73. **Attitude Update.** The Pew Research Center for the People and the Press studies public attitudes toward the press, politics, and public policy issues. Go to its Web site and find the latest survey about attitudes. Write a one-page summary of what Pew surveyed, how it conducted the survey, and what it found.

74. **Labor Statistics.** Use the Bureau of Labor Statistics Web page to learn about its monthly survey. Choose one aspect of the survey, such as how the sample is chosen or how it is used to compare unemployment rates over time. Write a short summary of what you learn.

75. **Professional Polling.** Visit the Web site of a national polling organization and report on a recent poll. Write a short description of the poll and its results, commenting on features such as sampling technique, sample size, and margin of error.

IN THE NEWS

76. **Statistics in the News.** Select three news stories from the past week that involve statistics in some way. In each case, write one or two paragraphs describing the role of statistics in the story.

77. **Statistics in Your Major.** Write two to three paragraphs describing the ways in which you think the science of statistics is important in your major field of study. (If you have not chosen a major, answer this question for a major that you are considering.)

78. **Statistics in Sports.** Choose a sport and describe three different statistics commonly tracked by participants in or spectators of the sport. In each case, briefly describe the importance of the statistic to the sport.

79. **Sample and Population.** Find a report in today's news concerning any type of statistical study. What is the population being studied? What is the sample? Why do you think the sample was chosen as it was?

80. **Poor Sampling.** In a recent newspaper or magazine, find an article about a study that attempts to describe some characteristic of a population, but that you believe involved poor sampling (for example, a sample that was too small or unrepresentative of the population under study). Describe the population, the sample, and what you think was wrong with the sample. Briefly discuss how you think the poor sampling affected the study results.

81. **Good Sampling.** In a recent newspaper or magazine, find an article that describes a statistical study in which the sample was well chosen. Describe the population, the sample, and why you think the sample was a good one.

82. **Margin of Error.** Find a report of a recent survey or poll. Interpret the sample statistic and margin of error quoted for the survey or poll.

5 UNIT 5B Should You Believe a Statistical Study?

Most statistical research is carried out with integrity and care. Nevertheless, statistical research is sufficiently complex that bias can arise in many different ways. We should always examine reports of statistical research carefully, looking for anything that might make us question the results. In this unit, we discuss eight guidelines that can help you answer the question "Should I believe a statistical study?"

Guideline 1: Identify the Goal, Population, and Type of Study

Before evaluating the details of a statistical study, we must know what it is about. Based on what you hear or read, try to answer basic questions such as these:

- What was the goal of the study?
- What was the population under study? Was the population clearly and appropriately defined?
- What type of study was used? Was the type appropriate for the goal?

If you can't find reasonable answers to these questions, it will be difficult to evaluate other aspects of the study.

By the Way

Surveys show that nearly half of Americans believe their horoscopes. However in controlled experiments, the predictions of horoscopes come true no more often than would be expected by chance.

✻EXAMPLE 1 *Appropriate Type of Study?*

A newspaper reports: "Researchers gave each of the 100 participants their astrological horoscopes, and asked them whether the horoscopes appeared to be accurate. Eighty-five percent of the participants reported that the horoscopes were accurate. The researchers concluded that horoscopes are valid most of the time." Analyze this study according to Guideline 1.

SOLUTION The goal of the study was to determine the validity of horoscopes. Based on the news report, it appears that the study was *observational:* The researchers simply asked the participants about the accuracy of the horoscopes. However, because the accuracy of a horoscope is somewhat subjective, this study should have been a controlled experiment in which some people were given their actual horoscope and others were given a fake horoscope. Then the researchers could have looked for differences between the two groups. Moreover, because researchers could easily influence the results by how they questioned the participants, the experiment should have been double-blind. In summary, the type of study was inappropriate to the goal and its results are meaningless.

Now try Exercises 19–20.

Guideline 2: Consider the Source

Statistical studies are supposed to be objective, but the people who carry them out and fund them may be biased. Thus, it is important to consider the source of a study and evaluate the potential for biases that might invalidate its conclusions.

✻EXAMPLE 2 *Is Smoking Healthy?*

By 1963, enough research on the health dangers of smoking had accumulated that the Surgeon General of the United States publicly announced that smoking is bad for health. Research done since that time has built further support for this claim. However, while the vast majority of studies show that smoking is unhealthy, a few studies found no dangers from smoking, and perhaps even health *benefits*. These studies generally were carried out by the Tobacco Research Institute, funded by the tobacco companies. Analyze the Tobacco Research Institute studies according to Guideline 2.

SOLUTION Tobacco companies had a financial interest in minimizing the dangers of smoking. Because the studies carried out at the Tobacco Research Institute were funded by the tobacco companies, there may have been pressure on the researchers to produce results to the companies' liking. This *potential* for bias does not mean their research *was* biased, but the fact that it contradicts virtually all other research on the subject should be cause for concern.

Now try Exercises 21–22.

Guideline 3: Look for Bias in the Sample

Look for bias that may prevent the sample from being representative of the population. There are two particularly common forms of bias that can affect sample selection.

By the Way

After decades of arguing to the contrary in 1999 the Philip Morris Company—the world's largest seller of tobacco products—publicly acknowledged that smoking causes lung cancer, heart disease, emphysema, and other serious diseases. Shortly thereafter, Philip Morris changed its name to Altria.

BIAS IN CHOOSING A SAMPLE

Selection bias occurs whenever researchers *select* their sample in a way that tends to make it unrepresentative of the population. For example, a pre-election poll that surveys only registered Republicans has selection bias because it is unlikely to reflect the opinions of all voters.

Participation bias occurs primarily with surveys and polls; it arises whenever people *choose* whether to participate. Because people who feel strongly about an issue are more likely to participate, their opinions may not represent the larger population that is less emotionally attached to the issue. (Surveys or polls in which people choose whether to participate are often called *self-selected* or *voluntary response surveys*.)

CASE STUDY *The 1936 Literary Digest Poll*

HISTORICAL NOTE

A young pollster named George Gallup conducted his own survey prior to the 1936 election. Sending postcards to only 3000 randomly selected people, he correctly predicted not only the outcome of the election, but also the outcome of the *Literary Digest* poll to within 1%. Gallup went on to establish a very successful polling organization.

The *Literary Digest*, a popular magazine of the 1930s, successfully predicted the outcomes of several elections using large polls. In 1936, editors of the *Literary Digest* conducted a particularly large poll in advance of the presidential election. They randomly chose a sample of 10 million people from various lists, including names in telephone books and rosters of country clubs. They mailed a postcard "ballot" to each of these 10 million people. About 2.4 million people returned the postcard ballots. Based on the returned ballots, the editors of the *Literary Digest* predicted that Alf Landon would win the presidency by a margin of 57% to 43% over Franklin Roosevelt. Instead, Roosevelt won with 62% of the popular vote. How did such a large survey go so wrong?

The sample suffered from both selection bias and participation bias. The selection bias arose because the *Literary Digest* chose its 10 million names in ways that favored affluent people. For example, selecting names from telephone books meant choosing only from those who could afford telephones back in 1936. Similarly, country club members are usually quite wealthy. The selection bias favored Landon because he was the Republican, and affluent voters of the 1930s tended to vote for Republican candidates.

The participation bias arose because return of the postcard ballots was voluntary. People who felt most strongly about the election were more likely to be among those who returned their postcard ballots. This bias also tended to favor Landon because he was the challenger—people who did not like President Roosevelt could express their desire for change by returning the postcards. Together, the two forms of bias made the sample results useless, despite the large number of people surveyed.

EXAMPLE 3 *Self-Selected Poll*

The television show *Nightline* conducted a poll in which viewers were asked whether the United Nations headquarters should be kept in the United States. Viewers could respond to the poll by paying 50 cents to call a "900" phone number with their opinions. The poll drew 186,000 responses, of which 67% favored moving the United Nations out of the United States. Around the same time, a poll using simple random sampling of 500 people found that 72% wanted the United Nations to *stay* in the United States. Which poll is more likely to be representative of the general opinions of Americans?

SOLUTION The *Nightline* sample suffered from severe participation bias. Not only did viewers choose whether to call in for the survey, but they had to *pay* to participate. This cost made it even more likely that respondents would be those who felt a need for change. Thus, despite its large number of respondents, the *Nightline* survey was too biased to be trusted. In contrast, a simple random sample of 500 people is quite likely to be representative, so the finding of this small survey has a better chance of representing the true opinions of all Americans. Now try Exercises 23–24.

By the Way

More than a third of all Americans routinely shut the door or hang up the phone when contacted for a survey, thereby making self-selection a problem for legitimate pollsters. One reason people hang up may be the proliferation of selling under the guise of market research (often called "sugging") in which a telemarketer pretends you are part of a survey in order to get you to buy something.

Guideline 4: Look for Problems in Defining or Measuring the Variables of Interest

Statistical studies usually attempt to measure *something*, and we call the things being measured the **variables** of interest in the study. The term *variable* simply refers to an item or quantity that can vary or take on different values. For example, variables in the Nielsen ratings include *show being watched* and *number of viewers*.

DEFINITION

A **variable** is any item or quantity that can vary or take on different values. The *variables of interest* in a statistical study are the items or quantities that the study seeks to measure.

Results of a statistical study may be especially difficult to interpret if the variables under study are difficult to define or measure. For example, imagine trying to conduct a study of how exercise affects resting heart rates. The variables of interest would be *amount of exercise* and *resting heart rate*. However, both variables are difficult to define and measure. In the case of *amount of exercise*, it's not clear what the definition covers: Does it include walking to class? Even if we specify the definition, how can we measure *amount of exercise* given that some forms of exercise are more vigorous than others? The following two examples describe real cases in which defining or measuring variables caused problems in statistical studies.

Time out to think

How would you measure your resting heart rate? Describe some difficulties in defining and measuring *resting heart rate*.

✻EXAMPLE 4 *Can Money Buy Love?*

A Roper poll reported in *USA Today* involved a survey of the wealthiest 1% of Americans. The survey found that these people would pay an average of $487,000 for "true love," $407,000 for "great intellect," $285,000 for "talent," and $259,000 for "eternal youth." Analyze this result according to Guideline 4.

SOLUTION The variables in this study are very difficult to define. How, for example, do you define "true love"? And does it mean true love for a day, a lifetime, or something else? Similarly, does the ability to balance a spoon on your nose constitute "talent"? Because the variables are so poorly defined, it's likely that different people interpreted them differently, making the results very difficult to interpret.

Now try Exercise 25.

✻EXAMPLE 5 *Illegal Drug Supply*

Law enforcement authorities try to stop illegal drugs from entering the country. A commonly quoted statistic is that they succeed in stopping only about 10% to 20% of the drugs entering the United States. Should you believe this statistic?

SOLUTION There are essentially two variables in the study: *quantity of illegal drugs intercepted* and *quantity of illegal drugs NOT intercepted.* It should be relatively easy to measure the quantity of illegal drugs that law enforcement officials intercept. However, because the drugs are illegal, it's unlikely that anyone is reporting the quantity of drugs that are *not* intercepted. How, then, can anyone know that the intercepted drugs are 10% to 20% of the total? In a *New York Times* analysis, a police officer was quoted as saying that his colleagues refer to this type of statistic as "P.F.A.," for "pulled from the air."

Now try Exercise 26.

B W

Many hardware stores sell simple kits that you can use to test whether radon gas is accumulating in your home. If it is, the problem can be eliminated by installing an appropriate "radon mitigation" system, which usually consists of a fan that blows the radon out from under the house before it can get in.

Guideline 3: Watch Out for Confounding Variables

Variables that are *not intended* to be part of the study can sometimes make it difficult to interpret results properly. Such variables are often called *confounding* variables, because they confound (confuse) a study's results.

It's not always easy to discover confounding variables. Sometimes they are discovered years after a study was completed, and sometimes they are not discovered at all. Fortunately, confounding variables are sometimes more obvious and can be discovered simply by thinking hard about factors that may have influenced a study's results.

✻EXAMPLE 6 *Radon and Lung Cancer*

Radon is a radioactive gas produced by natural processes (the decay of uranium) in the ground. The gas can leach into buildings through the foundation and can accumulate in relatively high concentrations if doors and windows are closed. Imagine a study that seeks to determine whether radon gas causes lung cancer by comparing the lung cancer rate in Colorado, where radon gas is fairly common, with the lung cancer rate in Hong Kong, where radon gas is less common. Suppose the study finds that the

lung cancer rates are nearly the same. Is it fair to conclude that radon is *not* a significant cause of lung cancer?

SOLUTION The variables under study are *amount of radon* and *lung cancer rate*. However, because smoking can also cause lung cancer, *smoking rate* may be a confounding variable in this study. In particular, the smoking rate in Hong Kong is much higher than the smoking rate in Colorado, so any conclusions about radon and lung cancer must take the smoking rate into account. In fact, careful studies have shown that radon gas *can* cause lung cancer, and the U.S. Environmental Protection Agency (EPA) recommends taking steps to prevent radon from building up indoors.

Now try Exercises 27–28.

Guideline 6: Consider the Setting and Wording in Surveys

Even when a survey is conducted with proper sampling and with clearly defined terms and questions, it's important to watch out for problems in the setting or wording that might produce inaccurate or dishonest responses. Dishonest responses are particularly likely when the survey concerns sensitive subjects, such as personal habits or income. For example, the question "Do you cheat on your income taxes?" is unlikely to elicit honest answers from those who cheat, especially if the setting does not guarantee complete confidentiality.

In other cases, even honest answers may not be accurate if the wording of questions invites bias. Sometimes just the order of the words in a question can affect the outcome. A poll conducted in Germany asked the following two questions:

- *Would you say that traffic contributes more or less to air pollution than industry?*
- *Would you say that industry contributes more or less to air pollution than traffic?*

With the first question, 45% answered traffic and 32% answered industry. With the second question, only 24% answered traffic while 57% answered industry. Thus, simply changing the order of the words *traffic* and *industry* dramatically changed the survey results.

By the Way

People are more likely to choose the item that comes first in a survey because of what psychologists call the *availability error*—the tendency to make judgments based on what is *available* in the mind. Professional polling organizations must be very careful to avoid this problem, sometimes by posing the question to some people in one order and to others in the opposite order.

✻EXAMPLE 7 *Do You Want a Tax Cut?*

The Republican National Committee commissioned a poll to find out whether Americans supported a tax-cut proposal. Asked whether they favored the tax cut, 67% of respondents answered yes. Should we conclude that Americans supported the proposal?

SOLUTION A question like "Do you favor a tax cut?" is biased because it does not give other options (much like the fallacy of *limited choice* discussed in Unit 1A). In fact, an independent poll conducted at the same time gave respondents a list of options for using surplus revenues. This poll found that 31% wanted the money devoted to Social Security, 26% wanted it used to reduce the national debt, and only 18% favored using it for a tax cut. (The remaining 25% of respondents chose a variety of other options.)

Now try Exercises 29–30.

Guideline 7: Check That Results Are Presented Fairly

Even when a statistical study is done well, it may be misrepresented in graphs or concluding statements. Researchers may occasionally misinterpret the results of their own studies or jump to conclusions that are not supported by the results, particularly when they have personal biases toward certain interpretations. In other cases, news reporters or others may misinterpret a survey or jump to unwarranted conclusions that make a story seem more spectacular. Misleading graphs are an especially common problem (see Unit 5D). In general, you should look for inconsistencies between the interpretation of a study (in pictures and words) and any actual data given with it.

EXAMPLE 8 *Does the School Board Need a Statistics Lesson?*

The school board in Boulder, Colorado, created a hubbub when it announced that 28% of Boulder school children were reading "below grade level," and hence concluded that methods of teaching reading needed to be changed. The announcement was based on reading tests on which 28% of Boulder school children scored below the national average for their grade. Do these data support the board's conclusion?

SOLUTION The fact that 28% of Boulder children scored below the national average for their grade implies that 72% scored at or above the national average. Thus, the school board's ominous statement about students reading "below grade level" makes sense only if "grade level" means the national average score for a particular grade. This interpretation of "grade level" is curious because it means that half the students in the nation are always below grade level—no matter how high the scores. The conclusion that teaching methods needed to be changed was not justified by these data.

Now try Exercises 31–32.

Guideline 8: Stand Back and Consider the Conclusions

Extraordinary claims require extraordinary evidence.

—Carl Sagan (1934–1996)

Finally, even if a study seems reasonable according to all the previous guidelines, you should stand back and consider the conclusions. Ask yourself questions such as these:

- Did the study achieve its goals?
- Do the conclusions make sense?
- Can you rule out alternative explanations for the results?
- If the conclusions do make sense, do they have any practical significance?

EXAMPLE 9 *Practical Significance*

An experiment is conducted in which the weight losses of people who try a new "Fast Diet Supplement" are compared to the weight losses of a control group of people who try to lose weight in other ways. After eight weeks, the results show that the treatment group lost an average of $\frac{1}{2}$ pound more than the control group. Assuming that it has no dangerous side effects, does this study suggest that the Fast Diet Supplement is a good treatment for people wanting to lose weight?

SOLUTION Compared to the average person's body weight, the difference of $\frac{1}{2}$ pound hardly matters at all. Thus, while the statistics in this case may be interesting, they don't seem to have much practical significance.

Now try Exercises 33–36.

SUMMARY Eight Guidelines for Evaluating a Statistical Study

1. Identify the goal of the study, the population considered, and the type of study.
2. Consider the source, particularly with regard to whether the researchers may be biased.
3. Look for bias that may prevent the sample from being representative of the population.
4. Look for problems in defining or measuring the variables of interest, which can make it difficult to interpret results.
5. Watch out for confounding variables that can invalidate the conclusions of a study.
6. Consider the setting and the wording of questions in any survey, looking for anything that might tend to produce inaccurate or dishonest responses.
7. Check that results are presented fairly in graphs and concluding statements, since both researchers and media often create misleading graphics or jump to conclusions that the results do not support.
8. Stand back and consider the conclusions. Did the study achieve its goals? Do the conclusions make sense? Do the results have any practical significance?

EXERCISES 5B

QUICK QUIZ

Choose the best answer to each of the following questions. Explain your reasoning with one or more complete sentences.

1. You read about an issue that was subject to an observational study when clearly it should have been studied with a double-blind experiment. The results from the observational study are therefore
   a. still valid, but a little less reliable.
   b. valid, but only if you first correct for the fact that the wrong type of study was done.
   c. essentially meaningless.
2. A study conducted by the oil company Exxon Mobil shows that there was no lasting damage from a large oil spill in Alaska. This conclusion
   a. is definitely invalid, because the study was biased.
   b. may be correct, but the potential for bias means that you should look very closely at how the conclusion was reached.
   c. could be correct if it falls within the confidence interval of the study.
3. Consider a study designed to learn about the social networks of all college freshmen, in which the researchers randomly interviewed students living in on-campus dormitories. The way this sample was chosen means the study will suffer from
   a. selection bias.
   b. participation bias.
   c. confounding variables.
4. The show *American Idol* selects winners based on votes cast by anyone who wants to vote. This means that the winner
   a. is the person most Americans want to win.
   b. may or may not be the person most Americans want to win, because the voting is subject to participation bias.
   c. may or may not be the person most Americans want to win, because the voting should have been double-blind.
5. Consider an experiment in which you measure the weights of 6-year-olds. The variable of interest in this study is
   a. the size of the sample.
   b. the weights of 6-year-olds.
   c. the ages of the children under study.