	❑	❑	❑
Management	❑	❑	❑	❑

Comments _____

What prompted your visit to us? _____

Please drop in suggestion box at entrance. Thank you.

For example, an interviewer might make a recording error, such as a transposition in writing the age of a 24-year-old person as 42, or the person answering an interview question might misinterpret the question and provide an incorrect response.

Experienced data analysts take great care in collecting and recording data to ensure that errors are not made. Special procedures can be used to check for internal consistency of the data. For instance, such procedures would indicate that the analyst should review the accuracy of data for a respondent shown to be 22 years of age but reporting 20 years of work experience. Data analysts also review data with unusually large and small values, called outliers, which are candidates for possible data errors. In Chapter 3 we present some of the methods statisticians use to identify outliers.

Errors often occur during data acquisition. Blindly using any data that happen to be available or using data that were acquired with little care can result in misleading information and bad decisions. Thus, taking steps to acquire accurate data can help ensure reliable and valuable decision-making information.

1.4 Descriptive Statistics

Most of the statistical information in newspapers, magazines, company reports, and other publications consists of data that are summarized and presented in a form that is easy for the reader to understand. Such summaries of data, which may be tabular, graphical, or numerical, are referred to as **descriptive statistics**.

TABLE 1.4 FREQUENCIES AND PERCENT FREQUENCIES FOR THE EXCHANGE VARIABLE

Exchange	Frequency	Percent Frequency
New York Stock Exchange	20	80
Nasdaq National Market	5	20
Totals	25	100

Refer again to the data set in Table 1.1 showing data on 25 S&P 500 companies. Methods of descriptive statistics can be used to provide summaries of the information in this data set. For example, a tabular summary of the data for the qualitative variable Exchange is shown in Table 1.4. A graphical summary of the same data, called a bar graph, is shown in Figure 1.5. These types of tabular and graphical summaries generally make the data easier to interpret. Referring to Table 1.4 and Figure 1.5, we can see easily that the majority of the stocks in the data set are traded on the New York Stock Exchange. On a percentage basis, 80% are traded on the New York Stock Exchange and 20% are traded on the Nasdaq National Market.

A graphical summary of the data for the quantitative variable Share Price for the S&P stocks, called a histogram, is provided in Figure 1.6. The histogram makes it easy to see that the share prices range from $0 to $100, with the highest concentrations between $20 and $60.

In addition to tabular and graphical displays, numerical descriptive statistics are used to summarize data. The most common numerical descriptive statistic is the average, or mean. Using the data on the variable Earnings per Share for the S&P stocks in Table 1.1, we can compute the average by adding the earnings per share for all 25 stocks and dividing

FIGURE 1.5 BAR GRAPH FOR THE EXCHANGE VARIABLE

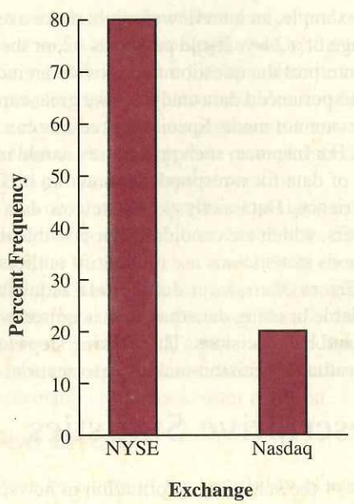
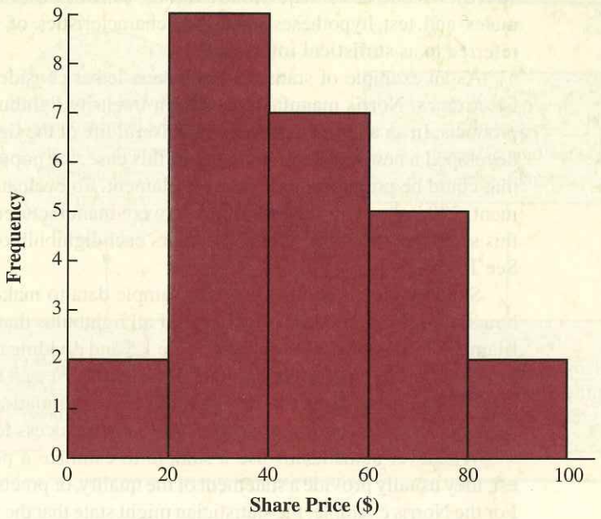


FIGURE 1.6 HISTOGRAM OF SHARE PRICE FOR 25 S&P STOCKS



the sum by 25. Doing so provides an average earnings per share of $2.49. This average demonstrates a measure of the central tendency, or central location, of the data for that variable.

In a number of fields, interest continues to grow in statistical methods that can be used for developing and presenting descriptive statistics. Chapters 2 and 3 devote attention to the tabular, graphical, and numerical methods of descriptive statistics.

1.5 Statistical Inference

Many situations require information about a large group of elements (individuals, companies, voters, households, products, customers, and so on). But, because of time, cost, and other considerations, data can be collected from only a small portion of the group. The larger group of elements in a particular study is called the **population**, and the smaller group is called the **sample**. Formally, we use the following definitions.

POPULATION

A population is the collection of all the elements of interest.

SAMPLE

A sample is a subset of the population.

The U.S. government conducts a census every 10 years. Market research firms conduct sample surveys every day.

The process of conducting a survey to collect data for the entire population is called a **census**. The process of conducting a survey to collect data for a sample is called a **sample survey**. As one of its major contributions, statistics uses data from a sample to make estimates and test hypotheses about the characteristics of a population through a process referred to as **statistical inference**.

As an example of statistical inference, let us consider the study conducted by Norris Electronics. Norris manufactures a high-intensity lightbulb used in a variety of electrical products. In an attempt to increase the useful life of the lightbulb, the product design group developed a new lightbulb filament. In this case, the population is defined as all lightbulbs that could be produced with the new filament. To evaluate the advantages of the new filament, 200 bulbs with the new filament were manufactured and tested. Data collected from this sample showed the number of hours each lightbulb operated before filament burnout. See Table 1.5.

Suppose Norris wants to use the sample data to make an inference about the average hours of useful life for the population of all lightbulbs that could be produced with the new filament. Adding the 200 values in Table 1.5 and dividing the total by 200 provides the sample average lifetime for the lightbulbs: 76 hours. We can use this sample result to estimate that the average lifetime for the lightbulbs in the population is 76 hours. Figure 1.7 provides a graphical summary of the statistical inference process for Norris Electronics.

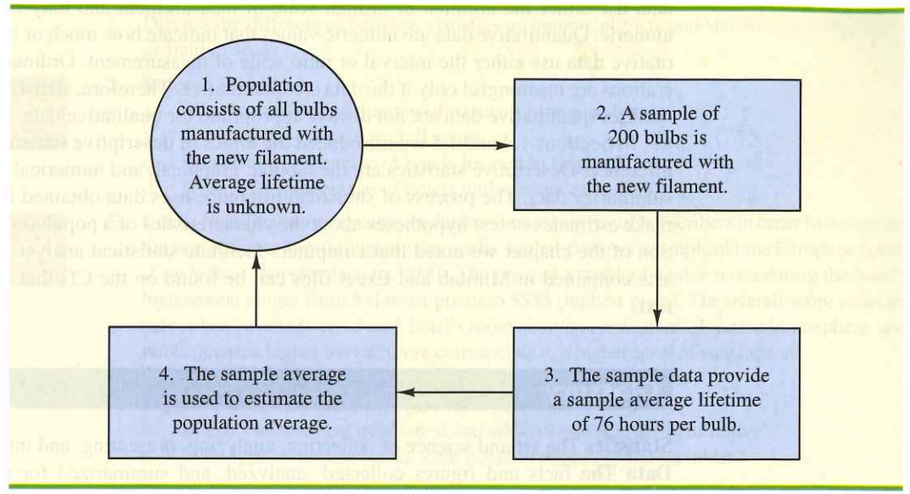
Whenever statisticians use a sample to estimate a population characteristic of interest, they usually provide a statement of the quality, or precision, associated with the estimate. For the Norris example, the statistician might state that the point estimate of the average lifetime for the population of new lightbulbs is 76 hours with a margin of error of ±4 hours. Thus, an interval estimate of the average lifetime for all lightbulbs produced with the new filament is 72 hours to 80 hours. The statistician can also state how confident he or she is that the interval from 72 hours to 80 hours contains the population average.

TABLE 1.5 HOURS UNTIL BURNOUT FOR A SAMPLE OF 200 LIGHTBULBS FOR THE NORRIS ELECTRONICS EXAMPLE

CD file 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

FIGURE 1.7 THE PROCESS OF STATISTICAL INFERENCE FOR THE NORRIS ELECTRONICS EXAMPLE



1.6 Computers and Statistical Analysis

Because statistical analysis typically involves large amounts of data, analysts frequently use computer software for this work. For instance, computing the average lifetime for the 200 lightbulbs in the Norris Electronics example (see Table 1.5) would be quite tedious without a computer. To facilitate computer usage, the larger data sets in this book are available on the CD that accompanies the text. A logo in the left margin of the text (e.g., Norris) identifies each of these data sets. The data files are available in both Minitab and Excel formats. In addition, we provide instructions in chapter appendixes for carrying out many of the statistical procedures using Minitab and Excel.

Summary

Statistics is the art and science of collecting, analyzing, presenting, and interpreting data. Nearly every college student majoring in business or economics is required to take a course in statistics. We began the chapter by describing typical statistical applications for business and economics.

Data consist of the facts and figures that are collected and analyzed. Four scales of measurement used to obtain data on a particular variable include nominal, ordinal, interval, and ratio. The scale of measurement for a variable is nominal when the data are labels or names used to identify an attribute of an element. The scale is ordinal if the data demonstrate the properties of nominal data and the order or rank of the data is meaningful. The scale is interval if the data demonstrate the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Finally, the scale of measurement is ratio if the data show all the properties of interval data and the ratio of two values is meaningful.

For purposes of statistical analysis, data can be classified as qualitative or quantitative. Qualitative data use labels or names to identify an attribute of each element. Qualitative data use either the nominal or ordinal scale of measurement and may be nonnumeric or numeric. Quantitative data are numeric values that indicate how much or how many. Quantitative data use either the interval or ratio scale of measurement. Ordinary arithmetic operations are meaningful only if the data are quantitative. Therefore, statistical computations used for quantitative data are not always appropriate for qualitative data.

In Sections 1.4 and 1.5 we introduced the topics of descriptive statistics and statistical inference. Descriptive statistics are the tabular, graphical, and numerical methods used to summarize data. The process of statistical inference uses data obtained from a sample to make estimates or test hypotheses about the characteristics of a population. In the last section of the chapter we noted that computers facilitate statistical analysis. The larger data sets contained in Minitab and Excel files can be found on the CD that accompanies the text.

Glossary

Statistics The art and science of collecting, analyzing, presenting, and interpreting data.
Data The facts and figures collected, analyzed, and summarized for presentation and interpretation.
Data set All the data collected in a particular study.
Elements The entities on which data are collected.
Variable A characteristic of interest for the elements.
Observation The set of measurements obtained for a particular element.
Nominal scale The scale of measurement for a variable when the data are labels or names used to identify an attribute of an element. Nominal data may be nonnumeric or numeric.
Ordinal scale The scale of measurement for a variable if the data exhibit the properties of nominal data and the order or rank of the data is meaningful. Ordinal data may be nonnumeric or numeric.
Interval scale The scale of measurement for a variable if the data demonstrate the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Interval data are always numeric.
Ratio scale The scale of measurement for a variable if the data demonstrate all the properties of interval data and the ratio of two values is meaningful. Ratio data are always numeric.
Qualitative data Labels or names used to identify an attribute of each element. Qualitative data use either the nominal or ordinal scale of measurement and may be nonnumeric or numeric.
Quantitative data Numeric values that indicate how much or how many of something. Quantitative data are obtained using either the interval or ratio scale of measurement.
Qualitative variable A variable with qualitative data.
Quantitative variable A variable with quantitative data.
Cross-sectional data Data collected at the same or approximately the same point in time.
Time series data Data collected over several time periods.
Descriptive statistics Tabular, graphical, and numerical summaries of data.
Population The collection of all the elements of interest.
Sample A subset of the population.
Census A survey to collect data on the entire population.
Sample survey A survey to collect data on a sample.
Statistical inference The process of using data obtained from a sample to make estimates or test hypotheses about the characteristics of a population.

Supplementary Exercises

1. Discuss the differences between statistics as numerical facts and statistics as a discipline or field of study.

SELF test

2. Refer to Table 1.6.
   a. What is the average number of rooms for the nine hotels?
   b. Compute the average overall score.
   c. What is the percentage of hotels located in England?
   d. What is the percentage of hotels with a room rate of $$?

SELF test

3. *Condé Nast Traveler* magazine conducts an annual survey of subscribers in order to determine the best places to stay throughout the world. Table 1.6 shows a sample of nine European hotels (*Condé Nast Traveler,* January 2000). The price of a standard double room during the hotel's high season ranges from $ (lowest price) to $$$$ (highest price). The overall score includes subscribers' evaluations of each hotel's rooms, service, restaurants, location/atmosphere, and public areas; a higher overall score corresponds to a higher level of satisfaction.
   a. How many elements are in this data set?
   b. How many variables are in this data set?
   c. Which variables are qualitative and which variables are quantitative?
   d. What type of measurement scale is used for each of the variables?
4. Consider the data set for the sample of 10 minisystems in Table 1.7.
   a. How many variables are in the data set?
   b. Which of the variables are quantitative and which are qualitative?
   c. What is the average CD capacity for the sample?
   d. What percentage of the minisystems provides an FM tuning rating of very good or excellent?
   e. What percentage of the minisystems includes two tape decks?
5. All-in-one sound systems, called minisystems, typically include an AM/FM tuner, a dual-cassette tape deck, and a CD changer in a book-sized box with two separate speakers. The data in Table 1.7 show the retail price, sound quality, CD capacity, FM tuning sensitivity and selectivity, and the number of tape decks for a sample of 10 minisystems (*Consumer Reports Buying Guide 2002*).
   a. How many elements does this data set contain?
   b. What is the population?
   c. Compute the average price for the sample.
   d. Using the results in part (c), estimate the average price for the population.

TABLE 1.6 RATINGS FOR NINE PLACES TO STAY IN EUROPE

CD file
Hotel

Name of Property	Country	Room Rate	Number of Rooms	Overall Score
Graveteye Manor	England	$$	18	83.6
Villa d'Este	Italy	$$$$	166	86.3
Hotel Prem	Germany	$	54	77.8
Hotel d'Europe	France	$$	47	76.8
Palace Luzern	Switzerland	$$	326	80.9
Royal Crescent Hotel	England	$$$	45	73.7
Hotel Sacher	Austria	$$$	120	85.5
Duc de Bourgogne	Belgium	$	10	76.9
Villa Gallici	France	$$	22	90.6

Source: Condé Nast Traveler, January 2000.

TABLE 1.7 A SAMPLE OF 10 MINISYSTEMS

CD file
Minisystems

Brand and Model	Price ($)	Sound Quality	CD Capacity	FM Tuning	Tape Decks
Aiwa NSX-AJ800	250	Good	3	Fair	2
JVC FS-SD1000	500	Good	1	Very Good	0
JVC MX-G50	200	Very Good	3	Excellent	2
Panasonic SC-PM11	170	Fair	5	Very Good	1
RCA RS 1283	170	Good	3	Poor	0
Sharp CD-BA2600	150	Good	3	Good	2
Sony CHC-CL1	300	Very Good	3	Very Good	1
Sony MHC-NX1	500	Good	5	Excellent	2
Yamaha GX-505	400	Very Good	3	Excellent	1
Yamaha MCR-E100	500	Very Good	1	Excellent	0

6. The Gallup organization conducted a telephone survey with a randomly selected national sample of 1005 adults, 18 years and older. The survey asked the respondents, "How would you describe your own physical health at this time?" (http://www.gallup.com, February 7, 2002). Response categories were Excellent, Good, Only Fair, Poor, and No Opinion.
   a. What was the sample size for this survey?
   b. Are the data qualitative or quantitative?
   c. Would it make more sense to use averages or percentages as a summary of the data for this question?
   d. Of the respondents, 29% said their personal health was excellent. How many individuals provided this response?
7. The Hawaii Visitors Bureau collects data on visitors to Hawaii. The following questions were among 16 asked in a questionnaire handed out to passengers during incoming airline flights in June 2003.
   - This trip to Hawaii is my: 1st, 2nd, 3rd, 4th, etc.
   - The primary reason for this trip is: (10 categories including vacation, convention, honeymoon)
   - Where I plan to stay: (11 categories including hotel, apartment, relatives, camping)
   - Total days in Hawaii

   a. What is the population being studied?
   b. Is the use of a questionnaire a good way to reach the population of passengers on incoming airline flights?
   c. Comment on each of the four questions in terms of whether it will provide qualitative or quantitative data.
8. The Commerce Department reported receiving the following applications for the Malcolm Baldrige National Quality Award: 23 from large manufacturing firms, 18 from large service firms, and 30 from small businesses.
   a. Is type of business a qualitative or quantitative variable?
   b. What percentage of the applications came from small businesses?
9. State whether each of the following variables is qualitative or quantitative and indicate its measurement scale.
   a. Annual sales
   b. Soft drink size (small, medium, large)
   c. Employee classification (GS1 through GS18)
   d. Earnings per share
   e. Method of payment (cash, check, credit card)

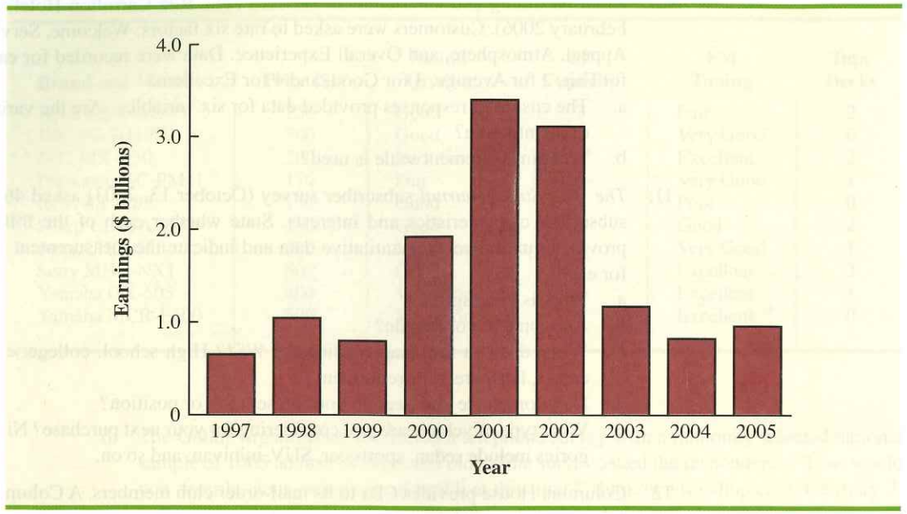
10. The Ritz-Carlton Hotel used a customer opinion questionnaire to obtain performance data about its dining and entertainment services (The Ritz-Carrolton Hotel, Naples, Florida, February 2006). Customers were asked to rate six factors: Welcome, Service, Food, Menu Appeal, Atmosphere, and Overall Experience. Data were recorded for each factor with 1 for Fair, 2 for Average, 3 for Good, and 4 for Excellent.
    a. The customer responses provided data for six variables. Are the variables qualitative or quantitative?
    b. What measurement scale is used?
11. *The Wall Street Journal* subscriber survey (October 13, 2003) asked 46 questions about subscriber characteristics and interests. State whether each of the following questions provided qualitative or quantitative data and indicate the measurement scale appropriate for each.
    a. What is your age?
    b. Are you male or female?
    c. When did you first start reading the *WSJ*? High school, college, early career, mid-career, late career, or retirement?
    d. How long have you been in your present job or position?
    e. What type of vehicle are you considering for your next purchase? Nine response categories include sedan, sports car, SUV, minivan, and so on.
12. Columbia House provides CDs to its mail-order club members. A Columbia House Music Survey asked new club members to complete an 11-question survey. Some of the questions asked were:
    a. How many CDs have you bought in the last 12 months?
    b. Are you currently a member of a national mail-order book club? (Yes or No)
    c. What is your age?
    d. Including yourself, how many people (adults and children) are in your household?
    e. What kind of music are you interested in buying? Fifteen categories were listed, including hard rock, soft rock, adult contemporary, heavy metal, rap, and country.

    Comment on whether each question provides qualitative or quantitative data.
13. CSM Worldwide forecasts global production for all automobile manufacturers. The following CSM data show the forecast of global auto production for General Motors, Ford, DaimlerChrysler, and Toyota for the years 2004 to 2007 (*USA Today,* December 21, 2005). Data are in millions of vehicles.

Manufacturer	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. Construct a time series graph for the years 2004 to 2007 showing the number of vehicles manufactured by each automotive company. Show the time series for all four manufacturers on the same graph.
    b. General Motors has been the undisputed production leader of automobiles since 1931. What does the time series graph show about who is the world's biggest car company? Discuss.
    c. Construct a bar graph showing vehicles produced by automobile manufacturer using the 2007 data. Is this graph based on cross-sectional or time series data?

FIGURE 1.8 EARNINGS FOR VOLKSWAGEN



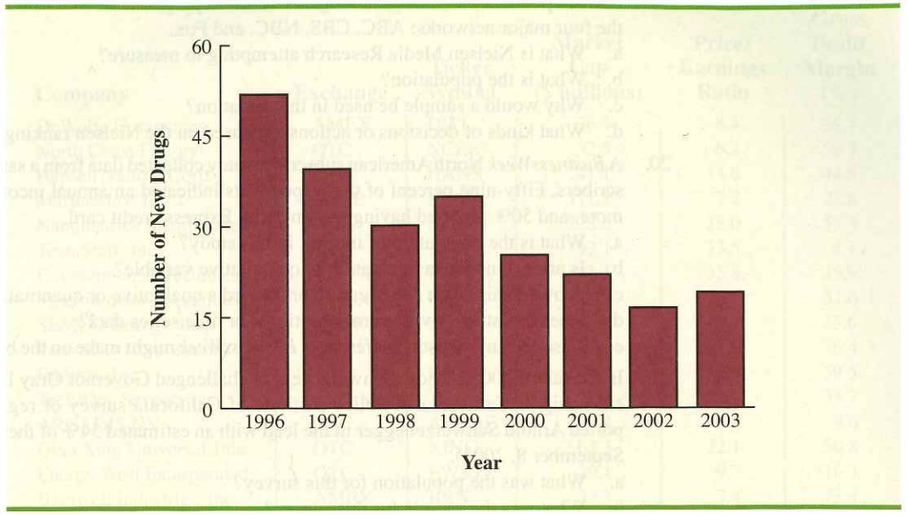
SELF test

14. Figure 1.8 provides a bar graph summarizing the earnings for Volkswagen for the years 1997 to 2005 (*BusinessWeek,* December 26, 2005).
    a. Are the data qualitative or quantitative?
    b. Are the data time series or cross-sectional?
    c. What is the variable of interest?
    d. Comment on the trend in Volkswagen's earnings over time. The *BusinessWeek* article (December 26, 2005) estimated earnings for 2006 at $600 million or $.6 billion. Does Figure 1.8 indicate whether this estimate appears to be reasonable?
    e. A similar article that appeared in *BusinessWeek* on July 23, 2001, only had the data from 1997 to 2000 along with higher earnings projected for 2001. What was the outlook for Volkswagen's earnings in July 2001? Did an investment in Volkswagen look promising in 2001? Explain.
    f. What warning does this graph suggest about projecting data such as Volkswagen's earnings into the future?

15. The marketing group at your company developed a new diet soft drink that it claims will capture a large share of the young adult market.
    a. What data would you want to see before deciding to invest substantial funds in introducing the new product into the marketplace?
    b. How would you expect the data mentioned in part (a) to be obtained?

16. The Food and Drug Administration (FDA) reported the number of new drugs approved over an eight-year period (*The Wall Street Journal,* January 12, 2004). Figure 1.9 provides a bar graph summarizing the number of new drugs approved each year.
    a. Are the data qualitative or quantitative?
    b. Are the data time series or cross-sectional?
    c. How many new drugs were approved in 2003?
    d. In what year were the fewest new drugs approved? How many?
    e. Comment on the trend in the number of new drugs approved by the FDA over the eight-year period.

FIGURE 1.9 NUMBER OF NEW DRUGS APPROVED BY THE FOOD AND DRUG ADMINISTRATION



17. A sample of midterm grades for five students showed the following results: 72, 65, 82, 90, 76. Which of the following statements are correct, and which should be challenged as being too generalized?
    a. The average midterm grade for the sample of five students is 77.
    b. The average midterm grade for all students who took the exam is 77.
    c. An estimate of the average midterm grade for all students who took the exam is 77.
    d. More than half of the students who take this exam will score between 70 and 85.
    e. If five other students are included in the sample, their grades will be between 65 and 90.

18. A survey of 131 investment managers in *Barron's* Big Money poll revealed the following (*Barron's,* October 28, 2002):
    - 43% of managers classified themselves as bullish or very bullish on the stock market.
    - The average expected return over the next 12 months for equities was 11.2%.
    - 21% selected health care as the sector most likely to lead the market in the next 12 months.
    - When asked to estimate how long it would take for technology and telecom stocks to resume sustainable growth, the managers' average response was 2.5 years.

    a. Cite two descriptive statistics.
    b. Make an inference about the population of all investment managers concerning the average return expected on equities over the next 12 months.
    c. Make an inference about the length of time it will take for technology and telecom stocks to resume sustainable growth.

19. Nielsen Media Research conducts weekly surveys of television viewing throughout the United States, publishing both rating and market share data. The Nielsen rating is the percentage of households with televisions watching a program, while the Nielsen share is the percentage of households watching a program among those households with televisions in use. For example, Nielsen Media Research results for the 2003 Baseball World Series between the New York Yankees and the Florida Marlins showed a rating of 12.8% and a share of 22% (Associated Press, October 27, 2003). Thus, 12.8% of households with televisions

were watching the World Series and 22% of households with televisions in use were watching the World Series. Based on the rating and share data for major television programs, Nielsen publishes a weekly ranking of television programs as well as a weekly ranking of the four major networks: ABC, CBS, NBC, and Fox.

a. What is Nielsen Media Research attempting to measure?
b. What is the population?
c. Why would a sample be used in this situation?
d. What kinds of decisions or actions are based on the Nielsen rankings?

20. A *BusinessWeek* North American subscriber study collected data from a sample of 2861 subscribers. Fifty-nine percent of the respondents indicated an annual income of $75,000 or more, and 50% reported having an American Express credit card.
    a. What is the population of interest in this study?
    b. Is annual income a qualitative or quantitative variable?
    c. Is ownership of an American Express card a qualitative or quantitative variable?
    d. Does this study involve cross-sectional or time series data?
    e. Describe any statistical inferences *BusinessWeek* might make on the basis of the survey.
21. In the fall of 2003, Arnold Schwarzenegger challenged Governor Gray Davis for the governorship of California. A Policy Institute of California survey of registered voters reported Arnold Schwarzenegger in the lead with an estimated 54% of the vote (*Newsweek,* September 8, 2003).
    a. What was the population for this survey?
    b. What was the sample for this survey?
    c. Why was a sample used in this situation? Explain.
22. A survey of 430 business travelers found 155 business travelers used a travel agent to make the travel arrangements (*USA Today,* November 20, 2003).
    a. Develop a descriptive statistic that can be used to estimate the percentage of all business travelers who use a travel agent to make travel arrangements.
    b. The survey reported that the most frequent way business travelers make travel arrangements is by using an online travel site. If 44% of business travelers surveyed made their arrangements this way, how many of the 430 business travelers used an online travel site?
    c. Are the data on how travel arrangements are made qualitative or quantitative?
23. A seven-year medical research study reported that women whose mothers took the drug DES during pregnancy were twice as likely to develop tissue abnormalities that might lead to cancer as were women whose mothers did not take the drug.
    a. This study involved the comparison of two populations. What were the populations?
    b. Do you suppose the data were obtained in a survey or an experiment?
    c. For the population of women whose mothers took the drug DES during pregnancy, a sample of 3980 women showed 63 developed tissue abnormalities that might lead to cancer. Provide a descriptive statistic that could be used to estimate the number of women out of 1000 in this population who have tissue abnormalities.
    d. For the population of women whose mothers did not take the drug DES during pregnancy, what is the estimate of the number of women out of 1000 who would be expected to have tissue abnormalities?
    e. Medical studies often use a relatively large sample (in this case, 3980). Why?
24. A manager of a large corporation recommends a $10,000 raise be given to keep a valued subordinate from moving to another company. What internal and external sources of data might be used to decide whether such a salary increase is appropriate?
25. Table 1.8 shows a data set containing information for 25 of the shadow stocks tracked by the American Association of Individual Investors (aaii.com, February 2002). Shadow

TABLE 1.8 DATA SET FOR 25 SHADOW STOCKS

CD file
Shadow02

Company	Exchange	Ticker Symbol	Market Cap ($ millions)	Price/ Earnings Ratio	Gross Profit Margin (%)
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

stocks are common stocks of smaller companies that are not closely followed by Wall Street analysts. The data set is also on the CD accompanying the text in the file named Shadow02.

a. How many variables are in the data set?
b. Which of the variables are qualitative and which are quantitative?
c. For the Exchange variable, show the frequency and the percent frequency for AMEX, NYSE, and OTC. Construct a bar graph similar to Figure 1.5 for the Exchange variable.
d. Show the frequency distribution for the Gross Profit Margin using the five intervals: 0–14.9, 15–29.9, 30–44.9, 45–59.9, and 60–74.9. Construct a histogram similar to Figure 1.6.
e. What is the average price/earnings ratio?