1

What Is Statistics?

GOALS

When you have completed this chapter you will be able to:

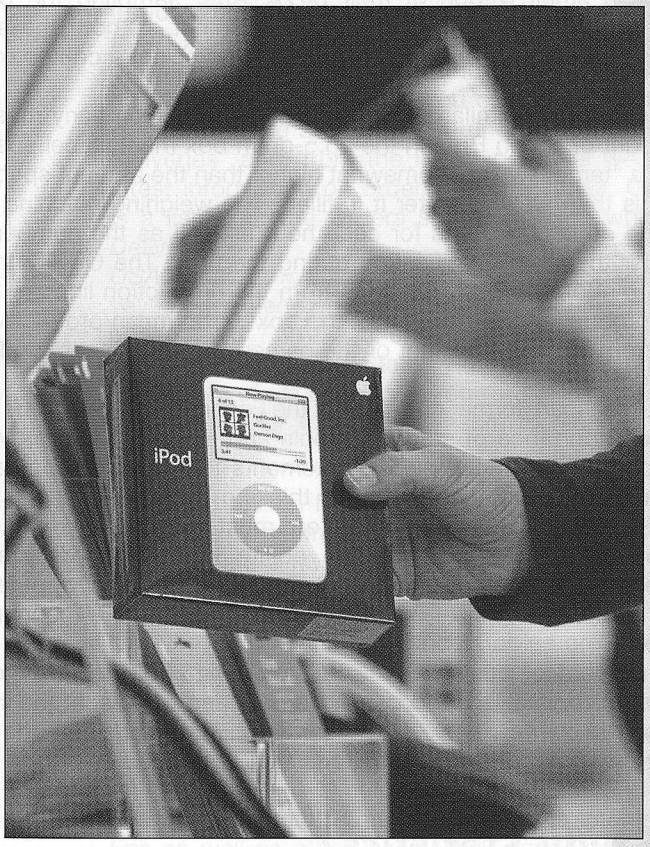
1 Understand why we study statistics.

2 Explain what is meant by *descriptive statistics* and *inferential statistics.*

3 Distinguish between a *qualitative variable* and a *quantitative variable.*

4 Describe how a *discrete variable* is different from a *continuous variable.*

5 Distinguish among the *nominal, ordinal, interval,* and *ratio* levels of measurement.



You are shopping for a new MP3 music player such as the Apple iPod. The manufacturers advertise the number of music tracks that can be stored in the memory. You, however, like Broadway musical tunes, which are much longer. You would like to estimate how many Broadway tunes will fit on your MP3 player. Would you collect information using a sample of a population? Why? (See Exercise 8d and Goal 2.)

Introduction

More than 100 years ago H. G. Wells, an English author and historian, suggested that one day quantitative reasoning will be as necessary for effective citizenship as the ability to read. He made no mention of business because the Industrial Revolution was just beginning. Mr. Wells could not have been more correct. While "business experience," some "thoughtful guesswork," and "intuition" are key attributes of successful managers, today's business problems tend to be too complex for this type of decision making alone.

One of the tools used to make decisions is statistics. Statistics is used not only by businesspeople; we all also apply statistical concepts in our lives. For example, to start the day you turn on the shower and let it run for a few moments. Then you put your hand in the shower to sample the temperature and *decide* to add more hot water or more cold water, or that the temperature is just right and to enter the shower. As a second example, suppose you are at the grocery store and wish to buy a frozen pizza. One of the pizza makers has a stand, and they offer a small wedge of their pizza. After sampling the pizza, you *decide* whether to purchase the pizza or not. In both the shower and pizza examples, you make a decision and select a course of action based on a sample.

Businesses face similar situations. The Kellogg Company must ensure that the mean amount of Raisin Bran in the 25.5-gram box meets label specifications. To do so, it sets a "target" weight somewhat higher than the amount specified on the label. Each box is then weighed after it is filled. The weighing machine reports a distribution of the content weights for each hour as well as the number "kicked-out" for being under the label specification during the hour. The Quality Inspection Department also randomly selects samples from the production line and checks the quality of the product and the weight of the box. If the mean product weight differs significantly from the target weight or the percent of kick-outs is too large, the process is adjusted.

Alan Greenspan, former chairman of the Federal Reserve Board, knows and understands the importance of statistical tools and techniques to provide accurate and timely information to make public statements that have the power to move global stock markets and influence political thinking. Dr. Greenspan, speaking before a National Skills Summit, stated: "Workers must be equipped not simply with technical know-how, but also with the ability to create, analyze, and transform information and to interact effectively with others. That is, separate the facts from opinions, and then organize these facts in an appropriate manner and analyze the information."

As a student of business or economics, you will need basic knowledge and skills to organize, analyze, and transform data and to present the information. In this text, we will show you basic statistical techniques and methods that will develop your ability to make good personal and business decisions.

Why Study Statistics?

If you look through your university catalog, you will find that statistics is required for many college programs. Why is this so? What are the differences in the statistics courses taught in the Engineering College, the Psychology or Sociology Departments in the Liberal Arts College, and the College of Business? The biggest difference is the examples used. The course content is basically the same. In the College of Business we are interested in such things as profits, hours worked, and wages. Psychologists are interested in test scores, and engineers are interested in how many units are manufactured on a particular machine. However, all three are

interested in what is a typical value and how much variation there is in the data. There may also be a difference in the level of mathematics required. An engineering statistics course usually requires calculus. Statistics courses in colleges of business and education usually teach the course at a more applied level. You should be able to handle the mathematics in this text if you have completed high school algebra.

Examples of why we study statistics

So why is statistics required in so many majors? The first reason is that numerical information is everywhere. Look in the newspapers *(USA Today),* news magazines *(Time, Newsweek, U.S. News and World Report),* business magazines *(BusinessWeek, Forbes),* or general interest magazines *(People),* women's magazines *(Ladies Home Journal* or *Elle),* or sports magazines *(Sports Illustrated, ESPN The Magazine),* and you will be bombarded with numerical information.

Here are some examples:

- In 2003 the typical household income in the United States was $43,318. For households in the Northeast the typical income was $46,742, $44,732 in the Midwest, $39,823 in the South, and $46,820 in the West. You can check the latest information by going to http://www.census.gov/hhes/income.
- In July 2005 Boeing reported that it had delivered 155 aircraft for the period from January 1, 2005, to June 30, 2005. There were a total of 113 737s delivered during the period, and Southwest Airlines was the largest purchaser, having bought 22 aircraft. You can see the latest information by going to the Boeing website at www.boeing.com, then in the search box type "orders and deliveries" and, from the list of potential websites, pick the one offering the latest information on orders and deliveries. On the lefthand side of this page is a listing of the Orders site map, from which you can select Current Year Deliveries.
- *USA Today* (www.usatoday.com) prints "Snapshots" that are the result of surveys conducted by various research organizations, foundations, and the federal government. For example, many believe that email is preferable to regular mail. However, the United States Postal Service reports from a recent survey that 67 percent of adults indicate regular mail is more personal than email, 56 percent indicate regular mail is a pleasure to receive, and 55 percent look forward to opening mail.

A second reason for taking a statistics course is that statistical techniques are used to make decisions that affect our daily lives. That is, they affect our personal welfare. Here are a few examples:

- Insurance companies use statistical analysis to set rates for home, automobile, life, and health insurance. Tables are available showing estimates that a 20-year-old female has 60.25 years of life remaining, an 87-year-old woman 4.56 years remaining, and a 50-year-old man 27.85 years remaining. Life insurance premiums are established based on these estimates of life expectancy. These tables are available at www.ssa.gov/OACT/STATS/table4cb.html. [This site is sensitive to capital letters.]
- The Environmental Protection Agency is interested in the water quality of Lake Erie as well as other lakes. They periodically take water samples to establish the level of contamination and maintain the level of quality.
- Medical researchers study the cure rates for diseases using different drugs and different forms of treatment. For example, what is the effect of treating a certain type of knee injury surgically or with physical therapy? If you take an aspirin each day, does that reduce your risk of a heart attack?

A third reason for taking a statistics course is that the knowledge of statistical methods will help you understand how decisions are made and give you a better understanding of how they affect you.

Statistics in Action

We call your attention to a feature title—*Statistics in Action*. Read each one carefully to get an appreciation of the wide application of statistics in management, economics, nursing, law enforcement, sports, and other disciplines.

- In 2005, *Forbes* published a list of the richest Americans. William Gates, founder of Microsoft Corporation, is the richest. His net worth is estimated at $46.5 billion. (www.forbes.com)
- In 2005, the four largest American companies, ranked by revenue were ExxonMobil, General Motors, Ford, and Chevron. (www.Industryweek.com)
- In the United States, a typical high school graduate earns $1.2 million in his or her lifetime, a typical college graduate with a bachelor's degree earns $2.1 million, and a typical college graduate with a master's degree earns $2.5 million. (usgovinfo.about.com/library/weekly/aa072602a.htm)

No matter what line of work you select, you will find yourself faced with decisions where an understanding of data analysis is helpful. In order to make an informed decision, you will need to be able to:

1. Determine whether the existing information is adequate or additional information is required.
2. Gather additional information, if it is needed, in such a way that it does not provide misleading results.
3. Summarize the information in a useful and informative manner.
4. Analyze the available information.
5. Draw conclusions and make inferences while assessing the risk of an incorrect conclusion.

The statistical methods presented in the text will provide you with a framework for the decision-making process.

In summary, there are at least three reasons for studying statistics: (1) data are everywhere, (2) statistical techniques are used to make many decisions that affect our lives, and (3) no matter what your career, you will make professional decisions that involve data. An understanding of statistical methods will help you make these decisions more effectively.

What Is Meant by Statistics?

How do we define the word *statistics?* We encounter it frequently in our everyday language. It really has two meanings. In the more common usage, statistics refers to numerical information. Examples include the average starting salary of college graduates, the number of deaths due to alcoholism last year, the change in the Dow Jones Industrial Average from yesterday to today, and the number of home runs hit by the Chicago Cubs during the 2005 season. In these examples statistics are a value or a percentage. Other examples include:

- The typical automobile in the United States travels 11,099 miles per year, the typical bus 9,353 miles per year, and the typical truck 13,942 miles per year. In Canada the corresponding information is 10,371 miles for automobiles, 19,823 miles for buses, and 7,001 miles for trucks.
- The mean time waiting for technical support is 17 minutes.
- The mean length of the business cycle since 1945 is 61 months.

The above are all examples of **statistics.** A collection of numerical information is called **statistics** (plural).

We often present statistical information in a graphical form. A graph is often useful for capturing reader attention and to portray a large amount of information. For example, Chart 1–1 shows Frito-Lay volume and market share for the major snack and potato chip categories in supermarkets in the United States. It requires only a quick glance to discover there were nearly 800 million pounds of potato chips sold and that Frito-Lay sold 64 percent of that total. Also note that Frito-Lay has 82 percent of the corn chip market.

The subject of statistics, as we will explore it in this text, has a much broader meaning than just collecting and publishing numerical information. We define statistics as:

STATISTICS The science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.

As the definition suggests, the first step in investigating a problem is to collect relevant data. They must be organized in some way and perhaps presented in a chart, such

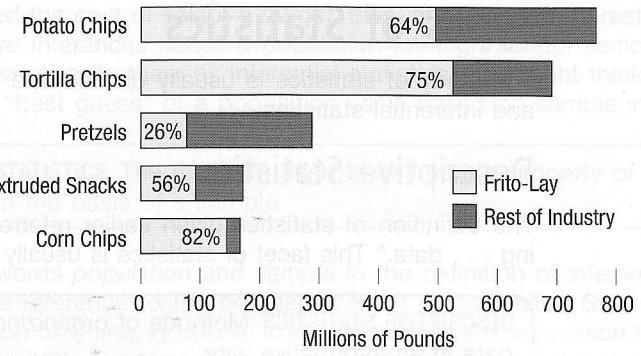
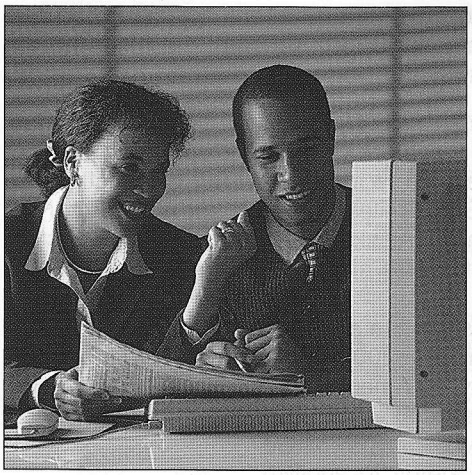


CHART 1–1 Frito-Lay Volume and Share of Major Snack Chip Categories in U.S. Supermarkets

as Chart 1–1. Only after the data have been organized are we then able to analyze and interpret them. Here are some examples of the need for data collection.



- Research analysts for Merrill Lynch evaluate many facets of a particular stock before making a "buy" or "sell" recommendation. They collect the past sales data of the company and estimate future earnings. Other factors, such as the projected worldwide demand for the company's products, the strength of the competition, and the effect of the new union–management contract, are also considered before making a recommendation.
- The marketing department at Colgate-Palmolive Co., a manufacturer of soap products, has the responsibility of making recommendations regarding the potential profitability of a newly developed group of face soaps having fruit smells, such as grape, orange, and pineapple. Before making a final decision, the marketers will test it in several markets. That is, they may advertise and sell it in Topeka, Kansas, and Tampa, Florida. On the basis of test marketing in these two regions, Colgate-Palmolive will make a decision whether to market the soaps in the entire country.
- The United States government is concerned with the present condition of our economy and with predicting future economic trends. The government conducts a large number of surveys to determine consumer confidence and the outlook of managers regarding sales and production for the next 12 months. Indexes, such as the Consumer Price Index, are constructed each month to assess inflation. Information on department store sales, housing starts, money turnover, and industrial production are just a few of the hundreds of items used to form the basis of the projections. These projections are used by banks to decide their prime lending rate and by the Federal Reserve Board to decide the level of control to place on the money supply.
- Managers must make decisions about the quality of their product or service. For example, customers call software companies for technical advice when they are not able to resolve an issue regarding the software. One measure of the quality of customer service is the time a customer must wait for a technical consultant to answer the call. A software company might set a target of one minute as the typical response time. The company would then collect and analyze data on the response time. Does the typical response time differ by day of the week or time of day? If the response times are increasing, managers might decide to increase the number of technical consultants at particular times of the day or week.

Types of Statistics

The study of statistics is usually divided into two categories: descriptive statistics and inferential statistics.

Descriptive Statistics

The definition of statistics given earlier referred to "organizing, presenting, analyzing . . . data." This facet of statistics is usually referred to as **descriptive statistics.**

DESCRIPTIVE STATISTICS Methods of organizing, summarizing, and presenting data in an informative way.

For instance, the United States government reports the population of the United States was 179,323,000 in 1960; 203,302,000 in 1970; 226,542,000 in 1980; 248,709,000 in 1990, and 265,000,000 in 2000. This information is descriptive statistics. It is descriptive statistics if we calculate the percentage growth from one decade to the next. However, it would **not** be descriptive statistics if we used these to estimate the population of the United States in the year 2010 or the percentage growth from 2000 to 2010. Why? Because these statistics are not being used to summarize past populations but to estimate future populations. The following are some other examples of descriptive statistics.

- There are a total of 42,796 miles of interstate highways in the United States. The interstate system represents only 1 percent of the nation's total roads but carries more than 20 percent of the traffic. The longest is I-90, which stretches from Boston to Seattle, a distance of 3,081 miles. The shortest is I-878 in New York City, which is 0.70 of a mile in length. Alaska does not have any interstate highways, Texas has the most interstate miles at 3,232, and New York has the most interstate routes with 28.
- According to the *Bureau of Labor Statistics,* the average hourly earnings of production workers were $17.73 for January 2006. You can review the latest information on wages and productivity of American workers by going to the Bureau of Labor Statistics website at: http://www.bls.gov/home.htm and selecting Average Hourly Earnings.

Masses of unorganized data—such as the census of population, the weekly earnings of thousands of computer programmers, and the individual responses of 2,000 registered voters regarding their choice for president of the United States—are of little value as is. However, statistical techniques are available to organize this type of data into a meaningful form. Data can be organized into a **frequency distribution.** (This procedure is covered in Chapter 2.) Various **charts** may be used to describe data; several basic chart forms are also presented in Chapter 4.

Specific measures of central location, such as the mean, describe the central value of a group of numerical data. A number of statistical measures are used to describe how closely the data cluster about an average. These measures of central tendency and dispersion are discussed in Chapter 3.

Inferential Statistics

The second type of statistics is **inferential statistics**—also called **statistical inference.** Our main concern regarding inferential statistics is finding something about a population from a sample taken from that population. For example, a recent survey showed only 46 percent of high school seniors can solve problems involving fractions, decimals, and percentages. And only 77 percent of high school seniors

correctly totaled the cost of salad, a burger, fries, and a cola on a restaurant menu. Since these are inferences about a population (all high school seniors) based on sample data, we refer to them as inferential statistics. You might think of inferential statistics as a "best guess" of a population value based on sample information.

INFERENTIAL STATISTICS The methods used to estimate a property of a population on the basis of a sample.

Note the words *population* and *sample* in the definition of inferential statistics. We often make reference to the population living in the United States or the 1.31 billion population of China. However, in statistics the word *population* has a broader meaning. A **population** may consist of *individuals*—such as all the students enrolled at Utah State University, all the students in Accounting 201, or all the CEOs from the Fortune 500 companies. A population may also consist of *objects,* such as all the Cobra G/T tires produced at Cooper Tire and Rubber Company in the Findlay, Ohio, plant; the accounts receivable at the end of October for Lorrange Plastics, Inc.; or auto claims filed in the first quarter of 2006 at the Northeast Regional Office of State Farm Insurance. The *measurement* of interest might be the scores on the first examination of all students in Accounting 201, the tread wear of the Cooper Tires, the dollar amount of Lorrange Plastics's accounts receivable, or the amount of auto insurance claims at State Farm. Thus, a population in the statistical sense does not always refer to people.

POPULATION The entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest.

To infer something about a population, we usually take a **sample** from the population.

SAMPLE A portion, or part, of the population of interest.

Reasons for sampling

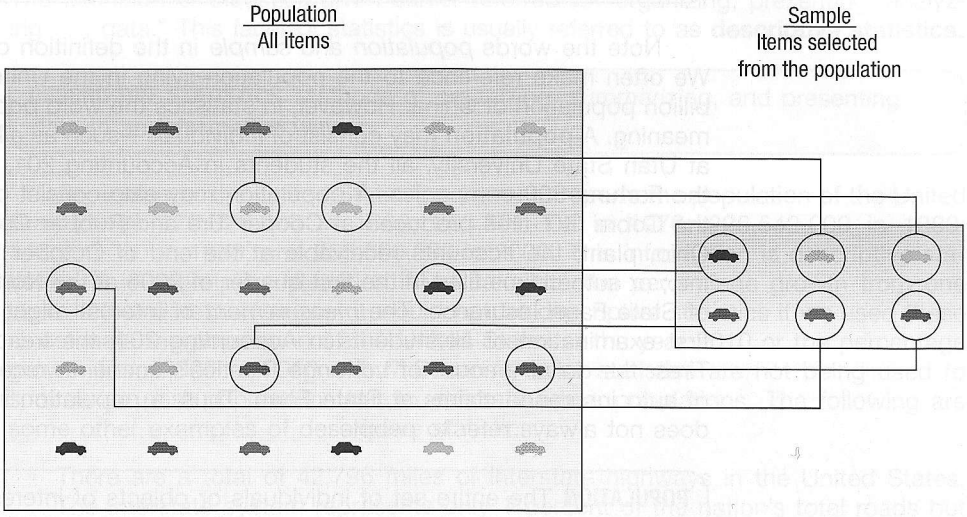
Why take a sample instead of studying every member of the population? A sample of registered voters is necessary because of the prohibitive cost of contacting millions of voters before an election. Testing wheat for moisture content destroys the wheat, thus making a sample imperative. If the wine tasters tested all the wine, none would be available for sale. It would be physically impossible for a few marine biologists to capture and tag all the seals in the ocean. (These and other reasons for sampling are discussed in Chapter 8.)

As noted, using a sample to learn something about a population is done extensively in business, agriculture, politics, and government, as cited in the following examples:

- Television networks constantly monitor the popularity of their programs by hiring Nielsen and other organizations to sample the preferences of TV viewers. For example, in a sample of 800 prime-time viewers, 320, or 40 percent, indicated they watched "CSI: Crime Scene Investigation" on CBS last week. These program ratings are used to set advertising rates or to cancel programs.
- Gamous and Associates, a public accounting firm, is conducting an audit of Pronto Printing Company. To begin, the accounting firm selects a random sample of 100 invoices and checks each invoice for accuracy. There is at least one error on five of the invoices; hence the accounting firm estimates that 5 percent of the population of invoices contain at least one error.
- A random sample of 1,260 marketing graduates from four-year schools showed their mean starting salary was $42,694. We therefore estimate the

mean starting salary for all accounting graduates of four-year institutions to be $42,694.

The relationship between a sample and a population is portrayed below. For example, we wish to estimate the mean miles per gallon of SUVs. Six SUVs are selected from the population. The mean MPG of the six is used to estimate MPG for the population.



We strongly suggest you do the Self-Review exercise.

Following is a self-review problem. There are a number of them interspersed throughout each chapter. They test your comprehension of the preceding material. The answer and method of solution are given at the end of the chapter. You can find the answer to the following Self-Review on page 19. We recommend that you solve each one and then check your answer.

Self-Review 1–1

The answers are at the end of the chapter.

The Atlanta-based advertising firm, Brandon and Associates, asked a sample of 1,960 consumers to try a newly developed chicken dinner by Boston Market. Of the 1,960 sampled, 1,176 said they would purchase the dinner if it is marketed.

(a) What could Brandon and Associates report to Boston Market regarding acceptance of the chicken dinner in the population?

(b) Is this an example of descriptive statistics or inferential statistics? Explain.

Types of Variables

Qualitative variable

There are two basic types of variables: (1) qualitative and (2) quantitative (see Chart 1–2). When the characteristic being studied is nonnumeric, it is called a **qualitative variable** or an **attribute.** Examples of qualitative variables are gender, religious affiliation, type of automobile owned, state of birth, and eye color. When the data are qualitative, we are usually interested in how many or what proportion fall in each category. For example, what percent of the population has blue eyes? How many Catholics and how many Protestants are there in the United States? What percent of the total number of cars sold last month were SUVs? Qualitative data are often summarized in charts and bar graphs (Chapter 2).

Quantitative variable

When the variable studied can be reported numerically, the variable is called a **quantitative variable.** Examples of quantitative variables are the balance in your

checking account, the ages of company presidents, the life of an automobile battery (such as 42 months), and the number of children in a family.

Quantitative variables are either discrete or continuous. **Discrete variables** can assume only certain values, and there are "gaps" between the values. Examples of discrete variables are the number of bedrooms in a house (1, 2, 3, 4, etc.), the number of cars arriving at Exit 25 on I-4 in Florida near Walt Disney World in an hour (326, 421, etc.), and the number of students in each section of a statistics course (25 in section A, 42 in section B, and 18 in section C). We count, for example, the number of cars arriving at Exit 25 on I-4, and we count the number of statistics students in each section. Notice that a home can have 3 or 4 bedrooms, but it cannot have 3.56 bedrooms. Thus, there is a "gap" between possible values. Typically, discrete variables result from counting.

Observations of a **continuous variable** can assume any value within a specific range. Examples of continuous variables are the air pressure in a tire and the weight of a shipment of tomatoes. Other examples are the amount of raisin bran in a box and the duration of flights from Orlando to San Diego. Grade point average (GPA) is a continuous variable. We could report the GPA of a particular student as 3.2576952. The usual practice is to round to 3 places—3.258. Typically, continuous variables result from measuring.

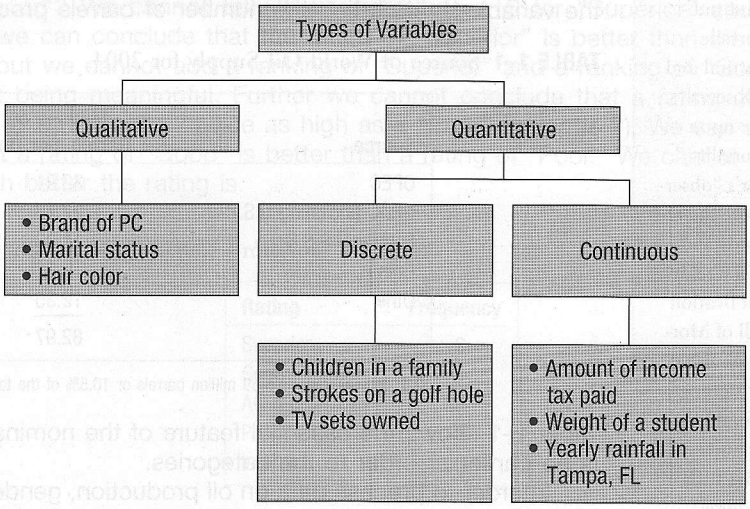
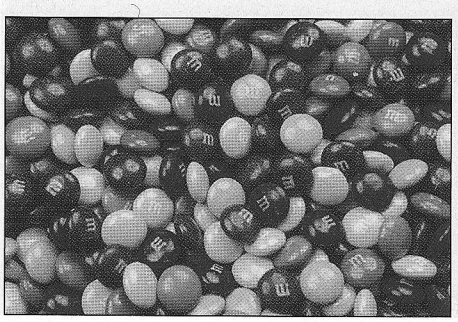


CHART 1–2 Summary of the Types of Variables

Levels of Measurement

Data can be classified according to levels of measurement. The level of measurement of the data dictates the calculations that can be done to summarize and present the data. It will also determine the statistical tests that should be performed. For example, there are six colors of candies in a bag of M&M's. Suppose we assign brown a value of 1, yellow 2, blue 3, orange 4, green 5, and red 6. From a bag of candies, we add the assigned color values and divide by the number of candies and report that the mean color is 3.56. Does this mean that the average color is blue or orange? Of course not! As a second example, in a high school track meet there are eight competitors in the 400-meter run. We report the order of finish and that the mean finish is 4.5. What does the mean finish tell us? Nothing! In both of these instances, we have not properly used the level of measurement.



There are actually four levels of measurement: nominal, ordinal, interval, and ratio. The lowest, or the most primitive, measurement is the nominal level. The highest, or the level that gives us the most information about the observation, is the ratio level of measurement.

Nominal-Level Data

For the **nominal level** of measurement observations of a qualitative variable can only be classified and counted. There is no particular order to the labels. The classification of the six colors of M&M's milk chocolate candies is an example of the nominal level of measurement. We simply classify the candies by color. There is no natural order. That is, we could report the brown candies first, the orange first, or any of the colors first. Gender is another example of the nominal level of measurement. Suppose we count the number of students entering a football game with a student ID and report how many are men and how many are women. We could report either the men or the women first. For the nominal level the only measurement involved consists of counts. Table 1–1 shows a breakdown of the sources of the world oil supply. The variable of interest is the country or region. This is a nominal-level variable because we record the information by source of the oil supply and there is no natural order. Do not be distracted by the fact that we summarize the variable by reporting the number of barrels produced per day.

Statistics in Action

Where did statistics get its start? In 1662 John Graunt published an article called "Natural and Political Observations Made upon Bills of Mortality." The author's "observations" were the result of a study and analysis of a weekly church publication called "Bill of Mortality," which listed births, christenings, and deaths and their causes. Graunt realized that the Bills of Mortality represented only a fraction of all births and deaths in London. However, he used the data to reach broad conclusions about the impact of disease, such as the plague, on the general population. His logic is an example of statistical inference. His analysis and interpretation of the data are thought to mark the start of statistics.

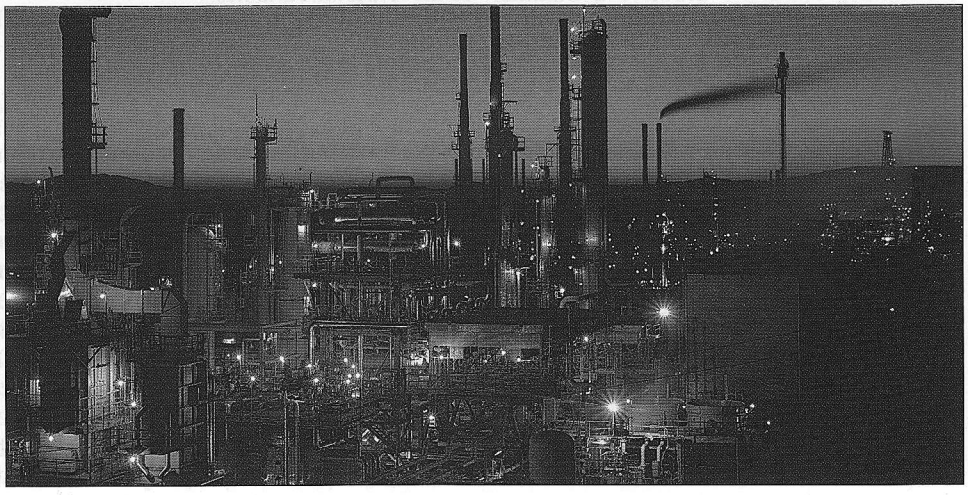
TABLE 1–1 Source of World Oil Supply for 2004.

Source	Millions of Barrels per Day	Percent
OPEC	32.91	39.7
OEDC (including U.S.)*	22.76	27.4
Former U.S.S.R.	11.33	13.7
China	3.62	4.4
Other	12.35	14.9
	82.97	100.1

*U.S. daily average is 8.69 million barrels or 10.5% of the total.

Table 1–1 shows the essential feature of the nominal scale of measurement: There is no particular order to the categories.

In order to process data on oil production, gender, employment by industry, and so forth, the categories are often numerically coded 1, 2, 3, and so on, with 1 rep-



resenting OPEC, 2 representing OECD, for example. This facilitates counting by the computer. However, because we have assigned numbers to the various categories, this does not give us license to manipulate the numbers. For example, 1 + 2 does not equal 3, that is, OPEC + OEDC does not equal former U.S.S.R. To summarize, the nominal-level data have the following properties:

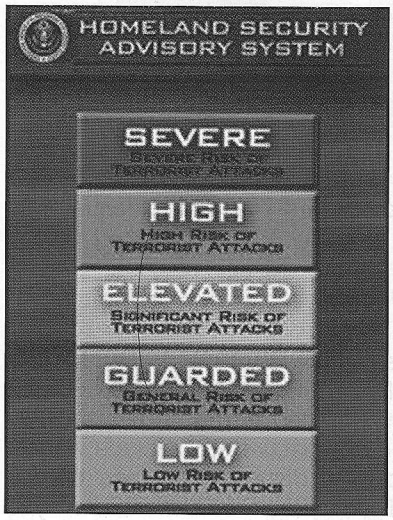
1. Data categories are represented by labels or names.
2. Even when the labels are numerically coded, the data categories have no logical order.

Ordinal-Level Data

The next higher level of data is the **ordinal level.** Table 1–2 lists the student ratings of Professor James Brunner in an Introduction to Finance course. Each student in the class answered the question "Overall, how did you rate the instructor in this class?" The variable rating illustrates the use of the ordinal scale of measurement. One classification is "higher" or "better" than the next one. That is, "Superior" is better than "Good," "Good" is better than "Average," and so on. However, we are not able to distinguish the magnitude of the differences between groups. Is the difference between "Superior" and "Good" the same as the difference between "Poor" and "Inferior"? We cannot tell. If we substitute a 5 for "Superior" and a 4 for "Good," we can conclude that the rating of "Superior" is better than the rating of "Good," but we cannot add a ranking of "Superior" and a ranking of "Good," with the result being meaningful. Further we cannot conclude that a rating of "Good" (rating is 4) is necessarily twice as high as a "Poor" (rating is 2). We can only conclude that a rating of "Good" is better than a rating of "Poor." We cannot conclude how much better the rating is.

TABLE 1–2 Rating of a Finance Professor

Rating	Frequency
Superior	6
Good	28
Average	25
Poor	12
Inferior	3



Another example of ordinal-level data is the Homeland Security Advisory System. The Department of Homeland Security publishes this information regarding the risk of terrorist activity to federal, state, and local authorities and to the American people. The five risk levels from lowest to highest including a description and color codes are shown to the left.

This is an example of the ordinal scale because we know the order or the ranks of the risk levels—that is, orange is higher than yellow—but the amount of the difference in risk is not necessarily the same. To put it another way, the difference in the risk level between yellow and orange is not necessarily the same as between green and blue. You can check the current risk level and read more about the various levels by going to: www.whitehouse.gov/homeland.

In summary the properties of the ordinal level of data are:

1. Data classifications are represented by sets of labels or names (high, medium, low) that have relative values.
2. Because of the relative values, the data classified can be ranked or ordered.

Interval-Level Data

The **interval level** of measurement is the next highest level. It includes all the characteristics of the ordinal level, but, in addition, the difference between values is a constant size. An example of the interval level of measurement is temperature. Suppose the high temperatures on three consecutive winter days in Boston are 28, 31, and 20 degrees Fahrenheit. These temperatures can be easily ranked, but we can also determine the difference between temperatures. This is possible because 1 degree Fahrenheit represents a constant unit of measurement. Equal differences between two temperatures are the same, regardless of their position on the scale. That is, the difference between 10 degrees Fahrenheit and 15 degrees is 5, the difference between 50 and 55 degrees is also 5 degrees. It is also important to note that 0 is just a point on the scale. It does not represent the absence of the condition. Zero degrees Fahrenheit does not represent the absence of heat, just that it is cold! In fact 0 degrees Fahrenheit is about −18 degrees on the Celsius scale.

Another example of the interval scale of measurement is women's dress sizes. Listed below is information on several dimensions of a standard U.S. women's dress.

Size	Bust (in)	Waist (in)	Hips (in)
8	32	24	35
10	34	26	37
12	36	28	39
14	38	30	41
16	40	32	43
18	42	34	45
20	44	36	47
22	46	38	49
24	48	40	51
26	50	42	53
28	52	44	55

Why is the "size" scale an interval measurement? Observe as the size changes by 2 units (say from size 10 to size 12 or from size 24 to size 26) each of the measurements increases by 2 inches. To put it another way the intervals are the same.

There is no natural zero point for dress size. A "size 0" dress does not have "zero" material. Instead it would have a 24-inch bust, 16-inch waist and 27-inch hips. Moreover, the ratios are not reasonable. If you divide a size 28 by a size 14, you do not get the same answer as dividing a size 20 by 10. Neither ratio is equal to two as the "size" number would suggest. In short, if the distances between the numbers make sense, but the ratios do not, then you have an interval scale of measurement.

The properties of the interval-level data are:

1. Data classifications are ordered according to the amount of the characteristic they possess.
2. Equal differences in the characteristic are represented by equal differences in the measurements.

Ratio-Level Data

Practically all quantitative data is recorded on the ratio level of measurement. The **ratio level** is the "highest" level of measurement. It has all the characteristics of the interval level, but in addition, the 0 point is meaningful and the ratio between two numbers is meaningful. Examples of the ratio scale of measurement include wages, units of production, weight, changes in stock prices, distance between branch offices, and height. Money is a good illustration. If you have zero dollars, then you have no money. Weight is another example. If the dial on the scale of a correctly calibrated device is at 0, then there is a complete absence of weight. The ratio of

two numbers is also meaningful. If Jim earns $40,000 per year selling insurance and Rob earns $80,000 per year selling cars, then Rob earns twice as much as Jim.

Table 1–3 illustrates the use of the ratio scale of measurement. It shows the incomes of four father-and-son combinations.

TABLE 1–3 Father–Son Income Combinations

Name	Father	Son
Lahey	$80,000	$ 40,000
Nale	90,000	30,000
Rho	60,000	120,000
Steele	75,000	130,000

Observe that the senior Lahey earns twice as much as his son. In the Rho family the son makes twice as much as the father.

In summary, the properties of the ratio-level data are:

1. Data classifications are ordered according to the amount of the characteristics they possess.
2. Equal differences in the characteristic are represented by equal differences in the numbers assigned to the classifications.
3. The zero point is the absence of the characteristic and the ratio between two numbers is meaningful.

Chart 1–3 summarizes the major characteristics of the various levels of measurement.

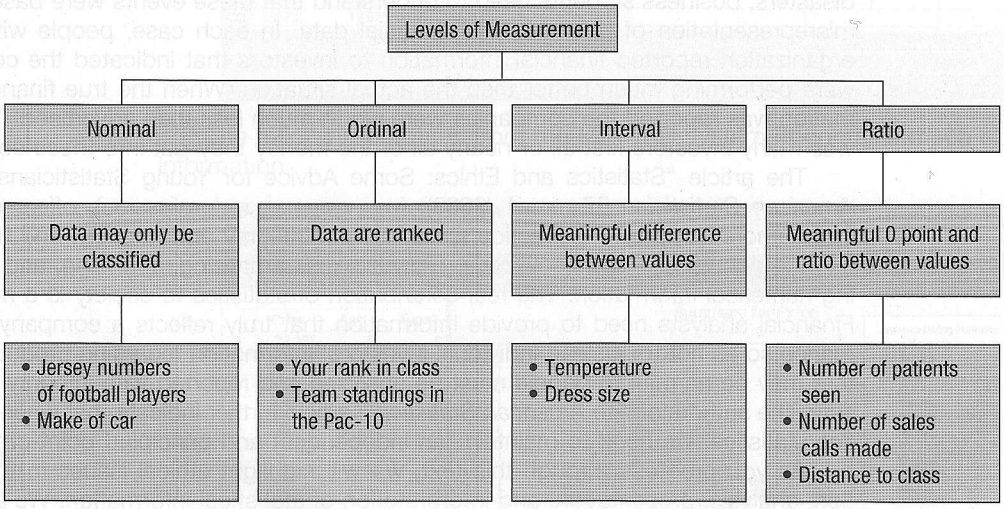


CHART 1–3 Summary of the Characteristics for Levels of Measurement

Self-Review 1–2

What is the level of measurement reflected by the following data?

(a) The age of each person in a sample of 50 adults who listen to one of the 1,230 talk radio stations in the United States is:

35	29	41	34	44	46	42	42	37	47
30	36	41	39	44	39	43	43	44	40
47	37	41	27	33	33	39	38	43	22
44	39	35	35	41	42	37	42	38	43
35	37	38	43	40	48	42	31	51	34

(b) In a survey of 200 luxury-car owners, 100 were from California, 50 from New York, 30 from Illinois, and 20 from Ohio.

Exercises

The answers to the odd-numbered exercises are at the end of the book.

1. What is the level of measurement for each of the following variables?
   a. Student IQ ratings.
   b. Distance students travel to class.
   c. Student scores on the first statistics test.
   d. A classification of students by state of birth.
   e. A ranking of students as freshman, sophomore, junior, and senior.
   f. Number of hours students study per week.
2. What is the level of measurement for these items related to the newspaper business?
   a. The number of papers sold each Sunday during 2006.
   b. The departments, such as editorial, advertising, sports, etc.
   c. A summary of the number of papers sold by county.
   d. The number of years with the paper for each employee.
3. Look in the latest edition of *USA Today* or your local newspaper and find examples of each level of measurement. Write a brief memo summarizing your findings.
4. For each of the following, determine whether the group is a sample or a population.
   a. The participants in a study of a new cholesterol drug.
   b. The drivers who received a speeding ticket in Kansas City last month.
   c. Those on welfare in Cook County (Chicago), Illinois.
   d. The 30 stocks reported as a part of the Dow Jones Industrial Average.

Ethics and Statistics

Following the events of Enron, Tyco, HealthSouth, WorldCom, and other corporate disasters, business students need to understand that these events were based on the misrepresentation of business and financial data. In each case, people within each organization reported financial information to investors that indicated the companies were performing much better than the actual situation. When the true financial information was reported the companies were worth much less than advertised. The result was many investors lost all or nearly all of the money they put into these companies.

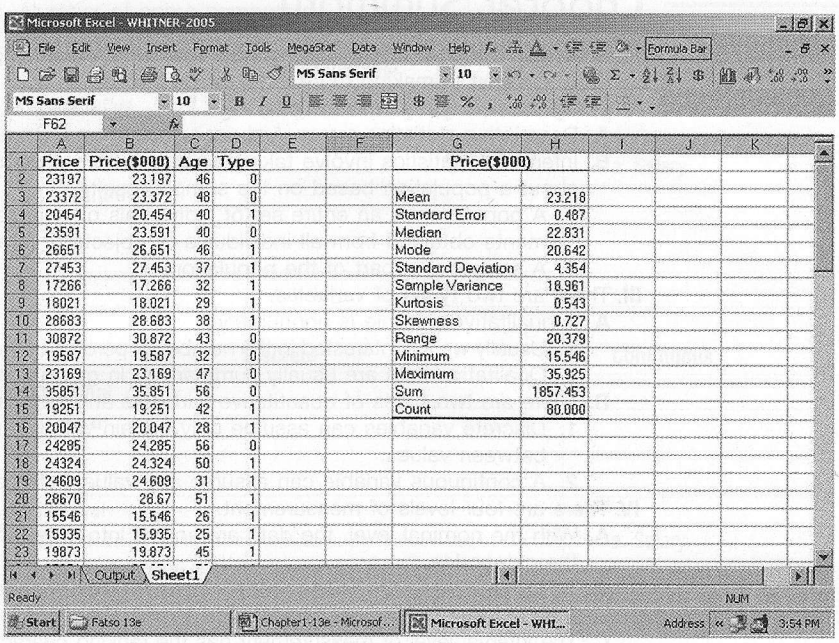
The article "Statistics and Ethics: Some Advice for Young Statisticians," in *The American Statistician* 57, no. 1 (2003), (www.amstat.org/profession), offers guidance. The authors advise us to practice statistics with integrity and honesty, and urge us to "do the right thing" when collecting, organizing, summarizing, analyzing, and interpreting numerical information. The real contribution of statistics to society is a moral one. Financial analysts need to provide information that truly reflects a company's performance so as not to mislead individual investors. Information regarding product defects that may be harmful to people must be analyzed and reported with integrity and honesty. The authors of *The American Statistician* article further indicate that when we practice statistics, we need to maintain "an independent and principled point-of-view."

As you progress through this text, we will highlight ethical issues in the collection, analysis, presentation, and interpretation of statistical information. We also hope that, as your learn about using statistics, you will become a more informed consumer of information. For example, you will question a report based on data that do not fairly represent the population, a report that does not include all relevant statistics, one that includes an incorrect choice of statistical measures, or a presentation that introduces the writer's bias in a deliberate attempt to mislead or misrepresent.

Computer Applications

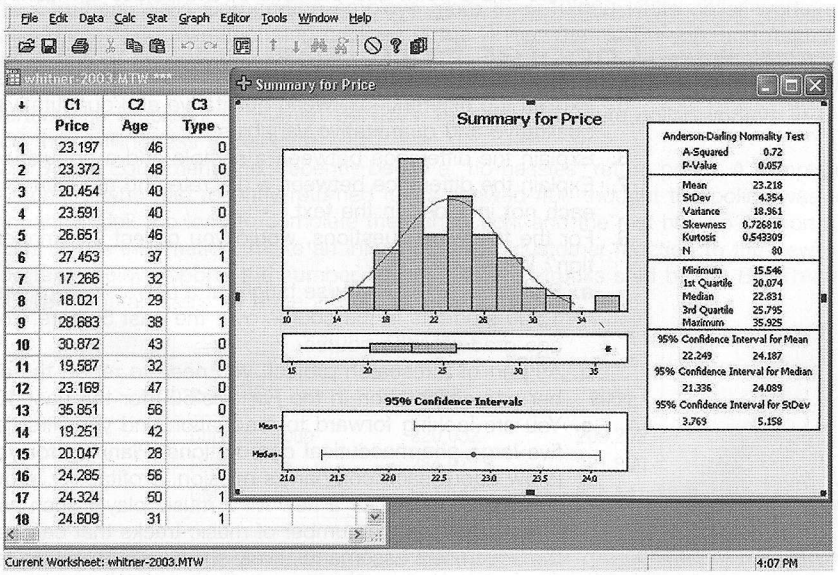
Computers are now available to students at most colleges and universities. Spreadsheets, such as Microsoft Excel, and statistical software packages, such as MINITAB, are available in most computer labs. The Microsoft Excel package is bundled with many home computers. In this text we use both Excel and MINITAB for the applications. We also use an Excel add-in called MegaStat. This add-in gives Excel the capability to produce additional statistical reports.

The following example shows the application of computers in statistical analysis. In Chapters 2, 3, and 4 we illustrate methods for summarizing and describing data. An example used in those chapters refers to the price reported in thousands of dollars of 80 vehicles sold last month at Whitner Autoplex. The following Excel output reveals, among other things, that (1) 80 vehicles were sold last month, (2) the mean (average) selling price was $23,218, and (3) the selling prices ranged from a minimum of $15,546 to a maximum of $35,925.



	A	B	C	D	E	F	G	H	I	J	K
1	Price	Price($000)	Age	Type			Price($000)				
2	23197	23.197	46	0							
3	23372	23.372	48	0			Mean	23.218			
4	20454	20.454	40	1			Standard Error	0.487			
5	23591	23.591	40	0			Median	22.831			
6	26651	26.651	46	1			Mode	20.642			
7	27453	27.453	37	1			Standard Deviation	4.354			
8	17266	17.266	32	1			Sample Variance	18.961			
9	18021	18.021	29	1			Kurtosis	0.543			
10	28683	28.683	38	1			Skewness	0.727			
11	30872	30.872	43	0			Range	20.379			
12	19587	19.587	32	0			Minimum	15.546			
13	23169	23.169	47	0			Maximum	35.925			
14	35851	35.851	56	0			Sum	1857.453			
15	19251	19.251	42	1			Count	80.000			
16	20047	20.047	28	1							
17	24285	24.285	56	0							
18	24324	24.324	50	1							
19	24609	24.609	31	1							
20	28670	28.67	51	1							
21	15546	15.546	26	1							
22	15935	15.935	25	1							
23	19873	19.873	45	1							

The following output is from the MINITAB system. It contains much of the same information.



	C1	C2	C3
	Price	Age	Type
1	23.197	46	0
2	23.372	48	0
3	20.454	40	1
4	23.591	40	0
5	26.651	46	1
6	27.453	37	1
7	17.266	32	1
8	18.021	29	1
9	28.683	38	1
10	30.872	43	0
11	19.587	32	0
12	23.169	47	0
13	35.851	56	0
14	19.251	42	1
15	20.047	28	1
16	24.285	56	0
17	24.324	50	1
18	24.609	31	1

Had we used a calculator to arrive at these measures and others needed to fully analyze the selling prices, hours of calculation would have been required. The likelihood of an error in arithmetic is high when a large number of values are concerned. On the other hand, statistical software packages and spreadsheets can provide accurate information in seconds.

At the option of your instructor, and depending on the software system available, we urge you to apply a computer package to the exercises in the **Data Set Exercises** section in each chapter. It will relieve you of the tedious calculations and allow you to concentrate on data analysis.

Chapter Summary

I. Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting data to assist in making more effective decisions.
II. There are two types of statistics.
  A. Descriptive statistics are procedures used to organize and summarize data.
  B. Inferential statistics involve taking a sample from a population and making estimates about a population based on the sample results.
    1. A population is an entire set of individuals or objects of interest or the measurements obtained from all individuals or objects of interest.
    2. A sample is a part of the population.
III. There are two types of variables.
  A. A qualitative variable is nonnumeric.
    1. Usually we are interested in the number or percent of the observations in each category.
    2. Qualitative data are usually summarized in graphs and bar charts.
  B. There are two types of quantitative variables and they are usually reported numerically.
    1. Discrete variables can assume only certain values, and there are usually gaps between values.
    2. A continuous variable can assume any value within a specified range.
IV. There are four levels of measurement.
  A. With the nominal level, the data are sorted into categories with no particular order to the categories.
  B. The ordinal level of measurement presumes that one classification is ranked higher than another.
  C. The interval level of measurement has the ranking characteristic of the ordinal level of measurement plus the characteristic that the distance between values is a constant size.
  D. The ratio level of measurement has all the characteristics of the interval level, plus there is a 0 point and the ratio of two values is meaningful.

Chapter Exercises

5. Explain the difference between qualitative and quantitative variables. Give an example of qualitative and quantitative variables.
6. Explain the difference between a sample and a population.
7. Explain the difference between a discrete and a continuous variable. Give an example of each not included in the text.
8. For the following questions, would you collect information using a sample or a population? Why?
  a. Statistics 201 is a course taught at a university. Professor A. Verage has taught nearly 1,500 students in the course over the past 5 years. You would like to know the average grade for the course.
  b. As part of a research project, you need to report the average profitability of the number one corporation in the Fortune 500 for the past 10 years.
  c. You are looking forward to graduation and your first job as a salesperson for one of five large pharmaceutical corporations. Planning for your interviews, you will need to know about each company's mission, profitability, products, and markets.
  d. You are shopping for a new MP3 music player such as the Apple iPod. The manufacturers advertise the number of music tracks that can be stored in the memory. Usually, the advertisers assume relatively short, popular music to estimate the number of tracks that can be stored. You, however, like Broadway musical tunes and they are much longer. You would like to estimate how many Broadway tunes will fit on your MP3 player.
9. Place these variables in the following classification tables. For each table, summarize your observations and evaluate if the results are generally true. For example, salary is reported as a continuous quantitative variable. It is also a continuous ratio scaled variable.

a. Salary
b. Gender
c. Sales volume of MP3 players
d. Soft drink preference
e. Temperature
f. SAT scores
g. Student rank in class
h. Rating of a finance professor
i. Number of home computers

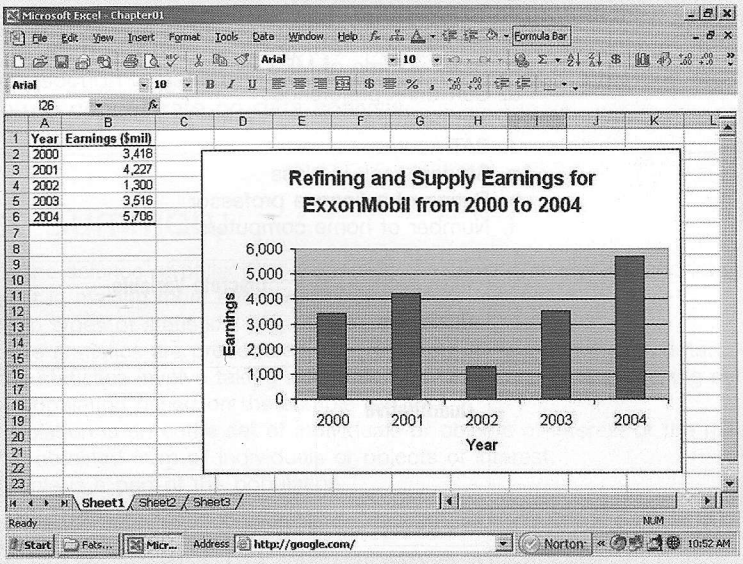
	Discrete Variable	Continuous Variable
Qualitative		
Quantitative		a. Salary

	Discrete	Continuous
Nominal		
Ordinal		
Interval		
Ratio		a. Salary

10. Using data from such publications as the *Statistical Abstract of the United States, The World Almanac, Forbes,* or your local newspaper, give examples of the nominal, ordinal, interval, and ratio levels of measurement.
11. The Struthers Wells Corporation employs more than 10,000 white collar workers in its sales offices and manufacturing facilities in the United States, Europe, and Asia. A sample of 300 of these workers revealed 120 would accept a transfer to a location outside the United States. On the basis of these findings, write a brief memo to Ms. Wanda Carter, Vice President of Human Services, regarding all white collar workers in the firm and their willingness to relocate.
12. AVX Stereo Equipment, Inc., recently began a "no-hassles" return policy. A sample of 500 customers who had recently returned items showed 400 thought the policy was fair, 32 thought it took too long to complete the transaction, and the rest had no opinion. On the basis of this information, make an inference about customer reaction to the new policy.
13. The table below reports the number of cars and light trucks sold by the Big Three automobile manufacturers for June 2004 and June 2005.

	Units	
Company	**2005**	**2004**
Chrysler Group	220,032	209,252
Ford	284,971	281,850
GM	551,141	375,141

   a. Compare the total sales in the two months. What do you conclude? Has there been an increase in sales?
   b. Compare the percent of the Big Three market for each company. Did the market increase or did GM steal sales from the other companies? Cite evidence.
14. The following chart depicts the earnings in millions of dollars for ExxonMobil for the period from 2000 to 2004.



Year	Earnings ($mil)
2000	3,418
2001	4,227
2002	1,300
2003	3,516
2004	5,706

Write a brief report discussing earnings at ExxonMobil during the period. Did earnings increase or decrease over the period?

exercises.com

These exercises use the World Wide Web, a rich and growing source of information. Because of the changing nature and the continuous revision of websites, you may well see different menus, and the exact addresses, or URLs, may change. When you visit a page, be prepared to search the link.

15. Suppose you recently opened an account with Ameritrade, Inc., an online broker. You have decided to purchase shares of either Johnson & Johnson (a pharmaceutical company) or PepsiCo (the parent company of Pepsi and Frito-Lay). For a comparison of the two companies go to http://finance.yahoo.com and in the space where it says **Get Quotes** enter the letters JNJ and PEP, which are the respective symbols for the two companies. Click on **Go** and you should receive some current information about the selling price of the two stocks. To the right of this information click on **More** and then click on **Analyst Opinion.** Here you will find information from stock analysts evaluating these stocks. Brokers rate the stock a 1 if it is a strong buy and a 5 if it is a strong sell. What level of measurement is this information? Which of the stocks would you recommend?

Data Set Exercises

16. Refer to the Real Estate data at the back of the text, which report information on homes sold in the Denver, Colorado, area last year. Consider the following variables: selling price, number of bedrooms, township, and distance from the center of the city.
- **a.** Which of the variables are qualitative and which are quantitative?
- **b.** Determine the level of measurement for each of the variables.

17. Refer to the Global Financial Performance Data set which reports financial information on 148 global companies. Consider the following variables: Name, economic sector, industry type, employees, sales, cost of sales, gross profit, gross profit margin, total taxes, net income, return on assets, and return on investment.
- **a.** Which of these variables are quantitative and which are qualitative?
- **b.** Determine the level of measurement for each variable.

18. Refer to the Wage data, which report information on annual wages for a sample of 100 workers. Also included are variables relating to industry, years of education, and gender for each worker.
- **a.** Which of the 12 variables are qualitative and which are quantitative?
- **b.** Determine the level of measurement for each variable.

19. Refer to the CIA data, which report demographic and economic information on 62 countries.
- **a.** Which of the variables are quantitative and which are qualitative?
- **b.** Determine the level of measurement for each of the variables.

Chapter 1 Answers to Self-Review

1–1 **a.** On the basis of the sample of 1,960 consumers, we estimate that, if it is marketed, 60 percent of all consumers will purchase the chicken dinner $(1{,}176/1{,}960) \times 100 = 60$ percent.

b. Inferential statistics, because a sample was used to draw a conclusion about how all consumers in the population would react if the chicken dinner were marketed.

1–2 **a.** Age is a ratio-scale variable. A 40-year-old is twice as old as someone 20 years old.

b. Nominal scale. We could arrange the states in any order.