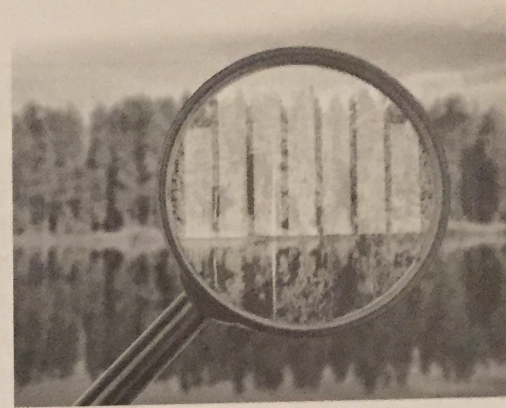
CHAPTER 1



# The Benefits and Risks of Using Statistics

### Thought Questions

1. A news story about drug use and grades concluded that smoking marijuana at least three times a week resulted in lower grades in college. How do you think the researchers came to this conclusion? Do you believe it? Is there a more reasonable conclusion?
2. It is obvious to most people that, on average, men are taller than women, and yet there are some women who are taller than some men. Therefore, if you wanted to "prove" that men were taller, you would need to measure many people of each sex. Here is a theory: On average, men have lower resting pulse rates than women do. How could you go about trying to prove or disprove that? Would it be sufficient to measure the pulse rates of one member of each sex? Two members of each sex? What information about men's and women's pulse rates would help you decide how many people to measure?
3. Suppose you were to learn that the large state university in a particular state graduated more students who eventually went on to become millionaires than any of the small liberal arts colleges in the state. Would that be a fair comparison? How should the numbers be presented in order to make it a fair comparison?
4. In a survey done in September, 2012, employers were asked a series of questions about whether colleges were preparing students adequately for careers in their companies. Of the 50,000 employers contacted, 704 responded. One of the questions asked was "How difficult is it to find recent college graduates who are qualified for jobs at your organization?" Over half (53%) of the respondents said that it was difficult or very difficult. Based on these results, can you conclude that about 53% of all employers feel this way? Why or why not?
(Source: http://chronicle.com/items/biz/pdf/Employers%20Survey.pdf.)

## 1.1 Why Bother Reading This Book?

If you have never studied statistics, you are probably unaware of the impact the science of statistics has on your everyday life. From knowing which medical treatments work best to choosing which television programs remain on the air, decision makers in almost every line of work rely on data and statistical studies to help them make wise choices.

We are exposed daily to information from surveys and scientific studies concerning our health, behavior, attitudes, and beliefs or revealing scientific and technological breakthroughs. This book's first objective is to help you understand this information and to sift the useful and the accurate from the useless and the misleading. (And there are plenty of both out there!) By the time you finish reading the book, you should be a statistical detective—able to read with a critical eye and to rely on your own interpretation of results emerging from surveys and statistical studies.

Another purpose of this book is to demystify statistical methods. Traditional statistics courses often place emphasis on how to compute rather than on how to understand. This book focuses on statistical ideas and their use in real life. Lastly, the book also contains information that can help you make better decisions. You will learn how psychological influences can keep you from making the best decisions, as well as new ways to think about coincidences, gambling, and other circumstances that involve chance events.

## 1.2 What is Statistics All About?

If we were all exactly the same—had the same physical makeup, the same behaviors and opinions, liked the same music and movies, and so on—most statistical methods would be of little use. But fortunately we aren't all the same! Statistical methods are used to analyze situations involving uncertainty and natural variation. They can help us understand our differences as well as find patterns and relationships that apply to all of us.

When you hear the word *statistics,* you probably either get an attack of math anxiety or think about lifeless numbers, such as the population of the city or town where you live, as measured by the latest census, or the per capita income in Japan. The goal of this book is to open a whole new world of understanding of the term *statistics,* and to help you realize that the invention of statistical methods is one of the most important developments of modern times. These methods influence everything from life-saving medical advances to the percent salary increase given to millions of people every year.

The word **statistics** is actually used to mean two different things. The better-known definition is that statistics are numbers measured for some purpose. A more appropriate, complete definition is the following:

> *Statistics is a collection of procedures and principles for gaining and analyzing information to educate people and help them make better decisions when faced with uncertainty.*

Using this definition, you have undoubtedly used statistics in your own life. For example, if you were faced with a choice of routes to get to school or work, or to get between one classroom building and the next, how would you decide which one to take? You would probably try each of them a number of times (thus gaining information) and then choose the best one according to criteria important to you, such as speed, fewer red lights, more interesting scenery, and so on. You might even use different criteria on different days—such as when the weather is pleasant versus when it is not. In any case, by sampling the various routes and comparing them, you would have gained and analyzed useful information to help you make a decision.

In addition to helping us make decisions, statistical studies also help us satisfy our curiosity about other people and the world around us. Do other people have the same opinions we do? Are they behaving the way we do—good or bad? Here is an example that may answer one of those questions for you.

**EXAMPLE 1.1** **Look Who's Talking!**

Texting or talking on a cell phone (other than hands-free) while driving is illegal in most of the United States and in many other countries. How many law-breakers are there? Does it differ by country? Does it differ by age group? The answers are lots, yes, and yes! Statistical surveys conducted in the United States and Europe in 2011 asked people, "In the past 30 days, how often have you talked on your cell phone while you were driving?" Response choices were "never," "just once," "rarely," "fairly often," and "regularly." The results showed that over two-thirds (68.7%) of adults surveyed in the United States admitted talking on their cell phone at least once in the past 30 days, but just over one-fifth (20.5%) of those in the United Kingdom admitting doing so. In the United States, almost 70% of respondents aged 18 to 24 admitted that they had talked on their phone, but only about 60% of those aged 55 to 64 did so. The age differences were even more striking for those who admitted texting, with slightly over 50% of the youngest age group doing so but under 10% of the oldest age group doing so.

Although it is interesting to know whether others are behaving like we are, these statistics also have implications for public health officials and lawmakers who are studying them to determine why drivers in the United Kingdom are so much more law-abiding than are drivers in the United States in this instance. Or are they? In this book, you will learn to ask other probing questions, such as whether the two groups are equally likely to be telling the truth in their responses and whether the difference could be partially explained by the fact that people in the United States drive more than people in the United Kingdom.

(Source: *Morbidity and Mortality Weekly Report,* March 15, 2013 Vol. 62, No. 10, Centers for Disease Control and Prevention.) ■

## 1.3 Detecting Patterns and Relationships

How do scientists decide what questions to investigate? Often, they start with observing something and become curious about whether it's a unique circumstance or something that is part of a larger pattern. In Case Study 1.1, we will see how one researcher followed a casual observation to a fascinating conclusion.

## CASE STUDY 1.1 Heart or Hypothalamus?

SOURCE: Salk (1973), pp. 26–29.

You can learn a lot about nature by observation. You can learn even more by conducting a carefully controlled experiment. This case study has both. It all began when psychologist Lee Salk noticed that despite his knowledge that the hypothalamus plays an important role in emotion, it was the heart that seemed to occupy the thoughts of poets and songwriters. There were no everyday expressions or song titles such as "I love you from the bottom of my hypothalamus" or "My hypothalamus longs for you." Yet, there was no physiological reason for suspecting that the heart should be the center of such attention. Why had it always been the designated choice?

Salk began wondering about the role of the heart in human relationships. He also noticed that when on 42 separate occasions he watched a rhesus monkey at the zoo holding her baby, she held the baby on the left side, close to her heart, on 40 of those occasions. He then observed 287 human mothers within 4 days after giving birth and noticed that 237, or 83%, held their babies on the left. Handedness did not explain it; 83% of the right-handed mothers and 78% of the left-handed mothers exhibited the left-side preference. When asked why they chose the left side, the right-handed mothers said it was so their right hand would be free. The left-handed mothers said it was because they could hold the baby better with their dominant hand. In other words, both groups were able to rationalize holding the baby on the left based on their own preferred hand.

Salk wondered if the left side would be favored when carrying something other than a newborn baby. He found a study in which shoppers were observed leaving a supermarket carrying a single bag; exactly half of the 438 adults carried the bag on the left. But when stress was involved, the results were different. Patients at a dentist's office were asked to hold a 5-inch rubber ball while the dentist worked on their teeth. Substantially more than half held the ball on the left.

Salk speculated, "It is not in the nature of nature to provide living organisms with biological tendencies unless such tendencies have survival value." He surmised that there must indeed be survival value to having a newborn infant placed close to the sound of its mother's heartbeat.

To test this conjecture, Salk designed a study in a baby nursery at a New York City hospital. He arranged for the nursery to have the continuous sound of a human heartbeat played over a loudspeaker. At the end of 4 days, he measured how much weight the babies had gained or lost. Later, with a new group of babies in the nursery, no sound was played. Weight gains were again measured after 4 days.

The results confirmed what Salk suspected. Although they did not eat more than the control group, the infants treated to the sound of the heartbeat gained more weight (or lost less). Further, they spent much less time crying. Salk's conclusion was that "newborn infants are soothed by the sound of the normal adult heartbeat." Somehow, mothers intuitively know that it is important to hold their babies on the left side. What had started as a simple observation of nature led to a further understanding of an important biological response of a mother to her newborn infant. ■

## How to Move From Noticing to Knowing

Some differences are obvious to the naked eye, such as the fact that the average man is taller than the average woman. If we were content to know about only such obvious relationships, we would not need the power of statistical methods. But had you noticed that babies who listen to the sound of a heartbeat gain more weight? Have you ever noticed that taking aspirin helps prevent heart attacks? How about the fact that people are more likely to buy jeans in certain months of the year than in others? The fact that men have lower resting pulse rates than women do? The fact that listening to Mozart improves performance on the spatial reasoning questions of an IQ test? All of these are relationships that have been demonstrated in studies using proper statistical methods, yet none of them are obvious to the naked eye.

Let's take the simplest of these examples—one you can test yourself—and see what's needed to properly demonstrate the relationship. Suppose you wanted to verify the claim that, on average, men have lower resting pulse rates than women do. Would it be sufficient to measure only your own pulse rate and that of a friend of the opposite sex? Obviously not. Even if the pair came out in the predicted direction, the singular measurements would certainly not speak for all members of each sex.

It is not easy to conduct a statistical study properly, but it is easy to understand much of how it should be done. We will examine each of the following concepts in great detail in the remainder of this book; here we just introduce them, using the simple example of comparing male and female pulse rates.

> *To conduct a statistical study properly, one must:*
>
> 1. Get a representative sample.
> 2. Get a large enough sample.
> 3. Decide whether the study should be an observational study or a randomized experiment.

**1. Get a representative sample.** Most researchers hope to extend their results beyond just the participants in their research. Therefore, it is important that the people or objects in a study be representative of the larger group for which conclusions are to be drawn. We call those who are actually studied a **sample** and the larger group from which they were chosen a **population.** (In Chapter 4, we will learn some ways to select a proper sample.) For comparing pulse rates, it may be convenient to use the members of your class. But this sample would not be valid if there were something about your class that would relate pulse rates and sex, such as if the entire men's track team happened to be in the class. It would also be unacceptable if you wanted to extend your results to an age group much different from the distribution of ages in your class. Often researchers are constrained to using such "convenience" samples, and we will discuss the implications of this later in the book.

**2. Get a large enough sample.** Even experienced researchers often fail to recognize the importance of this concept. In Part 4 of this book, you will learn how to detect the problem of a sample that is too small; you will also learn that such a sample can sometimes lead to erroneous conclusions. In comparing pulse rates, collecting one pulse rate from each sex obviously does not tell us much. Is two enough? Four? One hundred? The answer to that question depends on how much *natural variability* there is among pulse rates. If all men had pulse rates of 65 and all women had pulse rates of 75, it wouldn't take long before you recognized a difference. However, if men's pulse rates ranged from 50 to 80 and women's pulse rates ranged from 52 to 82, it would take many more measurements to convince you of a difference. The question of how large is "large enough" is closely tied to how diverse the measurements are likely to be within each group. The more diverse, or variable, the individuals within each group, the larger the sample needs to be to detect a real difference between the groups.

**3. Decide whether the study should be an observational study or a randomized experiment.** For comparing pulse rates, it would be sufficient to measure or "observe" both the pulse rate and the sex of the people in our sample. When we merely observe things about our sample, we are conducting an **observational study.** However, if we were interested in whether frequent use of aspirin would help prevent heart attacks, it would not be sufficient to simply observe whether people frequently took aspirin and then whether they had a heart attack. It could be that people who were more concerned with their health were both more likely to take aspirin and less likely to have a heart attack, or vice versa. Or, it could be that drinking the extra glass of water required to take the aspirin contributes to better health.

To be able to make a causal connection, we would have to conduct a **randomized experiment** in which we *randomly* assigned people to one of two groups. **Random assignments** are made by doing something akin to flipping a coin to determine the group membership for each person. In one group, people would be given aspirin, and in the other, they would be given a dummy pill that looked like aspirin. So as not to influence people with our expectations, we would not tell people which one they were taking until the experiment was concluded. In Case Study 1.2, we briefly examine the experiment that initially established the causal link between aspirin use and reduction of heart attacks. In Chapter 5, we discuss these ideas in much more detail.

## CASE STUDY 1.2 Does Aspirin Prevent Heart Attacks?

In 1988, the Steering Committee of the Physicians' Health Study Research Group released the results of a 5-year randomized experiment conducted using 22,071 male physicians between the ages of 40 and 84. The physicians had been randomly assigned to two groups. One group took an ordinary aspirin tablet every other day, whereas the other group took a "placebo," a pill designed to look just like an aspirin but with no active ingredients. Neither group knew whether they were taking the active ingredient.

The results, shown in Table 1.1, support the conclusion that taking aspirin does indeed help reduce the risk of having a heart attack. The rate of heart attacks in the

**TABLE 1.1** The Effect of Aspirin on Heart Attacks

| Condition | Heart Attack | No Heart Attack | Attacks per 1000 |
|---|---|---|---|
| Aspirin | 104 | 10,933 | 9.42 |
| Placebo | 189 | 10,845 | 17.13 |

group taking aspirin was only 55% of the rate of heart attacks in the placebo group, or just slightly more than half as big. Because the men were randomly assigned to the two conditions, other factors, such as amount of exercise, should have been similar for both groups. The only substantial difference in the two groups should have been whether they took the aspirin or the placebo. Therefore, we can conclude that taking aspirin caused the lower rate of heart attacks for that group.

Notice that because the participants were all male physicians, these conclusions may not apply to the general population of men. They may not apply to women at all because no women were included in the study. More recent evidence has provided even more support for this effect, however—something we will examine in more detail in an example in Chapter 27. ■

## 1.4 Don't Be Deceived by Improper Use of Statistics

Let's look at some examples representative of the kinds of abuses of statistics you may see in the media. The first example illustrates the danger of not getting a representative sample; in the second example, the statistics have been taken out of their proper context; and in the third and fourth examples, you will see how to stop short of making too strong a conclusion on the basis of an observational study.

**EXAMPLE 1.2** **Robotic Polls and Representative Samples**

Methods for polling voters to predict election results have become more complicated as more and more people rely exclusively on cell phones instead of landlines. But some polling firms were slow to change their methods and, consequently, made significant blunders in predicting the outcome of the 2012 Presidential Election. In particular, polling companies that exclusively used "robopolls," which choose their participants using randomized computer dialing and then use an automated script, missed the mark by an average of 4.3 percentage points in favor of the Republican candidate Mitt Romney. One firm even had a 15.7% Republican bias! What went wrong? By 2012, about one-third of Americans relied solely on cell phones and had no landline. But it was illegal to call cell phones using robopolls, so organizations that used them only reached homes with landlines. Younger voters, those less well-off financially and non-Caucasians, were less likely to have landlines and also more likely to vote for the Democratic candidate, Barack Obama. Therefore, the robopolls underestimated the support for Obama.

(*Source:* http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/#more-37396.)

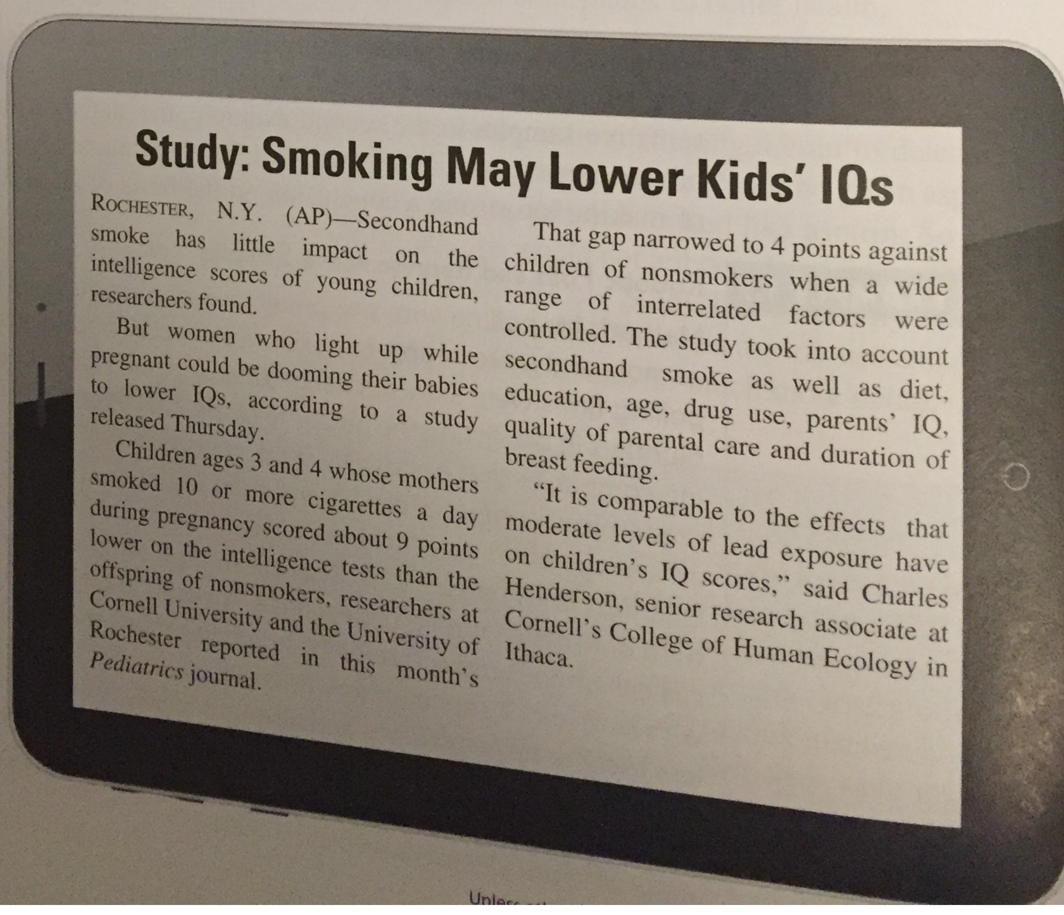
■

**EXAMPLE 1.3** Toxic Chemical Statistics

When a federal air report ranked the state of New Jersey as 22nd in the nation in its release of toxic chemicals, the New Jersey Department of Environmental Protection happily took credit (Wang, 1993, p. 170). The statistic was based on a reliable source, a study by the U.S. Environmental Protection Agency. However, the ranking had been made based on total pounds released, which was 38.6 million for New Jersey. When this total was turned into pounds per square mile in the state, it became apparent New Jersey was one of the worst—fourth on the list. Because New Jersey is one of the smallest states by area, the figures were quite misleading until adjusted for size. ■

**EXAMPLE 1.4** Mom's Smoking and Kid's IQ

Read the article in Figure 1.1, and then read the headline again. Notice that the headline stops short of making a causal connection between smoking during pregnancy and lower IQs in children. Reading the article, you can see that the results are based on an observational study and not an experiment—with good reason: It would clearly be unethical to randomly assign pregnant women to either smoke or not. With studies like this, the best that can be done is to try to measure and statistically adjust for other factors that might be related to both smoking behavior and children's IQ scores. Notice that when the researchers did so, the gap in IQ between the children of smokers and nonsmokers narrowed from nine points down to four points. There may be even more factors that the researchers did not measure that



**Study: Smoking May Lower Kids' IQs**

ROCHESTER, N.Y. (AP)—Secondhand smoke has little impact on the intelligence scores of young children, researchers found.

But women who light up while pregnant could be dooming their babies to lower IQs, according to a study released Thursday.

Children ages 3 and 4 whose mothers smoked 10 or more cigarettes a day during pregnancy scored about 9 points lower on the intelligence tests than the offspring of nonsmokers, researchers at Cornell University and the University of Rochester reported in this month's *Pediatrics* journal.

That gap narrowed to 4 points against children of nonsmokers when a wide range of interrelated factors were controlled. The study took into account secondhand smoke as well as diet, education, age, drug use, parents' IQ, quality of parental care and duration of breast feeding.

"It is comparable to the effects that moderate levels of lead exposure have on children's IQ scores," said Charles Henderson, senior research associate at Cornell's College of Human Ecology in Ithaca.

**Figure 1.1**
Don't make causal connections from observational studies

*Source:* "Study: Smoking May Lower Kids' IQs." Associated Press, February 11, 1994. Reprinted with permission.

would account for the remaining four-point difference. Unfortunately, with an observational study, we simply cannot make causal conclusions. We will explore this particular example in more detail in Chapter 6. ■

**EXAMPLE 1.5** **Does Marijuana Impair the Brain?**

An article headlined "New study confirms too much pot impairs brain" read as follows:

> *More evidence that chronic marijuana smoking impairs mental ability: Researchers at the University of Iowa College of Medicine say a test shows those who smoke seven or more marijuana joints per week had lower math, verbal and memory scores than non-marijuana users. Scores were particularly reduced when marijuana users held a joint's smoke in their lungs for longer periods.* (*San Francisco Examiner*, 13 March 1993, p. D-1)

This research was clearly based on an observational study because people cannot be randomly assigned to either smoke marijuana or not. The headline is misleading because it implies that there is a causal connection between smoking marijuana and brain functioning. All we can conclude from an observational study is that there is a relationship. It could be the case that people who choose to smoke marijuana are those who would score lower on the tests anyway. ■

In addition to learning how to evaluate statistical studies, in this book you will learn some simple methods for computing probabilities. Case Study 1.3 not only describes one practical application of probability but also illustrates that it is important to question what assumptions are made before computing the probability of a particular event.

## CASE STUDY 1.3 Using Probability to Detect Cheating

Professors and other students hate it when students cheat on exams, and some professors have devised methods that make it relatively easy to detect cheating on essays and similar questions. But detecting cheating on multiple choice exams is not as easy. That's where probability and statistics can help. Professor Robert Mogull thought he detected cheating by two students on multiple choice exams in his statistics class at Sacramento State University because they had identical questions wrong on all four 25-question multiple choice exams. He calculated the probability of that happening to be extremely small, failed the two students, and published a paper explaining his method (Mogull, 2003). But others have criticized his probability calculations because he assumed that all students were equally likely to miss any particular question (Actuarial Outpost, 2013). The critics pointed out that there are all sorts of reasons why two particular students might miss the same questions as each other, especially if they were friends. Perhaps they studied together, they had the same major and thus had similar knowledge, had the same statistics course in high school, and so on.

In a similar case, Klein (1992) described a situation in which two students were accused of cheating on a multiple-choice medical licensing exam. They had been observed whispering during one part of the 3-day exam and their answers to the questions they got wrong very often matched each other. The licensing board determined that the statistical evidence for cheating was overwhelming. They estimated that the odds of two people having answers as close as these two did were less than 1 in 10,000. Further, the students were husband and wife. Their tests were invalidated.

The case went to trial, and upon further investigation, the couple was exonerated. They hired a statistician who was able to show that the agreement in their answers during the session in which they were whispering was no higher than it was in the other sessions. What happened? The board assumed students who picked the wrong answer were simply guessing among the other choices. This couple had grown up together and had been educated together in India. Answers that would have been correct for their culture and training were incorrect for the American culture (for example, whether a set of symptoms was more indicative of tuberculosis or a common cold). Their common mistakes often would have been the right answers for India. So, the licensing board erred in calculating the odds of getting such a close match by using the assumption that they were just guessing. And, according to Klein, "with regard to their whispering, it was very brief and had to do with the status of their sick child" (p. 26). ■

## 1.5 Summary and Conclusions

In this chapter, we have just begun to examine both the advantages and the dangers of using statistical methods. We have seen that it is not enough to know the results of a study, survey, or experiment. We also need to know how those numbers were collected and who was asked. In the upcoming chapters, you will learn much more about how to collect and process this kind of information properly and how to detect problems in what others have done. You will learn that a relationship between two characteristics (such as smoking marijuana and lower grades) does not necessarily mean that one causes the other, and you will learn how to determine other plausible explanations. In short, you will become an educated consumer of statistical information.

### Thinking About Key Concepts

- Cause and effect conclusions cannot generally be made based on results of observational studies. The problem is that the groups being studied are likely to differ in lots of other ways, any of which could be causing the observed difference in outcome.
- In randomized experiments, cause and effect conclusions *can* generally be made. By randomly assigning individuals to receive different treatments, the groups should be similar on everything except the treatment they are given.

- If a study is conducted using a sample that is representative of a larger group for the question of interest, then the results from the sample can be considered to apply to the larger group (population) as well.
- If there is little natural variability in the responses being measured, then a sample with only a small number of individuals should be adequate for detecting differences among groups. But if there is lots of variability, differences are only likely to be detected if a large number of individuals are included in the measured sample.
- When comparing statistics for groups or locations of different sizes, it is more informative to report the results per person or per unit, rather than for the group as a whole.

## Exercises

*Exercises with numbers divisible by 3 (3, 6, 9, etc.) are included in the Solutions at the back of the book. They are marked with an asterisk (*).*

**1.** Explain why the relationship shown in Table 1.1, concerning the use of aspirin and heart attack rates, can be used as evidence that aspirin actually prevents heart attacks.

**2.** "People who often attend cultural activities, such as movies, sports events and concerts, are more likely than their less cultured cousins to survive the next eight to nine years, even when education and income are taken into account, according to a survey by the University of Umea in Sweden" (*American Health,* April 1997, p. 20).

**a.** Can this claim be tested by conducting a randomized experiment? Explain.

**b.** On the basis of the study that was conducted, can we conclude that attending cultural events causes people to be likely to live longer? Explain.

**c.** The article continued, "No one's sure how Mel Gibson and Mozart help health, but the activities may enhance immunity or coping skills." Comment on the validity of this statement.

**d.** The article notes that education and income were taken into account. Give two examples of other factors about the people surveyed that you think should also have been taken into account.

***3.** Explain why the number of people in a sample is an important factor to consider when designing a study.

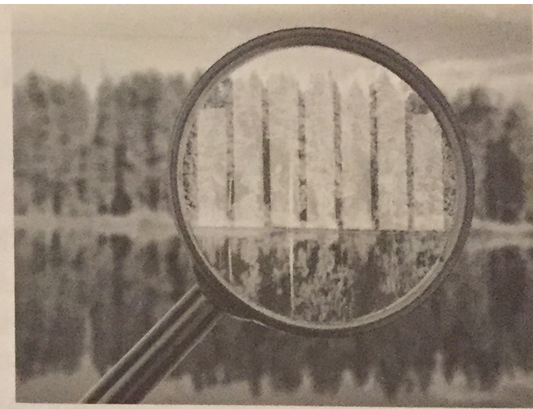
**4.** Explain what problems may arise in trying to make conclusions based on a survey mailed to the subscribers of a specialty magazine. Find or construct an example.

**5.** "If you have borderline high blood pressure, taking magnesium supplements may help, Japanese researchers report. Blood pressure fell significantly in subjects who got 400–500 milligrams of magnesium a day for four weeks, but not in those getting a placebo" (*USA Weekend,* 22–24 May 1998, p. 11).

**a.** Do you think this was a randomized experiment or an observational study? Explain.

**b.** Do you think the relationship found in this study is a causal

CHAPTER 2

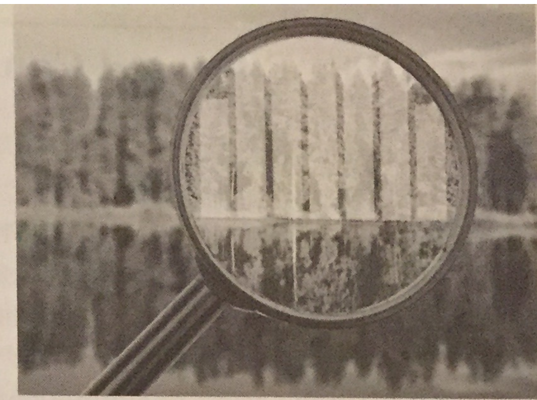


# Reading the News

## Thought Questions

1. Advice columnists sometimes ask readers to write and express their feelings about certain topics. For instance, Ann Landers once asked readers whether they thought engineers made good husbands. Do you think the responses are representative of public opinion? Explain why or why not.
2. Taste tests of new products are often done by having people taste both the new product and an old familiar standard. Do you think the results would be biased if the person handing the products to the respondents knew which was which? Explain why or why not.
3. Nicotine patches attached to the arm of someone who is trying to quit smoking dispense nicotine into the blood. Suppose you read about a study showing that nicotine patches were twice as effective in getting people to quit smoking as "placebo" patches (made to look like the real thing). Further, suppose you are a smoker trying to quit. What questions would you want answered about how the study was done and its results before you decided whether to try the patches yourself?
4. For a door-to-door survey on opinions about various political issues, do you think it matters who conducts the interviews? Give an example of how it might make a difference.

CHAPTER 2



# Reading the News

**Thought Questions**

1. Advice columnists sometimes ask readers to write and express their feelings about certain topics. For instance, Ann Landers once asked readers whether they thought engineers made good husbands. Do you think the responses are representative of public opinion? Explain why or why not.
2. Taste tests of new products are often done by having people taste both the new product and an old familiar standard. Do you think the results would be biased if the person handing the products to the respondents knew which was which? Explain why or why not.
3. Nicotine patches attached to the arm of someone who is trying to quit smoking dispense nicotine into the blood. Suppose you read about a study showing that nicotine patches were twice as effective in getting people to quit smoking as "placebo" patches (made to look like the real thing). Further, suppose you are a smoker trying to quit. What questions would you want answered about how the study was done and its results before you decided whether to try the patches yourself?
4. For a door-to-door survey on opinions about various political issues, do you think it matters who conducts the interviews? Give an example of how it might make a difference.

## 2.1 The Educated Consumer of Data

Visit any news-based website, pick up a popular magazine or tune into a news broadcast, and you are almost certain to find a story containing conclusions based on data. Should you believe what you read? Not always. It depends on how the data were collected, measured, and summarized. In this chapter, we discuss seven critical components of statistical studies. We examine the kinds of questions you should ask before you believe what you read. We go into further detail about these issues in subsequent chapters. The goal in this chapter is to give you an overview of how to be a more educated consumer of the data you encounter in your everyday life.

### What Are Data?

In statistical parlance, **data** is a plural word referring to a collection of numbers or other pieces of information to which meaning has been attached. For example, the numbers 1, 3, and 10 are not necessarily data, but they become so when we are told that these were the weight gains in grams of three of the infants in Salk's heartbeat study, discussed in Chapter 1. In Case Study 1.2, the data consisted of two pieces of information measured for each participant: (1) whether they took aspirin or a placebo, and (2) whether they had a heart attack.

### Don't Always Believe What You Read

When you encounter the results of a study in the news, you are rarely presented with the actual data. Someone has usually summarized the information for you, and he or she has probably already drawn conclusions and presented them to you. Don't always believe them. The meaning we can attach to data, and to the resulting conclusions, depends on how well the information was acquired and summarized.

In the remaining chapters of Part 1, we look at proper ways to obtain data. In Part 2, we turn our attention to how it should be summarized. In Part 4, we learn the power as well as the limitations of using the data collected from a sample to make conclusions about the larger population. In this chapter, we address seven features of statistical studies that you should think about when you read a news article. You will begin to be able to think critically and make your own conclusions about what you read.

## 2.2 Origins of News Stories

Where do news stories originate? How do reporters hear about events and determine that they are newsworthy? For stories based on statistical studies, there are several possible sources. The two most common of these sources are also the most common outlets for researchers to present the results of their work: academic conferences and scholarly journals.

## News from Academic Conferences

Every academic discipline holds conferences, usually annually, in which researchers can share their results with others. Reporters routinely attend these academic conferences and look for interesting news stories. For larger conferences, there is usually a "press room" where researchers can leave press releases for the media. If you pay attention, you will notice that in certain weeks of the year, there will be several news stories about studies with related themes. For instance, the American Psychological Association meets in August, and there are generally some news stories emerging from results presented there. The American Association for the Advancement of Science meets in February, and news stories related to various areas of science will appear in the news that week.

One problem with news stories based on conference presentations is that there is unlikely to be a corresponding written report by the researchers, so it is difficult for readers of the news story to obtain further information. News stories based on conference reports generally mention the name and date of the conference as well as the name and institution of the lead researcher, so sometimes it is possible to contact the researcher for further information. Some researchers make conference presentations available on their websites.

## News from Published Reports

Many news stories about statistical studies are based on published articles in scholarly journals. Reporters routinely read these journals when they are published, or they get advance press releases from the journal offices. News stories based on journal articles usually mention the journal and date of publication, so if you are interested in learning more about the study, you can obtain the original journal article. Journal articles are sometimes available on the journal's website or on the website of the author(s). You can also write to the lead author and request that the article be sent to you.

As a third source of news stories about statistical studies, some government and private agencies release in-depth research reports. Unlike journal articles, these reports are not necessarily "peer-reviewed" or checked by neutral experts on the topic. An advantage of these reports is that they are not restricted by space limitations imposed by journals and often provide much more in-depth information than do journal articles.

A supplementary source from which news stories may originate is a university media office. Most research universities have an office that provides press releases when faculty members have completed research that may be of interest to the public. The timing of these news releases usually corresponds to a presentation at an academic conference or publication of results in an academic journal, but the news release summarizes the information so that journalists don't have to be as versed in the technical aspects of the research to write a good story. When you read about a study in the news and would like more information, the news office of the lead researcher's institution is a good place to start looking. They may have issued a press release on which the story was based.

**News Stories and Original Sources in the Appendix and on the Companion Website, www.cengage.com/stats/UttsSTS4e**

To illustrate how the concepts in this book are used in research and eventually converted into news stories, there is a collection of examples included with this book. In each case, the example includes a story from a newspaper, magazine, or website, and these are printed or referenced in the Appendix and on the companion website accompanying the book. Sometimes there is also a press release. These are provided as an additional "News Story" and included on the companion website. Most of the news stories are based on articles from scholarly journals or detailed reports. Many of these articles are printed in full on the companion website, labeled as the "Original Source." Throughout this book, you will find a website icon when you need to refer to the material on the companion website. By comparing the news story and the original source, you will learn how to evaluate what is reported in the news.

## 2.3 How to Be a Statistics Sleuth: Seven Critical Components

Reading and interpreting the results of surveys or experiments is not much different from reading and interpreting the results of other events of interest, such as sports competitions or criminal investigations. If you are a sports fan, then you know what information should be included in reports of competitions and you know when crucial information is missing. If you have ever been involved in an event that was later reported in the news, you know that missing information can lead readers to erroneous conclusions.

In this section, you are going to learn what information should be included in news reports of statistical studies. Unfortunately, crucial information is often missing. With some practice, you can learn to figure out what's missing, as well as how to interpret what's reported. You will no longer be at the mercy of someone else's conclusions. You will be able to reach your own conclusions. To provide structure to our examination of news reports, let's list Seven Critical Components that determine the soundness of statistical studies. A good news report should provide you with information about all of the components that are relevant to that study.

**Component 1:** The *source* of the research and of the *funding.*

**Component 2:** The *researchers* who had *contact* with the participants.

**Component 3:** The *individuals* or objects studied and how they were *selected.*

**Component 4:** The exact nature of the *measurements* made or *questions* asked.

**Component 5:** The *setting* in which the measurements were taken.

**Component 6:** *Differences* in the groups being compared, *in addition* to the factor of interest.

**Component 7:** The *extent* or *size* of any claimed effects or differences.

Before delving into some examples, let's examine each component more closely. You will find that most of the problems with studies are easy to identify. Listing these components simply provides a framework for using your common sense.

**Component 1: The *source* of the research and of the *funding*** Studies are conducted for three major reasons. First, governments and private companies need to have data in order to make wise policy decisions. Information such as unemployment rates and consumer spending patterns are measured for this reason. Second, researchers at universities and other institutions are paid to ask and answer interesting questions about the world around us. The curious questioning and experimentation of such researchers have resulted in many social, medical, and scientific advances. Much of this research is funded by government agencies, such as the National Institutes of Health. Third, companies want to convince consumers that their programs and products work better than the competition's, or special-interest groups want to prove that their point of view is held by the majority.

Unfortunately, it is not always easy to discover who funded research. Many university researchers are now funded by private companies. In her book *Tainted Truth* (1994), Cynthia Crossen warns us:

> *Private companies, meanwhile, have found it both cheaper and more prestigious to retain academic, government, or commercial researchers than to set up in-house operations that some might suspect of fraud. Corporations, litigants, political candidates, trade associations, lobbyists, special interest groups—all can buy research to use as they like. (p. 19)*

If you discover that a study was funded by an organization that would be likely to have a strong preference for a particular outcome, it is especially important to be sure that correct scientific procedures were followed. In other words, be sure the remaining components have sound explanations.

**Component 2: The *researchers* who had *contact* with the participants** It is important to know who actually had contact with the participants and what message those people conveyed. Participants often give answers or behave in ways to comply with the desires of the researchers. Consider, for example, a study done at a shopping mall to compare a new brand of a certain product to an old familiar brand. Shoppers are asked to taste each brand and state their preference. It is crucial that both the person presenting the two brands and the respondents be kept entirely blind as to which is which until after the preferences have been selected. Any clues might bias the respondent to choose the old familiar brand. Or, if the interviewer is clearly eager to have them choose one brand over the other, the respondents will most likely oblige in order to please. As another example, if you discovered that a study on the prevalence of illegal drug use was conducted by sending uniformed police officers door-to-door, you would probably not have much faith in the results. We will discuss other ways in which researchers influence participants in Chapters 4 and 5.

**Component 3: The *individuals* or objects studied and how they were *selected*** It is important to know to whom the results can be extended. In general, the results of a study apply only to individuals similar to those in the study. For example, until recently, many medical studies included men only, so the results were of little value to women. When determining who is similar to those in the study, it is also important to know how participants were enlisted for the study. Many studies rely on volunteers recruited through newspapers or websites, who are usually paid a small amount for their participation. People who would respond to such recruitment efforts may differ in relevant ways from those who would not. Surveys relying on voluntary responses are likely to be biased because only those who feel strongly about the issues are likely to respond. For instance, some websites have a "question of the day" to which people are asked to voluntarily respond by clicking on their preferred answer. Only those who have strong opinions are likely to participate, so the results cannot be extended to any larger group.

**Component 4: The exact nature of the *measurements* made or *questions* asked** As you will see in Chapter 3, precisely defining and measuring most of the things researchers study isn't easy. For example, if you wanted to measure whether people "eat breakfast," how would you do so? What if they just have juice? What if they work until midmorning and then eat a meal that satisfies them until dinner? You need to understand exactly what the various definitions mean when you read about someone else's measurements.

In polls and surveys, the "measurements" are usually answers to specific questions. Both the wording and the ordering of the questions can influence answers. For example, a question about "street people" would probably elicit different responses than a question about "families who have no home." Ideally, you should be given the exact wording that was used in a survey or poll.

**Component 5: The *setting* in which the measurements were taken** The setting in which measurements were taken includes factors such as when and where they were taken and whether respondents were contacted by phone, mail, e-mail, or in person. A study can be easily biased by timing. For example, opinions on whether criminals should be locked away for life may change drastically following a highly publicized murder or kidnapping case. If a study is conducted by landline telephone and calls are made only in the evening, even certain groups of people who have landlines would be excluded, such as those who work the evening shift or who routinely eat dinner in restaurants, and people without landlines would not be represented at all.

Where the measurements were taken can also influence the results. Questions about sensitive topics, such as sexual behavior or income, might be more readily answered on a website, where respondents feel more anonymous. Sometimes research is done in a laboratory or university office, and the results may not readily extend to a natural setting. For example, studies of communication between two people are sometimes done by asking them to conduct a conversation in a university office with a voice recorder present. Such conditions almost certainly produce more limited conversation than would occur in a more natural setting.

**Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest** If two or more groups are being compared on a factor of interest, it is important to consider other ways in which the groups may differ that might influence the comparison. For example, suppose researchers want to know if smoking marijuana is related to academic performance. If the group of people who smoke marijuana has lower test scores than the group of people who don't, researchers may want to conclude that the lower test scores are due to smoking marijuana. Often, however, other disparities in the groups can explain the observed difference just as well. For example, people who smoke marijuana may simply be the type of people who are less motivated to study and thus would score lower on tests whether they smoked or not. Reports of research should include an explanation of any such possible differences that might account for the results. We will explore the issue of these kinds of extraneous factors, and how to control for them, in much more detail in Chapter 5.

**Component 7: The *extent* or *size* of any claimed effects or differences** Media reports about statistical studies often fail to tell you how large the observed effects were. Without that knowledge, it is hard for you to assess whether you think the results are of any practical importance. For example, if, based on Case Study 1.2, you were told simply that taking aspirin every other day reduced the risk of heart attacks, you would not be able to determine whether it would be worthwhile to take aspirin. You should instead be told that for the men in the study, the rate was reduced from about 17 heart attacks per 1000 participants without aspirin to about 9.4 heart attacks per 1000 with aspirin. Often news reports simply report that a treatment had an effect or that a difference was observed, but don't tell you the size of the difference or effect. We will investigate this issue in great detail in Part 4 of this book.

## 2.4 Four Hypothetical Examples of Bad Reports

Throughout this book, you will see numerous examples of real studies and news reports. So that you can get some practice finding problems without having to read unnecessarily long news articles, let's examine some hypothetical reports. These are admittedly more problematic than many real reports because they serve to illustrate several difficulties at once.

Read each article (at the top of pages 24, 26, 27 and 29) and see if your common sense gives you some reasons why the headline is misleading. Then read the commentary accompanying the article, discussing the Seven Critical Components.

Hypothetical News Article 1

## Study Shows Psychology Majors Are Smarter than Chemistry Majors

A fourth-year psychology student, for her senior thesis, conducted a study to see if students in her major were smarter than those majoring in chemistry. She handed out questionnaires in five advanced psychology classes and five advanced chemistry labs. She asked the students who were in class to record their grade-point averages (GPAs) and their majors. Using the data only from those who were actually majors in these fields in each set of classes, she found that the psychology majors had an average GPA of 3.05, whereas the chemistry majors had an average GPA of only 2.91. The study was conducted last Wednesday, the day before students were home enjoying Thanksgiving dinner.

### *Hypothetical News Article 1: "Study Shows Psychology Majors Are Smarter than Chemistry Majors"*

**Component 1: The *source* of the research and of the *funding*** The study was a senior thesis project conducted by a psychology major. Presumably, it was cheap to run and was paid for by the student. One could argue that she would have a reason to want the results to come out as they did, although with a properly conducted study, the motives of the experimenter should be minimized. As we shall see, there were additional problems with this study.

**Component 2: The *researchers* who had *contact* with the participants** Presumably, only the student conducting the study had contact with the respondents. Crucial missing information is whether she told them the purpose of the study. Even if she did not tell them, many of the psychology majors may have known her and known what she was doing. Any clues as to desired outcomes on the part of experimenters can bias the results.

**Component 3: The *individuals* or objects studied and how they were *selected*** The individuals selected are the crux of the problem here. The measurements were taken on advanced psychology and chemistry students, which would have been fine if they had been sampled correctly. However, only those who were in the psychology classes or in the chemistry labs that day were actually measured. Less conscientious students are more likely to leave early before a holiday, but a missed class is probably easier to make up than a missed lab. Therefore, perhaps a larger proportion of the students with low GPAs were absent from the psychology classes than from the chemistry labs. Due to the missing students, the investigator's results would overestimate the average GPA for psychology students more so than for chemistry students.

**Component 4: The exact nature of the *measurements* made or *questions* asked** Students were asked to give a "self-report" of their GPAs. A more accurate method would have been to obtain this information from the registrar at the university. Students may not know their exact GPA. Also, one group may be more likely to know the exact value than the other. For example, if many of the chemistry majors were planning to apply to medical school in the near future, they may be only too aware of their grades. Further, the headline implies that GPA is a measure of intelligence. Finally, the research assumes that GPA is a standard measure. Perhaps grading is more competitive in the chemistry department.

**Component 5: The *setting* in which the measurements were taken** Notice that the article specifies that the measurements were taken on the day before a major holiday. Unless the university consisted mainly of commuters, many students may have left early for the holiday, further aggravating the problem that the students with lower grades were more likely to be missing from the psychology classes than from the chemistry labs. Further, because students turned in their questionnaires anonymously, there was presumably no accountability for incorrect answers.

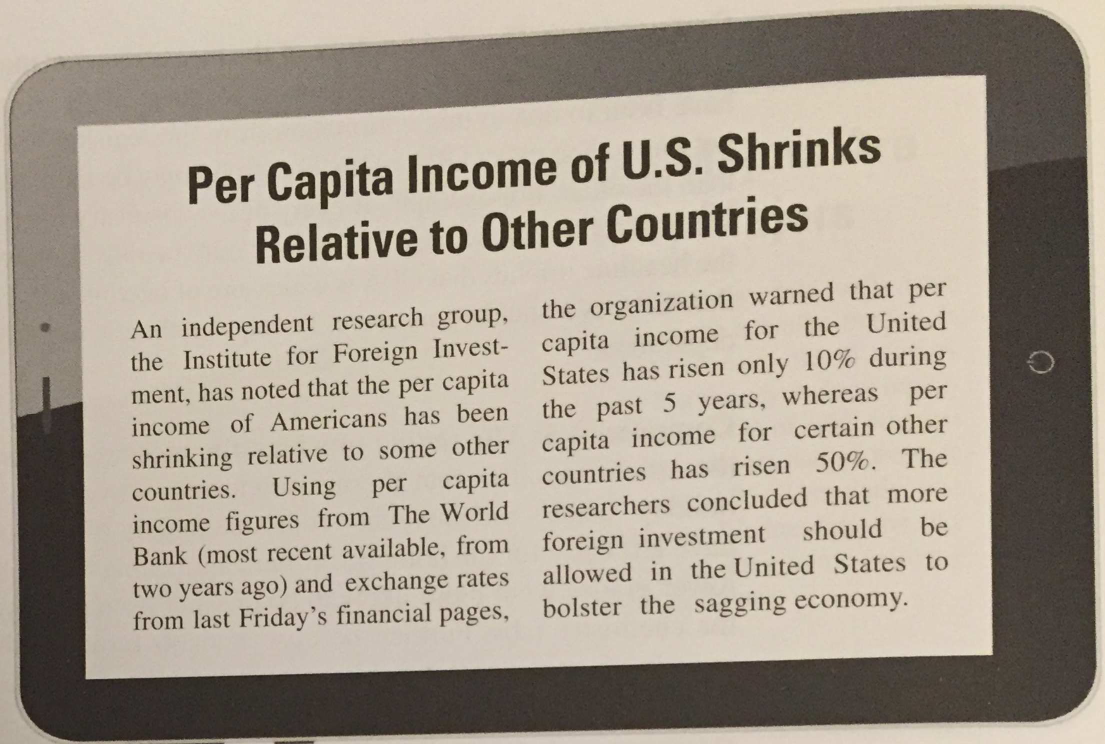
**Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest** The factor of interest is the student's major, and the two groups being compared are psychology majors and chemistry majors. This component considers whether the students who were interviewed for the study may differ in ways other than their choice of major. It is difficult to know what differences might exist without knowing more about the particular university. For example, because psychology is such a popular major, at some universities students are required to have a certain GPA before they are admitted to the major. A university with a separate premedical major might have the best of the science students enrolled in that major instead of chemistry. Those kinds of extraneous factors would be relevant to interpreting the results of the study.

**Component 7: The *extent* or *size* of any claimed effects or differences** The news report does present this information, by noting that the average GPAs for the two groups were 3.05 and 2.91. Additional useful information would be to know how many students were included in each of the averages given, what percentage of all students in each major were represented in the sample, and how much variation there was among GPAs within each of the two groups.

### *Hypothetical News Article 2: "Per Capita Income of U.S. Shrinks Relative to Other Countries"*

**Component 1: The *source* of the research and of the *funding*** We are told nothing except the name of the group that conducted the study, which should be fair warning. Being called "an independent research group" in the story does not mean that it is an unbiased research group. In fact, the last line of the story illustrates the probable motive for their research.

Hypothetical News Article 2



## Per Capita Income of U.S. Shrinks Relative to Other Countries

An independent research group, the Institute for Foreign Investment, has noted that the per capita income of Americans has been shrinking relative to some other countries. Using per capita income figures from The World Bank (most recent available, from two years ago) and exchange rates from last Friday's financial pages, the organization warned that per capita income for the United States has risen only 10% during the past 5 years, whereas per capita income for certain other countries has risen 50%. The researchers concluded that more foreign investment should be allowed in the United States to bolster the sagging economy.

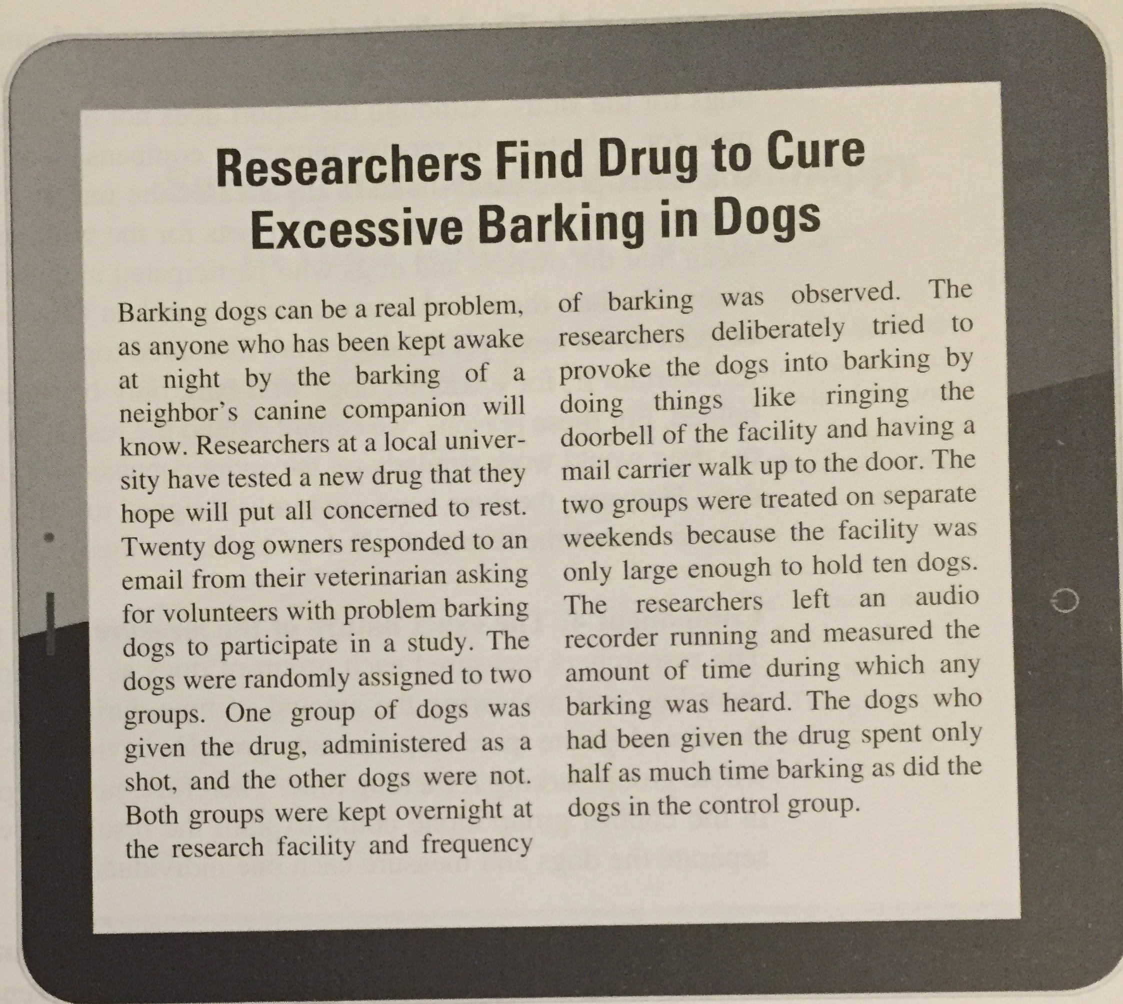
**Component 2: The *researchers* who had *contact* with the participants** This component is not relevant because there were no participants in the study.

**Component 3: The *individuals* or objects studied and how they were *selected*** The objects in this study were the countries used for comparison with the United States. We should have been told which countries were used, and why.

**Component 4: The exact nature of the *measurements* made or *questions* asked** This is the major problem with this study. First, as mentioned, we are not even told which countries were used for comparison. Second, current exchange rates but older per capita income figures were used. If the rate of inflation in a country had recently been very high, so that a large rise in per capita income did not reflect a concomitant rise in spending power, then we should not be surprised to see a large increase in per capita income in terms of actual dollars. In order to make a valid comparison, all figures would have to be adjusted to comparable measures of spending power, taking inflation into account. We will learn how to do that in Chapter 18.

**Components 5, 6, and 7: The *setting* in which the measurements were taken. *Differences* in the groups being compared, *in addition* to the factor of interest. The *extent* or *size* of any claimed effects or differences** These issues are not relevant here, except as they have already been discussed. For example, although the size of the difference between the United States and the other countries is reported, it is meaningless without an inflation adjustment.

Hypothetical News Article 3



## Researchers Find Drug to Cure Excessive Barking in Dogs

Barking dogs can be a real problem, as anyone who has been kept awake at night by the barking of a neighbor's canine companion will know. Researchers at a local university have tested a new drug that they hope will put all concerned to rest. Twenty dog owners responded to an email from their veterinarian asking for volunteers with problem barking dogs to participate in a study. The dogs were randomly assigned to two groups. One group of dogs was given the drug, administered as a shot, and the other dogs were not. Both groups were kept overnight at the research facility and frequency of barking was observed. The researchers deliberately tried to provoke the dogs into barking by doing things like ringing the doorbell of the facility and having a mail carrier walk up to the door. The two groups were treated on separate weekends because the facility was only large enough to hold ten dogs. The researchers left an audio recorder running and measured the amount of time during which any barking was heard. The dogs who had been given the drug spent only half as much time barking as did the dogs in the control group.

### *Hypothetical News Article 3: "Researchers Find Drug to Cure Excessive Barking in Dogs"*

**Component 1: The *source* of the research and of the *funding*** We are not told why this study was conducted. Presumably, it was because the researchers were interested in helping to solve a societal problem, but perhaps not. It is not uncommon for drug companies to fund research to test a new product or a new use for a current product. If that were the case, the researchers would have added incentive for the results to come out favorable to the drug. If everything were done correctly, such an incentive wouldn't be a major factor; however, when research is funded by a private source, that information should be provided when the results are announced.

**Component 2: The *researchers* who had *contact* with the participants** We are not given any information about who actually had contact with the dogs. One important question is whether the same handlers were used with both groups of dogs. If not, the difference in handlers could explain the results. Further, we are not told whether the dogs were primarily left alone or were attended most of the time. If researchers were present most of the time, their behavior toward the dogs could have had a major impact on the amount of barking.

**Component 3: The *individuals* or objects studied and how they were *selected*** We are told that the study used dogs whose owners volunteered them as problem dogs for the study. Although the report does not mention payment, it is quite common for volunteers to receive monetary compensation for their participation. The volunteers presumably lived in the area of the university. The dog owners had to be willing to be separated from their pets for the weekend. These and other factors mean that the owners and dogs who participated may differ from the general population. Further, the initial reasons for the problem behavior may vary from one participant to the next, yet the dogs were measured together. Therefore, there is no way to ascertain if, for example, dogs who bark only because they are lonely would be helped. For these reasons, we cannot extend the results of this study to conclude that the drug would work similarly on the entire population of dogs or even on all problem dogs. However, the dogs were randomly assigned to the two groups, so we should be able to extend the results to all dogs similar to those who participated.

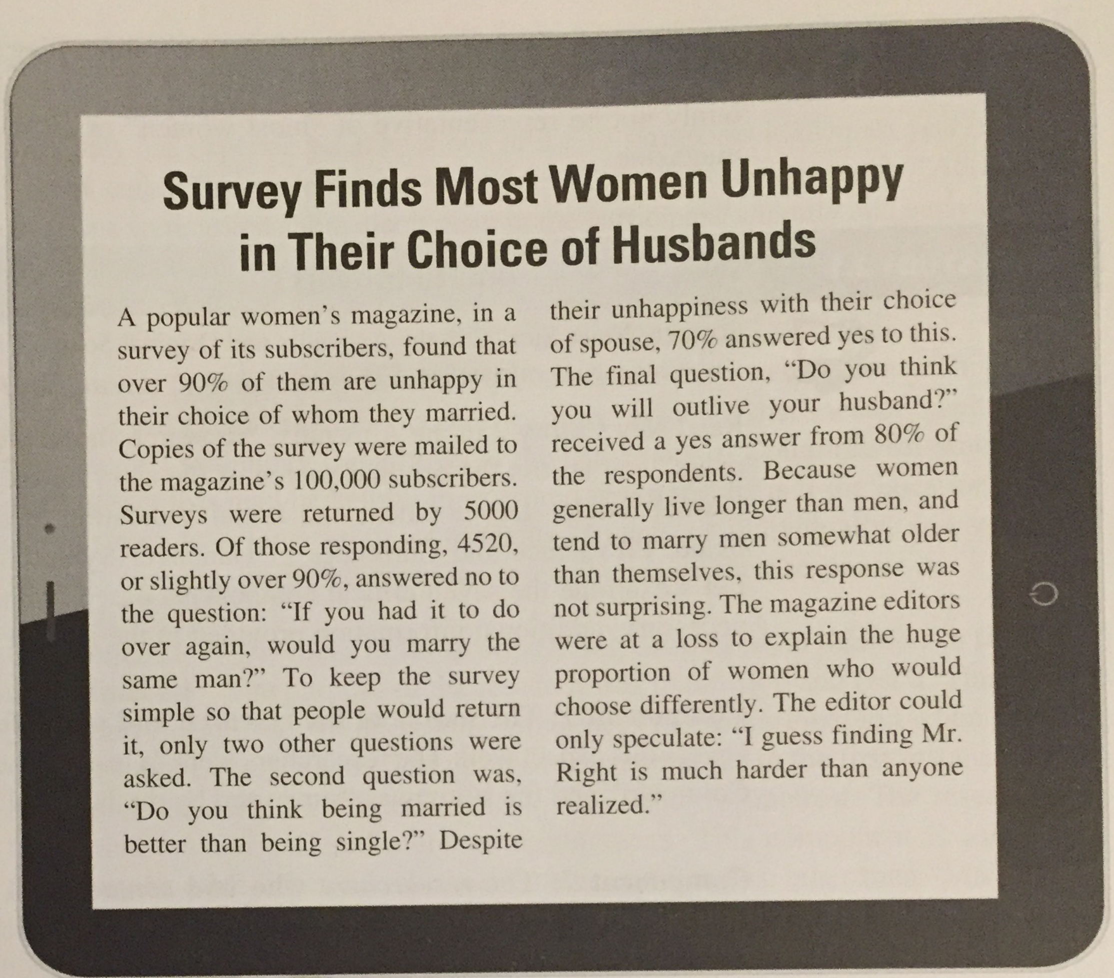
**Component 4: The exact nature of the *measurements* made or *questions* asked** The researchers measured each group of dogs as a group, by listening to an audio recording and measuring the amount of time during which there was any barking. Because dogs are quite responsive to group behavior, one barking dog could set the whole group barking for a long time. Therefore, just one particularly obnoxious dog in the control group alone could explain the results. It would have been better to separate the dogs and measure each one individually.

**Component 5: The *setting* in which the measurements were taken** The two groups of dogs (drug and no drug) were measured on separate weekends. This creates another problem. First, the researchers knew which group was which and may have unconsciously provoked the control group slightly more than the group receiving the drug. Further, conditions differed over the two weeks. Perhaps it was sunny one weekend and raining the next, or there were other subtle differences, such as more traffic one weekend than the next, small planes overhead, and so on. All of these could change the behavior of the dogs but might go unnoticed or unreported by the experimenters.

The measurements were also taken outside of the dogs' natural environments. The dogs in the experimental group in particular would have reason to be upset because they were first given a shot and then put together with nine other dogs in the research facility. It would have been better to put them back into their natural environment because that's where the problem barking was known to occur.

**Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest** The dogs were randomly assigned to the two groups (drug or no drug), which should have minimized overall differences in size, temperament, and so on for the dogs in the two groups. However, differences were induced between the two groups by the way the experiment was conducted. Recall that the groups were measured on different weekends—this could have created the difference in behavior. Also, the drug-treated dogs were given a shot to administer the drug, whereas the

Hypothetical News Article 4



**Survey Finds Most Women Unhappy in Their Choice of Husbands**

A popular women's magazine, in a survey of its subscribers, found that over 90% of them are unhappy in their choice of whom they married. Copies of the survey were mailed to the magazine's 100,000 subscribers. Surveys were returned by 5000 readers. Of those responding, 4520, or slightly over 90%, answered no to the question: "If you had it to do over again, would you marry the same man?" To keep the survey simple so that people would return it, only two other questions were asked. The second question was, "Do you think being married is better than being single?" Despite their unhappiness with their choice of spouse, 70% answered yes to this. The final question, "Do you think you will outlive your husband?" received a yes answer from 80% of the respondents. Because women generally live longer than men, and tend to marry men somewhat older than themselves, this response was not surprising. The magazine editors were at a loss to explain the huge proportion of women who would choose differently. The editor could only speculate: "I guess finding Mr. Right is much harder than anyone realized."

control group was given no shot. It could be that the very act of getting a shot made the drug group lethargic. A better design would have been to administer a placebo shot—that is, a shot with an inert substance—to the control group.

**Component 7: The *extent* or *size* of any claimed effects or differences** We are told only that the treated group barked half as much as the control group. We are not told how much time either group spent barking. If one group barked 8 hours a day but the other group only 4 hours a day, that would not be a satisfactory solution to the problem of barking dogs.

### *Hypothetical News Article 4: "Survey Finds Most Women Unhappy in Their Choice of Husbands"*

**Components 1 through 7** We don't even need to consider the details of this study because it contains a fatal flaw from the outset. The survey is an example of what is called a "volunteer sample" or a "self-selected sample." Of the 100,000 who received the survey, only 5% responded. The people who are most likely to respond to such a survey are those who have a strong emotional response to the question. In this case, it would be women who are unhappy with their current situation who would probably respond. Notice that the other two questions are more general and, therefore, not likely to arouse much emotion either way. Thus, it is the strong reaction to the first question that would drive people to respond. The results would certainly not be representative of "most women" or even of most subscribers to the magazine.

## CASE STUDY 2.1 Who Suffers from Hangovers?

SOURCE: News Story 2 in the Appendix and Original Source 2 on the companion website, www.cengage.com/stats/UttsSTS4e.

Read News Story 2 in the Appendix, "Research shows women harder hit by hangovers" and access the original source of the story on the companion website, the journal article "Development and Initial Validation of the Hangover Symptoms Scale: Prevalence and Correlates of Hangover Symptoms in College Students." Let's examine the seven critical components based on the news story and, where necessary, additional information provided in the journal article.

**Component 1: The *source* of the research and of the *funding*** The news story covers this aspect well. The researchers were "a team at the University of Missouri-Columbia" and the study was "supported by the National Institutes of Health."

**Component 2: The *researchers* who had *contact* with the participants** This aspect of the study is not clear from the news article, which simply mentions, "The researchers asked 1,230 drinking college students . . . ." However, the journal article says that the participants were enrolled in Introduction to Psychology courses and were asked to fill out a questionnaire. So it can be assumed that professors or research assistants in psychology had contact with the participants.

**Component 3: The *individuals* or objects studied and how they were *selected*** The news story describes the participants as "1,230 drinking college students, only 5 percent of whom were of legal drinking age." The journal article provides much more information, including the important fact that the participants were all enrolled in introductory psychology classes and were participating in the research to fulfill a requirement for the course. The reader must decide whether this group of participants is likely to be representative of all drinking college students, or some larger population, for severity of hangover symptoms. The journal article also provides information on the sex, ethnicity, and age of participants.

**Component 4: The exact nature of the *measurements* made or *questions* asked** The news story provides some detail about what was asked, noting that the participants were asked *"to describe how often they experienced any of 13 symptoms after drinking. The symptoms ranged from headaches and vomiting to feeling weak and unable to concentrate."* The journal article again provides much more detail, listing the 13 symptoms and explaining that participants were asked to indicate how often they were experienced on a 5-point scale (p. 1444 of the journal article). Further, participants were asked to provide a "hangover count" in which they noted how many times they had experienced at least one of the 13 symptoms in the past year, using a 5-point scale. This scale ranged from "never" to "52 times or more." Additional questions were asked about alcoholism in the participant's family and early experience with alcohol. Detailed information about all of these questions is included in the journal article.

**Component 5: The *setting* in which the measurements were taken** This information is not provided explicitly, but it can be assumed that measurements were taken in the Psychology Department at the University of Columbia-Missouri. One missing fact that may be helpful in interpreting the results is if the questions were administered to a large group of students at once, or individually, and whether students could be identified when the researchers read their responses.

**Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest** The purpose of the research was to develop and test a "Hangover Symptoms Scale" but two interesting differences in groups emerged when the researchers made comparisons. The groups being compared in the first instance were males and females; thus, Male/Female was the factor of interest. The researchers found that females suffered more from hangovers. This component is asking if there may be other differences between males and females, other than "Male" and "Female" that could help account for the difference. One possibility mentioned in the news article is body weight. Males tend to weigh more than females on average. An interesting question, not answered by the research, is if a group of males and females of the same weight, say 130 pounds, were to consume the same amount of alcohol, would the females suffer more hangover symptoms? The difference in weight between the two groups is in addition to the factor of interest, which is Male/Female. It may be the weight difference, and not the sex difference, that accounts for the difference in hangover severity.

The other comparison mentioned in the news article is between students who had alcohol-related problems, or whose biological parents had such problems, and students who did not have that history. In this case, the alcohol-related problems (of the student or parents) is the factor of interest. However, you can probably think of other differences in the two groups (those with problems and those without) that may help account for the difference in hangover severity between the two groups. For instance, students with a history of problems may not have as healthful diets in the past or present as students without such problems, and that may contribute to hangover severity. So the comparison of interest, between those with an alcohol problem in their background and those without, may be complicated by other differences in these two groups.

**Component 7: The *extent* or *size* of any claimed effects or differences** The news story does not report how much difference in hangover severity was found between men and women, or between those with and without a history of alcohol problems. Reading the journal article may explain why this is so—the article itself does not report a simple difference. In fact, simple comparisons don't yield much difference; for instance, 11% of men and 14% of women never experienced any hangover symptoms in the previous year. Differences only emerged when complicating factors such as amount of alcohol consumed were factored in. The researchers report, "After controlling for the frequency of drinking and getting drunk and for the typical quantity of alcohol consumed when drinking, women were significantly more likely than men to experience at least one of the hangover symptoms" (p. 1446). The article does not elaborate, such as explaining what would be the difference for a male and female who drank the same amount and equally often. ■

## 2.5 Planning Your Own Study: Defining the Components in Advance

Although you may never have to design your own survey or experiment, it will help you understand how difficult it can be if we illustrate the Seven Critical Components for a very simple hypothetical study you might want to conduct. Suppose you are interested in determining which of three local supermarkets has the best prices so you can decide where to shop. Because you obviously can't record and summarize the prices for all available items, you would have to use some sort of sample.

To obtain meaningful data, you would need to make many decisions. Some of the components need to be reworded because they are being answered in advance of the study, and obviously not all of the components are relevant for this simple example. However, by going through them for such a simple case, you can see how many ambiguities and decisions can arise when designing a study.

**Component 1: The *source* of the research and of the *funding*** Presumably you would be funding the study yourself, but before you start you need to decide why you are doing the study. Are you only interested in items you routinely buy, or are you interested in comparing the stores on the multitude of possible items?

**Component 2: The *researchers* who had *contact* with the participants** In this example, the question would be who is going to visit the stores and record the prices. Will you personally visit each store and record the prices? Will you send friends to two of the stores and visit the third yourself? If you use other people, you would need to train them so there would be no ambiguities. For example, if there are multiple brands and/or sizes of the same item, which price gets recorded? If there are organic and non-organic versions of an item, which price gets recorded?

**Component 3: The *individuals* or objects studied and how they were *selected*** In this case, the "objects studied" are items in the grocery store. The correct question is, "On what items should prices be recorded?" Do you want to use exactly the same items at all stores? What if one store offers its own brand but another only offers name brands? Do you want to choose a representative sampling of items you are likely to buy or choose from all possible items? Do you want to include nonfood items? How many items should you include? How should you choose which ones to select? If you are simply trying to minimize your own shopping bill, it is probably best to list the 20 or 30 items you buy most often. However, if you are interested in sharing your results with others, you might prefer to choose a representative sample of items from a long list of possibilities.

**Component 4: The exact nature of the *measurements* made or *questions* asked** You may think that the cost of an item in a supermarket is a well-defined measurement. But if a store is having a sale on a particular item on your list, should you use the sale price or the regular price? Should you use the price of the smallest possible size of the product? The largest? What if a store always has a sale on one brand or another of something, such as laundry soap, and you don't really care which brand you buy? Should you then record the price of the brand on sale that week? Should you record the prices listed on the shelves or actually purchase the items and see if the prices listed were accurate?

**Component 5: The *setting* in which the measurements were taken** When will you conduct the study? Supermarkets in university towns may offer sale prices on items typically bought by students at certain times of the year—for example, just after students have returned from vacation. Many stores also offer sale items related to certain holidays, such as ham or turkey just before Christmas or eggs just before Easter. Should you take that kind of timing into account?

**Component 6: *Differences* in the groups being compared, *in addition* to the factor of interest** The groups being compared are the groups of items from the three stores. There should be no additional differences related to the direct costs of the items. However, if you were conducting the study in order to minimize your shopping costs, you might ask if there are hidden costs for shopping at one store versus another. For example, do you always have to wait in line at one store and not at another, and should you therefore put a value on your time? Does one store make mistakes at check-out more often than another? Does one store charge for bags or give a discount for bringing your own bags? Does it cost more to drive to one store than another?

**Component 7: The *extent* or *size* of any claimed effects or differences** This component should enter into your decision about where to shop after you have finished the study. Even if you find that items in one store cost less than in another, the amount of the difference may not convince you to shop there. You would probably want to figure out approximately how much shopping in a particular store would save you over the course of a year. You can see why knowing the amount of a difference found in a study is an important component for using that study to make future decisions.

## CASE STUDY 2.2 Flawed Surveys in the Courtroom

SOURCES: http://openjurist.org/716/f2d/854, accessed May 11, 2013.
http://openjurist.org/615/f2d/252, accessed May 11, 2013.
Gastwirth (1988), pp. 517–520.

Companies sometimes sue other companies for alleged copying of names or designs that the first company claims are used to identify its products. In these kinds of cases, the court often wants to know whether it's likely that consumers are confusing the two companies based on the name or design. There is only one way to find that out—ask the consumers! But in the two instances presented here, the companies bringing the lawsuit made a crucial mistake—they did not ask a representative sample of people who were likely to buy the product.

In the first example, Brooks Shoe Manufacturing Company sued Suave Shoe Corporation for manufacturing shoes incorporating a "V" design used in Brooks's athletic shoes. Brooks claimed that the design was an unregistered trademark that people used to identify Brooks shoes. One of the roles of the court was to determine whether the design had "secondary meaning," which is a legal term indicating that "the primary significance of the term [or design] in the minds of the consuming public is not the product but the producer." (http://openjurist.org/596/f2d/111)

To show that the design had "secondary meaning" to buyers, Brooks conducted a survey of 121 spectators and participants at three track meets. Interviewers approached people and asked them a series of questions that included showing them a Brooks shoe with the name masked and asking them to identify it. Of those surveyed, 71% were able to identify it as a Brooks shoe, and 33% of those people said it was because they recognized the "V." When shown a Suave shoe, 39% of them thought it was a Brooks shoe, with 48% of those people saying it was because of the "V" design on the Suave shoe. Brooks Company argued that this was sufficient evidence that people might be confused and think Suave shoes were manufactured by Brooks.

Suave had a statistician as an expert witness who pointed out a number of flaws in the Brooks survey. Let's examine them using the Seven Critical Components as a guide. First, the survey was funded and conducted by Brooks, and the company's lawyer was instrumental in designing it. Second, the court determined that the interviewers who had contact with the respondents were inadequately trained in how to conduct an unbiased survey. Third, the individuals asked were not selected to be representative of the general public in the area (Baltimore/Washington, D.C.). For example, 78% had some college education, compared with 18.4% in Baltimore and 37.7% in Washington, D.C. Further, the settings for the interviews were track meets, where people were likely to be more familiar with athletic shoes. The questions asked were biased. For example, the exact wording used when a person was handed the shoes was: "I am going to hand you a shoe. Please tell me what brand you think it is." The way the question is framed would presumably lead respondents to think the shoe has a well-known brand name, which Brooks had at the time but Suave did not. Later in the questioning, respondents were asked, "How long have you known about Brooks Running Shoes?"

Suave introduced its own survey conducted on 404 respondents properly sampled from households in Baltimore, Maryland, and Greensboro, North Carolina, in which one or more individuals had purchased any type of athletic shoe during the previous year. Of those, only 2.7% recognized a Brooks shoe on the basis of the "V" design. The combination of the poor survey methods by Brooks and the proper survey by Suave convinced the court that the public did not make enough of an association between Brooks and the "V" design to allow Brooks to claim legal rights to the design. Brooks appealed the case to the Eleventh Circuit Court of Appeals, but the original decision was upheld, partly on the basis of the quality of the two surveys. (See http://openjurist.org/716/f2d/854.)

In the second example, Amstar Corporation, the makers of Domino sugar, sued Domino's Pizza, claiming that the use of the name was a trademark infringement. The original ruling was in favor of Amstar, and Domino's Pizza was told that they could no longer use the name, but the decision was overturned on appeal. Both companies had presented consumer surveys as part of their arguments. Although the first court thought that the Amstar survey was adequate and the Domino's Pizza survey was not, the appeals court bluntly stated, "Our own examination of the survey evidence convinces us that both surveys are substantially defective" (http://openjurist.org/615/f2d/252, note 38).

Who was surveyed? Amstar's survey consisted of female heads of households who were home during the day. They were shown a Domino's Pizza box and asked if they thought the company made any other products. If the answer was yes, they were asked "What products other than pizza do you think are made by the company that makes Domino's Pizza?" to which 71% responded "sugar." The court reasonably noted that surveying women who were home during the day was surely a biased sample, in favor of consumers who recognize Domino as the maker of sugar. But the appeals court noted an even more striking flaw: Eight of the 10 cities in which the surveys were conducted did not have a Domino's Pizza outlet, and in the other two cities, it had been open for less than 3 months. Therefore, participants in this survey were clearly more likely to know about Domino sugar than Domino's Pizza.

What about the survey conducted by Domino's Pizza? That one was conducted at Domino's Pizza outlets, and not surprisingly, found that almost no one was confusing the pizza company with the sugar maker. ■

## Thinking About Key Concepts

- Surveys that are open to anyone and rely completely on those who voluntarily respond are unlikely to represent any larger group. The only people likely to volunteer are those who feel strongly about the issues in the survey.
- People like to please others, and if they know the desired outcome in a survey or study, they are likely to comply and produce that response. Therefore, it's best if the person collecting the data does not know what the desired outcome is.