Samples, Good and Bad

2

David McNew/Staff/Getty Images

CASE STUDY As discussed in Chapter 1, in February 2014 the Michigan online news site MLive ran the story, "Take our online poll: Should Michigan legalize marijuana." Of 9684 respondents, 7906 (81.64%) said Yes, 1190 (12.29%) said No, and 588 (6.07%) said Decriminalize but not legalize. These results would seem to indicate overwhelming support for legalizing marijuana in Michigan. However, the Pew Research Center conducted a poll on March 25–29, 2015, in which they asked, "Do you think the use of marijuana should be made legal or not?" The Pew Research Center reported that the poll consisted of telephone interviews with 1500 randomly selected adult Americans and that 53% of those surveyed favor the legal use of marijuana. This is a majority of those surveyed but not the overwhelming majority that MLive found. There is a large discrepancy in the findings of these two polls. This may be because the polls were conducted at different times, the populations sampled were different (Michigan versus all adult Americans), the MLive poll had a much larger sample than the Pew poll, the questions asked were not identical, or perhaps the data from one or both polls are simply bad. By the end of this chapter, you will have learned how to assess this last issue, namely, whether the data from these polls are good or bad.

How to sample badly

For many years in Rapides Parish, Louisiana, only one company had been allowed to provide ambulance service. In 1999, the local paper, the *Town Talk*, asked readers to call in to offer their opinion on whether the company should keep its monopoly. Call-in polls are generally automated: call one

telephone number to vote Yes and call another number to vote No. Telephone companies often charge callers to these numbers.

The *Town Talk* got 3763 calls, which suggests unusual interest in ambulance service. Investigation showed that 638 calls came from the ambulance company office or from the homes of its executives. Many more, no doubt, came from lower-level employees. "We've got employees who are concerned about this situation, their job stability, and their families and maybe called more than they should have," said a company vice president. Other sources said employees were told to, as they say in Chicago, "vote early and often."

As the *Town Talk* learned, it is easier to sample badly than to sample well. The paper relied on *voluntary response*, allowing people to call in rather than actively selecting its own sample. The result was *biased*—the sample was overweighted with people favoring the ambulance monopoly. Voluntary response samples attract people who feel strongly about the issue in question. These people, like the employees of the ambulance company, may not fairly represent the opinions of the entire population.

There are other ways to sample badly. Suppose that we sell your company several crates of oranges each week. You examine a sample of oranges from each crate to determine the quality of our oranges. It is easy to inspect a few oranges from the top of each crate, but these oranges may not be representative of the entire crate. Those on the bottom are more often damaged in shipment. If we were less than honest, we might make sure that the rotten oranges are packed on the bottom, with some good ones on top for you to inspect. If you sample from the top, your sample results are again *biased*—the sample oranges are systematically better than the population they are supposed to represent.

Biased sampling methods

The design of a statistical study is **biased** if it systematically favors certain outcomes.

Selection of whichever individuals are easiest to reach is called **convenience sampling**.

A **voluntary response sample** chooses itself by responding to a general appeal. Write-in or call-in opinion polls are examples of voluntary response samples.

Convenience samples and voluntary response samples are often biased.

EXAMPLE 1 Interviewing at the mall


Frank Chmura/Alamy

Squeezing the oranges on the top of the crate is one example of convenience sampling. Mall interviews are another. Manufacturers and advertising agencies often use interviews at shopping malls to gather information about the habits of consumers and the effectiveness of ads. A sample of mall shoppers is fast and cheap. But people contacted at shopping malls are not representative of the entire U.S. population. They are richer, for example, and more likely to be teenagers or retired. Moreover, the interviewers tend to select neat, safe-looking individuals from the stream of customers. Mall samples are biased: they systematically overrepresent some parts of the population (prosperous people, teenagers, and retired people) and underrepresent others. The opinions of such a convenience sample may be very different from those of the population as a whole.

EXAMPLE 2 Write-in opinion polls

Ann Landers once asked the readers of her advice column, "If you had it to do over again, would you have children?" She received nearly 10,000 responses, almost 70% saying, "NO!" Can it be true that 70% of parents regret having children? Not at all. This is a voluntary response sample. People who feel strongly about an issue, particularly people with strong negative feelings, are more likely to take the trouble to respond. Ann Landers's results are strongly biased—the percentage of parents who would not have children again is much higher in her sample than in the population of all parents.

On August 24, 2011, Abigail Van Buren (the niece of Ann Landers) revisited this question in her column "Dear Abby." A reader asked, "I'm wondering when the information was collected and what the results of that inquiry were, and if you asked the same question today, what the majority of your readers would answer."

Ms. Van Buren responded, "The results were considered shocking at the time because the majority of responders said they would NOT have children if they had it to do over again. I'm printing your question because it will be interesting to see if feelings have changed over the intervening years."

In October 2011, Ms. Van Buren wrote that this time the majority of respondents would have children again. That is encouraging, but this was, again, a write-in poll.

"Hey, Pops, what was that letter you sent off to Ann Landers yesterday?"

Write-in and call-in opinion polls are almost sure to lead to strong bias. In fact, only about 15% of the public have ever responded to a call-in poll, and these tend to be the same people who call radio talk shows. That's not a representative sample of the population as a whole.

Statistics in Your World

Big data "Big data" is a vague term, but it is often used by some to emphasize the sheer size of data sets that now exist. Big data are often "found data," not a random sample. Proponents for big data claim that because every single data point can now be captured, old statistical sampling techniques are obsolete. In essence, they say, we have data on the entire population. Is this true? Many statisticians disagree. They challenge the notion that we could ever have all the data. The claim that we have the entire population is often an assumption, rather than a fact about the data. Although sample sizes are enormous, proponents for big data often ignore the bias that accompanies nonrandom sampling.

Simple random samples

In a voluntary response sample, people choose whether to respond. In a convenience sample, the interviewer makes the choice. In both cases, personal choice produces bias. The statistician's remedy is to allow impersonal chance to choose the sample. A sample chosen by chance allows neither favoritism by the sampler nor self-selection by respondents. Choosing a sample by chance attacks bias by giving all individuals an equal chance to be chosen. Rich and poor, young and old, black and white, all have the same chance to be in the sample.

The simplest way to use chance to select a sample is to place names in a hat (the population) and draw out a handful (the sample). This is the idea of *simple random sampling*.

Simple random sample

A **simple random sample (SRS)** of size n consists of n individuals from the population chosen in such a way that every set of n individuals has an equal chance to be the sample actually selected.

An SRS not only gives each individual an equal chance to be chosen (thus avoiding bias in the choice), but also gives every possible sample an equal chance to be chosen. Drawing names from a hat does this. Write 100 names on identical slips of paper and mix them in a hat. This is a population. Now draw 10 slips, one after the other. This is an SRS, because any 10 slips have the same chance as any other 10.

Are these random digits really random?

Not a chance. The random digits in Table A were produced by a computer program. Computer programs do exactly what you tell them to do. Give the program the same input, and it will produce exactly the same "random" digits. Of course, clever people have devised computer programs that produce output that *looks* like random digits. These are called "pseudo-random numbers," and that's what Table A contains. Pseudo-random numbers work fine for statistical randomizing, but they have hidden nonrandom patterns that can mess up more refined uses.

NOW IT'S YOUR TURN

2.1 Sampling my class. There are 20 students in my class. They are listed on my class roster in alphabetical order. There are blank rows between the first five names on the list, the second five names on the list, the third five names on the list, and the last five names on the list. Thus, the list appears as four groups of five names, each separated by blank rows.

I want to take a simple random sample consisting of four of the students in my class. To do this, I select a single student from each group of five as follows. I write the numbers 1 to 5 on identical slips of paper. I mix the slips in a hat and draw one at random. I count this number of students down in the first group of five and select this student. For example, if the number selected is 3, I select the third student in the first group of five on my class roster. I replace the slip in the hat, again mix the slips, and draw a new number. The student this many down on the list in the second group is selected. I repeat this process for the remaining two groups. Every student in the class has a 1-in-5 chance of being selected when I come to his or her group. Thus, every student has the same chance of being selected. Is the sample a simple random sample? Explain.

Drawing names from a hat makes clear what it means to give each individual and each possible set of n individuals the same chance to be chosen. That's the idea of an SRS. Of course, drawing slips from a hat would be a

bit awkward for a sample of the country's 117 million households. In practice, real sample surveys use computer-generated *random digits* to choose samples. Many statistical software packages have random number generators that generate random digits. Some also allow one to choose an SRS.

EXAMPLE 3 How to choose an SRS using software

Joan's small accounting firm serves 30 business clients. Joan wants to interview a sample of five clients to find ways to improve client satisfaction. To avoid bias, she chooses an SRS of size 5.

Step 1: Label. Give each client a numerical label between 1 and 30. Here is the list of clients, with labels attached, using 1 to 30:

1	A-1 Plumbing	16	JL Appliances
2	Accent Printing	17	Johnson Commodities
3	Action Sport Shop	18	Keiser Construction
4	Anderson Construction	19	Liu's Chinese Restaurant
5	Bailey Trucking	20	MagicTan
6	Balloons Inc.	21	Peerless Machine
7	Bennett Hardware	22	Photo Arts
8	Best's Camera Shop	23	River City Antiques
9	Blue Print Specialties	24	Riverside Tavern
10	Central Tree Service	25	Rustic Boutique
11	Classic Flowers	26	Satellite Services
12	Computer Answers	27	Scotch Wash
13	Darlene's Dolls	28	Sewer's Center
14	Fleisch Realty	29	Tire Specialties
15	Hernandez Electronics	30	Von's Video Games

Step 2: Software. Use statistical software to generate a random integer between 1 and 30. Repeat this process, ignoring any values that were previously generated, until you obtain five different integers between 1 and 30. Joan used software and generated the numbers 18, 9, 10, 3, 9, and 1. The five different integers are 18, 9, 10, 3, and 1, so the sample is the clients Keiser Construction, Blue Print Specialties, Central Tree Service, Action Sport Shop, and A-1 Plumbing.

Generating a random integer with value between 1 and 30 is equivalent to writing the numbers 1 to 30 on identical slips of paper, placing them in a hat, mixing the slips well, and drawing one at random. The computer does the mixing and drawing.

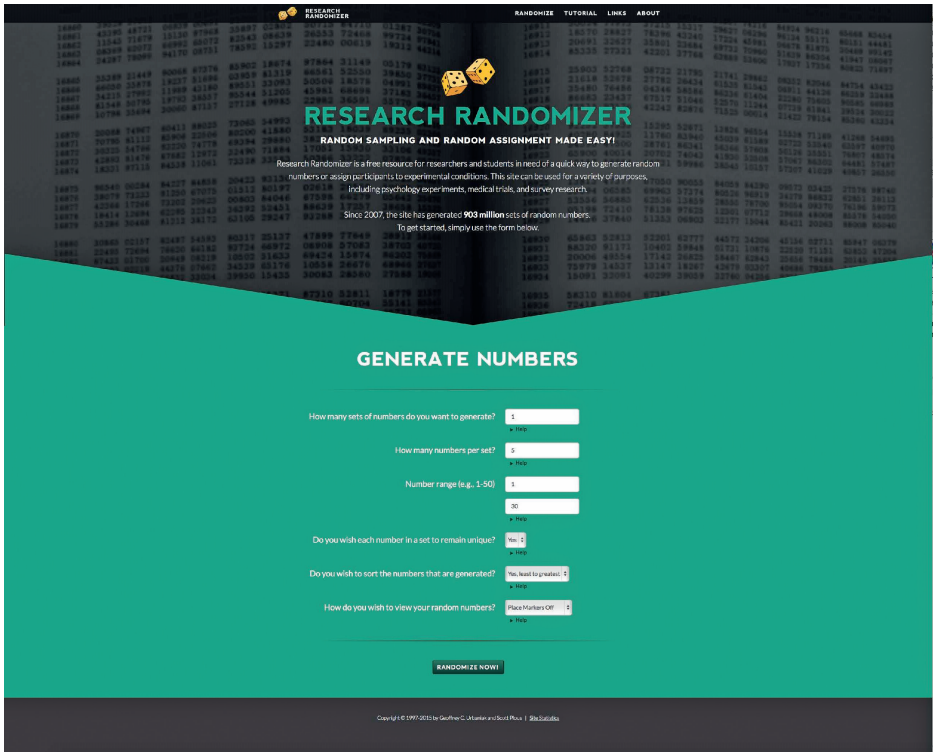

Figure 2.1 Using the Research Randomizer at www.randomizer.org. (Source: Randomizer .org. Copyright © 1997–2016 by Geoffrey C. Urbaniak and Scott Plous.)

Some statistical software may allow you to generate unique labels. A tool that is available on the Web is the Research Randomizer at www .randomizer.org. Fill in the boxes and click on the *Randomize Now* button. You can even ask the Randomizer to arrange your sample in order (see Figure 2.1).

If you don't use software, you can use a *table of random digits* to choose small samples by hand.

Random digits

A **table of random digits** is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these two properties:

1. Each entry in the table is equally likely to be any of the 10 digits 0 through 9.

2. The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

When is random too random? The streaming music service *Spotify* received complaints from listeners that the "shuffle" feature, used to play songs in a playlist in a random order, was not random enough. Listeners were hearing the same tracks two or three days in a row or the same artists back to back. The problem was not a lack of randomness, but too much randomness. *Spotify* developer Mattias Petter Johansson explained that "to humans, truly random does not feel random."

Table A at the back of the book is a table of random digits. You can think of Table A as the result of asking an assistant (or a computer) to mix the digits 0 to 9 in a hat, draw one, then replace the digit drawn, mix again, draw a second digit, and so on. The assistant's (or computer's) mixing and drawing save us the work of mixing and drawing when we need to randomize. Table A begins with the digits 19223950340575628713. To make the table easier to read, the digits appear in groups of five and in numbered rows. The groups and rows have no meaning—the table is just a long list of randomly chosen digits. Here's how to use the table to choose an SRS.

EXAMPLE 4 How to choose an SRS using a table of random digits

To repeat Example 3, we begin by assigning numerical labels to the 30 clients.

Step 1: Label. Give each client a numerical label, using as few digits as possible. Two digits are needed to label 30 clients, so we use labels

01, 02, 03, . . . , 28, 29, 30

It is also correct to use labels 00 to 29 or even another choice of 30 two-digit labels. Here is the list of clients, with labels attached, using 01 to 30:

01	A-1 Plumbing	16	JL Appliances
02	Accent Printing	17	Johnson Commodities
03	Action Sport Shop	18	Keiser Construction
04	Anderson Construction	19	Liu's Chinese Restaurant
05	Bailey Trucking	20	MagicTan
06	Balloons Inc.	21	Peerless Machine
07	Bennett Hardware	22	Photo Arts
08	Best's Camera Shop	23	River City Antiques
09	Blue Print Specialties	24	Riverside Tavern
10	Central Tree Service	25	Rustic Boutique
11	Classic Flowers	26	Satellite Services
12	Computer Answers	27	Scotch Wash
13	Darlene's Dolls	28	Sewer's Center
14	Fleisch Realty	29	Tire Specialties
15	Hernandez Electronics	30	Von's Video Games

Step 2: Table. Enter Table A anywhere and read two-digit groups. Suppose we enter at line 130, which is

69051 64817 87174 09517 84534 06489 87201 97245

The first 10 two-digit groups in this line are

69 05 16 48 17 87 17 40 95 17

Each two-digit group in Table A is equally likely to be any of the 100 possible groups, 00, 01, 02, …, 99. So two-digit groups choose two-digit labels at random. That's just what we want.

Joan used only labels 01 to 30, so we ignore all other two-digit groups. The first five labels between 01 and 30 that we encounter in the table choose our sample. Of the first 10 labels in line 130, we ignore five because they are too high (over 30). The others are 05, 16, 17, 17, and 17. The clients labeled 05, 16, and 17 go into the sample. Ignore the second and third 17s because that client is already in the sample. Now run your finger across line 130 (and continue to line 131 if needed) until five clients are chosen.

The sample is the clients labeled 05, 16, 17, 20, 19. These are Bailey Trucking, JL Appliances, Johnson Commodities, MagicTan, and Liu's Chinese Restaurant.

When using a table of random digits, as long as all labels have the same number of digits, all individuals will have the same chance to be chosen. Use the shortest possible labels: one digit for a population of up to 10 members, two digits for 11 to 100 members, three digits for 101 to 1000 members, and so on. As standard practice, we recommend that you begin with label 1 (or 01 or 001, as needed). You can read digits from Table A in any order—across a row, down a column, and so on—because the table has no order. As standard practice, we recommend reading across rows.

Using software or a table of random digits is much quicker than drawing names from a hat. As Examples 3 and 4 show, choosing an SRS has two steps.

Choose an SRS in two steps

Step 1: Label. Assign a numerical label to every individual in the population. Be sure that all labels have the same number of digits if you plan to use a table of random digits.

Step 2: Software or table. Use random digits to select labels at random.

NOW IT'S YOUR TURN

2.2 Evaluating teaching assistants. To assess how its teaching assistants are performing, the statistics department at a large university randomly selects three of its teaching assistants each week and sends a faculty member to visit their classes. The current list of 20 teaching assistants is given here. Use software, an online tool (for example, the Research Randomizer), or Table A at line 116 to choose three to be visited this week. Remember to begin by labeling the teaching assistants from 01 to 20.

Alexander	Park
Bean	Race
Book	Rodgers
Burch	Scarborough
Gogireddy	Siddiqi
Kunkel	Smith
Mann	Tang
Matthews	Twohy
Naqvi	Wilson
Ozanne	Zhang

Golfing at random

Random drawings give all the same chance to be chosen, so they offer a fair way to decide who gets a scarce good—like a round of golf. Lots of golfers want to play the famous Old Course at St. Andrews, Scotland. A few can reserve in advance. Most must hope that chance favors them in the daily random drawing for tee times. At the height of the summer season, only one in six wins the right to pay £170 (about $260) for a round.

Can you trust a sample?

The *Town Talk*, Ann Landers, and mall interviews produce samples. We can't trust results from these samples because they are chosen in ways that invite bias. We have more confidence in results from an SRS because it uses impersonal chance to avoid bias. The first question to ask of any sample is whether it was chosen at random. Opinion polls and other sample surveys carried out by people who know what they are doing use random sampling.

EXAMPLE 5 A Gallup Poll

A 2015 Gallup Poll on vaccinations asked the question, "From what you have heard or read, do you personally think certain vaccines do—or do not—cause autism in children, or are you unsure?" Gallup reported that the poll found that 52% of respondents were unsure. Is it actually

the case that a majority of Americans are unsure about whether certain vaccines cause autism in children? Ask first how Gallup selected its sample. Later in the article we read this: "Results for this Gallup poll are based on telephone interviews conducted Feb. 28–March 1, 2015, on the Gallup U.S. Daily survey, with a random sample of 1,015 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia." Gallup goes on to clarify that the sample included equal proportions of landline and cellular phone numbers selected using random-digit dialing.

This is a good start toward gaining our confidence. Gallup tells us what population it has in mind (people at least 18 years old living in the continental United States). We know that the sample from this population was of size 1015 and, most important, that it was chosen at random. There is more to consider in assessing a poll, and we will soon discuss this, but we have at least heard the comforting words "random sample."

STATISTICS IN SUMMARY

Chapter Specifics

- We select a **sample** in order to get information about some **population**.
- How can we choose a sample that fairly represents the population? **Convenience samples** and **voluntary response samples** are common but do not produce trustworthy data. These sampling methods are usually **biased**. That is, they systematically favor some parts of the population over others in choosing the sample.
- The deliberate use of chance in producing data is one of the big ideas of statistics. Random samples use chance to choose a sample, thus avoiding bias due to personal choice.
- The basic type of random sample is the **simple random sample**, which gives all samples of the same size the same chance to be the sample we actually choose.
- To choose an SRS by hand, use a **table of random digits** such as Table A in the back of the book, or use software.

The first step in reasoning from data to a conclusion is obtaining data. In Chapter 1, we discussed sample surveys as one way to collect data in an observational study. The method of selecting

the sample in a sample survey affects how well the sample represents the population. Biased sampling methods, such as convenience sampling and voluntary response samples, produce data that can be misleading, resulting in incorrect conclusions. Simple random sampling avoids bias and produces data that give us confidence that the first step in our argument is sound.

In the next chapter, we look more closely at what a simple random sample tells us about the population from which it is selected. And in Chapter 4, we discuss some of the problems faced by people who take surveys in the real world.

CASE STUDY EVALUATED To participate in the MLive poll described in the Case Study that opened the chapter, you had to choose to go online to the MLive website and click on one of the possible responses. Use what you have learned in this chapter to assess whether the data collected in such an online poll are good or bad. Your assessment should be written so that someone who knows no statistics will understand your reasoning.

LaunchPad Online Resources
macmillan learning

- The Snapshot Video *Sampling* discusses the importance of sampling, the basics of sampling, simple random sampling, and some practical issues in selecting random samples.
- The video technology manuals explain how to select an SRS using Excel, JMP, Minitab, and the TI 83/84.
- The Statistical Applet *Simple Random Sample* can be used to select a simple random sample when the number of labels is 144 or fewer.

CHECK THE BASICS

For Exercise 2.1, see page 25; for Exercise 2.2, see page 30.

2.3 Biased sampling methods? A method for selecting a sample is said to be biased if

(a) the race or gender of respondents is taken into account.

(b) it systematically favors certain outcomes.

(c) the race or gender of the person asking the questions is known by respondents.

(d) any of the above is true.

2.4 An online survey. You go to a website to access a news story. In order to access the story, you are asked to answer a brief survey. If you choose not to answer the survey, you can only access the article for a fee.

This method for obtaining a sample is an example of

(a) simple random sampling.

(b) random sampling, but not simple random sampling, because people visit the website at random.

(c) a convenience sample.

(d) a write-in opinion poll.

2.5 Simple random sample. I plan to take a sample of 10 students in my introductory statistics class. Which of the following is a simple random sample?

(a) I choose the 10 students sitting in the front row. Students select seats at random, so this would be a simple random sample.

(b) I choose the first 10 students who enter the classroom. Students arrive at random, so this would be a simple random sample.

(c) I write the names of all the students on similar slips of paper, put the slips of paper in a box, mix them well, and draw 10 slips from the box. The 10 names drawn are my sample.

(d) All of the above are simple random samples.

2.6 Random digits? Which of the following is true of a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 produced by software that generates random digits or listed in a table of random digits?

(a) Any number in the string of digits is equally likely to be any of the 10 digits 0 through 9.

(b) Knowledge of the values in one part of the string of digits gives no information about the numbers in any other part.

(c) Any pair of digits in the string of digits is equally likely to be any of the 100 digits 00 through 99.

(d) All of the above are true.

2.7 Choosing a simple random sample. Angela, Juan, Kevin, Lucinda, and Tanya are students in my capstone course. I wish to select a simple random sample of two of them to work on a project. I label Angela as 1, Juan as 2, Kevin as 3, Lucinda as 4, and Tanya as 5. I generate the following sequence of four random digits using statistical software: 1, 7, 8, 4. Based on these digits, my simple random sample is

(a) Juan and Tanya.

(b) Angela and Lucinda.

(c) any pair of students because all are equally likely.

(d) impossible to determine. I need to generate additional digits.

CHAPTER 2 EXERCISES

2.8 Letters to the editor. You work for a local newspaper that recently reported on a bill that would make it easier to create charter schools in the state. You report to the editor that 201 letters have been received on the issue, of which 171 oppose the legislation. "I'm surprised that most of our readers oppose the bill. I thought it would be quite popular," says the editor. Are you convinced that a majority of the readers oppose the bill? How would you explain the statistical issue to the editor?

2.9 Instant opinion. On January 30, 2015, the *Los Angeles Times* ran an online poll on its website and asked readers the question, "Which NFL team would you like to see come to L.A.?" Readers could click on one of three buttons (a picture of the Oakland Raiders logo, a picture of the San Diego Chargers logo, and a picture of the St. Louis Rams logo) to vote. In all, 11,908 (33%) selected the Oakland Raiders, 1939 (5%) selected the San Diego Chargers, and 22,138 (62%) selected the St. Louis Rams.

(a) What is the sample size for this poll?

(b) Explain why the poll may give unreliable information.

2.10 More instant opinion. In early 2015, the *Drudge Report* ran an online poll asking participants to vote for who they preferred for the Republican presidential candidate. The choices were Jeb Bush, Ben Carson, Chris Christie, Ted Cruz, Mike Huckabee, Sarah Palin, Rand Paul, Rick Santorum, Donald Trump, Rick Perry, Carly Florina, Marco Rubio, and Scott Walker. When the poll closed in February 2015, Scott Walker had received 199,095 votes (44%), Ted Cruz had received 58,844 votes (13%), Rand Paul had received 51,770 votes (12%), Ben Carson had received 37,945 votes (8%), Donald Trump had received 23,974 votes (5%), Sarah Palin had received 20,935 votes (5%), Jeb Bush had received 18,864 votes (4%), Marco Rubio had received 14,995 votes (3%), Rick Perry had received 6268 votes (1%), Mike Huckabee had received 6259 votes (1%), Chris Christie had received 5726 votes (1%), Rick Santorum had received 3038 votes (1%), and Carly Fiorina had received 2291 votes (1%).

(a) What is the sample size for this poll?

(b) The sample size for this poll is much larger than is typical for polls such as the Gallup Poll. Explain why the poll may give unreliable information, even with such a large sample size.

2.11 Ann Landers takes a sample. Advice columnist Ann Landers once asked her divorced readers whether they regretted their decision to divorce. She received approximately 30,000 responses, about 23,000 of which came from women. Nearly 75% said they were glad they divorced, and most of them said they wished they had done it sooner. Explain why this sample is certainly biased. What is the likely direction of the bias? That is, is 75% probably higher or lower than the truth about the population of all adults who have been divorced?

2.12 We don't like one-way streets. Highway planners decided to make a main street in West Lafayette, Indiana, a one-way street. The *Lafayette Journal and Courier* took a one-day poll by inviting readers to call a telephone number to record their comments. The next day, the paper reported:

Journal and Courier readers overwhelmingly prefer two-way traffic flow in West Lafayette's Village area to one-way streets. By nearly a 7–1 margin, callers on Wednesday complained about the one-way streets that have been in place since May. Of the 98

comments received, all but 14 said no to one-way.

(a) What population do you think the newspaper wants information about?

(b) Is the proportion of this population who favor one-way streets almost certainly larger or smaller than the proportion 14/98 in the sample? Why?

2.13 Design your own bad sample. Your college wants to gather student opinion about parking for students on campus. It isn't practical to contact all students.

(a) Give an example of a way to choose a sample of students that is poor practice because it depends on voluntary response.

(b) Give an example of a bad way to choose a sample that doesn't use voluntary response.

2.14 A call-in opinion poll. In 2005, the *San Francisco Bay Times* reported on a poll in New Zealand that found that New Zealanders opposed the nation's new gay-inclusive civil-unions law by a 3–1 ratio. This poll was a call-in poll that cost $1 to participate in. The *San Francisco Bay Times* article also reported that a scientific polling organization found that New Zealanders favor the law by a margin of 56.4% to 39.3%. Explain to someone who knows no statistics why the two polls can give such widely differing results and which poll is likely to be more reliable.

2.15 Call-in versus random sample polls. A national survey of TV network news viewers found that 48% said they would believe a phone-in poll of 300,000 persons rather than a random sample of 1000 persons. Of the viewers, 42% said they would believe the random sample poll. Explain to someone who knows no statistics why the opinions of only 1000 randomly chosen respondents are a better guide to what all people think than the opinions of 300,000 callers.

2.16 Choose an SRS. A firm wants to understand the attitudes of its minority managers toward its system for assessing management performance. Following is a list of all the firm's managers who are members of minority groups. Use software or Table A to choose five to be interviewed in detail about the performance appraisal system. If you use Table A, begin at line 132 to choose the five to be interviewed.

Baker	Liu
Berliner	Lu
Calder	MacEachern
Chkrebtii	Notz
Craigmile	Ozturk
Critchlow	Peruggia
Everson	Pratola
Hans	Rumsey
Herbei	Sivakoff
Holloman	Stasny
Hsu	Turkmen
Kaizar	VanZandt
Kubatko	Vu
Kurtek	Xu
Lee	Zhu
Lin	

2.17 Choose an SRS. Your class in ancient Ugaritic religion is poorly taught and the class members have decided to complain to the dean. The class

decides to choose six of its members at random to carry the complaint. The class list appears here. Choose an SRS of six using either software or the table of random digits, beginning at line 112.

Bartoszynski	Hsu	Scott
Blumenthal	Jacobs	Smith
Bower	Marker	Srivastava
Browne	Miller	Verducci
Cheng	Naber	Wang
Dean	Pearl	Whitney
Drake	Powers	Willke
Fligner	Rustagi	Wolfe
Goel	Santner	Woodall

2.18 An election day sample. You want to choose an SRS of 20 of Columbus, Ohio's, 636 voting precincts for special voting-fraud surveillance on election day.

(a) Explain clearly how you would label the 636 precincts. If you will use Table A to choose an SRS, be sure to explain how many digits make up each of your labels.

(b) Use either software or Table A to choose the SRS, and list the labels of the precincts you selected. If you use Table A, enter Table A at line 107.

2.19 Is this an SRS? A university has 30,000 undergraduate and 10,000 graduate students. A survey of student opinion concerning health care benefits for domestic partners of students selects 300 of the 30,000 undergraduate students at random and then separately selects 100 of the 10,000 graduate students at random. The 400 students chosen make up the sample.

(a) Explain why this sampling method gives each student an equal chance to be chosen.

(b) Nonetheless, this is not an SRS. Why not?

2.20 How much do students pay for rent? A university's housing and residence office wants to know how much students pay per month for rent in off-campus housing. The university does not have enough on-campus housing for students, and this information will be used in a brochure about student housing. The population contains 12,304 students who live in off-campus housing and have not yet graduated. The university will send a questionnaire to an SRS of 200 of these students, drawn from an alphabetized list.

(a) Describe how you would label the students in order to select the sample. If you will use Table A to choose an SRS, be sure to explain how many digits make up each of your labels.

(b) Use software or Table A to select the first five students in the sample. If you use Table A, begin at line 125.

2.21 Apartment living. You are planning a report on apartment living in a college town. You decide to select three apartment complexes at random for in-depth interviews with residents. Use software or Table A, starting at line 112, to select a simple random sample of three of the following apartment complexes.

Albany Commons	Canterbury Way
Apple Run	Chablis Villas
Bexley Court	Cherryblossom Way
Brooks Edge	

Dublin Plaza
English Village
Fairway Lakes
Forest Creek
Forest Park
Gaslight Village
Georgetowne
Golf Pointe
Hickory Mill
Highview Place
Indian Creek
Jefferson Commons
Kenbrook Village
Lawn Manor
Little Brook Place
Marble Cliff
Morse Glen
Oak Run
Old Nantucket
Parliament Ridge
Pheasant Run
Ravine Bluff
Rocky Creek
Scioto Commons
Stratford East
Walnut Knolls
Woodland Trace
York Terrace

2.22 How do random digits behave? Which of the following statements are true of a table of random digits, and which are false? Explain your answers.

(a) Each pair of digits has chance 1/100 of being 00.

(b) There are exactly four 4s in each row of 40 digits.

(c) The digits 12345 can never appear as a group because this pattern is not random.

2.23 Text in your vote. During the television broadcast on the Big Ten Network of a 2010 basketball game at Ohio State between Ohio State and Penn State, the announcers asked the following question: "Which player has meant the most to his team this year: Talor Battle of Penn State, Evan Turner of Ohio State, or other?" Viewers were asked to text in their votes. Later in the program, the results were announced. Evan Turner received 72% of the vote; Talor Battle, 26%; and "other," 2%. Explain why this opinion poll is almost certainly biased.

2.24 More randomization. Most sample surveys call residential telephone numbers at random. They do not, however, always ask their questions of the person who picks up the phone. Instead, they ask about the adults who live in the residence and choose one at random to be in the sample. Why is this a good idea?

2.25 Racial profiling and traffic stops. The Denver Police Department wants to know if Hispanic residents of Denver believe that the police use racial profiling when making traffic stops. A sociologist prepares several questions about the police. The police department chooses an SRS of 200 mailing addresses in predominantly Hispanic neighborhoods and sends a uniformed Hispanic police officer to each address to ask the questions of an adult living there.

(a) What are the population and the sample?

(b) Why are the results likely to be biased even though the sample is an SRS?

2.26 Random selection? Choosing at random is a "fair" way to decide who gets some scarce good, in the sense that everyone has the same chance to win. But random choice isn't always a good idea—sometimes we don't want to treat everyone the same because some people have a better claim. In each of the following situations, would you support choosing at random? Give your reasons in each case.

(a) The basketball arena has 4000 student seats, and 7000 students want tickets. Shall we choose 4000 of the 7000 at random?

(b) The list of people waiting for liver transplants is much larger than the number of available livers. Shall we let impersonal chance decide who gets a transplant?

(c) During the Vietnam War, young men were chosen for army service at random, by a "draft lottery." Is this the best way to decide who goes and who stays home?

EXPLORING THE WEB

Follow the QR code to access exercises.