Statistics Success Stories and Cautionary Tales

The eight stories in this chapter are meant to bring life to the term ***statistics.*** After reading these stories, if you think the subject of statistics is lifeless or gruesome, check your pulse!

Let's face it. You're a busy person. Why should you spend your time learning about statistics? In this chapter, we give eight examples of situations in which statistics provided either enlightenment or misinformation. After reading these examples, we hope you will agree that learning about statistics may be interesting and useful.

Each of the stories in this chapter illustrates one or more concepts that will be developed throughout the book. These concepts are given as "the moral of the story" after a case is presented. Definitions of some terms used in the story also are provided following each case. By the time you read all of these stories, you already will have an overview of what statistics is all about.

1.1 What Is Statistics?

When you hear the word *statistics* you probably think of lifeless or gruesome numbers, such as the population of your state or the number of violent crimes committed in your city last year. The word *statistics,* however, actually is used to mean two different things. The better-known definition is that statistics are numbers measured for some purpose. A more complete definition, and the one that forms the substance of this book, is the following:

DEFINITION **Statistics** is a collection of procedures and principles for gathering data and analyzing information to help people make decisions when faced with uncertainty.

The stories in this chapter are meant to bring life to this definition. After reading them, if you think the subject of statistics is lifeless or gruesome, check your pulse!

1.2 Eight Statistical Stories with Morals

The best way to gain an understanding of some of the ideas and methods used in statistical studies is to see them in action. Each of the eight stories presented in this section includes interesting lessons about how to extract information from data. The methods and ideas will be expanded throughout the book, but these stories will give you an excellent overview of why it is useful to study statistics. To help you understand some basic statistical principles, each case study is accompanied by a "moral of the story" and by some definitions. All of the ideas and definitions will be discussed in greater detail in subsequent chapters.

CASE STUDY 1.1 Who Are Those Speedy Drivers?

A survey taken in a large statistics class at Penn State University contained the question "What's the fastest you have ever driven a car? ____ mph." The *data* provided by the 87 males and 102 females who responded are listed here.

Males: 110 109 90 140 105 150 120 110 110 90 115 95 145 140 110 105 85 95 100 115 124 95 100 125 140 85 120 115 105 125 102 85 120 110 120 115 94 125 80 85 140 120 92 130 125 110 90 110 110 95 95 110 105 80 100 110 130 105 105 120 90 100 105 100 120 100 100 80 100 120 105 60 125 120 100 115 95 110 101 80 112 120 110 115 125 55 90

Females: 80 75 83 80 100 100 90 75 95 85 90 85 90 90 120 85 100 120 75 85 80 70 85 110 85 75 105 95 75 70 90 70 82 85 100 90 75 90 110 80 80 110 110 95 75 130 95 110 110 80 90 105 90 110 75 100 90 110 85 90 80 80 85 50 80 100 80 80 80 95 100 90 100 95 80 80 50 88 90 90 85 70 90 30 85 85 87 85 90 85 75 90 102 80 100 95 110 80 95 90 80 90

From these numbers, can you tell which sex tends to have driven faster and by how much? Notice how difficult it is to make sense of the *data* when you are simply presented with a list. Even if the numbers had been presented in numerical order, it would be difficult to compare the two groups.

Your first lesson in statistics is how to formulate a simple summary of a long list of numbers. The **dotplot** shown in Figure 1.1 helps us see the pattern in the data. In the plot, each dot represents the response of an individual student. We can see that the men tend to claim a higher "fastest ever driven" speed than do the women.

The graph shows us a lot, and calculating some statistics that summarize the data will provide additional insight. There are a variety of ways to do so, but for this example, we examine a **five-number summary** of the data for males and females. The five numbers are the lowest value; the cut-off points for one-fourth, one-half, and three-fourths of the data; and the highest value. The three middle values of the summary (the cutoff points for one-fourth, one-half, and three-fourths of the data) are called the *lower quartile, median,* and *upper quartile,* respectively. Five-number summaries can be represented like this:

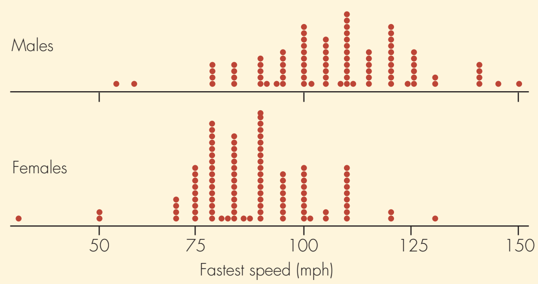

Figure 1.1 Responses to "What's the fastest you've ever driven?"

	Males (87 Students)			Females (102 Students)		
Median		110			89	
Quartiles	95		120	80		95
Extremes	55		150	30		130

Some interesting facts become immediately obvious from these summaries. By looking at the medians, you see that half of the men have driven 110 miles per hour or more, whereas the halfway point for the women is only 89 miles per hour. In fact, three-fourths of the men have driven 95 miles per hour or more, but only one-fourth of the women have done so. These facts were not at all obvious from the original lists of numbers.

Moral of the Story: *Simple summaries of data can tell an interesting story and are easier to digest than long lists.*

Definitions: Data is a plural word referring to numbers or nonnumerical labels (such as male/female) collected from a set of entities (people, cities, and so on). The **median** of a numerical list of data is the value in the middle when the numbers are put in order. For an even number of entities, the median is the average of the middle two values. The **lower quartile** and **upper quartile** are (roughly) the medians of the lower and upper halves of the data.

CASE STUDY 1.2 Safety in the Skies?

If you fly often, you may have been relieved to see the *New York Times* headline on October 1, 2007, proclaiming "Fatal airline crashes drop 65%" (Wald, 2007). And you may have been dismayed if you had seen an earlier headline in *USA Today* that read, "Planes get closer in midair as traffic control errors rise" (Levin, 1999). The details were even more disturbing: "Errors by air traffic controllers climbed from 746 in fiscal 1997 to 878 in fiscal 1998, an 18% increase."

So, are the risks of a fatal airline crash or an air traffic control error something that should be a major concern for airline passengers? Don't cancel your next vacation yet. A look at the statistics indicates that the news is actually pretty good! The low risk becomes obvious when we are told the *base rate* or *baseline risk* for these problems. According to the *New York Times* article, "the drop in the accident rate [from 1997 to 2007] will be about 65%, to one fatal accident in about 4.5 million departures, from 1 in nearly 2 million in 1997." And according to the 1999 *USA Today* story, "The errors per million flights handled by controllers climbed from 4.8 to 5.5." So the *rate* of fatal accidents changed from about 1 in 2 million departures in 1997 to 1 in 4.5 million departures in 2007, and the ominous rise in air traffic controller errors in 1998 still led to a very low rate of only 5.5 errors per million flights.

Fortunately, the rates for these problems were provided in both stories. This is not always the case in news reports of changes in rates or risk. For instance, an article may say that the risk of a certain type of cancer is doubled if you eat a certain unhealthful food. But what good is that information unless you know the actual risk? Doubling your chance of getting cancer from 1 in a million to 2 in a million is trivial, but doubling your chance from 1 in 50 to 2 in 50 is not.

Moral of the Story: *When you read about the change in the rate or risk of occurrence of something, make sure you also find out the base rate or baseline risk.*

Definitions: The **rate** at which something occurs is simply the number of times it occurs per number of opportunities for it to occur. In fiscal year 1998, the rate of air traffic controller errors was 5.5 per million flights. The **risk** of a bad outcome in the future can be estimated by using the past rate for that outcome, if it is assumed the future will be like the past. Based on recent data, the estimated risk of a fatal accident for any given flight is 1 in 4.5 million, which is 1/4,500,000 or about .00000022. The **base rate** or **baseline risk** is the rate or risk at a beginning time period or under specific conditions. For instance, the base rate of fatal airline crashes from which the 65% decrease for 2007 was calculated was about 1 crash per 2 million flights, for fiscal year 1997.

CASE STUDY 1.3 Did Anyone Ask Whom You've Been Dating?

"According to a new *USA Today*/Gallup Poll of teenagers across the country, 57% of teens who go out on dates say they've been out with someone of another race or ethnic group" (Peterson, 1997). That's over half of the dating teenagers, so of course it was natural for the headline in the *Sacramento Bee* to read, "Interracial dates common among today's teenagers." The article contained other information as well, such as "In most cases, parents aren't a major obstacle. Sixty-four percent of teens say their parents don't mind that they date interracially, or wouldn't mind if they did."

There are millions of teenagers in the United States whose experiences are being reflected in this story. How could the polltakers manage to ask so many teenagers these questions? The answer is that they didn't. The article states that "the results of the new poll of 602 teens, conducted Oct. 13–20, reflect the ubiquity of interracial dating today." They asked only 602 teens? Could such a small sample possibly tell us anything about the millions of teenagers in the United States? The answer is "yes" if those teens constituted a *random sample* from the *population* of interest.

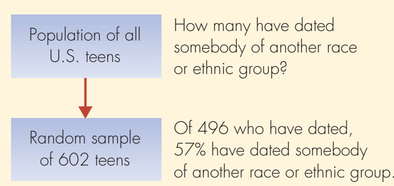

Figure 1.2 Population and sample for the survey.

The featured statistic of the article is that "57 percent of teens who go out on dates say they've been out with someone of another race or ethnic group." Only 496 of the 602 teens in the poll said that they date, so the 57% value is actually a percentage based on 496 responses. In other words, the pollsters were using information from only 496 teenagers to estimate something about all teenagers who date. Figure 1.2 illustrates this situation.

How accurate could this *sample survey* possibly be? The answer may surprise you. The results of this *poll* are accurate to within a *margin of error* of about 4.5%.

(continued)

As surprising as it may seem, the true percentage of all dating teens in the United States who date interracially is reasonably likely to be within 4.5% of the reported percentage that's based only on the 496 teens asked! We'll be conservative and round the 4.5% margin of error up to 5%. At the time the poll was taken, the percentage of all dating teenagers in the United States that would say they had dated someone of another race or ethnic group was likely to be in the range 57% $\pm$ 5%, or between 52% and 62%. (The symbol $\pm$ is read "plus and minus" and means that the value on the right should be added to and subtracted from the value on the left to create an interval.)

Polls and *sample surveys* are frequently used to assess public opinion and to estimate population characteristics such as the percent of teens who have dated interracially or the proportion of voters who plan to vote for a certain candidate. Many sophisticated methods have been developed that allow pollsters to gain the information they need from a very small number of individuals. The trick is to know how to select those individuals. In Chapter 5, we examine a number of other strategies that are used to ensure that sample surveys provide reliable information about populations.

Moral of the Story: *A representative sample of only a few thousand, or perhaps even a few hundred, can give reasonably accurate information about a population of many millions.*

Definitions: A **population** is a collection of all individuals about which information is desired. The "individuals" are usually people but could also be schools, cities, pet dogs, agricultural fields, and so on. A **random sample** is a subset of the population selected so that every individual has a specified probability of being part of the sample. In a **poll** or **sample survey,** the investigators gather opinions or other information from each individual included in the sample. The **margin of error** for a properly conducted survey is a number that is added to and subtracted from the sample information to produce an interval that is 95% certain to contain the true value for the population. In the most common types of sample surveys, the margin of error is approximately equal to 1 divided by the square root of the number of individuals in the sample.

Hence, a sample of 496 teenagers who have dated produces a margin of error of about $1/\sqrt{496} = .045$, or about 4.5%. In some polls the margin of error is called the **margin of sampling error** to distinguish it from other sources of errors and biases that can distort the results. The next Case Study illustrates a common source of bias that can occur in surveys, discussed more fully in Chapter 5.

CASE STUDY 1.4 Who Are Those Angry Women?

A well-conducted survey can be very informative, but a poorly conducted one can be a complete disaster. As an extreme example, Moore (1997, p. 11) reports that for her highly publicized book *Women and Love,* Shere Hite (1987) sent questionnaires to 100,000 women asking about love, sex, and relationships. Only 4.5% of the women responded, and Hite used those responses to write her book. As Moore notes, "The women who responded were fed up with men and eager to fight them. For example, 91% of those who were divorced said that they had initiated the divorce. The anger of women toward men became the theme of the book." Do you think that women who were angry with men would be likely to answer questions about love relationships in the same way as the general population of women?

The Hite sample exemplifies one of the most common problems with surveys: The sample data may not represent the population. Extensive *nonparticipation (nonresponse)* from a random sample, or the use of a *self-selected* (i.e., *all volunteer*) *sample,* will probably produce biased results. Those who voluntarily respond to surveys tend to care about the issue and therefore have stronger and different opinions than those who do not respond.

Moral of the Story: *An unrepresentative sample, even a large one, tells you almost nothing about the population.*

Definitions: Nonparticipation bias (also called **nonresponse bias**) can occur when many people who are selected for the sample either do not respond at all or do not respond to some of the key survey questions. This may occur even when an appropriate random sample is selected and contacted. The survey is then based on a nonrepresentative sample, usually those who feel strongly about the issues. Some surveys don't even attempt to contact a random sample but instead ask anyone who wishes to respond to do so. Magazines, television stations, and Internet websites routinely conduct this kind of poll, and those who respond are called a **self-selected sample** or a **volunteer sample.** In most cases, this kind of sample tells you nothing about the larger population at all; it tells you only about those who responded.

CASE STUDY 1.5 Does Prayer Lower Blood Pressure?

Read the original source on the companion website, http://www.cengage.com/statistics/Utts4e.

News headlines are notorious for making one of the most common mistakes in the interpretation of statistical studies: jumping to unwarranted conclusions. A headline in *USA Today* read, "Prayer can lower blood pressure" (Davis, 1998). The story that followed continued the possible fallacy it began by stating, "Attending religious services lowers blood pressure more than tuning into religious TV or radio, a new study says." The words "attending religious services lowers blood pressure" imply a direct cause-and-effect relationship. This is a strong statement, but it is not justified by the research project described in the article.

The article was based on an *observational study* conducted by the U.S. National Institutes of Health, which followed 2391 people aged 65 or older for 6 years. The article described one of the study's principal findings: "People who attended a religious service once a week and prayed or studied the Bible once a day were 40% less likely to have high blood pressure than those who don't go to church every week and prayed and studied the Bible less" (Davis, 1998). So the researchers did observe a relationship, but it's a mistake to think that this justifies the conclusion that prayer actually *causes* lower blood pressure.

When groups are compared in an observational study, the groups usually differ in many important ways that may contribute to the observed relationship. In this example, people who attended church and prayed regularly may have been less likely than the others to smoke or to drink alcohol. These could affect the results because smoking and alcohol use are both believed to affect blood pressure. The regular church attendees may have had a better social network, a factor that could lead to reduced stress, which in turn could reduce blood pressure. People who were generally somewhat ill may not have been as willing or able to go out to church. We're sure you can think of other possibilities for *confounding variables* that may have contributed to the observed relationship between prayer and lower blood pressure.

Moral of the Story: *Cause-and-effect conclusions cannot generally be made on the basis of an observational study.*

Definitions: An **observational study** is one in which participants are merely observed and measured. Comparisons based on observational studies are comparisons of naturally occurring groups. A **variable** is a characteristic that differs from one individual to the next. It may be numerical, such as blood pressure, or it may be categorical, such as whether or not someone attends church regularly. A **confounding variable** is a variable that is not the main concern of the study but may be partially responsible for the observed results.

(Source: International Journal of Psychiatry in Medicine *by Koenig, H.G., L.K. George, J.C. Hays, and D.B. Larson. [See p. 701 for complete credit.])*

CASE STUDY 1.6 Does Aspirin Reduce Heart Attack Rates?

Read the original source on the companion website, http://www.cengage.com/statistics/Utts4e.

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 purpose of the experiment was to determine whether or not taking aspirin reduces the risk of a heart attack. The physicians had been *randomly assigned* to one of the two *treatment* groups. One group took an ordinary aspirin tablet every other day, while the other group took a *placebo.* None of the physicians knew whether he was taking the actual aspirin or the placebo.

The results, shown in Table 1.1, support the conclusion that taking aspirin does indeed help to reduce the risk of having a heart attack. The rate of heart attacks in the group taking aspirin was only about half the rate of heart attacks in the placebo group. In the aspirin group, there were 9.42 heart attacks per 1000 participating doctors, while in the placebo group, there were 17.13 heart attacks per 1000 participants.

Table 1.1 The Effect of Aspirin on Heart Attacks

Treatment	Heart Attacks	Doctors in Group	Attacks per 1000 Doctors
Aspirin	104	11,037	9.42
Placebo	189	11,034	17.13

Because the men in this experiment were randomly assigned to the two conditions, other important risk factors such as age, amount of exercise, and dietary habits should have been similar for the two groups. The only important difference between the two groups should have been whether they took aspirin or a placebo. This makes it possible to conclude that taking aspirin actually *caused* the lower rate of heart attacks for that group. In a later chapter, you will learn how to determine that the difference seen in this sample is *statistically significant.* In other words, the observed sample difference probably reflects a true difference within the population.

To what population does the conclusion of this study apply? The participants were all male physicians, so the

(continued)

conclusion that aspirin reduces the risk of a heart attack may not hold for the general population of men. No women were included, so the conclusion may not apply to women at all. More recent evidence, however, has provided additional support for the benefit of aspirin in broader populations.

Moral of the Story: *Unlike with observational studies, cause-and-effect conclusions can generally be made on the basis of randomized experiments.*

Definitions: A **randomized experiment** is a study in which treatments are randomly assigned to participants. A **treatment** is a specific regimen or procedure assigned to participants by the experimenter. A **random assignment** is one in which each participant has a specified probability of being assigned to each treatment. A **placebo** is a pill or treatment designed to look just like the active treatment but with no active ingredients. A **statistically significant** relationship or difference is one that is large enough to be unlikely to have occurred in the sample if there was no relationship or difference in the population.

(Source: International Journal of Psychiatry in Medicine *by Koenig, H.G., L.K. George, J.C. Hays, and D.B. Larson. [See p. 701 for complete credit.])*

CASE STUDY 1.7 Does the Internet Increase Loneliness and Depression?

It was big news. Researchers at Carnegie Mellon University had found that "greater use of the Internet was associated with declines in participants' communication with family members in the household, declines in size of their social circle, and increases in their depression and loneliness" (Kraut et al., 1998, p. 1017). An article in the *New York Times* reporting on this study was titled "Sad, lonely world discovered in cyberspace" (Harmon, 1998). The study included 169 individuals in 73 households in Pittsburgh, Pennsylvania, who were given free computers and Internet service in 1995, when the Internet was still relatively new. The participants answered a series of questions at the beginning of the study and either 1 or 2 years later, measuring social contacts, stress, loneliness, and depression. The *New York Times* reported:

> In the first concentrated study of the social and psychological effects of Internet use at home, researchers at Carnegie Mellon University have found that people who spend even a few hours a week online have higher levels of depression and loneliness than they would if they used the computer network less frequently. . . . it raises troubling questions about the nature of "virtual" communication and the disembodied relationships that are often formed in cyberspace.
>
> *(Source: "Sad, Lonely World Discovered in Cyberspace," by A. Harmon, New York Times, August 30, 1998, p. A3. Reprinted with permission of the New York Times Company.)*

Given these dire reports, one would think that using the Internet for a few hours a week is devastating to one's mental health. But a closer look at the findings reveals that the changes were actually quite small, though statistically significant. Internet use averaged 2.43 hours per week for participants. The number of people in the participants' "local social network" decreased from an average of 23.94 people to an average of 22.90 people, hardly a noticeable loss. On a scale from 1 to 5, self-reported loneliness decreased from an average of 1.99 to 1.89 (lower scores indicate greater loneliness). And on a scale from 0 to 3, self-reported depression dropped from an average of .73 to an average of .62 (lower scores indicate higher depression).

The *New York Times* did report the magnitude of some of the changes, noting for instance that "one hour a week on the Internet was associated, on average, with an increase of .03, or 1% on the depression scale." But the attention the research received masked the fact that the impact of Internet use on depression, loneliness, and social contact was actually quite small, and thus may not have been of much practical significance.

As a follow-up to this study, in July 2001, *USA Today* (Elias, 2001) reported that in continued research, the bad effects had mostly disappeared. The article, titled "Web use not always a downer: Study disputes link to depression," began with the statement "Using the Internet at home doesn't make people more depressed and lonely after all." However, the article noted that the lead researcher, Robert Kraut of Carnegie Mellon University, believes that the earlier findings were correct but that "the Net has become a more social place since the study began in 1995." His explanation for the change in findings is that "either the Internet has changed, or people have learned to use it more constructively, or both." Research on this topic continues to develop. A study released in February 2010 (Morrison and Gore, 2010) identified 18 "Internet addicted" individuals out of 1319 study participants. They found that the Internet addicts scored in the "moderately-to-severely depressed range" on a test called the Beck Depression Inventory, while an equivalent group of non-addicts scored "firmly in the non-depressed range." As the authors point out, it is not clear whether Internet use causes depression, depression causes more Internet use, or some other factors lead to abnormal scores in both for some people.

Moral of the Story: *A statistically significant finding does not necessarily have practical significance or importance. When a study reports a statistically significant finding, find out the magnitude of the relationship or difference. A secondary moral to this story is that the implied direction of cause and effect may be wrong. In this case, it could be that people who were more lonely and depressed were more prone to using the Internet. And remember that, as the follow-up research makes clear, "truth" doesn't necessarily remain fixed across time. Any study should be viewed in the context of society at the time it was done.*

CASE STUDY 1.8 Did Your Mother's Breakfast Determine Your Sex?

Read the original source on the companion website, http://www.cengage.com/statistics/Utts4e.

You've probably heard that "you are what you eat," but did it ever occur to you that you might be who you are because of what your mother ate? A study published in 2008 by the British Royal Society seemed to find just that. The researchers reported that mothers who ate breakfast cereal prior to conception were more likely to have boys than mothers who did not (Mathews et al., 2008). But 9 months later, just enough time for the potential increased cereal sales to have produced a plethora of little baby boys, another study was published that dashed cold milk on the original claim (Young et al., 2009). The dispute was based on something statisticians call *multiple testing,* which can lead to erroneous findings of statistical significance. The authors of the original study had asked 740 women about 133 different foods they might have eaten just before getting pregnant. They found that 59% of the women who consumed breakfast cereal daily gave birth to a boy, compared to only 43% of the women who rarely or never ate cereal (http://www.cbsnews.com/stories/2008/04/22/health/webmd/main4036102.shtml). The result was highly statistically significant, but almost none of the other foods tested showed a statistically significant difference in the ratio of male to female births.

As previously discussed, statistical significance is how statisticians assess whether a difference found in a sample, in this case of 740 women, is large enough to conclude that the difference is likely to represent more than just chance. But sometimes what looks like a statistically significant difference is actually a *false positive*—a difference that looks like it wasn't due to chance when it really was. The more differences that are tested, the more likely it is that one of them will be a false positive. The criticism by Young et al. was based on this idea. When 133 food items that in fact do not affect the sex of a baby are all tested, it is likely that at least one of them will show up as a false positive, showing a big enough difference in the proportion of male to female births to be statistically significant when in fact the difference is due to chance.

The authors of the original study defended their work (Mathews et al., 2009). They noted that they only tested the individual food items after an initial test based on total pre-conception calorie consumption showed a difference in male and female births. They found that 56% of the mothers in the top third of calorie consumption had boys, compared with only 45% of the mothers in the bottom third of calorie consumption. That was one of only two initial tests they did; the other had to do with vitamin intake. With only two tests, it is unlikely that either of them would be a false positive. Unfortunately the media found the cereal connection to be the most interesting result in the study, and that's what received overwhelming publicity. The best way to resolve the debate, as in most areas of science, is to ask the same questions in a new study and see if the results are consistent. The authors of the original study have stated their intention to do that.

Moral of the Story: *When you read about a study that found a relationship or difference, try to find out how many different things were tested. The more tests that are done, the more likely it is that a statistically significant difference is a false positive that can be explained by chance. You should be especially wary if dozens of things are tested and only one or two of them are statistically significant.*

Definitions: Multiple testing or **multiple comparisons** in statistics refers to the fact that researchers often test many different hypotheses in the same study. This practice may result in statistically significant findings by mistake, called **false positive** results. Sometimes this practice is called **data snooping** because researchers snoop around in their data until they find something interesting to report.

1.3 The Common Elements in the Eight Stories

The eight stories were meant to bring life to our definition of statistics. Let's consider that definition again:

> **STATISTICS** is a collection of procedures and principles for gathering data and analyzing information to help people make decisions when faced with uncertainty.

Think back over the stories. In each of them, *data are used to make a judgment about a situation.* This common theme is what statistics is all about. The stories should also help you realize that you can be misled by the use of data, and learning to recognize how that happens is one of the themes of this book.

The Discovery of Knowledge

Each story illustrates part of the process of discovery of new knowledge, for which statistical methods can be very useful. The basic steps in this process are as follows:

1. *Asking the right question(s)*
2. *Collecting useful data,* which includes deciding how much is needed
3. *Summarizing and analyzing data,* with the goal of answering the questions
4. *Making decisions and generalizations* based on the observed data
5. *Turning the data and subsequent decisions into new knowledge*

We'll explore these five steps throughout the book, concluding with a chapter on "Turning Information into Wisdom." We're confident that your active participation in this exploration will benefit you in your everyday life and in your future professional career.

In a practical sense, almost all decisions in life are based on knowledge obtained by gathering and assimilating data. Sometimes the data are quantitative, as when an instructor must decide what grades to give on the basis of a collection of homework and exam scores. Sometimes the information is more qualitative and the process of assimilating it is informal, such as when you decide what you are going to wear to a party. In either case, the principles in this book will help you to understand how to be a better decision maker.

THOUGHT QUESTION 1.1 Think about a decision that you recently had to make. What "data" did you use to help you make the decision? Did you have as much information as you would have liked? If you could freely use them, how would you use the principles in this chapter to help you gain more useful information?*

***HINT:** As an example, how did you decide to live where you are living? What additional data, if any, would have been helpful?

IN SUMMARY Some Important Statistical Principles

The "moral of the story" items for the case studies presented in this chapter give a good overview of many of the important ideas covered in this book. Here is a summary:

- Simple summaries of data can tell an interesting story and are easier to digest than long lists.
- When you read about the change in the rate or risk of occurrence of something, make sure you also find out the base rate or baseline risk.
- A representative sample of only a few thousand, or perhaps even a few hundred, can give reasonably accurate information about a population of many millions.
- An unrepresentative sample, even a large one, tells you almost nothing about the population.
- Cause-and-effect conclusions cannot generally be made on the basis of an observational study.
- Unlike with observational studies, cause-and-effect conclusions can generally be made on the basis of randomized experiments.
- A statistically significant finding does not necessarily have practical significance or importance. When a study reports a statistically significant finding, find out the magnitude of the relationship or difference.
- When you read about a study that found a relationship or difference, try to find out how many different things were tested. The more tests that are done, the more likely it is that a statistically significant difference is a false positive that can be explained by chance.

Key Terms

Every term in this chapter is discussed more extensively in later chapters, so don't worry if you don't understand all of the terminology that has been introduced here. The following list indicates the page number(s) where the important terms in this chapter are introduced and defined.

Exercises

◆ Denotes that the dataset is available on the companion website, http://www.cengage.com/statistics/Utts4e, but is not required to solve the exercise.

Bold *exercises have answers in the back of the text.*

Note: Many of these exercises will be repeated in later chapters in which the relevant material is covered in more detail.

Skillbuilder Exercises

1.1 Refer to the data and five-number summaries given in Case Study 1.1. Give a numerical value for each of the following.

a. The fastest speed driven by anyone in the class.
b. The slowest of the "fastest speeds" driven by a male.
c. The speed for which one-fourth of the women had driven at that speed or faster.
d. The proportion of females who had driven 89 mph or faster.
e. The number of females who had driven 89 mph or faster.

1.2 A five-number summary for the heights in inches of the women who participated in the survey in Case Study 1.1 is as shown.

	Female Heights (inches)		
Median		65	
Quartiles	63.5		67.5
Extremes	59		71

a. What is the median height for these women?
b. What is the range of heights—that is, the difference in heights between the shortest woman and the tallest woman?
c. What is the interval of heights containing the shortest one-fourth of the women?
d. What is the interval of heights containing the middle one-half of the women?

1.3 In recent years, Vietnamese American women have had the highest rate of cervical cancer in the country. Suppose that among 200,000 Vietnamese American women, 86 developed cervical cancer in the past year.

a. Calculate the rate of cervical cancer for these women.
b. What is the estimated risk of developing cervical cancer for Vietnamese American women in the next year?
c. Explain the conceptual difference between the rate and the risk, in the context of this example.

1.4 The risk of getting lung cancer at some point in one's life for men who have never smoked is about 13 in 1000. The risk for men who smoke is just over 13 times the risk for nonsmokers. (Source: Villenueve and Lau, 1994)

a. What is the base rate for lung cancer in men over a lifetime?
b. What is the approximate lifetime risk of getting lung cancer for men who smoke?

1.5 Refer to Case Study 1.3, in which teens were asked about their dating behavior.

a. What population is represented by the random sample of 602 teens?
b. What population is represented by the 496 teens in the sample who had dated?

1.6 Using Case Study 1.6 as an example, explain the difference between a population and a sample.

1.7 A CBS News poll taken in December 2009, asked a random sample of 1048 adults in the United States, "In general, do you think the education most children are getting today in public schools is better, is about the same, or is worse than the education you received?" About 34% said "Better," 24% said "About the same," and 38% said "Worse." (The remaining 4% were unsure.)

a. What is the population for this survey?
b. What is the approximate margin of error for this survey?
c. Provide an interval that is 95% certain to cover the true percentage of U.S. adults in December 2009 who would have answered "Better" to this question if asked.

1.8 A telephone survey of 2000 Canadians conducted March 20–30, 2001, found that "Overall, about half of Canadians in the poll say the right number of immigrants are coming into the country and that immigration has a positive effect on Canadian communities. Only 16 percent view it as a negative impact while one third said it had no impact at all" (*The Ottawa Citizen,* August 17, 2001, p. A6.).

a. What is the population for this survey?
b. How many people were in the sample used for this survey?
c. What is the approximate margin of error for this survey?
d. Provide an interval of numbers that is 95% certain to cover the true percentage of Canadians who view immigration as having a negative impact.

1.9 In Case Study 1.3, the margin of error for the sample of 496 teenagers was about 4.5%. How many teenagers should be in the sample to produce an approximate margin of error of .05 or 5%?

1.10 About how many people would need to be in a random sample from a large population to produce an approximate margin of error of .30 or 30%?

1.11 A popular Sunday newspaper magazine often includes a yes-or-no survey question such as "Do you think there is too much violence on television?" or "Do you think parents should use physical discipline?" Readers are asked to mail their answers to the magazine, and the results are reported in a subsequent issue.

a. What is this type of sample called?
b. Do you think the results of these polls represent the opinions of all readers of the magazine? Explain.

1.12 A proposed study design is to leave 100 questionnaires by the checkout line in a student cafeteria. The questionnaire can be picked up by any student and returned to the cashier. Explain why this volunteer sample is a poor study design.

1.13 For each of the examples given here, decide whether the study was an observational study or a randomized experiment.

a. A group of students enrolled in an introductory statistics course were randomly assigned to take either a web-based course or a traditional lecture course. The two methods were compared by giving the same final examination in both courses.

b. A group of smokers and a group of nonsmokers who visited a particular clinic were asked to come in for a physical exam every 5 years for the rest of their lives to monitor and compare their health status.

c. CEOs of major corporations were compared with other employees of the corporations to see if the CEOs were more likely to have been the first child born in their families than were the other employees.

1.14 For each of the studies described, explain whether the study was an observational study or a randomized experiment.

a. A group of 100 students was randomly divided, with 50 assigned to receive vitamin C and the remaining 50 to receive a placebo, to determine whether or not vitamin C helps to prevent colds.

b. A random sample of patients who received a hip transplant operation at Stanford University Hospital during 2000 to 2010 will be followed for 10 years after their operation to determine the success (or failure) of the transplant.

c. Volunteers with high blood pressure were randomly divided into two groups. One group was taught to practice meditation and the other group was given a low-fat diet. After 8 weeks, reduction in blood pressure was compared for the two groups.

1.15 Read Case Study 1.5. Give an example of a confounding variable that might explain why elderly people who attended religious services might have lower blood pressure than those who did not. Do not use one of the variables already mentioned in the Case Study.

1.16 Suppose that an observational study showed that students who got at least 7 hours of sleep performed better on exams than students who got less than 7 hours of sleep. Which of the following are possible confounding variables, and which are not? Explain why in each case.

a. Number of courses the student took that term.

b. Weight of the student.

c. Number of hours the student spent partying in a typical week.

1.17 A randomized experiment was done in which overweight men were randomly assigned to either exercise or go on a diet for a year. At the end of the study there was a statistically significant difference in average weight loss for the two groups. What additional information would you need in order to determine if the difference in average weight loss had *practical* importance?

1.18 Explain the distinction between statistical significance and practical significance. Can the result of a study be statistically significant but not practically significant?

1.19 A (hypothetical) study of what people do in their spare time found that people born under the astrological sign of Aries were significantly more likely to be regular swimmers than people born under other signs. What additional information would you want to know to help you determine if this result is a false positive?

1.20 Explain what is meant by a "false positive" in the context of conclusions in statistical studies.

Chapter Exercises

1.21 Refer to Case Study 1.6, in which the relationship between aspirin and heart attack rates was examined. Using the results of this experiment, what do you think is the base rate of heart attacks for men like the ones in this study? Explain.

1.22 Students in a statistics class at Penn State were asked, "About how many minutes do you typically exercise in a week?" Responses from the *women* in the class were

60, 240, 0, 360, 450, 200, 100, 70, 240, 0, 60, 360, 180, 300, 0, 270

Responses from the *men* in the class were

180, 300, 60, 480, 0, 90, 300, 14, 600, 360, 120, 0, 240

a. Compare the women to the men using a dotplot. What does your plot show you about the difference between the men and the women?

b. For each sex, determine the median response.

c. Do you think there's a "significant" difference between the weekly amount that men and women exercise? Explain.

1.23 Refer to Exercise 1.22.

a. Create a five-number summary for the men's responses. Show how you found your answer.

b. Use your five-number summary to describe in words the exercise behavior of this group of male students.

1.24 Refer to Exercise 1.22.

a. Create a five-number summary for the women's responses. Show how you found your answer.

b. Use your five-number summary to describe in words the exercise behavior of this group of female students.

1.25 An article in the magazine *Science* (Service, 1994) discussed a study comparing the health of 6000 vegetarians and a similar number of their friends and relatives who were not vegetarians. The vegetarians had a 28% lower death rate from heart attacks and a 39% lower death rate from cancer, even after the researchers accounted for differences in smoking, weight, and social class. In other words, the reported percentages were the remaining differences after adjusting for differences in death rates due to those factors.

a. Is this an observational study or a randomized experiment? Explain.

b. On the basis of this information, can we conclude that a vegetarian diet causes lower death rates from heart attacks and cancer? Explain.

◆ Dataset available but not required **Bold** exercises answered in the back

c. Give an example of a potential confounding variable and explain what it means to say that it is a confounding variable.

1.26 Refer to Exercise 1.25, comparing vegetarians and nonvegetarians for two causes of death. Were base rates given for the two causes of death? If so, what were they? If not, explain what a base rate would be for this study.

1.27 An article in the *Sacramento Bee* (March 8, 1984, p. A1) reported on a study finding that "men who drank 500 ounces or more of beer a month (about 16 ounces a day) were three times more likely to develop cancer of the rectum than nondrinkers." In other words, the rate of cancer in the beer-drinking group was three times that of the non–beer drinkers in this study. What important numerical information is missing from this report?

1.28 Dr. Richard Hurt and his colleagues (Hurt et al., 1994) randomly assigned volunteers wanting to quit smoking to wear either a nicotine patch or a placebo patch to determine whether wearing a nicotine patch improves the chance of quitting. After 8 weeks of use, 46% of those wearing the nicotine patch but only 20% of those wearing the placebo patch had quit smoking.

a. Was this a randomized experiment or an observational study?
b. The difference in the percentage of participants who quit (20% versus 46%) was statistically significant. What conclusion can be made on the basis of this study?
c. Why was it advisable to assign some of the participants to wear a placebo patch?

1.29 Refer to the study in Exercise 1.28, in which there was a statistically significant difference in the percentage of smokers who quit using a nicotine patch and a placebo patch. Now read the two cautions in the "moral of the story" for Case Study 1.7. Discuss each of them in the context of this study.

1.30 Refer to the study in Exercises 1.28 and 1.29, comparing the percentage of smokers who quit using a nicotine patch and a placebo patch. Refer to the definition of statistics given on page 1, and explain how it applies to this study.

1.31 Case Study 1.6 reported that the use of aspirin reduces the risk of heart attack and that the relationship was found to be "statistically significant." Does either of the cautions in the "moral of the story" for Case Study 1.7 apply to this result? Explain.

1.32 A random sample of 1001 University of California faculty members taken in December 1995 was asked, "Do you favor or oppose using race, religion, sex, color, ethnicity, or national origin as a criterion for admission to the University of California?" (Roper Center, 1996). Fifty-two percent responded "favor."

a. What is the population for this survey?
b. What is the approximate margin of error for the survey?
c. Based on the results of the survey, could it be concluded that a majority (over 50%) of *all* University of California faculty members favor using these criteria? Explain.

1.33 The Roper Organization conducted a poll in 1992 (Roper, 1992) in which one of the questions asked was whether or not the respondent had ever seen a ghost. Of the 1525 people in the 18- to 29-year-old age group, 212 said "yes."

a. What is the approximate margin of error that accompanies this result?
b. What is the interval that is 95% certain to contain the actual proportion of people in this age group who have seen a ghost?

1.34 Refer to Exercise 1.33. What is the risk of someone in this age group having seen a ghost?

1.35 Refer to Exercise 1.33. The Roper Organization selected a random sample of adults in the United States for this poll. Suppose listeners to a late-night radio talk show were asked to call and report whether or not they had ever seen a ghost.

a. What is this type of sample called?
b. Do you think the proportion reporting that they had seen a ghost for the radio poll would be higher or lower than the proportion for the Roper poll? Explain.

1.36 The CNN website sometimes has a small box called "Quick vote" that contains a question about an interesting topic in the news that day. For example, one question in February 2010 asked "Should the U.S. military let gays and lesbians serve openly?" Visitors to the website are invited to click their response and to view the results. When the results are displayed they contain the message "This is not a scientific poll."

a. What type of sample is obtained in this Quick vote?
b. What do you think is meant by the message that "This is not a scientific poll?"

1.37 Explain what is meant by "data snooping."

1.38 A headline in a major newspaper read, "Breast-fed youth found to do better in school."

a. Do you think this statement was based on an observational study or a randomized experiment? Explain.
b. Given your answer in part (a), which of these two alternative headlines do you think would be preferable: "Breast-feeding leads to better school performance" or "Link found between breast-feeding and school performance"? Explain.

1.39 In this chapter, you learned that cause and effect can be concluded from randomized experiments but generally not from observational studies. Why don't researchers simply conduct all studies as randomized experiments rather than observational studies?

1.40 Why was the study described in Case Study 1.5 conducted as an observational study instead of an experiment?

1.41 Give an example of a question you would like to have answered, such as whether or not eating chocolate helps to prevent depression. Then explain how a randomized experiment or an observational study could be done to study this question.

◆ Dataset available but not required **Bold** exercises answered in the back

1.42 Suppose you were to read the following news story: "Researchers compared a new drug to a placebo for treating high blood pressure, and it seemed to work. But the researchers were concerned because they found that significantly more people got headaches when taking the new drug than when taking the placebo. Headaches were the only problem out of the 20 possible side effects the researchers tested."

a. Do you think the research used an observational study or a randomized experiment? Explain.

b. Do you think the researchers are justified in thinking the new drug would cause more headaches in the population than the placebo would? Explain.

1.43 Refer to Case Study 1.5. Explain what mistakes were made in the implementation of steps 4 and 5 of "The Discovery of Knowledge" when *USA Today* reported the results of this study.

1.44 Refer to Case Study 1.6. Go through the five steps listed under "The Discovery of Knowledge" in Section 1.3, and show how each step was addressed in this study.